Bachelor of Science in Computer Science & Engineering

A Comparative Analysis of Different Association Rule Mining Algorithms

by

Jaher Hassan Chowdhury

ID: 1504076

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)

Chattogram-4349, Bangladesh.

April, 2021

A Comparative Analysis of Different Association Rule Mining Algorithms

Submitted in partial fulfilment of the requirements for
Degree of Bachelor of Science
in Computer Science & Engineering

by

Jaher Hassan Chowdhury
ID: 1504076

Supervised by

Prof. Dr. Mohammad Shamsul Arefin
Professor
Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)
Chattogram-4349, Bangladesh.

The thesis titled '**A Comparative Analysis of Different Association Rule Mining Algorithms'** submitted by ID: 1504076, Session 2019-2020 has been accepted as satisfactory in fulfilment of the requirement for the degree of Bachelor of Science in Computer Science & Engineering to be awarded by the Chittagong University of Engineering & Technology (CUET).

Board of Examiners

____________________________________ Chairman

Prof. Dr. Mohammad Shamsul Arefin
Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

____________________________________ Member (Ex-Officio)

Dr. Asaduzzaman
Professor & Head
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

____________________________________ Member (External)

Dr. Mohammed Moshiul Hoque
Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

Declaration of Originality

This is to certify that I am the sole author of this thesis and that neither any part of this thesis nor the whole of the thesis has been submitted for a degree to any other institution.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. I am also aware that if any infringement of anyone's copyright is found, whether intentional or otherwise, I may be subject to legal and disciplinary action determined by Dept. of CSE, CUET.

I hereby assign every rights in the copyright of this thesis work to Dept. of CSE, CUET, who shall be the owner of the copyright of this work and any reproduction or use in any form or by any means whatsoever is prohibited without the consent of Dept. of CSE, CUET.

Signature of the candidate

Date:

Acknowledgements

The success and final result of this thesis involved a great deal of support and assistance from many people, and I consider myself incredibly fortunate to have received it all the way through my thesis. Professionally and socially, it has been a rewarding experience. Much of what I've accomplished has been possible only because of such oversight and assistance.Following that, I'd like to express my heartfelt gratitude to Prof. Dr. Mohammad Shamsul Arefin, Department of Computer Science Engineering, Chittagong University of Engineering Technology (CUET), for his guidance, motivation, and unwavering support during my thesis work. I appreciate his many probing questions, which force me to consider things from various angles throughout the entire time.

I thank Dr. Asaduzzaman, Head Professor, Department of Computer Science Engineering, Chittagong University of Engineering Technology (CUET) who took close interest in my work and directed me by providing the necessary details. I am grateful for all of the department's teaching workers' relentless motivation, assistance, and advice.

Finally, I want to thank my father and mother for their unconditional love, support, encouragement, and contribution all throughout my life and academic career in every aspect along the years.

Abstract

Data mining is an important aspect of developing association rules for a large number of itemsets. Association rule mining [ARM] is a data-processing technique that has two sub-processes.The first method is known as finding frequent itemsets, and the second method is known as association rule mining.The concepts relating to the use of frequent itemsets are extracted during this sub process.Many algorithms for locating frequent itemsets and association rules have been created by researchers. This paper provides a comprehensive analysis and evaluation of various Association Rule mining algorithms. The merits, demerits, data support of the ARM algorithms were also compared in this paper.

Keywords: Data Mining ,KDD, Apriori, FP-Growth, Fp-Growth with lift,RP-Growth,FP-Close,Indirect Association, MNR association rule, Sporadic Association rule,TOP-K Association rule,IGB Association Rule, Comparison of association rules.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Introduction

Normally data mining is the process of analyzing data and summarize the data into a valuable information. It is an analyzing tool that can allow a user to analyze the data and summarize the relationship among those data. The problem of finding the relationship amongst the data is first found by Agarwall.[1] The solution of this problem helped the researcher in market basket analysis. It also helps them to enhance the earning and optimized storage. The main task for analysing the algorithms is to choose the correct datasets in order to compare them and plot the data in a correct way.Though there are a lot of problems.

1.2 Design Overview

Data mining techniques created a partition of useful artificial intelligence. Throughout the primary periods, the primary contributions to computer systems have been directed toward the examination of new technologies for Network, established instruction. The rapid growth of databases necessitates the development of technologies that process data and information logically.Data mining techniques have become an increasingly important area of research.So, the best technique is to find out which algorithm should be used in a specific dataset to derive useful information. That's why comparison of association rule mining algorithm is needed.

The main steps of Comparing association rule mining algorithms are :

1. The first thing is to collect the datasets for applying the algorithms. Because, we can't apply association rules in every datasets and some of the dataset is hard

to find.

2. Pre-processing of dataset is important. Some datasets have missing values. For this, we have to pre-process the datasets for applying the algorithms.

3. Then we have to implement our algorithms with the help of suitable programming language. So, for this thesis paper we have chosen Java as our programming language.

4. Then after applying the algorithm in suitable datasets, we have to compare the results. Then after comparing amongst the results we will make a decision.

1.3 Difficulties

Nowadays, data mining and information disclosure are critical innovations for researchers and businesses across various industries. As data mining developed into a setup and gained control, new data mining challenges needed to be addressed. So, Comparing association rule mining algorithms have some same challenges as data mining. The challenges are :

01.Noisy and incomplete data : This information about present reality is noisy, incomplete, and heterogeneous. Massive amounts of data will be unreliable or inaccurate on a regular basis. These issues may arise as a result of human error, blunders, or errors in the instruments used to collect the data.

02.Distributed Data : True data is typically stored in stages during distributed processing. It is quite possible that it is stored on the internet, on individual systems, or even in databases. It is fundamentally difficult to move all data to a unified data archive, primarily for technical and organizational reasons.

03. Complex Data : True data is truly diverse, and it may very well be media data, such as natural language text, time series, spatial data, temporal data, complex data, audio or video, or images. It is extremely difficult to manage these disparate types of data and focus on the critical information. Frequently, new apparatuses and systems would be required to separate critical information.

04. Data Visualization : Data visualization is a critical cycle in data mining because it is the first interaction that demonstrates the output to the client in a respectable manner. The information extracted should convey the specific significance of what it truly intends to convey. However, it is frequently exceedingly difficult to communicate information to the end user in a precise and straightforward manner. To make it fruitful, the output information and input data should be extremely effective, successful, and complex data perception methods should be used.

05. Efficieny and Improvement of data mining algorithms : The Data Mining algorithm must be scalable and efficient in order to extract information from massive amounts of data contained in the data set. The distribution and development of parallel data mining algorithms must be motivated by factors such as the difficulty of data mining approaches, the enormous size of the database, and the entire data flow.

1.4 Applications

There is a wide range of areas where we can use the association rule mining techniques. Association rule mining algorithms can be used for a wide range of applications, such as -

1. Most commonly used for the market basket analysis, which is very important for the financial world.

2.The application of association rules in medical diagnosis can aid physicians in curing patients.

3.The application of data mining techniques to census data, and more broadly to official data, has enormous potential for advancing sound public policy and ensuring a democratic society's effective functioning.

4.The association rules can be used for Websites that have to work with tons of data.

1.5 Motivation

Market basket is a typical and commonly used example of the association rule mining. In supermarkets, for example, data is obtained using bar-code scanners. A significant number of transaction documents are composed of 'business basket' databases. Every record lists all products the consumer has purchased in a single sales account. Every record lists all products the consumer has purchased in a single sale. A manager will be curious to know if such classes of products are purchased regularly together. These data may be used to change store layouts (optimally position products with respect to each other), to cross-sell, to promotions, to catalog design and to define consumer segments based on trends of buying. Every product now days comes with bar code. The software which supports these Bar code-based purchasing / ordering systems generate large quantities of sales data, usually collected in 'baskets' (records in which products purchased by the customer at the time are grouped). The business world rapidly accepted this data as having enormous potential commercial value. Commercial companies are especially interested in finding association rules that define trends of transactions, since the existence of one object in a basket indicates the presence of one or more additional items. The result of this "Market Basket Analysis" can then be used to suggest product combinations for special promotions or sales, develop a more efficient store layout and provide insight into brand loyalty and co-branding.

Applying association rules in the area of medical care may be used to support doctors in treating patients. The general problem of inducing accurate diagnostic rules is difficult, because theoretically no method of inference alone can guarantee the correctness of induced hypotheses[2]. Diagnosis is not a straightforward task in practice, since it requires poor diagnosis testing and noise exposure in outstanding preparation. This can result in hypotheses with unsatisfactory accuracy of prediction that are too inaccurate for important medical applications[3]. Serban[2] has suggested a methodology focused on the laws of relational interaction and supervised methods of learning.

Proteins are essential constituents of any organism's cellular machinery. DNA technologies have provided instruments for fast DNA determination. By definition, the amino acid composed of structural gene proteins [4]. Protein has a unique 3-dimensional structure, which depends on the sequence of amino acids; minor change in sequence may alter protein function. A topic of great anxiety was the strong reliance of protein functioning on its amino acid sequence. Lots of study has gone into understanding the protein structure and nature; a number of questions still remain to be satisfactorily understood. The amino acid sequences of proteins are now widely assumed not to be random. A group of researchers deciphered the existence of associations between various amino acids in a protein.[5] Such association rules are useful in order to improve our understanding of protein structure and have the ability to provide information as to the global interactions between certain complex sets of amino acids in proteins. Awareness of these association rules is highly desirable for artificial protein synthesis.

Censuses make available to both researchers and the general public a wide array of general statistical knowledge about society[6]. Population and economic census details can be expected in the planning of public services (education , health, transportation, funds) as well as in public industry (for setting up new factories, shopping malls or banks and even marketing of specific products). Data mining techniques have great potential to promote good public policies and to promote the successful functioning of a democratic society in general, through census data and more general official data.[7]. On the other hand, it is not undemanding and involves a rigorous methodological analysis, which is still in the preliminary stages.

Customer Relationship Management (CRM), in which banks aim to recognize the desires of various customer groups, goods and services tailored to their tastes to promote synergy between credit card customers and the bank, has become a subject of great interest[8].Wang primarily explains how to integrate data mining into the marketing information management process.[9]

For these things i strongly think that the comparative association rule mining will play a vital role in our real life.

1.6 Contribution of the thesis

Thesis or research work is carried out to accomplish a specific set of objectives, such as defining a new methodology or improving an existing one. The thesis is a comparison-based work. In this thesis, we have tried to improve the results than the result of the other thesis. Moreover, we tried to work with many datasets and algorithms that are never used to compare the association rule mining algorithms. The primary contribution of this thesis is the following:-

01.To implement ten different association rule mining algorithms.

02.To implement the algorithms in five different datasets.

03.To compare these algorithms on the basis of association rules generated.

04.To compare these algorithms on the basis of their data structure and how they behave on different datasets.

1.7 Thesis Organization

This report is divided into five sections. Chapter one contains some introductory texts on comparative analysis of association rule mining algorithms and some challenges associated with implementing our work, our motivation for work, and our work's objectives. Chapter two discusses previously implemented works, their limitations, and their impact on our work. Additionally, it details the theories and algorithms that underpin our system.The proposed methodology for comparing association rule mining algorithms is described in Chapter 3. Chapter 4 describes the working dataset and conducts an analysis of the proposed framework's performance metrics.Chapter 5 summarizes this thesis work and makes some future recommendations.

1.8 Conclusion

The purpose of this chapter is to provide an overview of the comparative analysis of association rule mining algorithms. Along with the difficulties, this chapter summarizes the comparative analysis of association rule mining algorithms. Additionally, the motivation for this work and contributions is stated here. The following chapter will discuss the problem's history and current state.

Chapter 2

Literature Review

2.1 Introduction

So, in this section, we are going to see some of our related works. We also will discuss their merits and problems with their works. We also talked about how they could solve this problem also. Moreover, in this section, we will explain the difficulties we had to face during our implementation.

2.2 Related Terminology

Association Rule Mining : Association rules are if / then statements support uncovering connections in a database between unrelated data, relational database, or other repository of knowledge. Association Rules are used to assess the relationship between objects which are used together regularly. Practices of association rules are basket-data analysis, grouping, association rules cross-marketing, clustering, catalog and loss-leader architecture analysis et cetera. If the customer buys bread, for example, then he can also buy some butter. If the consumer purchases a laptop then he will even purchase memory card. There are two basic requirements which the association rules apply, support and confidence It sets out the relationships and rules created by the analysis of frequently used data with if / then Models.

Support :

Support provides a snapshot of how much an itemset is in all transactions. Consider set of items1 = bread and set2 = shampoo. There will be much more bread-containing transactions than those of shampoos. So, as you correctly inferred, itemset1 would have a higher help than itemset2 generally. Take itemset1

= broad, butter and itemset2 = bread, shampoo. A lot of purchases are going to have bread and butter on the cart but bread and shampoo? Not that many. Thus itemset1 would usually have a higher benefit in this case than itemset2.Support is mathematically a fraction of the total number of transactions in the itemset.
Support (P -> R)=(Transactions containing both P and R)/(Total Number of Transactions)
Support value allows one to define the rules which should be considered for further review. For example, one would want to consider only sets of products that occur at least 20 times out of a total of 10,000 transactions i.e. support = 0.002. If there is a very low support for an itemset, we do not have enough knowledge about the relationship between its items and thus no conclusions can be drawn from such rules.

Confidence:
Confidence says how likely item Y is to be purchased when buying item X, expressed as X-> Y. This is calculated by the proportion of transactions with item X which also includes item Y.
Confidence (X -> Y) = (Transactions containing both X and Y)/ Transactions containing X
For such a regular outcome it doesn't matter what you have in the antecedent. Trust in an association rule which has a very frequent effect will always be high.

Lift:
Lift says how likely item Y is to be bought when item X is bought, while controlling how popular item Y is.To this measure lift is a very simple term. Imagine X and Y item is in your cart. Think of it as the * lift * that X gives us confidence to have Y on your cart. To rephrase, lift is the increase in the likelihood of having Y on the cart with knowledge of X being present over the likelihood of having Y on the cart with no knowledge of X being present.
Lift (X -> Y) = (Transaction containing both X and Y)/(Transaction Containing X)/Fraction of transactions containing Y.
Value K: In the Top-K association Rule Mining, we have encountered a new term, and that is K. Instead of minimum support, we introduced it.
K = A parameter k representing the number of association rules to be discovered

(a positive integer)

Itempair Minimum Support :

In the indirect association rule we will see a term 'ts' that indicates itempair minimum support.First we have to know about the item pair then it can be easily understand.

Consider the following pair of items, x and y, which are rarely encountered in the same transaction. If both items are highly dependent on the presence of another itemset M, the pair (x, y) is said to be associated indirectly via M.

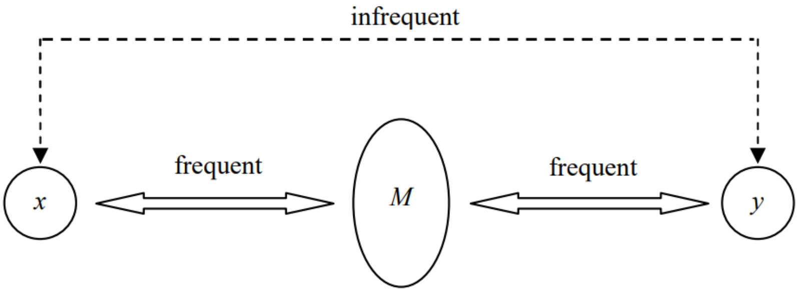


Figure 2.1: Indirect association between x and y via mediator M

Fig.1. Indirect association between x and y via mediator M

2.3 Related Literature Review

The first association rule mining algorithm is Apriori. That was proposed by Agarwal and Srikant in 1994 [1]. Here the association rule is generate by the means of support and confidence. In this algorithm, there are two processes. The first thing is to generate the candidate item sets and second thing is to produce a large item set which is based on the minimum threshold values of support. The normal thing is that if any k itemset is not frequent than (k+1) super-itemset will also be not frequent. In this, algorithm they use candidate generation process to mine the frequent itemset. Here, two things are very important. These two things are used to prune and calculate the accurate association rules. But the main problem of the association rule is that it is very time consuming and it requires a lot of space. It also searches the database several times to generate candidate item sets. To overcome the shortcomings of the Apriori, the data scientists came with a new algorithm. That is FP-Growth[10]. It is normally tree based and it just scan the whole database just twice. It constructs a frequent pattern tree and conditional pattern base from database which satisfy the minimum support. Due to compact structure it requires less execution time and space. It is also best for large itemset. After that, FP-Growth with lift measure is introduced.This is a variation of the algorithm for mining all association rules from a transaction database.Traditionally, association rule mining is performed using two interestingness measures: the support and confidence to evaluate rules. In this algorithm, they showed how to use another popular measure called the lift or interest[11]. Then the re-searcher wanted to find out the different types of association rules. So, L Szathmary worked with a new algorithm to find interesting association rules in datasets.It is a mining algorithm for closed association rules a compact subset of all association rules[12].In 2005, some re-searcher was interested to find the sporadic association rules in the datasets.They invent an algorithm for mining rules of association that are perfectly sporadic. The algorithm begins by generating perfectly rare itemsets using AprioriInverse. Then, it generates the association rules using these itemsets[13].In the same year, they came up with the zart algorithm to find the IGB association rules. This algorithm extracts from a transaction database a subset of all association rules known as IGB association rules (Informative and Generic Basis of Association Rules).This algorithm performs two steps to discover the IGB association rules: (1) It begins by applying the Zart algorithm to discover Closed itemsets and their associated generators. (2) Then, using closed itemsets and generators, association rules are generated[14].In 1998, Kryszkiewicz discovered the set of "minimal non redundant association rules," a set of association rules that is lossless and compact.He employs the Zart algorithm in this implementation to discover closed itemsets and their associated generators. Then, using this information, the "minimal non redundant association rules" were generated[15].Then, looking forward in time, the researcher discovers an indirect relationship between the data. As a result, they implemented an algorithm for discovering indirect associations between items in transaction databases called Indirect association rule mining. This algorithm is critical because conventional association rule mining algorithms are oriented around direct associations between itemsets. This algorithm is capable of discovering indirect associations, which is advantageous in fields such as biology. Indirect association rule mining is helpful for various purposes, including stock market analysis and competitor product analysis[16]. After working with these algorithms the scientist came up with a term to replace support in order to increase the performance of the algorithms .TopKRules is a search algorithm for locating the top-k association rules in a transaction database. Other association rule mining algorithms entail the specification of a difficult-to-set minimum support (minsup) parameter (usually users set it by trial and error, which is time consuming). TopKRules circumvents this issue by allowing users to specify k, the number of rules to be discovered, directly rather than through the use of minsup[17].The researcher proposes RP-growth, a fast algorithm for top-k mining of discriminative patterns that are highly relevant to the class of interest, based on FP-growth. RP-growth performs a branch-and-bound search with anti-monotonic upper bounds on relevance scores such as F-score and 2, and the pruning in branch-and-bound search is successfully translated to minimum support raising, a well-known and simple-to-implement pruning strategy for top-k mining. Additionally, by introducing the concept of weakness and an additional, aggressive pruning strategy based on weakness, RP-growth efficiently discovers k patterns with a high degree of diversity and relevance to the class of interest[18]. Different Re-searcher in the past years work with these algorithm and compare among the algorithms to see the relation and diffrences amongst them. Because if we can know the behaviour of the algorithm we can easily work with them.

In 2017, Vani worked with Apriori,Eclat and FP-Growth algorithms[19].He compared them in terms of memory uses and run time.After comparing among the algorithms he found that Eclat gives better results than the other algorithm. But in thier re-search work they focused on generating the frequent patterns rather than generating association rules. Moreover, the data volume is low because he worked with four small datasets and three algorithms.

Three re-searchers worked with the Apriori,Eclat, dEclat,FP-growth,FIN, AprioriTID, Relim and H-Mine algorithms[20].The performance factors considered are the number of frequent items generated (different thresholds), memory requirements, execution time, and the number of rules generated (different support and confidence threshold values), as well as the size of the datasets used. Finally, they observed that the DCLAT algorithm outperformed the other algorithms in this comparative analysis. But they didn't work with huge amount of data and they also focused on finding out the frequent pattern itemsets.

A paper was presented by Komal Khurana and Mrs. Simple Sharma[21]. This article compares five algorithms for mining association rules: AIS, SETM, Apriori, AprioriTID, and Apriori Hybrid. AprioriTID and Apriori Hybrid have been proposed as solutions to the Apriori algorithm's problem. They conclude that the Apriori Hybrid is superior to the Apriori and AprioriTID due to its reduced overall speed and increased accuracy.

Ziauddin and colleagues conducted research on association rule mining[22]. They presented an overview of research activity dating all the way back to its inception .He did, however, suggest that association rule mining is still in its infancy. There are still several critical issues to investigate in order to identify useful association rules.

