



MINISTRE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE ABBES LAGHROUR DE KHENCHELA
FACULTE DES SCIENCES ET DE LA TECHNOLOGIE



Département Mathématique et Informatique

N° de série :

Mémoire de fin d'études

Pour l'obtention du diplôme de Master (L.M.D)

Spécialité : Génie Logicielle Système Distribuée

Diagnostic Industriel

Présenté et soutenu publiquement par :

Ben Ammara Adel & Bechaa Nacer

Membres de jury :

Président Du Jury : Abd Elhadi Adel

Examineur: Mebarki Djemouai

Encadreur : Rafik Mahdaoui

Année universitaire : 2020/2021

Remerciements

Lorsque nous réussissons un travail et achevons nos buts c'est en grande partie grâce à toutes les personnes qui nous ont assistés durant notre travail.

C'est pour cela, on remercie Mr.rafik mahdaoui de nous avoir donné la plus grande aide pendant toute la durée de notre travail. De nous avoir procuré plus que des directions mais aussi d'avoir été disponible, clairvoyante et pour avoir été un vrai modèle par ses qualités humaines.

Nous présentons nos vifs remerciements à l'ensemble des membres du jury pour avoir pris le temps de rédiger et juger notre modeste travail et de nous aider à nous améliorer pour de futurs travaux.

Pour avoir été d'une aide précieuse, que ce soit professionnel ou émotionnel, nous réservons un remerciement spécial à nos parents en particulier et nos familles en général pour avoir été nos guides et pour avoir ajouté de la légèreté pendant la durée de notre travail.

nos remerciements ne seront pas complets sans mentionner nos amis nos remerciements aussi à toute personne qui nous a aidé de près ou de loin pour réussir ce modeste travail.

Merci de nous avoir inspirés et nous avoir encouragés à toujours faire de notre mieux.

Dédicaces

Avec un cœur plein de sentiments sincères et reconnaissants, je dédie ce modeste travail à :

La femme qui a dédié sa vie à moi en plus de donner naissance à mon existence, elle a tellement souffert pour garantir mon éducation et que je grandis pour être un homme honnête et responsable que dieu te protège de tout malheur ma chère mère

*L'homme qui a sacrifié son temps et que je porte son nom , mon père
a ma petite famille (ma femme et mes enfants)*

à ma grande famille benamara (mes frères et mes soeurs et leurs enfants)

*à toute personne que j'aime
je vous dédie ce modeste travail*

ADEL

Dédicaces

Je dédie ce travail :

A l'ensemble de ma famille et plus particulièrement à mes parents pour leur amour, leur confiance, leurs conseils ainsi que leur soutien inconditionnel qui m'a permis de réaliser les études pour lesquelles je me destine et par conséquent ce mémoire,

Merci Maman Merci papa pour être des parents modèles.

A mes chers frères et merci pour être la lumière de ma vie. A mon ami et binôme

*merci pour ton aide et support dans cette long
aventure qui a été une merveilleuse expérience dans ma vie.*

Merci à tous mes amis pour leur support et soutien moral.

Merci à tous.

NACEUR

Table des matières

Titre	N
Remerciement + dédicace	-
Introduction générale	1
CHAPITRE I : Le Diagnostic Industriel	2
I. Le diagnostic	4
I.1. Introduction	4
I.2. Définitions	6
I.3 Le diagnostic industriel dans le cadre de la supervision:	8
I.4. Comment aborder un problème de diagnostic des défaillances ?	9
I.5. Procédure de diagnostic	10
I.6. Conclusion	12
CHAPITRE II :LA FOUILLE DE DONNÉE ÉDUCATIVE	13
Introduction	14
1.1 Définitions	15
1.1.1 Fouille de données (Data maining)	16
1.1.2 Données	16
1.1.3 Information	17
1.1.4 Connaissance	17
1.1.5 Entrepôts de données	19
1.2 Processus d'Extraction des Connaissances à partir des Données (ECD)	19
1.3 Tâches d'exploration de données	20
1.3.1 Classification et prédiction	20
1.3.2 Clustérisation	20
1.3.3 Les règles d'association	20
1.4 Applications d'exploration de données	20
1.4.1 Affaires et argent	21
1.4.2 La médecine	21
1.4.3 Marketing et vente	21
1.4.4 L'aspect éducatif	21
1.5 Fouille de données éducatives (Educational Data Mining)	21
1.5.1 Définition	21
1.5.2 Objectifs de la fouille de données éducatives	22
1.5.3 Tâches EDM	22
1.5.4 les méthodes	23
1.6 outils de la Fouille de données	23

1.6.1 logiciels libres	27
1.6.2 logiciels commerciaux.....	28
1.7 conclusion	25
CHAPITRE III :LA FOUILLE DE DONNÉE ÉDUCATIVE	29
1. Introduction	30
2. Présentation de la SCIMAT	30
3. Etapes de Fabrication du ciment	31
4. Atelier de clinkérisation	37
5. La modélisation	39
6. CONCLUSION	49
CONCLUSION GENERALE	50
REFERNCES BIBLIOGRAPHIQUES	52

Résumé

L'extraction des connaissances à partir des données est définie comme le processus d'analyse des données sous différentes perspectives et de découverte des modèles à partir des ensembles de données utiles pour prédire les résultats qui nous aident à prendre la bonne décision.

Comme ce processus passe par plusieurs étapes, de la (collecte des données. Nettoyage et transmission) à l'obtention des résultats, leur validation et leur interprétation, et enfin la fusion des connaissances acquises.

Il existe de nombreuses techniques d'analyse de données qui peuvent être utilisées pour extraire des modèles, parmi lesquelles des techniques de classification qui génèrent des modèles (phase d'apprentissage) utilisés pour prédire les données futures d'un objet (phase de prédiction) à partir de données provenant d'autres objets similaires.

Notre travail est lié à cet aspect. Nous avons d'abord collecté les données et nous avons ensuite généré plusieurs modèles qui permettent de classer les performances en deux types (fonctionnement normal, présence de défaillances).

Notre travail est concrétisé par une application développée sous Java et Weka qui permette de faire la saisie des nouvelles données, l'apprentissage pour la création des modèles de classification.

Introduction générale

Introduction générale

L'un des grands problèmes dans les systèmes de production industriels et comment nous pouvons détecter une faute, défaillance et une panne dans un délai très court ainsi avec une crédibilité très élevée.

Donc pour cela les experts on essayer de faire des programmes informatique qui aident les experts ainsi que les utilisateurs et le techniciens des systèmes de production de détecter des défaillance d'un manière propre.

L'objectif des systèmes d'aide au diagnostic industriel est de trouver à chaque instant le système ou se trouve ,dans une situation de défaillance ou on bon fonctionnement,

Donc, pour cela plusieurs techniques utilisées pour détecter des défaillances dans les installations industrielles, parmi ces techniques on trouve les techniques de l'intelligence artificielle et exactement les techniques de data maining

L'objectif de ce mémoire est modéliser et simuler un systèmes d'aide au diagnostic industriel basé sur les techniques de data maining.

CHAPITRE I

Le diagnostic Industriel

I. Le diagnostic :

I.1. Introduction :

Dans le monde industriel gains de productivité représentent un souci quotidien pour les dirigeants des entreprises. Ainsi les petites et moyennes entreprises réalisent, dans un contexte concurrentiel très compétitif, un chiffre d'affaires qui représente plus de 60% du chiffre d'affaires de l'industrie manufacturière.

Le diagnostic de défaillance des systèmes industriels, s'il est réalisé avec efficacité ou s'il permet de détecter de façon précoce une dégradation, représente un des moyens pour contribuer à gagner des points de productivité ; a cause de l'importance de diagnostic dans la maintenance corrective, de sa pertinence et de sa rapidité dépend l'efficacité de l'intervention au sein de l'entreprise. Pour cela, des recherches ont été effectuées autour du problème du diagnostic, que ce soit pour des systèmes dynamiques, des réseaux de distribution ou des réseaux de télécommunications.

I.2. Définitions :

En médecine, où il trouve son origine, le diagnostic est « *la partie de l'acte médical qui vise à déterminer la nature de la maladie observée. ... ; devant l'urgence thérapeutique, le médecin devra alors décider si et quand les documents qui s'accumulent permettent l'arrêt de l'investigation et le passage à une conclusion, qui reste forcément révisable, car une recherche de plus en plus raffinée de signes pourrait parfois se poursuivre indéfiniment* ».

Non loin de cette définition, dans le domaine de l'industrie, Charbonnaud définit le diagnostic comme étant « *un acte intelligent qui se doit de trouver dans un système physique les dysfonctionnements et surtout les causes de la situation de panne* ».

Etymologiquement, le mot diagnostic vient du grecque et signifie (Dia : par, Gnosis : connaissance). La définition donnée par (Peng, 1990), s'en inspire et se formule ainsi :

«Etant donné un ensemble de manifestations observées (symptômes, constatations, etc.), il s'agit d'expliquer leur présence, de remonter aux causes, en utilisant un savoir sur le système considéré».

Selon la définition retenue par l'AFNOR (Association Française de Normalisation) le Diagnostic industriel est : «...l'identification de la cause probable de la (ou les) défaillance(s) à l'aide d'un raisonnement logique fondé sur un ensemble d'informations provenant d'une inspection, d'un contrôle, ou d'un test » [NF X60-010].

Cette définition résume les deux tâches essentielles en diagnostic :

- Observer les symptômes de la défaillance ;
- Identifier la cause de la défaillance à l'aide d'un raisonnement logique fondé sur des observations (figure .I.1)

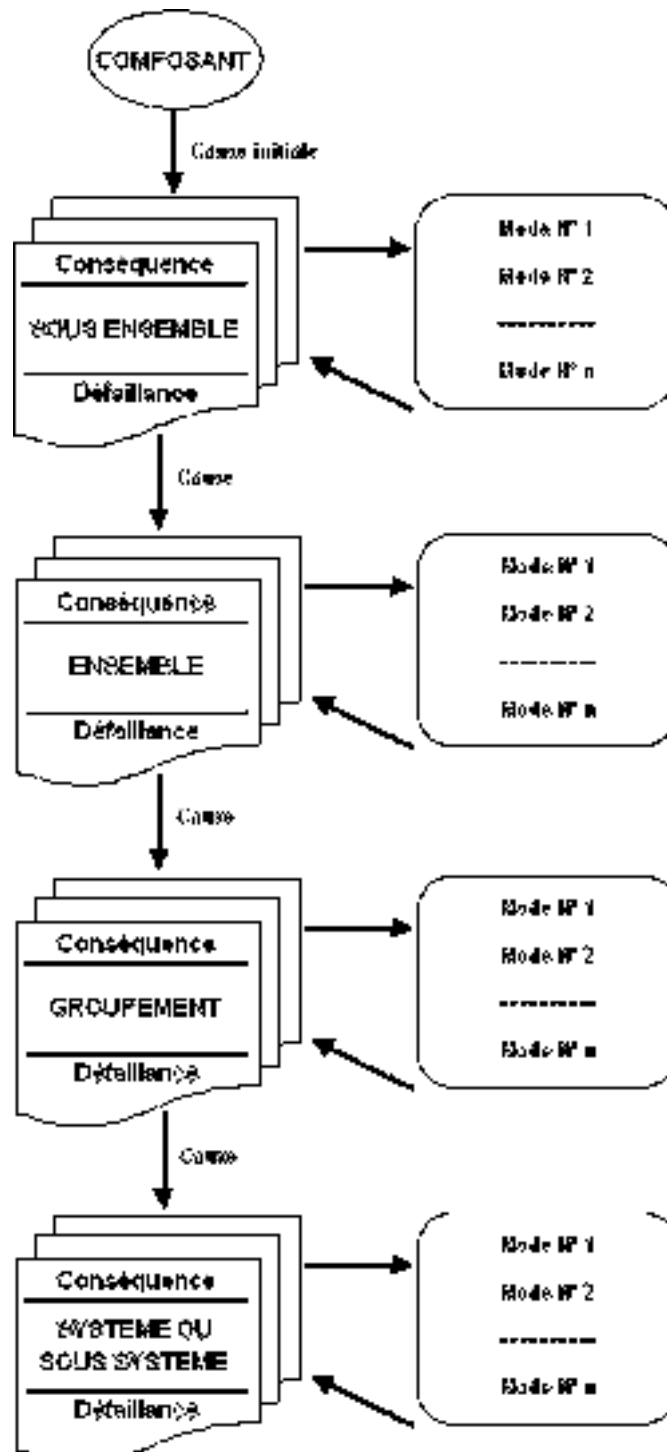


Fig.I.1. Le chemin de diagnostic

I.3 Le diagnostic industriel dans le cadre de la supervision:

