

African Journal *of* Computing & ICTs

Volume 8. No. 3. September, 2015

African Journal *of* Computing & ICTs

Volume 8. No. 3. September, 2015

www.ajocict.net

All Rights Reserved
© 2015

A Journal of the Institute of Electrical & Electronics Engineers (IEEE)
Computer Chapter Nigeria Section

ISSN- 2006-1781

CONTENTS

1-8 Secret Communication Using Image Steganography

E.P. Musa & Simon Philip

Department of Computer Science, Ramat Polytechnic, Maiduguri, Nigeria

9-16 Achieving Modified Bloom's Taxonomy with the Zolla Class Management and Learning Software.

M.I. Osigbemeh, M. Eze., F. Imouokhome & O. Omaka

Dept of Electrical & Electronics Engineering, Federal University of Ndufu-Alike Ikwo, Ebonyi State, Nigeria

17-22 Towards a practical framework for successful change planning and implementation

E.O. Ademola

Principal Consultant, Power-Age Management Consul, Edenbridge Close, Orpington, Kent BR5 3SL United Kingdom

23-28 Multi Disease Analysis System Using Content-based Image Retrieval

M. Lavanya, S. Hari, D.R.N. Kumar & S.O. Manoj

Department of Computer Science and Engineering, Sri Ramakrishna Institute of Technology, Coimbatore

29-34 AudioKrypt: Application of Blowfish for audio file encryption

N.C. Woods & A.I. Bhadmus

Department of Computer Science, University of Ibadan, Ibadan, Nigeria.

35-42 Invoking .asm Codes in Object Oriented Programming Platform for Embedded Systems

F.U. Onu, C.G. Ezekwe, M. Uto & E. Ugwu

Department of Computer Science, Ebonyi State University, Abakiliki

43-50 A Hybrid Cryptosystem Using Elgamal Algorithm and Matrix Encryption

N.J. Mamvong, T. Aboiyar & T. Gbaden

ICT Directorate, University of Agriculture, Makurdi, Nigeria

51-58 An Analytical Study of Various Ad-hoc Network Routing Protocols Under Certain Parameters using Qualnet-7.1

S.K. Gupta, M. Yadav & R. K. Saket

Systems Engineering, Department of Electrical Engineering,

Indian Institute of Technology (Banaras Hindu University), Varanasi – 221005, (Uttar Pradesh), India

59-68 A Comparative Study of ICT Maturity Measurement Models

V.A. Olutayo & G.O. Ekuobase,

Department of Computer Science, Joseph Ayo Babalola University, Osun State, Nigeria

69-76 Comparative Analysis of Association Rule Mining Techniques for Monitoring Behavioural Patterns of Customers in a Grocery Store

I.A. Adeniji, Y.K. Saheed, T.O. Oladele & J.O. Braimah

Department of Computer Science, Al-Hikmah University, Ilorin, Nigeria

77-88 Broadband infrastructure using Fiber Optic: The state of things in a Developing Economy

F.M. Dahunsi

Dept. of Electrical & Electronics Engineering, The Federal University of Technology, Akure, Ondo State, Nigeria

89-100 Harmonic Mitigation on a 33kV Distribution Line using Passive Filters

L.M. Adesina & O.A. Fakolujo

Department of Inspection and Quality Assurance, Eko Electricity Distribution Plc (EKEDP), Lagos – Nigeria

101-110 On the Adoption and Implementation of a Cashless Economy in Nigeria

R.O. Akinyede & F.M. Dahunsi

Department of Computer Science, Federal University of Technology, Akure, Nigeria

111-128 4th Generation (4G) Technological Infrastructure and Enhanced Mobile Learning: An Effective Tool for Open and Distance Education

M.E. Augustus
 Rivers State Senior Secondary Schools Board, Port Harcourt, Rivers State, Nigeria

129-134 A Cognitive Load Theory-Based Framework for Designing an E-Learning Environment

O.C. Agbonifo & E.O. Ibam
 Department of Computer Science, Federal University of Technology, Akure, Nigeria

135-140 Development of a Yoruba Language E-Tutor for Windows Phone

I. A. Adeyanju, I. Omotosho & B.M Fatunde
 Department of Computer Engineering, Federal University Oye-Ekiti, Ekiti State, Nigeria

141-154 Fuzzy Analytical Hierarchical Process Model & ICT Maturity Model of SMEs for ICT Maturity Measurement of Nigerian Service Firms

G.O. Ekuobase & V.A. Olutayo
 Department of Computer Science, University of Benin, Benin City, Nigeria.

