# INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATION & MANAGEMENT

IJRCM

A Monthly Double-Blind Peer Reviewed (Refereed/Juried) Open Access International e-Journal - Included in the International Serial Directories

*Indexed & Listed at:*

Ulrich's Periodicals Directory ©, ProQuest, U.S.A., EBSCO Publishing, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A.,
Open J-Gage, India [link of the same is duly available at Inflibnet of University Grants Commission (U.G.C.)],
Index Copernicus Publishers Panel, Poland with IC Value of 5.09 & number of libraries all around the world.

Circulated all over the world & Google has verified that scholars of more than 3130 Cities in 166 countries/territories are visiting our journal on regular basis.

Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA

http://ijrcm.org.in/

# CONTENTS

# CALL FOR MANUSCRIPTS

We invite unpublished novel, original, empirical and high quality research work pertaining to recent developments & practices in the areas of Computer Science & Applications; Commerce; Business; Finance; Marketing; Human Resource Management; General Management; Banking; Economics; Tourism Administration & Management; Education; Law; Library & Information Science; Defence & Strategic Studies; Electronic Science; Corporate Governance; Industrial Relations; and emerging paradigms in allied subjects like Accounting; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Rural Economics; Co-operation; Demography: Development Planning; Development Studies; Applied Economics; Development Economics; Business Economics; Monetary Policy; Public Policy Economics; Real Estate; Regional Economics; Political Science; Continuing Education; Labour Welfare; Philosophy; Psychology; Sociology; Tax Accounting; Advertising & Promotion Management; Management Information Systems (MIS); Business Law; Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labour Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; International Relations; Human Rights & Duties; Public Administration; Population Studies; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism & Hospitality; Transportation Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic; Web Design and emerging paradigms in allied subjects.

Anybody can submit the **soft copy** of unpublished novel; original; empirical and high quality **research work/manuscript** ***anytime*** in ***M.S. Word format*** after preparing the same as per our **GUIDELINES FOR SUBMISSION**; at our email address i.e. infoijrcm@gmail.com or online by clicking the link **online submission** as given on our website (***FOR ONLINE SUBMISSION, CLICK HERE***).

# GUIDELINES FOR SUBMISSION OF MANUSCRIPT

1. **COVERING LETTER FOR SUBMISSION:**

**DATED: _____________**

***THE EDITOR***
IJRCM

Subject: **SUBMISSION OF MANUSCRIPT IN THE AREA OF** .

**(e.g. Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/Engineering/Mathematics/other, please specify)**

**DEAR SIR/MADAM**

Please find my submission of manuscript entitled '___________________________________________' for possible publication in your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published elsewhere in any language fully or partly, nor is it under review for publication elsewhere.

I affirm that all the author (s) have seen and agreed to the submitted version of the manuscript and their inclusion of name (s) as co-author (s).

Also, if my/our manuscript is accepted, I/We agree to comply with the formalities as given on the website of the journal & you are free to publish our contribution in any of your journals.

**NAME OF CORRESPONDING AUTHOR**:
Designation:
Affiliation with full address, contact numbers & Pin Code:
Residential address with Pin Code:
Mobile Number (s):
Landline Number (s):
E-mail Address:
Alternate E-mail Address:

**NOTES**:
a) The whole manuscript is required to be in ***ONE MS WORD FILE*** only (pdf. version is liable to be rejected without any consideration), which will start from the covering letter, inside the manuscript.
b) The sender is required to mentionthe following in the **SUBJECT COLUMN** of the mail:
**New Manuscript for Review in the area of** (Finance/Marketing/HRM/General Management/Economics/Psychology/Law/Computer/IT/ Engineering/Mathematics/other, please specify)
c) There is no need to give any text in the body of mail, except the cases where the author wishes to give any specific message w.r.t. to the manuscript.
d) The total size of the file containing the manuscript is required to be below **500 KB**.
e) Abstract alone will not be considered for review, and the author is required to submit the complete manuscript in the first instance.
f) The journal gives acknowledgement w.r.t. the receipt of every email and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of manuscript, within two days of submission, the corresponding author is required to demand for the same by sending separate mail to the journal.

2. **MANUSCRIPT TITLE**: The title of the paper should be in a 12 point Calibri Font. It should be bold typed, centered and fully capitalised.
3. **AUTHOR NAME (S) & AFFILIATIONS**: The author (s) **full name**, **designation**, **affiliation** (s), **address**, **mobile/landline numbers**, and **email/alternate email address** should be in italic & 11-point Calibri Font. It must be centered underneath the title.
4. **ABSTRACT**: Abstract should be in fully italicized text, not exceeding 250 words. The abstract must be informative and explain the background, aims, methods, results & conclusion in a single para. Abbreviations must be mentioned in full.

5. **KEYWORDS**: Abstract must be followed by a list of keywords, subject to the maximum of five. These should be arranged in alphabetic order separated by commas and full stops at the end.

6. **MANUSCRIPT**: Manuscript must be in ***BRITISH ENGLISH*** prepared on a standard A4 size ***PORTRAIT SETTING PAPER***. It must be prepared on a single space and single column with 1" margin set for top, bottom, left and right. It should be typed in 8 point Calibri Font with page numbers at the bottom and centre of every page. It should be free from grammatical, spelling and punctuation errors and must be thoroughly edited.

7. **HEADINGS**: All the headings should be in a 10 point Calibri Font. These must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.

8. **SUB-HEADINGS**: All the sub-headings should be in a 8 point Calibri Font. These must be bold-faced, aligned left and fully capitalised.

9. **MAIN TEXT**: The main text should follow the following sequence:

   INTRODUCTION

   REVIEW OF LITERATURE

   NEED/IMPORTANCE OF THE STUDY

   STATEMENT OF THE PROBLEM

   OBJECTIVES

   HYPOTHESES

   RESEARCH METHODOLOGY

   RESULTS & DISCUSSION

   FINDINGS

   RECOMMENDATIONS/SUGGESTIONS

   CONCLUSIONS

   SCOPE FOR FURTHER RESEARCH

   ACKNOWLEDGMENTS

   REFERENCES

   APPENDIX/ANNEXURE

   It should be in a 8 point Calibri Font, single spaced and justified. The manuscript should preferably not exceed ***5000 WORDS***.

10. **FIGURES &TABLES**: These should be simple, crystal clear, centered, separately numbered &self explained, and **titles must be above the table/figure**. **Sources of data should be mentioned below the table/figure**. It should be ensured that the tables/figures are referred to from the main text.