The re-searchers collaborated on projects with SETM, AIS, Apriori, AprioriTID, and AprioriHybrid.AprioirHybrid is the superior algorithm after comparison.They compared these algorithm in terms of itemset generated,speed and also the memory it used.But the algorithm was intend to discover normal association rules. They did not focus on the variety of association rules[23].

AprioriTid and AprioriHybrid have been proposed to address the issue of apriori algorithms[24]. They infer from the comparison that AprioriHybrid is superior to Apriori and AprioriTid because it has a slower overall pace and increased precision. On the basis of these parameters, they concluded that the LogElcat algorithm outperforms all other algorithms.But,working with a small amount of algorithm can give them better decision and they have worked with the old algorithms rather than the new ones.

In 2015, two re-searcher worked with Apriori, Apriori hybrid, Fp-growth,AIS, SETM and AprioriTID .The algorithms are systematized, and their performance is evaluated from both a run time and theoretical perspective.Despite the fundamental differences in the strategies employed, the run time demonstrated by algorithms is nearly identical. By making only two passes through the data sets and eliminating the concept of candidate generation, FP growth outperformed Apriori in all cases[25].

So, in this work we worked with 10 different association rule mining algorithms and five different datasets. We also give emphasis on different association rules.As our data volume is high, we mainly worked with number of association rules generated in terms of support,lift and confidence. We tried to clear view on different association rule mining algorithm and how they perform on different datasets.

2.4 Conclusion

In this section,First, we discussed about our related terminology so that anyone can read the report easily. After that, we tried to give an overview of the related works.Moreover, we also discussed about their merits and why their work was not up to the mark. We also discussed about the new research that is conducted with the new strategy.

2.4.1 Implementation Challenges

We have faced some implementation challenges. The first challenge was to collect the datasets. Significantly, the Facebook dataset was challenging to find. Because Facebook is not giving their information in public nowadays. Then, the second problem was to pre-process the dataset according to our algorithms after collecting it. So, we had to convert the dataset into numeric so that the execution time is good. On the other hand, choosing a vast dataset was problematic. Because the Computer specification was not high enough to process these data. Moreover, The new association rule mining technique resources was not good enough. However, with the guidance of my supervisor, at last I completed the work.

Chapter 3

Methodology

3.1 Introduction

Association rule mining [ARM] has been initially presented as the most well known and exhilaratingly researched method of data mining.. Association rule mining is a high resolution designed for substitute rule mining since its objects completely implement rules on data, and can therefore arrange a full representation in a huge dataset of associations.. So, in this section we are going to explain our methodology for the work.How we compared the results of ten algorithms that were implemented in five datasets.

3.2 Diagram/Overview of Framework

Proposed framework. is depicted in Figure 3.1. The Architecture is comprised of the following major steps:
(1) Dataset Collection
(2) Pre-processing of Datasets
(3)Applying Association Rule mining Algorithms
(4) Evaluating Perfomance
(5) Decision

Our main work is to create a system so that we can compare algorithms of association rule mining. That's why we divided the system in five parts. First, we collected the datasets. After collecting the datasets we work on the algorithms and implemented it in our desired programming language. Then we implemented them in five different datasets. After that, we evaluate the performance of the

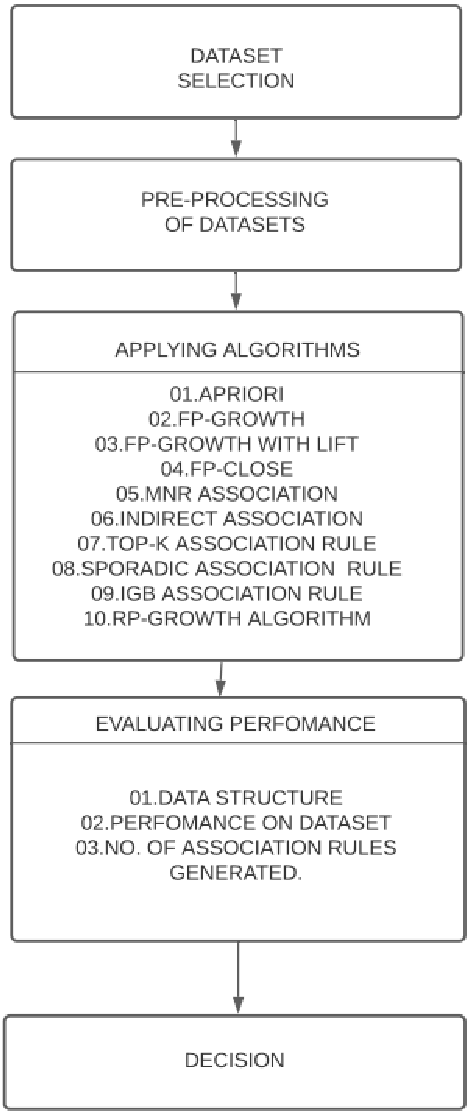


Figure 3.1: Proposed framework.

algorithms on the basis of association rule generated,their performance in the datasets,time needed to generate this dataset.Then,after evaluating the performance we know that which algorithm is suitable for which dataset.

3.3 Detailed Explanation

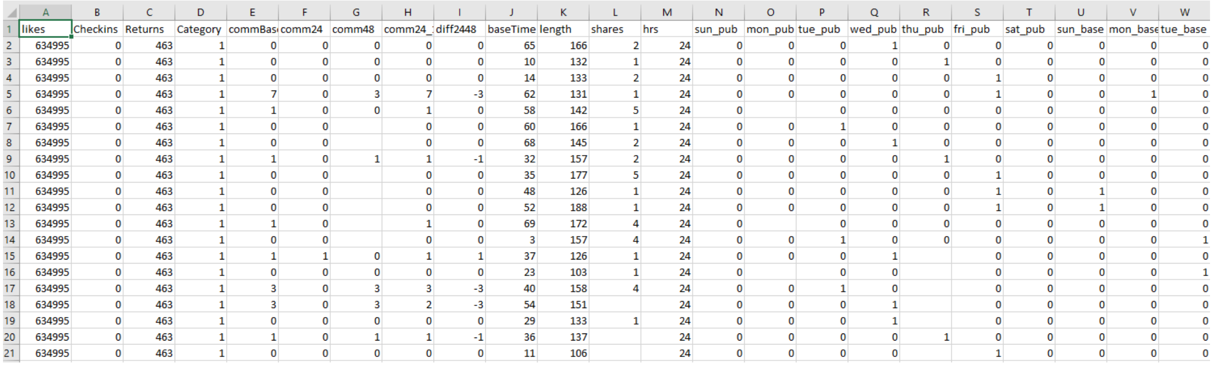
3.3.1 Dataset Collection

Before every data mining work, one has to look for the datasets. As we are working with the association rule mining algorithms, we have to look for those suitable datasets. We were looking for one transnational, one social media dataset, one school dataset and two real-life datasets.

Zoo Dataset : It has 17 Boolean valued attributed and it is collected from the UCI Machine Learning Repository.

Facebook Dataset : It has 27 attributes and have 40949 instances. It is collected from Kaggle.

Breast Cancer Dataset : This dataset has 10 attributes and it is collected from UCI machine learning repository.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	likes	Checkins	Returns	Category	commBas	comm24	comm48	comm24_	diff2448	baseTime	length	shares	hrs	sun_pub	mon_pub	tue_pub	wed_pub	thu_pub	fri_pub	sat_pub	sun_base	mon_base	tue_base
2	634995	0	463	1	0	0	0	0	0	65	166	2	24	0	0	0	1	0	0	0	0	0	0
3	634995	0	463	1	0	0	0	0	0	10	132	1	24	0	0	0	0	1	0	0	0	0	0
4	634995	0	463	1	0	0	0	0	0	14	133	2	24	0	0	0	0	0	1	0	0	0	0
5	634995	0	463	1	7	0	3	7	-3	62	131	1	24	0	0	0	0	0	1	0	0	1	0
6	634995	0	463	1	1	0	0	1	0	58	142	5	24	0		0	0	0	0	0	0	0	0
7	634995	0	463	1	0	0		0	0	60	166	1	24	0	0	1	0	0	0	0	0	0	0
8	634995	0	463	1	0	0		0	0	68	145	2	24	0	0	0	1	0	0	0	0	0	0
9	634995	0	463	1	1	0	1	1	-1	32	157	2	24	0	0	0	0	1	0	0	0	0	0
10	634995	0	463	1	0	0		0	0	35	177	5	24	0	0	0	0	0	1	0	0	0	0
11	634995	0	463	1	0	0		0	0	48	126	1	24	0	0	0	0	0	1	0	1	0	0
12	634995	0	463	1	0	0		0	0	52	188	1	24	0	0	0	0	0	1	0	1	0	0
13	634995	0	463	1	1	0		1	0	69	172	4	24	0		0	0	0	0	0	0	0	0
14	634995	0	463	1	0	0		0	0	3	157	4	24	0	0	1	0	0	0	0	0	0	1
15	634995	0	463	1	1	1	0	1	1	37	126	1	24	0	0	0	1		0	0	0	0	0
16	634995	0	463	1	0	0	0	0	0	23	103	1	24	0		0	0		0	0	0	0	1
17	634995	0	463	1	3	0	3	3	-3	40	158	4	24	0	0	1	0		0	0	0	0	0
18	634995	0	463	1	3	0	3	2	-3	54	151		24	0	0	0	1		0	0	0	0	0
19	634995	0	463	1	0	0	0	0	0	29	133	1	24	0	0	0	1		0	0	0	0	0
20	634995	0	463	1	1	0	1	1	-1	36	137		24	0	0	0	0	1	0	0	0	0	0
21	634995	0	463	1	0	0	0	0	0	11	106		24	0	0	0	0		1	0	0	0	0

Figure 3.2: A snapshot of Facebook Dataset

Sales Transaction Dataset : This dataset is collected from Kaggle and 52 attributes.

High School Dataset : This dataset is collected from the Kaggle and it has 33 attributes.

3.3.2 Pre-processing of Datasets

Mainly, data pre-processing is getting the dataset ready for the algorithms. The transaction dataset does not have any missing values. So, we did not have to clean the data that means we did not need to add the probable missing values for the dataset. However, we have to fill in the missing value of the breast dataset. The second thing was to convert the dataset into numerical values. We convert the dataset into the numerical dataset because association rule mining algorithms create many rules because of low support and confidence. Moreover, The Facebook dataset was too large, and we converted it into a smaller one for applying association rules. Moreover, the dataset have missing values and we have worked with that so that our result could be better.

3.3.3 Applying Association Rule mining Algorithms

We have selected ten different association rule mining algorithm for our work, and the algorithms are Apriori, FP-Growth, Fp-Growth with lift, RP-growth, MNR, Indirect, Sporadic, IGB, FP-Close, TOP K association rule mining algorithms. We implemented the association rule mining algorithms in the java programming language. Because java makes it easier for data scientists and programmers to scale their applications. As our work is based on comparison we focused on association rules generated,time and space complexity of the algorithm during implementing it.

In 1994, R. Agrawal and R. Srikant provided the Apriori algorithm for frequent itemsets searching on data set under the rule of the Boolean association rule. The algorithm is named Apriori as it uses a common set of elements by its predecessor. To find k+1 articles with k-frequent itemsets, we use an iterative or level approach. The Apriori property, which helps reduce the search area in order to improve the efficiency of generation of frequent items, is an important property. Apriori property – Non-empty sub-sets must often be used for all frequent objectsets. The Apriori algorithm's anti-monotonic supportive measures are key concepts.

Apriori Property – All frequently used non-empty subsets have to be frequent. His anti-monotonic support measures are a key concept of the Apriori algorithm. Apriori believes that- "Every subset of a common item must be frequent (Apriori propertry)."If an item is rare, all its supersets are rare.

The Apriori algorithm has two main disadvantages:-

01.Candidate sets must be constructed at every step.

02.The algorithm must repeatedly scan the database to create the candidate sets. The two characteristics inevitably slow down the algorithm. A new association mining algorithm called Frequent Pattern growth algorithm has been developed to overcome these redundant steps. It overcomes the Apriori algorithm's disadvantages by saving all transactions in a Trie Data Structure. This tree is called the FP-tree.

In the Fp-Growth with lift algorithm, we have taken into account a new parameter and that is called lift.If the lift is equal to 1, it means that the X and Y are independent for an association rule X ==> Y. If the lift is higher than 1, it is a positive correlation between X and Y. If the lift is less than 1, X and Y have a negative relationship.Subsequent lift controls (frequency) when calculating the conditional probability of Y X occurrence.

FPClose is an FPGrowth algorithm of algorithms, which has been designed to mine frequent closed items. One of the fastest closed mining algorithms is supposed to be FPClose. FPClose outputs Closed itemsets are often closed. A common item set is an item set which appears in the transaction databases in at least minsup transactions. A frequently closed item is a frequent item set which doesn't feature exactly the same support in a correct superset. The frequent closed array is therefore a subset of the common array. The set of frequently closed items is generally much smaller than the set of common items, and no information can be shown to be lost.

This algorithm extracts from a dataset a subset of association rules called IGB Association rules. It is an informative and generic set of association rules.This algorithm takes two steps to discover the IGB association rules: (1) first, the Closed Items and their related generators are found Then (2) the rules of association are generated through the use of closed object and generators.
Sporadic Association rule mining algorithm is an algorithm for mining perfectly sporadic association rules. First, the algorithm creates quite rare items. A rare item set (also a sporadic item set) is an item set not commonly used and all of its subsets are not common items. In addition, support must be higher or equal to the minimum aid threshold. Then the association rules are generated with these itemsets.

A closed association rule mining algorithm return a set of closed association rules .A closed association rule is an X ==> Y association rule that is a closed item-set of union of X and Y.The algorithm returns all closed association rules so that they are either higher or equal to the user's mininimum support and minimum confidence thresholds.

The "minimal non-redundant association rules," a series of loss-free and compact association rules (Kryszkievicz, 1998) are discovered by MNR algorithm. We find closed items and their associated generators in this implementation. This information is then used to produce the 'minimum non-redundant rules of association.' The minimum set of unredundant assignment rules shall be defined in the form P1 => P2/P1, where P1 is a generator of P2, P2 is a closed element and the rule shall have no less than support and confidence than Minimum support and Minimum confidence respectively.

Indirect is an algorithm for finding indirect links between items in transactions databases (Tan et al., KDD 2000; Tan, Steinbach Kumar, 2006, p. 469).

Traditional rule mining association algorithms focus on direct links between itemsets. This algorithm can find indirect associations, useful for fields like biology. Indirect association rule mining applies in various ways, for example stock analysis and competitive product analysis (Tan et al., 2000). The form x,y=>M of an indirect association is where x and y are individual items, M is a set called "mediator."

The following conditions shall be met by an indirect association:

01.The transaction count shall be greater or equal to minimum support with all items from x to M as separated by the total transaction number.

02.There shall be higher or equal to minus the number of transactions with all parts of y M divided by the total number of transactions.

03.The total transaction number must be smaller than itempair minimum support for transactions containing x,y divided into total transactions.

04.The Confidencex to M and y to M must be greater or equal to minimum confidence.

05.The Confidence x to M and y to M must be greater or equal to minconf. The trust of an item X in another item Y is defined as the number of transactions containing X and Y divisioned by the number of transactions containing X.

TopKRules is an algorithm for the discovery in a transaction database of the highest standards for associations. Other algorithms associated with mining rules require a minimum (minsup) support parameter to be set (usually users set it by trial and error, which is time consuming). TopKRules solves this problem by letting users indicate directly k, instead of using minsup, the number of rules to be discovered.

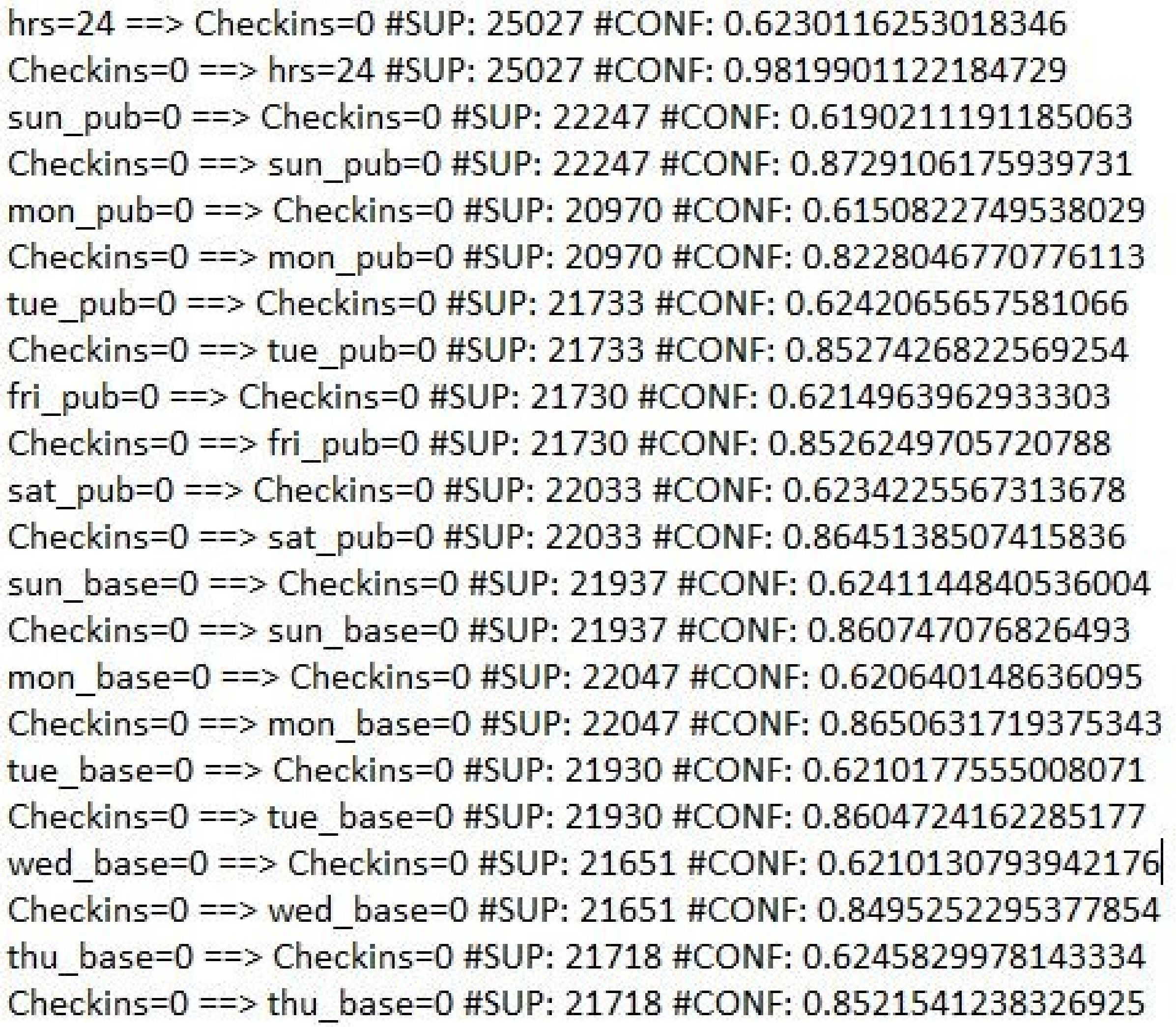
RPGrowth adapts the FPGrowth algorithm to detect rare items in the transaction database. Sidney Tsang, Yun Sing Koh, and Gillian Dobbie proposed the adjustment of the FPGrowth to identify scarce item sets (2011). It retains the speed, memory efficiency, and use of the fp tree as modified through the FP-Growth algorithm. A transaction data base RPGrowth is an algorithm for the finding of item sets (group of items) that occur seldom (rare item sets). A rare itemset is an itemset appearing within the minrare support and minimum support range.

3.3.4 Evaluating Performance

The central part of the work is Evaluating the performance of the algorithm. For this, we changed the value of support, confidence and K value in two different ways. At first, we changed the support/K and confidence value of the algorithms. In the second stage, we run the algorithm by changing the support value or K value and keeping the confidence value fixed. After the execution of the algorithm, we collected the value of no. of association rules generated. Plot the values in the graph so that we can visualise things better.