155-164 Problems Militating Against Computer Education and Its Effects on the Secondary School Curriculum in Umuahia North Local Government Area of Abia State, Nigeria.

E. Ogwo, C.E. Onwe & N.E. Maidoh
 Department of Computer Science, Akanu Ibiam Federal Polytechnic, Unwana, Afikpo, Nigeria.

165-174 Customer Churn Analysis In Banking Sector Using Data Mining Techniques

A. O. Oyeniyi & A.B. Adeyemo
 Department of Computer Science, University of Ibadan, Ibadan, Nigeria

175-180 Delay Prediction in Wireless Sensor Network Routing Using ART1 Neural Network

M. Mittal & K. Kumar
 Department of computer Science, Gurukul Kangri University, Haridwar, India

181-188 A Query Optimization Application in Database Management System Using Rough-Genetic Algorithm

P. Enyindah & P.O. Asagba
 Department of Computer Science, University of Port Harcourt, Rivers State, Nigeria

189-204 Performance improvement through interference mitigation techniques in Transmitted Reference UWB system used in WPAN overlay systems

B. Das, G.K. Mishra, S. Kanungo & B.K. Sahoo
 Department of Electronics & Telecomms Engineering, VSS University of Technology, Burla- 768018, Odisha, India

205-208 A Comparative Study of Consistency Theorems in Distributed Databases

C.K. Obasi, P.O. Asagba & A.I. Silas
 Department of Computer Science, University of Port Harcourt, Port Harcourt, Nigeria.

209-218 An Exploratory Study of the Risks and Motivating Factors of ICT Abuse among Students of Tertiary Institutions in Nigeria

O.T. Arogundade, F.S. Osholake, O.D. Aborisade & O. Olayiwola
 Department of Computer Science, Federal University of Agriculture, Abeokuta, Nigeria

219-227 CC3000 Wifi Base Configuration Of Gateway For Internet Accessibility in Monitoring Variables Via Wireless Sensor Systems

J. Agajo, I. C. Obiora-Dimson, O.B. Chukwuejekwu & N. Nathaniel
 Department of Computer Engineering, Federal University of Technology, Minna, Nigeria

229-238 Towards An Optimized Solution for Quality of Service Delivery

U.C. Ogude & L.N. Onyejebu
 Department of Computer Science, University of Port Harcourt, Port Harcourt, Nigeria

239-245 Performance Evaluation of Image Edge Detection Techniques in the Domain of Human Computer Interaction

Adigun, A.A.
 Department of Information and Communication Technology, Osun State University, Osogbo, Nigeria



Editorial Board

Editor-in-Chief

Prof. Dele Oluwade

Senior Member (IEEE) & Chair IEEE Nigeria – Computer Chapter.

Editorial Advisory Board

Prof. Gloria Chukwudebe - Senior Member & Chairman IEEE Nigeria Section

Engr. Tunde Salihu – Senior Member & Former Chairman IEEE Nigeria Section

Prof. Adenike Osofisan - University of Ibadan, Nigeria

Prof. Amos David – Universite Nancy2, France

Prof. Clement K. Dzidonu – President Accra Institute of Technology, Ghana

Prof. Adebayo Adeyemi – Vice Chancellor, Bells University, Nigeria

Prof. S.C. Chiemeké – University of Benin, Nigeria

Prof. Akaro Ibrahim Mainoma – DVC (Admin) Nasarawa State University, Nigeria

Dr. Richard Boateng – University of Ghana, Ghana.

Prof. Lynette Kvassny – Pennsylvania State University, USA

Prof. C.K. Ayo – Covenant University, Nigeria

Dr. Williams Obiozor – Bloomsburg University of Pennsylvania, USA

Prof Enoch Tangjong – University of Beau, Cameroon

Prof. Sulayman Sowe, United Nations University Institute of Advanced Studies, Japan

Dr. John Effah, University of Ghana Business School, Ghana

Mr. Colin Thakur - Durban University of Technology, South Africa

Mr. Adegoke, M.A. – Bells University of Technology, Ota, Nigeria

Managing/Production Editor

Dr. Longe Olumide PhD

Department of Computer Science

University of Ibadan, Ibadan, Nigeria

Comparative Analysis of Association Rule Mining Techniques for Monitoring Behavioural Patterns of Customers in a Grocery Store

I. A. Adeniji

Department of Computer Science,
University of Ilorin
Ilorin, Nigeria
herdeynig@gmail.com

