

UNIVERSITE KASDI MERBAH OUARGLA
Faculté Des Nouvelles Technologique De L’Information et De Télécommunications
Département D’Electronique et De Télécommunications



Mémoire
MASTER PROFESSIONNEL

Domaine : Science et technique

Filière : Automatique

Spécialité : Automatique et système

Présenté par : BASSA Mohammed Ziad et BENLAMNOUAR Mohammed Arslane

Méthodes d'intelligence artificielle appliquées au diagnostic de défauts

Soutenu publiquement

Le : 20/06/2023

Devant le jury :

Mr. CHAKOUR Chouaib

Directeur de Mémoire

Mr. KARA Fouzia

Président

Mr. ROUABEH Boubaker

Examinateur

Année Universitaire : 2022 /2023

Dédicace

Je dédie ce travail,

*À mes très chers parents, pour tous leurs sacrifices, leur amour,
leur tendresse, leur soutien et leurs prières tout au long de mes
études, et qui m'ont permis*

De devenir ce que je suis aujourd'hui.

*A mes chers frères, pour leur appui et leur encouragement,
A toute ma famille*

*À tous mes amis qui m'ont accordé leur
Soutien dans les instants les plus difficiles.*

Remerciements

Tout d'abord, merci à Dieu de m'avoir permis de mener à bien ce travail dans les meilleures conditions, et de nous avoir donné force, courage et patience.

Je tiens à remercier mon encadrant Mr. Chakour Chouaib pour avoir accepté la direction scientifique et technique de ce travail et pour son soutien et ses conseils tout au long de cette période, ainsi que pour sa patience.

Merci à tous les professeurs de la Faculté des Nouvelles Technologique de l'Information et de Télécommunications qui nous ont suivi durant toute ces années de l'université.

Enfin, je remercie tous ceux qui ont contribué à l'élaboration de ce travail, de près ou de loin.

Résumé

Ce mémoire aborde le diagnostic de défauts en utilisant des techniques d'intelligence artificielle, notamment qui sont en relation avec l'algorithme des k-plus proches voisins (kNN) et les méthodes de clustering C-means et Fuzzy C-means. L'objectif principal de cette étude est d'améliorer l'efficacité de l'approche diagnostic pour la prédiction des éventuels défauts dans les systèmes industriels en exploitant la puissance de l'intelligence artificielle.

Tout d'abord, le diagnostic de défauts basé sur l'approche conventionnelle KNN est considéré comme un problème exclusivement de classification, ce qui limite sa capacité à détecter des types de défauts spécifiques. La méthode étudiée dans ce mémoire consiste à la ré-exploration de l'algorithme KNN dans le but d'effectuer un diagnostic prédictif des défauts. Ainsi, les problèmes liés à sa complexité de calcul et sa consommation de l'espace mémoire sont améliorés en le combinant avec les méthodes de clustering C-moyen et C-moyen flou.

Mots clés : diagnostic de défauts, Intelligence artificielle, k-plus proches voisins, méthodes de clustering, C-means, Fuzzy C-means .

Abstract

This thesis addresses fault diagnosis using artificial intelligence techniques, particularly those related to the k-nearest neighbors (kNN) algorithm, and the clustering methods C-means and Fuzzy C-means. The main objective of this study is to enhance the efficiency of the diagnostic approach for predicting potential faults in industrial systems by harnessing the power of artificial intelligence.

Firstly, fault diagnosis based on the conventional kNN approach is considered exclusively as a classification problem, which limits its ability to detect specific types of faults. The method considered in this thesis involves re-exploring the kNN algorithm to perform predictive fault diagnosis. Thus, issues related to computational complexity and memory consumption are improved by combining it with the C-means and Fuzzy C-means clustering methods.

Keywords: Fault diagnosis, Artificial intelligence, k-Nearest Neighbors, Clustering, C means, Fuzzy C means.

ملخص

تتناول هذه الأطروحة تشخيص العيوب باستخدام تقنيات الذكاء الاصطناعي، ولا سيما تلك المتعلقة بخوارزمية أقرب الجيران

C-mean و Fuzzy C-mean و Fuzzy C-mean.(kNN)

الهدف الرئيسي لهذه الدراسة هو تحسين كفاءة النهج التشخيصي للتنبؤ بالعيوب المحتملة في الأنظمة الصناعية من خلال

استغلال قوة الذكاء الاصطناعي

أولاً، يعتبر تشخيص الأخطاء بناءً على نهج kNN التقليدي مشكلة تصنيف حصرية مما يحد من قدرتها على اكتشاف أنواع معينة من الأخطاء. تمثل الطريقة المدروسة في هذه الرسالة في إعادة استكشاف خوارزمية kNN من أجل إجراء تشخيص تنبؤي للأخطاء. وبالتالي، يتم تحسين المشكلات المتعلقة بتعقيدها الحسابي واستهلاك مساحة الذاكرة من خلال دمجها مع طرق التجميع الغامض Fuzzy C-means و C—means

الكلمات المفتاحية : تشخيص الأخطاء، الذكاء الاصطناعي، أقرب الجيران، طرق التجميع، الوسائل (C)، الوسائل الضبابية

Table des figures

Figure 1.1 - Différents types de défauts d'un système physique [3]	6
Figure 1.2 – Une classification générale des méthodes de diagnostic [4]	7
Figure 1.3 – Principe de base du diagnostic des défaut à base de données	8
Figure 2.1 - Fonctionnement de l'algorithme kNN [12].....	14
Figure 2.2 - Résultats de classification du nouveau point	15
Figure 2.3 - les 8 proches voisins du nouveau point.....	16
Figure 2.4 - k voisins les plus proches dans le cercle	17
Figure 2.5 - KNN pour diagnostic de défaut.....	18
Figure 2.6 - l'indice de détection D^2kNN dans les deux cas.....	21
Figure 2.7 - La contribution de chaque variable	22
Figure 3.1– le principe de clustering.....	24
Figure 3.2 – C-means KNN pour diagnostic de défaut.....	28
Figure 3.3 – FCM-kNN pour diagnostic de défaut	29
Figure 3.4 - L'indice de détection $D^2C\text{-means } kNN$ dans les deux cas	31
Figure 3.5 - L'indice de détection $D^2 FCM\text{-}kNN$ dans les deux cas	31
Figure 3.6 - La contribution de chaque variable (C-means kNN)	32
Figure 3.7 - La contribution de chaque variable (FCM-kNN).....	33
Figure 4.1 – Le centre de traitement du champ de Haoud Berkaoui [21].....	35
Figure 4.2- Logigramme DCS de compresseur K103-A1	37
Figure 4.3 - L'évolution des mesures des 5 capteurs.....	38
Figure 4.4 – Evolution des indices de détection sans default avec déférents paramètres (k et C)	39
Figure 4.5 – Evolution de la variable X3 (Pression aspiration compresseur) sans et avec défaut	40
Figure 4.6 – Evolution des indices de détection avec default dans X3 avec déférents paramètres (k et C)	41
Figure 4.7 – Contributions des variables fautives du défaut X3 : kNN	42
Figure 4.8 – Contributions des variables fautives du défaut X3 :C-means et kNN	42
Figure 4.9 – Contributions des variables fautives du défaut X3 : FCM-kNN	43
Figure 4.10 – Evolution de la variable X2 (Capteur de débit du compresseur) sans et avec défaut	44
Figure 4.11 – Evolution des indices de détection avec default (X2)	44
Figure 4.12 – Contributions des variables fautives du défaut X2 : kNN	45
Figure 4.13 – Contributions des variables fautives du défaut X2 : C-means kNN.....	45
Figure 4.14 – Contributions des variables fautives du défaut X2 : FCM-kNN	46

Liste des tableaux

Tableau 2.1 – Les variables des 5 mesures des capteurs	20
Tableau 4.1 – Les variables collectée pour l’application.....	38
Tableau 4.2 – Le temps d’exécution des algorithmes kNN, C-means kNN et FCM-kNN.....	46
Tableau 4.3 – Le taux de fausses alarmes (TFA) des algorithmes kNN, C-means kNN et FCM-kNN.....	47

Table des matières

1 Diagnostic de défaut à base des techniques de l'intelligence artificielle.

1.1	Introduction	3
1.2	C'est quoi le diagnostic ?	4
1.2.1	C'est quoi le but de diagnostic ?	4
1.2.2	Que se passe-t-il lorsque le diagnostic n'est pas utilisé ?.....	4
1.3	Terminologie	5
1.4	Les différents types de défaut.....	6
1.5	Les méthodes de diagnostic.....	6
1.5.1	Diagnostic à base de modèle :.....	7
1.5.2	Diagnostic sans modèle :	7
1.6	Les approches à base de traitement de données	8
1.6.1	Les différentes approches à base de traitement de données (Data-Driven Approches) [20]	9
1.7	Conclusion.....	10

2 La méthode des kNN pour la détection et l'isolation de défaut

2.1	Introduction	12
2.2	L'algorithme des KNN.....	12
2.3	Principe de fonctionnement du kNN.....	12
2.3.1	Distance Euclidienne	13
2.3.2	Distance de Cityblock.....	13
2.3.3	Distance Cosinus.....	13
2.3.4	Distance corrélation	13
2.4	Avantages et inconvénients du kNN	14
2.4.1	Les avantages	14
2.4.2	Les inconvénients.....	15
2.5	Exemple de simulation (kNN pour la classification des données).....	15
2.6	L'algorithme des KNN appliqué au diagnostic.....	17
2.6.1	Détection de défauts basée sur kNN	18
2.7	KNN pour localisation de défaut.....	19
2.7.1	Localisation par calcul des contributions.....	19

2.8	Exemple de simulation (kNN pour FDI).....	20
2.9	Conclusion.....	22
3	Les méthodes C-means-KNN et Fuzzy C-means-KNN pour diagnostic de défauts	
3.1	Introduction	24
3.2	Principe de clustering	24
3.3	Les avantages de clustering.....	25
3.4	La méthode C-means.....	25
3.4.1	Les étapes de l'algorithme C-means	25
3.5	Fuzzy C-means clustering	26
3.5.1	Les étapes de l'algorithme Fuzzy C-means	26
3.6	La différence entre C means et fuzzy C means	27
3.7	La méthode C-means-KNN pour détection et isolation de défauts.....	27
3.8	KNN et Fuzzy C-means pour détection et isolation défaut.....	29
3.9	Exemple de simulation	30
3.10	Conclusion.....	33
4	Application (compresseur centrifuge)	
4.1	Introduction	35
4.2	Description de champs de Haoud Berkaoui	35
4.2.1	Description de la station de compression de Haoud Berkaoui	36
4.2.2	Description de Compresseur K103	36
4.3	Diagnostic de défauts du compresseur K103-A1	37
4.3.1	Résultats de simulation	37
4.4	Conclusion.....	47

Introduction générale

Face aux contraintes économiques et aux avancées technologiques rapides, les systèmes industriels deviennent de plus en plus complexes. La sécurité et la fiabilité des processus sont essentielles pour éviter toute perturbation causée par des anomalies ou des dysfonctionnements. Ces problèmes peuvent entraîner des conséquences financières désastreuses pour l'entreprise. Heureusement, les avancées technologiques ont permis le développement de nouvelles techniques de surveillance et de diagnostic des défaillances. Il existe une vaste littérature sur le sujet, comprenant notamment des méthodes basées sur le concept de redondance de l'information.

L'intelligence artificielle (IA) a connu des progrès considérables ces dernières années et a trouvé de nombreuses applications dans divers domaines, notamment dans le domaine du diagnostic de défauts. Le diagnostic de défauts revêt une grande importance dans de nombreux secteurs industriels, car il permet une détection et une localisation efficaces des défauts dans un système ou un processus. L'IA offre des outils et des techniques avancés pour améliorer les capacités de diagnostic et contribue ainsi à améliorer la performance et la fiabilité des opérations industrielles. L'utilisation de techniques d'intelligence artificielle, telles que l'algorithme des k-plus proches voisins (kNN), les méthodes de clustering C-means et Fuzzy C-means, offre de nouvelles opportunités pour améliorer la précision et l'efficacité du diagnostic de défauts. Ce mémoire se concentre sur l'application de ces trois approches d'intelligence artificielle pour le diagnostic prédictif du fonctionnement des processus industriels. Le document est divisé en quatre chapitres, où chaque chapitre explore en détail une approche spécifique et ses implications dans le domaine du diagnostic industriel.

Le premier chapitre présente une introduction générale sur le diagnostic de défauts et souligne son importance dans divers domaines industriels. Il met également en évidence les défis associés au diagnostic de défauts et présente l'utilisation de l'intelligence artificielle comme une solution prometteuse pour améliorer les performances de diagnostic.

Le deuxième chapitre se consacre à l'algorithme des k-plus proches voisins (kNN). Il offre une explication détaillée du fonctionnement de l'algorithme kNN et de son application dans le cadre du diagnostic prédictif des défauts. Des exemples concrets et des études de cas sont présentés pour démontrer l'efficacité de kNN dans le domaine du diagnostic de défauts.

Le troisième chapitre explore la combinaison des techniques de clustering C-means et Fuzzy C-means avec l'algorithme kNN pour améliorer le diagnostic de défauts. Il explique comment ces techniques de clustering peuvent être utilisées pour regrouper les données en clusters significatifs, ce qui peut aider à améliorer l'efficacité de l'approche diagnostic en termes de complexité de calcul et de consommation de l'espace mémoire.

Enfin, le quatrième chapitre présente une application des méthodes développées sur un processus réel. Il met en évidence l'implémentation pratique des algorithmes kNN, kNN avec C-means et kNN avec Fuzzy C-means pour le diagnostic de fonctionnement d'un compresseur centrifuge d'une station de compression de gaz à Haoud Berkaoui.

Chapitre 1

1 Diagnostic de défauts à base de l'intelligence artificielle.

Sommaire

1.1	Introduction	3
1.2	C'est quoi le diagnostic ?	4
1.2.1	C'est quoi le but de diagnostic ?	4
1.2.2	Que se passe-t-il lorsque le diagnostic n'est pas utilisé ?	4
1.3	Terminologie	5
1.4	Les différents types de défaut.....	6
1.5	Les méthodes de diagnostic.....	6
1.5.1	Diagnostic à base de modèle :.....	7
1.5.2	Diagnostic sans modèle :	7
1.6	Les approches à base de traitement de données	8
1.6.1	Les différentes approches à base de traitement de données (Data-Driven Approches) [20]	9
1.7	Conclusion.....	10

Chapitre1

Diagnostic de défauts à base de l'intelligence artificielle.

1.1 Introduction

De nos jours, le diagnostic de systèmes est une discipline complexe et en évolution constante, qui intègre une combinaison de techniques et de technologies, allant des méthodes statistiques et informatiques traditionnelles aux approches plus récentes d'intelligence artificielle et d'apprentissage automatique. Grâce à ces outils et techniques, les ingénieurs et les techniciens peuvent rapidement identifier les anomalies dans le fonctionnement des systèmes et proposer des solutions pour les résoudre, garantissant ainsi la fiabilité, la sécurité et l'efficacité des systèmes.

1.2 C'est quoi le diagnostic ?

Comme dans son acceptation médicale, le diagnostic des systèmes automatiques consiste à remonter des symptômes perçus vers les causes. Il est également défini comme l'ensemble des actions mises en œuvre en vue de détecter, localiser et identifier tout phénomène anormal sur un système à partir de la connaissance disponible sur ce dernier. La connaissance disponible inclut, non seulement, les sorties du contrôleur, les mesures fournies par les capteurs mais aussi, le plus souvent, des informations sur la structure et/ou un modèle du système [1].

1.2.1 C'est quoi le but de diagnostic ?

Le but des diagnostics est de fournir une évaluation approfondie et précise de l'état de santé ou de fonctionnement d'un système, d'un équipement, d'une personne ou d'un environnement. Les diagnostics sont généralement utilisés pour identifier les problèmes, les dysfonctionnements ou les défauts, afin de permettre aux personnes concernées de prendre des mesures pour les corriger.

Les diagnostics peuvent être effectués dans de nombreux domaines différents, tels que la médecine, l'ingénierie, l'environnement, l'informatique, la psychologie, la finance et d'autres secteurs. Les résultats des diagnostics peuvent aider les professionnels à prendre des décisions éclairées sur les mesures à prendre pour résoudre les problèmes identifiés.

1.2.2 Que se passe-t-il lorsque le diagnostic n'est pas utilisé ?

Lorsque le diagnostic n'est pas utilisé, cela peut entraîner des conséquences négatives importantes. Les problèmes peuvent passer inaperçus : Si un diagnostic n'est pas effectué, il peut être difficile de détecter les problèmes qui existent. Les problèmes peuvent ainsi passer inaperçus et ne pas être traités, ce qui peut entraîner des conséquences graves et mettre en péril la sécurité des personnes,

Chapitre1 **Diagnostic de défauts à base de l'intelligence artificielle.**

de l'environnement et des biens. Voici quelques exemples de ce qui peut arriver si le diagnostic n'est pas utilisé :

Les coûts peuvent augmenter : L'absence de diagnostic peut entraîner des coûts plus élevés à long terme. Par exemple, si un équipement n'est pas diagnostiqué et que des réparations mineures sont négligées, cela peut entraîner des dommages plus importants qui nécessitent des réparations plus coûteuses.

La prise de décision peut être inexacte : Les diagnostics aident à identifier les problèmes, ce qui permet une prise de décision plus éclairée. En l'absence de diagnostic, les décisions peuvent être basées sur des hypothèses ou des spéculations, ce qui peut entraîner des erreurs de jugement.