3.3.5 Decision

In this section, we compare the algorithms in terms of the association rules generated for support and confidence values. We will see the algorithm's behaviour in different datasets for a slight change in the support and confidence. Then, we will give a decision basis on our comparison.



```
hrs=24 ==> Checkins=0 #SUP: 25027 #CONF: 0.6230116253018346
Checkins=0 ==> hrs=24 #SUP: 25027 #CONF: 0.9819901122184729
sun_pub=0 ==> Checkins=0 #SUP: 22247 #CONF: 0.6190211191185063
Checkins=0 ==> sun_pub=0 #SUP: 22247 #CONF: 0.8729106175939731
mon_pub=0 ==> Checkins=0 #SUP: 20970 #CONF: 0.6150822749538029
Checkins=0 ==> mon_pub=0 #SUP: 20970 #CONF: 0.8228046770776113
tue_pub=0 ==> Checkins=0 #SUP: 21733 #CONF: 0.6242065657581066
Checkins=0 ==> tue_pub=0 #SUP: 21733 #CONF: 0.8527426822569254
fri_pub=0 ==> Checkins=0 #SUP: 21730 #CONF: 0.6214963962933303
Checkins=0 ==> fri_pub=0 #SUP: 21730 #CONF: 0.8526249705720788
sat_pub=0 ==> Checkins=0 #SUP: 22033 #CONF: 0.6234225567313678
Checkins=0 ==> sat_pub=0 #SUP: 22033 #CONF: 0.8645138507415836
sun_base=0 ==> Checkins=0 #SUP: 21937 #CONF: 0.6241144840536004
Checkins=0 ==> sun_base=0 #SUP: 21937 #CONF: 0.860747076826493
mon_base=0 ==> Checkins=0 #SUP: 22047 #CONF: 0.620640148636095
Checkins=0 ==> mon_base=0 #SUP: 22047 #CONF: 0.8650631719375343
tue_base=0 ==> Checkins=0 #SUP: 21930 #CONF: 0.6210177555008071
Checkins=0 ==> tue_base=0 #SUP: 21930 #CONF: 0.860472416228517
wed_base=0 ==> Checkins=0 #SUP: 21651 #CONF: 0.6210130793942176
Checkins=0 ==> wed_base=0 #SUP: 21651 #CONF: 0.8495252295377854
thu_base=0 ==> Checkins=0 #SUP: 21718 #CONF: 0.6245829978143334
Checkins=0 ==> thu_base=0 #SUP: 21718 #CONF: 0.8521541238326925
```

Figure 3.3: A snapshot of Association rules after applying apriori rule mining algorithm in Facebook dataset

3.4 Implementation

The system requirement for the implementation of the project is mentioned below:

- Hardware requirements:

– Personal computer

- System configuration:

– Operating system: Windows

– A 64-bit Intel core i5 processor

– 8GB RAM

– 500MB free internal storage

- Software requirements:

– Language: JAVA

– IDE: Eclipse JAVA 2019-06

3.5 Conclusion

So, in this section showed the detailed explanation of the steps of methodology. Moreover, we give a small brief about our system overview. In the next section. we will discuss about our results and discussion.

Chapter 4

Results and Discussions

4.1 Introduction

In the previous chapter, a detailed description of the framework and its various components were discussed. The performance of the association rule mining algorithm on various datasets will be discussed in this chapter. Our main challenge is to observe how the algorithms work in our selected datasets. The proposed method was implemented in java language with an Intel Core i5 processor and an 8GB RAM. Data was collected from the UCI machine learning repository and Kaggle.

4.2 Dataset Description

The main thing in this research work is to choose the correct datasets. For working in this thesis paper we have selecteed five datasets. The description of the datasets are given below :

Breast cancer dataset :

Data Set Characteristics: Multivariate
Attribute Characteristics: Categorical
Number of Instances: 286
Number of Attributes: 09
Missing Values? : Yes
This dataset is collected from the uci repository.

Facebook Comment Prediction Dataset :

The number of instances in this datasets is 40949.28 columns content in this Dataset. The Attributes are described below :

01.Likes : It is a feature that defines users support for specific comments, pictures, wall posts, statuses, or pages.

02.Checkins : It is an act of showing presence at particular place and under the category of place,institution pages only.

03.Returns : How many people return after visiting the pages. 04.Category : Defines the category of the source of the document eg: place, institution, brand etc.

05. Length : Post length

06. Shares : How many shares the post get

07. Hours : Expaned time after the post .

from 8-13. This includes the pattern of the posting comment at different intervals w.r.t to the randomly chosen base date/time.

i)commBase

: Total comment count before selected base date/time.

ii)comm24

: Comment number to the chosen base date/time in the last 24 hours w.r.t.

iii)comm48

: Comment count is the last 48 hrs to last 24 hours with respect to base date/time.

iv)comm24-1

: Comment

count in 24 hours after the document is published, but before the selected basis date/time.

v)diff2448

: The differencebetween comm24 and comm48.

vi) baseTime

: the base time in hour selected.

From 13-26. Features on weekdays: Binary indicators (0, 1) are used to represent the day the post was posted and the day the base date/time is selected. This type identifies 14 characteristics. Example (Sunday pub means published Sunday, Sun base means that we want a comment about the post and if so)

27. Output : did we get the specific number of comments in that base time, if yes then how many ? if no then Zero.

High School Assesment Dataset :

Number of instances in this dataset is 395 nad there are 33 attributes in the dataset. So the atttributes are :

school - student's school (binary values that means 'GP' means that Gabriel Pereira or 'MS' means that Mousinho da Silveira)

sex –sex for a student (binary: 'F' means that women or 'M' means that men)

age - The age of the student (numerical values from 15 to 22)

address - Type of domicile of student (binary: "U" - urban or "R" - rural)

famsize - Family dimensions (binary: 'LE3'–less than or equal to 3 or 'GT3'-larger than 3)

Pstatus - Status of coexistence for parents (binary: 'T'-or 'A'-separate)

Medu – education to the mother (education:0-none,1-primary (fourth grade), 2-5th to 9th grade, 3-secondary or 4-high)

Fedu - father's training (numeric: 0 - none, 1 - primary school (4th degree), 2 - 5th to 9th degree, 3 - secondary school or 4 - university)

Mjob - mother's job (nominal values - 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')

Fjob - father's job (nominal values - 'teacher', 'health' care related, civil 'services' (e.g. administrative or police services), 'at_home' or 'other' work)

reason – Why the students choose this school (nominal values -close to 'home', school 'reputation', 'course' preference or 'other')

guardian – Who will be the student's guardian (nominal values - 'mother', 'father' or 'other')

traveltime – Time takes to go from home to school (numeric values 1 - 1 hour)

studytime - weekly study time of the students (numeric values 1 - 10 hours)

failures - number of failures in the past classes (numeric values :- n if 1<=n<3, else 4)

schoolsup - extra educational support (binary: yes or no)

famsup - family educational support (binary: yes or no)

paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no) activities - extra-curricular activities (binary: yes or no)

nursery - attended nursery school (binary: yes or no)

higher - wants to take higher education (binary: yes or no)

internet - Internet access at home (binary: yes or no)

romantic - with a romantic relationship (binary: yes or no)

famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)

freetime - free time after school (numeric: from 1 - very low to 5 - very high)

goout - going out with friends (numeric: from 1 - very low to 5 - very high)

Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)

Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)

health - current health status means that numerical values from 1 - very bad to 5 - very good absences - number of school absences that indicates numerical values from 0 to 93)

this dataset is collected from the uci ripository.

Super Market Dataset : We took this dataset as a transactional dataset. This dataset have 811 instances and it has 107 attributes. It is collected from kaggle and the attributs information are :

Product_Code

52 weeks: W0, W1, ..., W51. Normalised vlaues of weekly data that indicates Normalised 0, Normalised 1, ..., Normalised 51.

Zoo Dataset :

Data Set Characteristics: Multivariate

Attribute Characteristics: Categorical

Number of Instances: 101

Number of Attributes: 17

Missing Values? : NO

Then the attributes are :

1.animal name: Unique for each instance

2. hair: Boolean

3. feathers: Boolean

4. eggs: Boolean
5. milk: Boolean
6. airborne: Boolean
7. aquatic: Boolean
8. predator: Boolean
9. toothed: Boolean
10. backbone: Boolean
11. breathes: Boolean
12. venomous: Boolean
13. fins: Boolean
14. legs: Numeric (set of values: 0,2,4,5,6,8)
15. tail: Boolean
16. domestic: Boolean
17. catsize: Boolean
18. type: Numeric (integer values in range [1,7])

This dataset is also collected from the uci repository.
So, we have taken one transactional, one social media and three real-life datasets to see the behaviour of algorithms on different datasets.

So, we have taken one transactional, one social media and three real-life datasets to see the behaviour of algorithms on different datasets.

4.3 Impact Analysis

So, when we are doing any project or thesis, there will be some impact on society and the environment. There will be some ethical impact as data mining is concerned with the data of personal or business individuals.

4.3.1 Social and Environmental Impact

Like another thesis, this thesis will have some social and environmental impacts. My project is all about the data mining process. Data mining plays a vital role in our regular life. If a data scientist knows which algorithm is faster to extract the association rules in a dataset, he could easily find the interesting transaction relationship. This relation will help a business person to change his strategy in business.Moreover, the prediction of diseases will be much more straightforward for the association rule mining.

Classification of new species is the new term of the association rule. It will relate the things that are in common in a species and group them.Moreover, protein Sequencing is going to have a new look for association rule mining.

So, at last, we can say that if one can apply the association rules in a suitable dataset and find the mystery of association amongst the data, it will severely impact our social and environmental life.

4.3.2 Ethical Impact

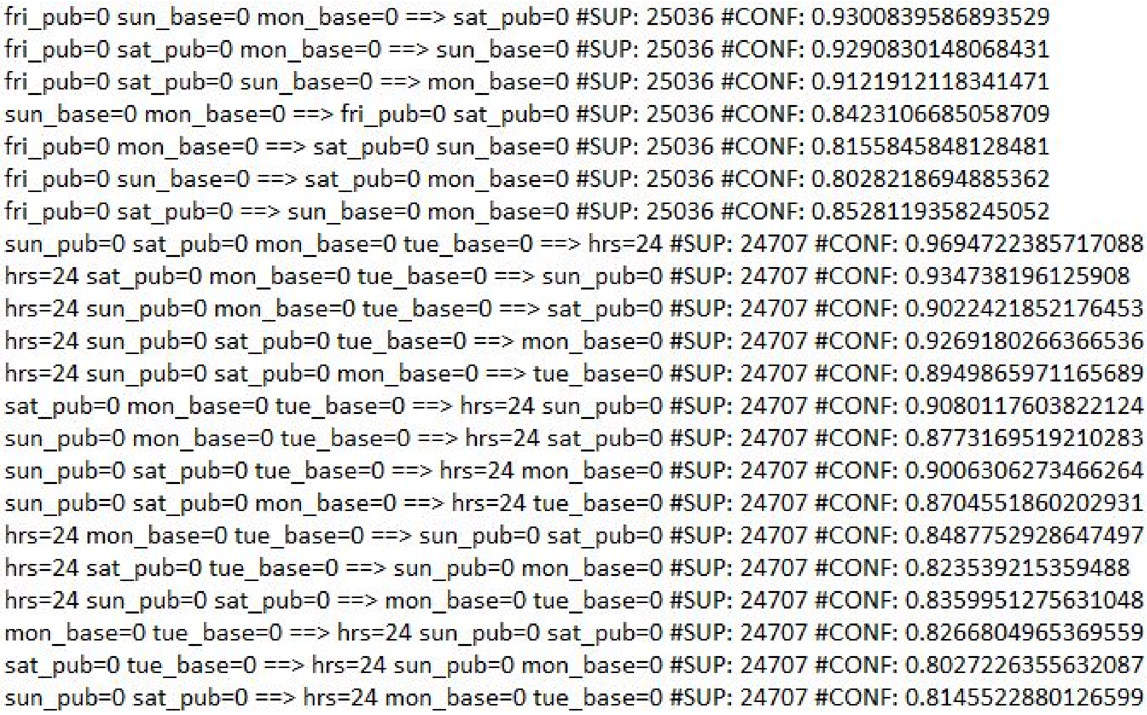
This digital world is just based on the data. In daily life, we are working with the data to make our life beautiful. The business giants analyze the transaction data, the scientist works the protein sequence, and cure the patients. So, if some person collects data without the permission of the individuals, then this is unethical. Moreover, the more dangerous thing is if someone applies the association rules in the business transaction data because it will reveal the strategies of the business individuals.

Moreover, with the student's and patient's data, one can easily find the person's information. So, the data integrity will be lost, and that people would be in a problem. So, this thesis will have some ethical impact too.

4.4 Result Analysis

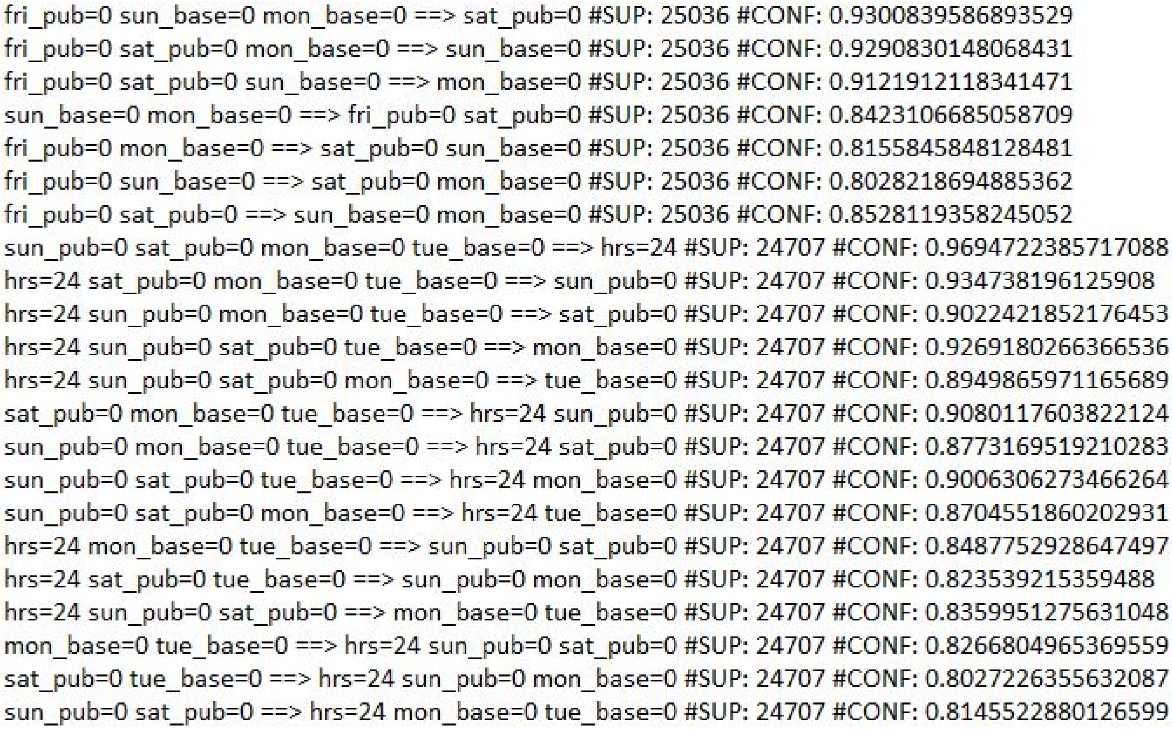
We have applied ten algorithms in five different datasets. The sample output of these algorithms are shown below : In fig-4.1, in the first rule we have SUP:25036 that means it has support of 25036 transactions. Moreover, it has confidence values of 0.93008.

In fig-4.2, in the first rule we have SUP:25036 that means it has support of 25036 transactions. Moreover, it has confidence values of 0.93008.



```
fri_pub=0 sun_base=0 mon_base=0 ==> sat_pub=0 #SUP: 25036 #CONF: 0.9300839586893529
fri_pub=0 sat_pub=0 mon_base=0 ==> sun_base=0 #SUP: 25036 #CONF: 0.9290830148068431
fri_pub=0 sat_pub=0 sun_base=0 ==> mon_base=0 #SUP: 25036 #CONF: 0.9121912118341471
sun_base=0 mon_base=0 ==> fri_pub=0 sat_pub=0 #SUP: 25036 #CONF: 0.8423106685058709
fri_pub=0 mon_base=0 ==> sat_pub=0 sun_base=0 #SUP: 25036 #CONF: 0.8155845848128481
fri_pub=0 sun_base=0 ==> sat_pub=0 mon_base=0 #SUP: 25036 #CONF: 0.8028218694885362
fri_pub=0 sat_pub=0 ==> sun_base=0 mon_base=0 #SUP: 25036 #CONF: 0.8528119358245052
sun_pub=0 sat_pub=0 mon_base=0 tue_base=0 ==> hrs=24 #SUP: 24707 #CONF: 0.9694722385717088
hrs=24 sat_pub=0 mon_base=0 tue_base=0 ==> sun_pub=0 #SUP: 24707 #CONF: 0.934738196125908
hrs=24 sun_pub=0 mon_base=0 tue_base=0 ==> sat_pub=0 #SUP: 24707 #CONF: 0.9022421852176453
hrs=24 sun_pub=0 sat_pub=0 tue_base=0 ==> mon_base=0 #SUP: 24707 #CONF: 0.9269180266366536
hrs=24 sun_pub=0 sat_pub=0 mon_base=0 ==> tue_base=0 #SUP: 24707 #CONF: 0.8949865971165689
sat_pub=0 mon_base=0 tue_base=0 ==> hrs=24 sun_pub=0 #SUP: 24707 #CONF: 0.9080117603822124
sun_pub=0 mon_base=0 tue_base=0 ==> hrs=24 sat_pub=0 #SUP: 24707 #CONF: 0.8773169519210283
sun_pub=0 sat_pub=0 tue_base=0 ==> hrs=24 mon_base=0 #SUP: 24707 #CONF: 0.9006306273466264
sun_pub=0 sat_pub=0 mon_base=0 ==> hrs=24 tue_base=0 #SUP: 24707 #CONF: 0.8704551860202931
hrs=24 mon_base=0 tue_base=0 ==> sun_pub=0 sat_pub=0 #SUP: 24707 #CONF: 0.8487752928647497
hrs=24 sat_pub=0 tue_base=0 ==> sun_pub=0 mon_base=0 #SUP: 24707 #CONF: 0.823539215359488
hrs=24 sun_pub=0 sat_pub=0 ==> mon_base=0 tue_base=0 #SUP: 24707 #CONF: 0.8359951275631048
mon_base=0 tue_base=0 ==> hrs=24 sun_pub=0 sat_pub=0 #SUP: 24707 #CONF: 0.8266804965369559
sat_pub=0 tue_base=0 ==> hrs=24 sun_pub=0 mon_base=0 #SUP: 24707 #CONF: 0.8027226355632087
sun_pub=0 sat_pub=0 ==> hrs=24 mon_base=0 tue_base=0 #SUP: 24707 #CONF: 0.8145522880126599
```