Pour la mise en place d'un système de supervision, du point de vue de la communauté du continu, trois fonctions doivent être prises en compte : la *détection*, le *diagnostic* et la *reconfiguration*, la surveillance du procédé traite les données disponibles en ligne, afin d'obtenir son état de fonctionnement. Dans la surveillance, nous retrouvons les fonctions de *détection* de défaillances et de *diagnostic* (figure. I.2). De façon générale, la détection consiste en l'identification des changements ou déviations des mesures du procédé par rapport au fonctionnement normal, ce qui se traduit par la génération des symptômes. Le diagnostic consiste à déterminer quelles sont l'origine et/ou la (les) cause(s) qui ont pu engendrer le symptôme détecté. A ce stade, le système doit avoir la capacité de décider quand le procédé se trouve dans une situation de fonctionnement normal, et quand une action corrective doit être appliquée. Cette action corrective correspond à l'étape de reconfiguration de la commande de façon à ramener le procédé dans un mode de fonctionnement normal.

Cependant, il existe d'autres approches pour la mise en place d'un système de *supervision*. En effet, pour la communauté de systèmes à événements discrets (SED), la supervision a pour but de contrôler l'exécution d'une opération effectuée par le système de commande sans rentrer dans les détails de cette exécution. La supervision a lieu dans une structure hiérarchique (au moins avec 2 niveaux), et recouvre l'aspect du fonctionnement normal et anormal :

- En fonctionnement normal, le rôle de la supervision est de prendre, en temps réel, les décisions correspondantes aux degrés de liberté exigés par la flexibilité décisionnelle;
- En présence de défaillances, la supervision prend toutes les décisions nécessaires pour le retour vers un fonctionnement normal. Après avoir déterminé un nouveau mode de fonctionnement, il peut s'agir de choisir une solution curative, d'effectuer des ré-ordonnements "locaux", ou même de déclencher des procédures d'urgence.

La surveillance est responsable de l'acquisition des signaux en provenance des ressources et de la commande. Ces informations sont utilisées pour la reconstitution de l'état réel du système commandé et pour faire les inférences nécessaires afin de produire des informations supplémentaires pour dresser des historiques de fonctionnement. Les activités de la surveillance sont donc limitées aux fonctions relatives aux informations et n'ont pas une

action directe sur le modèle ni sur le procédé. A priori, la surveillance a un rôle passif vis-à-vis de la commande. Parmi les fonctions de la surveillance nous trouvons donc, en plus de l'acquisition de données, la détection [COM00] qui caractérise le fonctionnement du système de normal ou anormal. Deux classes d'anomalies sont distinguées :

- La première regroupe les situations pour lesquelles le comportement du système devient anormal car les contraintes d'opération ne sont pas garanties;
- La deuxième regroupe les situations dans lesquelles le comportement est anormal par rapport à la loi de commande appliquée. Cette classe recouvre les anomalies de fabrication mise en évidence par des contrôles de qualité.

Le **suivi** fait partie aussi des fonctions de la surveillance. Cette fonction maintient en permanence un historique des traitements effectués, et une trace des événements que perçoit le système. La fonction **diagnostic** établit un lien de cause à effet entre un *symptôme* observé et la défaillance qui est survenue, ses causes et ses conséquences. La communauté des SED distingue trois sous-fonctions [COM00]:

- La localisation, qui détermine le sous-système responsable de la défaillance,
- L'identification, qui détermine les causes qui ont engendré la défaillance,
- L'explication, qui justifie les conclusions du diagnostic.

La fonction pronostic est également une fonction de surveillance qui a pour but de déterminer les conséquences d'une défaillance sur le fonctionnement futur du système. Il existe aussi des fonctions propres à la supervision :

- La reconfiguration qui agit sur le procédé en adaptant la configuration matérielle à la situation, ainsi que sur le système de commande en changeant la loi de commande,
- La décision, qui détermine l'état à atteindre pour le retour en opération normale et la séquence d'actions correctives à réaliser pour arriver à cet état.

Dans ces conditions, la supervision n'est plus un simple enchaînement des fonctions de surveillance (détection - diagnostic) et reconfiguration. Des "modèles de surveillance" beaucoup plus complexes peuvent être mis en place en fonction du procédé considéré, du type de défaillances et également de la politique de production de l'entreprise [ZAM98].

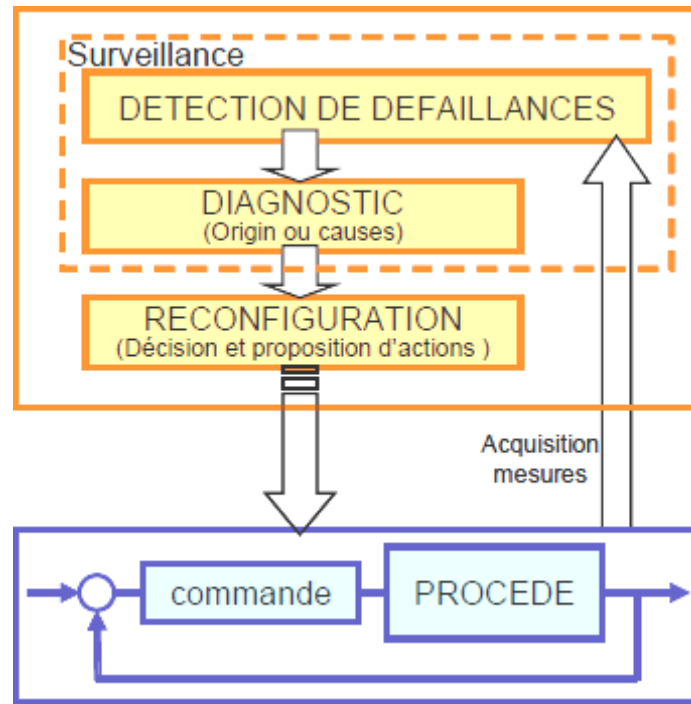


Fig.I.2. schéma général de supervision

I.4. Comment aborder un problème de diagnostic des défaillances ?

Le diagnostic est une chaîne complexe où aucune action n'est indépendante de celle qui la précède comme de celle qui la suit [Dubuisson, 90]. Il demande l'exploitation de toute la connaissance "accessible" existante sur le système. Il convient de distinguer la connaissance "accessible" et la connaissance "disponible". La première repose sur les instruments d'observation dont on dispose, qu'ils agissent sur le présent ou sur le passé, c'est -à- dire qu'elle est basée sur toute l'information que recueillent les capteurs. La seconde quant à elle, concerne l'ensemble des informations qu'on pourrait atteindre s'il n'existait aucune limitation physique ou technologique.

La conception des solutions relatives aux problèmes posés doit répondre aux trois points essentiels de la surveillance : la détection, le diagnostic, la reprise. Ceci nécessite l'élaboration de modèles de bon fonctionnement et, également de modèles de pannes dans la majorité des cas.

La détection est pratiquement fondée sur une comparaison, en temps réel, du comportement effectif du système surveillé avec celui d'un modèle de bon comportement.

La détection consiste à remarquer une incohérence entre le comportement du système physique et celui du modèle.

Le diagnostic, lui, est souvent fondé sur la recherche d'une cohérence entre le comportement (ou l'état) du système physique et celui d'un modèle de panne correspondant à des connaissances profondes.

La pratique du diagnostic industriel s'inscrit dans le processus de conduite de l'entreprise. En effet, les activités de conduite et de maintenance du processus industriel conduisent à entreprendre en pratique des opérations de diagnostic qui sont de nature très différentes et souvent conditionnées par la spécificité des systèmes et des moyens disponibles.

Les systèmes industriels sont caractérisés par une complexité importante (technologie d'automatisation, interactions importantes opérateurs-processus) qui complique davantage la tâche de l'opérateur chargé du diagnostic. Pour pallier cette difficulté, l'aide informatique devient indispensable et un large éventail de techniques ont été utilisées pour automatiser le diagnostic des systèmes industriels. A chaque type de diagnostic est associé un ensemble de techniques utilisables.

I.5. Procédure de diagnostic :

La procédure de diagnostic de défaillances et de dégradations susceptibles d'affecter les différentes entités d'un processus industriel s'articule autour des étapes suivantes :

- ↳ L'extraction des informations nécessaires à la mise en forme des caractéristiques associées aux fonctionnements normaux et anormaux, à partir de moyens de mesures appropriées ou d'observations réalisées lors des rondes par les personnels de surveillance,
- ↳ L'élaboration des caractéristiques et signatures associées à des symptômes révélateurs de défaillances et de dégradations en vue de la détection d'un dysfonctionnement,
- ↳ La détection d'un dysfonctionnement par comparaison avec des signatures associées à des états de fonctionnements normaux et la définition d'indicateurs de confiance dans la détection,
- ↳ La mise en œuvre d'une méthode de diagnostic de la défaillance ou de la dégradation à partir de l'utilisation des connaissances sur les relations de cause à effet,
- ↳ La prise de décision en fonction des conséquences futures des défaillances et des dégradations. Cette prise de décision peut conduire à un arrêt de l'installation si les

conséquences de la défaillance sont importantes pour la sécurité des personnes et des biens ou à une reconfiguration du fonctionnement du procédé pour éviter une perte de production en attendant le prochain arrêt de production le plus propice aux opérations de maintenance corrective.

La figure I.3 représente l'ensemble des tâches à réaliser pour assurer un fonctionnement satisfaisant d'un processus industriel.