1.3 Terminologie

Il existe une variété de termes utilisés dans la littérature. Nous utiliserons ceux trouvés dans le rapport de la référence [2] :

Défaut : est un écart entre la caractéristique observée et la caractéristique de référence. Un défaut peut conduire à une défaillance du système tout comme il peut exister sans altérer son fonctionnement général.

Défaillance : est une conséquence possible d'une altération importante des caractéristiques. Elle correspond à l'incapacité du système à réaliser la fonction souhaitée. Elle peut être passagère ou permanente.

Panne : est l'interruption permanente du système à réaliser la fonction souhaitée.

Résidus : Indicateurs d'anomalie. Exprient la différence entre les valeurs théoriques et celles mesurées. Présentent une valeur nulle lors du fonctionnement normal.

Monitoring : c'est la surveillance en temps réel du système, elle permet la détection précoce des défauts.

Supervision : c'est la surveillance du système et la prise de décisions rapide dans le cas de l'existence d'un défaut.

Détection de défaut : il s'agit de détecter rapidement les anomalies et de prendre une décision rapide quant à l'existence ou non d'un défaut dans le système.

Isolation des défauts : il s'agit de déterminer le composant qui cause le problème, par exemple, quel capteur ou actionneur est devenu défectueux.

Identification du défaut : il s'agit d'estimer la taille, le type ou la nature du défaut.

Chapitre1 Diagnostic de défauts à base de l'intelligence artificielle.

Diagnostic : il suit la détection de défauts et inclut l'isolation et l'identification.

1.4 Les différents types de défaut

Dans la littérature, trois types de défauts sont distingués :

- **Défauts sur les capteurs :**

Les capteurs représentent l'interface de sortie d'un système, et transmettent des informations sur le comportement de ce dernier. Par conséquent, les défauts du capteur peuvent entraîner une dégradation des performances du processus et compromettre l'intégrité des données ce qui revient à fausser la prise de décisions.

Les capteurs sont utilisés soit pour mesurer directement les états du système, soit pour générer des estimations d'état pour la loi de contrôle. Ainsi, la présence de défauts dans les capteurs peut détériorer les estimations d'état et par conséquent entraîner un diagnostic inexact.

- **Défaut sur les actionneurs :**

Dans de nombreux systèmes les signaux de commande du contrôleur ne peuvent pas être directement appliqués au système. Des actionneurs sont nécessaires pour les transformer en signaux d'actionnement appropriés. Cependant, l'apparition d'anomalies dans cet organe peut amener à une perte de contrôle totale.

- **Défauts sur le processus :**

Le défaut sur le processus survient généralement lorsque certaines conditions changent dans le système, rendant invalide l'équation dynamique nominale du système.

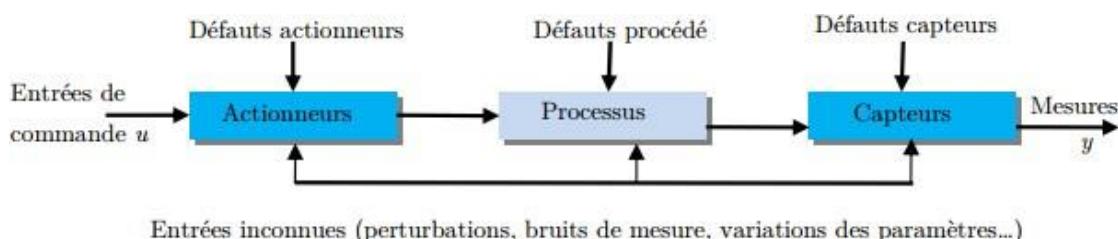


Figure 1.1 - Différents types de défauts d'un système physique [3]

1.5 Les méthodes de diagnostic

Les méthodes utilisées dans le cadre industriel pour le diagnostic des pannes sont très diverses. Une classification non exhaustive de ces méthodes est présentée dans la figure 1.2.

Chapitre1 **Diagnostic de défauts à base de l'intelligence artificielle.**

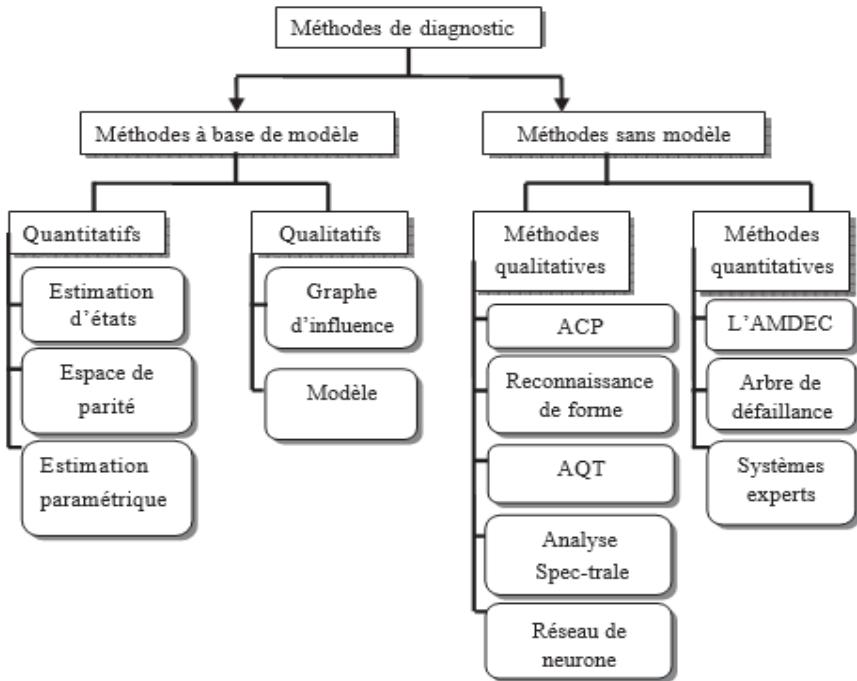


Figure 1.2 – Une classification générale des méthodes de diagnostic [4]

1.5.1 Diagnostic à base de modèle :

Les approches basées sur des modèles effectuent un diagnostic de panne basé sur la redondance analytique, où l'accord entre les valeurs mesurées et le comportement prédict par un modèle est vérifié. Les résidus sont évalués pour réaliser la détection, l'isolation et l'identification des défauts [5].

En règle générale, les approches basées sur un modèle peuvent être plus efficaces si un modèle correct et précis est construit. Cependant, un modèle mathématique parfaitement précis et complet d'un système physique n'est jamais disponible ou difficile à mettre en œuvre.

1.5.2 Diagnostic sans modèle :

Le diagnostic sans modèle est une approche qui consiste à utiliser des techniques statistiques et des algorithmes de Machine Learning pour diagnostiquer les systèmes sans avoir besoin d'un modèle préalable du système.

Chapitre1 **Diagnostic de défauts à base de l'intelligence artificielle.**

Cette approche peut être particulièrement utile lorsque les modèles de systèmes sont difficiles à construire. Dans le diagnostic sans modèle on a deux classes : la première classe est les méthodes qualitatives qui contient plusieurs approches comme l'ACP, Reconnaissance de forme, l'Analyse Spectrale et les Réseau de neurone) et la deuxième classe qui est les méthodes quantitatives qui contient des différentes techniques comme les Arbres de défaillance et les Systèmes experts).

1.6 Les approches à base de traitement de données

Pour les systèmes où l'obtention d'un modèle comportemental est difficile, des méthodes de détection et de diagnostic sans modèle de référence sont utilisées. Ils sont basés sur l'utilisation efficace à la fois des données générées par le système et des données générées par l'être humain (à partir de l'intelligence artificielle, de la reconnaissance des formes, des systèmes experts ou des systèmes basés sur la connaissance). Le traitement de ces données aboutit à la création d'un modèle à la manière d'une boîte noire. C'est l'approche qui est basée sur le traitement des données.

La croissance exponentielle des données recueillies sur le fonctionnement des procédés a été un moteur majeur du développement des méthodes de traitement des données. Ces données sont facilement accessibles et offrent une mine de connaissances qui peuvent aider les utilisateurs à prendre des décisions.

Le diagnostic de défauts dépend des modèles sous une forme ou une autre. Les modèles ne sont que des représentations condensées des connaissances. L'idée de base derrière ces stratégies est de transformer la base de données disponible en un modèle de processus implicite qui fournit des connaissances sur le fonctionnement du système.

Le principe de base de modélisation de ces approches est résumé sur la figure 1.3

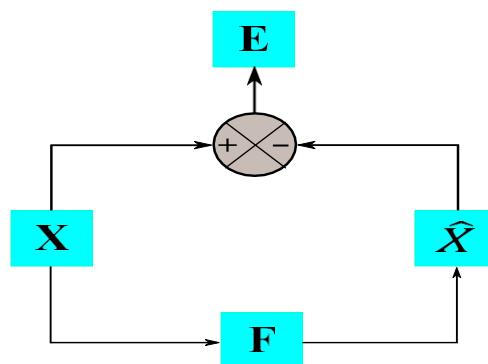


Figure 1.3 – Principe de base du diagnostic des défaut à base de données

Chapitre1 **Diagnostic de défauts à base de l'intelligence artificielle.**

On note que, X est une matrice de données qui regroupe l'historique de l'ensemble de données à surveiller. F est une matrice des caractéristiques extraites à partir de l'historique de données X . F est une matrice d'information qui représente toute information pertinente sur le fonctionnement processus. La matrice F est un modèle implicitement établi à partir de données historiques d'entrées et de sorties du système. \hat{X} est la matrice estimée. E est la matrice résiduelle, considérée aussi comme une matrice d'évaluation des résidus, qui aide à la décision.

1.6.1 Les différentes approches à base de traitement de données (Data-Driven Approches) [20]

Il existe plusieurs approches pour le traitement des données, chacune ayant ses propres avantages et inconvénients en fonction du contexte. Elles sont divisées en deux classes, la première concerne les approches statistiques multivarié et la deuxième classe est les approches de machine Learning.

- **Les approches statistiques multivarié :** Le traitement multivarié regroupe les méthodes statistiques qui s'attachent à l'observation et au traitement simultané de plusieurs variables statistiques en vue d'en dégager une information synthétique pertinente. Les deux grandes catégories de méthodes d'analyse statistique multivariées sont, d'une part, les méthodes dites descriptives et, d'autre part, les méthodes dites explicatives.

Les méthodes descriptives ont pour objectif d'aider à structurer et résumer un ensemble de données issues de plusieurs variables, sans privilégier particulièrement l'une de ces variables. Les méthodes explicatives ont, quant à elles, pour objectif d'expliquer l'une des variables (dite dépendante) à l'aide de deux ou plusieurs variables explicative (dites indépendantes).

- **Les approches de Machine Learning :** L'apprentissage automatique est devenu une branche majeure des mathématiques appliquées, à l'intersection des statistiques et de l'intelligence artificielle. Son objectif est de réaliser des modèles qui apprennent par l'exemple : il s'appuie sur des données numériques (résultats de mesures ou de simulations), contrairement aux modèles de connaissances qui s'appuient sur des équations issues des premiers principes de la physique, de la chimie, de la biologie, et de l'économie, etc.

L'apprentissage statistique est d'une grande utilité lorsque l'on cherche à modéliser des processus complexes, souvent non linéaires, pour lesquels les connaissances théoriques sont trop imprécises pour permettre des prédictions précises. Ses domaines d'applications sont

Chapitre1

Diagnostic de défauts à base de l'intelligence artificielle.

multiples : fouille de données, bio-informatique, génie des procédés, aide au diagnostic médical, télécommunications, vision par ordinateur, interface cerveau-machines, et bien d'autres. Les principaux types d'apprentissage automatique sont : l'apprentissage supervisé et l'apprentissage non supervisé. L'apprentissage automatique supervisé repose sur des données de formation d'entrée et de sortie étiquetées, tandis que l'apprentissage non supervisé traite des données non étiquetées ou brutes.

1.7 Conclusion

Au cours de ce chapitre, nous avons établi la nécessité et l'objectif de mettre en place une stratégie de diagnostic dans un environnement industriel. Nous avons brièvement présenté et expliqué les différentes approches utilisées dans le domaine du diagnostic, en mettant particulièrement l'accent sur les méthodes basées sur le traitement des données. Dans le prochain chapitre, nous examinerons plus en détail l'une de ces méthodes, à savoir la méthode des k plus proches voisins (k-Nearest Neighbors ou kNN), en nous concentrant spécifiquement sur son application dans le domaine du diagnostic de fonctionnement des systèmes.

Chapitre 2

2 La méthode des kNN pour la détection et l'isolation de défaut

Sommaire

2.1	Introduction	12
2.2	L'algorithme des KNN.....	12
2.3	Principe de fonctionnement du kNN.....	12
2.3.1	Distance Euclidienne	13
2.3.2	Distance de Cityblock.....	13
2.3.3	Distance Cosinus.....	13
2.3.4	Distance corrélation	13
2.4	Avantages et inconvénients du kNN	14
2.4.1	Les avantages	14
2.4.2	Les inconvénients.....	15
2.5	Exemple de simulation (kNN pour la classification des données).....	15
2.6	L'algorithme des KNN appliqué au diagnostic.....	17
2.6.1	Détection de défauts basée sur kNN	18
2.7	KNN pour localisation de défaut.....	19
2.7.1	Localisation par calcul des contributions.....	19
2.8	Exemple de simulation (kNN pour FDI).....	20
2.9	Conclusion.....	22

2.1 Introduction

L'algorithme des k plus proches voisins (kNN) est une méthode de classification non-paramétrique qui a été proposée pour la première fois par Fix et Hodges en 1951 [9]. Cette méthode est devenue l'une des techniques les plus populaires en apprentissage automatique et elle est utilisée dans de nombreux domaines, tels que la reconnaissance d'images, la bio-informatique, la classification de documents et la recommandation de produits.

Le principe de base de l'algorithme kNN est très simple. Pour classifier un nouvel échantillon de données, l'algorithme cherche les k échantillons les plus proches de celui-ci dans l'ensemble d'apprentissage et détermine la classe majoritaire parmi ces k voisins. La proximité est généralement mesurée à l'aide d'une distance euclidienne ou d'une mesure de similarité. Une des forces du kNN est qu'il ne nécessite pas de modèle ou d'hypothèses préalables sur la distribution des données, ce qui le rend particulièrement utile dans les situations où la structure sous-jacente des données est complexe ou inconnue. De plus, kNN peut être utilisé pour des problèmes de classification multi-classes et de régression.

Dans ce chapitre, nous allons explorer plus en détail le principe de base de l'algorithme kNN en classification, et sa mise en application pour la détection et la localisation de defaults.

2.2 L'algorithme des kNN

L'algorithme des k plus proches voisins (kNN) est une méthode d'apprentissage supervisé utilisée pour la classification et la régression. L'objectif principal de l'algorithme kNN est de déterminer la classe d'un nouvel exemple en se basant sur ses voisins les plus proches dans l'espace des caractéristiques. Cela signifie que l'algorithme va trouver les k échantillons les plus similaires à l'exemple donné et prédire la classe de l'exemple en fonction de la classe majoritaire parmi les k voisins.

2.3 Principe de fonctionnement du kNN

Le principe de fonctionnement de l'algorithme des k plus proches voisins (kNN) est relativement simple. L'algorithme kNN repose sur le principe que des échantillons similaires ont tendance à être regroupés dans des zones similaires de l'espace des caractéristiques. Ainsi, lorsqu'un nouvel exemple est présenté à l'algorithme kNN, celui-ci détermine d'abord les k échantillons les plus proches de l'exemple dans l'espace des caractéristiques. Ces échantillons sont appelés les "plus

Chapitre 2

"proches voisins" de l'exemple. La distance entre les échantillons est calculée à l'aide d'une mesure de distance appropriée, telle que la distance euclidienne.

Une fois que les k plus proches voisins ont été identifiés, l'algorithme kNN peut alors prédire la classe ou la valeur de la variable cible de l'exemple en utilisant une méthode de vote majoritaire. Cela signifie que la classe la plus fréquente parmi les k voisins sera attribuée à l'exemple (voir figure 2.1). Si l'algorithme kNN est utilisé pour la régression, la valeur moyenne des k plus proches voisins sera utilisée pour prédire la valeur de la variable cible.