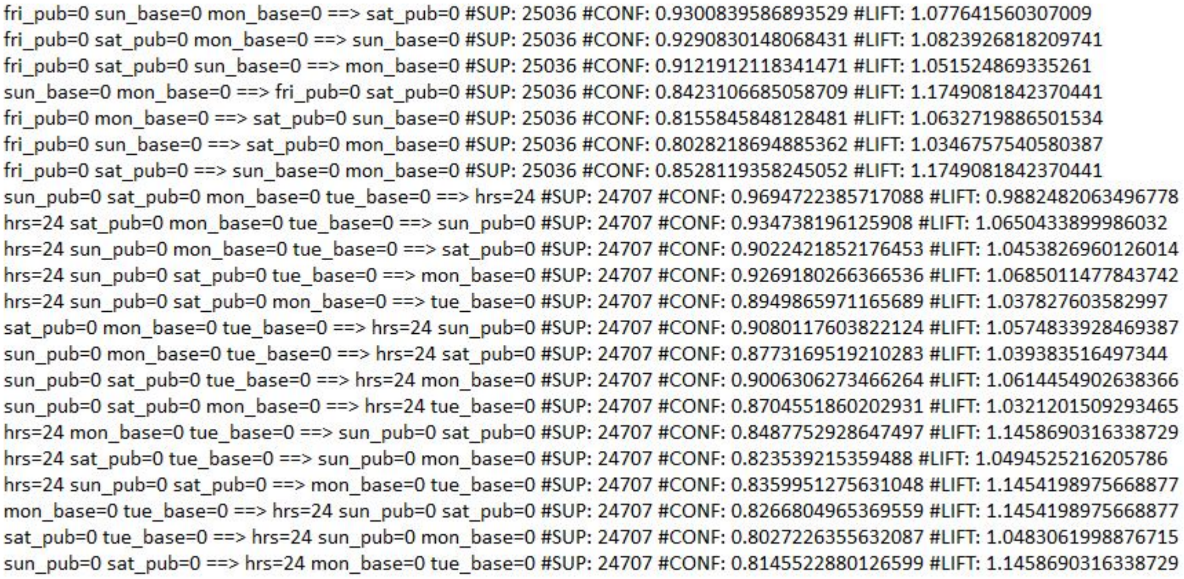
Figure 4.1: A snapshot of association rules generated in Facebook dataset after applying Apriori algorithm



```
fri_pub=0 sun_base=0 mon_base=0 ==> sat_pub=0 #SUP: 25036 #CONF: 0.9300839586893529
fri_pub=0 sat_pub=0 mon_base=0 ==> sun_base=0 #SUP: 25036 #CONF: 0.9290830148068431
fri_pub=0 sat_pub=0 sun_base=0 ==> mon_base=0 #SUP: 25036 #CONF: 0.9121912118341471
sun_base=0 mon_base=0 ==> fri_pub=0 sat_pub=0 #SUP: 25036 #CONF: 0.8423106685058709
fri_pub=0 mon_base=0 ==> sat_pub=0 sun_base=0 #SUP: 25036 #CONF: 0.8155845848128481
fri_pub=0 sun_base=0 ==> sat_pub=0 mon_base=0 #SUP: 25036 #CONF: 0.8028218694885362
fri_pub=0 sat_pub=0 ==> sun_base=0 mon_base=0 #SUP: 25036 #CONF: 0.8528119358245052
sun_pub=0 sat_pub=0 mon_base=0 tue_base=0 ==> hrs=24 #SUP: 24707 #CONF: 0.9694722385717088
hrs=24 sat_pub=0 mon_base=0 tue_base=0 ==> sun_pub=0 #SUP: 24707 #CONF: 0.934738196125908
hrs=24 sun_pub=0 mon_base=0 tue_base=0 ==> sat_pub=0 #SUP: 24707 #CONF: 0.9022421852176453
hrs=24 sun_pub=0 sat_pub=0 tue_base=0 ==> mon_base=0 #SUP: 24707 #CONF: 0.9269180266366536
hrs=24 sun_pub=0 sat_pub=0 mon_base=0 ==> tue_base=0 #SUP: 24707 #CONF: 0.8949865971165689
sat_pub=0 mon_base=0 tue_base=0 ==> hrs=24 sun_pub=0 #SUP: 24707 #CONF: 0.9080117603822124
sun_pub=0 mon_base=0 tue_base=0 ==> hrs=24 sat_pub=0 #SUP: 24707 #CONF: 0.8773169519210283
sun_pub=0 sat_pub=0 tue_base=0 ==> hrs=24 mon_base=0 #SUP: 24707 #CONF: 0.9006306273466264
sun_pub=0 sat_pub=0 mon_base=0 ==> hrs=24 tue_base=0 #SUP: 24707 #CONF: 0.8704551860202931
hrs=24 mon_base=0 tue_base=0 ==> sun_pub=0 sat_pub=0 #SUP: 24707 #CONF: 0.8487752928647497
hrs=24 sat_pub=0 tue_base=0 ==> sun_pub=0 mon_base=0 #SUP: 24707 #CONF: 0.823539215359488
hrs=24 sun_pub=0 sat_pub=0 ==> mon_base=0 tue_base=0 #SUP: 24707 #CONF: 0.8359951275631048
mon_base=0 tue_base=0 ==> hrs=24 sun_pub=0 sat_pub=0 #SUP: 24707 #CONF: 0.8266804965369559
sat_pub=0 tue_base=0 ==> hrs=24 sun_pub=0 mon_base=0 #SUP: 24707 #CONF: 0.8027226355632087
sun_pub=0 sat_pub=0 ==> hrs=24 mon_base=0 tue_base=0 #SUP: 24707 #CONF: 0.8145522880126599
```

Figure 4.2: A snapshot of association rules generated in Facebook dataset after applying FP-Growth algorithm

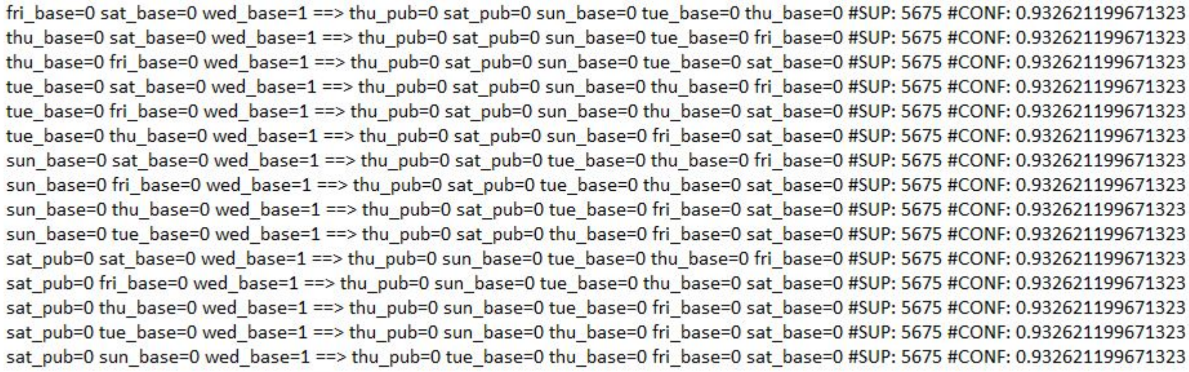
In fig-4.3, in the first rule we have SUP:25036 that means it has support of 25036 transactions. Moroever, it has confidence values of 0.93008 and lift value



```
fri_pub=0 sun_base=0 mon_base=0 ==> sat_pub=0 #SUP: 25036 #CONF: 0.9300839586893529 #LIFT: 1.077641560307009
fri_pub=0 sat_pub=0 mon_base=0 ==> sun_base=0 #SUP: 25036 #CONF: 0.9290830148068431 #LIFT: 1.0823926818209741
fri_pub=0 sat_pub=0 sun_base=0 ==> mon_base=0 #SUP: 25036 #CONF: 0.9121912118341471 #LIFT: 1.051524869335261
sun_base=0 mon_base=0 ==> fri_pub=0 sat_pub=0 #SUP: 25036 #CONF: 0.8423106685058709 #LIFT: 1.1749081842370441
fri_pub=0 mon_base=0 ==> sat_pub=0 sun_base=0 #SUP: 25036 #CONF: 0.8155845848128481 #LIFT: 1.0632719886501534
fri_pub=0 sun_base=0 ==> sat_pub=0 mon_base=0 #SUP: 25036 #CONF: 0.8028218694885362 #LIFT: 1.0346757540580387
fri_pub=0 sat_pub=0 ==> sun_base=0 mon_base=0 #SUP: 25036 #CONF: 0.8528119358245052 #LIFT: 1.1749081842370441
sun_pub=0 sat_pub=0 mon_base=0 tue_base=0 ==> hrs=24 #SUP: 24707 #CONF: 0.9694722385717088 #LIFT: 0.9882482063496778
hrs=24 sat_pub=0 mon_base=0 tue_base=0 ==> sun_pub=0 #SUP: 24707 #CONF: 0.934738196125908 #LIFT: 1.0650433899986032
hrs=24 sun_pub=0 mon_base=0 tue_base=0 ==> sat_pub=0 #SUP: 24707 #CONF: 0.9022421852176453 #LIFT: 1.0453826960126014
hrs=24 sun_pub=0 sat_pub=0 tue_base=0 ==> mon_base=0 #SUP: 24707 #CONF: 0.9269180266366536 #LIFT: 1.0685011477843742
hrs=24 sun_pub=0 sat_pub=0 mon_base=0 ==> tue_base=0 #SUP: 24707 #CONF: 0.8949865971165689 #LIFT: 1.037827603582997
sat_pub=0 mon_base=0 tue_base=0 ==> hrs=24 sun_pub=0 #SUP: 24707 #CONF: 0.9080117603822124 #LIFT: 1.0574833928469387
sun_pub=0 mon_base=0 tue_base=0 ==> hrs=24 sat_pub=0 #SUP: 24707 #CONF: 0.8773169519210283 #LIFT: 1.039383516497344
sun_pub=0 sat_pub=0 tue_base=0 ==> hrs=24 mon_base=0 #SUP: 24707 #CONF: 0.9006306273466264 #LIFT: 1.0614454902638366
sun_pub=0 sat_pub=0 mon_base=0 ==> hrs=24 tue_base=0 #SUP: 24707 #CONF: 0.8704551860202931 #LIFT: 1.0321201509293465
hrs=24 mon_base=0 tue_base=0 ==> sun_pub=0 sat_pub=0 #SUP: 24707 #CONF: 0.8487752928647497 #LIFT: 1.1458690316338729
hrs=24 sat_pub=0 tue_base=0 ==> sun_pub=0 mon_base=0 #SUP: 24707 #CONF: 0.823539215359488 #LIFT: 1.0494525216205786
hrs=24 sun_pub=0 sat_pub=0 ==> mon_base=0 tue_base=0 #SUP: 24707 #CONF: 0.8359951275631048 #LIFT: 1.1454198975668877
mon_base=0 tue_base=0 ==> hrs=24 sun_pub=0 sat_pub=0 #SUP: 24707 #CONF: 0.8266804965369559 #LIFT: 1.1454198975668877
sat_pub=0 tue_base=0 ==> hrs=24 sun_pub=0 mon_base=0 #SUP: 24707 #CONF: 0.8027226355632087 #LIFT: 1.0483061998876715
sun_pub=0 sat_pub=0 ==> hrs=24 mon_base=0 tue_base=0 #SUP: 24707 #CONF: 0.8145522880126599 #LIFT: 1.1458690316338729
```

Figure 4.3: A snapshot of association rules generated in Facebook dataset after applying FP-Growth with Lift algorithm

of 1.077. Here we can see lift greater than one, so they are positively co-related.



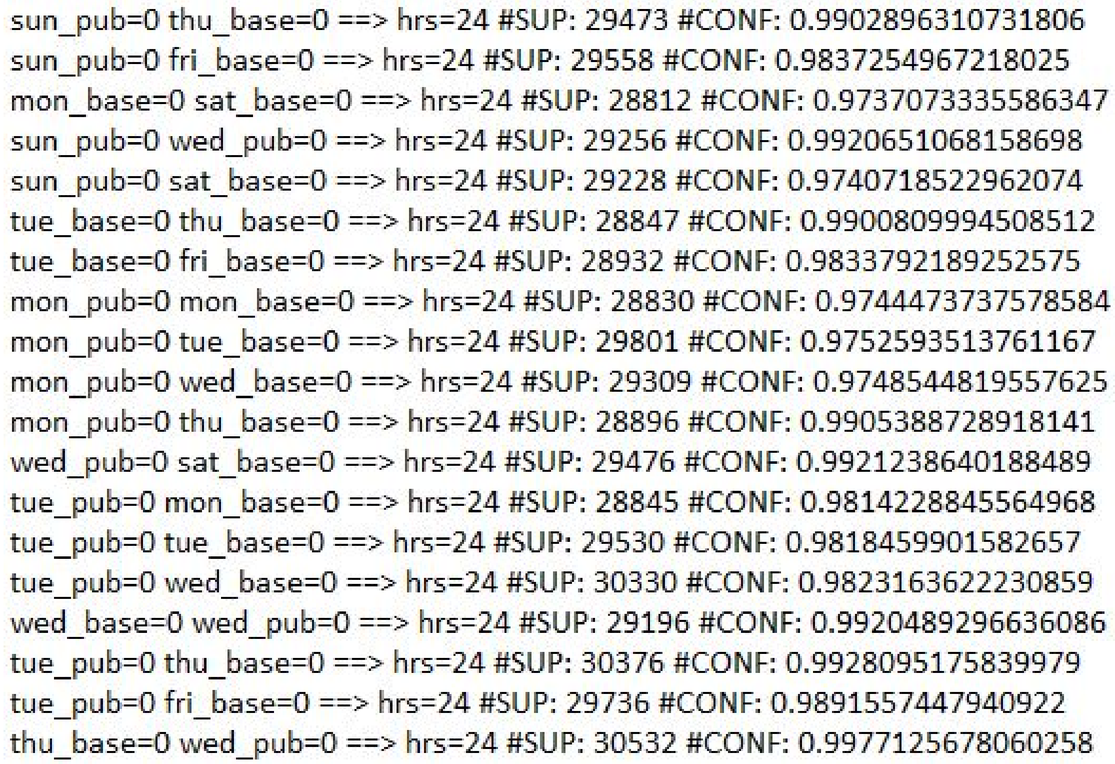
```
fri_base=0 sat_base=0 wed_base=1 ==> thu_pub=0 sat_pub=0 sun_base=0 tue_base=0 thu_base=0 #SUP: 5675 #CONF: 0.932621199671323
thu_base=0 sat_base=0 wed_base=1 ==> thu_pub=0 sat_pub=0 sun_base=0 tue_base=0 fri_base=0 #SUP: 5675 #CONF: 0.932621199671323
thu_base=0 fri_base=0 wed_base=1 ==> thu_pub=0 sat_pub=0 sun_base=0 tue_base=0 sat_base=0 #SUP: 5675 #CONF: 0.932621199671323
tue_base=0 sat_base=0 wed_base=1 ==> thu_pub=0 sat_pub=0 sun_base=0 thu_base=0 fri_base=0 #SUP: 5675 #CONF: 0.932621199671323
tue_base=0 fri_base=0 wed_base=1 ==> thu_pub=0 sat_pub=0 sun_base=0 thu_base=0 sat_base=0 #SUP: 5675 #CONF: 0.932621199671323
tue_base=0 thu_base=0 wed_base=1 ==> thu_pub=0 sat_pub=0 sun_base=0 fri_base=0 sat_base=0 #SUP: 5675 #CONF: 0.932621199671323
sun_base=0 sat_base=0 wed_base=1 ==> thu_pub=0 sat_pub=0 tue_base=0 thu_base=0 fri_base=0 #SUP: 5675 #CONF: 0.932621199671323
sun_base=0 fri_base=0 wed_base=1 ==> thu_pub=0 sat_pub=0 tue_base=0 thu_base=0 sat_base=0 #SUP: 5675 #CONF: 0.932621199671323
sun_base=0 thu_base=0 wed_base=1 ==> thu_pub=0 sat_pub=0 tue_base=0 fri_base=0 sat_base=0 #SUP: 5675 #CONF: 0.932621199671323
sun_base=0 tue_base=0 wed_base=1 ==> thu_pub=0 sat_pub=0 thu_base=0 fri_base=0 sat_base=0 #SUP: 5675 #CONF: 0.932621199671323
sat_pub=0 sat_base=0 wed_base=1 ==> thu_pub=0 sun_base=0 tue_base=0 thu_base=0 fri_base=0 #SUP: 5675 #CONF: 0.932621199671323
sat_pub=0 fri_base=0 wed_base=1 ==> thu_pub=0 sun_base=0 tue_base=0 thu_base=0 sat_base=0 #SUP: 5675 #CONF: 0.932621199671323
sat_pub=0 thu_base=0 wed_base=1 ==> thu_pub=0 sun_base=0 tue_base=0 fri_base=0 sat_base=0 #SUP: 5675 #CONF: 0.932621199671323
sat_pub=0 tue_base=0 wed_base=1 ==> thu_pub=0 sun_base=0 thu_base=0 fri_base=0 sat_base=0 #SUP: 5675 #CONF: 0.932621199671323
sat_pub=0 sun_base=0 wed_base=1 ==> thu_pub=0 tue_base=0 thu_base=0 fri_base=0 sat_base=0 #SUP: 5675 #CONF: 0.932621199671323
```

Figure 4.4: A snapshot of association rules generated in Facebook dataset after applying RP-Growth algorithm

In fig-4.4, in the first rule we have SUP:5675 that means it has support of 5675 transactions. Moroever, it has confidence values of 0.9326211.

In fig-4.5, in the first rule we have SUP:29473 that means it has support of 29473 transactions. Moreover, it has confidence values of 0.9902896310731806.

In fig-4.6, in the first rule we have SUP:22623 that means it has support of 22623



```
sun_pub=0 thu_base=0 ==> hrs=24 #SUP: 29473 #CONF: 0.9902896310731806
sun_pub=0 fri_base=0 ==> hrs=24 #SUP: 29558 #CONF: 0.9837254967218025
mon_base=0 sat_base=0 ==> hrs=24 #SUP: 28812 #CONF: 0.9737073335586347
sun_pub=0 wed_pub=0 ==> hrs=24 #SUP: 29256 #CONF: 0.9920651068158698
sun_pub=0 sat_base=0 ==> hrs=24 #SUP: 29228 #CONF: 0.9740718522962074
tue_base=0 thu_base=0 ==> hrs=24 #SUP: 28847 #CONF: 0.9900809994508512
tue_base=0 fri_base=0 ==> hrs=24 #SUP: 28932 #CONF: 0.9833792189252575
mon_pub=0 mon_base=0 ==> hrs=24 #SUP: 28830 #CONF: 0.9744473737578584
mon_pub=0 tue_base=0 ==> hrs=24 #SUP: 29801 #CONF: 0.9752593513761167
mon_pub=0 wed_base=0 ==> hrs=24 #SUP: 29309 #CONF: 0.9748544819557625
mon_pub=0 thu_base=0 ==> hrs=24 #SUP: 28896 #CONF: 0.9905388728918141
wed_pub=0 sat_base=0 ==> hrs=24 #SUP: 29476 #CONF: 0.9921238640188489
tue_pub=0 mon_base=0 ==> hrs=24 #SUP: 28845 #CONF: 0.9814228845564968
tue_pub=0 tue_base=0 ==> hrs=24 #SUP: 29530 #CONF: 0.9818459901582657
tue_pub=0 wed_base=0 ==> hrs=24 #SUP: 30330 #CONF: 0.9823163622230859
wed_base=0 wed_pub=0 ==> hrs=24 #SUP: 29196 #CONF: 0.9920489296636086
tue_pub=0 thu_base=0 ==> hrs=24 #SUP: 30376 #CONF: 0.9928095175839979
tue_pub=0 fri_base=0 ==> hrs=24 #SUP: 29736 #CONF: 0.9891557447940922
thu_base=0 wed_pub=0 ==> hrs=24 #SUP: 30532 #CONF: 0.9977125678060258
```

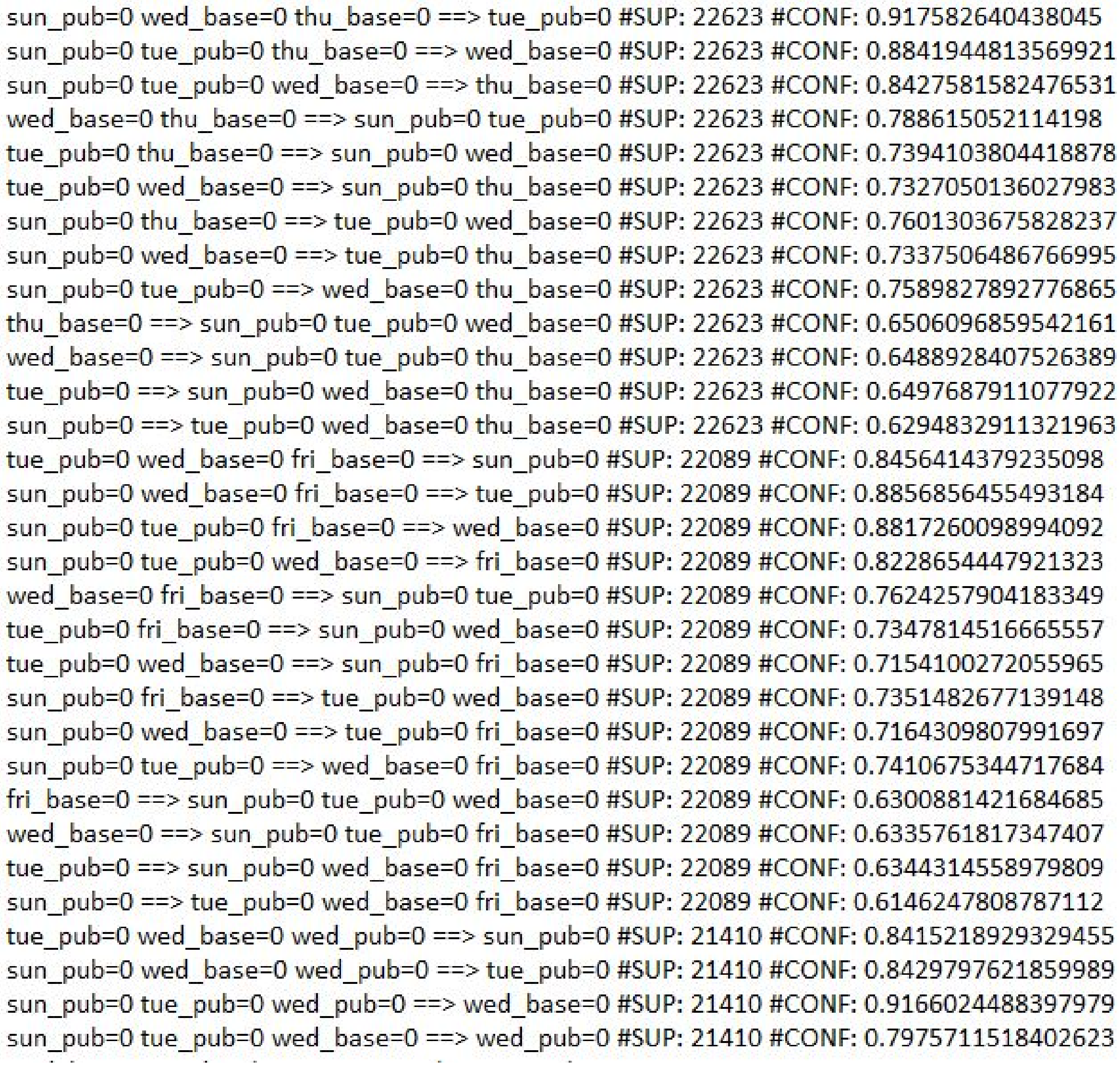
Figure 4.5: A snapshot of association rules generated in Facebook dataset after applying MNR algorithm

transactions. Moreover, it has confidence values of 0.917582640438045.