11. **EQUATIONS**:These should be consecutively numbered in parentheses, horizontally centered with equation number placed at the right.

12. **REFERENCES**: The list of all references should be alphabetically arranged. The author (s) should mention only the actually utilised references in the preparation of manuscript and they are supposed to follow **Harvard Style of Referencing**. The author (s) are supposed to follow the references as per the following:

- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
- Use (**ed.**) for one editor, and (**ed.s)** for multiple editors.
- When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc, in chronologically ascending order.
- Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
- The title of books and journals should be in italics. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
- For titles in a language other than English, provide an English translation in parentheses.
- The location of endnotes within the text should be indicated by superscript numbers.

**PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:**

**BOOKS**

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

**CONTRIBUTIONS TO BOOKS**

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

**JOURNAL AND OTHER ARTICLES**

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

**CONFERENCE PAPERS**

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–22 June.

**UNPUBLISHED DISSERTATIONS AND THESES**

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, Kurukshetra University, Kurukshetra.
- **ONLINE RESOURCES**
- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

**WEBSITES**

- Garg, Bhavet (2011): Towards a New Natural Gas Policy, Political Weekly, Viewed on January 01, 2012 http://epw.in/user/viewabstract.jsp

# DETERMINING APPROXIMATE FUNCTIONAL DEPENDENCIES USING ASSOCIATION RULE MINING

*SIKHA BAGUI*
*PROFESSOR*
*DEPARTMENT OF COMPUTER SCIENCE*
*UNIVERSITY OF WEST FLORIDA*
*PENSACOLA*

*ANTON ZAYNAKOV*
*IT SUPPORT TECHNICIAN*
*DEPARTMENT OF COMPUTER SCIENCE*
*UNIVERSITY OF WEST FLORIDA*
*PENSACOLA*

## ABSTRACT

*In this paper we present a unique way to analyze the support and confidence of association rules to come up with Approximate Functional Dependencies (AFDs). We also discuss how the nature of AFDs determined from association rule mining is different from functional dependencies (FDs) in the relational model.*

## KEYWORDS

Functional dependencies, approximate functional dependencies, association rule mining, relational databases, apriori algorithm, support, confidence.

## 1. INTRODUCTION

Association rule mining is used to find relationships between items or itemsets in *market basket* or transactional data representations (Agrawal, et al. 1993). Statistical measures like support and confidence are used to measure the strength of these relationships (Agrawal and Srikant, 1994). In this paper we are trying to determine if these relationships are FD relationships as defined in relational representations.

FDs exist between attributes in a relational representation. In a relational representation, there is a FD between the primary key attribute and other attributes. Association rule mining, however, finds correlations among data contents and works at the instance or attribute-value level rather than the attribute level. In this paper we would like to see if FDs or AFDs can be determined from association rule mining using association rule mining's statistical measures like support and confidence. It would appear that 100 % confidence in association rule mining would translate to FDs, but as we will see in this paper, this is not always true. Hence we use association rule mining to determine AFDs. We will also see that the FDs or AFDs determined from association rule mining have a different nature.

Data in market basket data representations (or in transaction datasets) is in the form represented in (Han and Kamber, 2012), as shown in Figure 1. There is no particular order, format or domain constraint for the items purchased. Transactions can also have duplicate items or *n* number of same items, and the number of items purchased is not fixed or limited.

Using association rule mining, figure 1 would show not only which items are purchased in which transaction but also which items are purchased when other items are purchased. The items in figure 1 are shirt, pen, book, etc. Itemsets are sets of items. Examples of 1-itemsets would be {shirt}, {pen}, {book}; examples of 2-itemsets would be {shirt, pen}, {shirt, book}, {pen, book}; examples of 3-itemsets would be {shirt, pen, book}, and so on. Data in transactional datasets do not follow the principles of database normalization.

**FIGURE 1: MARKET BASKET DATA REPRESENTATION**

| Transaction_ID | Items_Purchased |
|---|---|
| T100 | Shirt, pen |
| T200 | Shirt, book |
| T300 | Pen, book |
| T400 | Shirt, pen, book |
| T500 | Bread, cake, shoes, socks |

Data in relational databases, however, is more orderly and structured and follows principles of FDs and database normalization. A relational representation would be made of a finite number of attributes, and the attribute values would have to be within a valid domain. Relations within a relational representation would have a key to identify a unique tuple or row and there would be FDs between the key or keys of the table and the other attributes in that row. There would also be no multi-valued attributes in a relational representation. A formal relational database representation (Elmasri and Navathe (2007), Date (2003), Bagui and Earp, 2012) is very different in format from the representation presented in figure 1. A relational database representation is presented in figure 2.

**FIGURE 2: A RELATIONAL DATABASE REPRESENTATION**

| STUDENT | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| STNO | SNAME | MAJOR | CLASS | BDATE | AGE | HIGHSCHOOL | CAMPUSRESIDENT | HRSWORKED | GPA |
| 200 | John | ENG | 1 | 1/2/1990 | under_25 | Highland Park | YES | 21_40 | 2.90 |
| 300 | Mary | CS | 4 | 7/6/1987 | under_25 | Pensacola | NO | less_10 | 3.5 |

| | CLASSES | | | | | |
|---|---|---|---|---|---|---|
| STNO | CLASSID | SECID | INSTRID | ISBN | TIME | DAYS |
| 200 | COP3698 | 0456 | 96 | 2901558609012 | 12 - 1:15 | TR |
| 200 | ART2103 | 9876 | 70 | 5693256148965 | 8:30 - 9:15 | MW |
| 300 | COP3698 | 6789 | 96 | 2901558609012 | 5:30 - 8:45 | T |
| 300 | ART2103 | 0321 | 70 | 5693256148965 | 11:15 - 12:45 | TR |

As can be seen from figure 2, relational schemas have fixed columns (attributes) and values within the columns fall within a fixed domain. The data is also of a particular type. Every relational schema has a primary key, and the other attributes in the relation are functionally dependent on the primary key. Hence FDs, the key to relational representations, are determined at the attribute (columnar) level in relational representations. FD means that the value of an attribute is uniquely determined by some other attribute, usually the key, for example, from figure 2, the stno (student number) would determine the sname, major, class, etc.