Le choix de la valeur de k est important dans l'algorithme kNN car elle peut avoir un impact significatif sur les performances de l'algorithme. Si k est trop petit, l'algorithme peut être sensible aux bruits de données, tandis que si k est trop grand, l'algorithme peut ne pas être suffisamment flexible pour capturer les variations dans les données. Le choix de la mesure de distance est également important et dépendra des caractéristiques des données utilisées [10] [11]. Il existe plusieurs relations pour calculer les distances, parmi eux :

2.3.1 Distance Euclidienne

La distance Euclidienne entre les deux vecteurs X et Y est définie par :

2.3.2 Distance de Cityblock

Cette distance est définie comme suit :

$$d_E(X, Y) = \sum_{i=1}^m |x_i - y_i| \dots \dots \dots \quad (2.2)$$

2.3.3 Distance Cosinus

Cette distance est définie comme suit :

$$d_{cos}(X, Y) = 1 - \frac{\sum_{i=1}^m x_i y_i}{\sqrt{\sum_{i=1}^m x_i^2} \sqrt{\sum_{i=1}^m y_i^2}} \dots \dots \dots \quad (2.3)$$

2.3.4 Distance corrélation

Cette distance est définie comme suit :

$$d_{cos}(X, Y) = 1 - \frac{\sum_{i=1}^m (x_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^m (x_i - y_i)^2}} \cdot \frac{(x_i - y_i)}{\sqrt{\sum_{i=1}^m (x_i - y_i)^2}} \dots \dots \dots \quad (2.4)$$

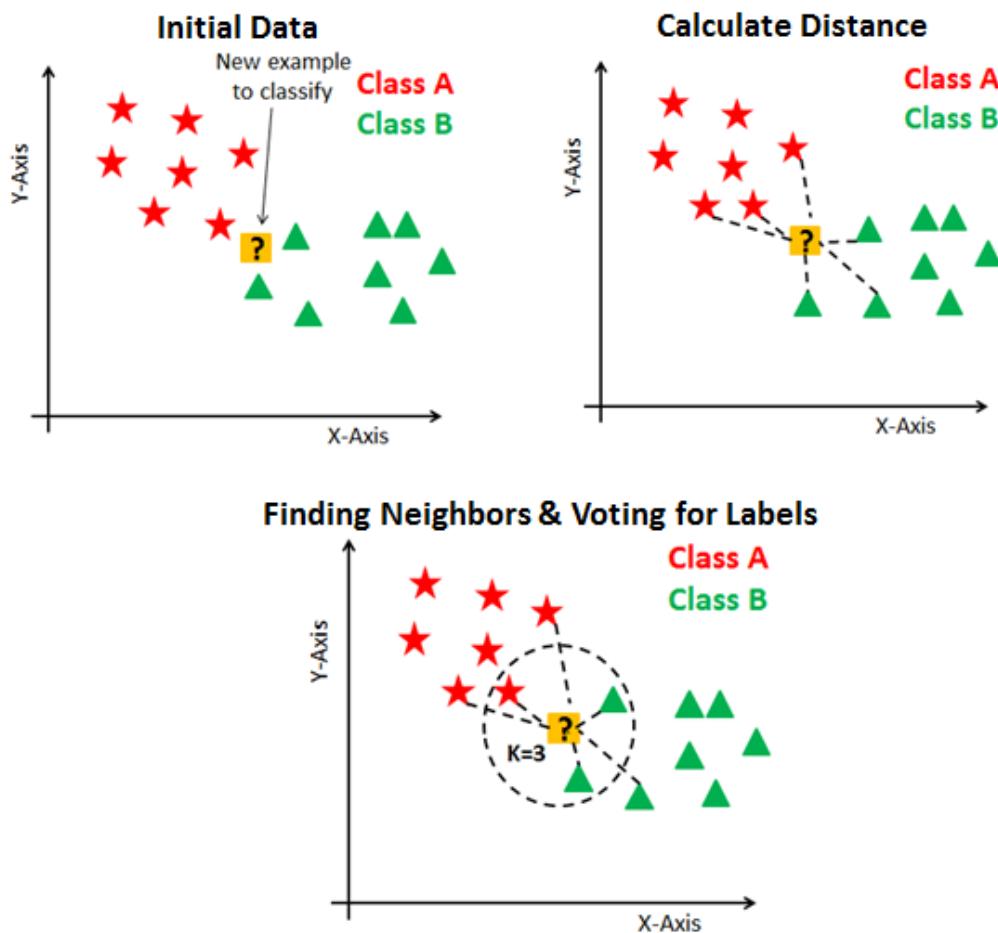


Figure 2.1 - Fonctionnement de l'algorithme kNN [12]

2.4 Avantages et inconvénients du kNN

L'algorithme des k plus proches voisins (kNN) est une méthode simple mais puissante pour la classification. Cependant, comme toutes les méthodes d'apprentissage automatique, elle a ses avantages et ses inconvénients.

2.4.1 Les avantages

- Facile à comprendre et à implémenter.
- Fonctionne bien pour les données non linéaires ou complexes.
- Peut être utilisé pour la classification et la régression.

Chapitre 2

La méthode des kNN pour la détection et l'isolation de défauts

- Peut être utilisé avec différents types de mesures de distance et de poids pour donner plus ou moins d'importance à certains voisins.
- Pas de phase d'apprentissage, ce qui signifie que l'algorithme peut être utilisé pour des problèmes en ligne.

2.4.2 Les inconvénients

- Peut-être lent pour des ensembles de données très volumineux ou avec un grand nombre de dimensions
- Sensible aux valeurs aberrantes et au bruit dans les données.
- Le choix de la valeur de k peut être crucial pour obtenir des résultats précis.
- Peut nécessiter un prétraitement des données pour normaliser les variables.
- La performance peut être affectée par le choix de la mesure de distance.

Il est important de noter que les avantages et les inconvénients de l'algorithme KNN peuvent varier selon le contexte d'utilisation et les caractéristiques des données [10,11].

2.5 Exemple de simulation (kNN pour la classification des données)

Voici un exemple trouvé dans Matlab d'une base de données très connu qu'il s'appelle Fisheriris qui implémente l'algorithme kNN comme classifieur avec une base de données simple qui est représenté sous la figure 2.2.

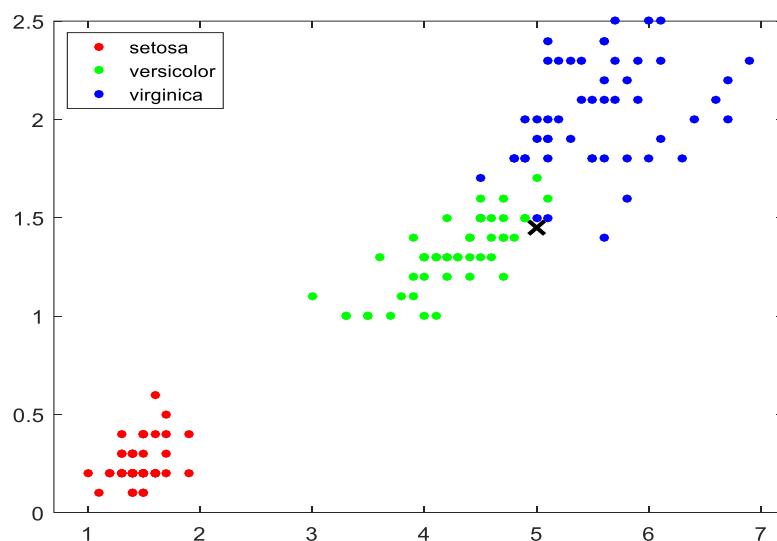


Figure 2.2 - Résultats de classification du nouveau point

Chapitre 2

La méthode des kNN pour la détection et l'isolation de défauts

Nous remarquons qu'il existe trois classes de données différentes en couleur, et la marque noire est le nouveau point que nous allons prédire sa classe.

Tout d'abord, nous choisirons le nombre de voisins, nous choisirons $k=8$. Ensuite, nous calculerons la distance euclidienne entre les points de données et le nouveau point. La distance euclidienne est la distance entre deux points, qui est représenté sous la relation suivante

$$d_E(X, Y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \dots \dots \dots (2.5)$$

Après avoir calculé la distance euclidienne entre le nouveau point et les données, nous avons obtenu les 8 proches voisins qu'ils sont représenté sous la figure 2.3

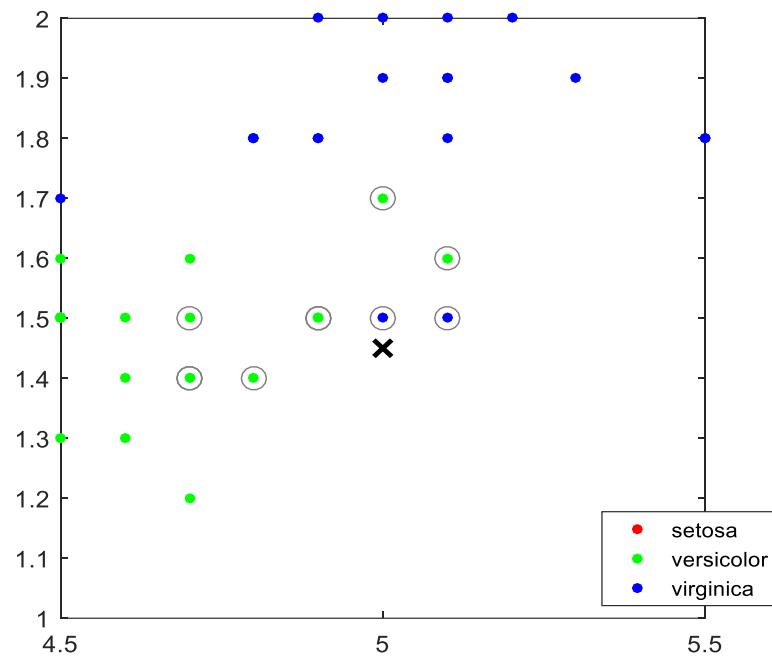


Figure 2.3 - les 8 proches voisins du nouveau point

Ensuite, pour déterminer la classe du nouveau point de données inconnues en se basant sur les classes des k voisins les plus proches dans un ensemble de données étiquetées, cette méthode s'appelle la règle de vote.

Chapitre 2 La méthode des kNN pour la détection et l'isolation de défauts

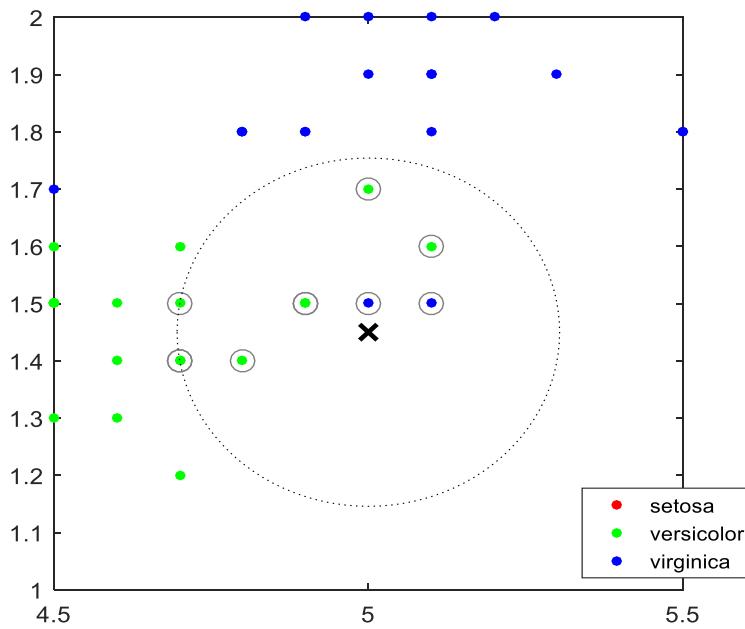


Figure 2.4 - k voisins les plus proches dans le cercle

La règle de vote est une étape cruciale dans l'algorithme kNN, car elle permet de classifier de nouveaux points en se basant sur les informations contenues dans les données étiquetées existantes. Elle est souvent utilisée dans des applications de classification, de reconnaissance de formes et d'apprentissage automatique.

2.6 L'algorithme des KNN appliqué au diagnostic

Le k-Nearest Neighbor (kNN) est un algorithme d'apprentissage supervisé qui peut être utilisé pour le diagnostic des défauts. Le principe de fonctionnement de l'algorithme KNN pour la prédiction des défauts est illustré sur la figure 2.5. Pour commencer, il est nécessaire de disposer d'une base de données contenant l'historique des données de fonctionnement normal du système pour tous les paramètres à surveiller. Ces données serviront d'ensemble d'apprentissage pour l'algorithme. Après, le nombre k de voisins les plus proches et la métrique de distance à utiliser sont choisis. Une fois que des données de test sont disponibles, une étape de détection de défauts est effectuée. Cette étape consiste à calculer un indice de détection, appelé D , basé sur les distances des k plus proches voisins. Si cet indice de détection dépasse un seuil prédéfini, cela indique la présence d'un défaut. À ce stade, une étape de localisation de défauts est déclenchée. Elle repose sur le calcul des

Chapitre 2

La méthode des kNN pour la détection et l'isolation de défauts

contributions des différents défauts à l'indice de détection D. Cette étape permet d'identifier la source probable du défaut.

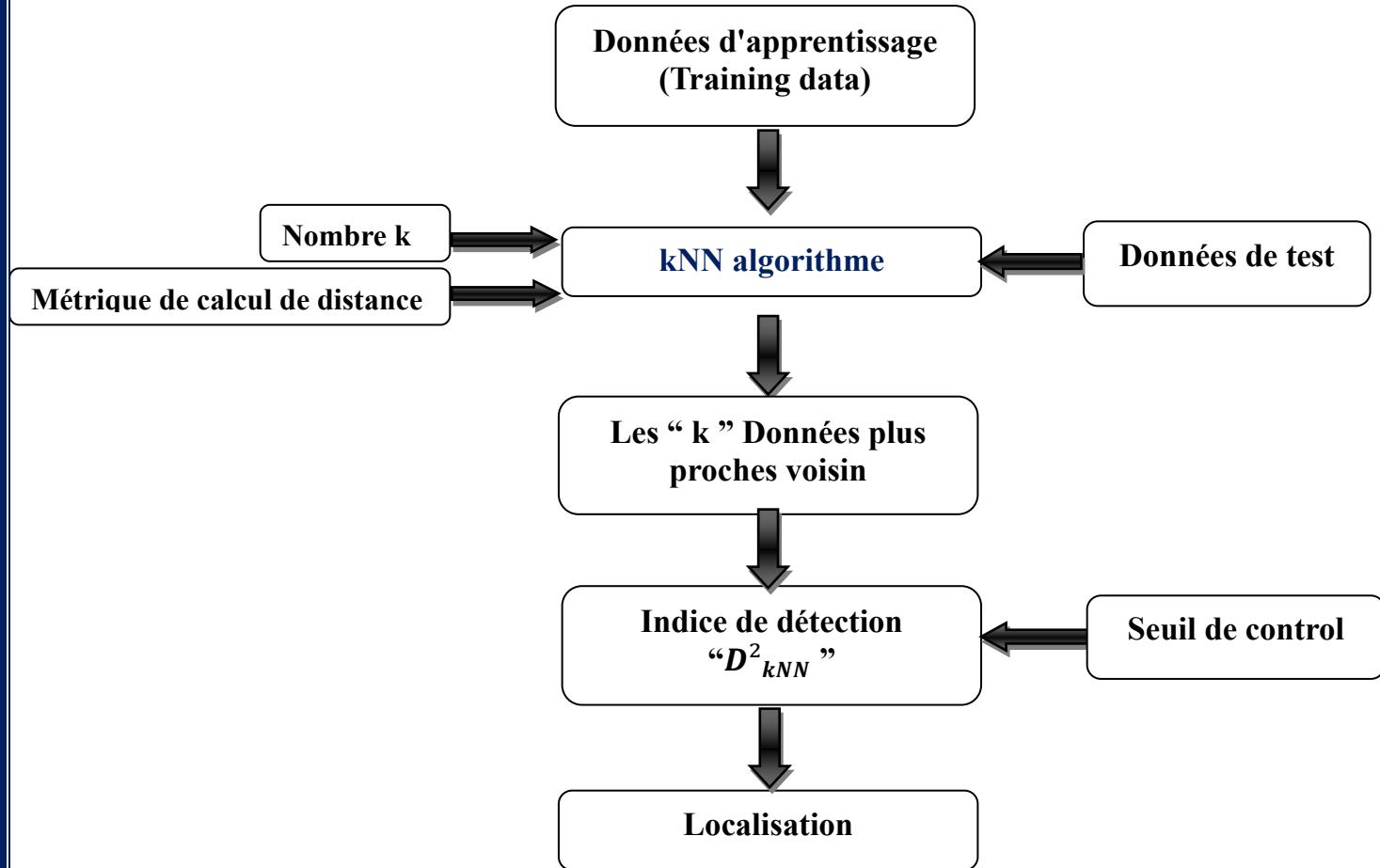


Figure 2.5 - KNN pour diagnostic de défaut

2.6.1 Détection de défauts basée sur kNN

He et Wang [13] proposent un indice de détection de défauts pour la méthode KNN. La détection de defaults est basée sur le calcul de la distance, qui est une somme des distances euclidiennes carrées d'un nouvel échantillon à ses voisins les plus proches, puis sa comparaison à la zone de fonctionnement normale.

La première étape consiste à calculer la distance quadratique moyenne entre l'échantillon de mesure nouvellement collecté (test data) $x_i \in \Re^m$ et ($k < N$) et ses proches voisins de la matrice de données du fonctionnement normal X comme suit :

Chapitre 2

$$D^2_{kNN}(\mathbf{i}) = \frac{1}{k} \sum_{j=1}^k d^2_{kNN}(i, j) \dots \dots \dots \quad (2.6)$$

Avec,

Ici d_{kNN} (i, j) est la distance euclidienne au carré entre l'échantillon et son j -ème proches voisins. La deuxième étape consiste à calculer la limite de contrôle utilisée pour la détection des anomalies, selon la littérature, il existe de nombreuses façons de calculer la limite de contrôle $D^2_{\alpha kNN}$ avec un niveau de signification α . Dans notre cas, un choix empirique de la valeur du seuil est effectué dans la phase de test, i.e., en absence de défauts.

Après cela, la détection de défaut en ligne peut être simplement appliquée en calculant la distance euclidienne au carré $D^2_{xkNN}(i)$ de l'échantillon entrant $x_i \in \mathbb{R}^m$ aux k voisins les plus proches comme dans Eq. (2.6) et ensuite appliquer le problème de détection des défauts dans le cadre statistique [17]

$$D^2_{xkNN} \leq D^2_{akNN} \text{ hypothèse nulle (cas sans faute)} \dots \dots \dots \quad (2.9)$$

$$D^2_{xkNN} > D^2_{akNN} \text{ hypothèse alternative (cas de défaut).} \dots \dots \dots \quad (2.10)$$

2.6.2 KNN pour localisation de défaut

La localisation de défaut consiste à déterminer l'emplacement du défaut, c'est-à-dire l'endroit où la défaillance s'est produite. Parmi les différentes méthodes de localisation de défaut, il existe une méthode basée sur le calcul des contributions à l'indice de détection. Cette méthode de localisation est couramment utilisée, en particulier dans le contexte des méthodes statistiques telles que l'ACP (Analyse en Composantes Principales). Une extension de cette approche est également proposée pour le cas spécifique de la méthode KNN (k-Nearest Neighbors).

