



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mohamed Khider – BISKRA

Faculté des Sciences Exactes, des Sciences de la Nature et de la Vie

Département d'informatique

N° d'ordre : SIOD18 /M2/2018

Mémoire

présenté pour obtenir le diplôme de master académique en

Informatique

Parcours : Système d'information et Optimisation et décision

Système data mining basé sur un SMA

Par :

ATIA SARA

Soutenu le 25/06/2018, devant le jury composé de :

HOUHOU Okba

MCB

Président

TOUIL Keltoum

MAA

Rapporteur

RAHMANI Salima

MAA

Examineur

Dédicace

Je dédie ce modeste travail :

*A ceux qui m'ont indiqué la bonne voie, et qui ont attendu avec patience
les fruits de leur bonne éducation ...mes parents,*

*C'est avec fierté, honneur et énorme joie que je dédie ce modeste travail :
A mon cher père qui a toujours été là pour moi et que j'espère avoir rendu
fier.*

A ma mère pour ses prières qui m'ont guidé jusqu'ici.

Et A toute ma famille

A mes amis, surtout Ramoul Amine et Hassni Khaoula

*A tous mes professeurs que j'ai connu durant mes études, surtout mon
promoteur: Mme. Touil Keltoum.*

-A mes connaissances de proche ou de loin

Remerciement

Je voudrais remercier en premier lieu notre grand Dieu qui m'a donné la force et le courage de continuer et de m'aider et d'ouvrir la voie à cet humble travail.

J'exprime ma gratitude et ma gratitude à Mme Touil Keltoum pour ses merveilleux conseils et son aide lors de l'obtention du diplôme.

Résumé :

La combinaison entre les agents et le datamining est fondée sur les défis rencontrés par les deux communautés, et la nécessité de développer des systèmes de traitement de données plus intelligents. La technologie des agents dont le but est de traiter des systèmes complexes a révélé des possibilités pour améliorer le domaine de datamining. Ce travail est consacré à l'intégration des SMA dans le datamining. Il présente un développement d'un système de classification par l'algorithme k-NN, d'une base de données répartie, basé sur les agents.

Mots Clés: *SMA, datamining, Intégration, classification, algorithme k-PPV.*

Abstract:

The combination between agents and data mining is based on the challenges faced by both communities, and the need to develop smarter data processing systems. The technology of agents whose purpose is to deal with complex systems has revealed opportunities to improve the data mining domain. This work concerns the integration of MAS in data mining. It presents a development of a system of classification by the algorithm k-NN, of a distributed database, based on agents.

Keywords: *MAS, data mining, Integration, classification, k-NN algorithm.*

Sommaire

Introduction générale.....	1
Chapitre 1: Systèmes data mining	
1. 1 INTRODUCTION	4
1. 2 DEFINITION DU DATA MINING	4
1.3 Les étapes du processus de data mining	4
1.4 Motivation du Data Mining	6
1.4.1 Problème de l'explosion de données	6
1.4.2 Améliorer la productivité	6
1.4.3 Croissance en puissance/coût des machines capables	6
1.5 Data mining sur quels types de données ?.....	7
1.6 Taches du Data Mining	8
1.6.1 La classification	9
1.6.1.1 Classification non supervisée	9
1.6.1.2 Classification supervisée (apprentissage)	9
1.6.2 La prédiction	Erreur ! Signet non défini.
1.6.3 L'optimisation	Erreur ! Signet non défini.
1.6.4 LA SEGMENTATION (ANALYSE DES CLUSTERS)	10
1.6.5 L'association	10
1.7 Les domaines d'application de Data Mining	Erreur ! Signet non défini.
1 .7.1 Le data mining dans la banque.....	11
1.7.2 LE DATA MINING DANS L'ASSURANCE ...	ERREUR ! SIGNET NON DEFINI.
1.7.3 Le data mining dans le commerce	11
1.7.4 Le data mining dans la détection de fraude	12
1.7.5 Le data mining dans la télécommunication	12
1.7.6 Le data mining dans le transport et voyage	12
1.7.7 LE DATA MINING DANS LA MEDICINE ET LA PHARMACIE .	12

Table de matière

1.8 TECHNIQUES DU DATA MINING	13
1.8.1 TECHNIQUES NON SUPERVISE	13
1.8.1.1 Les règles d'association	
1.8.1.2 Clustering	13
1.8.2 Techniques Supervisé	13
1.8.2.1 Les réseaux neurones	14
1.8.2.2 Les arbres de décision	14
1.8.2.3 L'algorithme des k-Plus proches voisins	14
1.9 CONCLUSION	15
Chapitre 2 : Système Multi Agents Erreur ! Signet non défini.	
2.1 Introduction	17
2.2 Définition d'un Agent	17
2.3 L'environnement	18
2.3.1 L'environnement d'un agent	18
2.3.2. L'environnement du système	18
2.4 Caractéristiques d'agents	18
2.5 TYPOLOGIE DES AGENTS	ERREUR ! SIGNET NON DEFINI.
2.5.1 Agents réactifs	19
2.5.2 Agents cognitifs	20
2.6 les systèmes multi agents	21
2.6.1 Définition d'un SMA	21
2.6.2 Les caractéristiques d'un SMA	22
2.6.3 COMMUNICATION ENTRE AGENTS	23
2.7 Domaines d'application des SMA	25
2.8 Quelques Systèmes existants du Data Mining basé Agents	26
2.8.1-Le Système MAD-IDS(ADMI) (Intrusion Détection System) 2011	26
2.8.2-Le système PAPYRUS 1999	28
2.8.3 Etude comparative des systèmes DMBA	30
2.9 Conclusion	32

Chapitre 3 : Conception

3.1 Introduction 34

3.2 Conception générale et détaillée du système 34

3.3 Algorithme K plus proche voisin 35

3.4 FONCTIONNEMENT DU SYSTEME 37

 3.4.1 DIAGRAMME DE CAS D'UTILISATION 37

 3.4.2 DIAGRAMME DE SEQUENCE 37

 3.4.3 DAGRAMME DE CLASSE..... 38

3. 5 CONCLUSION **ERREUR ! SIGNET NON DEFINI.**

Chapitre 4: Implémentation

4.1 INTRODUCTION..... 41

4.2 LANGAGE DE PROGRAMMATION 41

 4.2.1 JAVA 41

4.3 L'ENVIRONNEMENT DE DEVELOPPEMENT 42

 4.3.1 NETBEANS 42

 4 .3.2 LA PLATEFORME JADE (JAVA AGENT DEVELOPMENT
FRAMEWORK) 43

4.4PRINCIPALES FENETRES DE L'APPLICATION 44

 4.4.1 PRESENTATION DE LA COLLECTION 44

 4.4.2 PRINCIPALES FENETRES DE L'AGENT INTERFACE 46

 4.4.3 ÉCHANGE DE MESSAGE ENTRE LES AGENTS DU SYSTEME
..... 47

4.5 CONCLUSION 48

CONCLUSION GENERALE 49

Liste des figures

Table de figure

Figure 1.1 : Présentation classique du processus de l'ECD.....	5
Figure 1.2 : Masse importante de données– supports hétérogènes.....	7
Figure 1.3 : Tâches du Data Mining	9
Figure2.1 :Structure d'un agent réactif	19
Figure 2.2 : Structure d'un agent cognitif.....	20
Figure 2.3 : Schéma général d'un système multi-agent.....	22
Figure2 .4 :Communication entre agents par Partage d'information	24
Figure 25 : Communication par envoi de messages	25
Figure 2.6 : Domaines d'application des SMA	26
Figure 2.7 : L'architecture générale du système MAD-IDS.....	27
Figure 2.8 : Un cluster de stations de travail obtenu avec Papyrus.....	29
Figure 2.9 : Architecture générale d'un Système MADM	30
Figure 3.1 : Architecture générale du système.....	34
Figure 3.2 : Diagramme de cas d'utilisation AUML du modèle proposé.....	37
Figure 3.3 : Diagramme Séquence du modèle proposé.....	38
Figure 3.4 : Diagramme classe du modèle proposé.....	39
Figure 4.1 : langage de programmation JAVA	42
Figure 4.2 :Interface de Netbeans IDE 8.2.....	43
Figure4.3 : interface graphique de la plateforme JADE.....	44
Figure 4.4 : la base de données sous forme de chaines de caractères	45
Figure 4.5 : la base de données sous forme de chaines de Numérique	46
Figure 4.6 : Fenêtre principale de l'agent interface.....	47
Figure 4.7 : Echange de messages entre les agents du système.....	48

Liste des figures

Liste des tableaux

Table 2.1 : Les agents cognitifs et réactifs.....21

Table 2.2 : Tableau comparatif des systèmes DMBA candidats.....31

Introduction générale

Désormais, le Datamining est au cœur de toutes les préoccupations du monde des affaires. C'est un processus qui permet de découvrir, dans de grosses bases de données consolidées, des informations jusque là inconnues mais qui peuvent être utiles et lucratives et d'utiliser ces informations pour soutenir des décisions commerciales tactiques et stratégiques.

Les approches traditionnelles de la statistique ont des limites avec de grosses bases de données, car en présence de milliers ou de millions d'individus et de centaines ou de milliers de variables, on trouvera forcément un niveau élevé de redondance parmi ces variables. Les techniques de datamining interviennent et offrent des réponses à l'analyse de données volumineuse et nous permettra d'extraire des informations intéressantes et apportent de nouvelles connaissances jusque là inconnues, que les méthodes statistiques classiques n'ont pas mit en avant.

L'exploitation de ces nouvelles informations peut présenter un intérêt pour analyser et interpréter les comportements d'individus et ensemble d'individus. Les résultats obtenus s'insérant dans un dispositif d'analyse globale permettent alors de dresser dans des plans stratégiques ou politiques les axes d'effort à respecter.

Pour obtenir des informations, le processus de data mining travaille sur des bases de données très volumineuses, alors il est nécessaire de maîtriser des techniques assurant la fiabilité et la vitesse du parcours dans ces bases. Les techniques les plus classiques et courantes pour détecter les relations fonctionnelles entre les données sont: les techniques prédictives et les techniques descriptives.

Les principales caractéristiques des agents (autonomie, adaptabilité, coopération et communication) permettent d'une part de gérer efficacement des composants distribués hétérogènes et autonomes et d'autre part, de faciliter les échanges d'informations et le partage de ressources entre les composants (communication et coopération).

Dans le cadre de ce travail, on s'intéresse au développement d'un système de classification par l'algorithme k plus proches voisins, d'une base de données répartie, basé sur un système multi-agents.

Introduction générale

Organisation du mémoire :

Le plan suivi dans ce mémoire est le suivant:

Chapitre 1 : est consacré au concept de data mining.

Chapitre 2 : traite les agents et les systèmes multi-agents

Chapitre 3 : propose notre conception de la méthode suivie par le système.

Chapitre 4 : présente une mise en œuvre pratique.

Enfin la conclusion résume les différents travaux et ouvre quelques perspectives envisagées à la suite de ce travail.

Chapitre 1 :

**Systeme Data
Mining**

1.1 Introduction :

Le Data Mining (Fouille de Données) est un domaine qui consiste à comprendre les données, généralement par le moyen de méthodes statistiques. En d'autres termes, le data mining cherche à identifier des tendances parmi les données. Comme ce processus peut être très difficile, il est souvent comparé au minage de l'or dans les rivières: le gravier des alluvions représente l'énorme quantité de données et les pépites d'or représentent la connaissance cachées qu'on veut trouver.

1.2 Définition du data mining :

Le Data Mining englobe toutes les étapes d'extraction des connaissances à partir des données, il implique l'utilisation d'outils d'analyse des données recueillies sophistiqués, pour découvrir des connaissances précédemment inexploitées, cachées dans les bases de données. Ces outils peuvent inclure des modèles statistiques, informatiques...etc.

Ils existent d'autres définitions :

- Le Data Mining est l'analyse de grandes ensembles de données observationnelles pour découvrir des nouvelles relations entre elles et de les reformuler afin de les rendre plus utilisables de la part de ses propriétaires [1] ;
- Le Data Mining est un domaine interdisciplinaire utilisant dans le même temps des techniques d'apprentissage automatiques, de reconnaissance des formes, des statistiques, des bases de données et de visualisation pour déterminer les manières d'extraction des informations de très grandes bases de données [2] ;
- Le Data Mining est un processus inductif, itératif et interactif dont l'objectif est la découverte de modèles de données valides, nouveaux, utiles et compréhensibles dans de larges Bases de Données [3].

1.3 Les étapes du processus de data mining :

La création d'un modèle d'exploration de données fait partie d'un processus plus vaste qui va d'un ensemble de requêtes de consultation des données, la création d'un modèle afin d'y répondre et le déploiement du modèle obtenu dans un environnement de travail. Ce processus peut être décrit à l'aide des six étapes de base suivantes [13]:

- ❖ **Collecte des données** : la combinaison de plusieurs sources de données, souvent hétérogènes, dans une base de données. [4] [5]
- ❖ **Nettoyage des données** : la normalisation des données : l'élimination du bruit (les attributs ayant des valeurs invalides et les attributs sans valeurs) [4][5]
- ❖ **Sélection des données** : Sélectionner de la base de données les attributs utiles pour une tâche particulière du data mining [6]
- ❖ **Transformation des données** : le processus de transformation des structures des attributs pour être adéquates à la procédure d'extraction des informations [7]
- ❖ **Extraction des informations (Data mining)**: l'application de quelques algorithmes du Data Mining sur les données produites par l'étape précédente (*Knowledge Discovery in Data bases*, ou KDD) [5][6]
- ❖ **Visualisation des données** : l'utilisation des techniques de visualisation (histogramme, camembert, arbre, visualisation 3D) pour exploration interactive de données (la découverte des modèles de données) [7][5]
- ❖ **Evaluation des modèles** : l'identification des modèles strictement intéressants en se basant sur des mesures données [4]

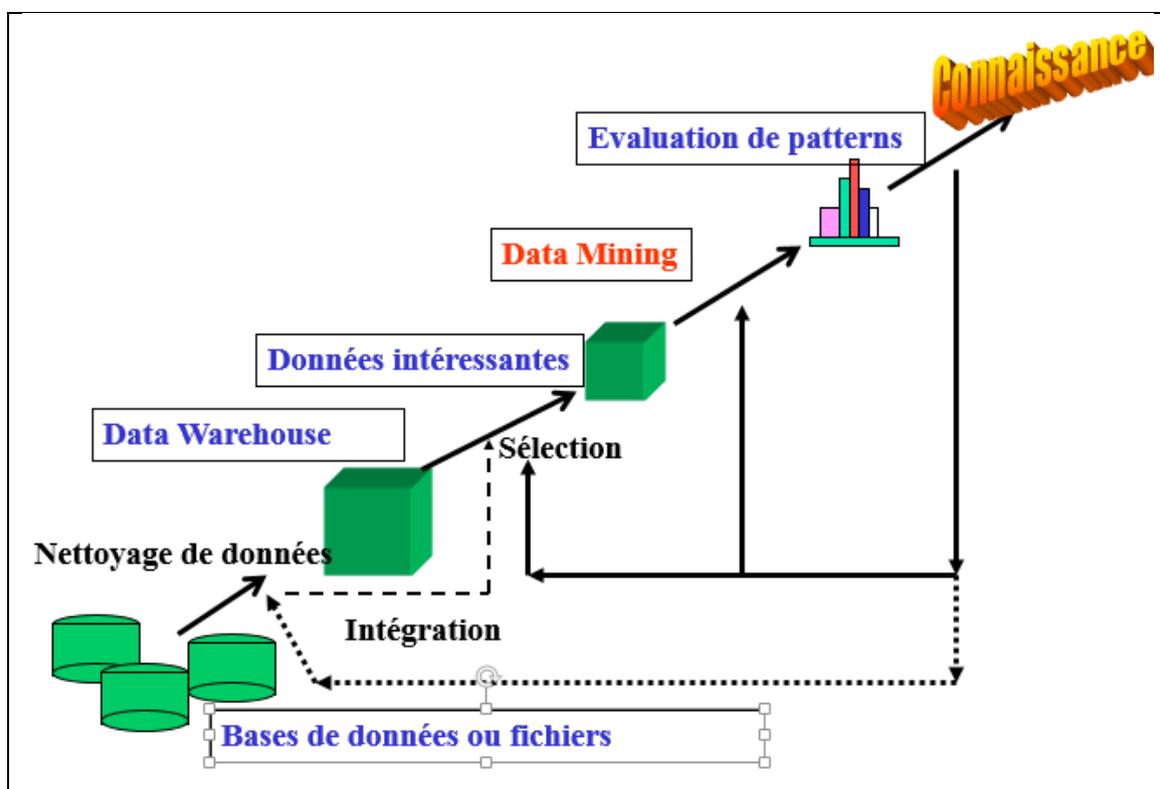


Figure 1.1 :Présentation classique du processus de l'ECD [1]

