



**UNIVERSITE MOHAMED BOUDIAF - M'SILA**  
**FACULTE DES MATHÉMATIQUES ET**  
**DE L'INFORMATIQUE**



**DEPARTEMENT D'INFORMATIQUE**

**MEMOIRE de fin d'étude**

**Présenté pour l'obtention du diplôme de MASTER**

**Domaine : Mathématiques et Informatique**

**Filière : Informatique**

**Spécialité : Technologie de l'Information et de Communication**

**Par: BENYOUNES Sami**

**SUJET**

**Etude comparative des Algorithmes de Génération des  
Règles d'Association**

**Soutenu publiquement le : 01/06/2016 devant le jury composé de :**

**Mme M.BOUDIA**  
**Mr GUESMIA Salah**  
**Mr M.BOUNIF**

**Université de M'sila**  
**Université de M'sila**  
**Université de M'sila**

**Président**  
**Rapporteur**  
**Examineur**

**Promotion : 2015/2016**

## TABLE DES MATIERES

<b>Introduction</b>		Page1
<b>Chapitre 1 : Fouille de donnée (DATA MINING)</b>		
1.1	Introduction	Page3
1.1.1	Qu'est ce que la fouille de données	Page3
1.2	Les tâches du Data-Mining :	Page5
1.3	Les techniques de Data-Mining	Page5
1.4	Qu'est ce qu'une donnée	Page6
1.4.1	Notations	Page6
1.4.2	Les différentes natures d'attributs	Page7
1.4.3	Les différentes natures de valeur d'attribut	Page7
1.4.4	Le bruit	Page8
1.4.5	Différentes tâches d'extraction d'information	Page8
1.4.6	Les étapes d'un processus d'Extraction de Connaissances à partir des données (ECD)	Page10
1.4.6.1	Nettoyage et intégration des données	Page10
1.4.6.2	Prétraitement des données	Page12
1.4.6.3	Fouille de données (Data-Mining)	Page12
1.4.6.4	Evaluation et présentation	Page12
1.5	Conclusion	Page14
<b>Chapitre 2 : Règles d'Association</b>		
2.1	Introduction	Page15
2.2	Cadre informel	Page15
2.3	Cadre formel	Page16
2.4	La découverte des règles d'association	Page18
2.4.1	Extraction des itemsets fréquents	Page18
2.5	Génération des règles d'association	Page23
2.6	Les améliorations de l'algorithme Apriori	Page26
2.7	Réduction de l'ensemble de règles d'association	Page27
2.7.1	Approche orientée données	Page27
2.7.2	Approche orientée utilisateur	Page31
2.8	Domaine d'applications	Page33
2.9	Conclusion	Page33
<b>Chapitre 3 : Les algorithmes de génération des Règles d'Association</b>		
3.1	Introduction	Page34
3.2	Génération des Itemset fréquents	Page35
3.2.1	Algorithme FP-GROWTH	Page35
3.2.2	Algorithme ECLAT	Page43
3.2.3	Algorithme FIN	Page48
3.2.4	Algorithme PrePost+	Page54
3.3	Génération des Règles d'association	Page59
3.3.1	Génération des règles d'association par deux étapes	Page60
3.3.1.1	L'algorithme Apriori pour génération des règles d'association	Page60
3.3.2	Génération des règles d'association par	Page63

		approche d'expansion de règles	
	3.3.2.1	Génération des Règles d'Association de Top-K	Page63
	3.3.2.2	Génération des Règles d'Association de Top-K Non-Redondantes	Page68
3.4		Conclusion	Page72
<b>Chapitre 4 : Modélisation et Implémentation</b>			
4.1		Introduction	Page73
4.2		Spécification des besoins	Page73
	4.2.1	Spécification des besoins fonctionnels	Page73
	4.2.2	Spécification des besoins Non fonctionnels	Page73
4.3		Méthodologie et Approche Adoptée	Page74
	4.3.1	Présentation d'UML	Page74
	4.3.2	Les avantages d'UML	Page75
	4.3.3	Les Diagrammes de Cas D'utilisation	Page75
	4.3.3.1	Identification des acteurs	Page75
	4.3.3.2	Diagramme de cas d'utilisation globale	Page76
	4.3.3.2.a	Description du cas d'utilisation « Gérer les transactions »	Page76
	4.3.3.2.b	Description du cas d'utilisation « Générer des fichiers d'entrée »	Page77
	4.3.3.2.c	Description du cas d'utilisation « Choisir et paramétrer l'algorithme à exécuter »	Page77
	4.3.3.2.d	Description du cas d'utilisation « Analyser les résultats » :	Page78
	4.3.3.2.e	Description du cas d'utilisation « Comparer les résultats » :	Page78
	4.3.3.2.f	Description du cas d'utilisation « S'authentifier»	Page79
4.4		La Conception Générale	Page80
	4.4.1	Le cycle de développement en v	Page80
4.5		La Conception Détaillée	Page81
	4.5.1	Le diagramme de déploiement	Page81
	4.5.2	les diagrammes de séquence	Page81
	4.5.2.1	le diagramme de séquence « s'authentifier »	Page82
	4.5.2.2	le diagramme de séquence « Gérer les transactions »	Page82
	4.5.2.3	le diagramme de séquence « Générer des fichiers d'entrée »	Page83
	4.5.2.4	le diagramme de séquence « Choisir et paramétrer l'algorithme à exécuter »	Page83
	4.5.2.5	le diagramme de séquence « Analyser les résultats »	Page84
	4.5.2.6	le diagramme de séquence « Comparer les résultats »	Page84
	4.5.3	Les diagrammes d'activité	Page84
	4.5.4	Diagramme de Classe	Page86
4.6		Implémentation	Page86
	4.6.1	Choix technique	Page86