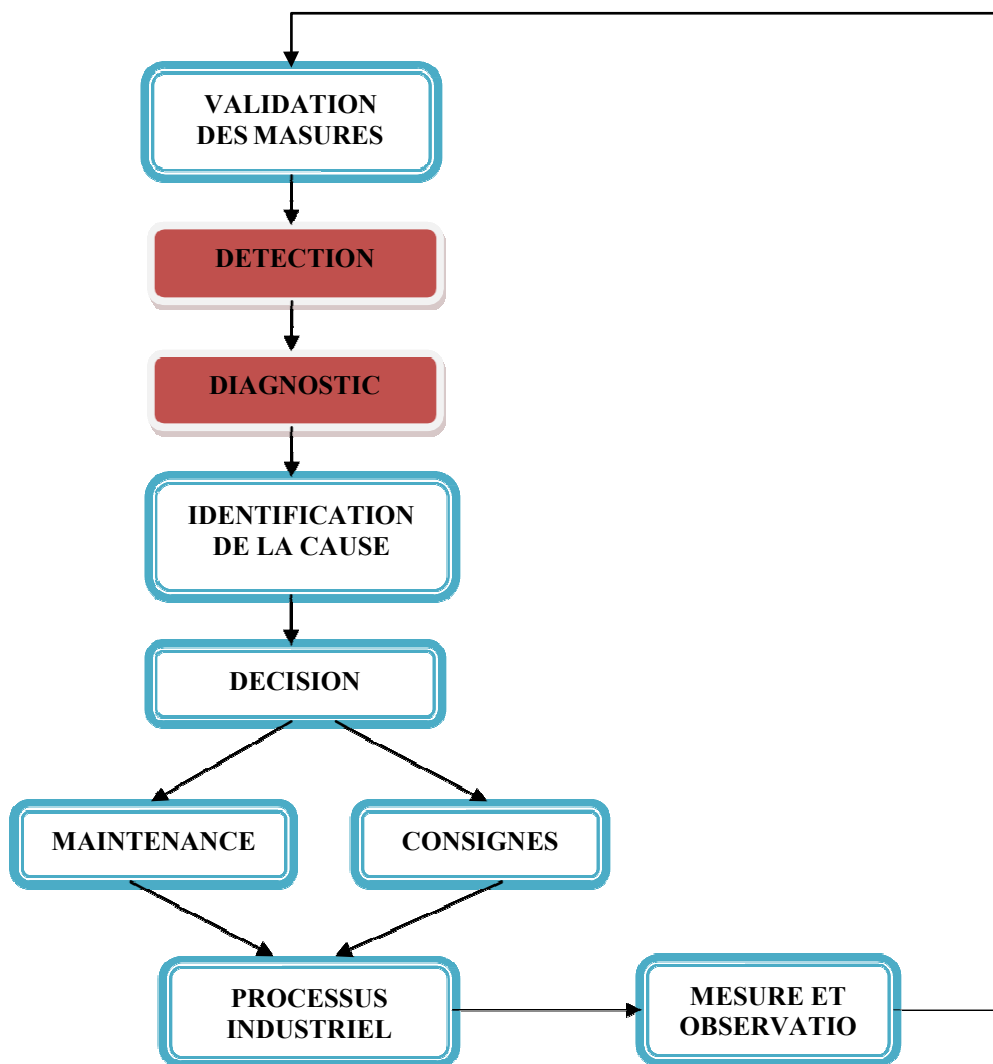


Fig.I.3. L'action de diagnostiquer

Ces étapes si elles sont nécessaires ne sont cependant pas suffisantes car il faut aussi prendre en compte les conséquences de leur mise en place dans l'organisation de l'entreprise. En effet, il convient au préalable de prendre des dispositions pour remplir les meilleures conditions pour réaliser un système d'aide au diagnostic efficace et utilisé.

- ↳ Etude de l'intérêt technico-économique de l'implantation du système d'aide au diagnostic en prenant en compte le retour d'expérience des coûts des défaillances et leurs impacts sur la productivité de l'investissement,
- ↳ Etude de la faisabilité technique du système d'aide au diagnostic pour s'assurer que des techniques et technologies sont disponibles et opérationnelles pour le problème à résoudre,
- ↳ Création d'une équipe de projet avec tous les intervenants pour définir et rédiger le cahier des spécifications détaillées,
- ↳ Validation du cahier des charges par les responsables de l'entreprise pour le lancement de la réalisation et de son implantation,
- ↳ Formation et informations des acteurs pour une utilisation effective du système d'aide au diagnostic,
- ↳ Réalisation et implantation du système d'aide au diagnostic sur site,
- ↳ Mise en place d'un retour d'expérience pour évaluer les impacts économiques et sociaux du système d'aide au diagnostic.

I.6. Conclusion :

Le diagnostic industriel des systèmes de production en est une solution qui pourrait aider les experts des systèmes à détecter et diagnostiquer des pannes, des défaillances et même des défauts dans une durée très courte afin de minimiser les coûts de la maintenance industrielle. Pour cela nous avons abordés dans ce premier chapitre les entraves de diagnostic industriel ainsi que l'ensemble des méthodes utilisées et d'autres techniques de fouille de données qui sera bien présentée dans le chapitre suivant.

CHAPITRE II

La fouille de donnée éducative

Introduction

Durant ces dernières années, avec l'augmentation sans cesse de capacité de stockage des ordinateurs, les quantités de données collectées, dans divers domaines d'application de l'informatique, deviennent de plus en plus importantes, donc, nous parlons de Big Data, nous parlons de quantités inimaginables de données de nombreux types et sources de tailles énormes. Nous demandons ici à quel point ces données sont importantes à la lumière du fait qu'elles indiquent que les informations structurées pour ces données ne constituent qu'une petite fraction de 10% par rapport aux informations non organisées qui composent le reste. Cela a accru le besoin de développer des outils puissants pour l'analyse des données et l'extraction d'informations et de connaissances. Les méthodes conventionnelles et statistiques ne peuvent pas gérer cette énorme quantité, donc des outils intelligents sont utilisés pour traiter ces données. Ce chapitre a pour objet est de présenter dans un premier temps les concepts liés à la fouille de données. Dans un second temps, il présente les techniques de data mining qu'on peut utiliser pour l'extraction des connaissances à partir des données éducatives.

1.1 Définitions

1.1.1 Fouille de données (Data Mining)

C'est une technologie qui vise à extraire des connaissances à partir d'énormes quantités de données, basées sur des algorithmes mathématiques qui constituent la base de l'extraction de données et dérivent de nombreuses sciences telles que les statistiques, les mathématiques, la logique, les sciences de l'apprentissage, l'intelligence artificielle et les systèmes experts. L'exploration de données est apparue à la fin des années 1980 et s'est révélée être une solution efficace pour analyser de grandes quantités de données et les convertir à partir d'informations accumulées et simplement incompréhensibles en informations précieuses qui pourraient être utiles pour une exploitation et une utilisation ultérieures.[1]

Une autre définition

La fouille de données est le processus d'analyse des données sous différentes perspectives et de découverte des déséquilibres, des modèles et des connexions dans les ensembles de données qui est perspicace et utile pour prédire les résultats qui vous aide à prendre une décision correcte.

Le processus d'exploration de données est basé sur les trois éléments de base suivant :

1.1.2 Données

Selon wikipedia : « Une donnée est une description élémentaire d'une réalité. C'est par exemple une observation ou une mesure ». Donc, les données sont des faits, des nombres, ou des textes pouvant être traités par un ordinateur. Aujourd'hui, les entreprises accumulent de vastes quantités de données sous différents formats, dans différentes quantités de données. Parmi ces données, on distingue[4] :

- Les données opérationnelles ou transactionnelles telles que les données de

ventes, de coûts, d'inventaire, de tickets de caisse ou de comptabilité.

- Les données non opérationnelles, telles que les ventes industrielles, les données prévisionnelles, les données macro-économiques.
- Les métadonnées, à savoir les données concernant les données elles-mêmes, telles que les définitions d'un dictionnaire de données.

Type de données

- Données quantitatives : Valeurs numériques et sommables, discrètes, continues.
- Données qualitatives :

Ordinales : Ex : petit, moyen, grand, très grand.

Nominales : (catégories ou modalités), Ex : féminin, masculin, célibataire, marié, divorcé, veuf. . .

1.1.3 Information

Selon wikipedia : Au sens étymologique, l'information est ce qui donne une forme à l'esprit. Elle vient du verbe latin « informare », qui signifie « donner forme à » ou « se former une idée de ». L'information est aussi une notion abstraite, mais d'un niveau d'abstraction supérieur à celui de la donnée. On peut dire pour simplifier que l'information est une donnée + un sens

1.1.4 Connaissance

Selon wikipedia : « La connaissance est une notion aux sens multiples à la fois utilisée dans le langage courant et objet d'étude poussée de la part des philosophes contemporains ». La connaissance est aussi une notion abstraite, d'un niveau d'abstraction supérieur à celui de l'information. La connaissance à la différence de l'information est partagée et s'appuie sur un référentiel collectif. Une connaissance est une information nouvelle, apprise par association d'informations de base, de règles, de raisonnement, d'expérience, d'expertise, etc.[6]

Processus de découverte des connaissances

Certaines personnes ne différencient pas l'exploration de données de la découverte de connaissances tandis que d'autres considèrent l'exploration de données comme une étape essentielle du processus de découverte de connaissances. Voici la liste des étapes impliquées dans le processus de découverte des connaissances :

- **Nettoyage des données** : Dans cette étape, le bruit et les données incohérentes sont supprimés.
- **Intégration des données** : Dans cette étape, plusieurs sources de données sont combinées.
- **Sélection des données** : Dans cette étape, les données pertinentes pour la tâche d'analyse sont extraites de la base de données.
- **Transformation des données** : Dans cette étape, les données sont transformées ou consolidées sous des formes appropriées pour l'exploration en effectuant des opérations de résumé ou d'agrégation.
- **Exploration de données** : Dans cette étape, des méthodes intelligentes sont appliquées afin d'extraire des modèles de données.
- **Évaluation de modèle** : Dans cette étape, les modèles de données sont évalués.
- **Présentation des connaissances** : Dans cette étape, les connaissances sont représentées.

1.1.5 Entrepôts de données

Il est utilisé dans les analyses temporelles, la découverte des connaissances et la prise de décision. Une énorme quantité de données peut être stockée qui peut être de différentes sources, par exemple plusieurs bases de données de plusieurs modèles, elles sont conçues pour extraire des données, les traiter, les représenter et les présenter de manière appropriée à ces fins.[2]

1.2 Processus d'Extraction des Connaissances à partir des Données (ECD)

Le processus d'extraction des connaissances à partir de données présenté en Figure[5]

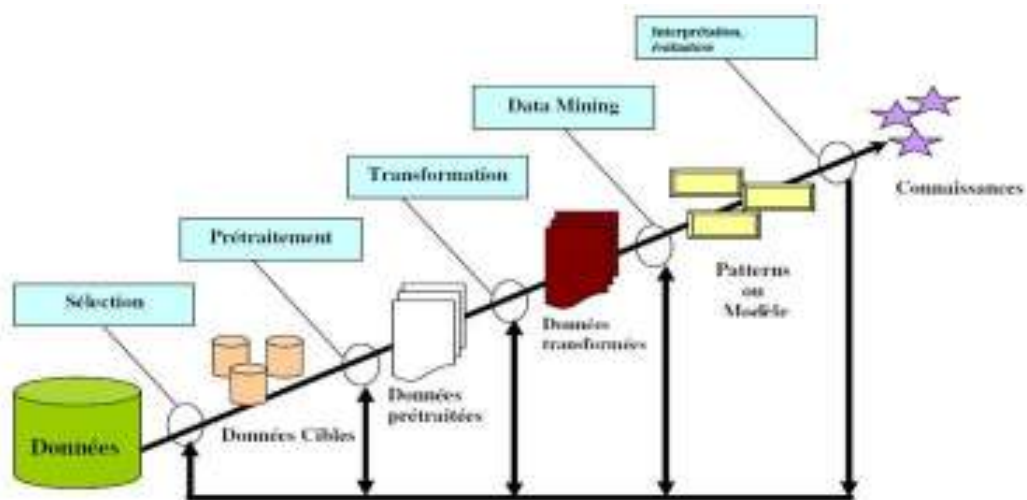


Figure 1.1 – Processus (ECD)

Pour extraire des connaissances des données, nous suivons plusieurs étapes :

- ♠ La sélection des données.
- ♠ Le prétraitement des données.

- ♠ La transformation de données.
- ♠ Fouille de données (Data Mining).
- ♠ Evaluation et interprétation des connaissances.

1. **La sélection des données** : L'objectif de l'extraction de connaissance est de déduire des nouvelles connaissances valides et utiles. Cet aspect est très important, mais on ne peut pas appliquer le processus d'ECD sur toutes les données qu'on a, donc le besoin est exprimé par l'utilisateur. Ce dernier fait la sélection des données selon l'objectif visé.