1.4 Motivation du Data Mining :

1.4.1 Problème de l'explosion de données :

- Masse importante de données (millions de milliards d'instances) : elle double tous les 20 mois.
 - BD très larges - Very Large Data bases (VLDB)
- Données multi-dimensionnelles (milliers d'attributs)
 - BD denses
- Inexploitables par les méthodes d'analyse classiques
- Collecte de masses importantes de données (Gbytes/heure)
 - Données satellitaires, génomiques (micro-arrays, ...),
 - simulations scientifiques, etc.
- Besoin de traitement en temps réel de ces données. [8]

1.4.2 Améliorer la productivité :

- Forte pression due à la concurrence du marché ;
- Brièveté du cycle de vie des produits ;
- Besoin de prendre des décisions stratégiques efficaces ;
 - Exploiter le vécu (données historiques) pour prédire le futur et anticiper le marché
 - individualisation des consommateurs (dé-massification).[8]

1.4.3 Croissance en puissance/coût des machines capables :

- de supporter de gros volumes de données ;
- d'exécuter le processus intensif d'exploration ;
- hétérogénéité des supports de stockage. [8]

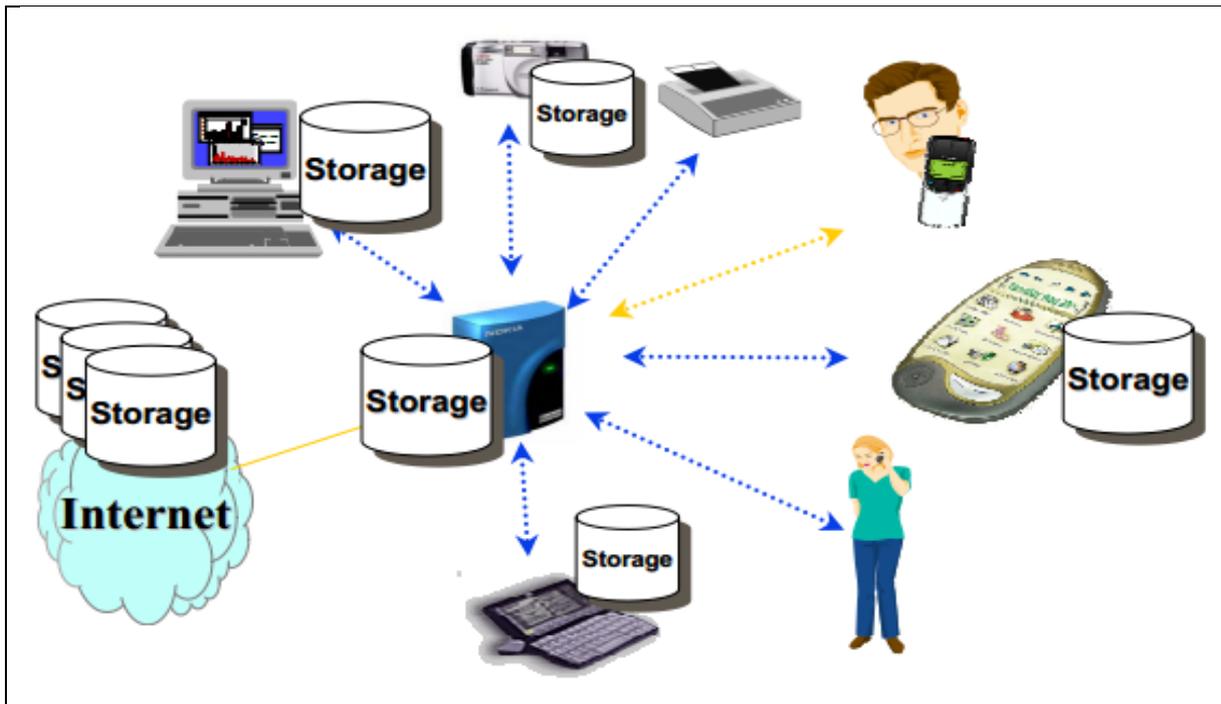


Figure1. 2 : Masse importante de données– supports hétérogènes [8]

1.5 Data mining sur quels types de données ?

Le data mining est applicable à n'importe quel type d'information, il n'est pas spécifique à un type particulier de données. Il est utilisé dans plusieurs types de base de données tel que :

- **Les bases de données multimédia** : Les bases de données multimédia incluent la vidéo, les images et l'audio. Elles peuvent être stockées sur des bases de données orientées objets ou relationnelles. Le multimédia est caractérisé par sa dimensionnalité élevée ce qui rend le datamining sur ce type de données très difficile. [9]
- **Le World Wide Web** : Le World Wide Web (WWW) est le dépôt le plus hétérogène et le plus dynamique disponible. Un nombre massif d'utilisateurs accèdent à ses ressources quotidiennement. Les données dans le World Wide Web sont organisées dans des documents inter-connectés. Ces documents peuvent être des textes, audio, vidéos, données brutes, et même des applications. [4]
- **Les fichiers plats** : Les fichiers plats sont actuellement la source de donnée la plus commune pour les algorithmes du data mining et particulièrement au niveau de recherches. Les fichiers plats sont des fichiers de données simples dans un format binaire ou textuelle avec une structure connue par l'algorithme du data mining qui va-t-être y appliqué. [9]

- **Les bases de données relationnelles** : Une base de données relationnelle se compose d'un ensemble de tables. Les tableaux ont des colonnes et des lignes, où les colonnes représentent les attributs et les lignes représentent les tuples. Les algorithmes de Data mining utilise les bases de données relationnelles peuvent être plus souples que des algorithmes de data mining spécifiquement écrits pour les fichiers plats, puisqu'ils peuvent tirer avantage de la structure inhérente aux bases de données relationnelles. Le Data Mining peut profiter du SQL (Structured Query Language) pour la sélection, la transformation et la consolidation. [9]
- **Les data warehouses** : Un Data Warehouse est un support de données dans laquelle est centralisé un volume important de données consolidées à partir des différentes sources de données (souvent hétérogènes). [4]
- **Les bases de données transactionnelles** : en général, une base de données transactionnelle est un fichier où chaque enregistrement représente une transaction. Une transaction contient un identifiant unique de transaction (transaction ID) et une liste d'items composant la transaction (par exemple : les achats d'un client lors d'une visite). [4]

1.6 Taches du Data Mining :

Le Data Mining remplit l'une des tâches suivantes :

- La classification
- L'estimation
- La prédiction
- Le groupement par similitude
- L'analyse des clusters.

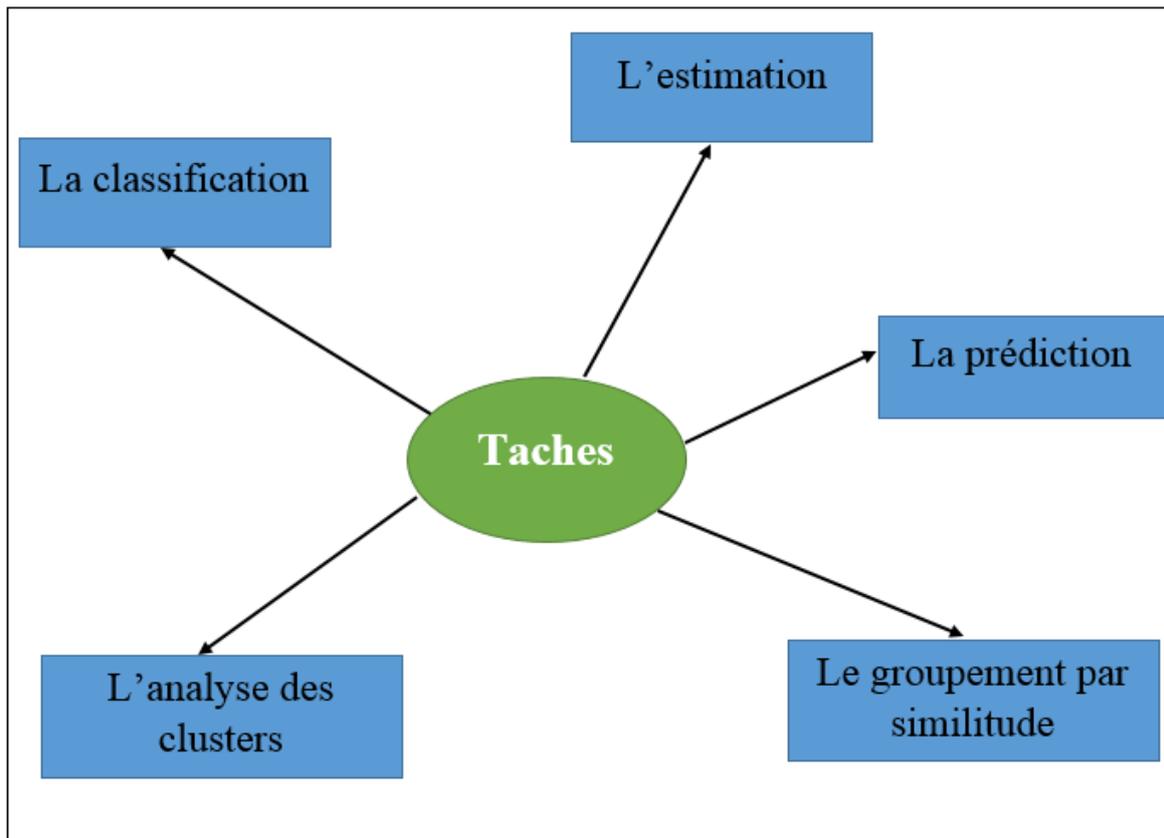


Figure1.3 : Tâches du Data Mining

1.6.1 La classification :

La classification consiste à examiner les caractéristiques d'un élément nouvellement présenté afin de l'affecter à une classe d'un ensemble prédéfini. Elle se rapporte à des événements discrets (homme/femme, client bon payeur/ mauvais payeur, etc.)[10]

1.6.1.1 Classification non supervisée :

- Établir des représentations des données dans des espaces à faible dimensions pour y lire des typologies d'individus
 - Nombre de classes initialement inconnu
- Méthodes :
 - Analyse en composante principales (ACP),[11]

1.6.1.2 Classification supervisée (apprentissage) :

- Obtenir un critère de séparation destiné à prédire l'appartenance à une classe
- Nombre de classes initialement connu

- Méthodes :
 - Analyse discriminante
 - k-Plus proches voisins
 - Arbre de décision
 - Réseau de neurones

Quelques exemples de l'utilisation des tâches de classification dans les domaines de Recherche et commerce sont les suivants :

- Déterminer si l'utilisation d'une carte de crédit est frauduleuse.
- Diagnostiquant si une certaine maladie est présente.
- Déterminer quels numéros de téléphone correspondent aux fax.
- Déterminer quelles lignes téléphoniques sont utilisées pour l'accès à Internet. [10]

1.6.2 La prédiction :

La prédiction vise à prédire la valeur future d'un champ. Tout comme la tâche précédente, elle s'appuie sur le passé et le présent mais son résultat se situe dans un futur généralement précisé.

1.6.3 L'optimisation :

L'optimisation est le problème qui consiste à optimiser un ou plusieurs paramètres du système selon un ensemble de contraintes. Pour résoudre de nombreux problèmes, il est courant pour chaque solution potentielle d'y associer une fonction d'évaluation. Le but de l'optimisation est de maximiser ou minimiser cette fonction.

1.6.4 La segmentation (analyse des clusters) :

L'analyse des clusters consiste à segmenter une population hétérogène en sous-populations homogènes (groupes). Il appartient ensuite à un expert du domaine de déterminer l'intérêt et la signification des groupes ainsi constitués. Contrairement à la classification, les sous-populations ne sont pas préétablies.

1.6.5 L'association :

L'association est la tâche qui consiste à rechercher les relations ou les dépendances existantes entre plusieurs caractéristiques d'un individu. [11]

1.7 Les domaines d'application de Data Mining:

La technologie de data mining a une grande importance économique grâce aux possibilités qu'elle offre pour optimiser la gestion des ressources (humaines et matérielles).

Les domaines d'application actuels du data mining sont les suivants [13] :

1.7.1 Le data mining dans la banque :

- Prédire la réaction des clients aux changements des taux d'intérêt.
- Identifier les clients qui seront les plus réceptifs aux nouvelles offres de produits.
- Identifier les clients "fidèles".
- Déterminer les clients qui posent le risque le plus élevé de manquer à leurs engagements aux prêts.
- Détecter les activités frauduleuses dans les transactions par cartes de crédit.
- Prédire les clients qui sont susceptibles de changer leurs cartes d'affiliation au cours du prochain trimestre.
- Déterminez les préférences des clients pour les différents modes de transaction à savoir par le biais de guichets ou par l'intermédiaire de cartes de crédit, etc.

1.7.2 Le data mining dans l'assurance :

- Des produits obligatoires (automobile, habitation) :
- Analyse des sinistres
- Recherche des critères explicatifs du risque ou de la fraude,
- Extraction de caractéristiques à partir de ces déclarations (type d'accident, de blessures, etc.)

1.7.3 Le data mining dans le commerce :

- La prédiction de la propension du client à acheter.
- L'évaluation des risques pour chaque transaction.
- Connaitre la distribution et l'emplacement géographique des clients.
- L'analyse de la fidélité des clients dans les opérations à base de crédit.
- Évaluation de la menace concurrentielle dans une région.

1.7.4 Le data mining dans la détection de fraude :

- Détection de fraude de cartes de crédits.
- Détection de fraude dans les listes des électeurs en utilisant les réseaux de neurones en combinaison avec le data mining.
- La détection des fraudes dans les demandes de passeport par la conception d'un système de diagnostic par apprentissage en ligne.
- Détection de fausses demandes de remboursement médicale.

1.7.5 Le data mining dans la télécommunication :

- Analyse des achats de services de télécommunication.
- Prédiction de modèles d'appels téléphoniques.
- Gestion des ressources et de trafic réseau.
- Automatisation de la gestion du réseau et de la maintenance en utilisant l'intelligence artificielle pour diagnostiquer et réparer les problèmes de transmission du réseau, etc.

1.7.6 Le data mining dans le transport et voyage:

- Optimisation des tournées,
- Prédiction de carnets de commande,
- Marketing relationnel dans le cadre de programmes de fidélité.

1.7.7 Le Data mining dans la médecine et la pharmacie :

- Déterminer des segments de patients susceptibles d'être soumis à des protocoles thérapeutiques déterminés, chaque segment regroupant tous les patients réagissant identiquement
- Mettre en évidence des facteurs de risque ou de rémission dans certaines maladies. Choisir le traitement le plus approprié
- Pronostic des infarctus et des cancers (décès, survie)
- Optimisation des plans d'action des visiteurs médicaux pour le lancement de nouvelles molécules.
- Prédire les effets sur la peau humaine de nouveaux cosmétiques, en limitant le nombre de tests sur les animaux.[13]

1.8 Techniques du Data Mining:

Pour effectuer les tâches du Data Mining il existe plusieurs techniques issus de disciplines scientifiques diverses (statistiques, intelligence artificielle, base de données) afin de faire apparaître des corrélations cachées dans des gisements de données pour construire des modèles à partir de ces données. Dans ce chapitre, nous présentons les techniques du data mining les plus connues qui répartissent en deux grandes familles.

1.8.1 Techniques Non Supervisées :

Les techniques descriptives réduisent, résument, synthétisent les données sans existence d'un variable cible à prédire. [14] Donc le but de ces méthodes consiste à expliquer les relations existantes dans les données et identifient les associations entre des variables. Nous citons par exemple le clustering, les règles d'association...etc.

1.8.1.1 Les règles d'association :

Les règles d'association consistent à trouver quelles valeurs des variables vont ensemble. Par exemple, telle valeur d'une variable va avec telle valeur d'une autre variable.

On appelle aussi ce type d'analyse une « analyse d'affinité ». [15]

1.8.1.2 Clustering :

Le clustering ou segmentation est un apprentissage non supervisé (on ne définit pas ce qui est "sorties") qui vise à identifier des ensembles d'éléments qui partagent certaines similarités. Les algorithmes de clustering minimisent l'homogénéité à l'intérieur de chaque ensemble et maximisent l'hétérogénéité entre les ensembles.

Autrement dit, il s'agit pour n variables de créer des sous-ensembles disjoints de données. [15]

1.8.2 Techniques Supervisées :

Les techniques prédictives permettent d'expliquer une variable cible appelée : variable privilège [14]. Elles ont pour but de déterminer une fonction qui associe des entrées et des sorties, Ces techniques sont constituées de différentes méthodes. Nous citons par exemple les réseaux de neurones, les arbres de décision...etc.

1.8.2.1 Les réseaux de neurones :

Un réseau de neurones est un outil disponible dans les environnements de Data Mining. Le réseau de neurone est défini par un ensemble d'unités de traitement qui peuvent être des unités d'entrée, de sortie et caché. Le réseau de neurones se compose d'un ensemble de nœuds connectés entre eux par des liens orientés, ou connexions. Bien que n'importe quelle structure soit imaginable, la forme la plus utilisée est l'organisation en couches successives, le réseau est un système non linéaire.

