

République Algérienne Démocratique et Populaire.  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique.

Université A. Mira de Béjaia.  
Faculté des Sciences et des Sciences de l'Ingénieur

Département de Recherche Opérationnelle

## *Mémoire de fin d'études*

En  
Recherche Opérationnelle

Thème

*Systemes d'Aide à la Décision basés sur le Data Mining*

Présenté par :

Mr RAGAB HASSEN Hani.

Devant le jury composé de :

Président	P <sup>r</sup> Radjef M.S.
Rapporteur	M <sup>me</sup> El-Maouheb A.
Examinatrice	M <sup>me</sup> Aoudia Z.
Examineur	M <sup>r</sup> Tari A.K.
Invitée	M <sup>me</sup> Imloul S.
Invité	M <sup>r</sup> Touazi D.

Promotion 2002 – 2003

# Table des matières

<b>Introduction Générale</b>	<b>10</b>
<b>1 Systèmes Informatiques d'Aide à La Décision</b>	<b>17</b>
<b>Introduction</b>	<b>18</b>
<b>1 Historique des systèmes d'aide à la décision</b>	<b>19</b>
<b>2 Systèmes Informatiques dans l'Industrie</b>	<b>23</b>
2.1 Systèmes informatiques opérationnels . . . . .	23
2.2 Systèmes Informatiques Décisionnels . . . . .	24
2.3 Cohabitation des Systèmes Opérationnels et Décisionnels . . . . .	25
2.4 Le Data Warehouse . . . . .	26
2.4.1 Architecture d'un Data Warehouse . . . . .	26
2.4.2 Niveaux de schémas de données . . . . .	30
2.4.3 Magasin de données . . . . .	31
2.4.4 Modélisation multidimensionnelle des données . . . . .	31
<b>3 Outils d'Aide à la Décision</b>	<b>35</b>
3.1 Les Requêteurs et les Outils de Reporting . . . . .	35
3.2 Les Outils OLAP . . . . .	36
3.3 Les Outils du Data Mining . . . . .	37
<b>II Le Data Maining</b>	<b>42</b>
<b>Introduction</b>	<b>43</b>
<b>4 Modèles Descriptifs et Modèles Prédictifs</b>	<b>45</b>
4.1 Modèles descriptifs . . . . .	45

4.1.1	Groupage ou Agrégation (Clustering) . . . . .	45
4.1.2	Règles associatives (association rules, association discovery) . . . . .	46
4.1.3	Découverte de Séquences (Sequencing, Sequence Discovery) . . . . .	47
4.2	Modèles prédictifs . . . . .	48
4.2.1	Classification . . . . .	48
4.2.2	Régression . . . . .	48
4.2.3	Séries chronologique (Time series) . . . . .	49
<b>5</b>	<b>Modèles de Data Mining</b> . . . . .	<b>53</b>
5.1	Les modèles du Data Mining . . . . .	53
5.1.1	Réseaux de neurones . . . . .	53
5.1.2	Les arbres de décision . . . . .	57
5.1.3	Les réseaux bayésiens et bayésiens naïfs . . . . .	59
5.1.4	Réseaux Bayésiens Naïfs . . . . .	60
5.1.5	K-Plus Proches Voisins (K-Nearest Neighbors) . . . . .	61
5.1.6	Algorithme Apriori . . . . .	61
5.1.7	Les algorithmes génétiques . . . . .	62
5.2	Etude comparative . . . . .	63
	Conclusion . . . . .	66
<b>6</b>	<b>Processus du Data Mining</b> . . . . .	<b>67</b>
6.1	Définition du Problème . . . . .	68
6.2	Construction de la base de données du Data Mining . . . . .	69
6.2.1	Collecte des données . . . . .	70
6.2.2	Description des données . . . . .	71
6.2.3	Sélection . . . . .	71
6.2.4	Evaluation de la qualité des données et nettoyage des données . . . . .	71
6.2.5	Intégration et consolidation . . . . .	72
6.2.6	Construction de la méta Data . . . . .	72
6.2.7	Chargement de la base de données du Data Mining . . . . .	72
6.2.8	Maintenance de la base de données du Data Mining . . . . .	72
6.3	Exploration des données . . . . .	72
6.4	Préparation des données pour la modélisation . . . . .	73
6.4.1	Sélection des variables . . . . .	73
6.4.2	Sélection des lignes . . . . .	73
6.4.3	Création de nouvelles variables . . . . .	74
6.4.4	Transformation des variables . . . . .	74

6.5	Construction du modèle . . . . .	74
6.6	Evaluation et Interprétation du modèle . . . . .	76
6.7	Déploiement du modèle et de ses résultats . . . . .	78
<b>III Conception et Réalisation de quelques outils de Data Mining</b>		<b>79</b>
<b>Introduction</b>		<b>80</b>
<b>7</b>	<b>Les réseaux de neurones.</b>	<b>82</b>
7.1	Les fondements . . . . .	82
7.1.1	Un cadre de traitement distribué . . . . .	83
7.1.2	Topologies des réseaux de neurones. . . . .	85
7.1.3	Apprentissage d'un réseau de neurone. . . . .	86
7.2	Notation et terminologie. . . . .	87
7.3	Le Perceptron. . . . .	89
7.3.1	Réseaux avec fonctions d'activation à seuil. . . . .	89
7.3.2	Règle d'apprentissage du Perceptron. . . . .	91
7.3.3	La règle Delta. . . . .	92
7.4	Le problème du OU Exclusive : XOR. . . . .	93
7.4.1	Pouvoir représentatif des perceptrons multicouches. . . . .	96
7.5	La rétropropagation. . . . .	96
7.5.1	Réseaux feed-forward multicouches. . . . .	97
7.5.2	La règle delta généralisée. . . . .	97
7.6	Travailler avec la rétropropagation. . . . .	101
Conclusion . . . . .		102
<b>8</b>	<b>Combiner les Réseaux de Neurones et les Algorithmes Génétiques.</b>	<b>104</b>
8.1	Les Algorithmes Génétiques. . . . .	104
8.1.1	Crossover . . . . .	106
8.1.2	La sélection. . . . .	107
8.1.3	Algorithmes Génétiques Parallèles. . . . .	108
8.1.4	Réseaux de Neurones et Algorithmes Génétiques. . . . .	109
8.2	Les différentes approches de combinaison. . . . .	110
8.2.1	Les axes principaux de recherche. . . . .	110
8.2.2	Codage Direct. . . . .	112
8.2.3	Codage indirect. . . . .	116

Conclusion . . . . .	117
<b>9 Implémentation d'un logiciel de Data Mining : Smart Miner.</b>	<b>118</b>
9.1 Choix des langages et du SGBD. . . . .	118
9.2 Types de modèles disponibles. . . . .	119
9.3 Algorithmes utilisés. . . . .	119
9.3.1 La rétropropagation. . . . .	120
9.3.2 NNUGA. . . . .	121
9.3.3 L'algorithme des centres mobiles. . . . .	124
9.4 Les interfaces graphiques. . . . .	125
9.5 Problèmes pratiques. . . . .	134
9.5.1 La fonction sigmoïde dans NNUGA. . . . .	134
9.5.2 Choix des centres initiaux dans l'algorithme des centres mobiles.	135
Conclusion . . . . .	136
<b>Conclusion Générale et Perspectives</b>	<b>137</b>
<b>Bibliographie</b>	<b>142</b>