2.7 Localisation par calcul des contributions

Le principe de contribution est généralement quantifié en fonction de la part de chaque variable dans le calcul d'un indice de détection donné. La méthode est basée sur l'idée que la variable qui contribue le plus à la métrique de détection de défaut est la plus susceptible de porter le défaut. Elle consiste à déterminer la contribution de chaque variable aux métriques de détection de fautes utilisées.

Inspirés par l'idée de la méthode des contributions d'analyse en composantes principales (ACP), nous décomposons la distance KNN (indice de détection en FD-KNN) d'un échantillon \mathbf{x} en la somme de m composantes :

Chapitre 2

$$D^2 x = \sum_{i=1}^m \sum_{j=1}^k [\xi_i^T (x - x_j)]^2 \dots \dots \dots \quad (2.11)$$

Ainsi, nous définissons la contribution de la i ème variable de \mathbf{x} à l'indice de détection D^2_x comme

$$Cont_{kNN}(i) = \sum_{j=1}^k [\xi_i^T (x - x_j)]^2, \quad i=1,\dots,m \quad \dots \dots \dots \quad (2.12)$$

2.8 Exemple de simulation (kNN pour FDI)

Voici un exemple de simulation du kNN pour FDI (Fault Détection Isolation). Afin d'illustrer les différentes étapes de l'algorithme kNN pour la détection et la localisation de default, nous avons utilisé une base de données utilisée dans une compétition de machine Learning appelée "Artificial Intelligence for Industries" (AI4I) en 2020. Cette compétition a été organisée dans le but de promouvoir l'utilisation de l'intelligence artificielle dans les industries et de stimuler les avancées dans le domaine, c'est une base de données pour prédire la panne de la machine (binaire) et le type (multi classe), c'est une base de données synthétique qui reflète les données réelles de maintenance prédictive rencontrées dans l'industrie. L'ensemble de données contient 10000 cas des 5 mesures des capteurs présentés dans la table 2.1. Les premières 7000 échantillons sont utilisés comme des données d'apprentissage tandis que le reste est considéré comme des données de test.

N°var	Description des variables	Unité
X1	Température de l'air	[K]
X2	Température de processus	[K]
X3	Vitesse rotationnelle	[Tr/min]
X4	Couple	[Nm]
X5	Usure des outils	[Min]

Tableau 2.1 – Les variables des 5 mesures des capteurs

On a choisi un nombre des proches voisins $k=4$ puis on a calculé l'indice de détection D2 pour toutes les données de test. La figure 2.6 montre l'évolution de l'indice de détection dans le cas d'absence de default avec la couleur bleu. Après simulation un défaut d'amplitude d'environ 40% sur la quatrième mesure (Couple) des échantillons de test à partir de l'échantillon 1000. Le

Chapitre 2

La méthode des kNN pour la détection et l'isolation de défauts

dépassement de seuil est clair sur le deuxième indice de la même figure, ce qui explique la présence d'un défaut.

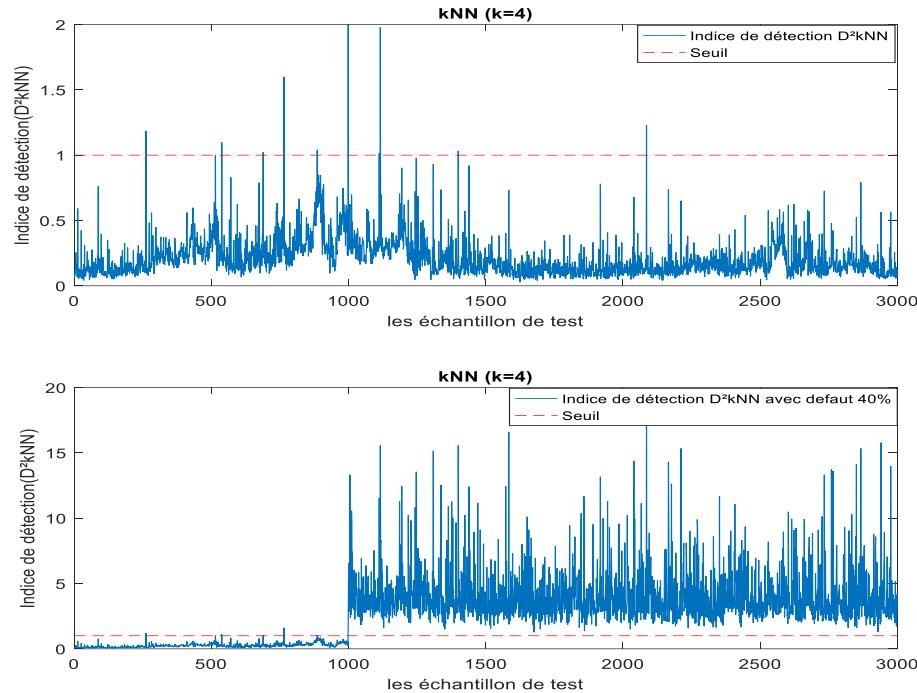


Figure 2.6 - l'indice de détection D^2kNN dans les deux cas

Pour faire localiser le défaut, on calcule les contributions de chaque variable à l'indice de détection D^2kNN . Les contributions sont affichées et représentées graphiquement sous la figure 2.7, où on remarque que la variable qui contribue le plus est la variable X4 alors que c'est la X4 qui est en défaut.

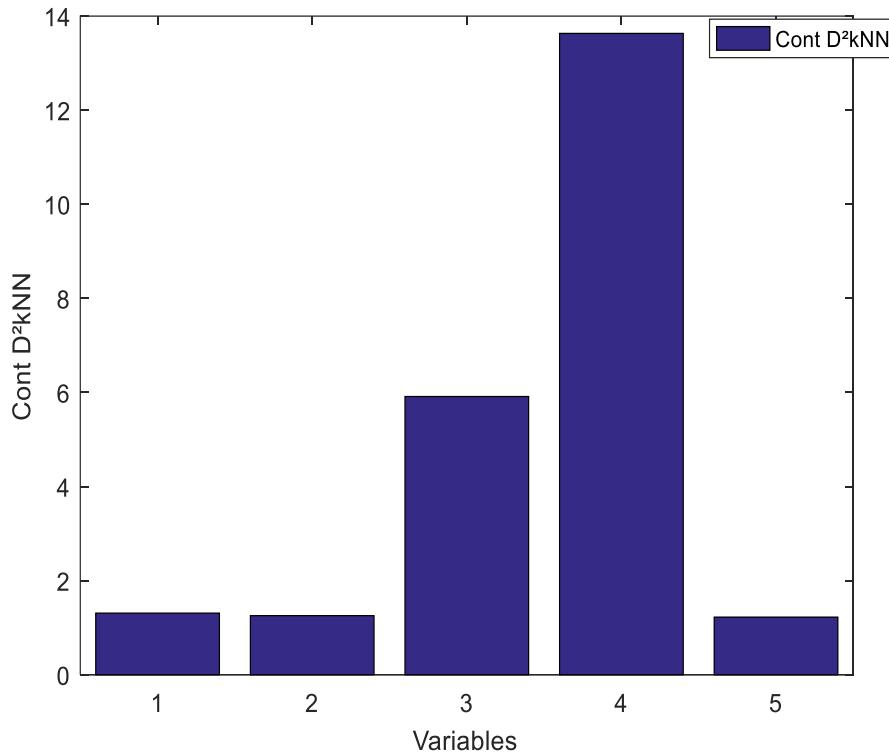


Figure 2.7 - La contribution de chaque variable

2.9 Conclusion

Dans ce chapitre, nous avons présenté la méthode de kNN pour la classification de données, puis sa mise en application pour des problème de diagnostic prédictif de défauts. Dans sa version conventionnelle, de nombreux problèmes peuvent être rencontrés, parmi lesquels l'algorithme devient beaucoup plus lent lorsque le nombre de données d'apprentissage augmente. Cela rendre l'étape de prédiction de défauts lente. Ainsi que sa complexité de calcul est de l'ordre de $O(n)$ avec ($k \ll n$). Ce qui augmente la consommation de l'espace mémoire.

Le chapitre suivant se concentre sur le clustering et ses différentes méthodes et leurs combinaisons avec kNN dans le diagnostic de default.

Chapitre 3

3 Les méthodes C-means-KNN et Fuzzy C-means-KNN pour diagnostic de défauts

Sommaire

3.1	Introduction	24
3.2	Principe de clustering	24
3.3	Les avantages de clustering.....	25
3.4	La méthode C-means.....	25
3.4.1	Les étapes de l'algorithme C-means	25
3.5	Fuzzy C-means clustering	26
3.5.1	Les étapes de l'algorithme Fuzzy C-means	26
3.6	La différence entre C means et fuzzy C means	27
3.7	La méthode C-means-KNN pour détection et isolation de défauts.....	27
3.8	KNN et Fuzzy C-means pour détection et isolation défaut.....	29
3.9	Exemple de simulation	30
3.10	Conclusion.....	33

3.1 Introduction

L'identification de groupes similaires dans des ensembles de données est connue sous le nom de clustering. Le clustering est une technique d'apprentissage non supervisée. Il est utilisé dans divers domaines, notamment l'exploration de données, la biologie, la psychologie, la reconnaissance de formes, etc. Le regroupement permet la découverte de structures, et l'identification des caractéristiques de données.

Ce chapitre est consacré à étudier deux méthodes de clustering, telles que le clustering avec C-means et le clustering avec Fuzzy C-means, et leur combinaison avec la méthode KNN pour obtenir des meilleures performances de l'approche diagnostic.

3.2 Principe de clustering

Le clustering est une méthode d'apprentissage automatique qui consiste à regrouper des points de données par similarité ou par distance. C'est une méthode d'apprentissage non supervisée et une technique populaire d'analyse statistique des données. Pour un ensemble donné de points, vous pouvez utiliser des algorithmes de classification pour classer ces points de données individuels dans des groupes spécifiques. En conséquence, les points de données d'un groupe particulier présentent des propriétés similaires. Alors que les points de données de différents groupes ont des caractéristiques différentes.

Le but des algorithmes de clustering est de donner un sens aux données et d'extraire de la valeur à partir de grandes quantités de données structurées et non structurées. Ces algorithmes vous permettent de séparer les données en fonction de leurs propriétés ou fonctionnalités et de les regrouper dans différents clusters en fonction de leurs similitudes.

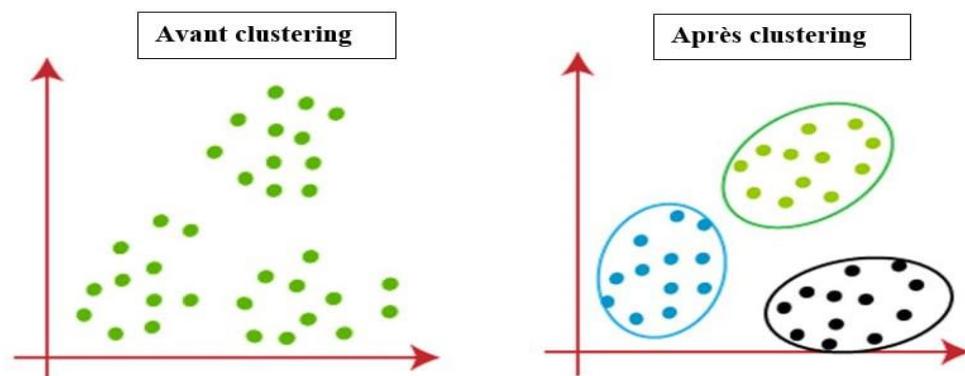


Figure 3.1 – le principe de clustering

3.3 Les avantages de clustering

- Le clustering permet d'explorer et de découvrir des structures intrinsèques dans les données sans a priori sur les classes ou les étiquettes. Cela peut révéler des informations cachées et faciliter la compréhension des données.
- Le clustering permet d'explorer et de découvrir des structures intrinsèques dans les données sans a priori sur les classes ou les étiquettes. Cela peut révéler des informations cachées et faciliter la compréhension des données.
- Les algorithmes de clustering peuvent être utilisés dans diverses applications, telles que la segmentation d'images, la segmentation de clients et la détection d'anomalies.
- Les algorithmes de clustering peuvent identifier des modèles dans les données qui peuvent ne pas être immédiatement apparents

3.4 La méthode C-means

Le clustering C-means est un algorithme d'apprentissage automatique non supervisé qui est utilisé pour regrouper un ensemble de points de données en clusters en fonction de leurs similitudes, il s'agit d'un type d'algorithme de clustering qui attribue chaque point de données à l'un des K clusters en fonction de sa distance par rapport aux centres de cluster.

Le "c" dans C-means fait référence aux centroïdes des k clusters créés par l'algorithme. Il commence par initialiser aléatoirement k centroïdes, puis ajuste itérativement les positions des centroïdes jusqu'à ce qu'elles convergent vers des positions stables. Les points de données sont ensuite attribués au centroïde le plus proche, qui forme les clusters k.

Le clustering C-means est un algorithme populaire pour la segmentation d'images, la compression de données et les applications d'exploration de données. L'un des principaux avantages du clustering C-means est qu'il est simple à mettre en œuvre et efficace en termes de calcul.

Cependant, il peut être sensible au choix initial des centroïdes et le nombre de clusters K doit être spécifié à l'avance.

3.4.1 Les étapes de l'algorithme C-means

- **Etape 1 :** Spécifiez le nombre de clusters C que vous souhaitez obtenir. Choisissez C points aléatoires dans l'ensemble de données comme centres initiaux des clusters.

- **Etape 2 :** Calculez la distance entre ce point et chaque centroïde
- **Etape 3 :** Calculez les nouveaux centroïdes pour chaque cluster en trouvant la moyenne des points qui lui sont associés
- **Etape 4 :** Répétez les étapes 2 et 3 jusqu'à ce que les points ne changent plus de cluster.
- **Etape 5 :** L'algorithme converge lorsque les points ne changent plus de cluster, ce qui signifie que les centroïdes ne se déplacent plus.

3.5 Fuzzy C-means clustering

Fuzzy C-Means (FCM) est un algorithme de clustering (ou classification non supervisée) qui permet de regrouper un ensemble de données en clusters en fonction de leur similarité. Contrairement aux méthodes de clustering traditionnelles qui assignent chaque point de données à un cluster unique et exclusif, FCM attribue à chaque point un degré d'appartenance pour chaque cluster, reflétant ainsi la probabilité que ce point appartienne à chaque cluster. FCM est basé sur l'idée de la théorie des ensembles flous, qui permet de représenter la notion de degré d'appartenance d'un élément à un ensemble plutôt que la notion binaire d'appartenance ou non-appartenance. L'algorithme FCM utilise une fonction objective pour minimiser la distance entre chaque point de données et les centres des clusters pondérés par les degrés d'appartenance des points. L'objectif est d'obtenir des centres de clusters qui minimisent l'erreur globale entre les données et les centres de clusters. FCM est largement utilisé dans divers domaines tels que la reconnaissance de formes, l'analyse d'image, la biologie, la finance, la médecine, etc. en raison de sa capacité à gérer des données complexes et ambiguës, ainsi que sa flexibilité et sa facilité d'utilisation.

La méthode de clustering Fuzzy C-means où elle partitionne les ensembles de données en c-clusters en minimisant la fonction de coût suivante :

$$\text{Minimize } J = \sum_{j=1}^c \sum_{i=1}^N \pi_{ij}^q \| x_i - C_j \|^2 \dots \dots \dots \quad (3.1)$$

$$\text{s.t. } \sum_{j=1}^c \pi_{ij} = 1, i=1,2,\dots,N \dots \dots \dots \quad (3.2)$$

3.5.1 Les étapes de l'algorithme Fuzzy C-means

- **Etape 1 :** Spécifier le nombre de clusters c et un degré de flouté m
- **Etape 2 :** Choisissez les centroïdes initiaux pour chaque cluster au hasard ou en utilisant une méthode spécifique.

Chapitre 3

Les méthodes C-means-KNN et Fuzzy C-means-KNN pour diagnostic de défauts

- **Etape 3 :** Répétez les étapes suivantes (a et b) jusqu'à ce que les centroïdes convergent et que les degrés d'appartenance soient stables.
- **Etape a :** Affecter chaque point de données à un cluster en calculant sa "flouïté" (degré d'appartenance) à chaque cluster en fonction de la distance euclidienne par rapport aux centroïdes.
- **Etape b :** Mettre à jour les centroïdes de chaque cluster en calculant la moyenne pondérée des points de données en fonction de leur degré d'appartenance.

3.6 La différence entre C means et fuzzy C means

C-means et Fuzzy C-means sont deux algorithmes de clustering qui visent à regrouper des points de données en clusters en fonction de leur similarité. La principale différence entre eux est que Fuzzy C-means permet à un point de données d'appartenir à plus d'un cluster avec un certain degré d'appartenance.