In fig-4.7, in the first rule we have SUP:5673 that means it has support of 5673 transactions. Moreover, it has confidence values of 0.8488702678437827.

In fig-4.8, in the first rule we have SUP:31185 that means it has support of 31185 transactions. Moreover, it has confidence values of 0.8872229650914677.

In fig-4.9,Consider the first line. It represents that items sun_pub and mon_pub are indirectly associated by the hrs_24 as mediator. Furthermore, it indicates that the support of sun_pub, mon_pub is 35161 transactions, the support of mon_pub,hrs_24 is 33337 transactions, the confidence of item sun_pub with respect to item hrs_24 is 0.978 and the confidence of item mon_pub with respect to item hrs_24 is 0.9777



```
sun_pub=0 wed_base=0 thu_base=0 ==> tue_pub=0 #SUP: 22623 #CONF: 0.917582640438045
sun_pub=0 tue_pub=0 thu_base=0 ==> wed_base=0 #SUP: 22623 #CONF: 0.8841944813569921
sun_pub=0 tue_pub=0 wed_base=0 ==> thu_base=0 #SUP: 22623 #CONF: 0.8427581582476531
wed_base=0 thu_base=0 ==> sun_pub=0 tue_pub=0 #SUP: 22623 #CONF: 0.788615052114198
tue_pub=0 thu_base=0 ==> sun_pub=0 wed_base=0 #SUP: 22623 #CONF: 0.7394103804418878
tue_pub=0 wed_base=0 ==> sun_pub=0 thu_base=0 #SUP: 22623 #CONF: 0.7327050136027983
sun_pub=0 thu_base=0 ==> tue_pub=0 wed_base=0 #SUP: 22623 #CONF: 0.7601303675828237
sun_pub=0 wed_base=0 ==> tue_pub=0 thu_base=0 #SUP: 22623 #CONF: 0.7337506486766995
sun_pub=0 tue_pub=0 ==> wed_base=0 thu_base=0 #SUP: 22623 #CONF: 0.7589827892776865
thu_base=0 ==> sun_pub=0 tue_pub=0 wed_base=0 #SUP: 22623 #CONF: 0.6506096859542161
wed_base=0 ==> sun_pub=0 tue_pub=0 thu_base=0 #SUP: 22623 #CONF: 0.6488928407526389
tue_pub=0 ==> sun_pub=0 wed_base=0 thu_base=0 #SUP: 22623 #CONF: 0.6497687911077922
sun_pub=0 ==> tue_pub=0 wed_base=0 thu_base=0 #SUP: 22623 #CONF: 0.6294832911321963
tue_pub=0 wed_base=0 fri_base=0 ==> sun_pub=0 #SUP: 22089 #CONF: 0.8456414379235098
sun_pub=0 wed_base=0 fri_base=0 ==> tue_pub=0 #SUP: 22089 #CONF: 0.8856856455493184
sun_pub=0 tue_pub=0 fri_base=0 ==> wed_base=0 #SUP: 22089 #CONF: 0.8817260098994092
sun_pub=0 tue_pub=0 wed_base=0 ==> fri_base=0 #SUP: 22089 #CONF: 0.8228654447921323
wed_base=0 fri_base=0 ==> sun_pub=0 tue_pub=0 #SUP: 22089 #CONF: 0.7624257904183349
tue_pub=0 fri_base=0 ==> sun_pub=0 wed_base=0 #SUP: 22089 #CONF: 0.7347814516665557
tue_pub=0 wed_base=0 ==> sun_pub=0 fri_base=0 #SUP: 22089 #CONF: 0.7154100272055965
sun_pub=0 fri_base=0 ==> tue_pub=0 wed_base=0 #SUP: 22089 #CONF: 0.7351482677139148
sun_pub=0 wed_base=0 ==> tue_pub=0 fri_base=0 #SUP: 22089 #CONF: 0.7164309807991697
sun_pub=0 tue_pub=0 ==> wed_base=0 fri_base=0 #SUP: 22089 #CONF: 0.7410675344717684
fri_base=0 ==> sun_pub=0 tue_pub=0 wed_base=0 #SUP: 22089 #CONF: 0.6300881421684685
wed_base=0 ==> sun_pub=0 tue_pub=0 fri_base=0 #SUP: 22089 #CONF: 0.6335761817347407
tue_pub=0 ==> sun_pub=0 wed_base=0 fri_base=0 #SUP: 22089 #CONF: 0.6344314558979809
sun_pub=0 ==> tue_pub=0 wed_base=0 fri_base=0 #SUP: 22089 #CONF: 0.6146247808787112
tue_pub=0 wed_base=0 wed_pub=0 ==> sun_pub=0 #SUP: 21410 #CONF: 0.8415218929329455
sun_pub=0 wed_base=0 wed_pub=0 ==> tue_pub=0 #SUP: 21410 #CONF: 0.8429797621859989
sun_pub=0 tue_pub=0 wed_pub=0 ==> wed_base=0 #SUP: 21410 #CONF: 0.9166024488397979
sun_pub=0 tue_pub=0 wed_base=0 ==> wed_pub=0 #SUP: 21410 #CONF: 0.7975711518402623
```

Figure 4.6: A snapshot of association rules generated in Facebook dataset after applying IGB Association rule mining algorithm

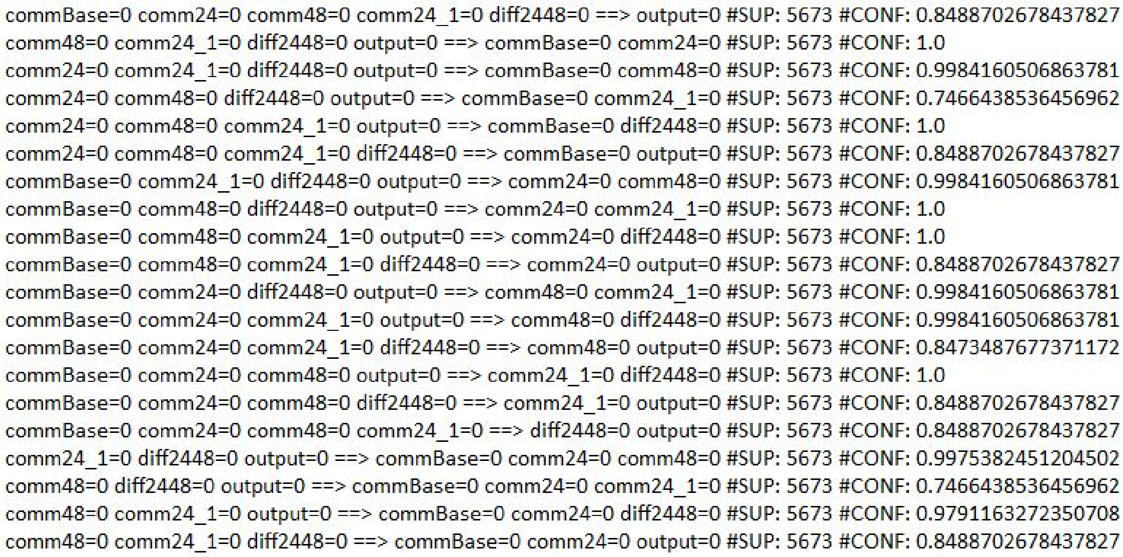
In fig-4.10, in the first rule we have SUP:25036 that means it has support of 25036 transactions. Moreover, it has confidence values of 0.93008395.

4.5 Evaluation of Performance

The main part of our thesis work is the Comparison and we can compare the algorithms bu evaluating their performance. We compared the algorithms in three different ways.

1) Changing all the values (K,min sup,minconf)

2) Confidence Value Fixed



```
commBase=0 comm24=0 comm48=0 comm24_1=0 diff2448=0 ==> output=0 #SUP: 5673 #CONF: 0.8488702678437827
comm48=0 comm24_1=0 diff2448=0 output=0 ==> commBase=0 comm24=0 #SUP: 5673 #CONF: 1.0
comm24=0 comm24_1=0 diff2448=0 output=0 ==> commBase=0 comm48=0 #SUP: 5673 #CONF: 0.9984160506863781
comm24=0 comm48=0 diff2448=0 output=0 ==> commBase=0 comm24_1=0 #SUP: 5673 #CONF: 0.7466438536456962
comm24=0 comm48=0 comm24_1=0 output=0 ==> commBase=0 diff2448=0 #SUP: 5673 #CONF: 1.0
comm24=0 comm48=0 comm24_1=0 diff2448=0 ==> commBase=0 output=0 #SUP: 5673 #CONF: 0.8488702678437827
commBase=0 comm24_1=0 diff2448=0 output=0 ==> comm24=0 comm48=0 #SUP: 5673 #CONF: 0.9984160506863781
commBase=0 comm48=0 diff2448=0 output=0 ==> comm24=0 comm24_1=0 #SUP: 5673 #CONF: 1.0
commBase=0 comm48=0 comm24_1=0 output=0 ==> comm24=0 diff2448=0 #SUP: 5673 #CONF: 1.0
commBase=0 comm48=0 comm24_1=0 diff2448=0 ==> comm24=0 output=0 #SUP: 5673 #CONF: 0.8488702678437827
commBase=0 comm24=0 diff2448=0 output=0 ==> comm48=0 comm24_1=0 #SUP: 5673 #CONF: 0.9984160506863781
commBase=0 comm24=0 comm24_1=0 output=0 ==> comm48=0 diff2448=0 #SUP: 5673 #CONF: 0.9984160506863781
commBase=0 comm24=0 comm24_1=0 diff2448=0 ==> comm48=0 output=0 #SUP: 5673 #CONF: 0.8473487677371172
commBase=0 comm24=0 comm48=0 output=0 ==> comm24_1=0 diff2448=0 #SUP: 5673 #CONF: 1.0
commBase=0 comm24=0 comm48=0 diff2448=0 ==> comm24_1=0 output=0 #SUP: 5673 #CONF: 0.8488702678437827
commBase=0 comm24=0 comm48=0 comm24_1=0 ==> diff2448=0 output=0 #SUP: 5673 #CONF: 0.8488702678437827
comm24_1=0 diff2448=0 output=0 ==> commBase=0 comm24=0 comm48=0 #SUP: 5673 #CONF: 0.9975382451204502
comm48=0 diff2448=0 output=0 ==> commBase=0 comm24=0 comm24_1=0 #SUP: 5673 #CONF: 0.7466438536456962
comm48=0 comm24_1=0 output=0 ==> commBase=0 comm24=0 diff2448=0 #SUP: 5673 #CONF: 0.9791163272350708
comm48=0 comm24_1=0 diff2448=0 ==> commBase=0 comm24=0 output=0 #SUP: 5673 #CONF: 0.8488702678437827
```

Figure 4.7: A snapshot of association rules generated in Facebook dataset after applying Sporadic Association Rule mining algorithm

3) Time taken to generate Association rules

4.5.1 Changing all the values

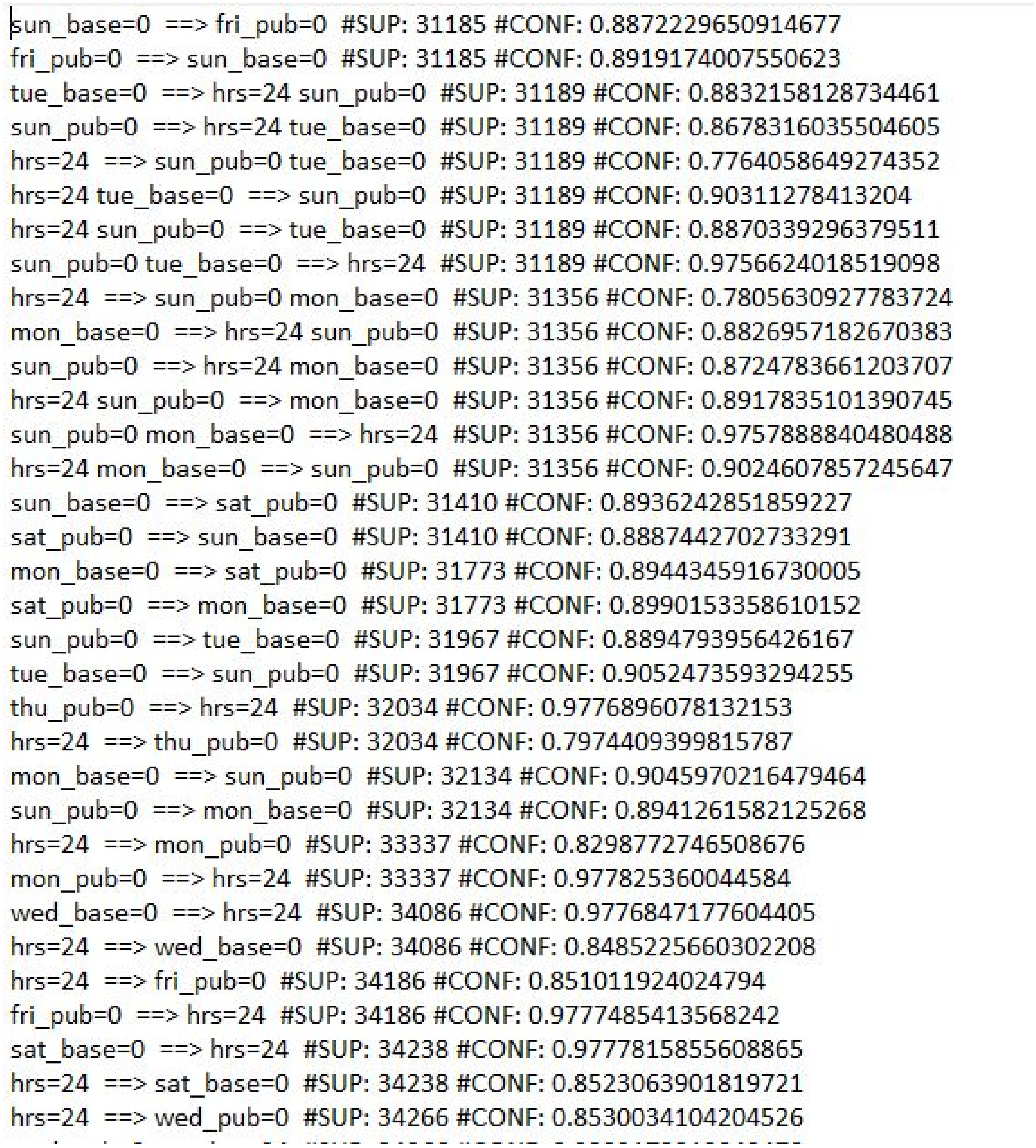
Apriori

In this algorithm we have take 0.3 and 0.4 as support and 0.4 and 0.5 as the confidence. At first we applied (0.3,0.4), means the support is 0.3 and the confidence is 0.4. so after applying the algorithm in two different ways in five datasets we can see the clear comparison here :

We can see that the association rules generated for the breast cancer dataset is low because the number of instances is not much in this dataset. Though having a low instance, the zoo dataset is generating high amount of association rules in small value of support and confidence. But when we increased the support and confidence value, we can see the drastically fall of number of associations rules. Moreover, we can see this type of behavior in the Facebook dataset also.

FP-Growth

If we can look down in the below graph, we can see that the results are same as the apriori algorithm in terms of number of association rules generated from the datasets.In Figure 4.12, we also took the 0.3 and 0.4 as support and 0.4 and 0.5



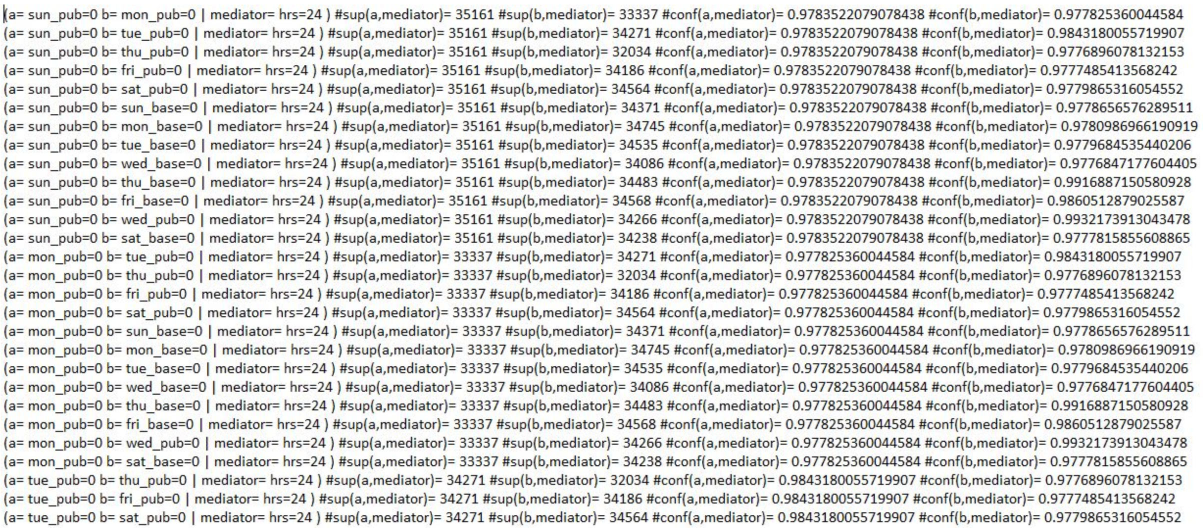
```
sun_base=0  ==> fri_pub=0  #SUP: 31185 #CONF: 0.8872229650914677
fri_pub=0  ==> sun_base=0  #SUP: 31185 #CONF: 0.8919174007550623
tue_base=0  ==> hrs=24 sun_pub=0  #SUP: 31189 #CONF: 0.8832158128734461
sun_pub=0  ==> hrs=24 tue_base=0  #SUP: 31189 #CONF: 0.8678316035504605
hrs=24  ==> sun_pub=0 tue_base=0  #SUP: 31189 #CONF: 0.7764058649274352
hrs=24 tue_base=0  ==> sun_pub=0  #SUP: 31189 #CONF: 0.90311278413204
hrs=24 sun_pub=0  ==> tue_base=0  #SUP: 31189 #CONF: 0.8870339296379511
sun_pub=0 tue_base=0  ==> hrs=24  #SUP: 31189 #CONF: 0.9756624018519098
hrs=24  ==> sun_pub=0 mon_base=0  #SUP: 31356 #CONF: 0.7805630927783724
mon_base=0  ==> hrs=24 sun_pub=0  #SUP: 31356 #CONF: 0.8826957182670383
sun_pub=0  ==> hrs=24 mon_base=0  #SUP: 31356 #CONF: 0.8724783661203707
hrs=24 sun_pub=0  ==> mon_base=0  #SUP: 31356 #CONF: 0.8917835101390745
sun_pub=0 mon_base=0  ==> hrs=24  #SUP: 31356 #CONF: 0.9757888840480488
hrs=24 mon_base=0  ==> sun_pub=0  #SUP: 31356 #CONF: 0.9024607857245647
sun_base=0  ==> sat_pub=0  #SUP: 31410 #CONF: 0.8936242851859227
sat_pub=0  ==> sun_base=0  #SUP: 31410 #CONF: 0.8887442702733291
mon_base=0  ==> sat_pub=0  #SUP: 31773 #CONF: 0.8944345916730005
sat_pub=0  ==> mon_base=0  #SUP: 31773 #CONF: 0.8990153358610152
sun_pub=0  ==> tue_base=0  #SUP: 31967 #CONF: 0.8894793956426167
tue_base=0  ==> sun_pub=0  #SUP: 31967 #CONF: 0.9052473593294255
thu_pub=0  ==> hrs=24  #SUP: 32034 #CONF: 0.9776896078132153
hrs=24  ==> thu_pub=0  #SUP: 32034 #CONF: 0.7974409399815787
mon_base=0  ==> sun_pub=0  #SUP: 32134 #CONF: 0.9045970216479464
sun_pub=0  ==> mon_base=0  #SUP: 32134 #CONF: 0.8941261582125268
hrs=24  ==> mon_pub=0  #SUP: 33337 #CONF: 0.8298772746508676
mon_pub=0  ==> hrs=24  #SUP: 33337 #CONF: 0.977825360044584
wed_base=0  ==> hrs=24  #SUP: 34086 #CONF: 0.9776847177604405
hrs=24  ==> wed_base=0  #SUP: 34086 #CONF: 0.8485225660302208
hrs=24  ==> fri_pub=0  #SUP: 34186 #CONF: 0.851011924024794
fri_pub=0  ==> hrs=24  #SUP: 34186 #CONF: 0.9777485413568242
sat_base=0  ==> hrs=24  #SUP: 34238 #CONF: 0.9777815855608865
hrs=24  ==> sat_base=0  #SUP: 34238 #CONF: 0.8523063901819721
hrs=24  ==> wed_pub=0  #SUP: 34266 #CONF: 0.8530034104204526
```

Figure 4.8: A snapshot of association rules generated in Facebook dataset after applying TOP-K Association Rule Mining algorithm

as the confidence. At first we applied (0.3,0.4), means the support is 0.3 and the confidence is 0.4. But the Fp-growth is faster than the apriori algorithm.