Y. K. Saheed

Department of Computer Science,
Al-Hikmah University
Ilorin, Nigeria
yksaheed@alhikmah.edu.ng

T. O. Oladele

Department of Computer Science,
University of Ilorin
Ilorin, Nigeria
oladele.to@unilorin.edu.ng

J. O. Braimah

Department of Statistics,
Abdul-Raheem College of Advanced Studies
Igbaha, Kwara State, Nigeria
ojbraimah2012@gmail.com

ABSTRACT

The amount of data being generated and stored is growing exponentially, due to the continuing advances in computer technology. This presents tremendous opportunities for those who can unlock the information embedded within this data, but also introduces new challenges. The primary challenge is how to discover the hidden knowledge or pattern from the large sets of data in order to be able to make intelligent decisions that would shape the future of the store and also to determine the best tool to use in mining. This study proposed to use two different algorithms in association rule mining. The main motivation for carrying out this study is to compare the two algorithms and determine the time complexity of the two algorithms in mining association rules. In this paper, association rule techniques were compared and analyzed for monitoring behavioural pattern of customers in a grocery store. The FP-growth Algorithm and Apriori Algorithm were applied on sixty three (63) datasets from a grocery store. The time complexity of the two algorithms were considered and it was observed that FP-growth Algorithm is more efficient within the resource constraints than Apriori Algorithm. The comparison and analysis were implemented using Weka tool. The results revealed that FP-growth Algorithm is currently one of the fastest approaches for frequent data item set mining.

Keywords: Time Complexity, Fp-growth, Apriori, Association rule mining, Grocery store.

African Journal of Computing & ICT Reference Format:

I.A. Adeniji, Y.K. Saheed, T.O. Oladele & J.O. Braimah (2015): Comparative Analysis of Association Rule Mining Techniques for Monitoring Behavioural Patterns of Customers in a Grocery Store. Afr J. of Comp & ICTs. Vol 8, No. 3. Pp 69-76.

1. INTRODUCTION

Across a wide variety of fields, data are being collected and accumulated at a dramatic pace. There is an urgent need for a new generation of computational techniques and tools to assist humans in extracting useful information (knowledge) from the rapidly growing volumes of data. These techniques and tools are the subject of the emerging field of Knowledge Discovery in Databases (KDD) [1].

The amount of data being generated and stored is growing exponentially, due to in large part to the continuing advances in computer technology. This presents tremendous opportunities for those who can unlock the information embedded within this data, but also introduces new challenges [9]. Worldwide enormous amounts of data is being produced and stored at an amazing pace.

Meanwhile, the steadily dropping prices for storage devices and the increasing capacity of each unit have made it into a simple choice to postpone the decision on what to do with all this data being produced. With hardware now having these sophistication, capability and capacity to store almost unlimited amount of data, we are rich in data but poor in information, the biggest challenge now confronting data analysts is how to use this large pool or quantity data to gain business advantage. Through knowledge discoveries in databases, interesting patterns could be discovered and analyzed from different dimensions to gain useful insight into the vast amount of data. Companies can use these discoveries to gain a competitive edge or advantage over their competitors, and thereby creating an opportunity to make higher profit. With the advancements in information technology and database technologies, data warehouses have emerged as a powerful tool to build decision support systems (DSS). Along with data mining, the power of data warehouses to support organizational decision making has increased vastly. The data warehouse provides opportunities for performing data mining tasks.

Data mining tools actively search for previously unrecognized patterns in the data. The patterns identified by these tools gives data analyst useful and unexpected insights that can be more carefully investigated subsequently, perhaps using other decision support tools. It is therefore a general expectation that data mining tools should be able to identify or find these patterns in the data with minimal user input. Data mining is related to the sub-area of statistics called exploratory data analysis, which has similar goals and relies on statistical measures. It is also closely related to the sub-areas of artificial intelligence called knowledge discovery and machine learning. The last important distinguishing characteristic of data mining is that the volume of data is very large.

One of the major challenges faced by most grocery stores is the issue of business intelligence as a result of their movement from manual storage of their records to automated storage. These challenges became more pronounced with the great rate at which data continue to grow. In this wise the primary challenge is how to discover the hidden knowledge or pattern from the large sets of data in order to be able to make intelligent decisions that would shape the future of the store and also to determine the best tool to use in mining. It has been observed over time that customer buying pattern in a shopping trip can lead to multiple product usually purchased from multiple categories. Efforts have been made by researchers to analyze the multi-category purchase information in order to plan marketing activities accordingly so as to maximize profit.