In this paper we determine a way to use association mining rules to find AFDs in large datasets. Determining AFDs in large datasets using association rule mining will serve the following purposes: (i) it will help in reverse engineering (Alashqur, 2009); Since relational databases are so widely used (Ceri, et al., 2000; Kappel, et al., 2001a; Kappel, et al., 2001b; Shanmugasundaram, et al., 2001), it might be necessary to reverse engineer to a relational database to take advantage of the benefits of relational databases; (ii) it can help in data prediction (Wolf, et al., 2007); (iii) It can help in further analyzing association mining rules.

The rest of the paper is organized as follows: Section 2 presents the relational representation; section 3 defines FDs; section 4 presents association rule mining; section 5 defines AFDs; section 6 discusses related works; section 7 shows how we calculated the AFDs using some real datasets; section 8 presents a discussion of the results; and section 9 presents the conclusions.

## 2. RELATIONAL REPRESENTATION

A relational schema R, denoted by R($A_1$, $A_2$, …, $A_n$), is composed of a relation name, R and a list of attributes $A_1$, $A_2$,…$A_n$. Each attribute $A_i$ has a domain made of a set of atomic values. Atomic means that the values are not divisible into components within the framework of R. A relational state, r, is made of n-tuples, where r = {$t_1$, $t_2$, …, $t_n$}. Each tuple is an ordered list of values, so t = <$v_1$, $v_2$, …, $v_n$>. Each value $v_i$ is within a specified domain (Elmasri and Navathe, 2007), and must have a value or will be null. Multivalued attributes are not allowed in the relational model and composite attributes are represented by their simple component attributes.

The relational representation is typically made up of more than one relational schema, as shown in figure 2. Figure 2 has two tables or relations, STUDENT and CLASSES. In the STUDENT table, each tuple represents an entity or student and in the CLASSES table each tuple represents an entity or class. The STUDENT table has stno (student number) as the primary key and the CLASSES table has secID as the primary key. In both tables, the rest of the attributes are fully functionally dependent on the primary key. FDs are explained next.

## 3. FUNCTIONAL DEPENDENCIES (FDs)

FDs can be determined by the semantics of attributes, but they can also be inferred or deduced. FDs are the basis for relational database theory. A FD can be defined as a relationship between two attributes or sets of attributes in a relation. Given a relation R, with *n* attributes, $A_1$, $A_2$, $A_3$, …, $A_n$, attribute $A_y$ of R is functionally dependent on attribute $A_x$ of R, ($A_x \rightarrow A_y$) (we will use "→" to show FD), if and only if each $A_x$ in R is associated with precisely one $A_y$ in R (in a particular database state). So, any two tuples, $t_1$ and $t_2$ in R in the form $t_1[X] = t_2[X]$ must also have $t_1[Y] = t_2[Y]$. That is, the values of the Y component of a tuple in R depends on, or are determined by, the values of the X component of the tuple; or the values of the X component of a tuple functionally determine the values of the Y component (Earp and Bagui, 2012; Elmasri and Navathe, 2007), hence Y is functionally dependent on X.

FDs cannot necessarily be reversed. That is, X → Y in a relation R, does not imply Y → X in a relation R. A functional dependency may also be between two sets of attributes, that is, between composite attributes. In relational databases, FDs hold all the time, which is in 100% of the cases.

## 4. ASSOCIATION RULES

Association rules are presented in the form $A \Rightarrow B$, where the rule body *A* (Left Hand Side (LHS)) and the head *B* (Right Hand Side (RHS)) are subsets of the set of items $I = \{i_1, i_2, \ldots, i_n\}$from a set of transactions $D = \{t_1, t_2, \ldots, t_n\}$, where $t_i (i \in [1,N])$ is a transaction and $t_i \subseteq I$, and $A \cap B = \emptyset$. Every subset of *I* is called an itemset. If an itemset contains *k* items, then it is called a *k*-itemset. The strength of an association rule is measured by a rule's support and confidence.

A rule's support measures the number of times ($A \cup B$) occurs together in a dataset. That is, the probability, $P(A \cup B)$. (Han and Kamber, 2012).

$support(A \Rightarrow B) = P(A \cup B)$

Confidence is taken to be the conditional probability, $P(B|A)$. (Han and Kamber, 2012). That is, the number of times A and B occur when A occurs.

$confidence\ (A \Rightarrow B) = P(B|A) = support(A \cup B)/support(A) = support\_count(A \cup B)/support\_count(A)$

Rules with high confidence and strong (reasonably large or high) support are referred to as strong rules (Agrawal, et al.1993; Han and Kamber, 2012; Park, Chen and Yu, 1995; Tan, Steinbach and Kumar, 2006). A rule with very low support may occur simply by chance. Confidence, on the other hand, measures the reliability of an inference rule. So, the higher the confidence, the more likely it is for *B* to be present in transactions that contains *A*.

One of the most commonly used algorithms for association rule mining is the Apriori algorithm. Next we explain the Apriori algorithm. The Apriori algorithm can be decomposed into the following two step process (Han and Kamber, 2006):

1. Find all frequent itemsets. An itemset that contains *k* items is a *k*-itemset. All frequent itemsets will occur at least as frequently as a pre-determined minimum support count.
2. Generate strong association rules from the frequent itemsets – these rules must satisfy a minimum support and minimum confidence.

The overall performance of mining association rules is determined by the first step.

### 4.1 ALGORITHM TO MINE ASSOCIATION RULES

The Apriori algorithm finds frequent itemsets using an iterative approach based on candidate generation. Below we present the pseudocode for the Apriori algorithm, as presented in (Han & Kamber, 2006):

```
L1 := {frequent 1-itemsets} D;
for (k=2; Lk-1 ≠ ∅ ; k++)
        Ck = apriori_gen(Lk-1, min_sup);
        for each transaction t ∈ D { //scan D for counts
                Ct = subset(Ck, t); // get the subsets of t that are candidates
                for each candidate c ∈ Ct
                        c.count++;
}
        Lk = { c ∈ Ck | c.count ≥ min_sup}
}
return L= Uk Lk;

procedure apriori_gen(Lk-1; frequent(k-1)-itemsets; min_sup: minimum support threshold)
        for each itemset l1 ∈ Lk-1
                for each itemset l2 ∈ Lk-1
                        if (l1[1] = l2[1]) ∧ (l1[2] = l2[2]) ∧ … ∧ (l1[k-2] = l2[k-2]) ∧ (l1[k-1] = l2[k-1]) then {
                                c = l1 |X| l2 ;            // join step: generates candidates
                                if has_infrequent_subset(c, Lk-1) then
                                        delete c;    //prune step: remove unfruitful candidate
```

**else add** *c to $C_k$;*

}

**return** $C_k$;

**procedure has_infrequent_subset**(c:candidate k-itemset; $L_{k-1}$: frequent ($k$-1)-itemsets);

//use prior knowledge

**for each** ($k$-1) – subset $s$ of $c$

**if** $s \notin L_{k-1}$ then

**return** TRUE;

**return** FALSE;

This Apriori algorithm employs an iterative approach known as a level-wise search, where $k$-itemsets are used to explore ($k$+1) – itemsets. First, the set of frequent 1-itemsets is found, denoted by $L_1$. All these frequent 1-itemsets have to have *support* above a user specified minimum. The frequent 1-itemsets are generated by counting item occurrences and then those that turn out to be frequent after computing their support are used.