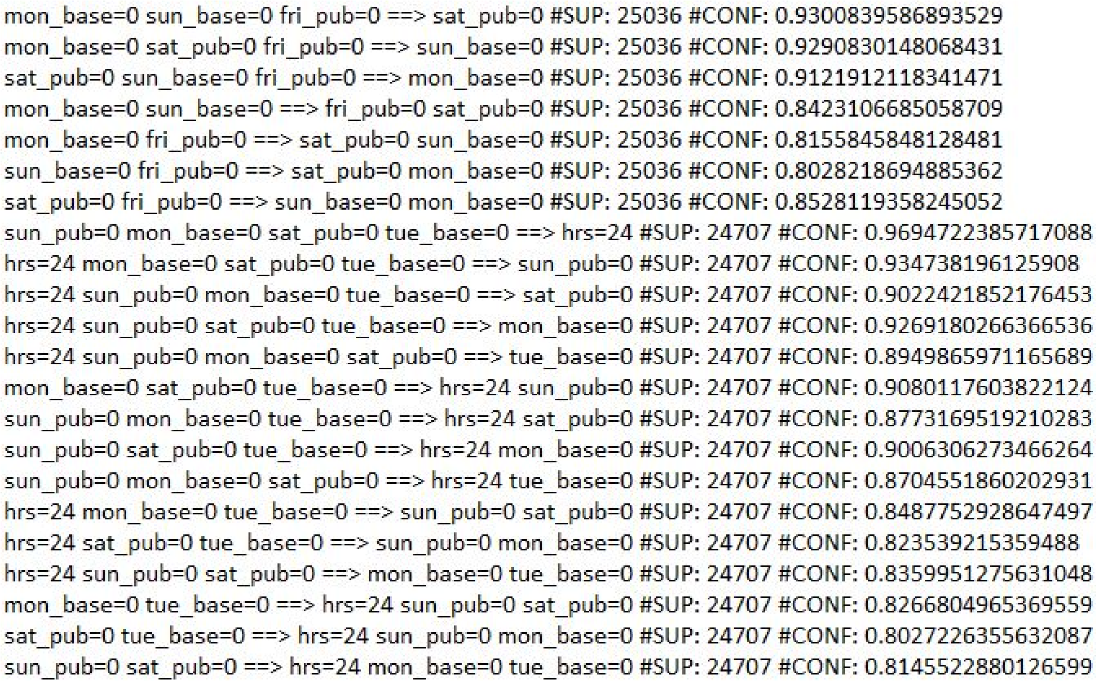
FP-Growth with lift

In Figure 4.13, we have take 0.2 and 0.3 as support,0.3 and 0.4 as the confidence and lift as 0.3,0.4. At first we applied (0.2,0.3,0.3), means the support is 0.2,the confidence is 0.3 and the lift is also 0.3. so after applying the algorithm in two different ways in five datasets we can see the clear comparison that in lower values



```
(a= sun_pub=0 b= mon_pub=0 | mediator= hrs=24 ) #sup(a,mediator)= 35161 #sup(b,mediator)= 33337 #conf(a,mediator)= 0.9783522079078438 #conf(b,mediator)= 0.977825360044584
(a= sun_pub=0 b= tue_pub=0 | mediator= hrs=24 ) #sup(a,mediator)= 35161 #sup(b,mediator)= 34271 #conf(a,mediator)= 0.9783522079078438 #conf(b,mediator)= 0.9843180055719907
(a= sun_pub=0 b= thu_pub=0 | mediator= hrs=24 ) #sup(a,mediator)= 35161 #sup(b,mediator)= 32034 #conf(a,mediator)= 0.9783522079078438 #conf(b,mediator)= 0.9776896078132153
(a= sun_pub=0 b= fri_pub=0 | mediator= hrs=24 ) #sup(a,mediator)= 35161 #sup(b,mediator)= 34186 #conf(a,mediator)= 0.9783522079078438 #conf(b,mediator)= 0.9777485413568242
(a= sun_pub=0 b= sat_pub=0 | mediator= hrs=24 ) #sup(a,mediator)= 35161 #sup(b,mediator)= 34564 #conf(a,mediator)= 0.9783522079078438 #conf(b,mediator)= 0.9779865316054552
(a= sun_pub=0 b= sun_base=0 | mediator= hrs=24 ) #sup(a,mediator)= 35161 #sup(b,mediator)= 34371 #conf(a,mediator)= 0.9783522079078438 #conf(b,mediator)= 0.9778656576289511
(a= sun_pub=0 b= mon_base=0 | mediator= hrs=24 ) #sup(a,mediator)= 35161 #sup(b,mediator)= 34745 #conf(a,mediator)= 0.9783522079078438 #conf(b,mediator)= 0.9780986966190919
(a= sun_pub=0 b= tue_base=0 | mediator= hrs=24 ) #sup(a,mediator)= 35161 #sup(b,mediator)= 34535 #conf(a,mediator)= 0.9783522079078438 #conf(b,mediator)= 0.9779684535440206
(a= sun_pub=0 b= wed_base=0 | mediator= hrs=24 ) #sup(a,mediator)= 35161 #sup(b,mediator)= 34086 #conf(a,mediator)= 0.9783522079078438 #conf(b,mediator)= 0.9776847177604405
(a= sun_pub=0 b= thu_base=0 | mediator= hrs=24 ) #sup(a,mediator)= 35161 #sup(b,mediator)= 34483 #conf(a,mediator)= 0.9783522079078438 #conf(b,mediator)= 0.9916887150580928
(a= sun_pub=0 b= fri_base=0 | mediator= hrs=24 ) #sup(a,mediator)= 35161 #sup(b,mediator)= 34568 #conf(a,mediator)= 0.9783522079078438 #conf(b,mediator)= 0.9860512879025587
(a= sun_pub=0 b= wed_pub=0 | mediator= hrs=24 ) #sup(a,mediator)= 35161 #sup(b,mediator)= 34266 #conf(a,mediator)= 0.9783522079078438 #conf(b,mediator)= 0.9932173913043478
(a= sun_pub=0 b= sat_base=0 | mediator= hrs=24 ) #sup(a,mediator)= 35161 #sup(b,mediator)= 34238 #conf(a,mediator)= 0.9783522079078438 #conf(b,mediator)= 0.9777815855608865
(a= mon_pub=0 b= tue_pub=0 | mediator= hrs=24 ) #sup(a,mediator)= 33337 #sup(b,mediator)= 34271 #conf(a,mediator)= 0.977825360044584 #conf(b,mediator)= 0.9843180055719907
(a= mon_pub=0 b= thu_pub=0 | mediator= hrs=24 ) #sup(a,mediator)= 33337 #sup(b,mediator)= 32034 #conf(a,mediator)= 0.977825360044584 #conf(b,mediator)= 0.9776896078132153
(a= mon_pub=0 b= fri_pub=0 | mediator= hrs=24 ) #sup(a,mediator)= 33337 #sup(b,mediator)= 34186 #conf(a,mediator)= 0.977825360044584 #conf(b,mediator)= 0.9777485413568242
(a= mon_pub=0 b= sat_pub=0 | mediator= hrs=24 ) #sup(a,mediator)= 33337 #sup(b,mediator)= 34564 #conf(a,mediator)= 0.977825360044584 #conf(b,mediator)= 0.9779865316054552
(a= mon_pub=0 b= sun_base=0 | mediator= hrs=24 ) #sup(a,mediator)= 33337 #sup(b,mediator)= 34371 #conf(a,mediator)= 0.977825360044584 #conf(b,mediator)= 0.9778656576289511
(a= mon_pub=0 b= mon_base=0 | mediator= hrs=24 ) #sup(a,mediator)= 33337 #sup(b,mediator)= 34745 #conf(a,mediator)= 0.977825360044584 #conf(b,mediator)= 0.9780986966190919
(a= mon_pub=0 b= tue_base=0 | mediator= hrs=24 ) #sup(a,mediator)= 33337 #sup(b,mediator)= 34535 #conf(a,mediator)= 0.977825360044584 #conf(b,mediator)= 0.9779684535440206
(a= mon_pub=0 b= wed_base=0 | mediator= hrs=24 ) #sup(a,mediator)= 33337 #sup(b,mediator)= 34086 #conf(a,mediator)= 0.977825360044584 #conf(b,mediator)= 0.9776847177604405
(a= mon_pub=0 b= thu_base=0 | mediator= hrs=24 ) #sup(a,mediator)= 33337 #sup(b,mediator)= 34483 #conf(a,mediator)= 0.977825360044584 #conf(b,mediator)= 0.9916887150580928
(a= mon_pub=0 b= fri_base=0 | mediator= hrs=24 ) #sup(a,mediator)= 33337 #sup(b,mediator)= 34568 #conf(a,mediator)= 0.977825360044584 #conf(b,mediator)= 0.9860512879025587
(a= mon_pub=0 b= wed_pub=0 | mediator= hrs=24 ) #sup(a,mediator)= 33337 #sup(b,mediator)= 34266 #conf(a,mediator)= 0.977825360044584 #conf(b,mediator)= 0.9932173913043478
(a= mon_pub=0 b= sat_base=0 | mediator= hrs=24 ) #sup(a,mediator)= 33337 #sup(b,mediator)= 34238 #conf(a,mediator)= 0.977825360044584 #conf(b,mediator)= 0.9777815855608865
(a= tue_pub=0 b= thu_pub=0 | mediator= hrs=24 ) #sup(a,mediator)= 34271 #sup(b,mediator)= 32034 #conf(a,mediator)= 0.9843180055719907 #conf(b,mediator)= 0.9776896078132153
(a= tue_pub=0 b= fri_pub=0 | mediator= hrs=24 ) #sup(a,mediator)= 34271 #sup(b,mediator)= 34186 #conf(a,mediator)= 0.9843180055719907 #conf(b,mediator)= 0.9777485413568242
(a= tue_pub=0 b= sat_pub=0 | mediator= hrs=24 ) #sup(a,mediator)= 34271 #sup(b,mediator)= 34564 #conf(a,mediator)= 0.9843180055719907 #conf(b,mediator)= 0.9779865316054552
```

Figure 4.9: A snapshot of association rules generated in Facebook dataset after applying Indirect association rule mining algorithm



```
mon_base=0 sun_base=0 fri_pub=0 ==> sat_pub=0 #SUP: 25036 #CONF: 0.9300839586893529
mon_base=0 sat_pub=0 fri_pub=0 ==> sun_base=0 #SUP: 25036 #CONF: 0.9290830148068431
sat_pub=0 sun_base=0 fri_pub=0 ==> mon_base=0 #SUP: 25036 #CONF: 0.9121912118341471
mon_base=0 sun_base=0 ==> fri_pub=0 sat_pub=0 #SUP: 25036 #CONF: 0.8423106685058709
mon_base=0 fri_pub=0 ==> sat_pub=0 sun_base=0 #SUP: 25036 #CONF: 0.8155845848128481
sun_base=0 fri_pub=0 ==> sat_pub=0 mon_base=0 #SUP: 25036 #CONF: 0.8028218694885362
sat_pub=0 fri_pub=0 ==> sun_base=0 mon_base=0 #SUP: 25036 #CONF: 0.8528119358245052
sun_pub=0 mon_base=0 sat_pub=0 tue_base=0 ==> hrs=24 #SUP: 24707 #CONF: 0.9694722385717088
hrs=24 mon_base=0 sat_pub=0 tue_base=0 ==> sun_pub=0 #SUP: 24707 #CONF: 0.934738196125908
hrs=24 sun_pub=0 mon_base=0 tue_base=0 ==> sat_pub=0 #SUP: 24707 #CONF: 0.9022421852176453
hrs=24 sun_pub=0 sat_pub=0 tue_base=0 ==> mon_base=0 #SUP: 24707 #CONF: 0.9269180266366536
hrs=24 sun_pub=0 mon_base=0 sat_pub=0 ==> tue_base=0 #SUP: 24707 #CONF: 0.8949865971165689
mon_base=0 sat_pub=0 tue_base=0 ==> hrs=24 sun_pub=0 #SUP: 24707 #CONF: 0.9080117603822124
sun_pub=0 mon_base=0 tue_base=0 ==> hrs=24 sat_pub=0 #SUP: 24707 #CONF: 0.8773169519210283
sun_pub=0 sat_pub=0 tue_base=0 ==> hrs=24 mon_base=0 #SUP: 24707 #CONF: 0.9006306273466264
sun_pub=0 mon_base=0 sat_pub=0 ==> hrs=24 tue_base=0 #SUP: 24707 #CONF: 0.8704551860202931
hrs=24 mon_base=0 tue_base=0 ==> sun_pub=0 sat_pub=0 #SUP: 24707 #CONF: 0.8487752928647497
hrs=24 sat_pub=0 tue_base=0 ==> sun_pub=0 mon_base=0 #SUP: 24707 #CONF: 0.823539215359488
hrs=24 sun_pub=0 sat_pub=0 ==> mon_base=0 tue_base=0 #SUP: 24707 #CONF: 0.8359951275631048
mon_base=0 tue_base=0 ==> hrs=24 sun_pub=0 sat_pub=0 #SUP: 24707 #CONF: 0.8266804965369559
sat_pub=0 tue_base=0 ==> hrs=24 sun_pub=0 mon_base=0 #SUP: 24707 #CONF: 0.8027226355632087
sun_pub=0 sat_pub=0 ==> hrs=24 mon_base=0 tue_base=0 #SUP: 24707 #CONF: 0.8145522880126599
```

Figure 4.10: A snapshot of association rules generated in Facebook dataset after applying FP-close algorithm

the zoo and transaction datasets produce much more association rules.But if we increase the values, the amount of association rules is drastically falls in both the datasets.

FP-CLOSE

In Figure 4.14, we have take 0.2 and 0.3 as minimum support and 0.3 and 0.4 as the minimum confidence. At first we applied (0.2,0.3), means the minimum

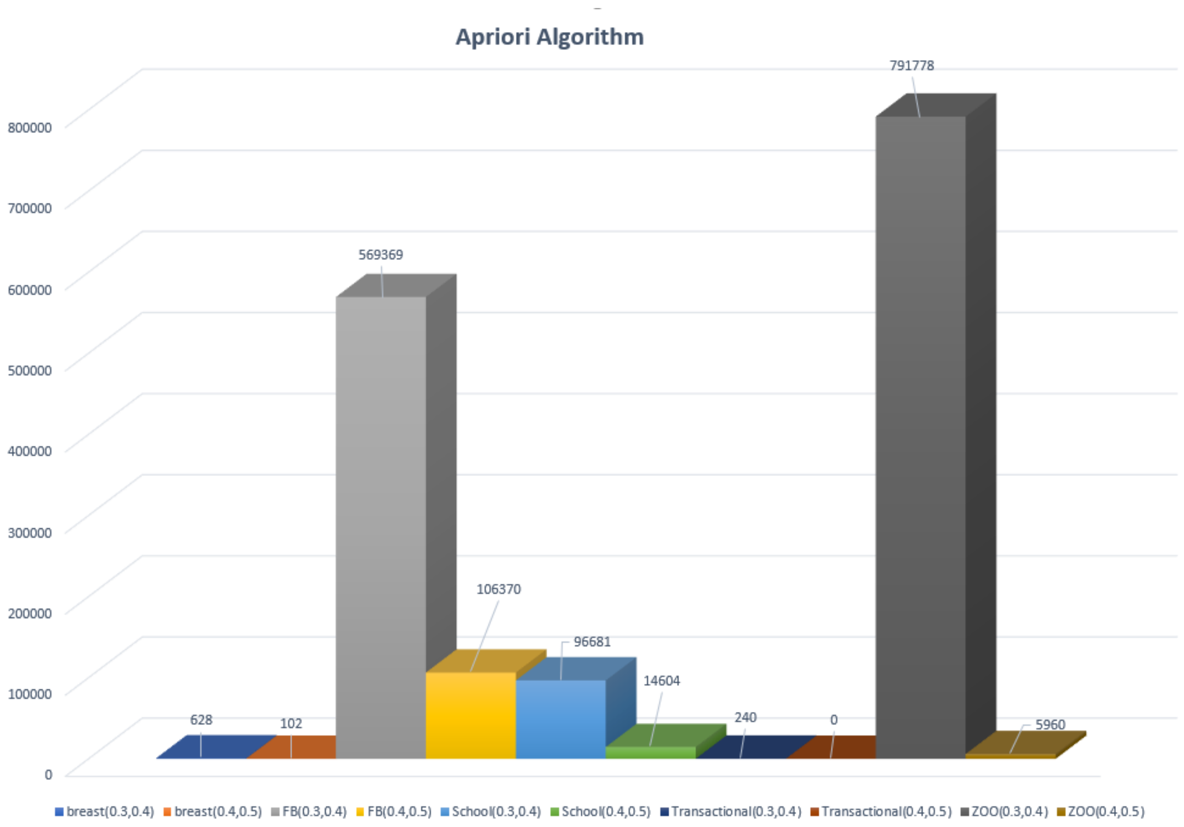


Figure 4.11: Association rules generated in different datasets for applying Apriori algorithm

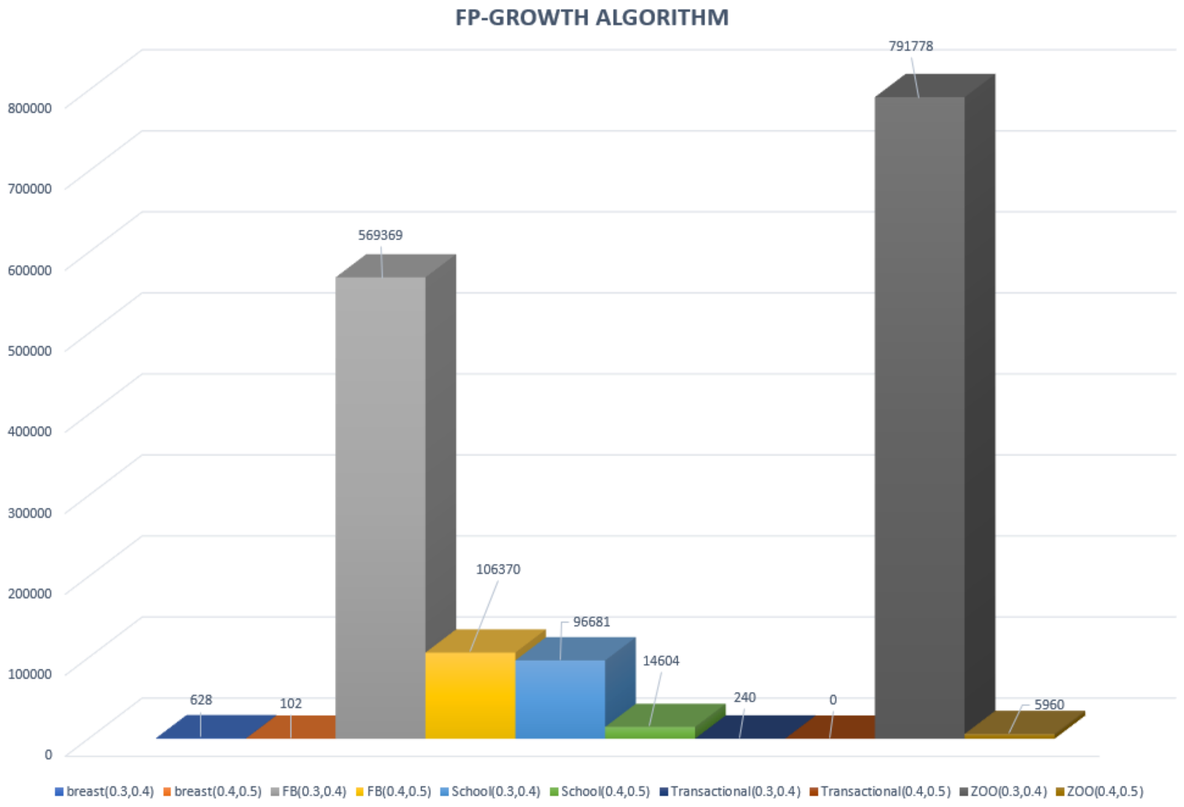


Figure 4.12: Association Rules Generated in Different Datasets for Applying FP-Growth Algorithm