Dans C-means, chaque point de données est assigné à un seul cluster, et l'assignation est déterminée en minimisant la distance entre le point de données et le centre du cluster. Le centre du cluster est calculé comme la moyenne de tous les points de données assignés à ce cluster. Dans Fuzzy C-means, contrairement à la méthode des K-means qui assigne de manière rigide chaque point de données à un seul cluster, la méthode de C-means floue permet une certaine "flouïté" en attribuant des degrés d'appartenance partiels à chaque point pour chaque cluster. Cela permet une représentation plus fine des transitions entre les clusters, ce qui améliore la qualité globale de la classification. Cela peut également influencer la qualité de l'approche de diagnostic en renforçant sa résistance aux fausses alertes et en augmentant sa sensibilité à la détection des défauts.

3.7 La méthode C-means-KNN pour détection et isolation de défauts

La combinaison de l'approche C-means avec l'algorithme KNN a pour objectif principal de réduire les problèmes liés à la complexité de calcul et la consommation de l'espace mémoire. L'algorithme de la méthode combiné C-means KNN est illustré dans la figure suivante :

Chapitre 3
Les méthodes C-means-KNN et Fuzzy C-means-KNN pour diagnostic de défauts

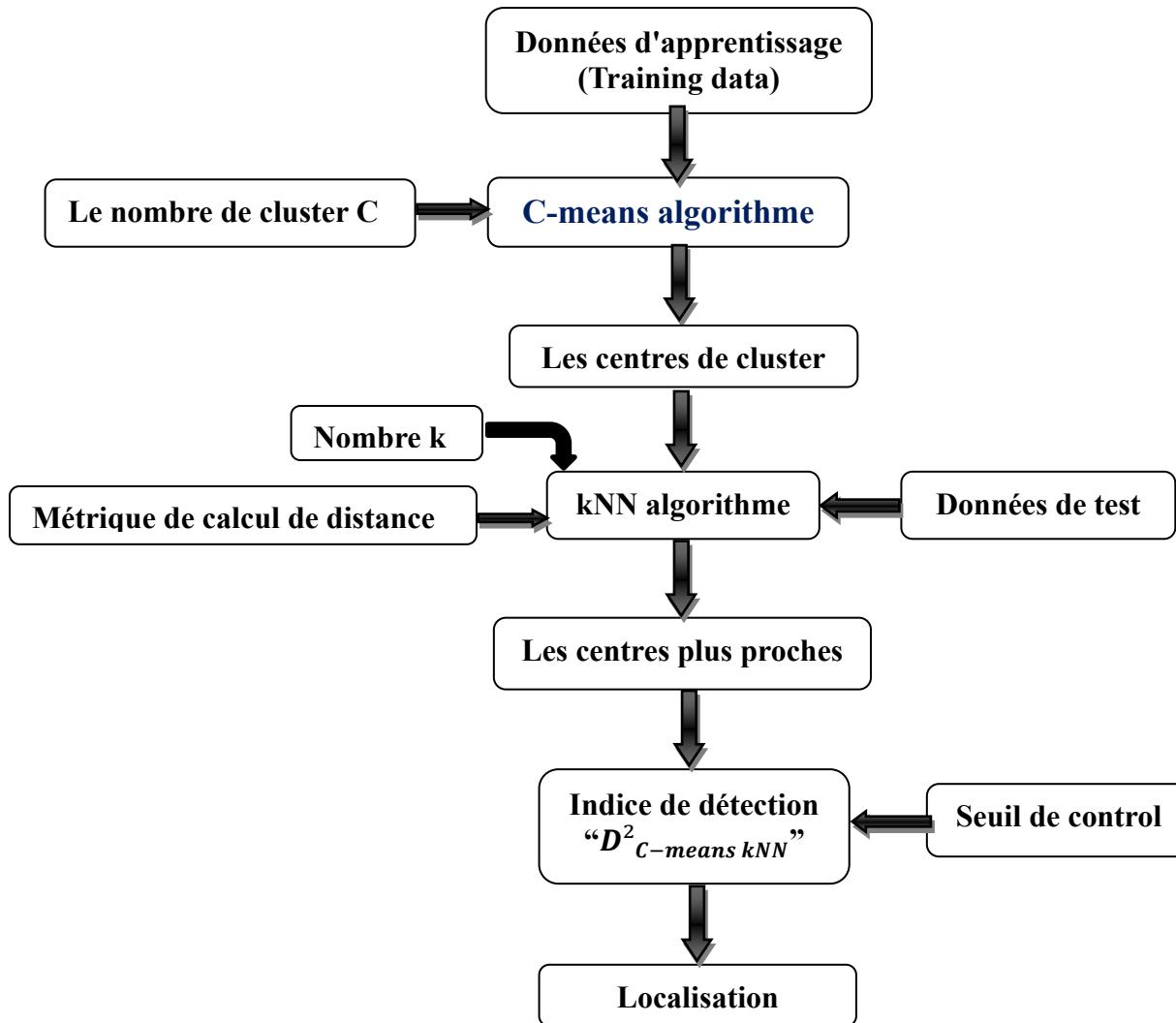


Figure 3.2 – C-means KNN pour diagnostic de défaut

La méthode C-means KNN utilise la distance euclidienne entre les centres des clusters et le nouveau point pour déterminer sa classe. Les distances sont calculées entre les centres des clusters étiquetés et la donnée de test pour classer le nouveau point de données. Ainsi, le calcul des k voisins les plus proches est limité aux centres des clusters plutôt que d'inclure l'ensemble des données d'apprentissage, ce qui réduit la complexité de l'algorithme (de O(n) à O(c), où c représente le nombre de clusters choisi).

3.8 KNN et Fuzzy C-means pour détection et isolation défaut

L'application de l'approche conventionnelle KNN pour la prédiction de défauts présente plusieurs faiblesses. C'est pourquoi l'algorithme kNN n'est pas largement utilisé dans sa forme originale, mais plutôt avec des versions améliorées limitant partiellement ces lacunes [18,19]. Cette nouvelle technique est proposée pour traiter les lacunes de la règle kNN en combinant les avantages de l'algorithme de clustering FCM dans la classification des données pour la détection de défauts.

Après l'étape d'apprentissage, à chaque instant, lorsqu'un nouvel échantillon est présent, il suffit de calculer la distance du nouvel échantillon au centre de chaque cluster sans avoir à réexplorer tous les apprentissages précédents puis l'hypothèse de détection est examinée pour prendre une décision. En utilisant l'algorithme FCM-KNN, les performances de surveillance du processus sont améliorées sous deux aspects, la vitesse de calcul et la complexité de la mémoire de stockage. L'organigramme qui résume les différentes étapes de cette méthode est présenté dans la figure suivante :

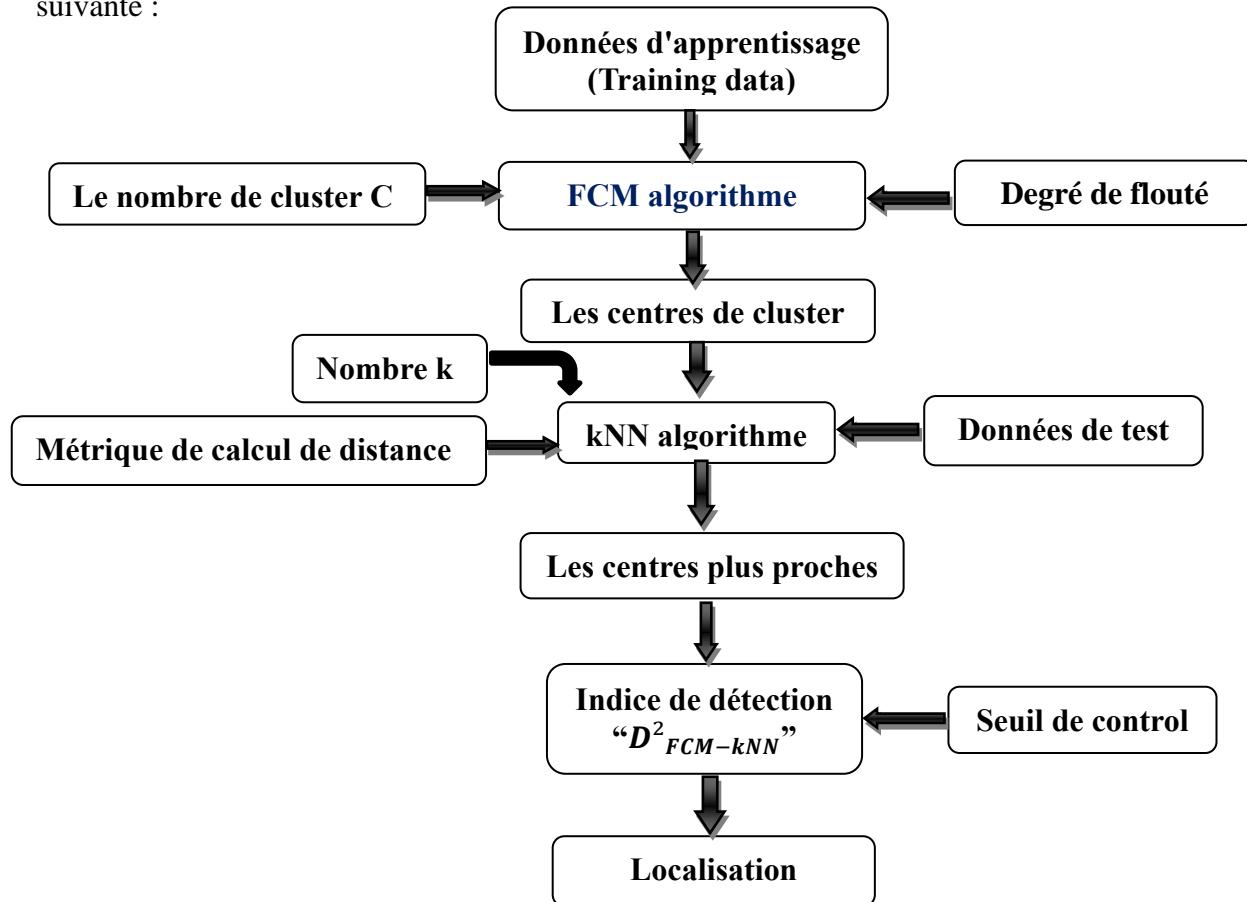


Figure 3.3 – FCM-kNN pour diagnostic de défaut

Chapitre 3

Les méthodes C-means-KNN et Fuzzy C-means-KNN pour diagnostic de défauts

Après qu'un nouvel échantillon $x \in m$ est présent, ses kNN aux centres des classes dans les ensembles de données d'apprentissage sont déterminés.

Ensuite, en calculez la distance euclidienne quadratique moyenne de l'échantillon entrant x à ses k voisins les plus proches comme suit

$$D^2_{FCM-kNN}(x) = \frac{1}{k} \sum_{j=1}^k d^2_{FCM-kNN}(x, j) \dots \dots \dots \quad (3.4)$$

Lorsque l'hypothèse de défaut est vérifiée, le diagnostic de défaut est exécuté en décomposant le $D^2_{FCM-kNN}(x)$ en une somme de termes où chaque terme associé à une variable est

$$D^2_{FCM-kNN}(x) = \sum_{i=1}^m \sum_{j=1}^k \| \zeta_i(x - C_j) \| \dots \dots \dots \quad (3.5)$$

Par conséquent, la contribution de chaque variable dans l'indice de détection de défaut $D^2_{FCM-kNN}(x)$ est déterminée comme suit

3.9 Exemple de simulation

Dans cet exemple, nous utilisons les mêmes données que celles utilisées dans le deuxième chapitre et appliquons la méthode de C-means kNN et FCM-kNN pour le diagnostic de default, la base de données contient 10000 échantillons pour 5 mesures des capteurs présenter dans la table 2.1 du chapitre précédent. Dans la phase d'apprentissage automatique, nous avons utilisé un jeu de données d'une taille égale à 7000 échantillons et 3000 pour le test, on prend le nombre de plus proches voisins $k=4$ et le nombre de cluster $C=1000$ dans les deux méthodes, on a créé un défaut avec un amplitude de 40% dans la variable X4. Les résultats obtenus sont présentés sous les figures (3.4) (3.5).

Chapitre 3

Les méthodes C-means-KNN et Fuzzy C-means-KNN pour diagnostic de défauts

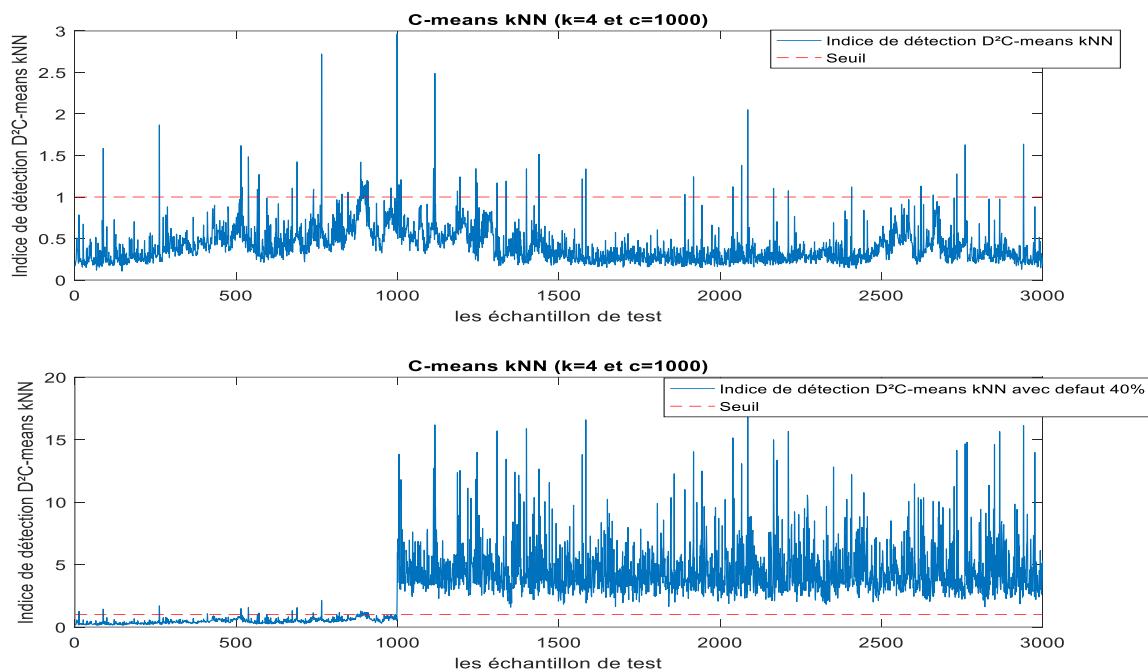


Figure 3.4 - L'indice de détection $D^2C\text{-means kNN}$ dans les deux cas

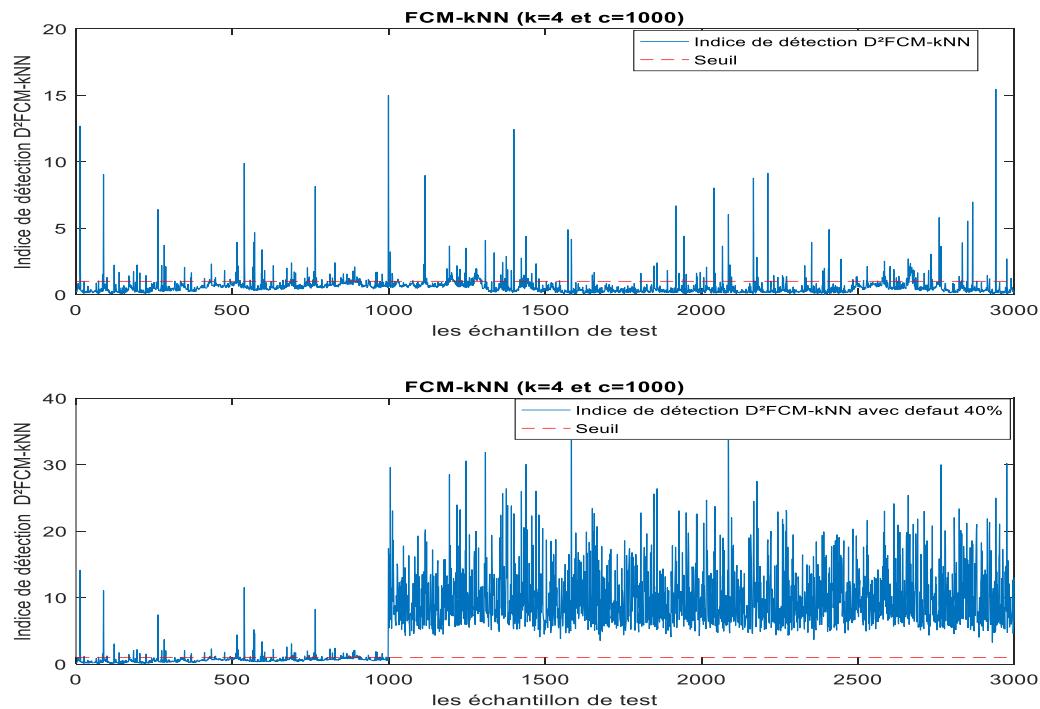


Figure 3.5 - L'indice de détection $D^2 FCM\text{-kNN}$ dans les deux cas

Chapitre 3

Les méthodes C-means-KNN et Fuzzy C-means-KNN pour diagnostic de défauts

Les figures (3.4) et (3.5) représentent l'indice de détection D^2 C-means kNN et D^2 FCM-kNN avec un défaut d'amplitude 40%. La ligne bleue représente l'indice de détection pour chaque échantillon de test, tandis que la ligne rouge en pointillés représente le seuil.

L'indice de détection D^2 C-means kNN et D^2 FCM-kNN mesure la distance entre chaque nouvel échantillon de test et les plus proches centres dans l'espace des caractéristiques. Plus l'indice de détection est élevé, plus il est probable que l'échantillon soit considéré comme un défaut.