Les réseaux de neurones sont pourtant largement diffusés et utilisés de manière industrielle dans de nombreux secteurs d'activité. [16]

1.8.2.2 Les arbres de décision :

Un arbre de décision, en général, a pour but de trouver les attributs explicatifs et les critères précis sur ces attributs donnant le meilleur classement vis-à-vis d'un attribut à expliquer. L'arbre de décision est un enchaînement hiérarchique de règles logiques construites automatiquement à partir d'une base d'exemple. Un exemple est constitué d'une liste d'attributs, dont la valeur détermine l'appartenance à classe donnée. La construction de l'arbre de décision consiste à utiliser les attributs pour subdiviser progressivement l'ensemble d'exemples en sous-ensembles de plus en plus fins.

1.8.2.3 L'algorithme des k-Plus proches voisins :

L'algorithme des k plus proches voisins (K-PPV, k nearest neighbor en anglais ou kNN) est un algorithme de raisonnement à partir de cas qui est dédié à la classification qui peut être étendu à des tâches d'estimation. Le but de cet algorithme est de prendre des décisions en se basant sur un ou plusieurs cas similaires déjà résolus en mémoire. Dans ce cadre, et contrairement aux autres méthodes de classification (arbres de décision, réseaux de neurones, algorithmes génétiques, ...etc.) l'algorithme de KNN ne construit pas de modèle à partir d'un échantillon d'apprentissage, mais c'est l'échantillon d'apprentissage, la fonction de distance et la fonction de choix de la classe en fonction des classes des voisins les plus proches, qui constituent le modèle. [17] [18]

❖ Les avantages et inconvénients :

- **Avantage :**

La méthode des K plus proche voisins présentent plusieurs avantages parmi lesquelles nous citons :

- Mise en œuvre simple

- Pas de modèle à construire
- Domaine de compétence bien délimité
- Efficace pour des classes réparties de manière irrégulière
- Utilisable pour des données hétérogènes (ou incomplètes)
- Nouvelles données aisément incorporables Pas de modèle à construire. [18]
- **Inconvénients :**

Les inconvénients de cette méthode sont plusieurs, parmi lesquelles on peut citer les deux principales :

- Nécessite capacité de stockage et puissance de calcul
- Besoin de nombreuses données de référence
- Sensible à la malédiction de la dimension (espace vide)
- Choisir K (alternative: choisir la taille du voisinage). [18]

1.9 Conclusion :

Le data mining est un domaine de recherche pour l'exploitation des grandes quantités de données collectées utilisées dans divers domaines d'application de l'informatique, de l'intelligence artificielle, des statistiques et des bases de données. L'idée principale est d'extraire les connaissances cachées à partir d'un tas de données disponible. Elles peuvent être sous forme de règles, de modèles, de régularités, de concepts etc...

Dans ce chapitre, nous avons présenté le domaine de data mining, et nous avons détaillé les principales techniques utilisées dans ce dernier.

Chapitre 2:

Systeme

Multi_Agents

2.1 Introduction :

Les systèmes multi-agents proposent une approche nouvelle pour le développement de systèmes à plusieurs composantes autonomes pouvant coopérer entre elles. Puisant ses fondements dans de nombreuses disciplines, ce domaine de recherche est extrêmement riche. Mais cette richesse induit une grande complexité et une grande multiplicité des approches proposées qui conduit à de très nombreux modèles d'agents, de l'environnement, des interactions, de l'organisation. Ces modèles sont d'ailleurs souvent combinés au sein d'un même système multi-agent.

Dans ce chapitre nous allons aborder présentons le paradigme agent. Avant de commencer, il serait intéressant de comprendre le concept agent, les caractéristiques et la typologie d'agents, nous présenterons par la suite le système multi-agent, les architectures et les domaines d'application.

2.2 Définition d'un Agent :

Un agent est un programme informatique qui est situé dans un environnement et qui est doté de comportements autonomes (actions) lui permettant d'atteindre, dans cet environnement, les objectifs qui lui ont été fixés lors de sa conception. Un agent logiciel est une entité autonome capable de communiquer, disposant d'une connaissance partielle de ce qui l'entoure et d'un comportement privé, ainsi que d'une capacité d'exécution propre. Un agent agit pour le compte d'un tiers (un autre agent, un utilisateur) qu'il représente sans être obligatoirement connecté à celui-ci, il réagit et interagit avec d'autres agents.

On appelle agent une entité physique ou virtuelle :

1. qui est capable d'agir dans un environnement ;
2. qui peut communiquer directement avec d'autres agents ;
3. qui est mue par un ensemble de tendances (sous forme d'objectifs individuels ou d'une fonction de satisfaction, voire de survie, qu'elle cherche à optimiser) ;
4. qui possède des ressources propres ;
5. qui est capable de percevoir (mais de manière limitée) son environnement ;
6. qui ne dispose que d'une représentation partielle de cet environnement (et éventuellement aucune) ;
7. qui possède des compétences et offre des services ;

8. qui peut éventuellement se reproduire ;
9. dont le comportement tend à satisfaire ses objectifs, en tenant compte des ressources et des compétences dont elle dispose, et en fonction de sa perception, de ses représentations et des communications qu'elle reçoit. [19]

2.3 L'environnement :

L'environnement dans lequel se déplacent les agents est important. On peut le définir comme un système duquel l'agent doit apprendre, au fil des expériences, les éléments nécessaires à la réalisation de son objectif. L'agent fait partie de ce système et peut le modifier.

2.3.1 L'environnement d'un agent :

Il est constitué de tout ce qui l'entoure : les objets ou les autres agents. On distingue deux types d'environnement, physique et social, pour un agent :

- ❖ **L'environnement physique** : il correspond à tous les objets que l'agent peut percevoir ou sur lesquels il peut agir. Par exemple, si l'on considère une fourmi comme étant un agent, son environnement physique est composé de nourriture et d'obstacles.
- ❖ **L'environnement social** : il correspond aux autres agents avec lesquels un agent est en interaction par envoi de messages par exemple. On parle d'environnement social surtout dans le cas d'agents communicants.

2.3.2. L'environnement du système :

Il est composé de tout ce qui n'est pas dans le système : les objets ou d'autres agents faisant partie d'un autre système multi-agent. L'environnement du système est-ce qui va le guider vers une adéquation fonctionnelle dans le cas de systèmes auto-organiseurs. [21]

2.4 Caractéristiques d'agents :

- **Autonome** : l'agent est capable d'agir sans l'intervention d'un tiers (humain ou agent) et contrôle ses propres actions ainsi que son état interne
- **Social** : l'agent doit être capable d'interagir avec des autres agents (logiciels ou humains) afin d'accomplir des tâches ou aider ces agents à accomplir les leurs.
- **Proactif** : l'agent doit exhiber un comportement proactif et opportuniste, tout en étant capable de prendre l'initiative au bon moment.

- **Coopération** : capable de faire la coordination avec d'autres agents pour atteindre l'objectif commun
- **Mobilité** : l'agent peut être mobile, capable de se déplacer vers un autre environnement
- **Rationalité** : l'agent est capable d'agir en fonction de ses objectifs internes et ses connaissances
- **Apprentissage** : l'agent est capable d'évoluer et d'apprendre ; en fonction de cet apprentissage, il est capable de changer de comportement. [22]

2.5 Typologie des agents :

2.5.1 Agents réactifs : agents sans intelligence (sans anticipation, sans planification) qui réagissent par stimulus-réponse à l'état courant de l'environnement. Des comportements intelligents peuvent émerger de leur association.

- Pas de représentation explicite.
- Organisation implicite/induite
- Communication via l'environnement

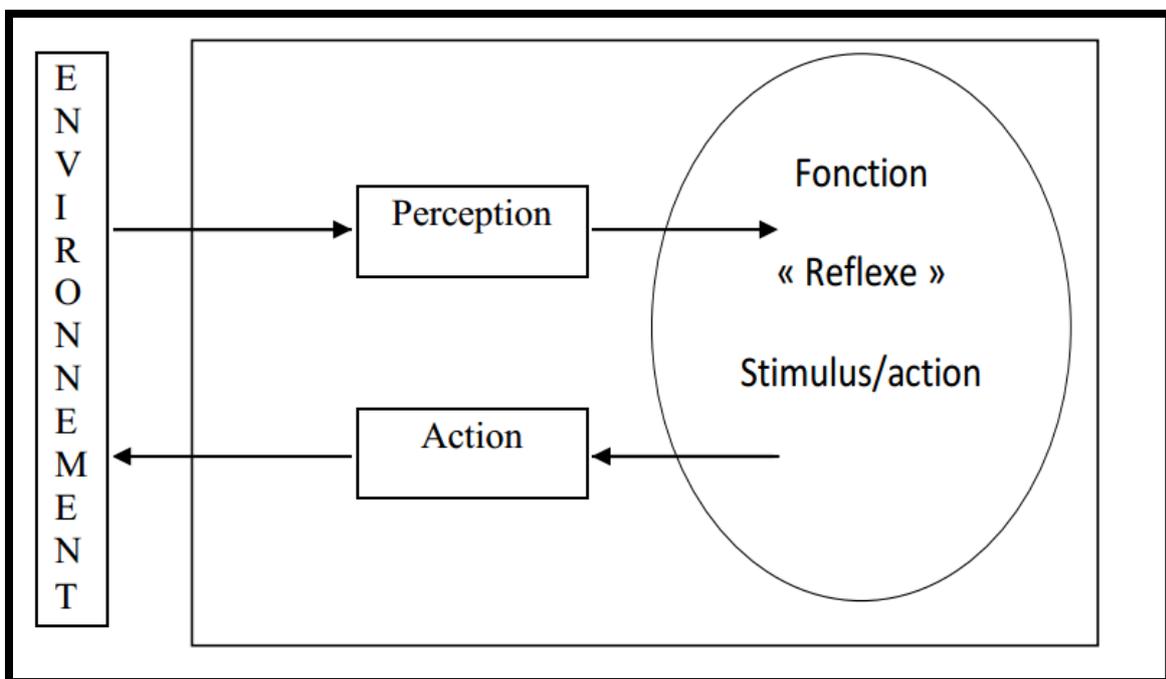


Figure2.1: Structure d'un agent réactif

2.5.2 Agents cognitifs : chaque agent est spécialisé dans un domaine et sait communiquer avec les autres. Ils possèdent des buts et des plans explicites leur permettant d'accomplir leurs buts. (Analogie avec les groupes en sociologie).

- Représentation explicite de soi, environnement et les autres agents
- Organisation explicite.
- Interaction explicite et élaborée. [23]

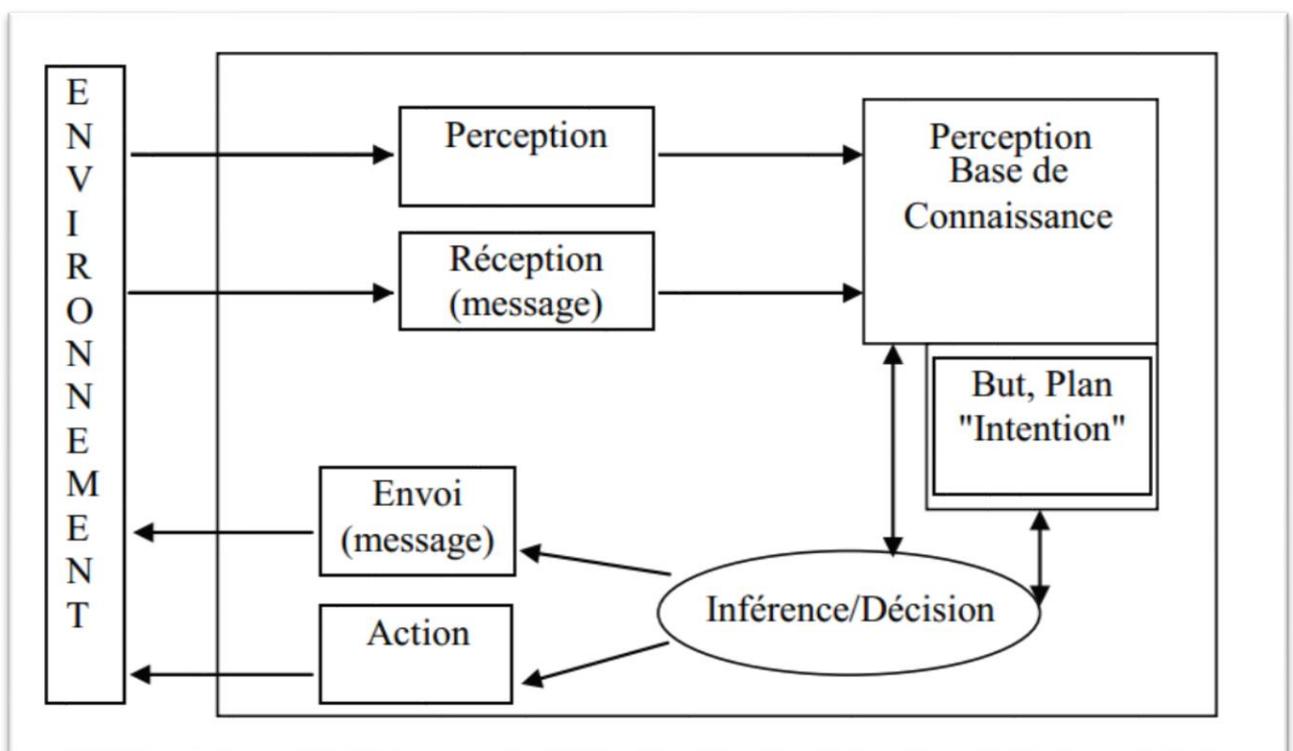


Figure2.2 : Structure d'un agent cognitif

Systèmes d'agents cognitifs	Systèmes d'agents réactifs
Représentation explicite de l'environnement Petit nombre d'agents	Pas de représentation explicite
L'agent peut tenir compte de son passé	Pas de mémoire de son histoire
Agents complexes	Fonctionnement stimulus/Reponse
Petit nombre d'agent	Grand nombre d'agent

Tableau 2. 1: Les agents cognitifs et réactifs [13]

2.6 Les systèmes multi-agents :

2.6.1 Définition d'un SMA :

En informatique, un système multi-agents (SMA) est un système composé d'un ensemble d'agents, situés dans un certain environnement et interagissant selon certaines relations. Un agent est une entité caractérisée par le fait qu'elle est, au moins partiellement, autonome. Ce peut être un processus, un robot, un être humain, etc.

Objet de longue date de recherches en intelligence artificielle distribuée, les systèmes multi-agents forment un type intéressant de modélisation de sociétés, et ont à ce titre des champs d'application larges, allant jusqu'aux sciences humaines. [24]

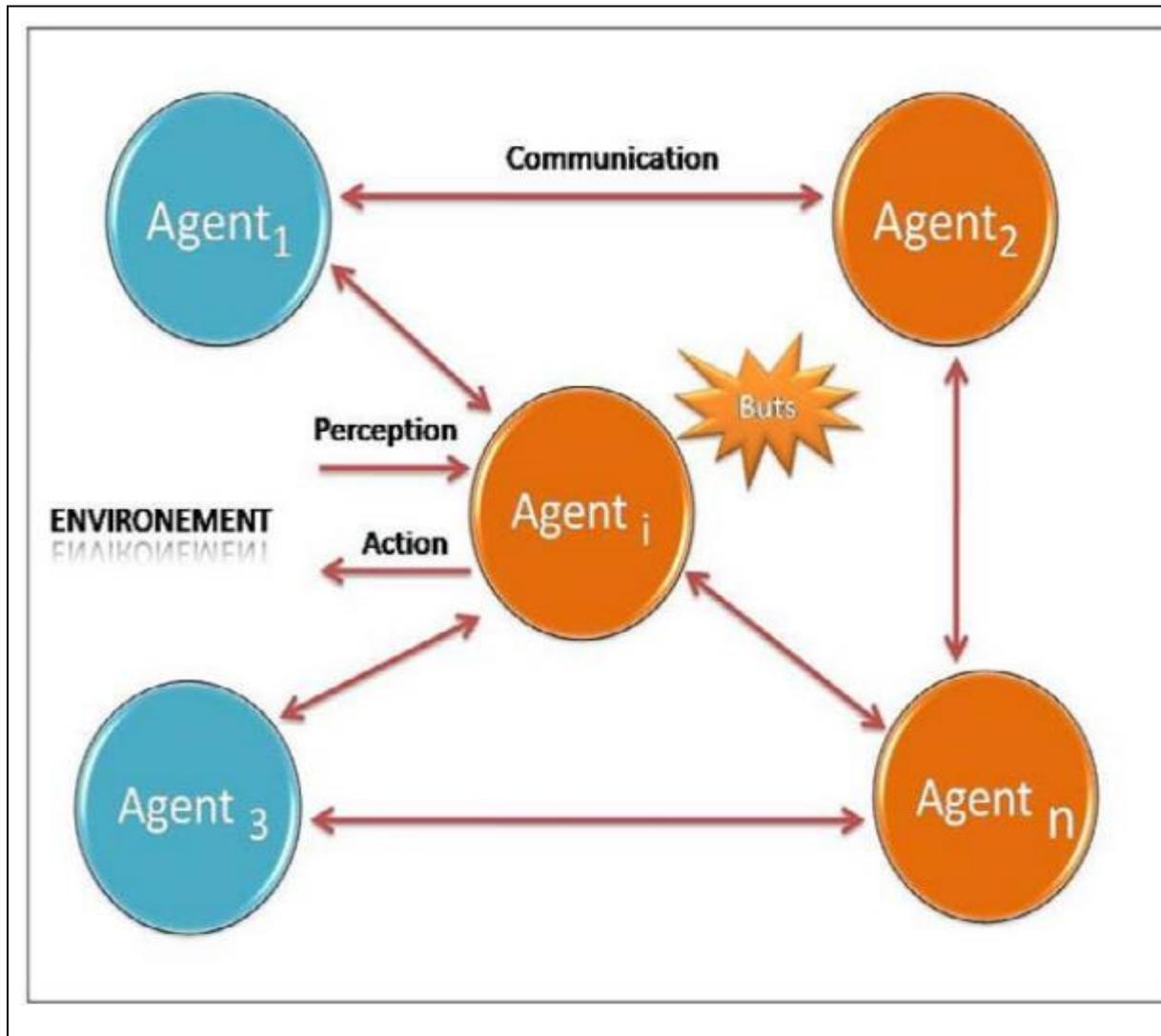


Figure2.3 : Schéma général d'un système multi-agent.