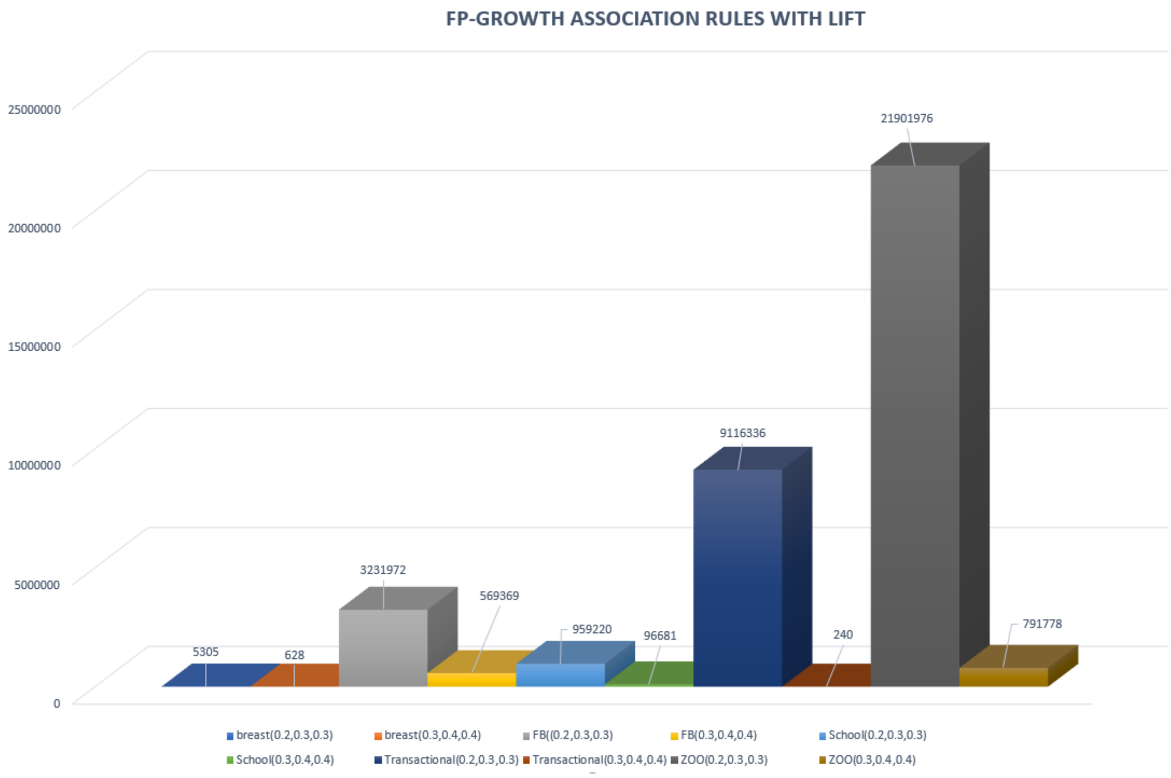


Figure 4.13: Association Rules Generated in Different Datasets for Applying FP-Growth with lift Algorithm

support is 0.2 and the minimum confidence is 0.3. so after applying the algorithm in two different ways in five datasets we can see that we have lots of closed association rules in FB dataset. Moreover, This algorithm shows similar kind of behaviour in four datasets like ZOO,Transactions,School and FB. Little amount of increase in values costs drastically fall in association rules generated.

RP-Growth

In Figure 4.15. we have take 0.8 and 0.9 as minimum support,0.2 as the minimum rare support and Minimum confidence as 0.9,1. At first we applied (0.8,0.2,0.9), means the minimum support is 0.8,the Minimum rare support is 0.2 and the minimum confidence is 0.9. So, after applying the algortihm in two different ways in five datasets we can see that in lower values the zoo and transactional datasets produce much more association rules.But if we increase the values, the amount of assoiation rules is falls in a small volume in both the datasets. So, finding we found lots of rare patterns in both the Social media and Genetical Datasets.

Sporadic Association Rule

In figure 4.16, we have take 0.2 and 0.3 as minimum support,0.3 and 0.4 as

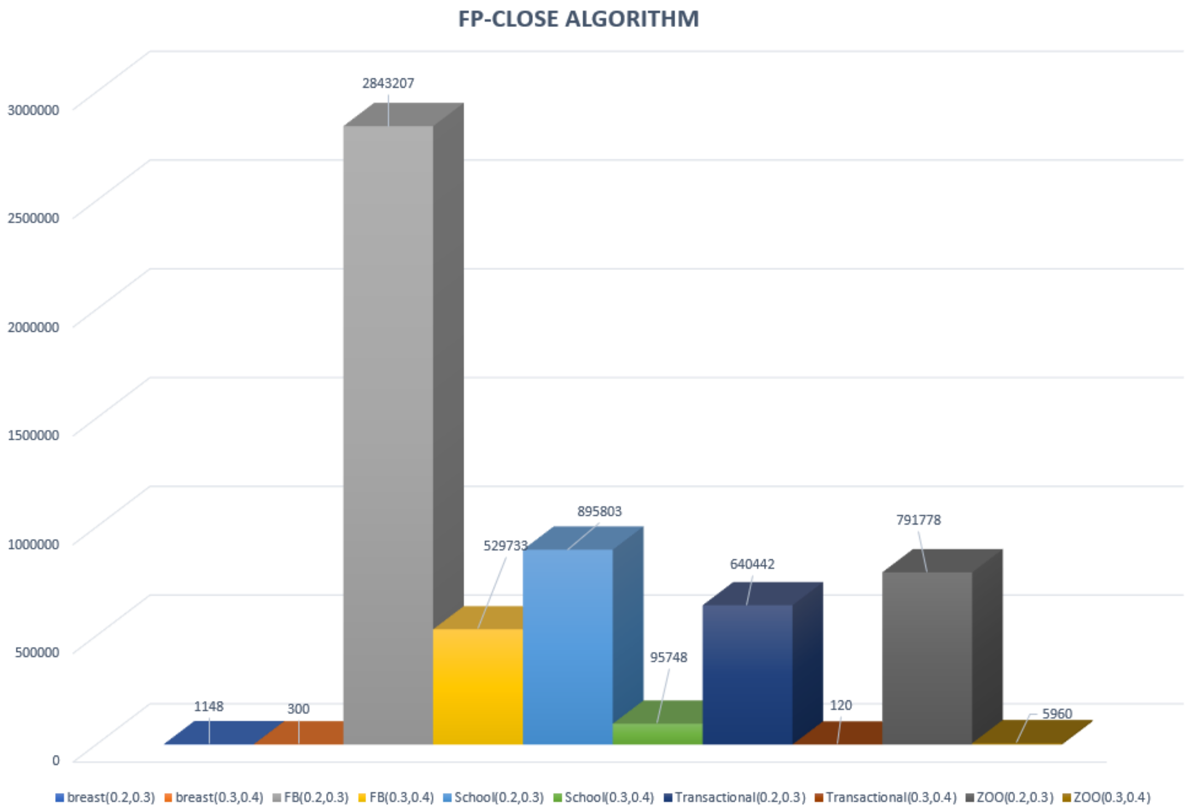


Figure 4.14: Association Rules Generated in Different Datasets for Applying FP-Close Algorithm

the maximum support and Minimum confidence as 0.3,0.4. At first we applied (0.2,0.3,0.3), means the minimum support is 0.2,the maximum support is 0.3 and the minimum confidence is 0.3. So, after applying the algortihm in two different ways in five datasets we can see that in lower values all the datasets except the transactional dataset ,giving less amount of sporadic association rules. But increasing the values will give less amount of association rules. So,transactional datsabases have higher amount of sporadic association rules.

Top-K Association Rule Mining Algorithm

In figure 4.17, instead of the minimum support we introduced a new term as the K. Which is not a float type and it's an integer. In this algorithm we have take 2 and 3 as the value of K and 0.2 and 0.3 as the minimum confidence. At first we applied (2,0.3), means the K value is 2 and the confidence is 0.2. so after applying the algortihm in two different ways in five datasets we can see that an unusual occurance happened. Increasing the values of wont't give you less amount of Top-K association rules. Moreover, we have a lot more association rules in the breast cancer,school and Zoo datasets. And incresing the values also kept the

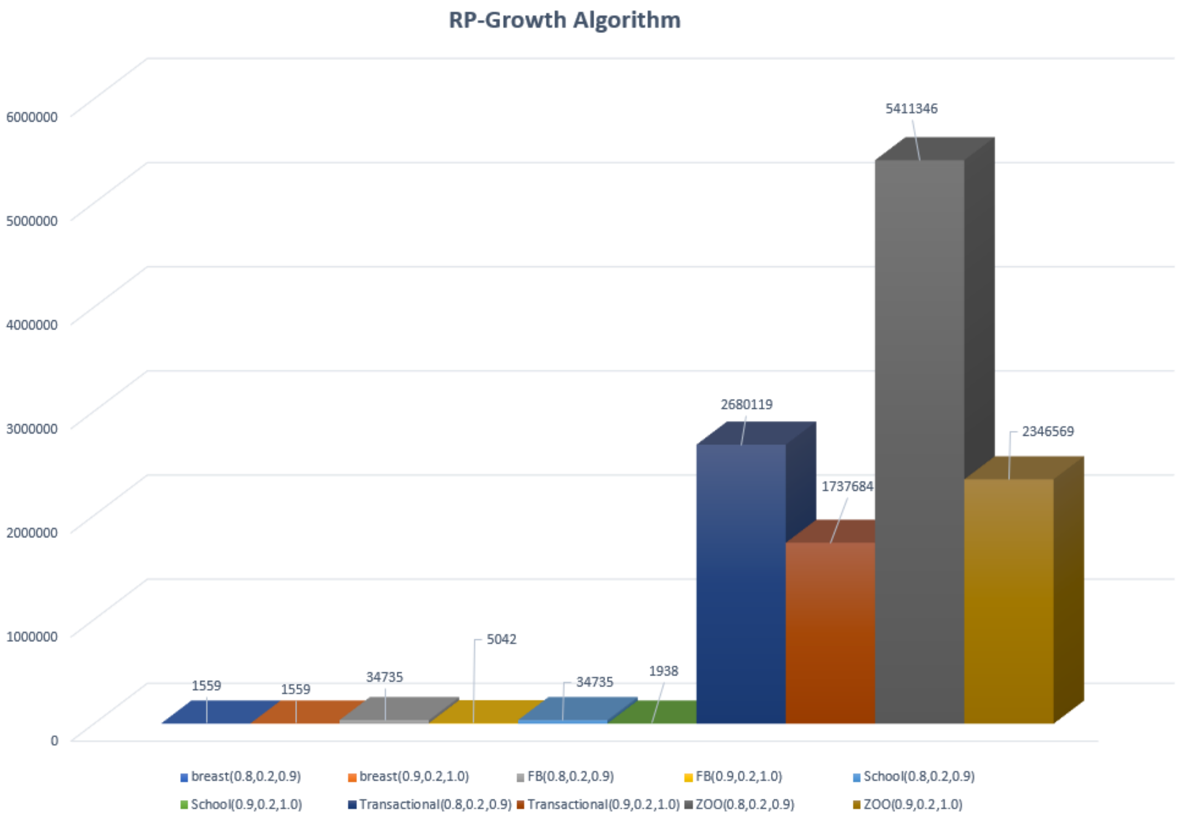


Figure 4.15: Association Rules Generated in Different Datasets for Applying RP-GROWTH Algorithm

number of association rules same in transactional dataset.

MNR (Minimal Non-Redundant) Association Rule Mining Algorithm: In Figure 4.18, we have take 0.2 and 0.3 as minimum support and 0.3 and 0.4 as the minimum confidence. At first we applied (0.2,0.3), means the support is 0.2 and the confidence is 0.3. So after applying the algortihm in two different ways in five datasets we can see the uusual behaviour of algorithms. Increasing the number of values will decrease the number of association rules genrated from the datasets. But this algorithm gives better result in facebook and Trnsactional dataset in small values.

Indirect Association Rule mining Algorithm : In figure 4.19, we have take 0.4 and 0.5 as mediator minimum support,0.5 and 0.6 as the itempair minimum support and minimum confidence as 0.5,0.6. At first we applied (0.4,0.5,0.5), means the mediator minimum support is 0.4,the itempair minimum support is 0.5 and the minimum confidence is 0.5. So, after applying

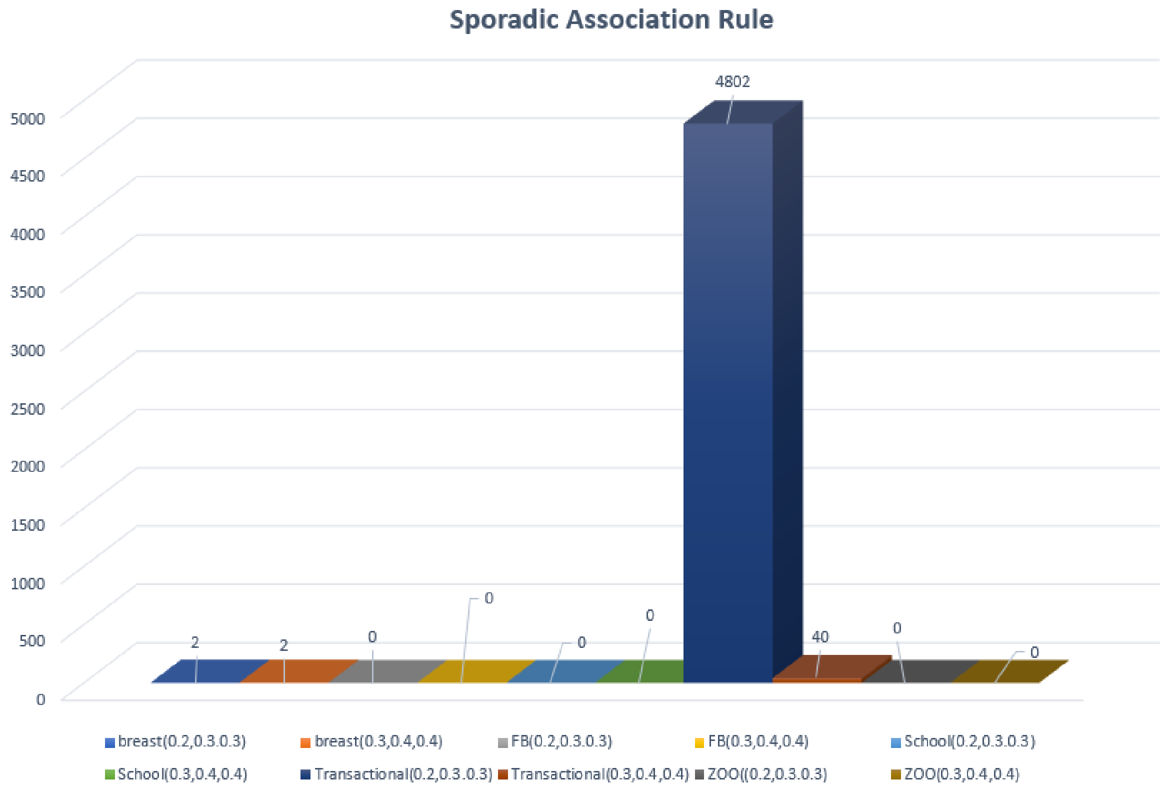


Figure 4.16: Association Rules Generated in Different Datasets for Applying Sporadic Association Rule mining algorithm

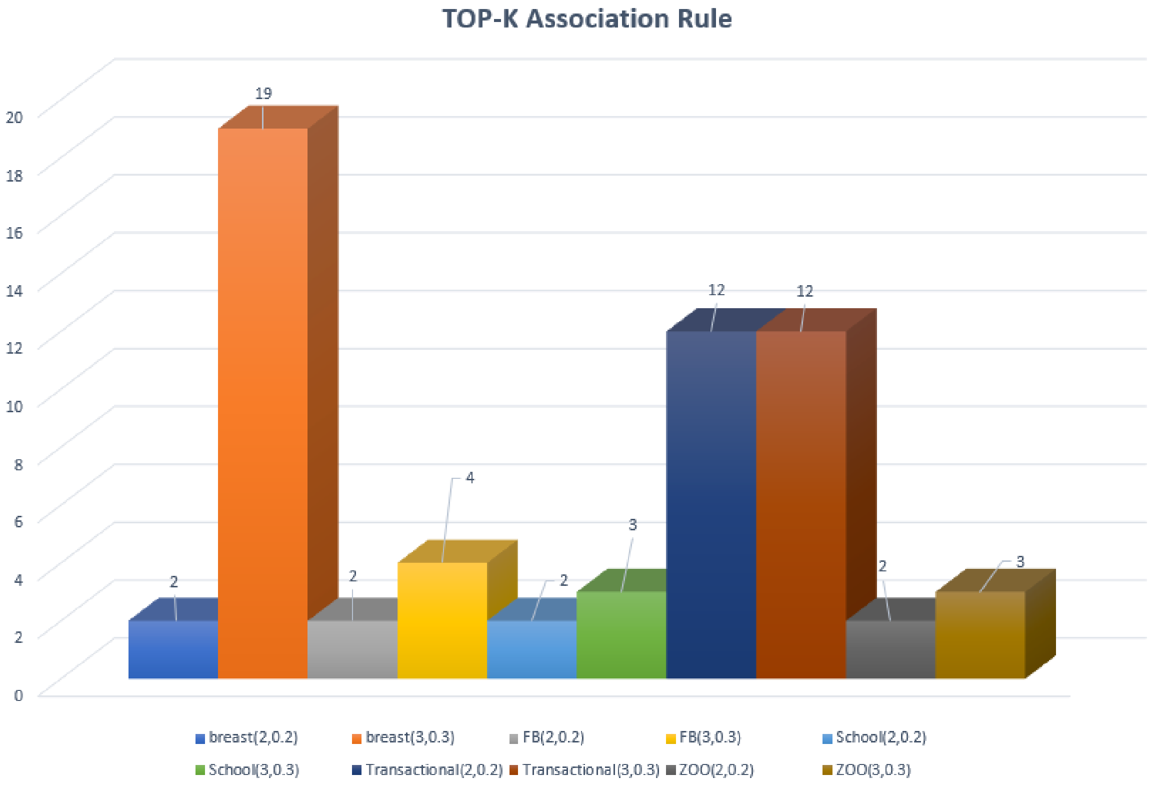


Figure 4.17: Association Rules Generated in Different Datasets for Applying Top-K Association Rule Mining Algorithm

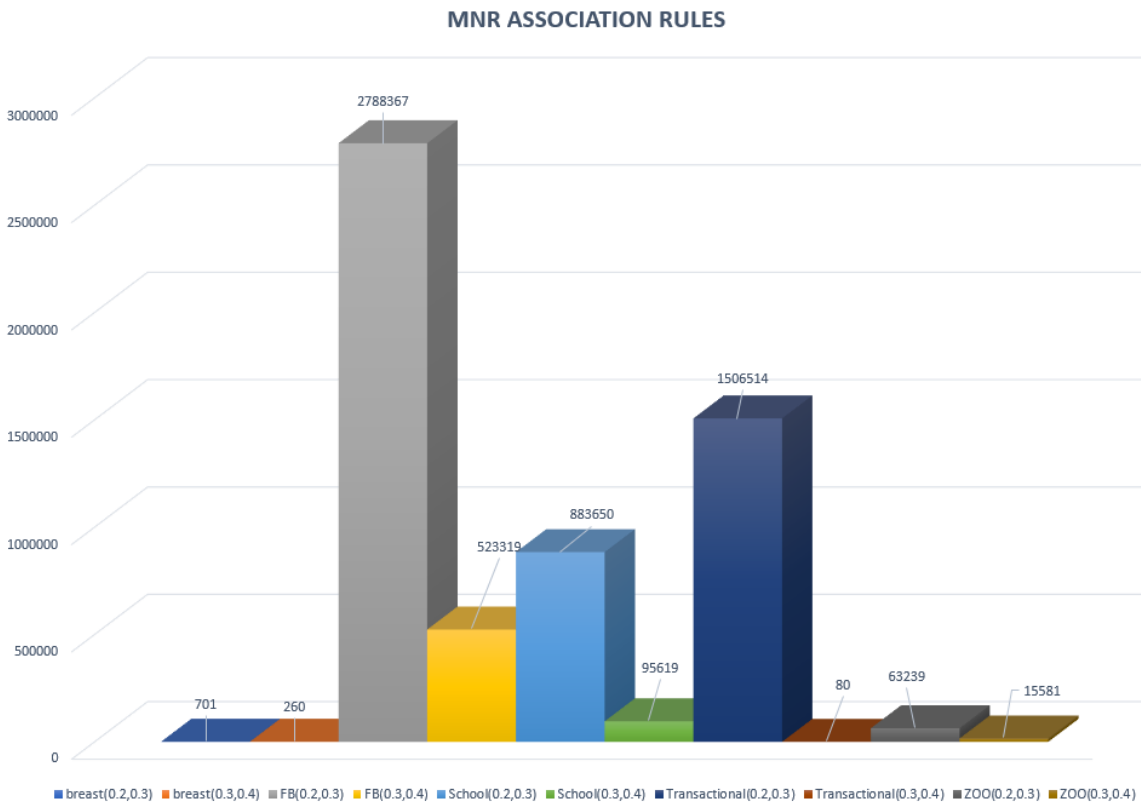


Figure 4.18: Association Rules Generated in Different Datasets for Applying MNR Association Rule Mining Algorithm

the algortihm in two different ways in five datasets we can see that in lower values all the datasets except the transactional dataset ,giving handsome amount of indirect association rules. But increasing the values will give less amount of association rules. But,we can see that transactional datsabases have no sporadic association rules.