Cette étape concerne donc le filtrage des données qui comprend deux opérations nécessaires :

- *La réduction de la dimensionnalité des données* : l'élimination d'attributs sans intérêt, ou ayant beaucoup de valeurs erronées ou manquantes.
 - *La réduction de la taille des données* : l'application des techniques du Data Mining est très coûteuse en terme de temps CPU et d'espace mémoire, c'est pour cela qu'on ne peut pas l'appliquer sur la totalité des données.
2. **Le prétraitement des données** : Le rôle de cette étape est de préparer les données afin qu'elles soient de meilleures qualités afin d'arriver à des résultats de qualité. Le prétraitement des données concerne, entre autres, le nettoyage des données, c'est-à-dire l'élimination du bruit, ainsi que le traitement des valeurs manquantes, ou erronées. Il faudrait alors définir les méthodes à utiliser pour le remplacement de ces valeurs.
 3. **La transformation de données** : Cette étape consiste à préparer les données brutes, et à les convertir en données appropriées. La transformation se fait par attribut, c'est-à-dire toutes les valeurs d'un attribut doivent être transformées en un format unique. Formellement, un attribut « A » est transformé en « B » qui serait utilisable par la tâche de la fouille de données choisie.
 4. **Fouille de données (Data Mining)** : Dans cette étape, des méthodes intelligentes sont utilisées afin d'extraire des modèles ou patterns. Cette étape est aussi désignée comme l'étape cœur du processus d'ECD . Il est clair que les étapes qui précèdent

la fouille de données sont très importantes, car la qualité des modèles ou patterns extraites, ainsi que leur coût d'extraction sont liés directement à ces étapes.

5. **Evaluation et interprétation des connaissances** : Les modèles ou patterns extraits ne sont pas dans la plupart du temps exploitables. En effet, il est difficile d'avoir directement des connaissances valides et utiles, à ce point-là. Il existe, cependant, des méthodes d'évaluation des modèles extraits. Ces méthodes peuvent aussi aider à corriger les modèles, et à les ajuster aux données.

Ces principales étapes de processus (ECD) visent à partir de données volumineuses l'extraction des connaissances, qui peuvent être exprimés sous forme d'un concept général qui enrichit le champ sémantique de l'utilisateur par rapport à une question qui le préoccupe. Elles peuvent prendre la forme (figure 1.1)

1.3 Tâches d'exploration de données

1.3.1 Classification et prédiction

La classification consiste à traiter un ensemble de modèles (ou fonctions) qui décrivent et distinguent les classes de donnée, afin de pouvoir utiliser le modèle pour prédire la classe d'objets dont l'étiquette de classe est inconnue. Le modèle dérivé est basé sur l'analyse d'un ensemble de données d'apprentissage (c'est-à-dire des objets de données dont l'étiquette de classe sont connus). La classification et la prévision peuvent devoir être précédées d'une analyse de pertinence qui tente d'identifier les attributs qui ne contribuent pas au processus de classification ou de prédiction. Ces attributs peuvent ensuite être exclus

1.3.2 Clustérisation

Contrairement à la classification et à la prédiction, qui analyse les objets de données nommés dans une classe, l'agrégation analyse les objets de données sans référence à une étiquette de classe connue. En général, les étiquettes de classe n'étaient pas présentes dans les données de formation simplement parce qu'elles n'étaient pas connues au début. L'agrégation peut être utilisée pour créer de telles étiquettes. Les objets sont regroupés selon le principe de maximiser la similitude entre les catégories et de réduire la similitude entre les classes.

Autrement dit, des groupes d'objets sont formés de sorte que les objets d'un groupe sont très similaires aux autres groupes, mais sont très différents des autres objets du groupe.

1.3.3 Les règles d'association

Les règles d'association sont parmi les méthodes de recherche des implications (relation) entre les attributs d'une base de données. Les règles d'association ont été introduites par Agrawal et al. L'étymologie des règles d'associations, aussi connues sous le nom de l'analyse du panier de la ménagère (Market Basket Analysis), vient des travaux qui ont été réalisés à partir des données provenant des supermarchés. L'objectif de ces travaux est d'identifier les items ou les groupes d'items (Itemset), fréquemment achetés ensemble par un client lors d'une même transaction où son support est supérieur ou égal à un seuil minimum, (Minsupp), fixé par l'utilisateur.[5]

1.4 Applications d'exploration de données

Il existe de nombreux exemples d'applications d'exploration de données dans divers domaines, notamment les suivants :

1.4.1 Affaires et argent

L'exploration de données, par exemple, est utilisée pour anticiper la capacité des clients à rembourser leurs prêts financiers. Cela se fait en appliquant des algorithmes d'exploration de données aux enregistrements historiques des clients précédents, résultant en une forme ou un ensemble de règles qui déterminent si le client peut rembourser son prêt ou non.

1.4.2 La médecine

Une grande présence de l'exploration de données est les solutions que vous fournissez, telles que l'attente d'un patient atteint de certaines maladies en fonction de ses données et de son dossier médical, ou l'étendue des effets des médicaments et des médicaments sur les patients en fonction des dossiers des patients précédents.

1.4.3 Marketing et vente

Les techniques d'exploration de données aident les sociétés de marketing à créer des modèles basés sur des données historiques pour de nouvelles campagnes de marketing telles que le publipostage et les campagnes de marketing Internet. Grâce aux résultats, les sociétés de marketing disposeront d'une méthode appropriée pour vendre des produits rentables à des clients cibles. L'exploration de données aide les entreprises de vente au détail, grâce à l'analyse du panier de marché, à ce que le magasin dispose d'une gestion appropriée afin que les clients puissent toujours acheter ensemble les produits qu'ils achètent fréquemment. De plus, il aide les détaillants à offrir certaines remises pour des produits spécifiques susceptibles d'attirer plus de clients.

1.4.4 L'aspect éducatif

L'exploration de données est utilisée pour anticiper la performance académique des étudiants sur la base des données de performance des étudiants précédents. Il est également possible de prédire la performance des enseignants ou leur capacité à donner des cours spécifiques. De nombreuses études et recherches sont menées dans ce domaine, et les applications d'exploration de données sont appelées dans le domaine éducatif (Educational Data Mining). *L'exploration de données a un rôle pivot dans tous les domaines scientifiques et pratiques. Son existence est une nécessité des systèmes et des applications.*

1.5 Fouille de données éducatives (Educational Data Mining)

1.5.1 Définition

C'est un domaine moderne qui vise à développer des moyens d'explorer les types de données utiles obtenues à partir des environnements d'apprentissage et à utiliser ces méthodes pour mieux comprendre les élèves et les environnements dans lesquels ils apprennent. Les principales utilisations de l'extraction de données éducatives comprennent la prévision des performances des élèves et l'étude du processus d'apprentissage afin de recommander des améliorations aux pratiques éducatives actuelles. Cela peut être considéré comme l'extraction de données éducatives, un domaine des sciences de l'éducation et de l'exploration de données, et l'analyse du processus d'apprentissage est un domaine connexe. Il aide à la prise de décision basée sur les données pour améliorer les pratiques et conditions d'enseignement actuelles.

1.5.2 Objectifs de la fouille de données éducatives

Les objectifs de la recherche rapportée sont de justifier les potentiels des algorithmes d'exploration de données dans le contexte de l'enseignement supérieur en développant un modèle d'exploration de données. Dans cette recherche, l'accent est mis sur l'extraction des connaissances cachées de la base de données des étudiants et la prévision des performances des étudiants en fonction des paramètres de dépendance. L'objectif de la recherche est de développer un modèle d'exploration de données avec des techniques de classification et de regroupement. La technique de classification classe les données en fonction de l'ensemble d'apprentissage et applique le modèle pour classer. L'EDM a été mis en œuvre ces dernières années pour atteindre un grand nombre d'objectifs éducatifs, notamment :

Objectifs pédagogiques

Ce sont les objectifs académiques qui aident à développer et à améliorer le niveau d'éducation d'un étudiant, mais aussi à guider les étudiants dans le cheminement scolaire en fonction de leur niveau et des résultats obtenus.

1.5.3 Tâches EDM

Les tâches EDM de base peuvent être mappées à des techniques d'exploration de données telles que : [3]

- > **Classification** :catégorisation des élèves pour déterminer les styles d'apprentissage et les préférences.
- > **La modélisation prédictive** : prédit la performance d'un étudiant à l'examen semestriel.
- > **Regroupement** :regrouper des étudiants similaires en fonction des résultats scolaires pour l'apprentissage collaboratif.
- > **Exploration de modèles** :recherche de modèles.
- > **Analyse visuelle** : raisonnement sur les processus éducatifs.

1.5.4 les méthodes

La fouille de données englobe plusieurs méthodes, qui permettent de créer des modèles ou patterns afin de les utiliser pour la découverte de connaissances. Nous nous concentrerons sur les méthodes d'explications et de prédictions. Ces méthodes peuvent être classifiées en trois catégories.

- Les méthodes de visualisation et de description.
- Les méthodes de classification et de structuration.
- Les méthodes d'explication et de prédictions.

les méthodes de visualisation et de description

Les méthodes de visualisation et de description, sont issues de la statistique descriptive et de l'analyse des données, ainsi que de la visualisation graphique. En effet, les méthodes de visualisation et de description sont fondées sur des graphiques, qui facilitent l'interprétation à l'utilisateur. La visualisation d'un graphique sert principalement à explorer les données, ou à confirmer des hypothèses.

Les méthodes de classification et de structuration

Les méthodes de classification et de structuration connus sous le nom classification automatique ou apprentissage non supervisé. Ces méthodes proviennent de l'analyse des données, de la reconnaissance des formes, de l'apprentissage automatique et du connexionnisme.

Les méthodes d'explication et de prédictions

Les méthodes d'explication et de prédictions ont pour objectif de relier un phénomène à expliquer à un phénomène explicatif, elles sont utilisées pour prévoir un comportement, ou bien pour classer de nouveaux cas dans des catégories prédéfinies. Ces méthodes sont issues de la statistique, de l'économétrie, de la reconnaissance de formes, de l'apprentissage automatique et du connexionnisme.

1.6 Outils de la fouille de données

1.6.1 Logiciels libres

Parmi les logiciels libres : KNIME et weka, ces deux logiciels sont décrits ci-dessous.

- **KNIME** :acronyme de Konstanz Information Miner, est un logiciel libre édité par un laboratoire de l'université de Constance dénommé Nycomed Chair for Bioinformatics and Information Mining. Il intègre notamment tous les modules d'analyse de Weka et permet de créer des scripts en langage R. KNIME s'exécute sur Linux, Windows et MacOS. Comme tous les logiciels libres, KNIME est extensible.

- **Weka** :est un logiciel libre de fouille de données développé en java et créé par l'université de Waikato (Nouvelle-Zélande). C'est une collection d'algorithmes d'apprentissage automatique mis en place pour effectuer des tâches d'exploration de données. Les algorithmes peuvent soit être appliqués directement à un ensemble de données soit être appelés directement par un code Java. Weka contient des outils pour les prétraitements des données, la classification, la régression, le clustering, les règles d'association et la visualisation. Comme KNIME, weka est un logiciel open source.
- **RapidMiner** :est un logiciel open source dédié au data mining. Il contient de nombreux outils pour traiter des données : lecture de différents formats d'entrée, préparation et nettoyage des données, statistiques, tous les algorithmes de data mining, évaluation des performances et visualisations diverses. C'est un logiciel puissant, il n'est pas facile à manipuler au premier abord, mais avec un peu de pratique, il permet de mettre en place rapidement une chaîne complète de traitement de données, de la saisie des données à leur classification.