The main aim of this study is to carry out a comparative analysis of association rule mining techniques for monitoring behavioural patterns of customers in a grocery store and the objectives are to:

1. Investigate the two techniques (algorithms) of association rule mining by applying them to actual data.
2. study how data mining technique works while mining the data from data warehouse.
3. determine the data mining technique that is most effective, efficient, and accurate by comparing their results.

2. RELATED WORKS

Association rule mining represents a data mining technique and its goal is to find interesting association or correlation relationships among a large set of data items. With massive amounts of data continuously being collected and stored in databases, many companies are becoming interested in mining association rules from their databases to increase their profits. This process analyzes customer buying habits by finding associations between the different items that customers place in their "shopping baskets" [8].

The abundance of data generates the appearance of a new field named data mining. Data collected in large databases become raw material for these knowledge discovery techniques and mining tools for "gold" were necessary. The current expert system technologies, which typically rely on users or domain experts to manually, input knowledge into knowledge bases. This procedure contains errors, and it is extremely time-consuming and costly. Data mining tools which perform data analysis may uncover important data patterns, contributing greatly to business strategies, knowledge bases, and scientific and medical research [6].

Data is one of the valuable resources for organization, and database management systems are gradually becoming ubiquitous in many small and medium scale companies. Although, some of the benefits of database management systems have been explored, however, many companies have not been able to exploit the advantages of gaining business intelligence from their databases. This has led to inadequate business decision making based on the data contained in the databases. In this study, association rule mining also known as market basket analysis using Apriori algorithm is presented for extracting valuable knowledge embedded in the database of a supermarket [2].

Data stream is a continuous, real time, ordered sequence of items. In data stream, data arrives endlessly and the volume of data can be potentially infinite. In recent years advances in hardware and software technologies have resulted in automated storage of data from a variety of process. Data mining techniques are applied in data streams to find out the significant knowledge.

The term data mining refers, to find relevant and useful information from large database. Some of the important techniques in data mining are association rule, classification, clustering, frequent episodes, and deviation detection. Frequent pattern mining is used to find important frequent patterns from the large dataset. Click stream analysis, market basket analysis, web link enquiry, genome study, network monitoring and medicine designing are some of the important areas where frequent pattern mining is used. Most commonly used frequent pattern mining algorithms are Apriori, partition algorithm, pincer- search algorithm, fp-growth algorithm, dynamic item set counting algorithm and so on.

Association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. With wide applications of computers and automated data collection tools, massive amounts of transaction data have been collected and stored in databases. Discovery of interesting association relationships among huge amounts of data will help marketing, decision making, and business management. Therefore, mining association rules from large data sets has been a focused topic in recent research into knowledge discovery in databases [3].

One of the currently fastest and most popular algorithms for frequent item set mining is the FP-growth algorithm [4]. It is based on a prefix tree representation of the given database of transactions (called an FP-tree), which can save considerable amounts of memory for storing the transactions. The basic idea of the FP-growth algorithm can be described as a recursive elimination scheme: in a pre-processing step delete all items from the transactions that are not frequent individually, that is, do not appear in a user-specified minimum number of transactions. Data mining is a new concept and an exploration and analysis of large data sets, in order to discover meaningful patterns and rules. Many organizations are now using the data mining techniques to find out meaningful patterns from the database.

3. METHODOLOGY

Association rule mining has proven to be an effective method in mining market basket analysis to determining the relationship between products bought by individual (transaction done) in grocery stores. To establish the effectiveness of association rule mining we have proposed to use two different algorithms in association rule mining. The main motivation for carrying out this study is to compare the two algorithms and determine the time complexity of the two algorithms in mining association rules.

According to [5], an association rule is an implication expression of the Form $X \Rightarrow Y$, where XI , and YI , and X and Y are disjoint item sets, i.e. $X \cap Y = \phi$. X is called antecedent while Y is called consequent, the rule means X implies Y . The strength of an association rule can be measured in terms of its support and confidence. The rule $X \Rightarrow Y$ holds in the transaction set D with confidence c and support s , if $c\%$ of the transactions in D that contains X also contains Y , and $s\%$ of

transactions in D contains XY . Both the antecedent and the consequent of the rule could have more than one Item. The formal definitions of these two metrics are:

Support(s) of an association rule is defined as the percentage/fraction that contain $X \Rightarrow Y$ to the number of records in the database.