2.6.2 Les caractéristiques d'un SMA :

- **interactions**

Interaction au niveau supérieur pour :

- coordination
- communication
- organisation
- résolution distribuée de problèmes

- **Coordination**

- agents coopératifs

- agents individualistes
- agents antagonistes – compétition

Comment réaliser la coordination – négociation, coalitions

- **Communication**

- langage de communication
- protocole de communication
- sémantique de la communication

- **Organisation**

- centralisée vs décentralisée
- hiérarchique

Apprentissage / adaptation [25]

2.6.3 Communication entre agents :

Les agents communiquent entre eux à l'aide de protocoles de communication. Leur communication peut être scindée en deux types :

- Partage d'information
- Envoi de messages.

a) Communication par partage d'information :

Les agents ne sont pas en liaison directe mais communiquent via une structure de données partagée, où on trouve les connaissances relatives à la résolution (état courant du problème) qui évoluent durant les processus d'exécution. Cette manière de communiquer est l'une des plus utilisées dans la conception des systèmes multi-agents. Le meilleur exemple qui utilise ce type de communication est les systèmes à tableau noir.

Ce mode de communication n'existe pas dans les SMA où l'on ne dispose qu'une vision partielle du système, car la communication par partage d'informations suppose l'existence d'une base partagée sur laquelle les agents viennent lire et écrire.

Ce type de communication est utilisé quand:

- il y a recouvrement des domaines d'expertise de chaque agent.

- Il suppose également que les agents ne possèdent qu'une connaissance limitée sur les domaines d'activité des autres agents et pose des problèmes de synchronisation. [26]

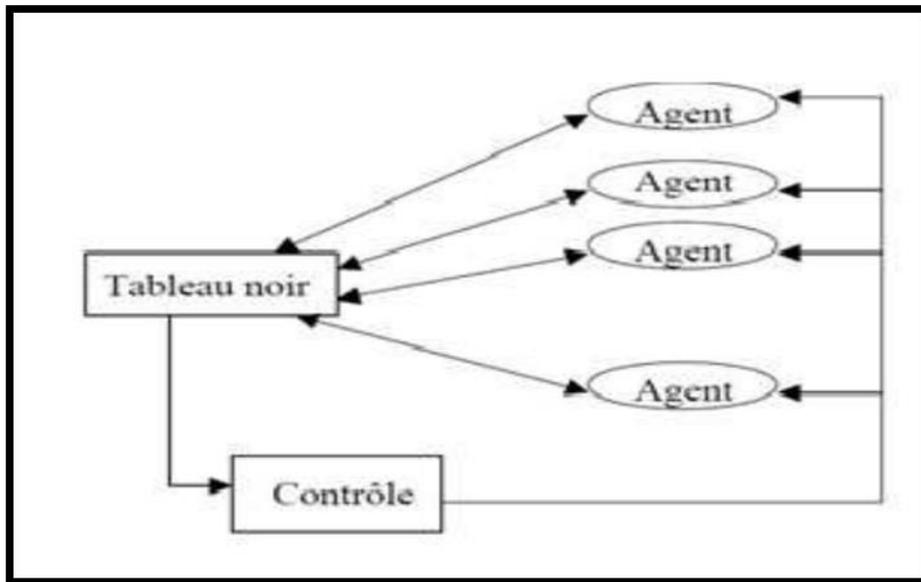


Figure2.4:Communication entre agents par Partage d'information [26]

b) Communication par envoi de messages :

Les agents sont en liaison directe et envoient leurs messages directement et explicitement au destinataire. La seule contrainte est la connaissance de l'agent destinataire: « Si un agent A connaît l'agent B, alors il peut entrer en communication avec lui » [26]

- **Mode point à point** : l'agent émetteur du message connaît et précise l'adresse de ou des agent(s) destinataire(s). Ce type de communication est généralement le plus employé par les agents cognitif.

- **Mode par diffusion** : le message est envoyé à tous les agents du système. Ce type de transmission est très utilisé dans les systèmes dynamiques ainsi que les systèmes d'agent réactif. En fait, ceci suppose en général une messagerie : un agent spécialisé gère autant de files d'attente que de destinataires, chaque agent peut traiter le premier message de sa file.

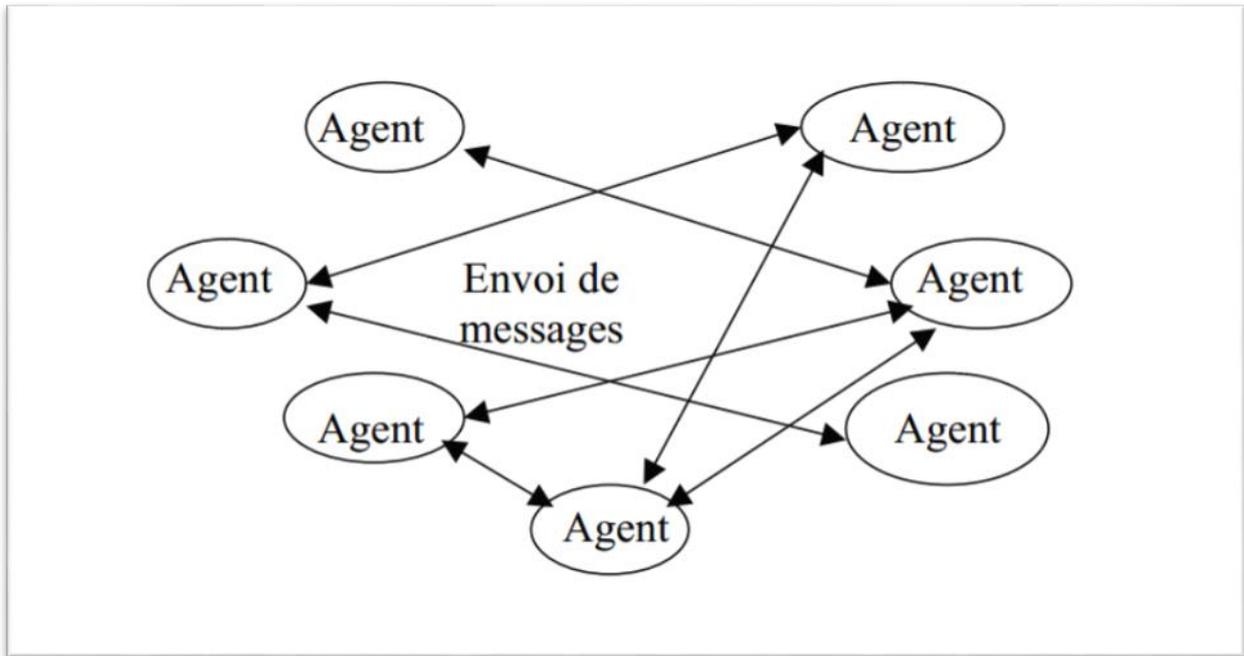


Figure2.5:Communication par envoi de messages

2.7 Domaines d'application des SMA :

- Résolution de problèmes
- Modélisation, Simulation et Analyse d'entités distribuées
- Applications Distribuées
- Assistance aux humains [27]

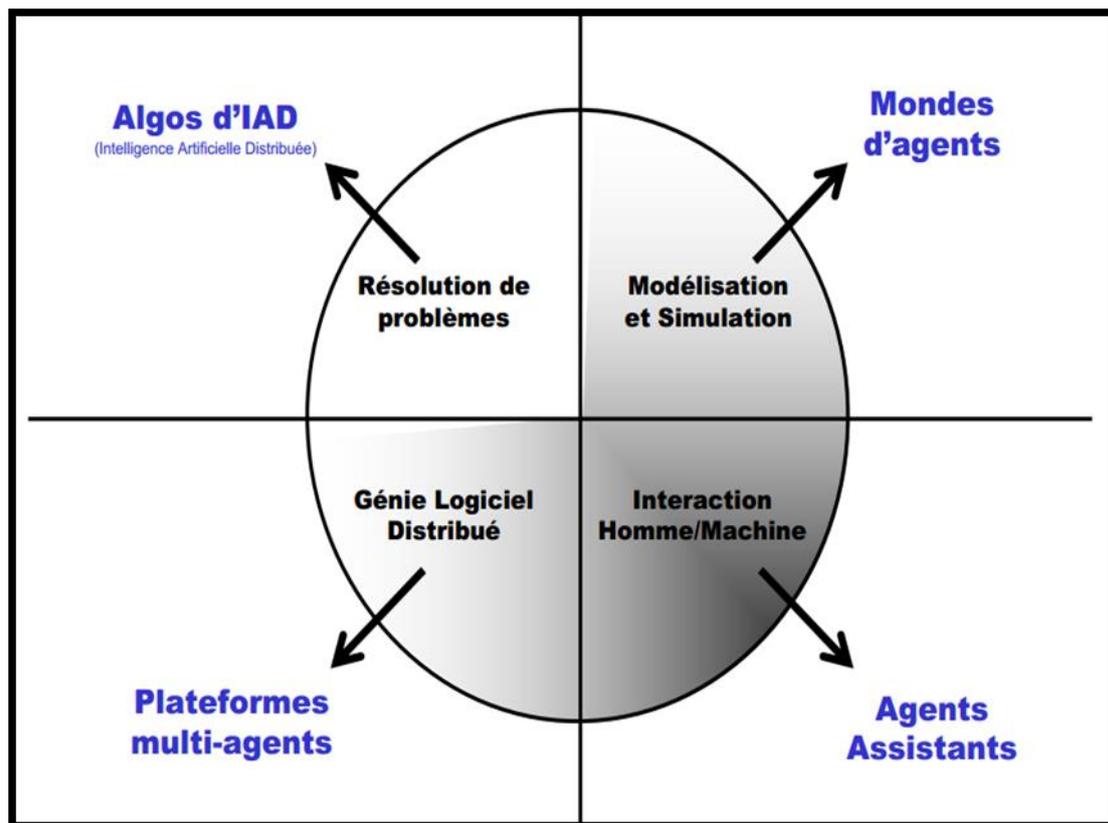


Figure2.6: Domaines d'application des SMA [27]

2.8 Quelques Systèmes existants du Data Mining basé Agents :

2.8.1 Le Système MAD-IDS(ADMI) (Intrusion Détection System) 2011 :

a)Présentation :

Un système qui aide à surveiller les événements qui se produisent dans un système distribué d'ordinateurs ou d'un réseau et analyse les signes d'intrusions est connu comme le système de détection d'intrusion (IDS). Les IDS doivent être dotés de précision, d'adaptation et extensibilité. Bien que bon nombre des techniques établies et de produits commerciaux existent, leur efficacité laisse place à l'amélioration. Un grand nombre de recherches ont été menées sur la détection de l'intrusion dans un environnement distribué pour pallier les inconvénients de l'approche centralisée. Cependant, l'IDS distribué souffre d'un certain nombre d'inconvénients par exemple, des taux élevés de faux positifs, la faiblesse de l'efficacité...etc. Dans cet exemple, Le système présenté repose sur un ensemble d'agents intelligents qui recueillent et analysent les connexions réseau et les techniques de data mining sont avérés utiles pour détecter les intrusions, Effectuer des expertises, ce qui a montré des performances supérieures des IDS distribués par rapport aux systèmes de détection décentralisés classiques. [28]

b) Le principe de fonctionnement:

Fondamentalement, le système se décompose en deux IDS principales qui peuvent être distinguées : La détection d'anomalies et l'utilisation abusive. L'objectif du projet est de jouer les deux rôles décrits précédemment (détection d'anomalies, détection d'abus) à la fois, et le principe de fonctionnement du système est expliqué comme suit [28] :

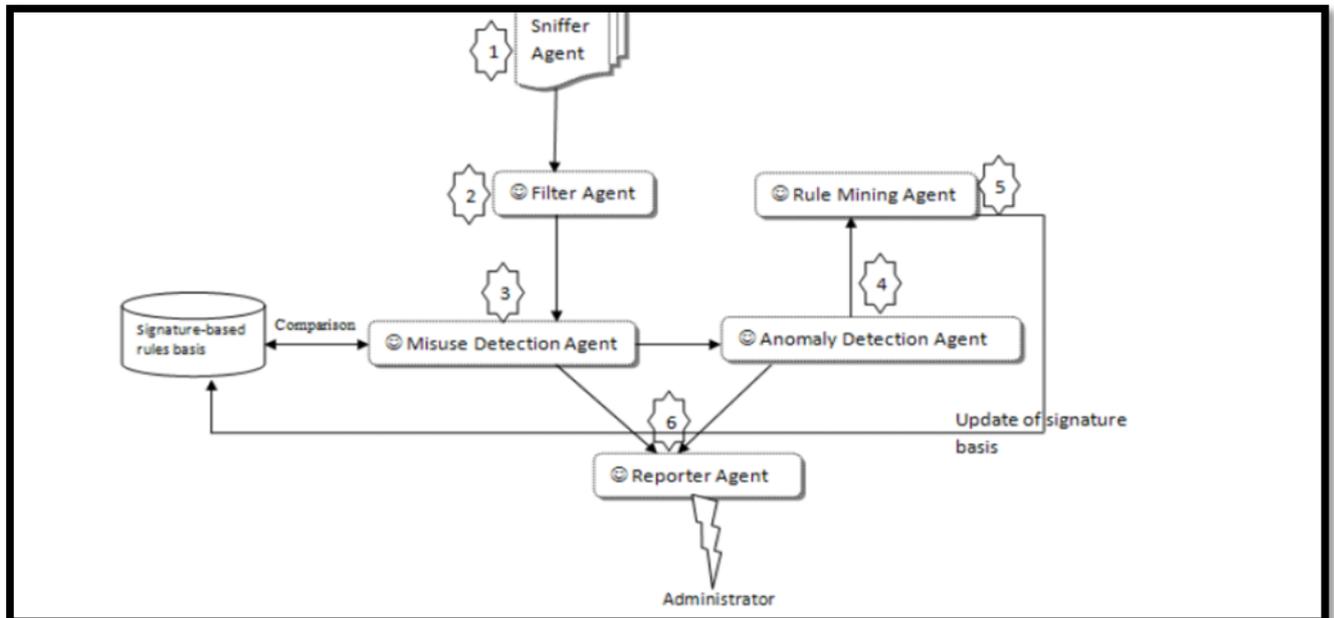


Figure2.7: Architecture générale du système MAD-IDS

1 : Sniffer Agent est un agent qui joue le rôle d'éclaireur, c'est lui qui lance la connexion à la source de données (locale ou distante) et rassemble des informations concernant une transaction prédéfinie.

2 : Les paquets sont vérifiés par le Filter Agent.

3 : Les paquets jugés correctes sont passés à l'Agent détecteur d'abus (Amusus Détection Agent) Dans le cas de test positif, les paquets victimes d'abus sont marqués et l'adresse de l'agresseur est retenue par l'agent de détection d'abus.

4 : les paquets jugés abimés ou anormaux sont examinés par l'agent de détection des anomalies (Anomalies détection Agent) qui va exécuter un algorithme de clustering afin de déterminer le type d'anomalies et l'adresse d'où vient l'intervention malveillante.