Dans ce cas, on peut observer que certains échantillons de test dépassent le seuil, indiquant qu'ils sont considérés comme des défauts détectés. Pour déterminer la position du défaut, nous utilisons la méthode du calcul des contributions, qui consiste à identifier la variable ayant la plus grande contribution, car elle est plus susceptible d'être l'origine du défaut. Les figures (3.6) et (3.7) illustrent le calcul des contributions des variables en fonction de l'indice de détection. Il est démontré que la variable ayant la plus grande contribution sur l'indice de détection est la variable X4.

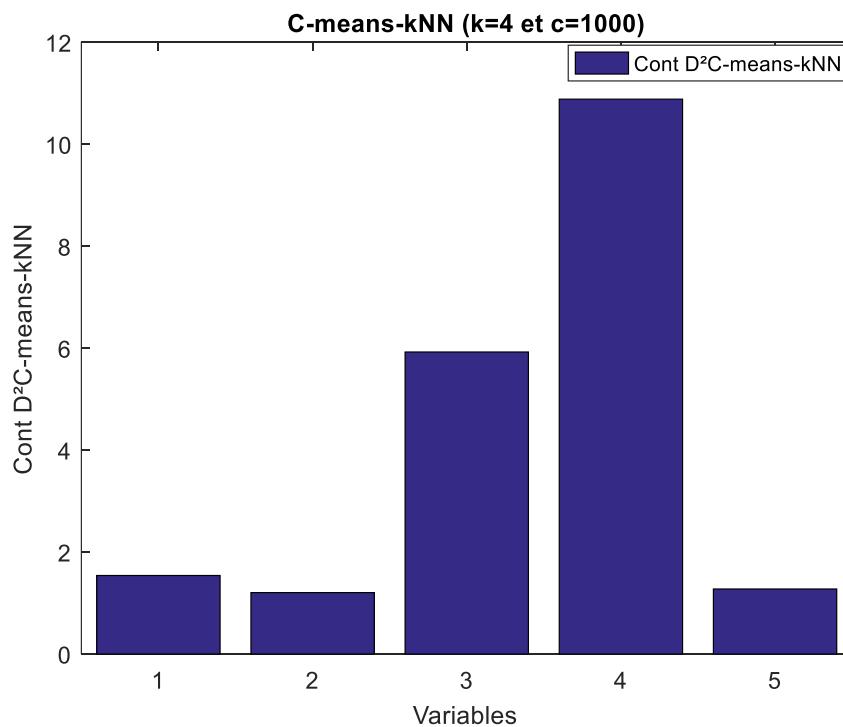


Figure 3.6 - La contribution de chaque variable (C-means kNN)

Chapitre 3
Les méthodes C-means-KNN et Fuzzy C-means-KNN pour diagnostic de défauts

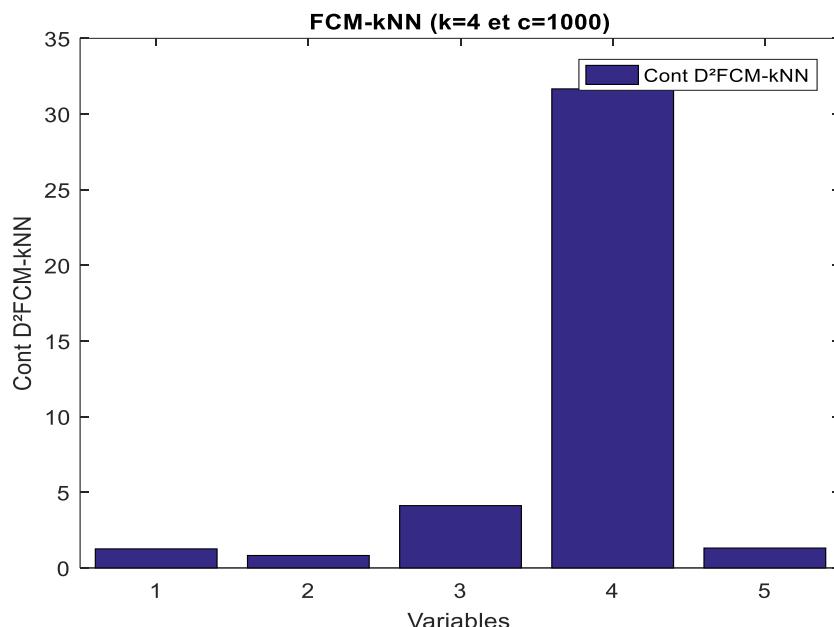


Figure 3.7 - La contribution de chaque variable (FCM-kNN)

3.10 Conclusion

Dans ce chapitre, nous avons abordé les concepts fondamentaux de l'analyse en cluster, ainsi que son objectif et ses applications dans le domaine du diagnostic de défauts. Nous avons examiné en particulier les approches C-means et Fuzzy C-means. De plus, nous avons effectué une simulation pour démontrer l'efficacité des approches combinées dans la prise de décision ou pour des raisons de détection et d'isolation de défauts.

Dans le chapitre suivant, nous appliquerons les connaissances acquises jusqu'à présent à l'analyse des résultats obtenus sur un processus réel

Chapitre 4

4 Application (compresseur centrifuge)

Sommaire

<u>4.1</u>	<u>Introduction</u>	35
<u>4.2</u>	<u>Description de champs de Haoud Berkaoui</u>	35
<u>4.2.1</u>	<u>Description de la station de compression de Haoud Berkaoui</u>	36
<u>4.2.2</u>	<u>Description de Compresseur K103</u>	36
<u>4.3</u>	<u>Diagnostic de défauts du compresseur K103-A1</u>	37
<u>4.3.1</u>	<u>Résultats de simulation</u>	37
<u>4.4</u>	<u>Conclusion</u>	47

4.1 Introduction

L'application de techniques de détection et de localisation de défauts dans les processus industriels est devenue de plus en plus importante ces dernières années, car elle permet aux entreprises de maintenir une production de haute qualité tout en réduisant les coûts associés aux défauts. Les défauts dans un processus industriel peuvent avoir des conséquences graves, allant de la perte de temps et d'argent à des accidents mortels. Par conséquent, il est essentiel d'identifier rapidement les défauts et de les localiser avec précision pour minimiser leur impact.

Dans ce chapitre, nous présentons l'application des différentes techniques qui ont été développées dans les chapitres précédents pour la détection et la localisation des défauts dans un processus dans la station de compression Haoud Berkaoui, ils sont appliqués pour montrer leur intérêt et aussi bien expliciter leur mise en œuvre.

4.2 Description de champs de Haoud Berkaoui

Le champ de pétrole couvre une superficie de 175 kilomètres carrés, il a été découvert en 1965 et mis en production en janvier 1967.

Le centre de production de Haoud Berkaoui (HBK) se compose d'un dispositif de séparation d'huile de 8 000 m³/jour, d'un self-stockage de 13 000 m³, d'un dispositif de compression de gaz de 1 million de mètres cubes/jour, d'un dispositif d'injection d'eau de 2 000 m³/jour et d'une station de déshuilage d'une capacité de production de 2 400 m³/jour.



Figure 4.1 – Le centre de traitement du champ de Haoud Berkaoui [21].

Chapitre 4 Application (compresseur centrifuge)

4.2.1 Description de la station de compression de Haoud Berkaoui

L'usine de récupération des gaz de torche du site de production de la zone de Haoud Berkaoui récupère les gaz d'hydrocarbures séparés du pétrole lors du processus de flash physiquement adiabatique.

Le gaz est séparé de l'huile à l'intérieur de trois systèmes de séparation existants, dont la différence repose sur le niveau de pression de travail. De chaque système (nommés BP, MP, HP, basse pression, moyenne pression, haute pression) le gaz est envoyé vers la nouvelle installation de récupération de gaz qui les comprime à une pression de 35bar environ.

Remarque les lettres A et B servent de repère ; pendant l'exploitation le train en marche sera celui principal (A ou B) alors que celui à l'arrêt sera celui de réserve (A ou B).

La station de récupération, de compression et d'envoie du gaz torché se compose de deux lignes de compression (1ère BP+MP et 2ème ligne HP). Chaque ligne est dotée de deux trains de compression : le train principal indiquée par la lettre A, et le train de réserve indiquée par la lettre B.

Les deux lignes de compression (1ère et 2ème) sont connectées à la même ligne de distribution de gaz vers Guellala.

Nous sommes intéressés par la deuxième ligne dans laquelle il se trouve le compresseur k103-A1

4.2.2 Description de Compresseur K103

Le compresseur centrifuge K103 est utilisé en deuxième ligne, et il comprime le gaz provenant des séparateurs HP jusqu'à la pression du collecteur de réception (35barg). Il fournit deux étages de pression qui sont entraînés par un moteur électrique.

Le compresseur centrifuge K103-A est contrôlé par leur propre UCP. Ce dernier peut être exercés en contrôle de pression ou de débit par moyen de la vanne de contrôle, positionnée en amont du séparateur initial. La valeur de set point de pression ou de débit peuvent être imposées à l'UCP du compresseur

Dans le 1er cas (contrôle de pression), la vanne de contrôle maintient à une valeur constante la pression des séparateurs HP existants en amont des limites de batterie de la nouvelle installation. Le débit du gaz alimenté au compresseur varie en fonction du gaz produit par les systèmes de séparation en amont : si la production de ce gaz augmente, la pression en amont de la vanne de contrôle tend à augmenter ; pour garder la pression constante la vanne de contrôle s'ouvre

Chapitre 4

Application (compresseur centrifuge)

d'avantage et laisse passer un débit plus important. La limite étant représentée par la zone de fonctionnement stable au-delà de laquelle l'arrêt de la machine est prévu

Dans le 2ème cas (contrôle de débit), la vanne de contrôle maintient à une valeur constante du débit alimenté à la deuxième ligne de compression. Le fonctionnement du compresseur est fixé sur le point déterminé par la courbe caractéristique du compresseur (débit et de rapport de compression β défini comme : P_{ref}/P_{asp}). Le débit alimenté au compresseur étant fixé, si la production de ce gaz en amont augmente, la pression en amont de la vanne de contrôle augmente jusqu'à la valeur maximale consentie au-delà de laquelle les vannes d'évent de l'installation existante s'ouvrent vers torche.

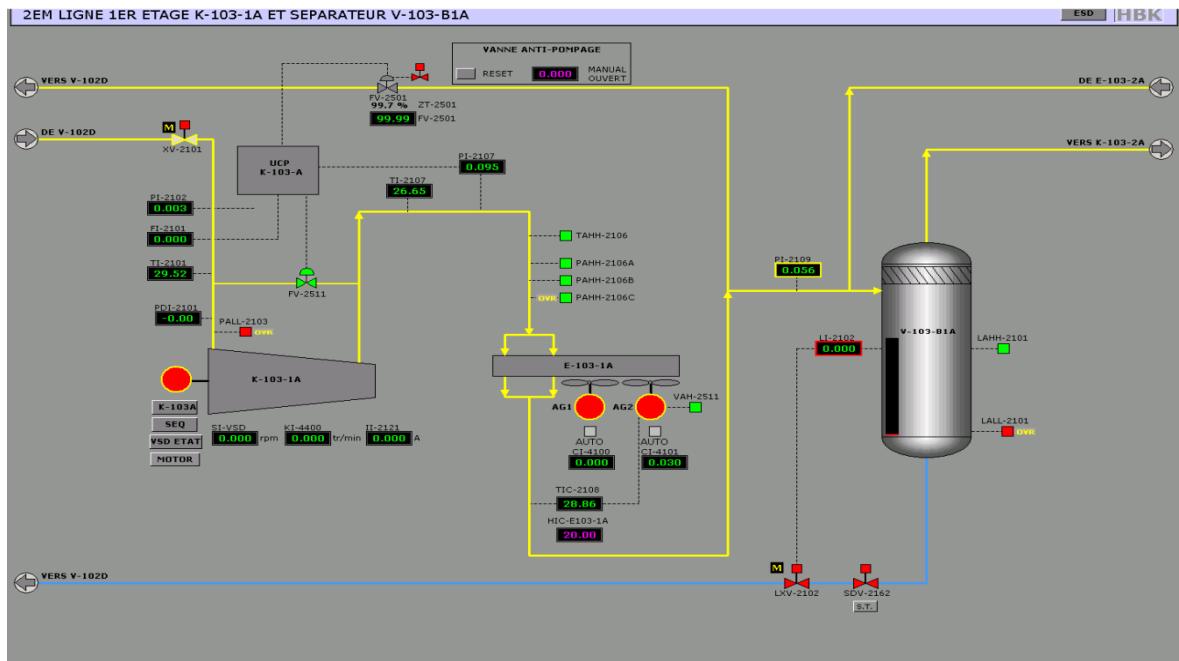


Figure 4.2- Logigramme DCS de compresseur K103-A1

4.3 Diagnostic de défauts du compresseur K103-A1

4.3.1 Résultats de simulation

Dans ce travail, les 5 variables mesurées en continu ont été utilisées pour construire la matrice de données, ces variables sont répertoriées dans le tableau (4.1). Les algorithmes décrits dans les chapitres précédents telle que kNN, kNN C-means et kNN Fuzzy C-means sont utilisées pour diagnostiquer les défauts au temps réel.

Chapitre 4 Application (compresseur centrifuge)

N°var	Description des variables	Unité
X1	Température gaz (entrée)	[°C]
X2	Capteur de débit du compresseur (entrée)	[Bar]
X3	Pression aspiration compresseur	[Nm ³ /h]
X4	Température gaz (sortie)	[°C]
X5	Pression refoulement compresseur	[Bar]

Tableau 4.1 – Les variables collectée pour l’application

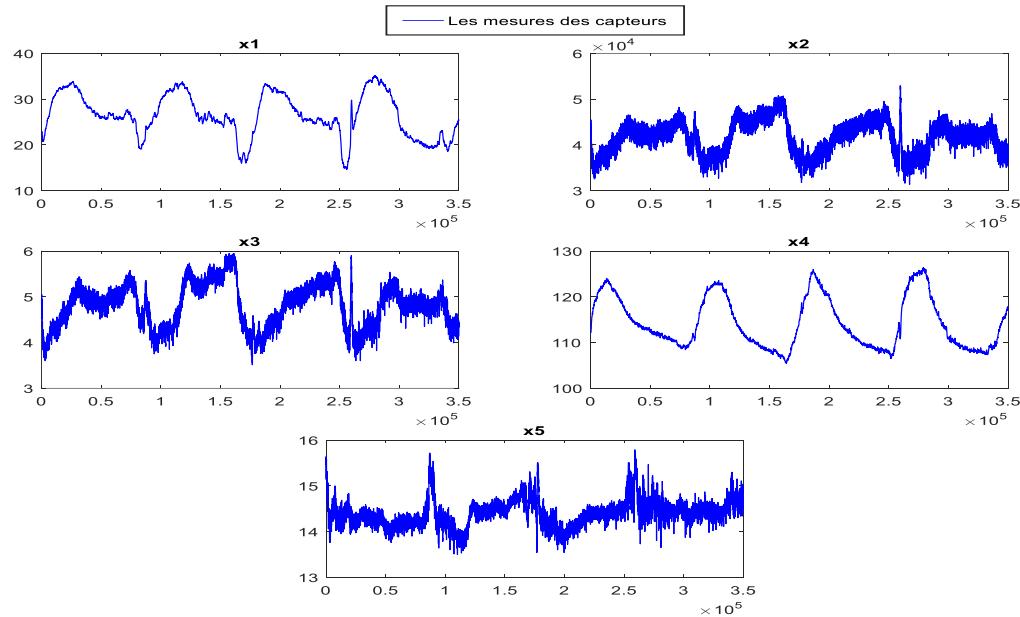


Figure 4.3 - L’évolution des mesures des 5 capteurs

Nous avons 350000 échantillons de mesure pour chaque variable qu’ils sont illustrés sous la figure (4.3). Dans la phase d’apprentissage automatique, nous avons utilisé un jeu de données d’une taille égale à 245000 échantillons de mesure. Les 105000 échantillons restants de la phase de test ont

Chapitre 4

Application (compresseur centrifuge)

été utilisés pour détecter et localiser les erreurs. Nous avons choisi le nombre de voisins les plus proches et le nombre de cluster différents. Les seuils de l'indice de détection d'erreurs sont prendre paramétrique (sans calcul).