$$\text{Support, } s(X \Rightarrow Y) = \frac{\sum (X \cup Y)}{N} \dots \dots \dots (1)$$

Confidence of an association rule is defined as the percentage/fraction of the number of transactions that contain $X \Rightarrow Y$ to the total number of records that contain X , where if the percentage exceeds the threshold of confidence an interesting association rule $X \Rightarrow Y$ can be generated.

$$\text{Confidence, } c(X \Rightarrow Y) = \frac{\sum (X \cup Y)}{\sum X} \dots \dots \dots (2)$$

Association rule mining is to find out association rules that satisfy the predefined minimum support and confidence from a given database [6]. The problem is usually decomposed into two sub-problems. One is to find those item sets whose occurrences exceed a predefined threshold in the database; those item sets are called frequent or large item sets. The second problem is to generate association rules from those large item sets with the constraints of minimal confidence. Suppose one of the large item sets is L_k , $L_k = \{I_1, I_2, I_{k-1}, I_k\}$, association rules with this item sets are generated in the following way: the first rule is $\{I_1, I_2, \dots, I_{k-1}\} = \{I_k\}$, by checking the confidence this rule can be determined as interesting or not. Then other rule are generated by deleting the last items in the antecedent and inserting it to the consequent, further the confidences of the new rules are checked to determine the interestingness of them. Those processes iterated until the antecedent becomes empty. Since the second sub-problem is quite straight forward, most of the researches focus on the first sub-problem.

Association rule mining process could be divided into two main phases to enhance the implementation of the algorithm. The phases are:

1. **Frequent Item Generation:** This is to find all the item sets that satisfy the minimum support threshold. The item sets are called frequent item sets.
2. **Rule Generation:** This is to extract all the high confidence rules from the frequent item sets found in the first step. These rules are called strong rules. According to [7], association analysis is based on the rule that is specified in the form: If item A is part of an event then $X\%$ of the time (confidence factor) item B is part of the same event.

This study is based on two different algorithm use in mining association rule. The algorithms are apriori algorithm and frequent pattern growth (fp-growth) algorithm.

Apriori algorithm

Apriori algorithm [3] is the most fundamental and important algorithm for mining frequent things. Apriori is used to find all frequent things in a given database, that is, it provides a way to find association rules on large scale. The importance of Apriori algorithm is to produce multiple searches over the database. It uses a repetitive approach called a breadth-first search (level-wise search) through the search room, where K -things are used to explain $(K+1)$ -things. Apriori is designed to operate on databases containing transactions. Each transaction is seen as a collection of items or data known as item sets. Given a threshold C , (confidence) the Apriori algorithm identifies the item sets which are subsets of at least C transactions in the database [6].

Apriori Principle:

- i. If an item set is frequent, then all of its subsets must also be frequent, or
- ii. If an item set is infrequent then all its supersets must also be infrequent

Apriori principle holds due to the following property of the support measure:

$$\forall X, Y: (X \subseteq Y) \Rightarrow s(X) \geq s(Y)$$

- i. Support of an item set never exceeds the support of its subsets.
- ii. This is known as the anti-monotone property of support.

General Process

Association rule generation is usually split up into two separate steps:

1. First, minimum support is applied to find all frequent item sets in a database.
2. Second, these frequent item sets and the minimum confidence constraint are used to form rules.

While the second step is straight forward, the first step needs more attention. Finding all frequent item sets in a database is difficult since it involves searching all possible item sets (item combinations). The set of possible item sets is the power set over I and has size $2^n - 1$ (excluding the empty set which is not a valid item set). Although the size of the power set grows exponentially in the number of items n in I , efficient search is possible using the downward-closure property of support (also called anti-monotonicity) which guarantees that for a frequent item set, all its subsets are also frequent and thus for an infrequent item set, all its supersets must also be infrequent. Exploiting this property, efficient algorithms (e.g., Apriori and Eclat) can find all frequent item sets.

Apriori Algorithm Pseudocode

```

procedure Apriori ( $T$ ,  $\text{minSupport}$ ) { //  $T$  is the database and
 $\text{minSupport}$  is the minimum support
 $L_1 = \{\text{frequent items}\}$ ;
For ( $k = 2$ ;  $L_{k-1} \neq \emptyset$ ;  $k++$ ) {
 $C_k = \text{candidates generated from } L_{k-1}$ 
//that is cartesian product  $L_{k-1} \times L_{k-1}$  and eliminating any  $k-1$ 
size itemset that is not
//frequent
for each transaction  $t$  in database do{
#increment the count of all candidates in  $C_k$  that are contained
in  $t$ 
 $L_k = \text{candidates in } C_k \text{ with } \text{minSupport}$ 
} //end for each
} //end for
Return  $\bigcup_k L_k$ ;
}

```

As is common in association rule mining, given a set of item sets (for instance, sets of retail transactions, each listing individual items purchased), the algorithm attempts to find subsets which are common to at least a minimum number C of the item sets. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found.