$L_1$ is then used to find $L_2$ , the set of frequent 2-itemsets, which is used to find $L_3$, and so on until no more frequent $k$-itemsets can be found. The size of the itemsets is incremented by one at each iteration, and the finding of each $L_k$ requires one full scan of the database. This phase stops when there are no frequent itemsets.

The apriori_gen procedure performs two steps – a join and a prune. In the join part, $L_{k-1}$ is joined with $L_{k-1}$ to generate potential candidates. The prune portion employs the Apriori property to remove candidates that have a subset that is not frequent. The test for infrequent subsets is shown in procedure has_infrequent_subset (Han & Kamber, 2006).

## 5. APPROXIMATE FUNCTIONAL DEPENDENCIES (AFDs)

An AFD will hold most of the time, but not all the time. Though FDs form the basis for database theory, AFDs can also have applications in database design (Bra and Paredaens, 1984) and the discovery of unexpected but meaningful approximate dependencies can also be used in data mining applications (Huhtala, et al., 1999). For example, in an environmental dataset an AFD could point to the causes for air pollution, and these can then be further investigated by domain experts.

We will define an AFD as: Given a relational representation R, attribute Y of R is approximately dependent on attribute X of R if and only if each X in R has associated with it one Y in R in at least (Z – T) cases. Z is the number of tuples or rows in R, and T is the number of tuples that have to be removed for each X in R to have associated with it precisely one Y. We will denote the AFD with a "$\rightsquigarrow$".

## 6. RELATED WORK

AFDs have been studied by a few. Huhtala, et al. (1999) presented the Tane algorithm to determine functional and approximate dependencies from large databases. Tane is based on partitioning sets of rows by their attribute values.

Kivinen and Mannila (1995) discussed several measures to determine the error of dependencies and derived bounds for discovering dependencies with errors.

Ilyas, et al. (2004) developed a system called CORDS to determine statistical correlations and soft FDs. One of the drawbacks of CORDS is that it works with a sample of data, hence we cannot assert with complete certainty if a functional dependency always holds.

Giannella and Robertson (2004) examined how to measure, based on information theory, the degree to which a FD is approximate. Their measure is compared with the other two standard measures, g3 and Tau.

Kalavagattu (2008) measured for AFDs (derived from association rules) and also presented an algorithm for generating AFDs according to measures of confidence and specificity with derivations.

Alashqur (2009) describes the similarities and differences between FDs and association rules and introduces a formal definition of FDs in terms of association rules. But Alashqur (2009) only talks about FDs whose confidence is 100%. We deal with AFDs whose confidence may be less that 100%.

Sanchez et al (2008) provide a methodology to adapt existing association rule mining algorithms to the task of discovering Approximate Dependencies. The adapted algorithms obtain the set of Approximate Dependencies that hold in a relation with accuracy and support greater than user-defined thresholds.

Approximate functional dependencies were also studied as fuzzy functional dependencies by some (Berzal, et al (2005); Sanchez et al. (2003)). Calero, et al. (2003, 2004a, b) introduced a methodology that employed fuzzy approximate dependencies for perform a high-level analysis of data.

Though AFDs have been studied by a few, none of them have studies related to determining AFDs from association rule mining using the statistical measures of support and confidence (where confidence is below 100%) using the approach we took. We present a unique way to analyze the support and confidence of association rules to come up with AFDs.

## 7. EXPERIMENTAL RESULTS

Our aim in this work is to determine AFDs between attribute-value pairs of association rules in large datasets using 2-item rules. This work can be extended to more than 2-item rules, but we do not consider that scenario in this paper. Our proposed method works for 2-item rules, that is, one attribute-value pair on either side of an association rule.

We tested our ideas using five datasets. We will present details of the work using the first dataset, Colleges, available at **[**ftp://85.158.30.137/lib.stat.cmu.edu/datasets/colleges/aaup.data**]**. This dataset has 1161 rows.

**STEP 1**

Our first step was to categorize the data to make it ready for association rule mining. We then ran the Apriori algorithm using Weka on the categorized dataset using a minimum support of 0.01 and minimum confidence of 1. The reason for the low minimum support and high confidence numbers was to get all possible 2-item rules so that we could create 2-itemsets out of 2-item rules with 100% confidence.

From this initial run we selected two 2-item rules with 100% confidence:

- Type=IIB 618 ==> NFP=NFPlowest 618
- Average Salary Assistant Proffessors=ASASPlow 415 ==> Number of Associate Professors=NAPlowest 415

The next step was to run Weka's Apriori algorithm using the attributes from the 2-item rules, hence we first ran the Apriori algorithm on the attributes from the first rule and then on the attributes from the second rule.