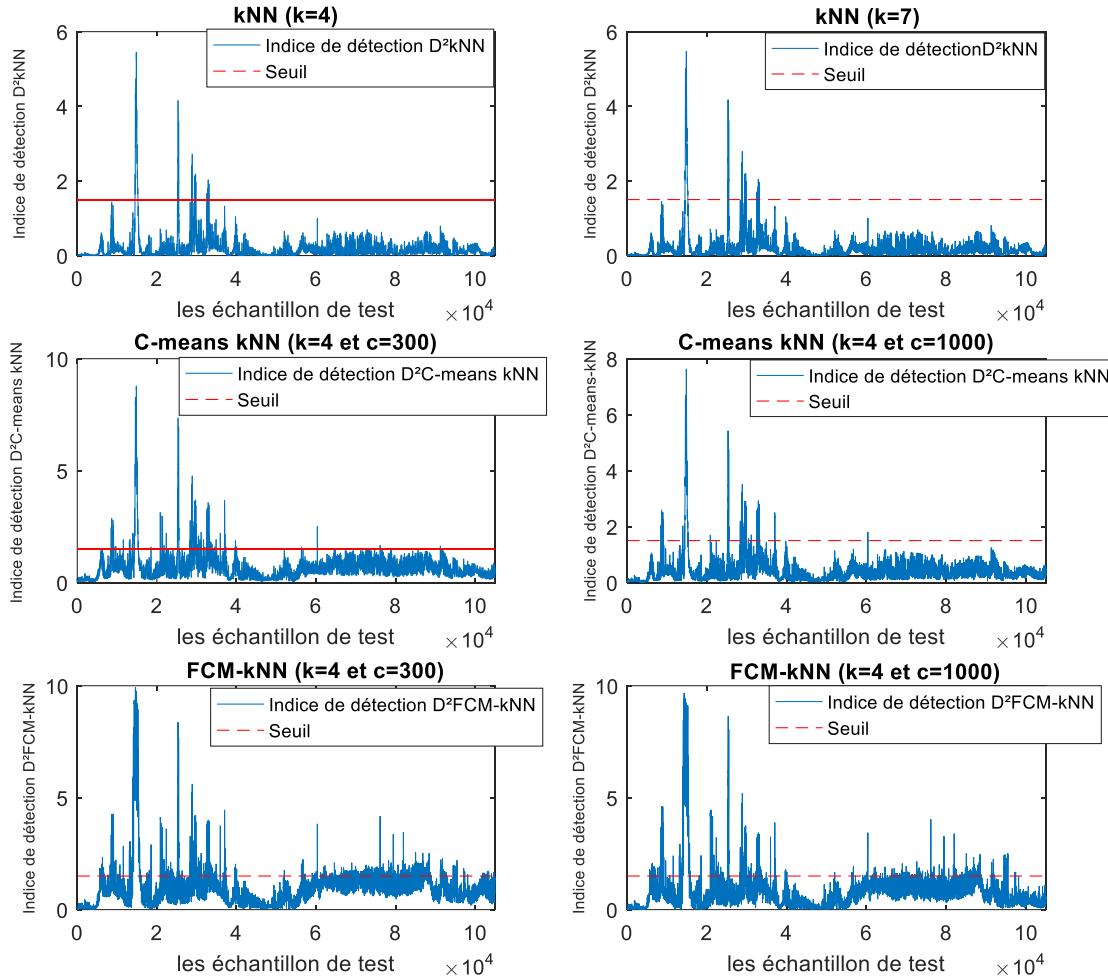


Figure 4.4 – Evolution des indices de détection sans default avec déférents paramètres (k et C)

La figure (4.4) montre les indices de détection, que signifier la distance entre chaque échantillon de test et leurs proches voisins ou bien les proches centres dans les algorithmes kNN C-means et FCM-kNN, il est bien montré que le taux de fausse alarme est énorme, et ça diminue quand on augmente le nombre de clusters, le dépassement de seuil fourni par les trois indices est indésirable.

Chapitre 4 Application (compresseur centrifuge)

Ce dernier montre que le système est défectueux, alors que le système fonctionne correctement dans cette plage de fonctionnement.

Afin d'illustrer la procédure de détection, un default sur la variable X3 (Pression aspiration compresseur) a été simulée d'amplitude d'environ 40 % de sa plage de variance apparaissant du temps 60000 à 105000 considéré comme le dernier échantillon dans les données de test comme le montre la figure (4.5).

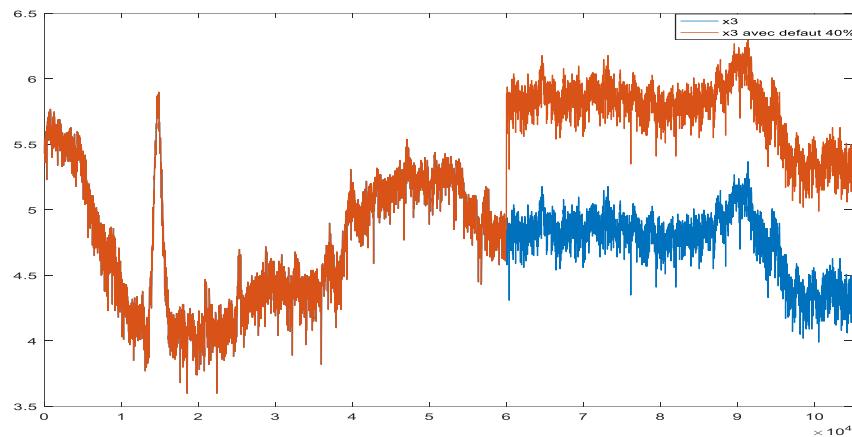


Figure 4.5 – Evolution de la variable X3 (Pression aspiration compresseur) sans et avec défaut

Chapitre 4

Application (compresseur centrifuge)

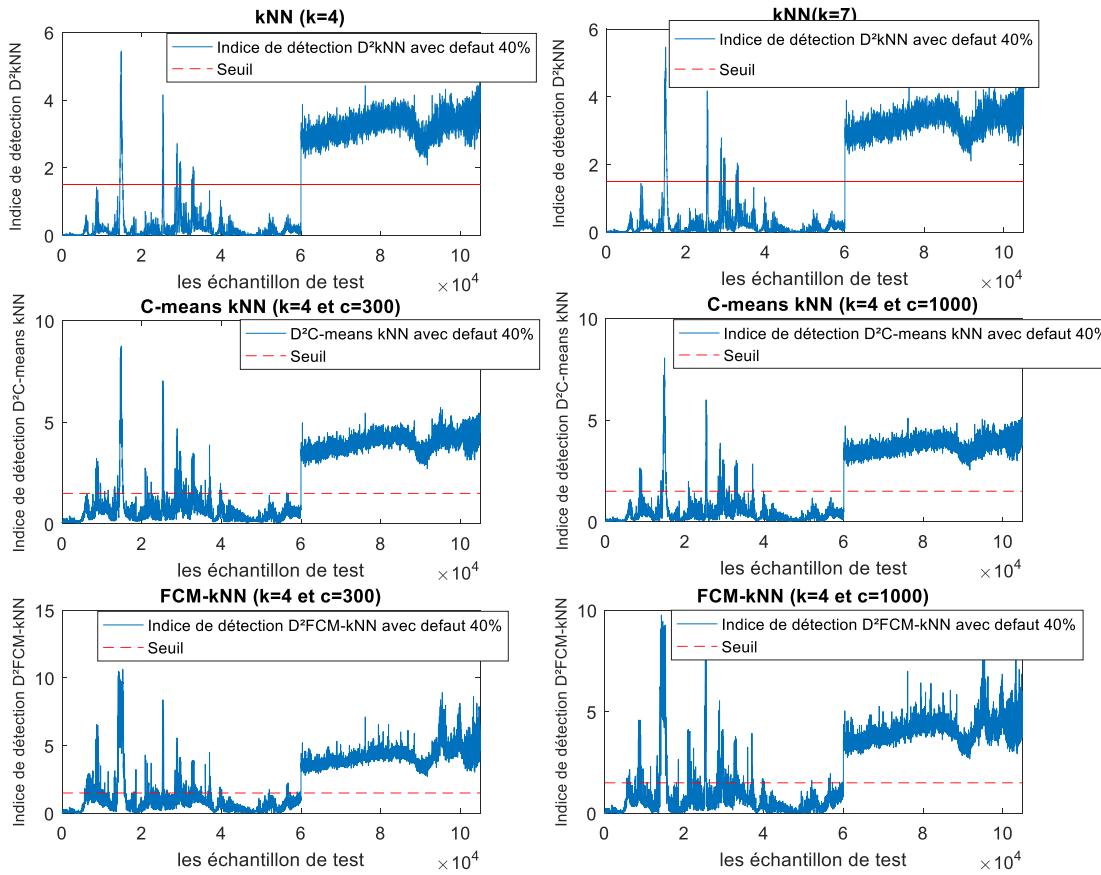


Figure 4.6 – Evolution des indices de détection avec default dans X3 avec déférents paramètres (k et C)

D'après la figure (4.6) qui montre évolution des indices de détection on remarque que le dépassement de seuil fournit par les indices de l'échantillon 60 000 jusqu'au dernier échantillon est énorme donc on a détecté un default.

Une fois qu'un défaut est détecté grâce à les indices de détection, une étape est nécessaire pour localiser le défaut. Pour la position de défaut, nous prenons la méthode de calcul des contributions, donc la variable avec la plus grande contribution est la plus susceptible de faire un défaut. Les figures (4.7), (4.8), et (4.9) montrent le calcul des contributions des variables selon l'indice de détection. Il est bien montré sur les figures (4.7), (4.8), et (4.9) que la variable qui possède la plus grande contribution selon l'indice de détection est la variable X3.

Chapitre 4
Application (compresseur centrifuge)

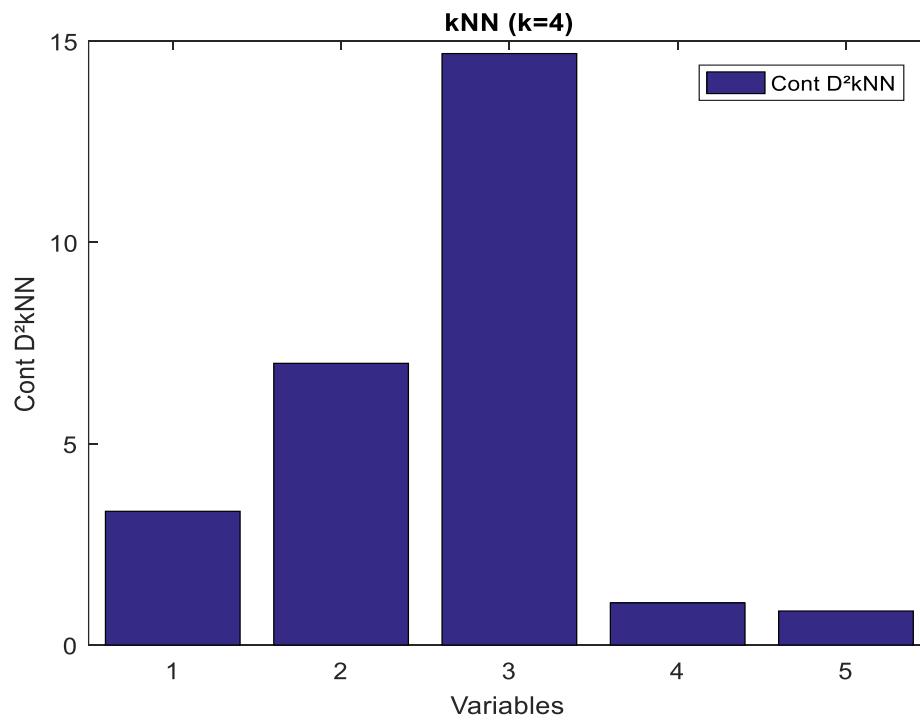


Figure 4.7 – Contributions des variables fautives du défaut X3 : kNN

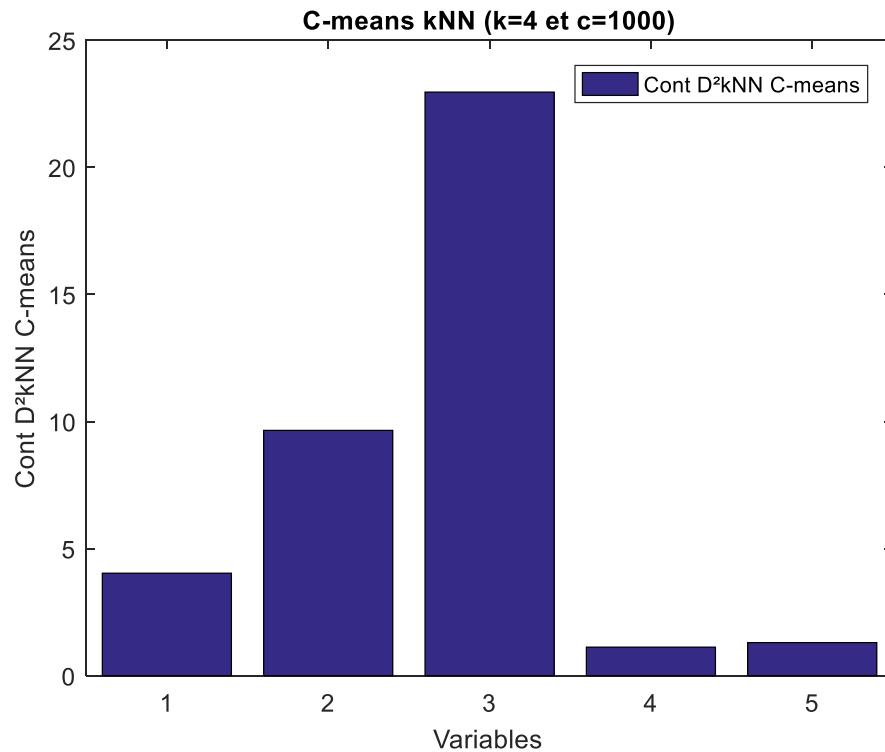


Figure 4.8 – Contributions des variables fautives du défaut X3 :C-means et kNN

Chapitre 4
Application (compresseur centrifuge)

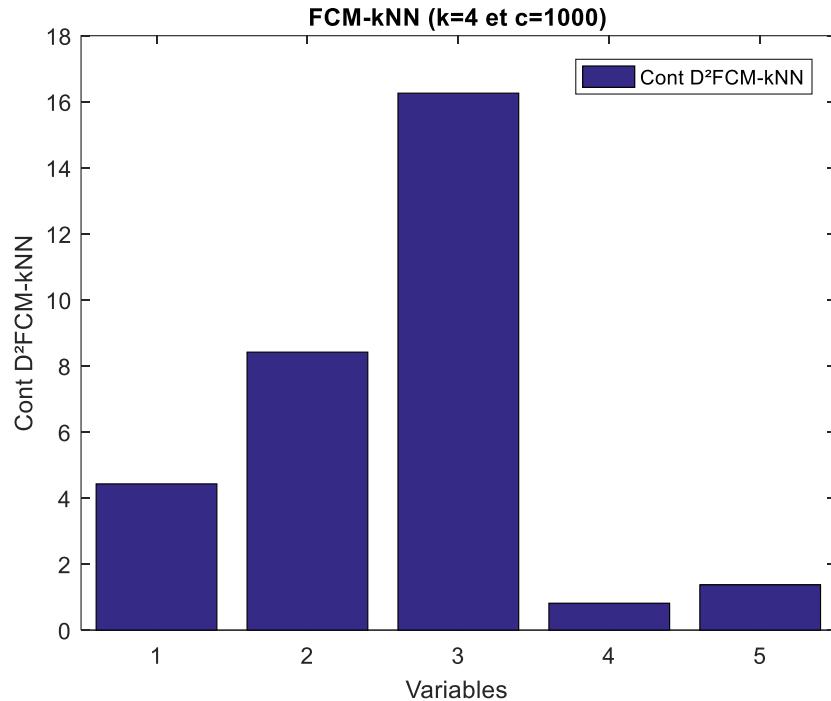


Figure 4.9 – Contributions des variables fautives du défaut X3 : FCM-kNN

Sur la même base de données qui est représenté dans le cas précédent, la même partition qu'avant, on change l'emplacement de défaut et on le met dans la variable X2 d'amplitude d'environ 40 % de sa plage de variance apparaissant du temps 60000 jusqu'au dernier échantillon comme le montre la figure (4.10), la figure (4.11) montre les indices de détection, il est bien représenté que le défaut a été détecter avec les trois méthodes à cause de dépassement de seuil. Les figures (4.12), (4.13), et (4.14) montrent le calcul des contributions des variables.