3.1 Frequent Item Generation

Apriori algorithm for frequent item generation is a two-step process of join and prune as follows:

JOIN STEP: Let the set of frequent item sets of size k be f_k and their candidates be C_k . Apriori first scans the database and searches for frequent item sets of size 1 by accumulating the count for each item and collecting those that satisfy the minimum support requirement. It then iterates on the following three steps and extracts all the frequent item sets.

1. Generate C_{k+1} , candidates of frequent item sets of size $k+1$, from the frequent item sets of size k .
2. Scan the database and calculate the support of each candidate of frequent item sets.
3. Add those item sets that satisfies the minimum support requirement to F_{k+1} .

PRUNE STEP: Check if all the item sets of size k in f_{k+1} are frequent and generate C_{k+1} by removing those that do not pass this requirement from f_{k+1} . This is because any subset of size k of C_{k+1} that is not frequent cannot be a subset of a frequent item sets of size $k + 1$. This step removes a candidate if it contains any infrequent k -pattern.

Method (level-wise search):

i. Initially, scan DB once to get frequent 1-itemset.

ii. For each level k :

•Generate length $(k+1)$ candidates from length k frequent patterns

•Scan DB and remove the infrequent candidates

–Terminate when no candidate set can be generated

Table 1: Transactional Table

Transaction id	Items
1	Apple, chocolate
2	Apple, Biscuit, Chocolate
3	Biscuit, Black Jam, Cream Soda
4	Cream Soda, Apple, Chocolate, Blue Band
5	Apple, Cream soda

Step 1 scan the database and generate the frequent item set as shown in Table 1 above

Table 2 : C1 (candidate table)

Itemsets	Frequency or support
Apple	4
Chocolate	3
Biscuit	2
Black Jam	1
Cream soda	3
Blueband	1

From table 2 assuming the set $\min_support$ is 2, apriori would generate, L1 (large dataset of frequent item sets) that meet the set $\min_support$ and prune out the items that does not meet the $\min_support$ as shown in table 3 below

Table 3: L1 (large dataset)

Itemsets	Support
Apple	4
Chocolate	3
Biscuit	2
Cream soda	3

From the large dataset, the apriori generate another candidate set C2 and its $\min_support$ as shown in table 4 below.

Table 4: L2 (large dataset)

Item sets	Frequency
Apple , Chocolate	3
Apple , Biscuit	1
Apple, Cream soda	1
Chocolate, Biscuit	0
Chocolate, Cream soda	0
Biscuit, Cream soda	0

Since no more candidate set can be generated again from the table, Terminate when no candidate set can be generated

3.2 Rule Generation

From the above frequent item generated and based on the $\min_support$. We now proceed to determine if the 2-itemset {Apple, Chocolate} lead to association rules with required confidence of 75%. Every 2-itemset {X, Y} can lead to two rules $X \Rightarrow Y$ and $Y \Rightarrow X$, if both satisfy the required confidence. As defined earlier, confidence of $X \Rightarrow Y$ is given by the support for X and Y together divided by the support for X.

The study therefore have two possible rules and their confidence as follows:

Apple $\Rightarrow$ Chocolate with confidence of $\frac{3}{4} = 75\%$

Chocolate $\Rightarrow$ Apple with confidence of $\frac{3}{3} = 100\%$

Therefore, both rules have confidence of minimum of 75% required and qualify. Rules that have more than the user-specified minimum confidence are called confident.

3.3 FP-Growth Algorithm

FP-GROWTH, takes a clear different approach to discovering or finding frequent item sets. The algorithm does not align to the GENERATE AND TEST model of the APRIORI, instead, it forms a compact datasets using a data structure called FP-TREE and then extract the frequent item sets directly from this structure. It compresses large database into a compact Frequent-Pattern tree (FP-tree) structure highly condensed, but complete for frequent pattern mining and avoid costly database scans. The divide-and-conquer methodology is used by FP-GROWTH in arriving at its result. The FP-growth algorithm mining frequent patterns without candidate generation was proposed by [4].