**STEP 2**

We ran Weka's Apriori algorithm on the Type and Number of Full Professors attributes (the attributes from the first rule), with the lowest minSupport (0.001) and minConfidence (0.001) settings. We got the following rules:

1. Type=IIB 618 ==> NFP=NFPlowest 618 conf:(1)
2. NFP=NFPhigh 11 ==> Type=I 11 conf:(1)
3. NFP=NFPhighest 9 ==> Type=I 9 conf:(1)
5. Type=IIA 363 ==> NFP=NFPlowest 337 conf:(0.93)
6. NFP=NFPmed 45 ==> Type=I 37 conf:(0.82)
7. NFP=NFPlow 88 ==> Type=I 70 conf:(0.8)
8. NFP=NFPlowest 1008 ==> Type=IIB 618 conf:(0.61)

9. Type=I 180 ==> NFP=NFPlow 70 conf:(0.39)
10. NFP=NFPlowest 1008 ==> Type=IIA 337 conf:(0.33)
11. Type=I 180 ==> NFP=NFPlowest 53 conf:(0.29)
12. Type=I 180 ==> NFP=NFPmed 37 conf:(0.21)
13. NFP=NFPlow 88 ==> Type=IIA 18 conf:(0.2)
14. NFP=NFPmed 45 ==> Type=IIA 8 conf:(0.18)
15. Type=I 180 ==> NFP=NFPhigh 11 conf:(0.06)
16. NFP=NFPlowest 1008 ==> Type=I 53 conf:(0.05)
17. Type=I 180 ==> NFP=NFPhighest 9 conf:(0.05)
18. Type=IIA 363 ==> NFP=NFPlow 18 conf:(0.05)
19. Type=IIA 363 ==> NFP=NFPmed 8 conf:(0.02)

The possible values of Type were Type = IIB, Type = IIA and Type = I. We kept one rule for each value of the Type attribute. The rule with the highest confidence was kept. For example, Type = IIA had 3 rules with confidences of 93%, 5% and 2% respectively. We kept the rule with the 93% confidence and did not use the rest of the rules. For Type = I, since the rule with the highest confidence had a confidence of 39%, we did not use it since we would only keep it if the rule's confidence was above 50% (this confidence is a user-defined confidence and is selected arbitrarily). For Type IIB, since this is a rule with confidence of 100%, we kept this one; so we ended up with the following:

| | | |
|---|---|---|
| Type = IIB | 618 tuples out of 618 were retained | |
| Type = IIA | 337 tuples out of 363 were retained | 26 tuples were removed |
| Type = I | 0 tuples were retained | 180 tuples were removed |

This means that, in this dataset, out of 1161 rows, if Type IIB occurs, then this always leads to NFPlowest, since this has 100% confidence, and this happened in 618 cases or 53% of the time. Similarly, when Type IIA occurred, NFPlowest occurred 93% of the time (shown by the confidence) and this happened 29% in the whole dataset (the support). In this study we are trying to get a combined support of at least 80%. Since Type IIB and Type IIA accounted for 82% of the data we will continue with AFD calculation.

We calculated the AFD's as follows:

***Step 2.1***

Calculate the total number of tuples removed. We based this on confidence of the association rules. If the confidence is below 50%, the rules are removed. Since Type = I had confidence less than 50%, these rules were not used. Also we try to obtain a combined support of at least 80%.

HR = Highest rule from each attribute-value combination with confidence > 50%

Total number of tuples retained = Total_retained

Total_retained% = 955/1161 = 82%

Total_retained = $\sum_{n=1}^{\infty}(HR)$

Total_retained = 618 + 337 = 955

Total number of tuples removed = Total_removed

Total_removed = Dataset size – Total_retained

Total_removed = 1161 – 955 = 206

***Step 2.2***

Next we calculate the impurity.

Impurity % = (Total_removed /dataset size) * 100

Impurity % = 206/1161 * 100

Impurity = 17.8%

***Step 2.3***:

The Approximate Functional Dependency (AFD):

AFD = 100 – Impurity%

AFD = 100 – 17.8

AFD = 82.2%

Therefore, we can conclude that dependency (Type → Number of Full Professors) has AFD with the strength=82.2% and that the impurity of this AFD is 17.8% or 206 tuples. This would imply that for the attributes Type and Number of Full Professors, if we know the Type, we can predict the Number of Full Professors with an accuracy of 82.2%

**Step 3**

Next we will consider the second 2-item rule. The attributes are Average Salary Assistant Professor (ASASP) and Number of Associate Professors (NAP). We ran Weka's Apriori algorithm using these two attributes with the lowest minSupport and minConfidence settings. We got the following rules:

1. ASASP=ASASPlow 415 ==> NAP=NAPlowest 415 conf:(1)
2. NAP=NAPhigh 9 ==> ASASP=ASASPmed 9 conf:(1)
3. NAP=NAPhighest 2 ==> ASASP=ASASPmed 2 conf:(1)
4. NAP=NAPlow 117 ==> ASASP=ASASPmed 106 conf:(0.91)
5. NAP=NAPmed 35 ==> ASASP=ASASPmed 31 conf:(0.89)
6. ASASP=ASASPmed 709 ==> NAP=NAPlowest 561 conf:(0.79)
7. ASASP=ASASPhigh 37 ==> NAP=NAPlowest 22 conf:(0.59)
8. NAP=NAPlowest 998 ==> ASASP=ASASPmed 561 conf:(0.56)
9. NAP=NAPlowest 998 ==> ASASP=ASASPlow 415 conf:(0.42)
10. ASASP=ASASPhigh 37 ==> NAP=NAPlow 11 conf:(0.3)
11. ASASP=ASASPmed 709 ==> NAP=NAPlow 106 conf:(0.15)
12. NAP=NAPmed 35 ==> ASASP=ASASPhigh 4 conf:(0.11)
13. ASASP=ASASPhigh 37 ==> NAP=NAPmed 4 conf:(0.11)
14. NAP=NAPlow 117 ==> ASASP=ASASPhigh 11 conf:(0.09)
15. ASASP=ASASPmed 709 ==> NAP=NAPmed 31 conf:(0.04)
16. NAP=NAPlowest 998 ==> ASASP=ASASPhigh 22 conf:(0.02)
17. ASASP=ASASPmed 709 ==> NAP=NAPhigh 9 conf:(0.01)
18. ASASP=ASASPmed 709 ==> NAP=NAPhighest 2 conf:(0)

The values of Average Salary Assistant Professors were ASASPlow, ASASPmed, and ASASPhigh. Again, keeping one rule for each value of the ASASP attribute (and only the rules with confidence > 50%), we have:

| | | |
|---|---|---|
| ASASP = ASASPlow | 415 tuples out of 415 were retained | |
| ASASP = ASASPmed | 561 tuples out of 709 were retained | 148 tuples were removed |
| ASASP = ASASPhigh | 22 tuples out of 37 were retained | 15 tuples were removed |

In this case, ASASPlow accounts for 35.74% of the data (the support) and ASASPmed accounts for 48.32% of the data. ASASPlow and ASASPmed taken together account for 84.06% of the data. ASASPhigh had a very low support of 1.89%, but we kept it since this rule's confidence was 59% (in this study we are keeping rules with confidence > 50%). So, the process is, we first check for the confidence, and if the confidence is above 50%, we keep the rule, even if the support is very low.