## Table des matières

---

	4.6.1.1	Choix du langage de programmation	Page86
	4.6.1.2	Choix de l'architecture de l'application	Page87
	4.6.2	Environnement Logistique	Page88
	4.6.2.1	Environnement de développement	Page88
	4.6.2.2	système de gestion de base de données	Page90
	4.6.2.3	Les données et structure de données	Page90
4.7		Travail réalisée	Page95
	4.7.1	Description de l'application	Page95
	4.7.2	Description des tables de la base de données	Page101
4.8		Discussion des Résultats	Page102
	4.8.1	Algorithme TopKRules pour confiacne = 0,8	Page102
	4.8.2	Algorithme TopKRules pour confiacne = 0,5	Page102
	4.8.3	Algorithme TopKRules Non redunadante pour confiance = 0.5 et delta=200	Page103
	4.8.4	Algorithme TopKRules Non redunadante pour confiance = 0.5 et delta=3600	Page104
	4.8.5	Fichier de sortie	Page105
4.9		Comparaison entre les deux algorithmes (TopKRules et TNR)	Page106
4.10		Conclusion	Page107
		<b>Conclusion Générale</b>	Page108

---

## INTRODUCTION GENERALE

Dans le domaine du Data-Mining la recherche des règles d'association est une méthode populaire étudiée d'une manière approfondie dont le but est de découvrir des relations ayant un intérêt pour le statisticien entre deux ou plusieurs variables stockées dans de très importantes bases de données[18].

Un exemple de règle d'association extraite d'une base de données de ventes de supermarché est : « céréales, sucre → lait (support 7%, confiance 50%) ». Cette règle indique que les clients qui achètent des céréales et du sucre ont également tendance à acheter du lait. La mesure de support définit la portée de la règle, c'est à dire la proportion de clients qui ont acheté les trois articles, et la mesure de confiance définit la précision de la règle, c'est à dire la proportion de clients qui ont acheté du lait parmi ceux qui ont acheté des céréales et du sucre. L'extraction de règles d'association consiste à extraire les règles dont le support et la confiance sont au moins égaux à des seuils minimaux de support et de confiance définis par l'utilisateur. Les règles d'association ont été utilisées avec succès dans de nombreux domaines, parmi lesquels l'aide à la planification commerciale, l'aide au diagnostic et en recherche médicale, l'amélioration des processus de télécommunications, l'organisation et l'accès aux sites Internet, et l'analyse d'images, de données spatiales, de données géographiques et de données statistiques [19].

Les règles d'association peuvent être générées par deux approches connus comme suit :

- ❖ Approche classique qui consiste à extraire les règles d'associations par deux étapes, la première précède à extraire les items fréquents, après, ces items seront utilisés pour générer les règles d'association en deuxième étape.
- ❖ Nouvelle approche qui consiste à extraire les règles d'associations en une seule étape, cette approche est nommée expansion des règles qui élimine la première étape pour avoir une bonne performance.

La problématique de notre travail s'est d'évaluer et de comparer la performance des algorithmes de génération des règles d'association en fonction de plusieurs paramètres définis par l'utilisateur.

Notre Travail est composé de plusieurs chapitres, on va les citer ci-dessous :

- ❖ Nous consacrerons le premier chapitre à présenter des principes fondamentaux de Data-Mining, ainsi le processus d'extraction des connaissances.
- ❖ Le second chapitre portera sur la présentation des différents concepts nécessaires pour

extraire des règles d'association.