Chapitre 4

Application (compresseur centrifuge)

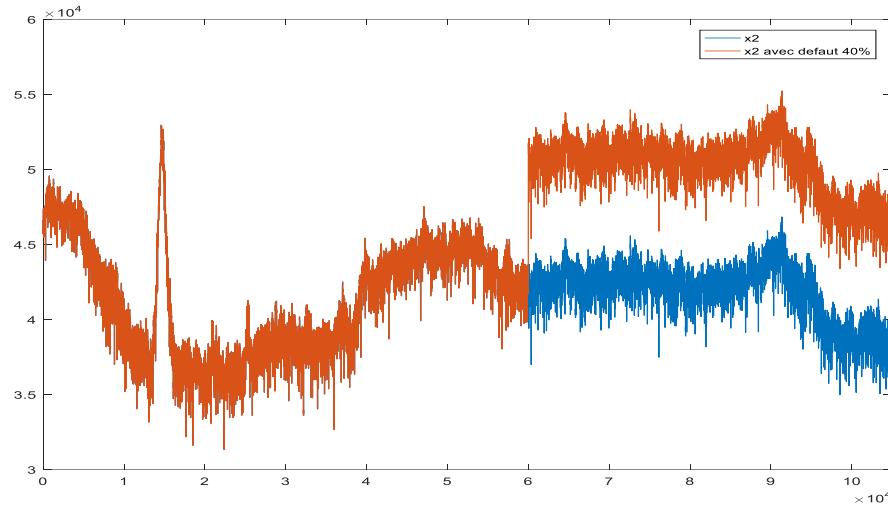


Figure 4.10 – Evolution de la variable X2 (Capteur de débit du compresseur) sans et avec défaut

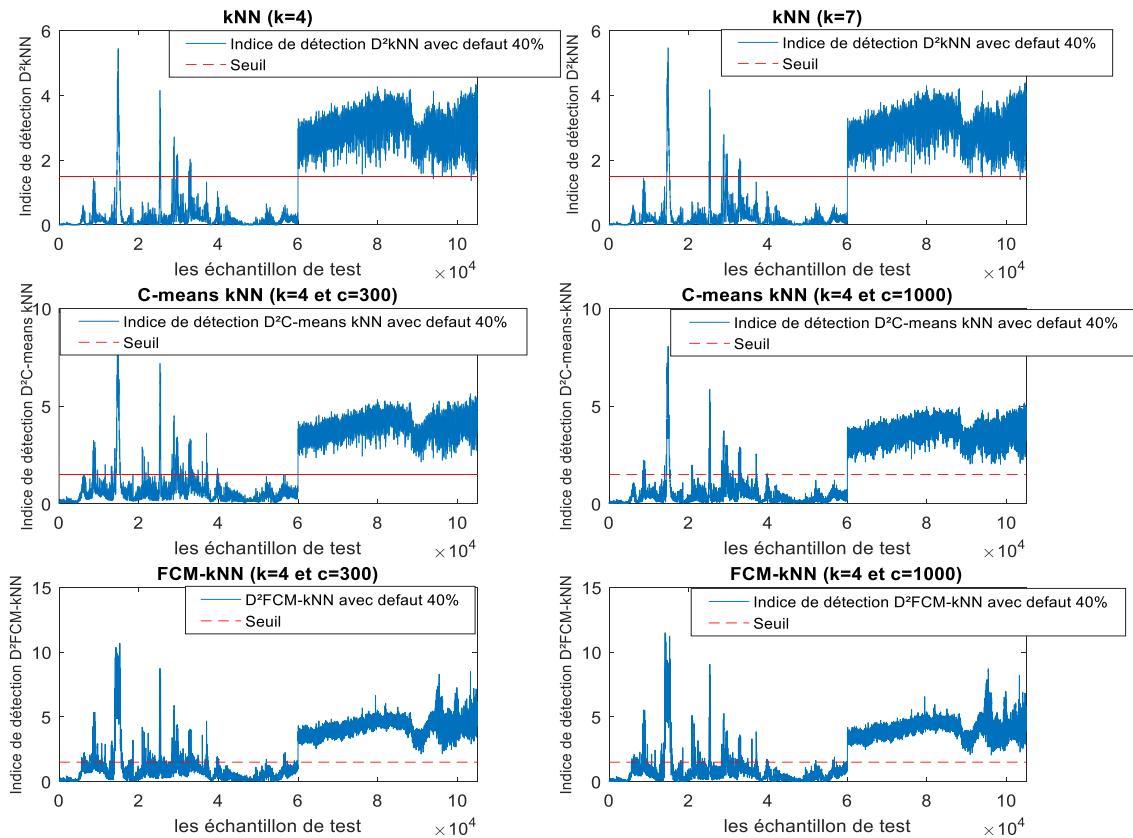


Figure 4.11 – Evolution des indices de détection avec default (X2)

Chapitre 4
Application (compresseur centrifuge)

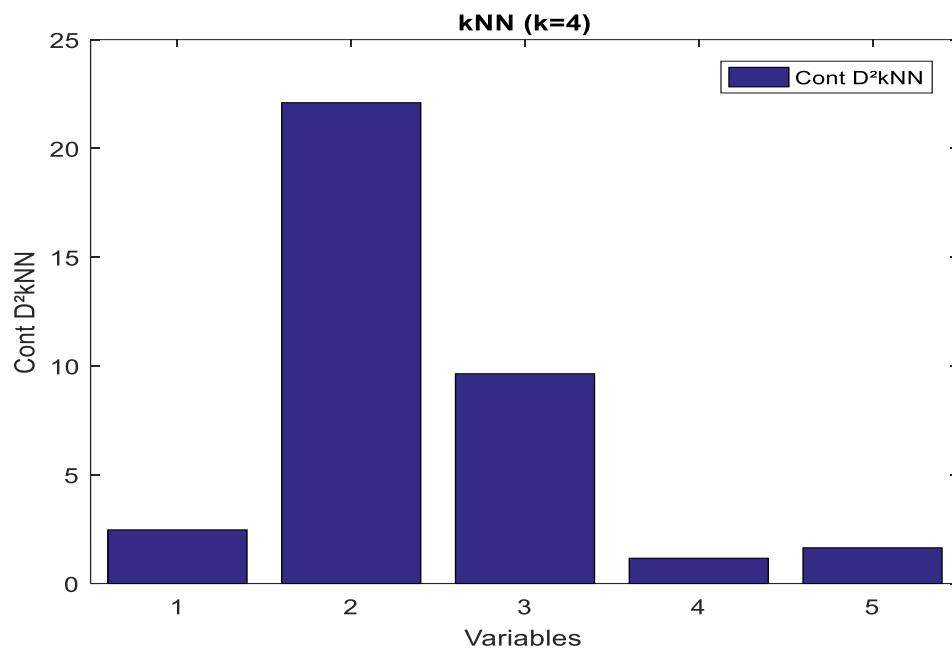


Figure 4.12 – Contributions des variables fautives du défaut X2 : kNN

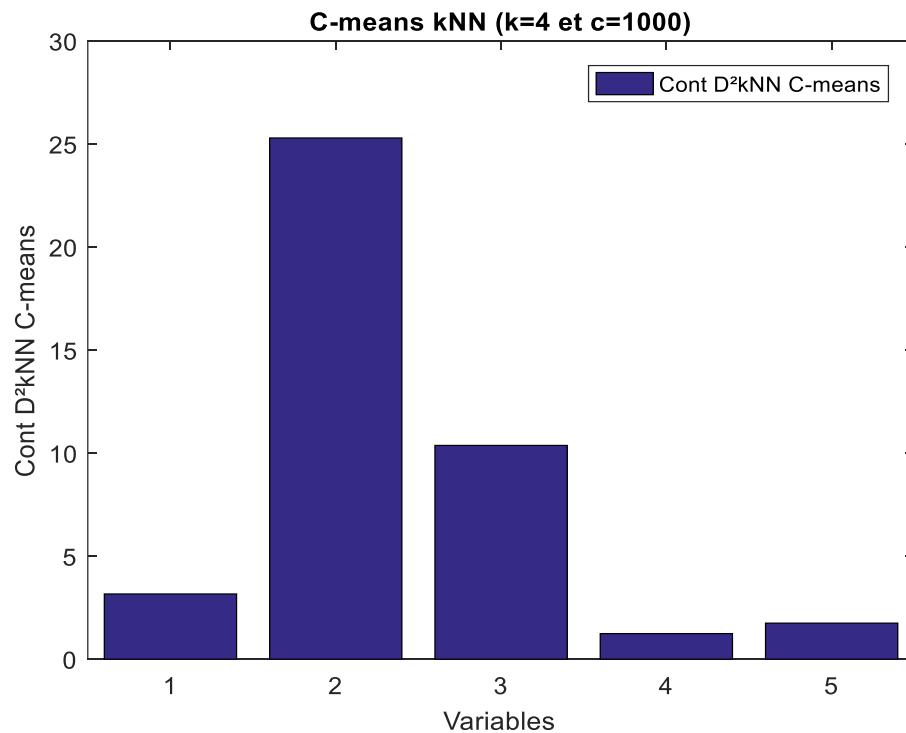


Figure 4.13 – Contributions des variables fautives du défaut X2 : C-means kNN

Chapitre 4
Application (compresseur centrifuge)

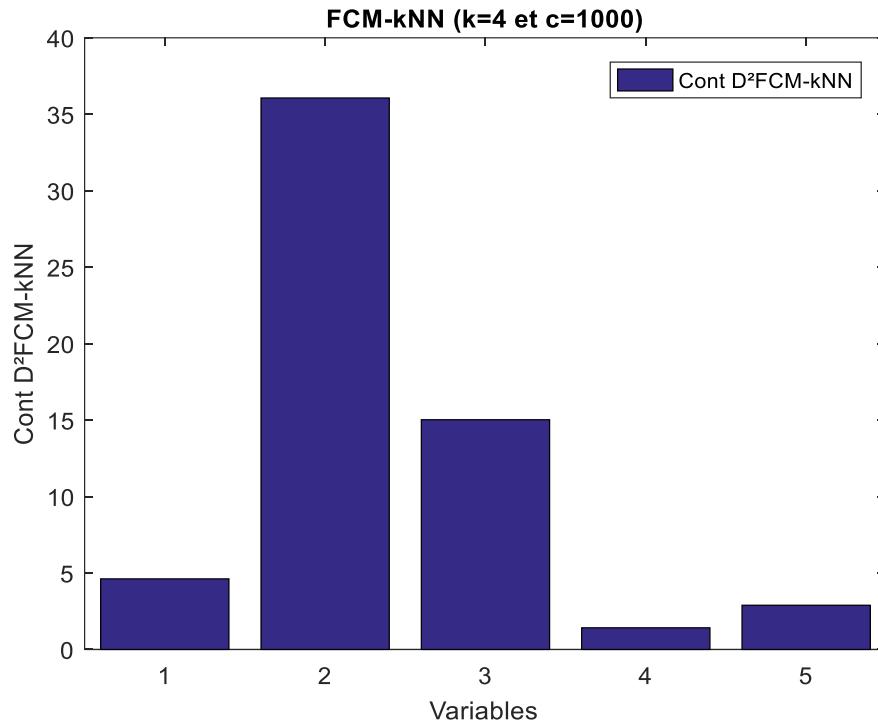


Figure 4.14 – Contributions des variables fautives du défaut X2 : FCM-kNN

Une étude comparative des algorithmes de mise à jour en ligne pour le diagnostic de default est menée sur les tableaux (4.2) et (4.3), la comparaison se concentrant sur le temps d'exécution de chaque algorithme et le taux de fausses alarmes (TFA).

- Le taux de fausses alarmes (TFA) : Cela reflète le scepticisme à l'égard du système de diagnostic, il est exprimé comme :

$$TFA = \frac{\text{nombre d'alarmes}}{\text{nombre d'observations saines}} \dots \dots \dots \dots \quad (4.1)$$

Nombre de k	kNN	C-means kNN			FCM-kNN		
		c=100	c=500	c=1000	c=100	c=500	c=1000
k=4	634.85s	16.76s	17.20s	18.32s	16.45s	17.20s	18.17s
k=7	637.94s	16.51s	17.59s	18.52s	16.35s	17.64s	18.14s
k=15	666.04s	16.98s	17.71s	18.69s	16.58s	17.40s	18.28s

Tableau 4.2 – Le temps d'exécution des algorithmes kNN, C-means kNN et FCM-kNN

Chapitre 4 Application (compresseur centrifuge)

TFA %	kNN	C-means kNN			FCM-kNN		
		c=100	c=500	c=1000	c=100	c=500	c=1000
k=4	1.21%	14.96%	2.89%	2.47%	47.34%	8.33%	7.31%
k=7	1.24%	38.80%	3.93%	3.04%	58.67%	15.07%	11.92%
k=15	1.35%	77.97%	6.68%	4.05%	74.51%	42.03%	29.07%

Tableau 4.3 – Le taux de fausses alarmes (TFA) des algorithmes kNN, C-means kNN et FCM-kNN

D'après les résultats présentés dans les tableaux (4.2) et (4.3), on remarque que lorsque le nombre de clusters augmente on obtient un bon taux de fausse alarme, et le taux de fausse alarme dans la méthode kNN être très faible par rapport aux autres méthodes, on a bien remarqué aussi qu'il y a une grande différence dans le temps d'exécution entre la méthode kNN classique et la méthode kNN combiner avec les méthodes de clustering, les deux méthodes qu'ils sont combiner avec kNN sont très rapide.

Donc l'approche kNN est vraiment lente, nous essayons de le faire rapidement, et au lieu de faire le calcul pour toutes la base de données, nous faisons le calcul pour les centres de cluster. Nous avons remarqué que le diagnostic étaient bons, nous avons donc amélioré le temps d'exécution tout en conservant les mêmes performances de diagnostic.

4.4 Conclusion

Dans ce chapitre, nous nous sommes consacrés à l'étude du comportement d'un compresseur de gaz K103-A1. Au cours de ce travail, une partie des bases de données sur ce sous-système est prise en compte, en a appliqué la méthode de kNN et la combinaison entre kNN et les méthodes de clusterings et déduire les différences entre les méthodes dans le diagnostic de default.

Conclusion générale

Dans le domaine du diagnostic des défauts, des méthodes basées sur le concept de redondance des informations ont été développées. Le diagnostic des défauts de processus à l'aide de méthodes d'analyse de données est une option plus prometteuse, car le but du diagnostic est de remarquer l'apparition de défauts, puis de découvrir la cause des defaults.

Un de nos objectifs a été d'utiliser la méthode de k plus proches voisins (kNN) pour appliquer cet algorithme dans le diagnostic de défauts, notamment dans la détection et la localisation des défauts, nous avons rencontré plusieurs inconvénients qui peuvent entraver le processus de diagnostic donc on a proposé des approches appelée les approches de clustering pour le combiner avec kNN, comme C-means et Fuzzy C-means , nous les avons combinés pour débarrasser les inconvénients que nous avons rencontré et garder la performance dans le diagnostic de default, dans lequel nous avons réduit les calculs et aussi on a amélioré le temps d'exécution dans lequel l'algorithme fonctionne rapidement et précisément. Nous avons appliqué des algorithmes développés sur un process réel pour l'étude du comportement du compresseur de gaz spécifique, le K103-A1. Nous avons utilisé la méthode du kNN ainsi que la combinaison entre le kNN et les méthodes de clustering pour diagnostiquer les défauts dans ce système. L'analyse des résultats obtenus nous a permis de tirer des conclusions sur les différences entre les méthodes et leur efficacité dans le diagnostic de défauts.

Dans les travaux futurs, il est recommandé d'explorer davantage les possibilités d'amélioration des modèles du kNN et du clustering en utilisant des algorithmes combinant différentes approches et en optimisant les procédures. L'intégration d'autres techniques d'intelligence artificielle, telles que l'apprentissage automatique, pourrait également être explorée pour améliorer les performances du diagnostic de défauts. En outre, l'application des méthodes étudiées à d'autres systèmes industriels et l'analyse comparative des résultats pourraient fournir des informations supplémentaires sur leur applicabilité et leur efficacité.

Bibliographie

- [1] Abdouramane Moussa Ali, Cédric Join, and Frédéric Hamelin. Diagnostic des systèmes dynamiques linéaires sans modèle explicite. *Journal Européen des Systèmes Automatisés (JESA)*, 45(Spécial : Nouvelles tendances en automatique) :339–362, January 2011
- [2] Ouabib Guenounou. Méthodologie de conception de contrôleurs intelligents par l'approche génétique. Application à un bioprocédé. *Automatique / Robotique*. Université Paul Sabatier - Toulouse III, 2009
- [3] Salowa METHNANI. Diagnostic, reconstruction et identification des défauts capteurs et actionneurs : application aux stations d'épurations des eaux usées. PhD thesis, École Doctorale E Mathématiques et Informatique [Équipe Estimation-Commande-Diagnostic (ESCODI)], 2012.
- [4] Y. A. Khulief, A. Khalifa, R. Ben Mansour, and M. A. Habib. Acoustic detection of leaks in water pipelines using measurements inside pipe. *Journal of Pipeline Systems Engineering and Practice*, 3(2) :47–54, may 2012.
- [5] Rolf Isermann. *Fault-Diagnosis Systems*. Springer Berlin Heidelberg, 2006.
- [6] Nassim Belkharchi. Contribution à l'étude du diagnostic et de la commande tolérante aux fautes par l'approche structurelle, PhD thesis, Université de Toulouse, 2011.
- [7] R. Isermann et P. Balle, “Trends in the Application of Model-Based Fault Detection and Diagnosis of Technical Processes”. *Control Engineering Practice*. Article 5(5), pp709-719, 1997.
- [8] Z. Ge, Review on data-driven modeling and monitoring for plant-wide industrial processes, *Chemom. Intell. Lab. Syst.* 171 (2017) 16–25, doi:10.1016/j.chemolab.2017.09.021.
- [9] Fix, E. & Hodges, J.L. Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties. Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.
- [10] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- [11] Alpaydin, E. (2010). *Introduction to machine learning* (2nd ed.). MIT Press.
- [12] Cours d'apprentissage supervisé de Marie Chavent | Université de Bordeaux
<https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn>
- [13] Q. Wang, Y.B. Liu, X. He, S.Y. Liu, J.H. Liu, Fault diagnosis of bearing based on KPCA and KNN method, in: *Advanced Materials Research*, vol. 986, Trans Tech Publ, 2014, pp. 1491–1496.

- [14] L.M. Elshenawy, C. Chakour and T.A. Mahmoud, Fault detection and diagnosis strategy based on k-nearest neighbors and fuzzy C-means clustering algorithm for industrial processes, Journal of the Franklin Institute, <https://doi.org/10.1016/j.jfranklin.2022.06.022>
- [15] E.L. Russell, L.H. Chiang, R.D. Braatz, Data-driven Methods for Fault Detection and Diagnosis in Chemical Processes, Springer-Verlag, London, 2000.
- [16] G. Verdier, A. Ferreira, Adaptive mahalanobis distance and k-nearest neighbor rule for fault detection in semiconductor manufacturing, IEEE Trans. Semicond. Manuf. 24 (1) (2010) 59–68.
- [17] S. Ding, Model-Based Fault Diagnosis Techniques, second ed., Springer-Verlag, London, 2013
- [18] W. Zhu, W. Sun, J. Romagnoli, Adaptive k-nearest-neighbor method for process monitoring, Ind. Eng. Chem. Res. 57 (7) (2018) 2574–2586.
- [19] Y. Wu, K. Ianakiev, V. Govindaraju, Improved k-nearest neighbor classification, Pattern Recognit. 35 (10) (2002) 2311–2318.
- [20] [Régression multiple \(soft-concept.com\)](#)
- [21] [Google Lens](#)