5 : le Rule Mining Agent est un agent qui récupère le résultat du détecteur d'abus et du détecteur d'anomalies aussi pour appliquer l'algorithme d'extraction de règles qui va aider à cerner les sites malveillants selon l'abus ou l'anomalie trouvée et composer une carte

représentant la situation actuelle du réseau et la mise à jour de la table des signatures des agressions connues.

6 : Le Reporter Agent va rédiger un journal ou un rapport de détection et un aperçu sur la situation pour le présenter à l'administrateur (l'utilisateur) du réseau ou à celui qui a lancé le processus de détection.

2.8.2-Le système PAPHYRUS 1999 :

a)-Présentation :

L'exploration de données est un problème pour le quelle le clustering fournit une alternative compétitive aux ordinateurs spécialisés de haute performance pour les grands ensembles de données distribuées. Des Nœuds (grappes, points) fournissent une infrastructure naturelle pour les grands ensembles de données distribués. Ces nœuds peuvent être connectés par des réseaux de matières premières pour former ce que nous appelons méta-clusters et des réseaux de haute performance pour former ce que nous appelons des super-amas. Papyrus est conçu pour les données des processus de fouille réparties sur des clusters, des méta-clusters, et super-amas. Nous décrivons dans un paragraphe le principe de fonctionnement de ce système.

[28]

b)- principe de fonctionnement :

Papyrus est un système de type Standard MADM (Multi Agents Data Mining avec architecture distribuées standard) constitué de couches d'outils logiciels et de services réseau dédiés aux systèmes distribués pour l'exploration de données et le calcul intensif. Les applications Papyrus peuvent être conçues pour accéder à l'une des quatre couches suivantes, selon les besoins. Les différentes techniques de datamining utilisent des stratégies possibles, ces dernières dépendent des données elles-mêmes, de la distribution appliquée, des ressources disponibles et des occurrences requises tel que [28]:

- MR : Move Result, la possibilité de déplacer l'ensemble des résultats volumineux et extensibles d'un processus de fouille locale vers un site central d'un réseau.
- MM : Move Model, la capacité à déplacer des modèles prédictifs d'un site vers un autre dans un réseau.
- MD : Move Data, la possibilité de déplacer des volumes gigantesque de données à traiter d'un site à un autre.

Le but principal du système est de réduire la charge aux applications qui peuvent déplacer des données (MD) d'un nœud à l'autre à l'aide de la couche inférieure. L'application des modèles de déplacement (MM) et les résultats (MR) des nœuds obtenus sont obtenus à l'aide de la couche supérieure.

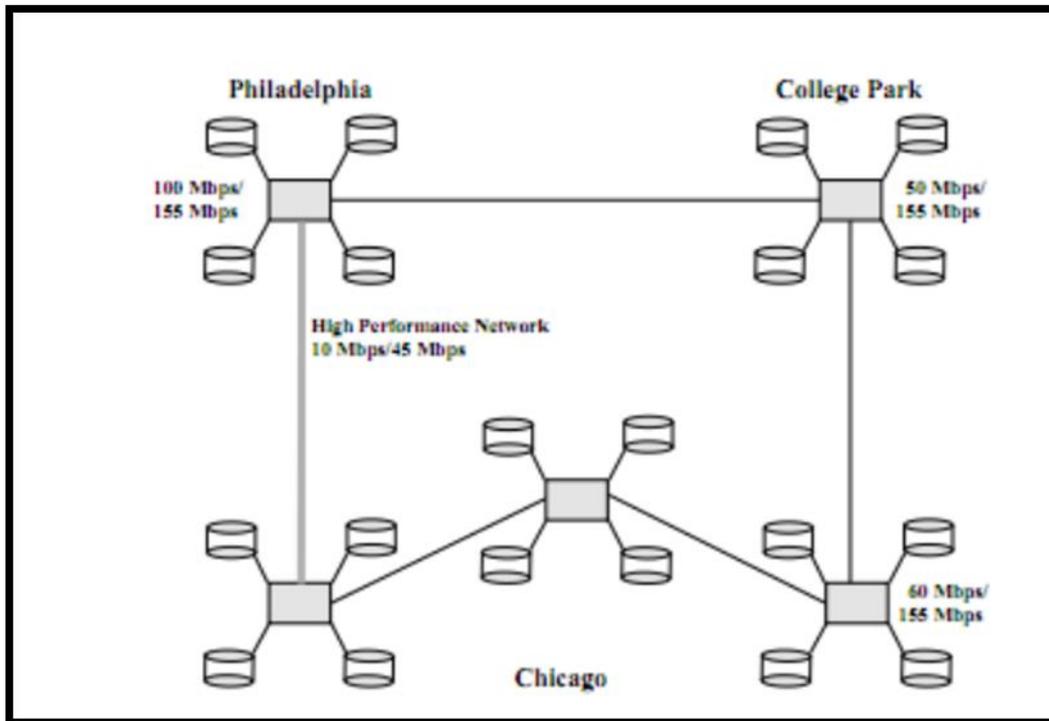


Figure2.8:Un cluster de stations de travail obtenu avec Papyrus

Les autres systèmes se concentrent sur l'application du MR, MM mais le système Papyrus s'occupe d'appliquer les trois opérations au niveau d'un réseau dans un cadre distribué à base d'agents pour bien alléger la transmission afin, par conséquent on pourra lancer d'autres processus sans craindre la saturation du réseau. [28]

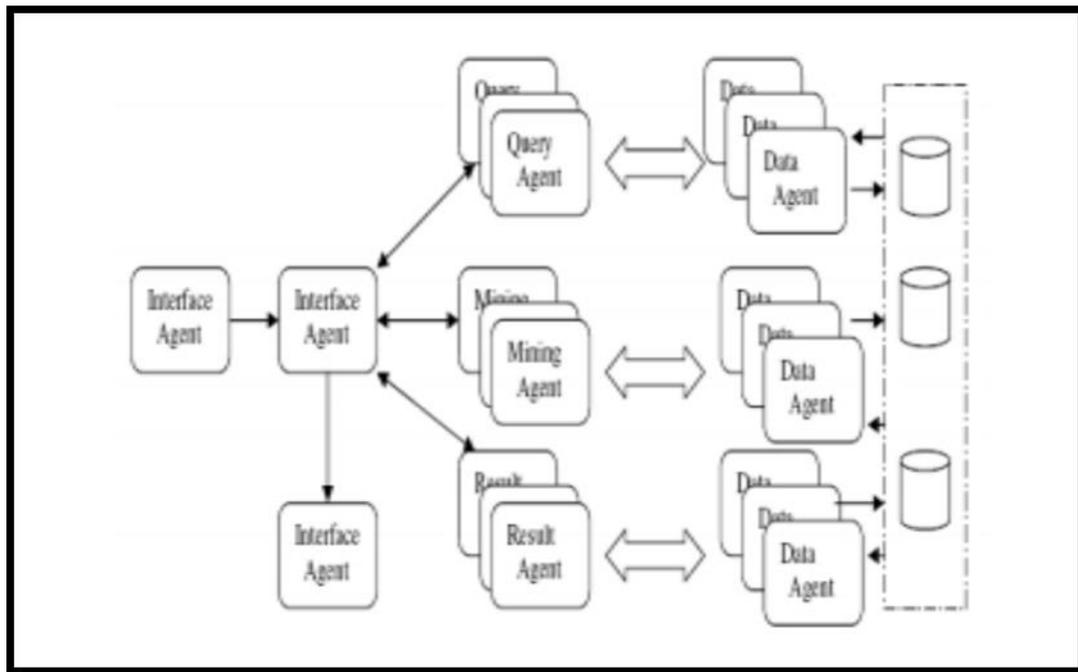


Figure 2.9: Architecture générale d'un Système MADM

2.8.3 Etude comparative des systèmes DMBA :

a) Les Critères de comparaison :

Pour pouvoir comparer les systèmes de Data Mining basé agents présentés précédemment, on présente l'ensemble des critères suivants :

- 1) Tâches Data Mining supportées par le système ?
- 2) Communication entre agents pendant ou après le processus Data Mining (pendant ou après l'extraction des connaissances) ?
- 3) Est-ce que le système peut réutiliser des algorithmes Data Mining existants sans modification ?
- 4) Qui fait le choix de la technique et l'algorithme réalisant une tâche données
- 5) Le système est-il interactif ?
- 6) Le type de sources de données ?
- 7) Le type de symbiose entre les deux technologies (Agents – Data Mining) ?
- 8) Le type de synchronisation appliquée entre les agents ?

Le Tableau 1 répond aux questions concernant les critères pour chaque projet en ce qui concerne l'interactivité, la distribution des tâches, la réutilisabilité et le nombre de sources de

données manipulables. Ensuite, on pourra choisir l'architecture la plus adaptée à notre système.

b) Tableau comparatif des systèmes étudiés :

Le Critère étudié	PAPYRUS	MAD-IDS
Tâches supportées par le système ?	clustering	Clustering Règles d'Association
Communication avant, pendant, après le processus data mining ?	Après Chaque étape Du processus	Au Début et fin Du processus
Réutilisation des algorithmes de data mining ?	OUI	NON
Qui fait le choix de la technique et l'algorithme réalisant une tâche data mining ?	La technique et l'algorithme sont adaptatifs	La technique et l'algorithme sont prédéterminés
Interactivité avec l'utilisateur ?	Oui	Oui
Le type de sources de données ?	Multi-Sources (4 supports)	Multi-sources Web
Le type de symbiose entre les deux technologies (Agents - Data Mining) ?	Contribution des Agents	Contribution des Agents
Type de Synchronisation ? (nombre d'agents qui prouvent interagir simultanément)	1	7 agents

Tableau 2.2 : Tableau comparatif des systèmes DMBA candidats

2.9 Conclusion :

Dans ce chapitre, nous avons présenté les principes de la technologie agent. Nous avons vu les caractéristiques des systèmes multi-agents.

Le prochain chapitre abordera la conception et l'implémentation de l'approche proposée.

Chapitre 3 :

Conception

3.1 Introduction :

Dans ce chapitre nous allons présenter la conception générale et détaillée de notre système où nous avons appliqué l'algorithme des K plus proches voisins pour effectuer une classification des données par un système multi-agents.

3.2 Conception générale et détaillée du système:

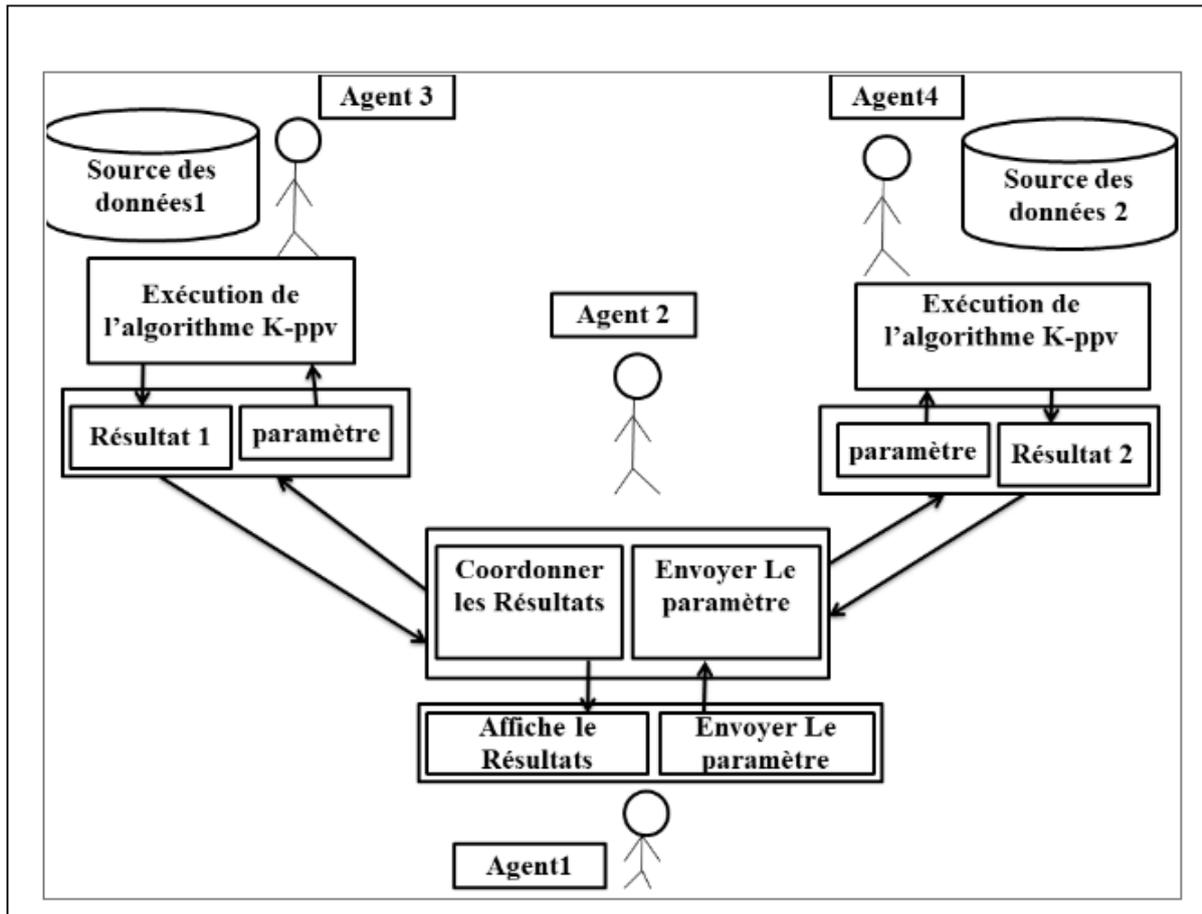


Figure 3.1 : Architecture générale du système

Dans notre projet, on va tester si une personne peut prendre un prêt dans une Banque ou non.

On va utiliser la combinaison entre la méthode des K plus proches voisins et le système multi-agent pour la résolution de ce problème.

➤ **Agent Interface (Agent 1):**

Il sélectionne une ligne, où la ligne dans la base de données représente une personne. Nous choisissons la valeur K qui est le nombre des plus proches voisins, puis envoyons la ligne et la valeur K à l'agent coordinateur (Agent 2) pour le test.

➤ **Agent coordinateur (Agent 2):**

Il reçoit des informations de l'Agent interface (Agent1), puis les envoie à l'Agent banque1 (Agent3) et à l'Agent banque2 (Agent4) .

Puis il reçoit le résultat local de Agent banque 1 (Agent 3) et Agent banque 2 (agent 4)

Après il fait la comparaison entre les deux résultats, si les résultats sont contraires, c-à-d l'un est un 'accepte' et l'autre 'refusé', il choisit le résultat 'accepte'. Si les résultats sont similaires, c-à-d un 'refus', il prend la valeur 'refus' et si le résultat est 'accepte' et l'autre 'accepte', il prend 'accepte'

Ensuite, Il envoie le résultat final à l'Agent interface (Agent 1)

➤ **Agent banque1 (Agent 3):**

Reçoit les informations de l'agent coordinateur, et applique la méthode K plus proches voisins. Le résultat local est la décision prise au niveau de banque1 puis il envoie le message à l'agent coordinateur (Agent 2).

➤ **Agent banque2 (Agent 4) :**

Reçoit les informations de l'agent coordinateur, et applique la K plus proches voisins. Le résultat local est la décision prise au niveau de banque2 puis il envoie le message à l'agent coordinateur (Agent 2).

3.3 Algorithme K plus proche voisin:

C'est une approche très simple et directe. Elle ne nécessite pas d'apprentissage mais simplement le stockage des données d'apprentissage. Son principe est le suivant. Une donnée de classe inconnue est comparée à toutes les données stockées. On choisit pour la nouvelle donnée la classe majoritaire parmi ses K plus proches voisins (Elle peut donc être lourde pour des grandes bases de données) au sens d'une distance choisie.

a) Distance :

Afin de trouver les K plus proches d'une donnée à classer, on peut choisir la distance euclidienne. Soient deux données représentées par deux vecteurs x_i et x_j , la distance entre ces deux données est donnée par.

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{k=1}^d (x_{ik} - x_{jk})^2} \quad (1)$$

b) Travail à réaliser :

Il s'agit d'implémenter en C l'algorithme des K-plus proches voisins pour prédire les classes de nouvelles données à partir de données étiquetées (données d'apprentissage). On commencera par le k-ppv. Le pseudo-code de cet algorithme est le suivant :

c) Algorithme K plus proche voisin :

Début

{

 For i=1 to m do

 {

 For j=1 to n do

 {

 Calculer la distance euclidienne entre mat [i][j] et entrer[i] en utilisant l'équation (1)

 dj->d(mat[i][j], entrer[i]) ;

 }

 }

 Trier les distances ;

 Choisir le K ;

 Calculer l'occurrence de classes ;

 Donner la classe qui à le max occurrence ;

```

    Fin
  }
}
    
```

3.4 Fonctionnement du système :

Cette partie consiste à présenter notre application en se basant sur le langage StarUML.