1.6.2 Logiciels commerciaux

Les logiciels commerciaux sont édités par des sociétés bien connues sur le marché :

- **KXEN** :Analytic Framework est un logiciel commercial édité par la société KXEN basée en Californie et fondée en 1998. Les modules de KXEN Analytic Framework permettent la prédiction, la segmentation, les associations, la fouille de textes et l'analyse des réseaux sociaux.
- **SAS Enterprise Miner** :est un outil commercial édité par la société SAS Institute Inc. C'est un logiciel offrant toutes les facettes de l'exploration de données dont le processus est facilité par son interface homme-machine bien conçue.
- **SPSS (Statistical Package for the Social Sciences)** :est un logiciel de statistiques, édité par la filiale d'IBM du même nom, qui se décompose en plusieurs modules dont SPSS Modeler pour le Data mining, SPSS Amos pour les modèles d'équation structurelle et Predictive Analytics pour l'analyse prédictive.

- **CORICO** :est un logiciel commercial intégrant l' Iconographie des corrélations et les Interactions logiques, qui se prêtent bien à l'analyse multi relationnelle. Il intègre aussi une technique de modélisation prédictive fondée sur les modèles de régression multiple postulés et non postulés.

1.7 Conclusion

Le développement des sciences, de l'économie et des technologies de l'information et de la communication a accru la quantité de données numériques et, avec ces énormes quantités de données, les méthodes d'analyse traditionnelles ne sont plus en mesure de les gérer.

Par conséquent, pour résoudre ces problèmes, l'exploration de données a été fournie avec des outils et des logiciels qui aident à explorer la quantité considérable et croissante de données, car ces données sont produites au milieu de sites de réseautage social ; et diverses institutions telles que les banques et les compagnies d'assurance, il est nécessaire d'explorer cette quantité de données pour en bénéficier afin d'accéder à des connaissances qui nous aident à prendre des décisions.

CHAPITRE III

Implémentation et Réalisation d'un système industriel

1. Introduction

L'objectif principal de l'implémentation de l'application de datamining après une série de plusieurs étapes dans le processus de développement est de développer des modèles de classification de système de diagnostic industriel utilisant les différents algorithmes offerts par l'outil Weka. Après une présentation de l'unité de clinkerisation de la cimenterie la CIMAT d'Ain-Touta-BATNA, nous allons donner un aperçu sur le processus de fabrication des ciments et plus précisément la zone de clinkérisation.

Dans les chapitres précédents, nous avons fourni les concepts nécessaires à la conception et à la mise en œuvre de la classification et de la prédiction. Ce chapitre donne un aperçu de notre système et les outils utilisés pour le développement. Dans ce chapitre nous allons présenter en premier temps, l'environnement du développement avec les différentes bibliothèques utilisées, ainsi que les techniques de datamining utilisées pour implémenter ce type de système. Ensuite, on va présenter notre application de classification développée illustrée par quelque résultat obtenu et enfin, nous terminons ce chapitre par une conclusion.

2. Présentation de la SCIMAT

2.1. Identité

- Raison Sociale : SCIMAT (Société des Ciments d'Ain-Touta), Filiale du Groupe ERCE.
- Forme Juridique : Société par Action (SPA).
- Siège sociale : Rue Benflis –BP : 67-05000 –El –Boustène, Batna –ALGERIE

2.2. Situation Géographique

- La Société des Ciments de Ain Touta (SCIMAT) se compose de :
- Une Direction Générale : située à Batna.
- Une Unité Cimenterie: située dans la commune de Tilatou, Daïra de Seggana, à 50 Km à l'ouest de Batna elle couvre une superficie de 20 hectares et implantée sur l'axe routier Batna-
- Bari ka (Route Nationale N° 5) et sur la rocade ferroviaire Ain-Touta-M'sila.

2.3. Fiche Technique

- Domaine d'activité : Fabrication et vente des ciments.
- Capacité de production : 1.500.000 de tonnes de ciment par an.

3. Etapes de Fabrication du ciment

Le Ciment produit par la cimenterie Ain-Touta est fabriqué à partir d'un processus à voie sèche c'est la procédure la plus utilisée et la plus économique.

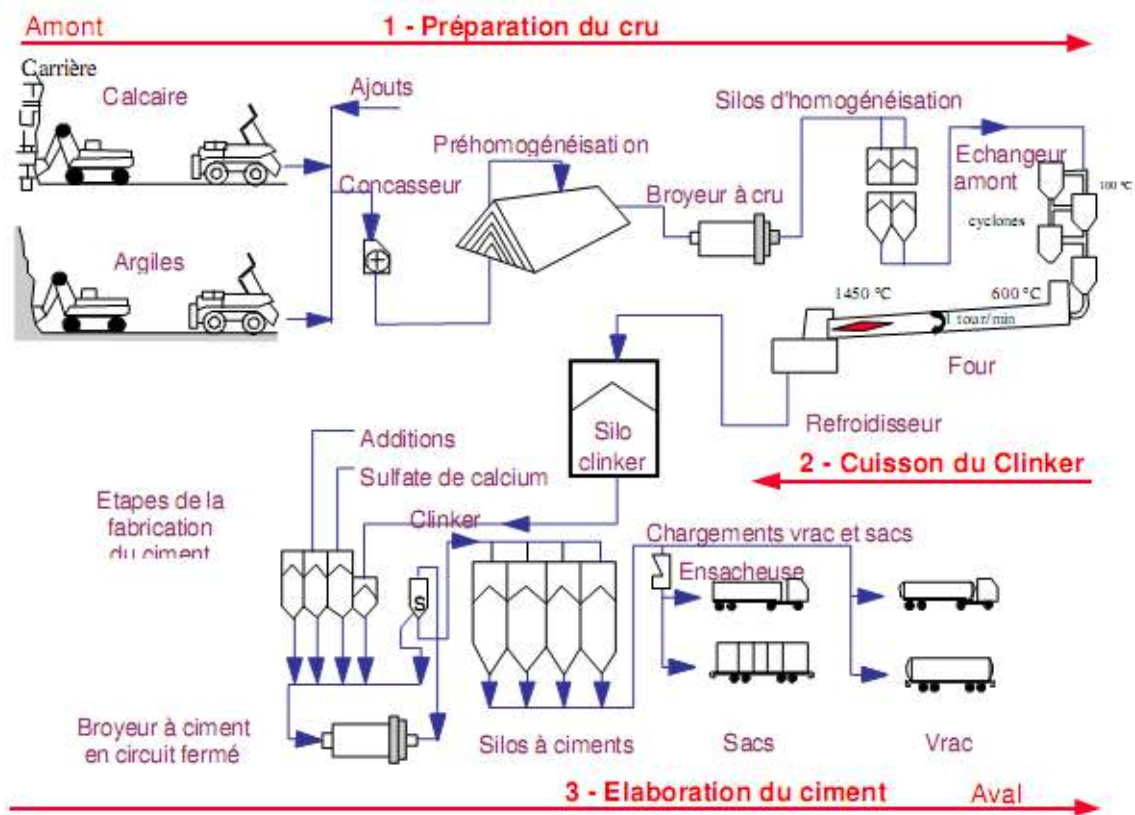


Figure2.1: Processus de fabrication du ciment

3.1. Matières premières

Calcaire: Roche constituée principalement de carbonates de calcium (75 à 92% de CaCO_3) extraite de la carrière par abattage à l'explosif.

Argile : Roche constituée principalement de silice, d'alumine et de fer extraite du Gisement par ripage

Minerai de fer: Roche qui se rencontre le plus souvent sous forme d'oxyde de fer (Fe_2O_3) et qui joue un rôle utile de fondant.

Gypse: Roche constituée principalement des sulfates de calcium ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$) et qui joue un rôle utile de retardateur de prise.

Pouzzolane: Roche d'origine volcanique constituée principalement de silice, utilisée comme ajout actif.

3.2. Les étapes du processus

Le procédé décomposé en 12 étapes clés

La carrière de cimenterie

- La pré-homogénéisation des matières premières
- Le broyage du cru
- L'homogénéisation du cru
- Le préchauffage du cru
- La cuisson
- Le refroidissement du clinker
- Le stockage du clinker
- Le broyage du ciment
- Le stockage du ciment
- Le contrôle qualité du ciment
- Conditionnement et expédition du ciment

La carrière de cimenterie : Toute l'histoire de la fabrication du ciment, commence à la carrière...

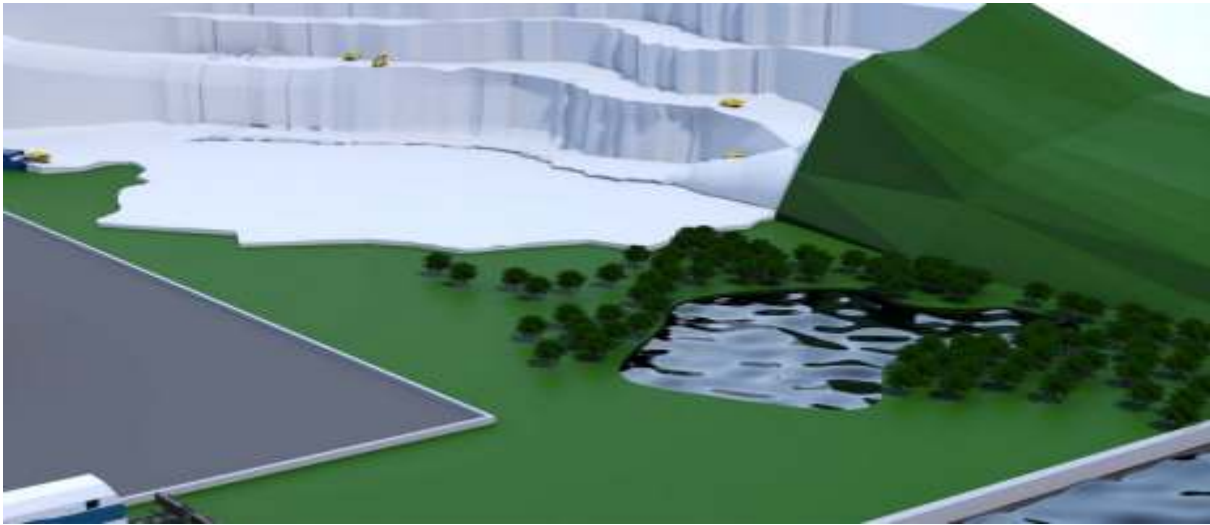


Figure 2.2 : La carrière de cimenterie

La pré-homogénéisation des matières premières : Les roches extraites de la carrière sont concassées et acheminées jusqu'à la cimenterie.