Fp-growth Algorithm Pseudocode

Input: FP-tree constructed based on Algorithm 1, using DB and a minimum support threshold ϵ .

Output: The complete set of frequent patterns.

Method: Call FP-growth (FP-tree; null), which is implemented as follows.

Procedure FP-growth (Tree; α)

```

{
  IF Tree contains a single path P
  THEN FOR EACH combination (denoted as  $\beta$ ) of the nodes in the path P DO
    generate pattern  $\beta \cup \alpha$  with support = minimum support of nodes in  $\beta$ ;
  ELSE FOR EACH  $a_i$  in the header of Tree DO {
    generate pattern  $\beta = a_i \cup \alpha$  with support =  $a_i$ .support;
    Construct  $\beta$ 's conditional pattern base and then  $\beta$ 's conditional FP-tree  $Tree_{\beta}$ ;
    IF  $Tree_{\beta} \neq \emptyset$ 
    THEN Call FP-growth ( $Tree_{\beta}$ ;  $\beta$ )
  }
}

```

3.3.1 Frequent Item Generation

FP-growth algorithm allows frequent item sets generation without candidate item sets generation. Fp-growth uses a two-step approach known as:

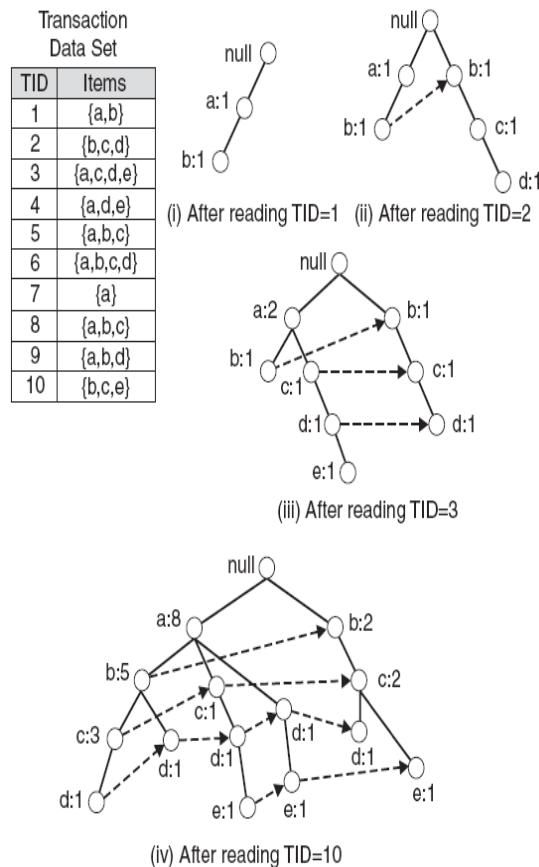


Figure 1: FP tree construction

(i) Building the Fp tree

(ii) Extraction of the frequent item sets from the tree.

Building The FP-Tree:

The Fp tree is constructed with two passes over the database

Pass 1:

- Scan data and find support for each item.
- Discard infrequent items.
- Sort frequent items in decreasing order based on their support.
- For our example: a; b; c; d; e Use this order when building the FP-Tree, so common prefixes can be shared.

Pass 2: construct the FP-Tree

- Read transaction 1: a; b, then create 2 nodes a and b and the path null ! a ! b. Set counts of a and b to 1.
- Read transaction 2: b; c; d, then create 3 nodes for b, c and d and the path null ! b ! c ! d. Set counts to 1. Note that, although transaction 1 and 2 share b, the paths are disjoint as they don't share a common prefix. Add the link between the b's.
- Read transaction 3: a; c; d; e. It shares common prefix item a with transaction 1 so the path for transaction 1 and 3 will overlap and the frequency count for node a will be incremented by 1. Add links between the c's and d's.
- Continue until all transactions are mapped to a path in the FP-tree.

Extraction Of The Frequent Itemsets From The Tree

FP-Growth extracts frequent item sets from the FP-tree.