- ❖ Le troisième chapitre fait le point sur les algorithmes de génération des règles d'association avec une étude approfondie, ainsi les différentes approches utilisées par ces algorithmes.
- ❖ Le quatrième chapitre est consacré à la présentation des outils utilisés pour développer notre application ainsi une discussion sur notre travail réalisé et l'évaluation des résultats obtenus par ces algorithmes.

Enfin, nous indiquerons quelques remarques sur ce travail en concluant par un bilan et nous exposerons les perspectives pour de futurs travaux.

## Conclusion générale :

Dans ce mémoire, nous avons étudié les performances de quelques algorithmes de génération des règles d'association, ces algorithmes donnent des résultats différentes par rapport aux données déclarer par l'utilisateur

En effet, ces méthodes de génération des règles d'associations nécessite une optimisation pour diminuer le temps d'exécution ainsi la consommation de l'espace mémoire, Alors, il existe plusieurs facteurs qui contrôlent ces résultats obtenus :

- Le choix des paramètres par l'utilisateur comme le support minimal et confiance minimale.
- Le choix de nombre k des règles d'association qui doit être recherché.

Dans notre travail, nous avons basés sur l'étude comparative entre les deux algorithmes **Top-K-Rules** et **Top-K-Rules-Non Redondante** en fonction de plusieurs paramètres donner par l'utilisateur, ces méthodes se diffèrent en terme de redondance, cet redondance est contrôlée par deux stratégies :

- Eliminer une règle déjà existe dans la liste des règles trouvées.
- Eliminer une règle de la liste par rapport à une nouvelle règle trouvée.

Pour avoir des résultat exactes, nous avons ajoutés un paramètre delta ( $\Delta$ ) qui nous a permet d'augmenter l'espace de recherche mais d'avoir des résultats exacte, si le nombre des règles trouvés ne sont pas supérieure au nombre delta des règles d'association redondante, les résultats sont exacte.

Et comme perspective de notre travail, nous envisageons d'étendre notre travail pour la prise en compte des valeurs manquantes (missing values), car ce cas est très largement rencontré dans des bases de données réelles. Nous proposons aussi d'étendre ce travail afin de faire développer notre application par :

- Faire des optimisations dans la recherche des règles d'association.
- Réaliser un module d'analyse qui nous a permet de comparer les résultats d'une façon automatique et d'avoir traiter les différents règles trouvés par extraire les différentes connaissances qui seront exploités par l'utilisateur.

## Bibliographie :

- [1] : [https://fr.wikipedia.org/wiki/Exploration\\_de\\_donn%C3%A9es](https://fr.wikipedia.org/wiki/Exploration_de_donn%C3%A9es) consulté le : 12/12/2015
- [2] : **Philippe preux**, Fouille de données Notes de cours Ph. PREUX Université de Lille 3, philippe.preux@univ-lille3.fr, 26 mai 2011
- [3] : **MOUNA Azzeddine**, Mémoire de magister en Informatique, titre : datamining distribue dans les grilles :approche règles d'association, Université des sciences et technologie d'Oran, 2012/2013
- [4] : [https://fr.wikipedia.org/wiki/R%C3%A8gle\\_d%27association](https://fr.wikipedia.org/wiki/R%C3%A8gle_d%27association) consulté le : 25/12/2015
- [5] : **Alouane Basma**, Recherche de partitions floues optimales par segmentation floue pour la fouille de données quantitatives, Université M'hamed BOUGARA de BOUMERDES, 2007/2008
- [6]:<http://blog.khaledtannir.net/2012/07/lalgorithme-fp-growth-les-bases-13/#.VsBGM3pGRDs> consulté le : 28/12/2015
- [7] : **Gaël Le Mahec**, Utilisation des arbres de radicaux pour les algorithmes de Data-Mining sur grille de calcul, Université de Picardie Jules Verne - 24 juillet 2004
- [8]: **Zhi-Hong Deng, Sheng-Long Lv**, Fast mining frequent itemsets using Nodesets, Key Laboratory of Machine Perception (Ministry of Education), 2014
- [9]: **Zhi-Hong Deng, Sheng-Long Lv**, An efficient N-lists-based algorithm for mining frequent itemsets via Children-Parent Equivalence pruning Key Laboratory of Machine Perception (Ministry of Education), 2015
- [10] **Philippe Fournier-Viger**, Cheng-Wei Wu and Vincent S. Tseng, Mining Top-K Association Rules, Philippe Fournier-Viger, Cheng-Wei Wu and Vincent S. Tseng, Dept. of Computer Science, University of Moncton, canada, philippe.fv@gmail.com, 2012
- [11] **Philippe Fournier-Viger1 and Vincent S**, Mining Top-K Non-Redundant Association Rules, Dept. of Computer Science, University of Moncton, Canada, Dept. of Computer Science and Info. Engineering, Natiotsengsm@mail.ncku.edu.tw, nal Cheng Kung University, Taiwan philippe.fournier-viger@umoncton.ca, 2012
- [12]:<https://www.techopedia.com/definition/17374/one-tier-architecture> consulté le : 12/03/2016
- [13] : **Aymen HEDIDAR**, CONCEPTION ET REALISATION D'UNE APPLICATION MOBILE M-BANKING, université virtuelle de Tunis, 2011/2012
- [14] : <https://fr.wikipedia.org/wiki/PostgreSQL> consulté le : 10/04/2016