3.4.1 Diagramme de cas d'utilisation :

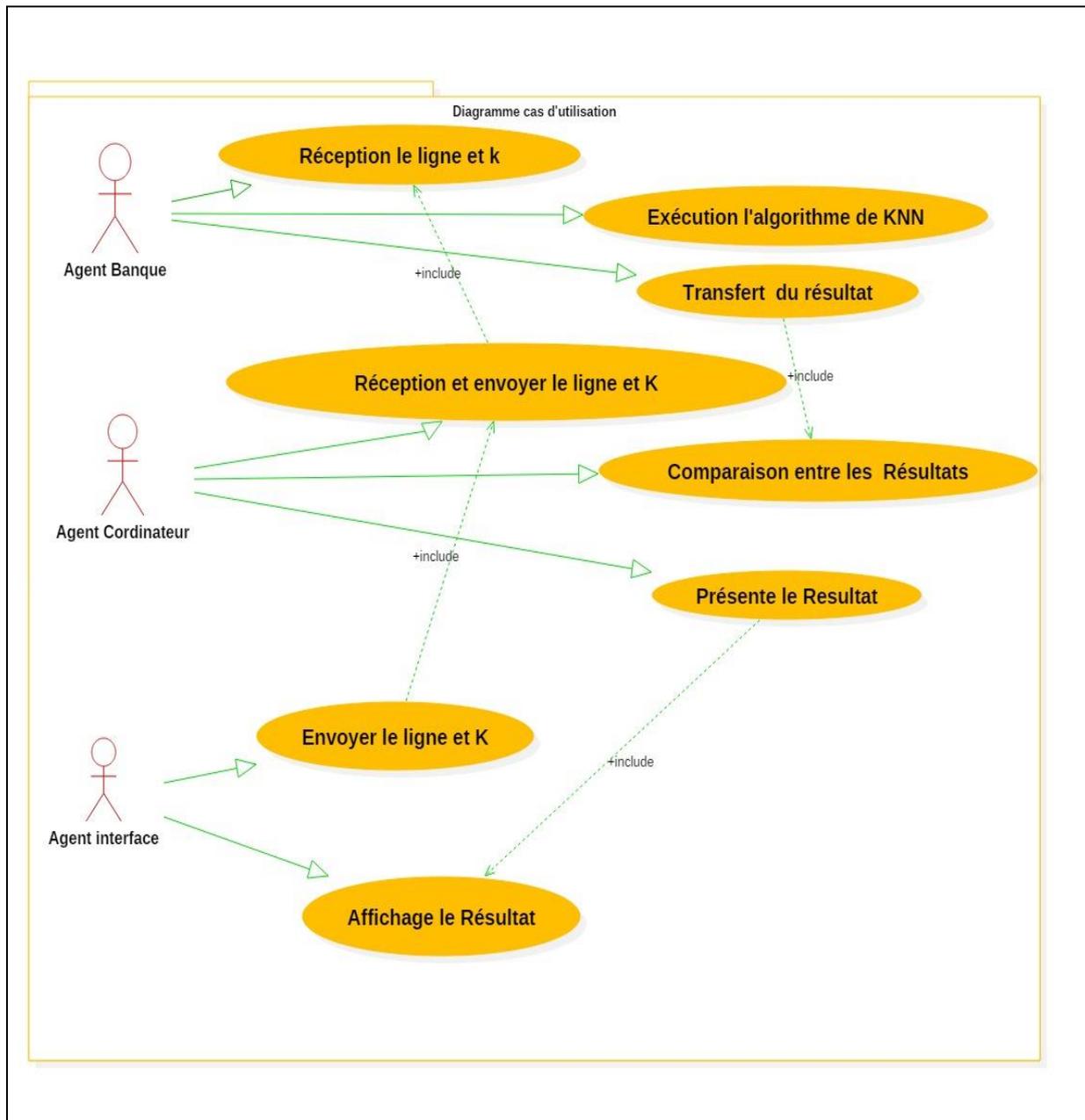


Figure3.2: Diagramme de cas d'utilisation AUML du modèle proposé.

3.4.2 Diagramme de séquence :

Le diagramme de séquences suivant décrit le processus de fonctionnement général de notre système.

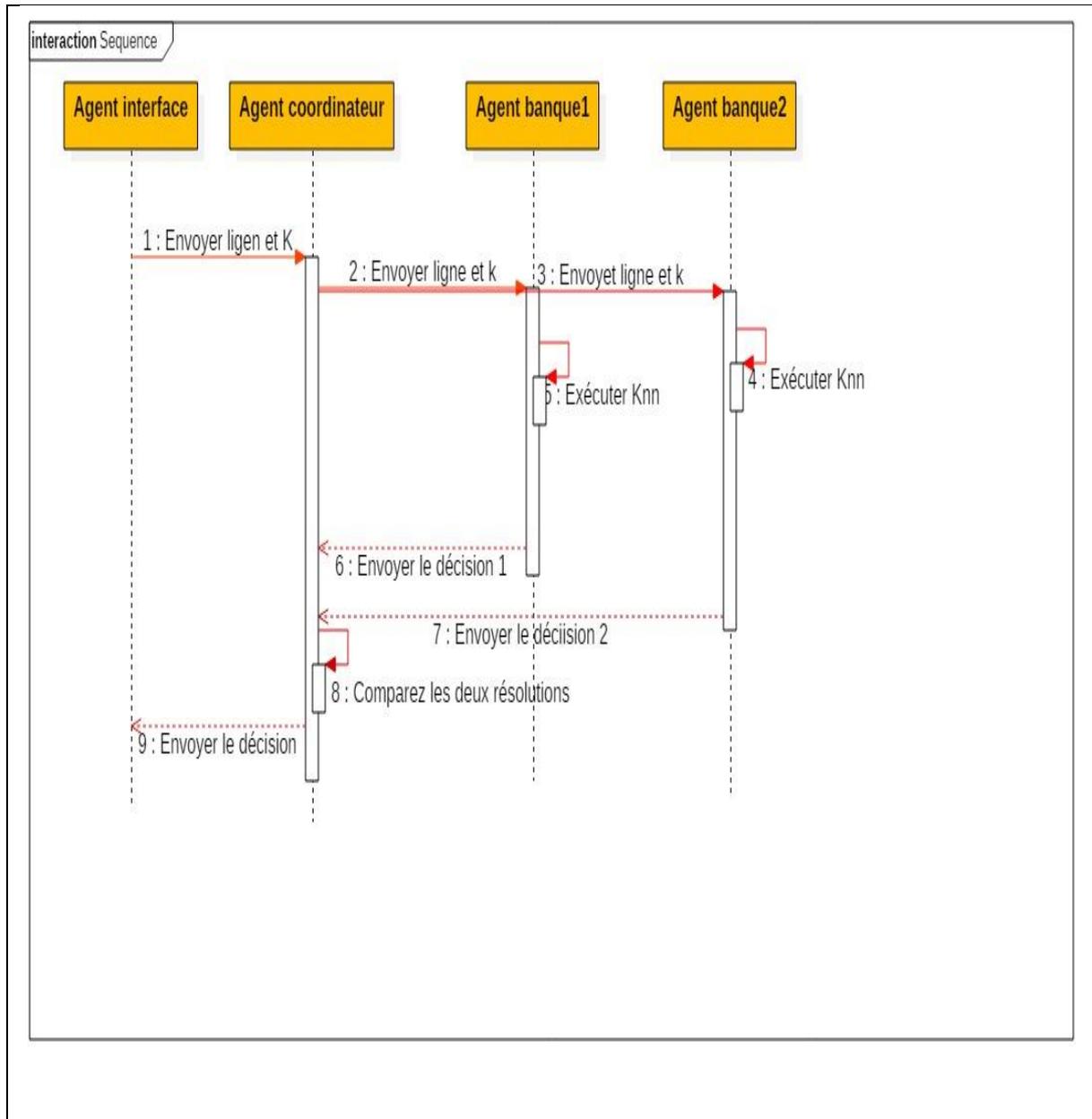


Figure3.3: Diagramme de séquences du modèle proposé.

3.4.3 Diagramme de classes :

La figure suivante présente le diagramme de toutes les classes de notre système et toutes les relations entre elles.

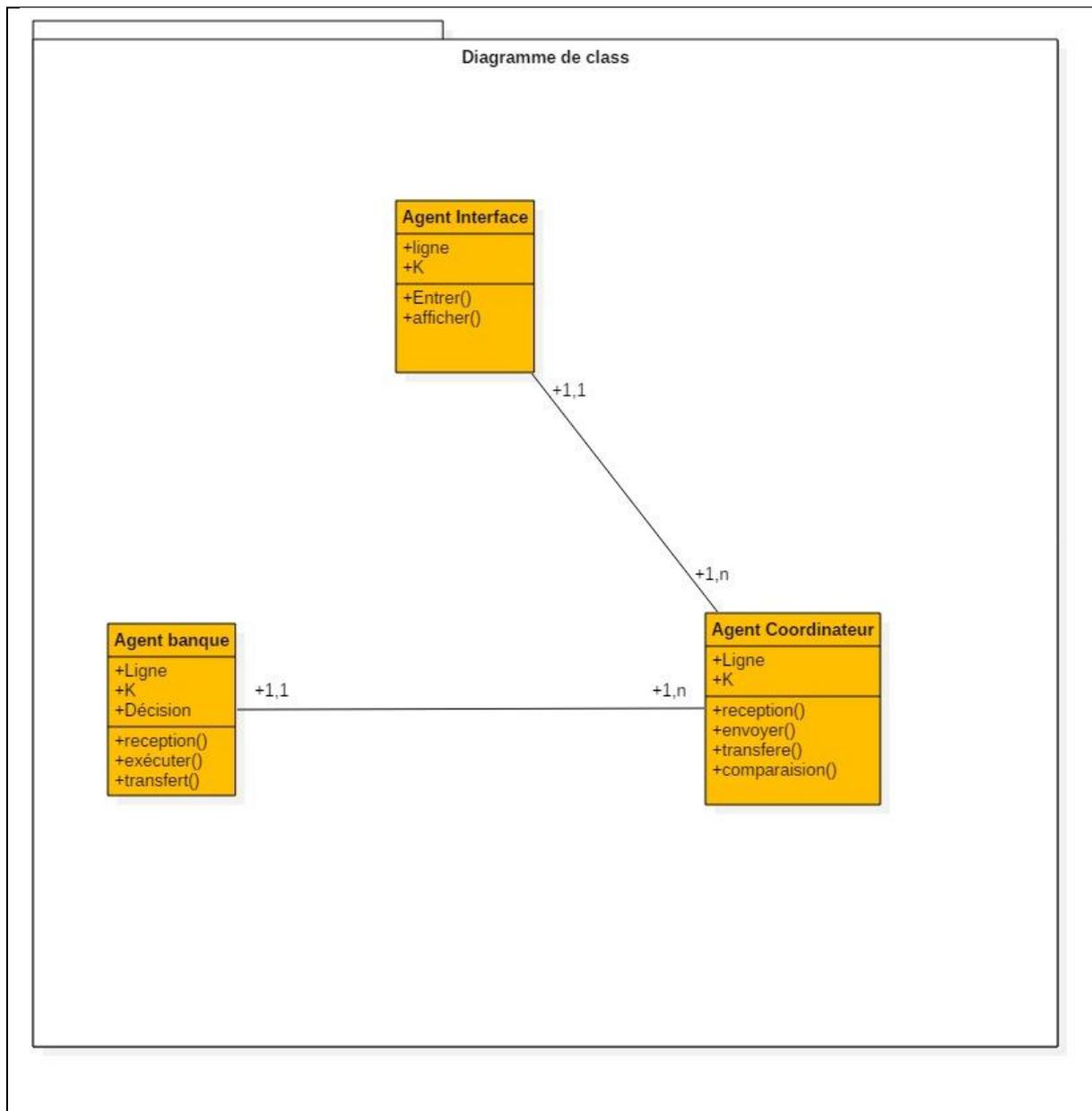


Figure 3.4 : Diagramme de classes du modèle proposé.

3.5 Conclusion :

Nous avons présenté dans ce chapitre notre architecture d'un système de data mining d'un grand ensemble de données en utilisant la méthode de K plus proches voisins et en se basant sur un système multi-agents.

Le chapitre suivant, présente l'implémentation de notre système et des résultats obtenus.

Chapitre 4:

Implémentation

4.1 Introduction :

Dans ce chapitre, nous allons présenter les différents outils et langages de programmation que nous avons utilisés pour la réalisation de notre système. Puis, nous donnerons quelques résultats de son fonctionnement.

Ensuite nous allons présenter quelques interfaces et fenêtres de notre système de système data mining basée sur SMA.

4.2 Langage de programmation :

Notre application est implémentée en utilisant le langage JAVA, sous l'environnement de développement Netbeans. Pour le système multi-agents nous avons utilisé la plateforme JADE.

4.2.1 JAVA :

Le langage Java est un langage généraliste de programmation synthétisant les principaux langages existants lors de sa création en 1995 par Sun Microsystems. Il permet une programmation orientée-objet (à l'instar de Small Talk et, dans une moindre mesure, C++), modulaire (langage ADA) et reprend une syntaxe très proche de celle du langage C. [29]

Le choix de ce langage a été motivé par les caractéristiques qu'il fournit :

- Un accès simplifié aux bases de données.
- Une programmation orientée objet et une bibliothèque immense d'objets : dès sa naissance, les programmeurs de Sun ont doté leur langage d'une des plus grandes bibliothèques d'objets prêts à l'emploi.
- Les classes d'extraction et d'indexation de la plateforme Lucene sont entièrement écrites en Java.

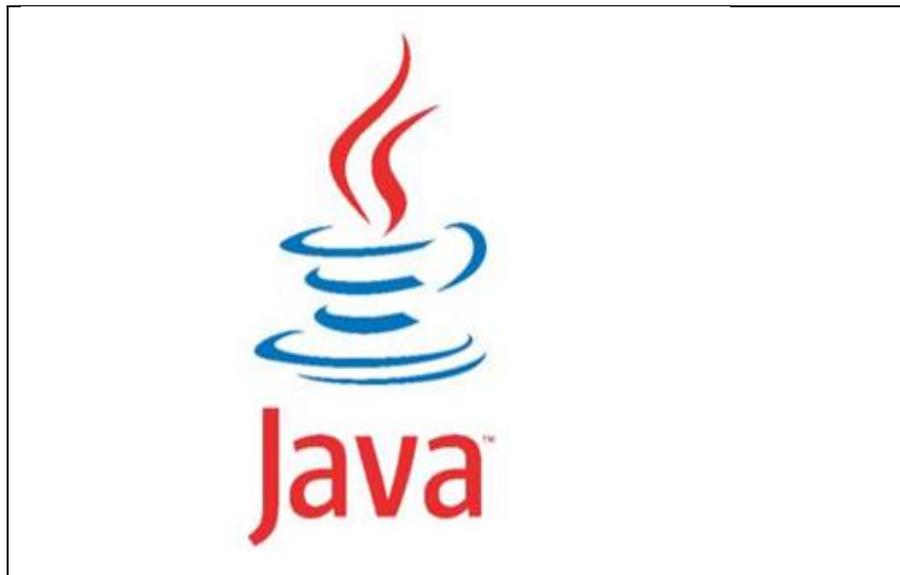


Figure 4.1: Le langage de programmation JAVA

4.3 L'environnement de développement :

L'environnement de développement de notre application est le suivant :

4.3.1 Netbeans :

Netbeans est un EDI (Environnement de Développement Intégré), un environnement libre et facile à utiliser. Il supporte plusieurs langages de programmation à savoir JAVA, C++, PHP, et bien d'autres. Il fournit plusieurs outils tel qu'un éditeur de texte doté d'un pré-compilateur avancé, un gestionnaire de projets. Ainsi que des outils de débogage et de test des programmes. C'est un outil qui facilite énormément la phase de développement et des tests.