***Step 3.1:***

Total_retained = $\sum_{n=1}^{\infty}(HR)$

Total_retained = 415 + 561 + 22 = 998

Total_retained% = 998/1161 = 85.96%

Total_removed = 1161 – 998 = 163

***Step 3.2:***

Impurity % = (Total_removed /dataset size) * 100

Impurity% = (163/1161) * 100 = 14%

***Step 3.3:***

The Approximate Functional Dependency (AFD):

AFD = 100 – Impurity%

AFD = 100 – 14

AFD = 86%

Therefore, we can conclude that dependency Average Salary Assistant Professors → Number of Associate Professors has AFD with the strength 86% and this AFD's impurity = 14% or 163 tuples. This means that if we know the ASASP, we can predict the NAP with an accuracy of 86%.

Using the same steps, we calculated the impurities and AFDs for the other four datasets. Figure 3 presents the 2-item rules extracted from each dataset. These were obtained using the minSupport of 0.01 and minConfidence of 1. Medical data set did not have 2-item rules with 100% confidence; therefore, we had to lower the confidence value to 95%.

**FIGURE 3: 2-ITEM RULES**

| |
|---|
| **Dataset: Colleges**<br>ftp://85.158.30.137/lib.stat.cmu.edu/datasets/colleges/aaup.data<br>**2-item rules with 100% confidence** |
| • Type=IIB 618 ==> NFP=NFPlowest 618 |
| • Average Salary Assistant Proffessors=ASASPlow 415 ==> Number of Associate Professors=NAPlowest 415 |
| **Dataset: Forest Fires**<br>http://archive.ics.uci.edu/ml/datasets/Forest+Fires<br>**2-item rules with 100% confidence** |
| • RH=Rhlow 305 ==> rain=RAINlow 305 conf:(1) |
| • temp=TEMPmed 236 ==> rain=RAINlow 236 conf:(1) |
| • DMC=DMClow 210 ==> rain=RAINlow 210 conf:(1) |
| • DMC=DMClow 210 ==> area=AREAsmallest 210 conf:(1) |
| • wind=WINDlow 209 ==> ISI=ISIlow 209 conf:(1) |
| • wind=WINDlow 209 ==> rain=RAINlow 209 conf:(1) |
| • wind=WINDlow 209 ==> area=AREAsmallest 209 conf:(1) |
| • X=Xcentral 207 ==> rain=RAINlow 207 conf:(1) |
| • X=Xeast 176 ==> area=AREAsmallest 176 conf:(1) |
| • X=Xwest 134 ==> FFMC=FFMChigh 134 conf:(1) |
| **Dataset: Green Vehicle data**<br>https://explore.data.gov/Transportation/Green-Vehicle-Guide-Data-Downloads/9un4-5bz7<br>**2-item rules with 100% confidence** |
| • Eng Displ=med_Eng_Displ 290 ==> # Cyl=med_cyl 290 conf:(1) |
| • # Cyl=high_cyl 234 ==> Eng Displ=large_Eng_Displ 234 conf:(1) |
| **Dataset: Contraceptive Method Choice**<br>http://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice<br>**2-item rules with 100% confidence** |
| • Number of Children ever born=Ten_to_Thirteen 28 ==> Wife Age=Adult 28 conf:(1) |
| **Dataset: Medical Data set**<br>**2-item rules with more than 90% confidence** |
| • Systolic Pressure=SP_low 4317 ==> Blood Pressure Medication=Bpmed_NO 4314 conf:(0.99) |
| • Body Mass Index=BMI_Normal 4405 ==> Blood Pressure Medication=Bpmed_NO 4379 conf:(0.99) |
| • Weight=Weight_average 4488 ==> Blood Pressure Medication=Bpmed_NO 4456 conf:(0.99) |
| • Current Smoker=CurrentSmoker_NO 3648 ==> Blood Pressure Medication=Bpmed_NO 3617 conf:(0.99) |

The 2-item rules presented in figure 3 were used to create 2-itemsets presented in figure 4. Attributes from each 2-item rule create one 2-itemset. Therefore, the number of 2-item rules should correspond to the number of 2-itemsets, but not in all cases. For example, in the Green Vehicle data set, we have two 2-item rules that create only one 2-itemset because the rules use the same pair of attributes.

**FIGURE 4: 2-ITEMSETS GENERATED**

**Dataset: Colleges**
ftp://85.158.30.137/lib.stat.cmu.edu/datasets/colleges/aaup.data
**2-itemsets extracted**

- Type, Number of Full Professors
- Average Salary Assistant Professors, Number of Associate Professors

**Dataset: Forest Fires**
http://archive.ics.uci.edu/ml/datasets/Forest+Fires
**2-itemsets extracted**

- RH, Rain
- Temp, Rain
- DMC, Rain
- DMC, Area
- Wind, ISI
- Wind, Rain
- Wind, Area
- X, Area
- X, Rain
- X, FFMC

**Dataset: Green Vehicle data**
https://explore.data.gov/Transportation/Green-Vehicle-Guide-Data-Downloads/9un4-5bz7
**2-itemsets extracted**

- Eng Displ, # Cyl

**Dataset: Contraceptive Method Choice**
http://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice
**2-itemsets extracted**

- Number of Children ever born, Wife Age

**Dataset: Medical Data set**
**2-itemsets extracted**

- Current Smoker, Blood Pressure Medication
- Body Mass Index, Blood Pressure Medication
- Systolic Pressure, Blood Pressure Medication
- Weight, Blood Pressure Medication

Figure 5 shows the resulting AFDs, the value and percentage of impurity, and the AFD's strength.