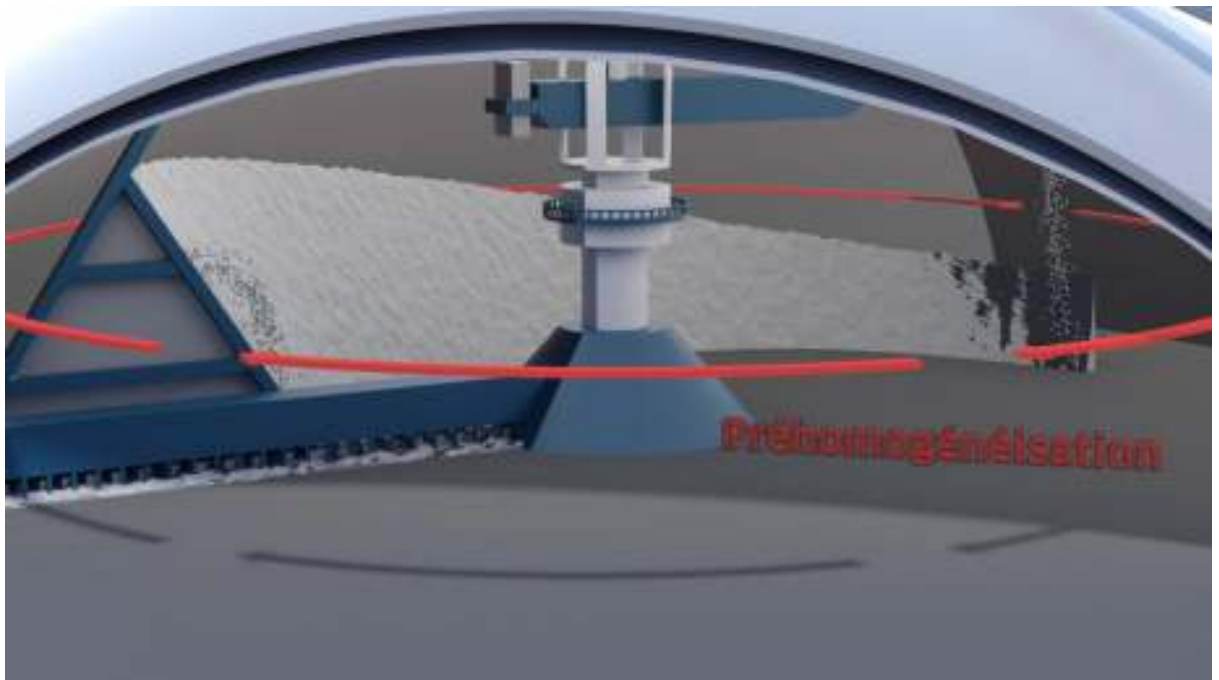


Figure 2.3 : La pré-homogénéisation des matières premières

Le broyage du cru : Le mélange de roches minérales, préparé dans le hall de pré-homogénéisation, est dosé, séché et broyé afin de réduire la roche en une poudre très fine, de taille micrométrique.



Figure 2.4 : Le broyage du cru

L'homogénéisation du cru : Après avoir été finement broyée, la farine crue est très largement brassée dans des silos d'homogénéisation, de manière à obtenir une composition.

Le préchauffage du cru : La farine crue dosée passe par deux air-lifts qui la transportent en haut des préchauffeurs à cyclones.

- Chaque tour de préchauffage est constituée de quatre étages de cyclones :
 - L'étage supérieur comporte un cyclone double de 4,35 m de diamètre.
 - Les trois étages au dessous comportant chacun un cyclone simple de 6,6m de diamètre.
- La farine crue alimentée aux prés chauffeurs subit une décarbonatation partielle suite échange thermique à contre courant à des températures variant de 320°C - 850°C selon étages cyclones.

Four rotatif : La farine, précédemment préchauffée à 850°C, est ensuite chauffée à 1450°C sous une flamme à 2 000°C, dans un four rotatif garni de briques réfractaires, qui fonctionne 24h/24 et 7j/7.

Le refroidissement du clinker : Le mélange en fusion est ensuite trempé (c'est-à-dire refroidi rapidement) par soufflage d'air, afin d'obtenir la chimie cristalline appropriée pour les propriétés.

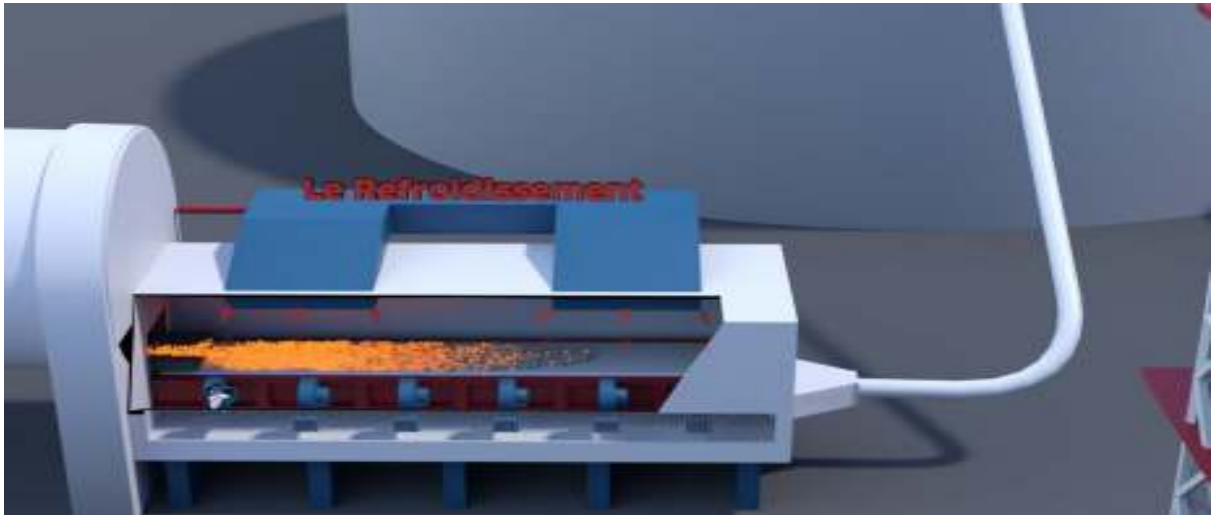


Figure 2.4 : Le refroidissement du clinker

Le stockage du clinker : Le clinker refroidit à 100°C est transporté et stocké dans un hall de plusieurs dizaines de milliers de tonnes, permettant ainsi d'assurer une continuité de la production.



Figure 2.5 : Le stockage du clinker

Le broyage du ciment : Le clinker est ensuite broyé avec du gypse, qui joue le rôle de régulateur de prise, et un ou plusieurs constituants principaux autre que le clinker :

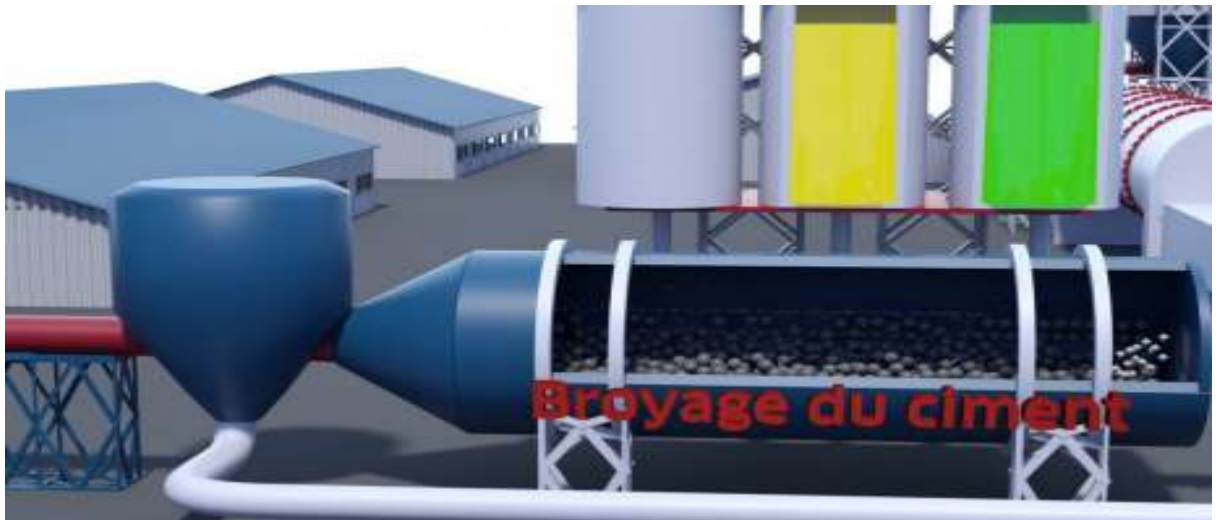


Figure 2.6 : Le broyage du ciment

Le stockage du ciment : Le ciment est stocké dans des silos d'une capacité de plusieurs milliers de tonnes, qui peuvent être parfois divisés en plusieurs compartiments.



Figure 2.7 : Le stockage du ciment

Le contrôle qualité du ciment : Lors de son stockage, le ciment est prélevé, analysé chimiquement et physiquement et des essais de résistance sur mortiers normalisés sont réalisés.



Figure 2.7 : Le contrôle qualité du ciment

Conditionnement et expédition du ciment : Le ciment est expédié en vrac ou en sac par différents modes de transport.



Figure 2.7 : Conditionnement et expédition du ciment

4. Atelier de clinkérisation

L'atelier de clinkérisation regroupe deux lignes de fours rotatifs dont le débit de clinker est 1560 t/h. Le four, de la SCIMAT, est un cylindre d'acier rotatif posé avec une inclinaison de 3 à 5% vers la sortie. Il est calé à intervalles par des trains porteurs à galets. La rotation lui est transmise à travers une couronne dentée, un pignon et un réducteur de vitesse. Le four aura tendance à se déplacer vers l'aval au cours de sa rotation. Il est protégé contre les très hautes températures de service par revêtement de briques réfractaires maçonnées à l'intérieur sur toute longueur. Le revêtement est sensible à la déformation élastique du four et les variations de température.

La température de cuisson est aux environs de 1450°C. La chauffe est faite soit au fuel, soit au gaz, soit au charbon pulvérisé ou une combinaison de ces trois. Les matières premières sont enfournées à l'entrée du four et se déplacent à contre-courant des gaz chauds, à travers le four jusqu'à ce qu'elles en ressortent par le refroidisseur sous forme de clinker cuit.



Figure 2.8 : Four Rotatif de la SCIMAT

4.1. Caractéristiques du Four rotatif

- Type du four : Four rotatif UNAX à 04 étages de préchauffeurs
- Capacité journalière : 2X1500 t/j de clinker
- Diamètre : 4,55 mètres
- Longueur : 68 m
- Vitesse de rotation : 2,0 tr/mn
- Revêtement : Réfractaire à base d'alumine et de magnésie chrome.
- Température de cuisson : Jusqu'à 1400°C
- Température Clinker sortie refroidisseurs : de 150 °C à 200 °C

5. La modélisation

5.1.Définition

La modélisation consiste à définir les points suivants : Le système, Le modèle, L'objectif, Un critère de rentabilité [1].

Le système : il y a de nombreuses définitions ont été attribuées au system

- Ensemble de composants reliés entre eux.
- Ensemble organisé d'éléments fonctionnels.
- Assemblage d'éléments fonctionnant de manière unitaire et en interaction Permanente [1].

Le modèle : structure simplifiée de la réalité qui ne contient pas que les caractéristiques estimées importantes pour l'étude du système [1].

L'objectif : le but pour lequel de modèle a été élaboré [1].

Un critère de rentabilité : un critère économique qui justifie l'utilisation d'un modèle [1].

figure suivante montre le processus global de notre système :

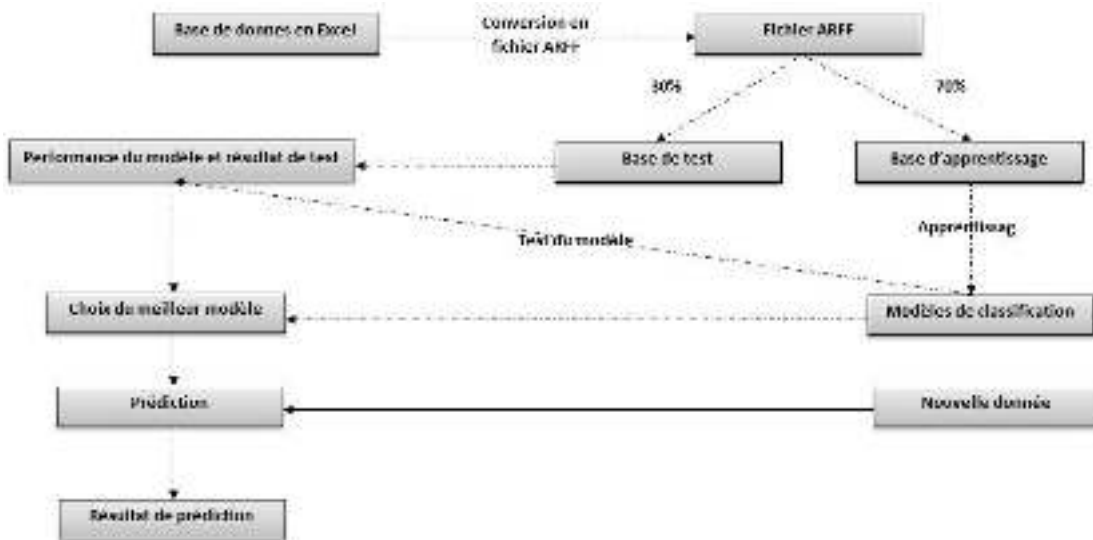


Figure 2.9 : Processus global de notre système

5.2 Environnement et outils de mise en œuvre

Notre système était développé sur un ordinateur de processeur Core i3 avec une RAM de 4 Go sous Windows 10, mais on peut implémenter ce système sur n'importe quel machine grâce au virtuel machine de Java.