La figure 4.2 présente une capture écran de Netbeans IDE 8.2

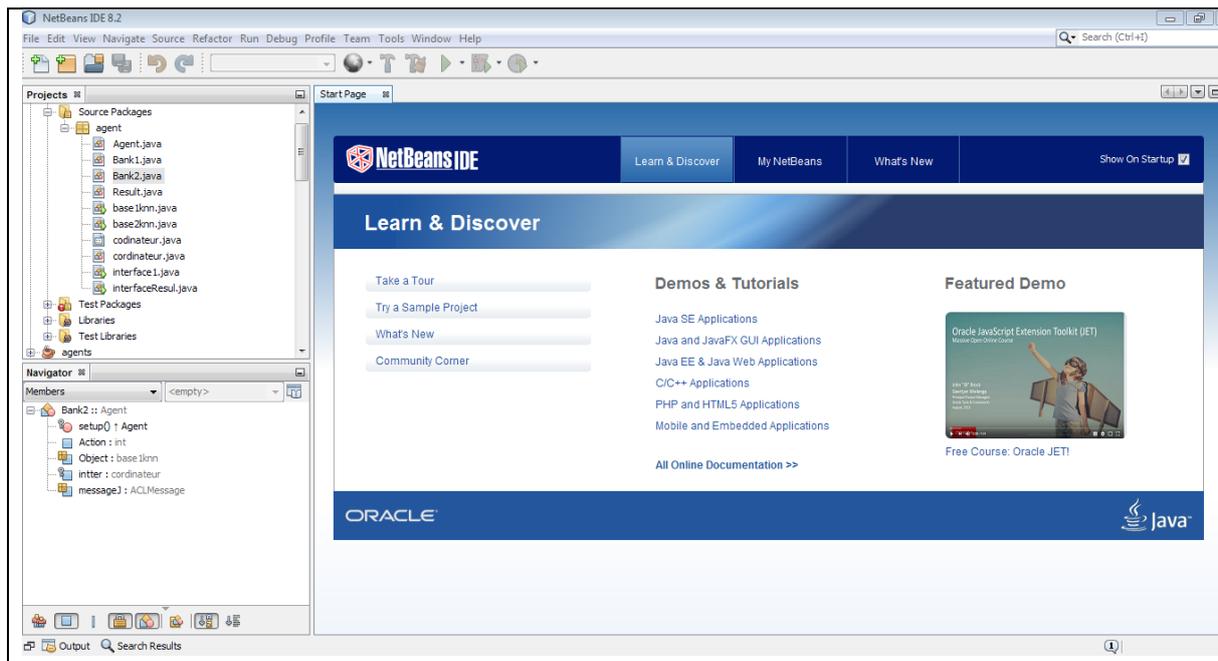


Figure 4.2: Interface de Netbeans IDE 8.2

4.3.2 La plateforme JADE (Java Agent Development Framework) :

JADE est une plate-forme de développement d'agents gratuite et Open Source développée par CSELT.

Le choix de cette plateforme est justifié par les points suivants:

- JADE simplifie l'implémentation d'un SMA à travers un Middleware répondant aux spécifications de la FIPA une librairie de classes que les utilisateurs peuvent utiliser et étendre ainsi un ensemble d'outils graphiques qui permettent le débogage et l'administration du SMA à concevoir ;
- JADE assure une communication transparente par l'échange de messages dans le langage normalisé FIPA-ACL ;
- JADE diminue l'effort de programmation car elle implémente deux agents (DF et AMS) dont les fonctionnalités sont utiles à notre application.

La figure 4.3 présente une capture écran de l'interface graphique de la Plateforme JADE

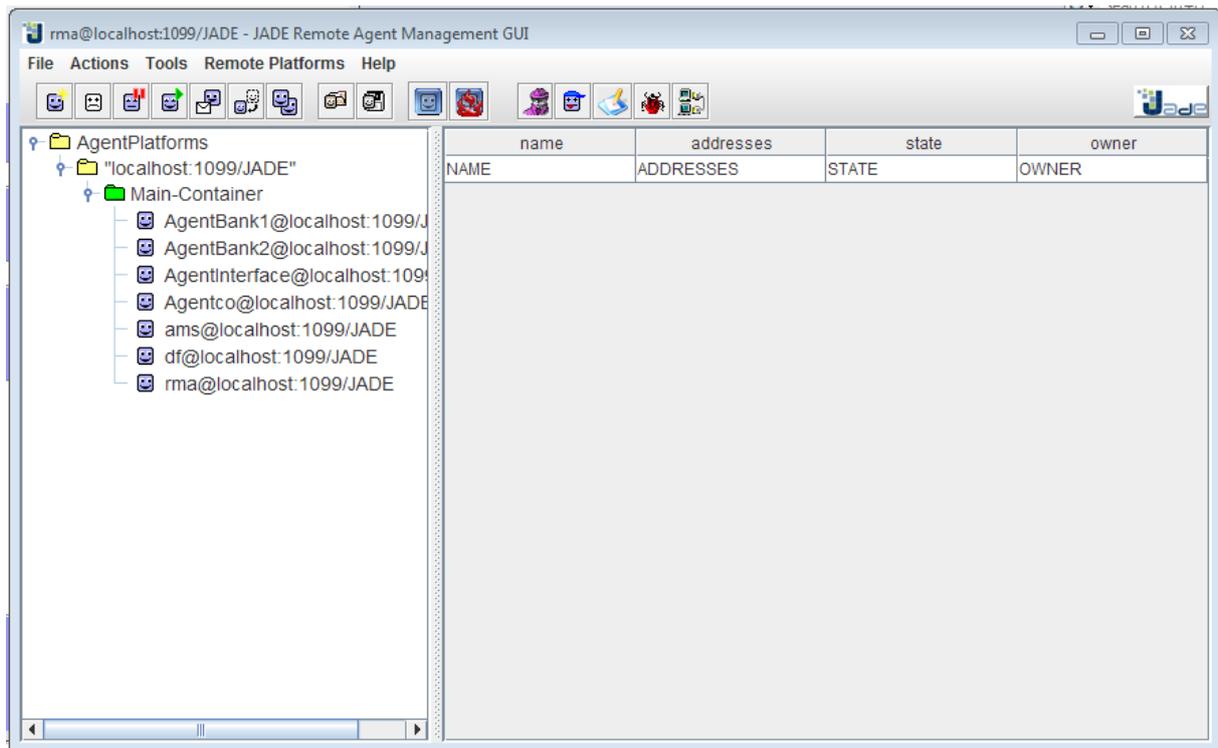


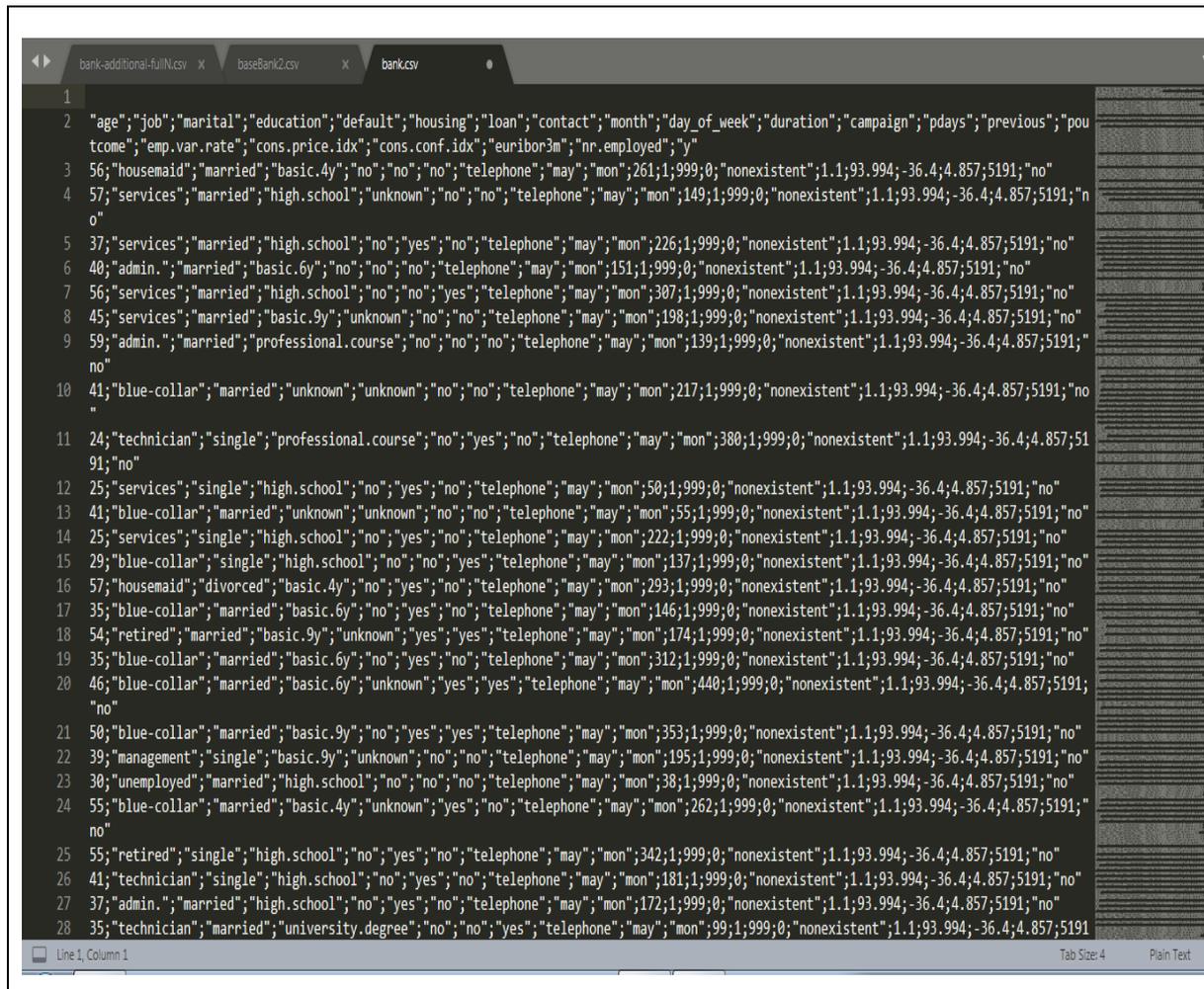
Figure 4.3 : interface graphique de la plateforme JADE

4.4 Principales fenêtres de l'application :

Nous allons présenter les différentes fonctionnalités de notre application.

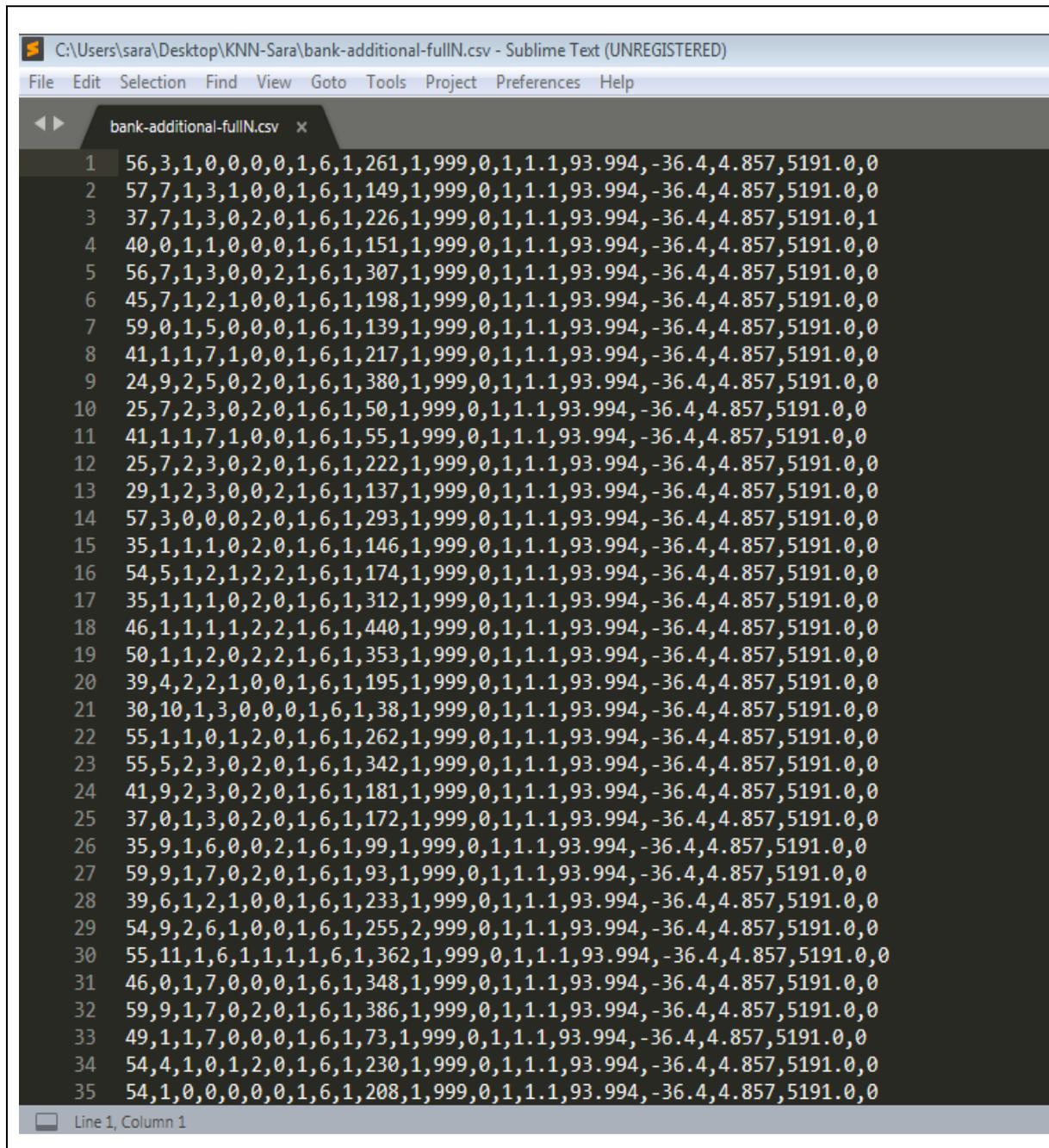
4.4.1 Présentation de la collection :

Nous avons téléchargé une base de données d'une banque [30], qui contient les informations de plusieurs personnes qui veulent faire un prêt d'une banque. Les champs de cette base de données qui étaient dans le format chaîne de caractères, nous les avons transformés vers des données numériques dans le but de mieux appliquer la méthode de K plus proches voisins.



```
1
2 "age";"job";"marital";"education";"default";"housing";"loan";"contact";"month";"day_of_week";"duration";"campaign";"pdays";"previous";"pou
tcome";"emp.var.rate";"cons.price.idx";"cons.conf.idx";"euribor3m";"nr.employed";"y"
3 56;"housemaid";"married";"basic.4y";"no";"no";"no";"telephone";"may";"mon";261;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
4 57;"services";"married";"high.school";"unknown";"no";"no";"telephone";"may";"mon";149;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"n
o"
5 37;"services";"married";"high.school";"no";"yes";"no";"telephone";"may";"mon";226;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
6 40;"admin."; "married";"basic.6y";"no";"no";"no";"telephone";"may";"mon";151;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
7 56;"services";"married";"high.school";"no";"no";"yes";"telephone";"may";"mon";307;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
8 45;"services";"married";"basic.9y";"unknown";"no";"no";"telephone";"may";"mon";198;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
9 59;"admin."; "married";"professional.course";"no";"no";"no";"telephone";"may";"mon";139;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"
no"
10 41;"blue-collar";"married";"unknown";"unknown";"no";"no";"telephone";"may";"mon";217;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
11 24;"technician";"single";"professional.course";"no";"yes";"no";"telephone";"may";"mon";380;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;51
91;"no"
12 25;"services";"single";"high.school";"no";"yes";"no";"telephone";"may";"mon";50;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
13 41;"blue-collar";"married";"unknown";"unknown";"no";"no";"telephone";"may";"mon";55;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
14 25;"services";"single";"high.school";"no";"yes";"no";"telephone";"may";"mon";222;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
15 29;"blue-collar";"single";"high.school";"no";"no";"yes";"telephone";"may";"mon";137;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
16 57;"housemaid";"divorced";"basic.4y";"no";"yes";"no";"telephone";"may";"mon";293;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
17 35;"blue-collar";"married";"basic.6y";"no";"yes";"no";"telephone";"may";"mon";146;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
18 54;"retired";"married";"basic.9y";"unknown";"yes";"yes";"telephone";"may";"mon";174;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
19 35;"blue-collar";"married";"basic.6y";"no";"yes";"no";"telephone";"may";"mon";312;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
20 46;"blue-collar";"married";"basic.6y";"unknown";"yes";"yes";"telephone";"may";"mon";440;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;
"no"
21 50;"blue-collar";"married";"basic.9y";"no";"yes";"yes";"telephone";"may";"mon";353;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
22 39;"management";"single";"basic.9y";"unknown";"no";"no";"telephone";"may";"mon";195;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
23 30;"unemployed";"married";"high.school";"no";"no";"no";"telephone";"may";"mon";38;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
24 55;"blue-collar";"married";"basic.4y";"unknown";"yes";"no";"telephone";"may";"mon";262;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"
no"
25 55;"retired";"single";"high.school";"no";"yes";"no";"telephone";"may";"mon";342;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
26 41;"technician";"single";"high.school";"no";"yes";"no";"telephone";"may";"mon";181;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
27 37;"admin."; "married";"high.school";"no";"yes";"no";"telephone";"may";"mon";172;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191;"no"
28 35;"technician";"married";"university.degree";"no";"no";"yes";"telephone";"may";"mon";99;1;999;0;"nonexistent";1.1;93.994;-36.4;4.857;5191
```

Figure 4.4 : la base de données sous forme de chaînes de caractères.