**FIGURE 5: RESULTING AFDS OF THE EXPERIMENTAL DATASETS**

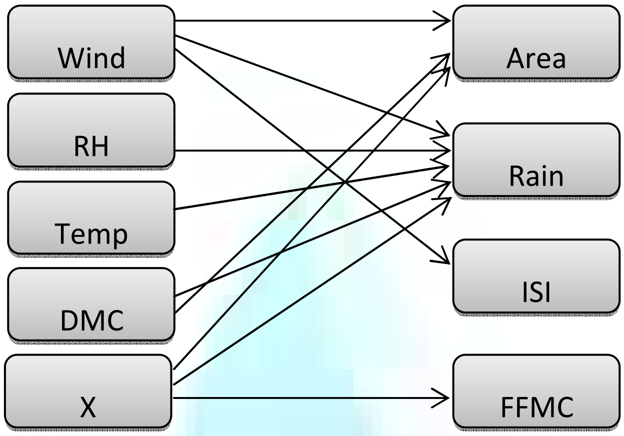
| | | | | |
|---|---|---|---|---|
| **Dataset: Colleges** ftp://85.158.30.137/lib.stat.cmu.edu/datasets/colleges/aaup.data | | | | |
| **Size** | **AFDs** | **Impurity (Tuples)** | **Impurity (Perc.)** | **AFD strength (Perc.)** |
| 1161 | Type ⇝ Number of Full Professors | 206 | 17.8 % | 82.2 % |
| 1161 | Average Salary Assistant Professor ⇝ Number of Associate Professors | 163 | 14 % | 86 % |
| **Dataset: Forest Fires** http://archive.ics.uci.edu/ml/datasets/Forest+Fires | | | | |
| **Size** | **AFDs** | **Impurity (Tuples)** | **Impurity (Perc.)** | **AFD strength (Perc.)** |
| 517 | Wind ⇝ Area | 3 | 0.5 % | 99.5 % |
| 517 | RH ⇝ Rain | 1 | 0.2 % | 99.8 % |
| 517 | Temperature ⇝ Rain | 1 | 0.2 % | 99.8 % |
| 517 | DMC ⇝ Rain | 1 | 0.2 % | 99.8 % |
| 517 | DMC ⇝ Area | 3 | 0.6 % | 99.4 % |
| 517 | Wind ⇝ ISI | 10 | 1.9 % | 98.1 % |
| 517 | Wind ⇝ Rain | 1 | 0.2 % | 99.8 % |
| 517 | X ⇝ Area | 3 | 0.6 % | 99.4 % |
| 517 | X ⇝ Rain | 1 | 0.2 % | 99.8 % |
| 517 | X ⇝ FFMC | 7 | 1.4 % | 98.6 % |
| **Dataset: Green Vehicle Data** https://explore.data.gov/Transportation/Green-Vehicle-Guide-Data-Downloads/9un4-5bz7 | | | | |
| **Size** | **AFDs** | **Impurity (Tuples)** | **Impurity (Perc.)** | **AFD strength (Perc.)** |
| 840 | Number of Cylinders ⇝ Engine Displacement | 41 | 5 % | 95 % |
| **Dataset: Contraceptive Method Choice** http://archive.ics.uci.edu/ml/datasets/Contraceptive+Method+Choice | | | | |
| **Size** | **AFDs** | **Impurity (Tuples)** | **Impurity (Perc.)** | **AFD strength (Perc.)** |
| 1473 | Wife Age ⇝ Number of Children Ever Born | 620 | 42 % | 58 % |
| **Dataset: Medical data set** | | | | |
| **Size** | **Relation** | **Impurity (Tuples)** | **Impurity (Perc.)** | **AFD strength (Perc.)** |
| 5945 | Current Smoker → Blood Pressure Medication | 44 | 0.7 % | 99.3 % |
| 5945 | Body Mass Index → Blood Pressure Medication | 44 | 0.7 % | 99.3 % |
| 5945 | Systolic Pressure → Blood Pressure Medication | 44 | 0.7 % | 99.3 % |
| 5945 | Weight → Blood Pressure Medication | 44 | 0.7 % | 99.3 % |

## 8. DISCUSSION OF THE RESULTS

In the first dataset, Colleges, there were two AFDs with relatively high strengths, 82.2% and 86% respectively.

In the second dataset, Forest Fires, there were quite a few AFDs with relatively high strengths. In fact, all of these strengths were over 99%, with just one at 98%, therefore most of these AFDs hold most of the time. Figure 6 visualizes the AFDs in the Forest fires data set. The attributes on the left would be the LHS of an association rule and the attributes on the right would be the RHS of the rule.

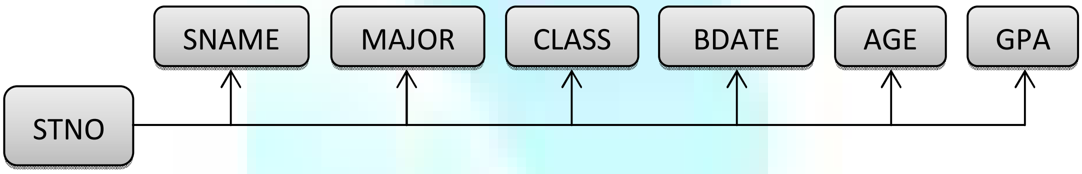
**FIGURE 6: AFDS DETERMINED FROM ASSOCIATION RULE MINING**




From this mapping we can see that AFDs determined from association rule mining would take a different nature. In FDs in relational database theory, a RHS cannot be mapped by more than one LHS. That is, a set of attributes in a row (RHS) is dependent on one key value (LHS), as shown in figure 7. This same set of attributes could not be dependent on more than one key value. From figure 7, sname, major, class, bdate, age, and GPA would be dependent on the student number (stno).

We can see from figure 6 that this is clearly not the case. Area, Rain, ISI, FFMC (the RHS) can be defined by more than one LHS. Hence, in AFDs determined from association rule mining, one LHS can map to more than one RHS. And, one RHS side can be mapped from more than one LHS. So, the nature of the AFDs determined from association rule mining is different from the definition of FDs used in relational database theory.

**FIGURE 7: FDS IN A RELATIONAL REPRESENTATION**




The third dataset also had an AFD with high strength (95%). Just as in the other datasets, there are many AFDs in this dataset too, however, we cannot discover them using our criteria (filtering only the rules with 100% confidence). If we lower the confidence threshold for association rule mining, we should be able to create much more 2-itemsets out of the rules with the confidence >50% and support >80%. In the fourth dataset, however, we found only one AFD with the strength 58%. So, this AFD would happen only about 58% of the time. In the last dataset, the Medical dataset, however, there were some really high AFDs.

## 9. CONCLUSION

From this study we can conclude that 100% confidence obtained from association rule mining does not necessarily mean a FD. To determine AFDs, in addition to the rules with 100% confidence, we have to determine what percentage of the data (the support) the rules with or close to 100% confidence cover. The higher the support (the closer the combined total of the support of the rules selected is to 100%) and the higher the confidence of the rules (and the closer this is to 100%), the higher strength of the AFD.