- [15]: **Ramakrishnan srikant**, Fast algorithms for mining association rules and sequential patterns by the university of wisconsin madison 1996.
- [16] : <http://www.jfree.org/jfreechart/> consulté le : 18/04/2016
- [17]: **Thomas H.cormen**, Introduction to Algorithms, Second Edition,
- [18] : [https://fr.wikipedia.org/wiki/R%C3%A8gle\\_d%27association](https://fr.wikipedia.org/wiki/R%C3%A8gle_d%27association) consulté le : 21/04/2016
- [19] : **Nicolas Pasquier**, Extraction de Bases pour les Règles d'Association à partir des Itemsets Fermés Fréquents, Laboratoire d'Informatique (LIMOS) - Université Clermont-Ferrand II,

## Résumé :

Dans le domaine de data-mining, La génération des règles d'association est la méthode d'extraction des informations insignifiant et des connaissances utiles qui sont cachés dans une base de données volumineuse.

Dans ce contexte, un nombre d'algorithmes s'inspirant de deux approches pour la génération des règles d'association :

- Un certain nombre d'algorithmes, s'inspirant de l'approche classique, basés sur l'extraction des itemsets fréquents puis la génération des règles d'association. La particularité de ces algorithmes est qu'ils génèrent un nombre énorme de règles rendant leur exploitation quasiment impossible.
- D'autres part, Un certain algorithme utilisent une nouvelle approche qui est basé sur la génération des règles d'association avec l'expansion des règles.

Dans ce projet, Nous exécutons des algorithmes pour générer les règles d'association en basant sur des données réels importée depuis une base de données, Ainsi, d'évaluer les performances de chaque d'algorithmes.

**Mots clés :** *Data mining, itemset fréquents, règles d'association, expansion des règles.*

## Abstract :

In data mining field, Association rules mining is the method of extracting the insignificant information and useful knowledge, which is hidden in a large database.

In this context, a number of algorithms based on two approaches for the generation of association rules:

- A number of algorithms, inspired by the traditional approach, based on the extraction of frequent itemsets and generate association rules. The special feature of these algorithms is that they generate a huge number of rules making them almost impossible to exploit.
- Other hand, a certain algorithms using a new approach that is based on the generation of association rules with the rules expansion.

In this project, we execute algorithms to generate association rules based on real data imported from a database, thus to evaluate the performance of each algorithm.

**Keywords :** *Data mining, frequent itemsets, association rules, rules expansion*

## ملخص :

في مجال التنقيب عن معلومات, استخراج قواعد الارتباط هو طريقة تنقيب على المعلومات والمعطيات, التي هي مخبأة في قواعد بيانات ضخمة, لذلك هناك عدة خوارزميات للبحث عن هذه القواعد و التي تستعمل نهجين للبحث عنها وهي كالآتي:

- عدد من خوارزميات مستوحاة من النهج التقليدي, استنادا إلى استخراج العناصر المتكررة, وتوليد قواعد الارتباط, الميزة الخاصة لهذه الخوارزميات هي أنها تولد عدد كبير من قواعد يجعلها عملية شبه مستحيلة لاستخدامها وتحليلها.
  - ومن ناحية أخرى, هناك بعض خوارزميات من تستخدم نهج جديد يقوم على توليد قواعد الارتباط عن طريق قواعد التوسع.
- في هذا مشروع, نقوم على تنفيذ خوارزميات لتوليد قواعد الارتباط بناء على بيانات حقيقية مستوردة من قواعد بيانات, وبالتالي تقييم أداء كل خوارزمية.
- كلمات مفتاحية: قاعدة بيانات, العناصر المتكررة, قواعد الارتباط, قواعد التوسع.