Line	Col 1	Col 2	Col 3	Col 4	Col 5	Col 6	Col 7	Col 8	Col 9	Col 10	Col 11	Col 12	Col 13	Col 14	Col 15	Col 16	Col 17	Col 18	Col 19	Col 20	Col 21	Col 22	Col 23	Col 24	Col 25	Col 26	Col 27	Col 28	Col 29	Col 30	Col 31	Col 32	Col 33	Col 34	Col 35
1	56,3,1,0,0,0,0,1,6,1,261,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
2	57,7,1,3,1,0,0,0,1,6,1,149,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
3	37,7,1,3,0,2,0,0,1,6,1,226,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,1																																		
4	40,0,1,1,0,0,0,0,1,6,1,151,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
5	56,7,1,3,0,0,2,1,6,1,307,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
6	45,7,1,2,1,0,0,0,1,6,1,198,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
7	59,0,1,5,0,0,0,0,1,6,1,139,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
8	41,1,1,7,1,0,0,0,1,6,1,217,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
9	24,9,2,5,0,2,0,0,1,6,1,380,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
10	25,7,2,3,0,2,0,0,1,6,1,50,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
11	41,1,1,7,1,0,0,0,1,6,1,55,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
12	25,7,2,3,0,2,0,0,1,6,1,222,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
13	29,1,2,3,0,0,2,1,6,1,137,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
14	57,3,0,0,0,2,0,0,1,6,1,293,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
15	35,1,1,1,0,2,0,0,1,6,1,146,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
16	54,5,1,2,1,2,2,1,6,1,174,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
17	35,1,1,1,0,2,0,0,1,6,1,312,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
18	46,1,1,1,1,2,2,1,6,1,440,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
19	50,1,1,2,0,2,2,1,6,1,353,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
20	39,4,2,2,1,0,0,0,1,6,1,195,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
21	30,10,1,3,0,0,0,0,1,6,1,38,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
22	55,1,1,0,1,2,0,0,1,6,1,262,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
23	55,5,2,3,0,2,0,0,1,6,1,342,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
24	41,9,2,3,0,2,0,0,1,6,1,181,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
25	37,0,1,3,0,2,0,0,1,6,1,172,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
26	35,9,1,6,0,0,2,1,6,1,99,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
27	59,9,1,7,0,2,0,0,1,6,1,93,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
28	39,6,1,2,1,0,0,0,1,6,1,233,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
29	54,9,2,6,1,0,0,0,1,6,1,255,2,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
30	55,11,1,6,1,1,1,1,6,1,362,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
31	46,0,1,7,0,0,0,0,1,6,1,348,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
32	59,9,1,7,0,2,0,0,1,6,1,386,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
33	49,1,1,7,0,0,0,0,1,6,1,73,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
34	54,4,1,0,1,2,0,0,1,6,1,230,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		
35	54,1,0,0,0,0,0,0,1,6,1,208,1,999,0,1,1.1,93.994,-36.4,4.857,5191.0,0																																		

Figure 4.5: la base de données sous forme Numérique

4.4.2 Principales fenêtres de l'agent interface :

Cette interface (Figure 4.6) représente interface agent, On va remplir les champs qui représentent la ligne des informations du personne dans la base de données et on donne dans le 2eme champ le K qu' il faut être un nombre impair. Et cette interface affiche la résultat

finale si la réponse est 0 ça veut dire refus d'offre prêt si la réponse est 1 veut dire accepter d'offrir le prêt.

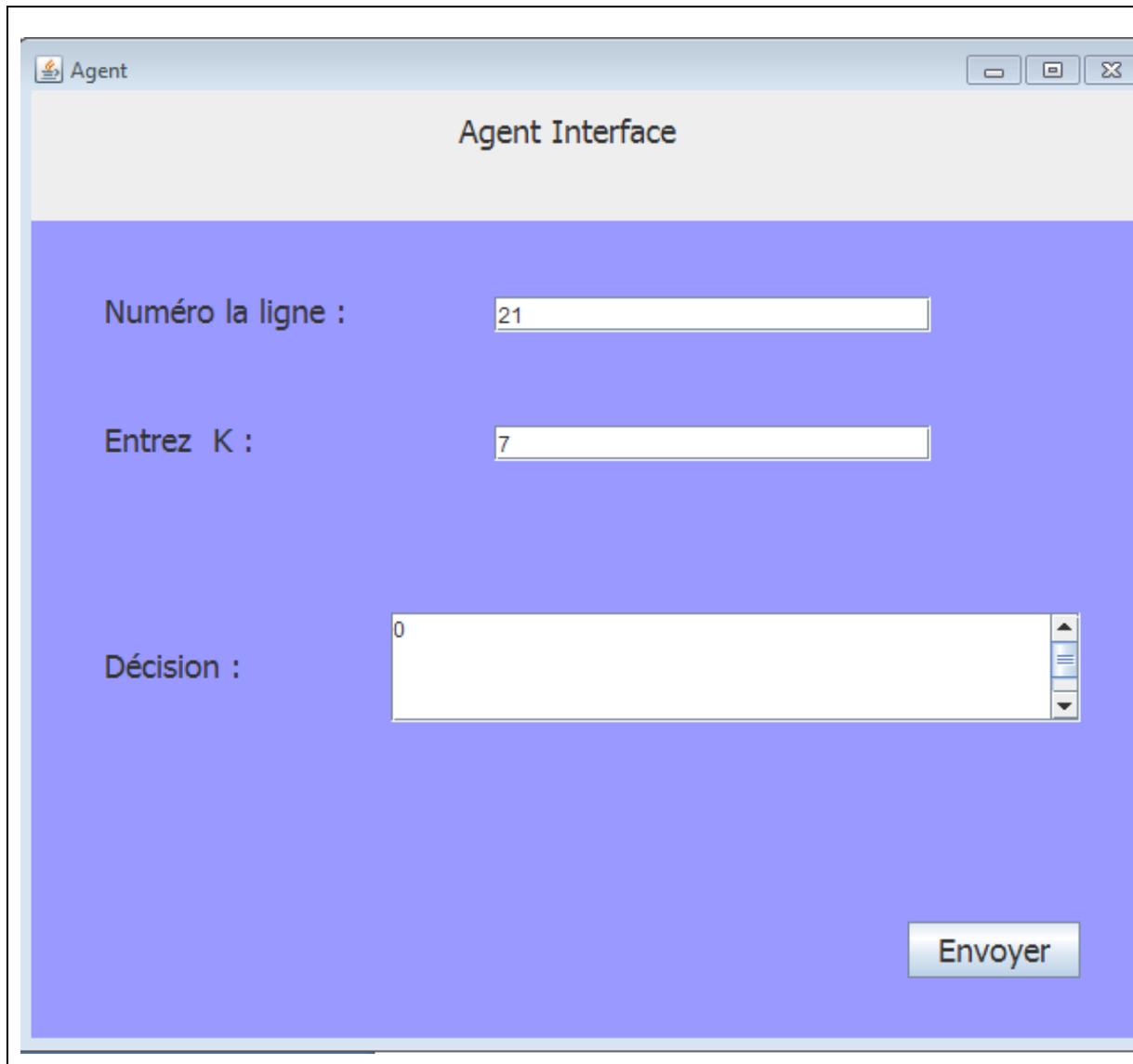


Figure 4.6: Fenêtre principale de l'agent interface.

4.4.3 Échange de message entre les agents du système :

La figure suivante montre la communication entre les agents. Cette figure est extraite à partir de l'outil de la plateforme JADE.

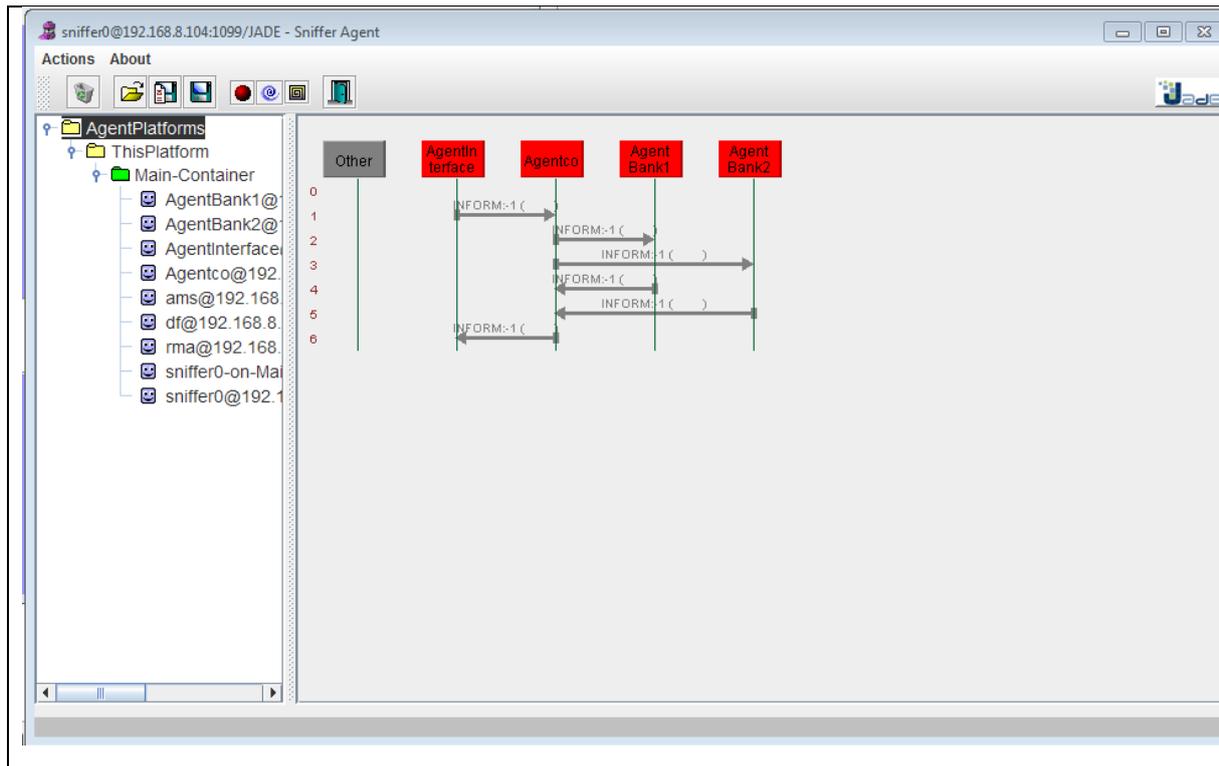


Figure 4.7 : échange de messages entre les agents du système

4.5 Conclusion :

Dans ce chapitre, nous avons décrit les différents outils et technologies utilisées pour le développement de notre application. Nous avons présenté quelques définitions des langages utilisés pour réaliser notre application.

A la fin nous avons présenté quelques interfaces de navigation que nous considérons comme étant plus importantes.

Conclusion générale

Dans ce travail, nous nous sommes focalisés sur l'étude de trois domaines différents qui sont le Data Mining, le paradigme agent et l'utilisation des agents dans le Data Mining dans l'objectif final de proposer une approche Data Mining basée agents et de concevoir le système multi-agents permettant de la mettre en œuvre.

Le data mining basé agents est une nouvelle discipline qui combine deux grandes technologies qui sont le Data Mining, et les systèmes multi-agents. De cette discipline dérivent deux nouveaux axes de recherche. L'un s'intéresse à l'amélioration du processus Data Mining en utilisant les avantages paradigme agent, et l'autre tire parti des techniques Data Mining pour le développement et la réalisation de l'intelligence dans la technologie agent.

Dans notre travail, nous nous sommes positionné dans la première vision, dont l'objectif est le développement et à la réalisation d'une architecture Data Mining basée agents, qui permet de prendre en charge le processus Data Mining durant chacune de ses étapes. Notre premier chapitre a été consacré au concept Data Mining, sa définition, son processus, leurs tâches et leurs techniques. L'étude de la discipline Data Mining nous a permis de découvrir un domaine très récent et très vague actuellement qui est l'extraction automatique des connaissances à partir des données.

Ce travail bien sûr est loin d'être achevé, et nécessite encore d'efforts et de persévérance. Comme perspectives, on peut se concentrer sur les points suivants :

- Utiliser d'autres techniques hors que la classification;
- Utiliser d'autres algorithmes et pas seulement K plus proches voisins

Bibliographies

- [1] : D.HAND, H.MANNILA et P.SMYTH, « **Principles of Data Mining**», MIT Press, Cambridge, MA, 2001
- [2] : P.CABENA, P.HADJINIAN, R.STADLER, J. VERHEES et A. ZANASI, «**Discovering Data Mining: From Concept to Implementation** », Prentice Hall, Upper Saddle River, NJ, 1998.
- [3] : E-G.TALBI, « **Fouille de données (Data Mining) : Un tour d’horizon** »,Laboratoire d’Informatique Fondamentale de Lille.
- [4]: O. R. ZAÏANE, Principles of Knowledge Discovery in Databases, CMPUT690, University of Alberta, 1999.
- [5] : Ph. PREUX, Fouille de données : Notes de cours, Université de Lille 3, 9 octobre 2008.
- [6] : G. DONG, J. PEI, Sequence Data Mining, Springer Edition, 2007
- [7] : S. PRABHU, N. VENKATESAN, Data Mining and Warehousing, New Age International (P) Ltd., Publishers, New Delhi, 2007.
- [8] : Bernard ESPINASSE, l'Université d'Aix-Marseille, ''Introduction aux méthodes de Fouille de données, 2009
- [9] : CHAMI Djazia, « **Une plate forme orientée agent pour le data mining**» Université HADJ LAKHDAR–BATNA Faculté des sciences de l’ingénieur Département d’informatique 2009
- [10] : M. J. BERRY, G. S. LINOFF, Data Mining Techniques For Marketing, Sales, and Customer Relationship, Management, Second Edition, 2004
- [11] : M. J. BERRY, G. S. LINOFF, Mastering Data Mining: The Art and Science of Customer Relationship Management, 2000
- [12] : D.T. LAROSE, Discovering Knowledge In Data: An Introduction to DataMining, Central Connecticut State University, 2005
- [13] : TOUAHRIA Lakhdar « **Segmentation D’un Environnement Data Mining Par Les Agents**» Université Mohamed Khider – BISKRA Département d’informatique 2013
- [14] : Stéphane Tufféry, « **Data Mining et statistique décisionnelle l’intelligence des données** », Editions Technip 2007.
- [15] : LIAUDET«**Cours de Data Mining:3 Modélisation- présentation générale-**» EPF – 4/ 5ème année - Option Ingénierie d’Affaires et de Projets - Finance

Références

- [16] : Mémoire Technique « **Le Data Mining** » -www.softcomputing.com
- [17] : R.GILLERON, M. TOMMASI, Découverte de connaissances à partir de données, 2000
- [18] : K. TEKNOMO, What is K Nearest Neighbors Algorithm?,
<http://people.revoledu.com/kardi/tutorial/KNN/What-is-K-NearestNeighborAlgorithm.html>,
2006.
- [19] : https://fr.wikipedia.org/wiki/Agent_logiciel#D.C3.A9finition 11 /10/2017 ||20:30
- [20] : Chaib-Draa, B., Jarras, I., & Moulin, B. (2001). Systèmes multi-agents: principes généraux et applications. Edition Hermès.
- [21] : BAHMED ABDELGHANI « Un système multi agent pour l'appariement et la recherche formes » mémoire de master ,Université Mohamed Khider – BISKRA ,2017
- [22] : <http://turing.cs.pub.ro/auf2/html/chapters/chapter1/chapitre-1.pdf> 11/10/2017 ||
22:23
- [23] : <https://www.irit.fr/~Chihab.Hanachi/Cours/SMA/CoursAgentsI.pdf> 12/10 /2017
||21 :09
- [24] : https://fr.wikipedia.org/wiki/Syst%C3%A8me_multi-agents 25/11/2017 ||22 :10
- [25] : Adina Magda Florea « **Agents et Systèmes Multi-agents** » Université 'Politehnica' de Bucarest
- [26] : Iffenger, C. (1992). Un système multi-agent pour le support des activités de conception de produit. Thèse de l'Université Paris VI
- [27] : Jean-Paul Sansonnet **Introduction aux Systèmes Multi Agents« Etude des interactions médiatisées entre gens et agents cognitifs »**
- [28] : Mr SALAH Djilali «**La contribution des systèmes multi agents dans la technologie du data mining distribue** » Université des Sciences et de la Technologie d'Oran Mohamed Boudiaf 2016
- [29] : <http://dspace.univ-tlemcen.dz/bitstream/112/8301/5/CHAPITRE3.pdf> 15/05 /2018
||23 :36
- [30] : <http://archive.ics.uci.edu/ml/index.php> 20/11 /2017 ||15 :18