## REFERENCES

1. Agrawal, R. and Srikant, R. 1994. Fast algorithms for mining association rules in large databases. *Proceedings of the 20th VLDB Conference*, p. 487-499.
2. Agrawal, R., Imielinski, T. and Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Conference*, ACM Press, 207-216.
3. Alashqur, A. (2009). Expressing Database Functional Dependencies in Terms of Association Rules. *European Journal of Scientific Research,* 32(2), 260-267.
4. Bagui S., & Earp, R. (2012) *Database Design Using Entity-Relationship Diagrams*, 2nd edition, Francis and Taylor, Boca Raton, FL.
5. Berzal, F., Blanco, I., Sánchez, D., Serrano, J., Vila, M.A. (2005) A Definition For Fuzzy Approximate Dependencies. *Fuzzy Set Syst*. 149, 105–129.
6. Bra, P. D. and Paredaens, J. (1984). Horizontal Decompositions for Handling Exceptions to Functional Dependencies. In Gallaire, H., Minker, J. and Nicolas, J. –M. (eds), *Advances in Database Theory*, 2, 123-141.
7. Calero, J., Delgado, G., Sánchez-Marañón, M., Sánchez, D., Serrano, J., Vila, M.A. (2003). Helping User To Discover Association Rules. A case In Soil Color As Aggregation Of Other Soil Properties. *Proceedings of the 5th International Conference On Enterprise Information Systems, ICEIS'03*, 533–540.
8. Calero, J., Delgado, G., Sánchez-Marañón, M., Sánchez, D., Vila, M.A., Serrano, J. (2004a). An Experience In Management Of Imprecise Soil Databases By Means Of Fuzzy Association Rules and Fuzzy Approximate Dependencies. *Proceedings of the 6th International Conference On Enterprise Information Systems, ICEIS'04,* 138–146.
9. Calero, J., Delgado, G., Serrano, J., Sánchez, D., Vila, M.A. (2004b). Fuzzy Approximate Dependencies Over Imprecise Domains. An Example In Soil Data Management. *Proceedings of the IADIS International Conference On Applied Computing 2004,* 396–403.
10. Ceri, S., Fraternali, P., and Paraboschi, S. (2000) 'XML: Current Developments and Future Challenges for the Database Community', *Proceedings of the 7th International Conference on Extending Database Technology (EDBT),* Springer, LNCS 1777, Konstanz, March 2000.
11. Date, C. 2003. *An Introduction to Database Systems.* Addison Wesley.
12. Elmasri R. & Navathe, S. B. (2007) *Fundamentals of Database Systems*, Fifth Edition, Pearson Education, Boston, MA.
13. Giannella, C. and Robertson, E. (2004). On approximation measures for functional dependencies. *Inf. Syst.,* 29(6), 483-507.
14. Han, J. and Kamber, M. (2012). *Data Mining: Concepts and Techniques.* Morgan Kaufmann Publishers, USA.

15. Huhtala, Y., Karkkainen, J., Porkka, P., and Toivonen, H., (1999). Tane: An Efficient Algorithm for Discovering Functional and Approximate Dependencies. *The Computer Journal*, 42(2), 100-111.
16. Ilyas, I. F., Markl, V., Haas, P., Brown, P. and Aboulnaga, A. (2004). Cords: automatic discovery of correlations and soft functional dependencies. *SIGMOD 2004: Proceedings of the 2004 ACM SIGMOD International Conference of Management of data,* New York, NY, USA, 647-658.
17. Kalavagattu, A. K. (2008). *Mining Approximate Functional Dependencies as Condensed Representations of Association Rules,* Master's Thesis, Arizona State University.
18. Kappel, G., Kapsammer, E., and Retschitzegger, W. (2001a) 'Architectural Issues for Integrating XML and Relational Database Systems – The X-Ray Approach', *Proceedings of the Workshop on XML Technologies and Software Engineering,* Toronto.
19. Kappel, G., Kapsammer, E., and Retschitzegger, W. (2001b) 'XML and Relational Database Systems – A Comparison of Concepts', *Proceedings of the 2nd International Conference on Internet Computing (IC),* CSREA Press, Las Vegas, USA.
20. Kivinen, J. and Mannila, H. (1995). Approximate dependency inference from relations. *Theor. Comp. Sci*., 149, 129-149.
21. Park, J. S., Chen, M. –S., and Yu, P.S. 1995. An effective hash based algorithm for mining association rules. *Proceedings of the 1995 ACM SIGMOD Conference*, ACM Press, 175-186.
22. Sanchez, D., Jose, M. S., Blanco, I., Maria, J. M-B., Vial, M-A. (2008). Using Association Rules to Mine for Strong Approximate Dependencies. *Data Mining and Knowledge Discovery*, 16, 313-348.
23. Sánchez, D., Serrano, J., Vila, M., Aranda, V., Calero, J., Delgado, G. (2003). Using Data Mining Techniques To Analyze Correspondences Between User and Scientific Knowledge In An Agricultural Environment. In Piattini, M., Filipe, J., Braz, J. (eds) *Enterprise Information Systems IV*. Kluwer Academic Publishers, Hingham, MA, USA, 75–89.
24. Shanmugasundaram, J., Shekita, E., Barr, R., Carey, M., Lindsay, B., Pirahesh, H., and Reinwald, B. (2001). 'Efficiently publishing relational data as XML document', *VLDB Journal*, Vol. 19, No. 2-3, 133-154.
25. Tan, P-N, SteinBach, M., and Kumar, V. 2006. *Introduction to Data Mining*, Addison Wesley.
26. Wolf, G., Khatri, H., Chen, Y., and Kambhampati (2007). Quic: A System for Handling Imprecision and Incompleteness in Autonomous Databases (demo), *CIDR*, 263-268.

# REQUEST FOR FEEDBACK

**Dear Readers**

At the very outset, International Journal of Research in Computer Application & Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you tosupply your critical comments and suggestions about the material published in this issue as well as on the journal as a whole, on our E-mail**infoijrcm@gmail.com** for further improvements in the interest of research.

If youhave any queries please feel free to contact us on our E-mail **infoijrcm@gmail.com**.

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward an appropriate consideration.

With sincere regards

Thanking you profoundly

**Academically yours**

Sd/-
**Co-ordinator**

# DISCLAIMER

# ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

## Our Other Journals