Java

Java est un langage de programmation orienté objet, inventé par James Gosling en 1992

alors qu'il travaillait dans les laboratoires Sun Microsystems - pour l'utiliser comme esprit pensant utilisé pour alimenter des dispositifs d'application intelligents tels que la télévision interactive, et Java était un développement de C ++, Et à sa naissance, son créateur l'a appelé "chêne", c'est-à-dire le chêne, qui est l'arbre qu'il a vu depuis la fenêtre de son bureau alors qu'il travaillait dans les laboratoires de Sun Microsystems, puis le nom a été changé en Java, et ce nom (inhabituellement dans les langages de programmation de dénomination) n'est pas les premières lettres de Mots d'une phrase ou d'une expression particulière avec une signification spécifique, et pour vous C'est juste un nom développé par les développeurs de cette langue pour rivaliser avec d'autres noms.

Il est fourni avec un ensemble d'outils (le JDK Java Développement Kit) et un ensemble de packages : ensemble de classes. Ces différentes classes de base couvrent beaucoup de domaine (entrées/sorties, interface graphique, réseau, etc.) Cette richesse en "bibliothèques standards" explique sûrement en partie le succès de Java. Le langage lui-même se trouve dans le package java. Lang .

Java est donc :

- Un langage de programmation orienté objets.
- Une architecture de machine virtuelle.
- Un ensemble d'outils.

Ses avantages : Le langage Java a des caractéristiques spéciales, ce qui en fait le langage de programmation le plus excitant. Ce qui le distingue sont les suivants :

Facilité.

Facilité d'accès.

Transférable et exécutable.

Jeux d'écriture et utilitaires.

Créer des programmes avec une interface utilisateur graphique.

Conception logicielle qui tire parti de tous les avantages d'Internet.

Le langage Java fournit un environnement interactif via le World Wide Web et est donc utilisé pour écrire des programmes éducatifs pour Internet via un logiciel de simulation informatique pour les expériences scientifiques et un logiciel de classe virtuelle pour le e-learning et l'enseignement à distance. L'efficacité de Java ne se limite pas au Web uniquement. Elle nous permet également de créer des programmes pour un usage personnel et professionnel. Ces programmes sont mis en œuvre à travers un certain nombre de programmes qui facilitent l'écriture de commandes telles que le programme NetBeans et Eclipse .

NetBeans

La programmation peut se faire pour des exemples simples avec le compilateur java, mais pour avoir plus de confort il est préférable d'utiliser un environnement de développement

intégré ou IDE, comme Eclipse ou NetBeans. Dans notre projet nous avons utilisé NetBeans.

NetBeans est un environnement de développement intégré (IDE) pour Java, placé en open source par Sun en juin 2000 sous licence CDDL (Common Développement and Distribution License). En plus de Java, NetBeans permet également de supporter différents autres langages, comme Python, C, C++, XML et HTML. Il comprend toutes les caractéristiques d'un IDE moderne (éditeur en couleur, projets multi-langage, refactoring, éditeur graphique d'interfaces et de pages web). NetBeans est disponible sous Windows, Linux, Solaris (sur x86 et

SPARC), Mac OS X et Open VMS. NetBeans est lui-même développé en Java, ce qui peut le rendre assez lent et gourmand en ressources mémoirees.[22]

Version utilisée La version utilisée pour le développement de notre application est la version IDE 8.0.2

l'outil WEKA

WEKA (WAIKATO Environment for Knowledge Analysis) est un outil de fouille de données (licence GNU) développé en Java. [21] Il a été créé à l'université de Waikato, en Nouvelle-Zélande, par un groupe de chercheurs issus de l'apprentissage automatique, de la reconnaissance de formes et de la fouille de données.

WEKA permet de prétraiter des données (onglet Preprocess dans l'interface graphique), faire de la classification supervisée (Classify) et non-supervisée (Cluster), des régressions (Select Attributes), rechercher des règles d'association (Associate), et de visualiser différentes représentations graphiques des données (Visualize).[20]

Il s'agit d'un logiciel « open source » gratuit dédié à la classification et à la fouille de données. Il s'adresse à deux types de publics. D'un côté, il présente une interface graphique, le rendant ainsi accessible à une utilisation de type « chargé d'études » sur des données réelles. De l'autre, du fait que le code source est librement disponible et l'architecture interne très simplifiée, il se prête à une utilisation de chercheurs qui veulent avant tout expérimenter de nouvelles techniques en améliorant celles déjà implémentées ou en introduisant de nouvelles.



Figure 4.2 – Interface graphique de WEKA

.Caractéristiques principales

Plus de 49 outils de prétraitement de données.

Plus de 76 algorithmes de classification régression regroupés en 07 familles.

Plus de 8 algorithmes de "clustering".

Plus de 15 évaluateurs d'attributs et plus de 10 algorithmes pour la sélection d'attribut.

3 algorithmes de recherche de règles d'association.

3 interfaces graphiques GUI.

« Explorer » (explorateur d'analyse de données).

« Expérimenter » (environnement expérimental).

« KnowledgeFlow » (le nouveau modèle de processus avec interface).

Structure de données

WEKA traite des données contenues dans des fichiers respectant le format ARFF Attribute-Relation File Format. Il s'agit de fichiers de type texte, décrivant des ensembles de "tuples" caractérisés par un certain nombre d'attributs communs.

Format d'un fichier ARFF (Attribute-Relation File Format) WEKA utilise (entre autres) le format de fichier arff pour enregistrer les données. Un fichier arff est composé d'une liste d'exemples définis par leurs valeurs d'attributs. Un fichier arff comprend toujours trois types d'informations : un nom pour la base de données, des attributs et des données. La chaîne de caractères @RELATION permet de donner un nom à la base de données. Par exemple, dans le cas du fichier Data.arff, le nom donné est Data. @RELATION Data. **La chaîne de caractères @ATTRIBUTE permet de définir un attribut. Un attribut peut être de 4 types :**

réel (NUMERIC ou REAL).

Nominal (valeurs-possible) **par exemple** : @attribute Sexe

FEM,MAL signifie que l'attribut Sexe peut avoir comme valeur soit Sexe-FEM ou soit Sexe-Mal.

Chaîne de caractère (STRING).

Date (date [<date-format>] @data : suivi d'une instance par ligne.

Les valeurs d'instance sont séparées par une virgule.

Remarque préliminaire : le caractère « % » marque les lignes de commentaires. WEKA propose un éditeur de fichier arff (tools → arffViewer) permettant de visualiser les fichiers arff sous la forme d'un tableau, et éventuellement de les modifier.

Exemple de fichier ARFF

```

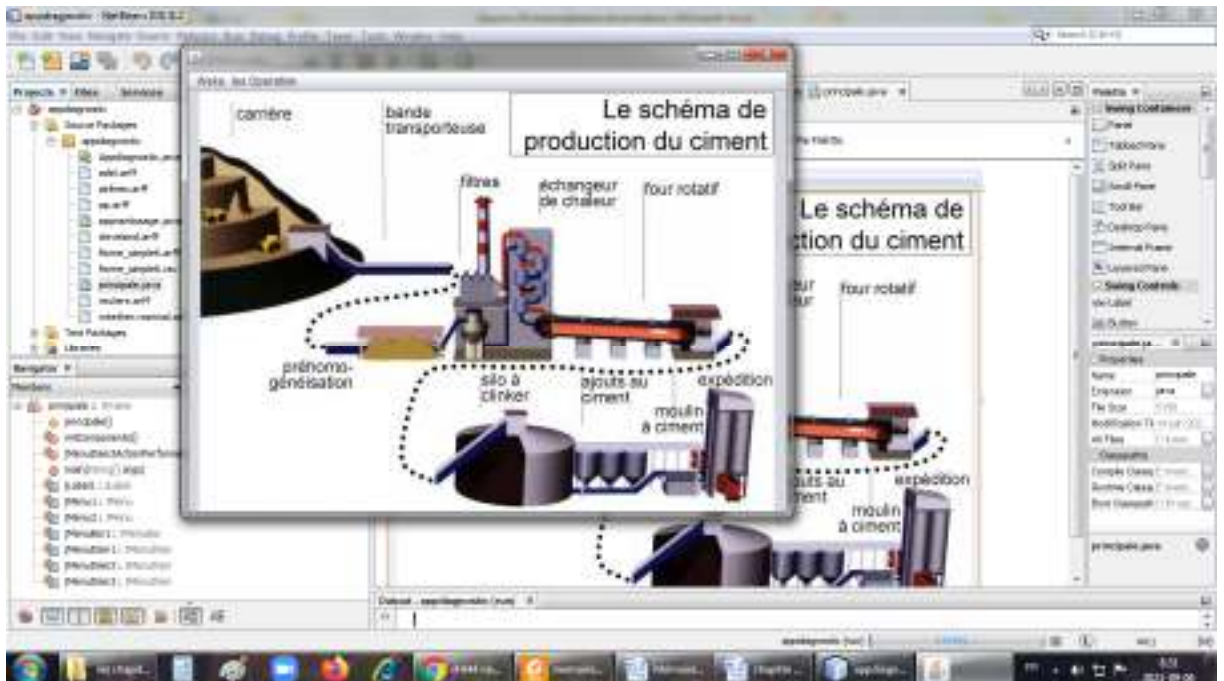
@relation
Data.
@attribute Sexe
FEM,MAL.

@attribute Age
>=20,19,'<=18 '.@attribute
Civil_State SIG.
@attribute Residence
DAI,WIL,COM,VIL.
@attribute Fam_Problems
NO,YES. @data.
FEM , >=20 , SIG , DAI
, NO , NO.FEM , 19 ,
SIG , DAI , NO , NO.
MAL , 19 , SIG , DAI ,
NO , NO. MAL , 19 ,
SIG , COM , YES ,NO.
    
```

4.1 Les interfaces de l'application développée

Nous présentons dans cette section des captures d'écran de l'application développée.