Bottom-up algorithm from the leaves towards the root

Divide and conquer: First look for frequent item sets ending in e, then de, etc. . . then d, then cd, etc. First, extract prefix path sub-trees ending in an item(set). (hint: using the linked lists)

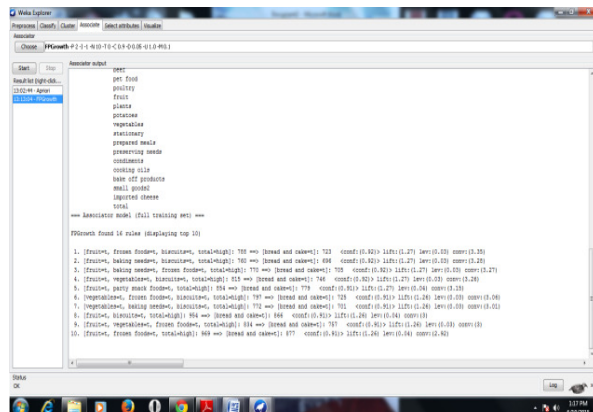


Figure 3: This environment depicts the result of FP growth implemented with data set. It also shows the relationship of Fp growth Algorithm among data items.

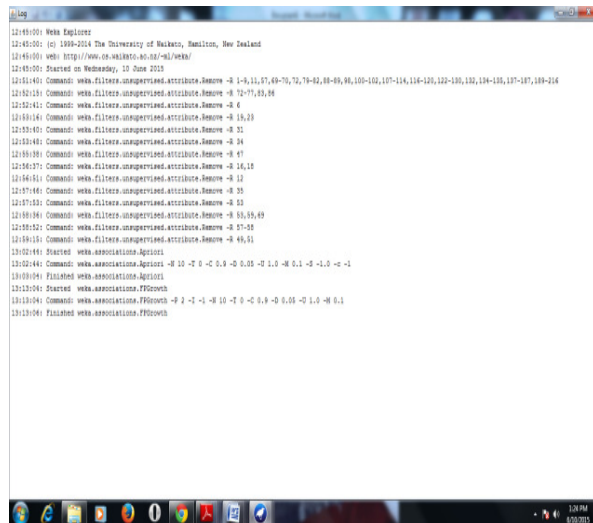


Figure 4: This environment revealed the time taken for both Fp-growth Algorithm and Apriori Algorithm and it was observed that Fp-growth Algorithm is the fastest approach to frequent data item set mining.

5. CONCLUSION

This study focused on comparison and analysis of association rule mining for monitoring behavioural pattern of customer in a grocery store. Two association rule techniques were used, FP-growth Algorithm and Apriori algorithm were implemented on sixty three (63) data sets in a Weka tool. It was observed that FP-growth Algorithm took faster time of execution as compared to Apriori Algorithm. This indicates that the amount of time it requires to run to completion is less than the amount of time required in Apriori algorithm. The study supports the fact that FP-growth algorithm is the fastest approach for frequent item set mining.

Future research will involve comparing the two Algorithms and their time complexity and further refining the method presented. This study would help in generating the efficient Algorithm for frequent data item set mining. Other data mining techniques and tools can also be explored and comparing their results would also be a promising area in future works.

REFERENCES

- [1] F'ayyad, U.M., Piatetsky-Shapiro, G., and Smyth, P. (1996). From Data Mining to Knowledge Discovery: An Overview, in AI (DDM, AAAI/MIT Press, pp. 1-30.
- [2] Abdulsalam S.O., Adewole, K.S., Akintola, A.G., Hambali, M.A.(2014). Data Mining in Market Basket Transaction: An Association Rule Mining Approach. International Journal of Applied Information Systems (IJ AIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 7– No.10, October 2014
- [3] Agrawal, R., Imielinski, T., and Swami, A. N. (1993). Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data
- [4] Han, J., Pei, H., and Yin, Y. (2000). Mining Frequent Patterns without Candidate Generation. In: Proc. Conf. on the Management of Data (SIGMOD'00, Dallas.
- [5] Oladipupo, O. O. and Oyelade, O. J. (2010). Knowledge Discovery from Students' Result Repository: Association Rule Mining Approach, International Journal of Computer Science and Security, 4(2), pp.199-207.
- [6] Agrawal, R., and Srikant, R. (1994) "Fast Algorithms for Mining Association Rules", Proceedings of the 20th International Conference on Very Large Data Bases, 1994, pp. 487-499.
- [7] Larissa, T. M. (2003). Introduction to Data Mining, London: Oxford University Press.
- [8] Irina Tudor (2008). Association Rule mining as a Data Mining technique. Buletinul Universitatii Petrol-Gaze din Ploiesti. Volume LX No.1/2008 49-56 Seria Mathematica-Informatica-Fizica
- [9] Weiss, Gary M., Davison and Brain D. (2010). Data Mining, Retrieved from the internet on 10-08-2014