4.1.1 L'Interface d'accueil



Fenêtre principale de l'application

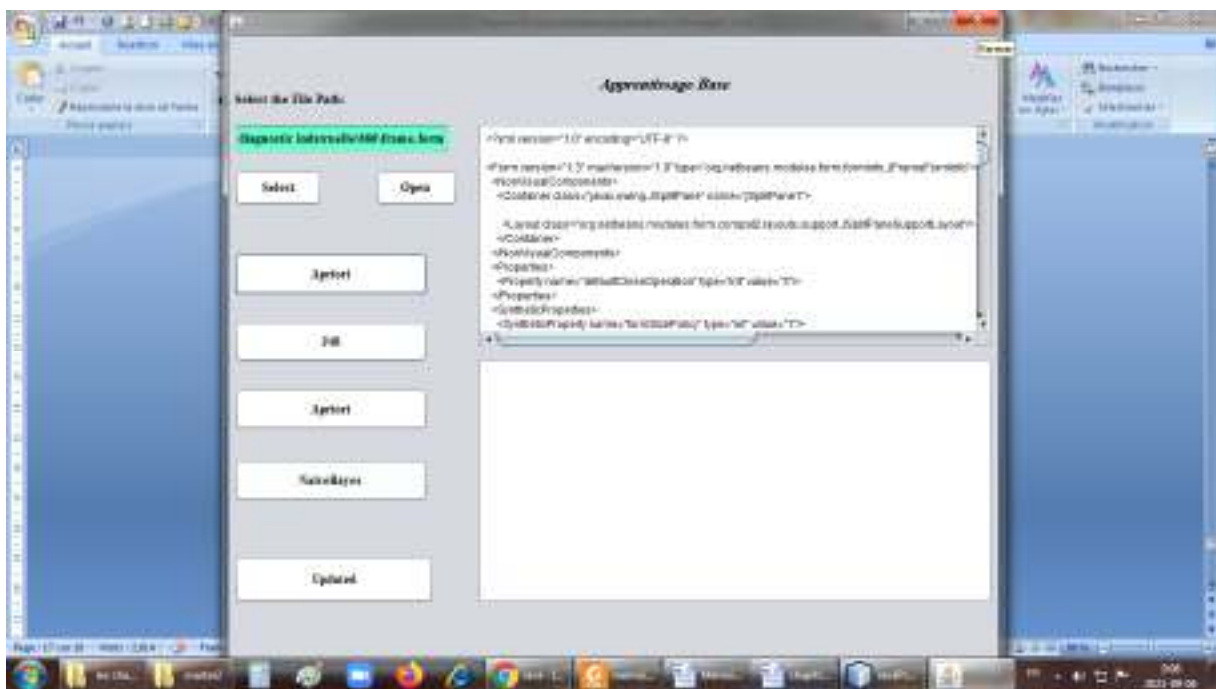
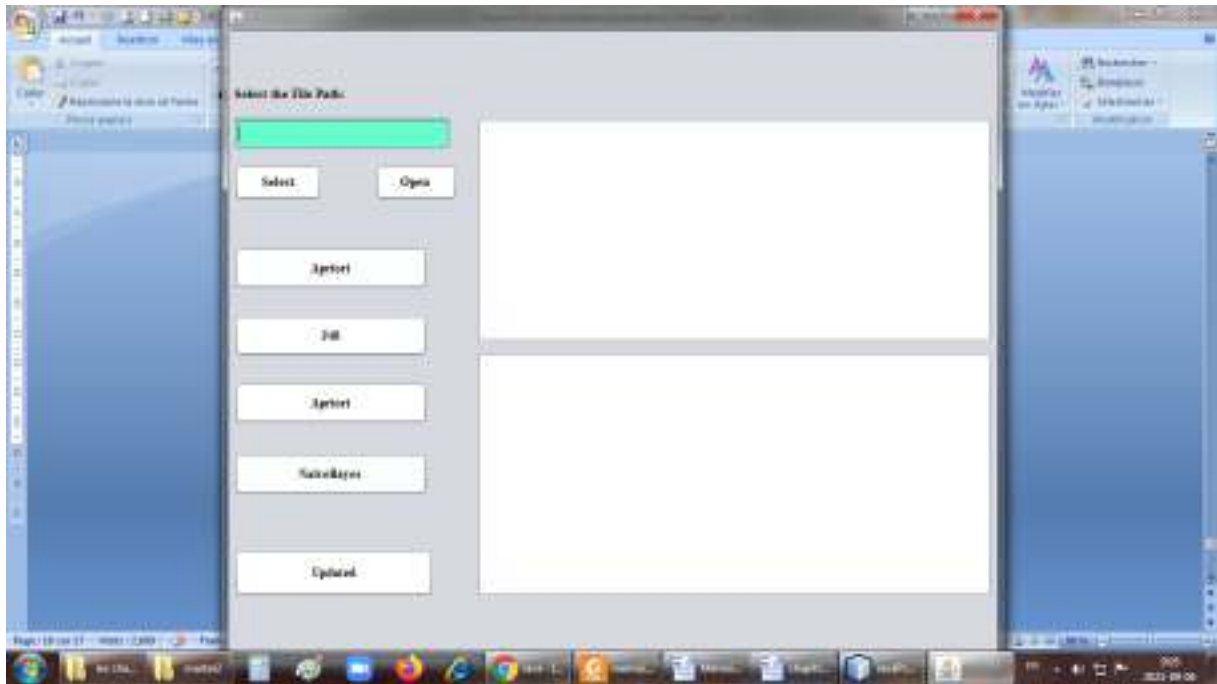
Cette fenêtre contient deux boutons principaux :

- Le bouton « Apprentissage » permet de sélectionner la base d'apprentissage et de lancer l'apprentissage pour construire le modèle de classification.

Le bouton « Prédiction » implémente la tâche de prédiction des résultats .

4.1.2 Sélection de la base d'apprentissage

La fenêtre de sélection de fichiers ARFF pour sélectionner la base de données ,comme illustré dans la figure suivante.



Sélection de la base d'apprentissage

6. Résultats de classification

L'objectif de cette section est de donner quelques résultats expérimentaux de notre application au logiciel Weka, l'idée est très simple de faire une classification de modes de défaillances de notre système de production,

pour cela nous avons choisi un ensemble d'algorithmes de classification

6.1 L'algorithme J48 : Le résultat obtenu est un Arbre de décision généré par l'algorithme J48 pour la classification de 4 modes de défaillances de notre système de production étudié

les autres modes de défaillance sont :

1	Décarbonatation perturbée (mauvaise)	Mode de défaillance (MD1)
2	Mauvaise clinkerisation	Mode de défaillance (MD2)
3	mauvais refroidissement	Mode de défaillance (MD3)
4	Insuffisance ou absence du gaz	Mode de défaillance (MD4)

La figure suivante illustre un classement des modes de défaillances de votre système de production

```

Classifier output
D_s_c_pp
T_s_c_pp
D_e_c_v
D_s_c_v
Etat_vanne
Capa_silo
D_e_c_si
D_s_c_si
Mode_d_def
Test mode:10-fold cross-validation
--- Classifier model (full training set) ---
ZeroR predicts class value: 2.52
Time taken to build model: 0 seconds
--- Cross-validation ---
--- Summary ---
Correlation coefficient          -0.1533
Mean absolute error             1.008
Root mean squared error         1.1254
Relative absolute error         100 %
Root relative squared error     100 %
Total Number of Instances      200
    
```

La figure suivante illustre l'ensemble des cluster ainsi que le taux de classification de chaque attribut qui influ sur le taux de classification globale de modes de défaillances du système de production

```

Clusterer output
  mean      1568.7554 1572.8728 1569.6382 1575.8433
  std. dev.   31.7352  33.0567  30.1358  27.428

D_s_c_si
  mean      1574.5344 1572.7982 1575.6033 1560.3669
  std. dev.   31.8378  30.9487  30.5658  31.6345

Mode_d_def
  mean       2.1951   3.0842   2.3044   3.6468
  std. dev.   0.9868   1.0606   1.1031   0.6941

Time taken to build model (full training data) : 5.22 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      105 ( 53%)
1       40 ( 20%)
2       53 ( 27%)
3        2 (  1%)

Log likelihood: -148.08664
    
```

6. Conclusion

Dans ce chapitre on a présenté brièvement les étapes de fabrication de Ciment et spécialement la phase de clinkérisation qui est le but de notre travail et comme notre travail est une Diagnostic du système , nous avons présenté les résultats obtenus grâce aux expériences que nous avons menées pour tester notre système développé pour évaluer les classifications fournies par les outils d'exploration de données Weka selon des critères spécifiques.

Conclusion générale

Conclusion general

L'objectif principal de notre mémoire de fin d'étude était de modéliser simuler et implémenter un système d'aide au diagnostic industriel des systèmes de production dans notre cas est une cimenterie de Aintouta Batna, par les méthodes de data mining ,
Après avoir donné une aperçue sur les systèmes de diagnostics méthodes et techniques nous avons ensuite fait tout un chapitre sur les concepts de bases de fouille de données industrielles,
le troisième chapitre est le vif de notre travail as voir de faire une classification des modes de défaillances et des pannes industriels par un ensemble des algorithmes qui nous aident à comprendre d'une manière approfondie le fonctionnement du système de production ainsi que le diagnostic,

Références bibliographiques

- [1] **Microsoft Experiences**, *Tout savoir sur l'Intelligence Artificielle*, (consulté le 09/02/2019), disponible sur : <https://experiences.microsoft.fr/business/intelligence-artificielle-ia-business/comprendre-utiliser-intelligence-artificielle/>
- [2] **Frédéric Fürst**, *Histoire de l'Intelligence Artificielle – 3- naissance de l'Intelligence Artificielle*, (consulté le 10/04/2019), disponible sur : https://home.mis.u-picardie.fr/~furst/docs/3-Naissance_IA.pdf
- [3] **Conseil de l'Europe**, *L' IA, C'est quoi – Histoire de l'Intelligence Artificielle*, (consulté le 10/04/2019), disponible sur : <https://www.coe.int/fr/web/artificial-intelligence/history-of-ai>
- [4] **Le Big Data**, *Le marché du big data atteindrait 67 milliards de dollars en 2021*, (consulté le 12/04/2019), disponible sur : <https://www.lebigdata.fr/marche-big-data-atteindrait-67-milliards-de-dollars-2021>
- [5] **Olivier Bison**, *Comprendre le data discovery*, (consulté le 15/04/2019), disponible sur : <https://tutos-du-web.fr/comprendre-le-data-discovery/>
- [6] **Le Big Data**, *La Dataviz – qu'est-ce que c'est ? – Définition, outils essentiels*, (consulté le 15/04/2019), disponible sur : <https://www.lebigdata.fr/dataviz-qu-est-ce-que-c-est>
- [7] **Le Big Data**, *Définition DATA CENTER: Qu'est-ce qu'un centre de données ?*, (consulté le 15/04/2019), disponible sur : <https://www.lebigdata.fr/definition-data-center-centre-donnees>
- [8] **Le Big Data**, *Business Intelligence ou Informatique décisionnelle : définition et outils*, (consulté le 15/04/2019), disponible sur : <https://www.lebigdata.fr/business-intelligence-definition>
- [9] **Inserm**, *Big Data en santé – des défis techniques, humains et éthiques à relever*, (consulté le 20/04/2019), disponible sur : <https://www.inserm.fr/information-en-sante/dossiers-information/big-data-en-sante>
- [10] **DATA ACADEMIE**, *Le Big Data et l'Intelligence Artificielle*, (consulté le 12/05/2019), disponible sur : <https://www.datacademie.com/blog/le-big-data-et-lintelligence-artificielle/>

Résumé

L'extraction des connaissances à partir des données est définie comme le processus d'analyse des données sous différentes perspectives et de découverte des modèles à partir des ensembles de données utiles pour prédire les résultats qui nous aident à prendre la bonne décision.

Comme ce processus passe par plusieurs étapes, de la (collecte des données. Nettoyage et transmission) à l'obtention des résultats, leur validation et leur interprétation, et enfin la fusion des connaissances acquises.

Il existe de nombreuses techniques d'analyse de données qui peuvent être utilisées pour extraire des modèles, parmi lesquelles des techniques de classification qui génèrent des modèles (phase d'apprentissage) utilisés pour prédire les données futures d'un objet (phase de prédiction) à partir de données provenant d'autres objets similaires.

Notre travail est lié à cet aspect. Nous avons d'abord collecté les données et nous avons ensuite généré plusieurs modèles qui permettent de classer les performances en deux types (fonctionnement normal, présence de défaillances).

Notre travail est concrétisé par une application développée sous Java et Weka qui permette de faire la saisie des nouvelles données, l'apprentissage pour la création des modèles de classification.