IGB (Informative and Generic Basis) Association Rule Mining Algorithm :

In this algorithm we have take 0.3 and 0.4 as support and 0.4 and 0.5 as the confidence. At first we applied (0.3,0.4), means the support is 0.3 and the confidence is 0.4. so after applying the algortihm in two different ways in five datasets we can see that this algorithm is doing very well in school and facebook datasets. But failed to generate good amount of association rules in Transactional database. Moreover, the behaviour of this algorithm is same like others in other four different dataset amd we can see that in fig-4.20.

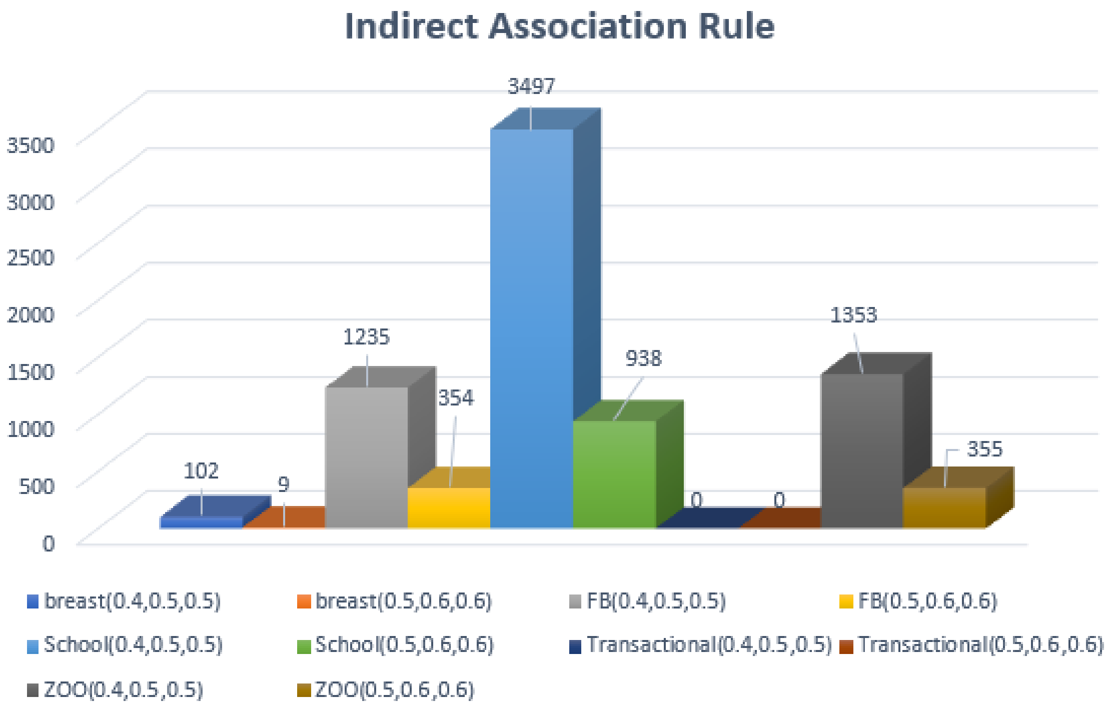


Figure 4.19: Association Rules Generated in Different Datasets for Applying Indirect Association Rule mining Algorithm

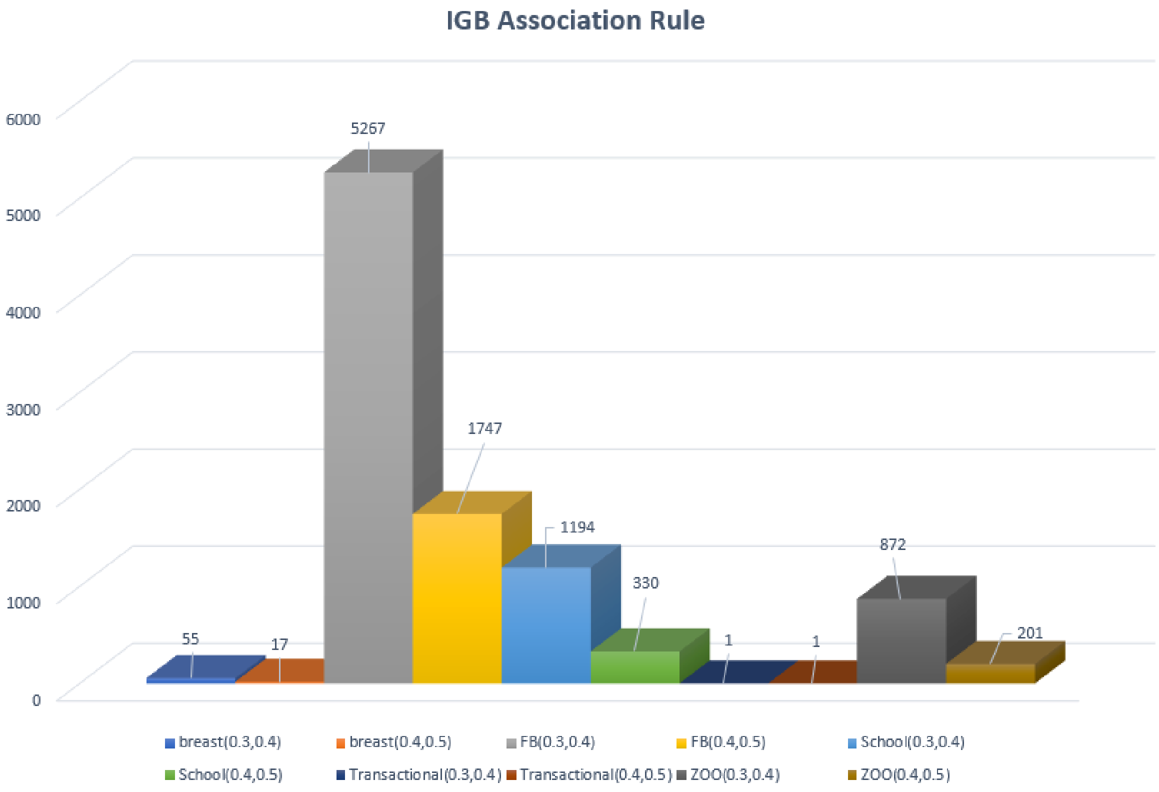


Figure 4.20: Association Rules Generated in Different Datasets for Applying IGB Association Rule mining Algorithm

4.5.2 Keeping Confidence Value Fixed

In this comparison, we are going to keep the confidence value fixed. After that we will apply the algorithm in different datasets.

Breast cancer Dataset

In Figure 4.21, we can see that the RP-Growth algorithm is not performing good in this dataset.Moreover, Apriori, FP-Growth, FP-Growth with lift are generating same amount of association rules. But which is the faster one we will see it in our other comparison.

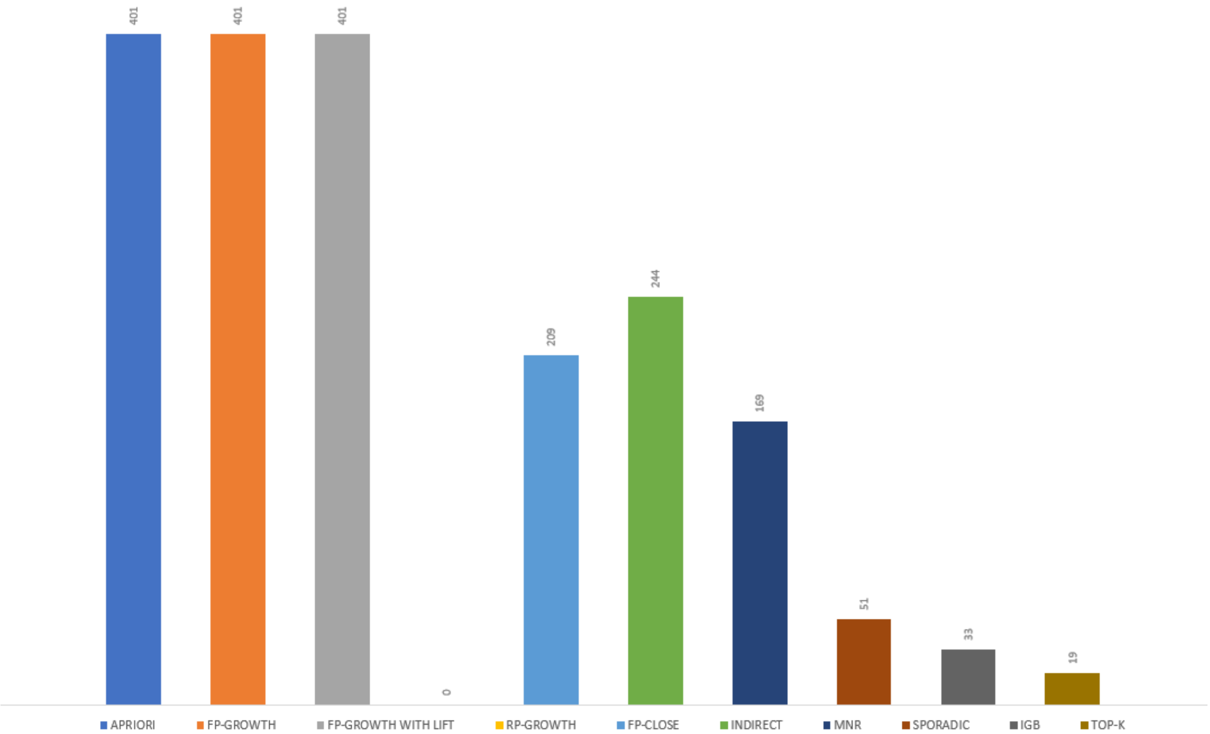


Figure 4.21: Association Rules generated in breast cancer dataset after keeping the confidence value fixed

Facebook Dataset

In Figure 4.22, we can see that the RP-Growth algorithm is not performing good in this dataset too.Moreover, Apriori, FP-Growth, FP-Growth with lift are generating same amount of association rules again.But, with these 3 algorithm we can see that MNR and FP-close algorithm also performing well in this dataset.

School Dataset

In Figure 4.23, we can see that the RP-Growth,Sporadic and Top-K association

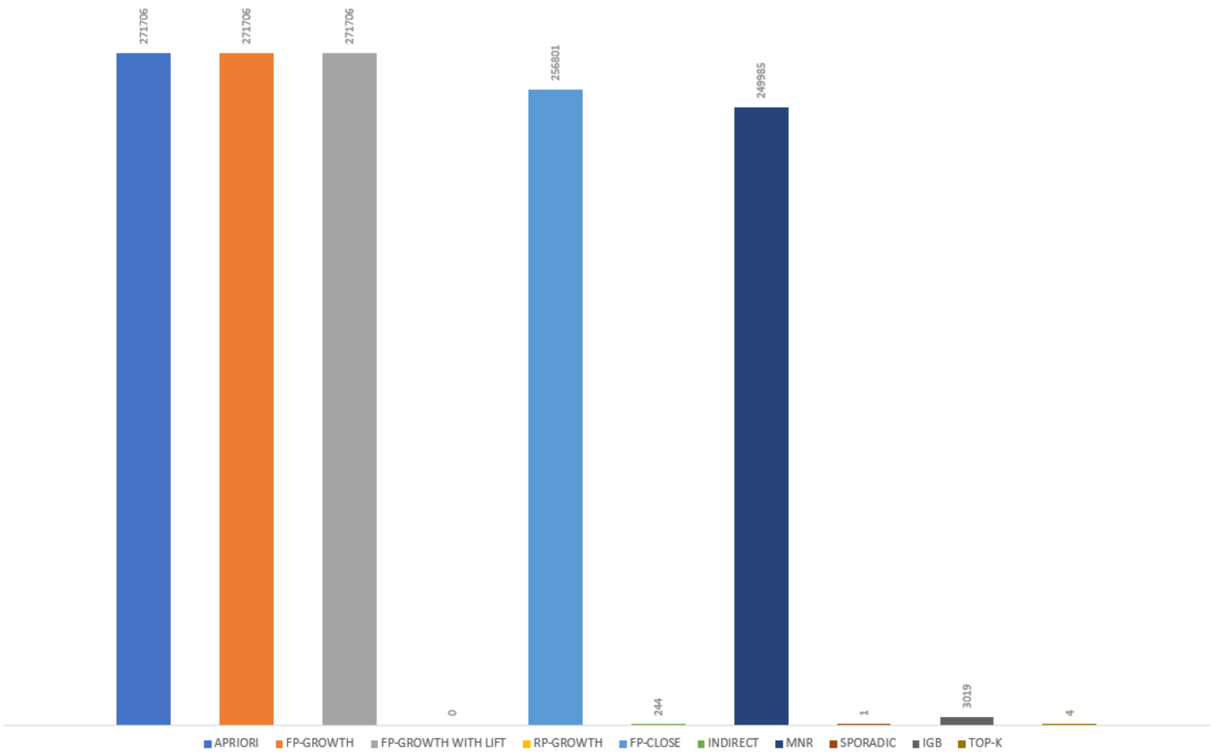


Figure 4.22: Association Rules generated in Facebook dataset after keeping the confidence value fixed

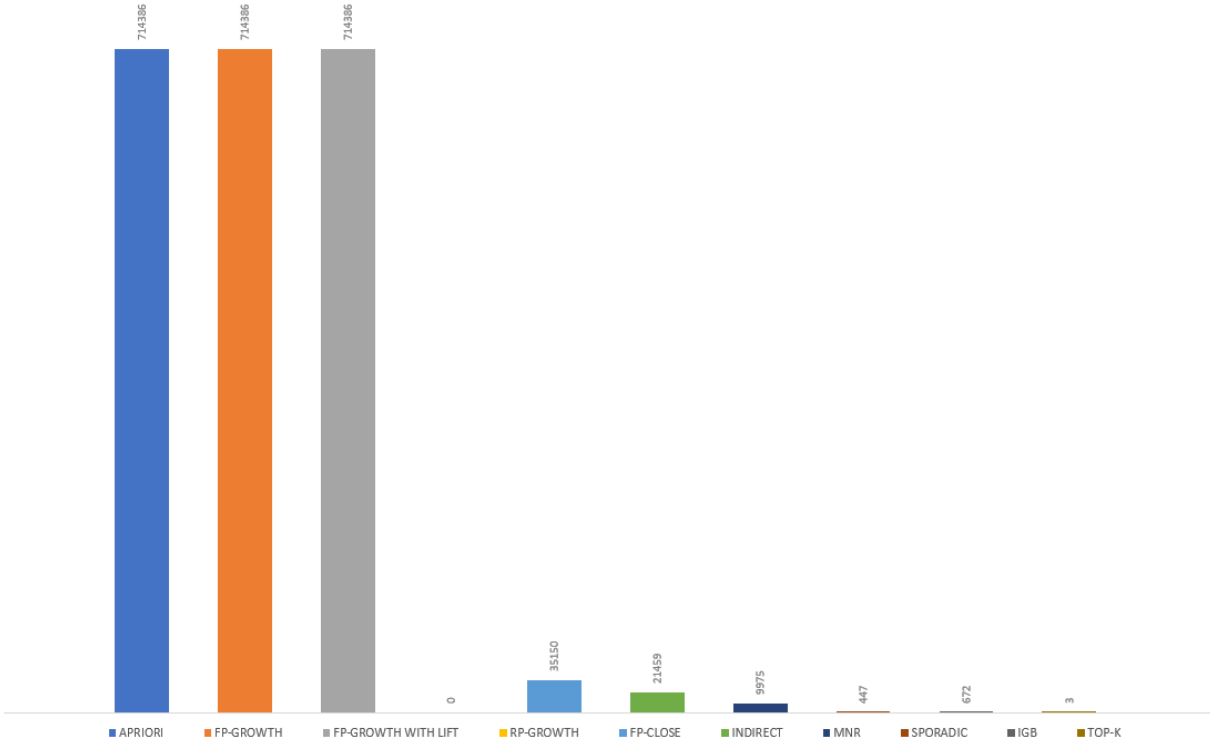


Figure 4.23: Association Rules generated in Student dataset after keeping the confidence value fixed

rule mining algorithm is not performing good in this dataset.Moreover, Apriori, FP-Growth, FP-Growth with lift are generating same amount of association rules again.But, with these 3 algorithm we can see that MNR and FP-close algorithm also performing well in this dataset.

Sales Transaction Dataset

In Figure 4.24, we can see that the RP-Growth is outperforming the other algorithms and generating a huge amount of association rules. Other algorithm in this dataset with the high fixed confidence value are not performing at all.

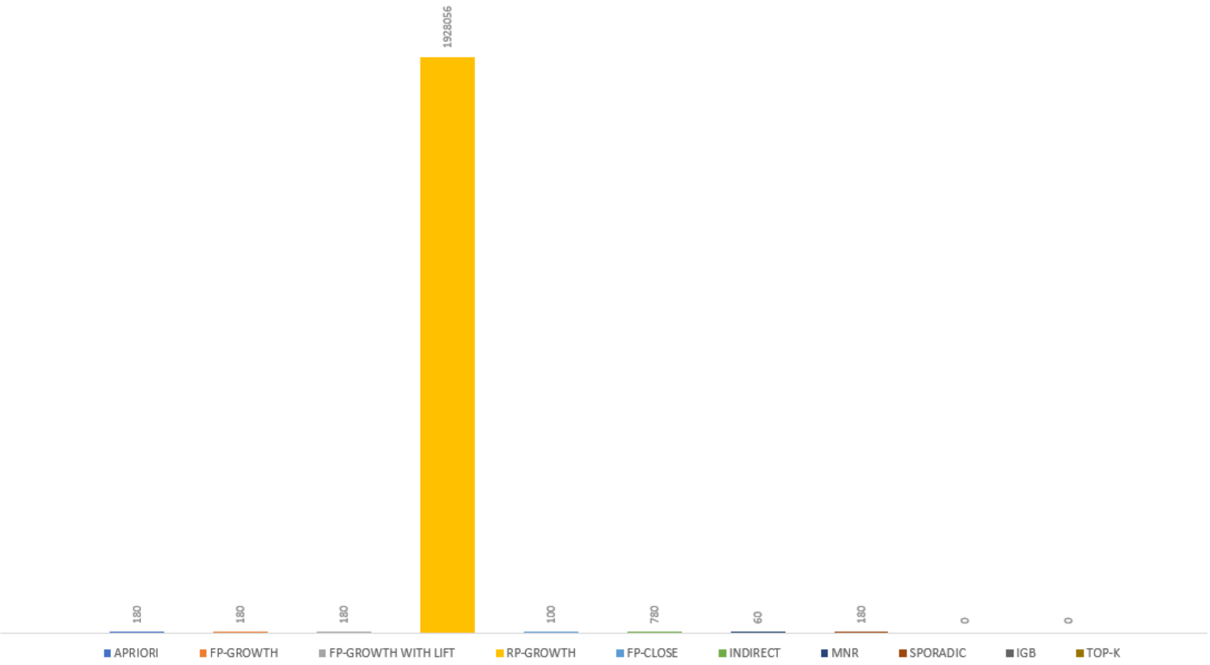


Figure 4.24: Association Rules generated in Sales Transaction Dataset after keeping the confidence value fixed

Zoo Dataset

In Figure 4.25, we can see same performance of algorithm we have seen in the dataset of Sales Transaction, Facebook and Breast cancer dataset.

4.5.3 Time taken to generate Association rules

From the number of association rules generated , we can't differentiate among the FP-Growth,FP-Growth with lift and Apriori algorithm. Because, They were

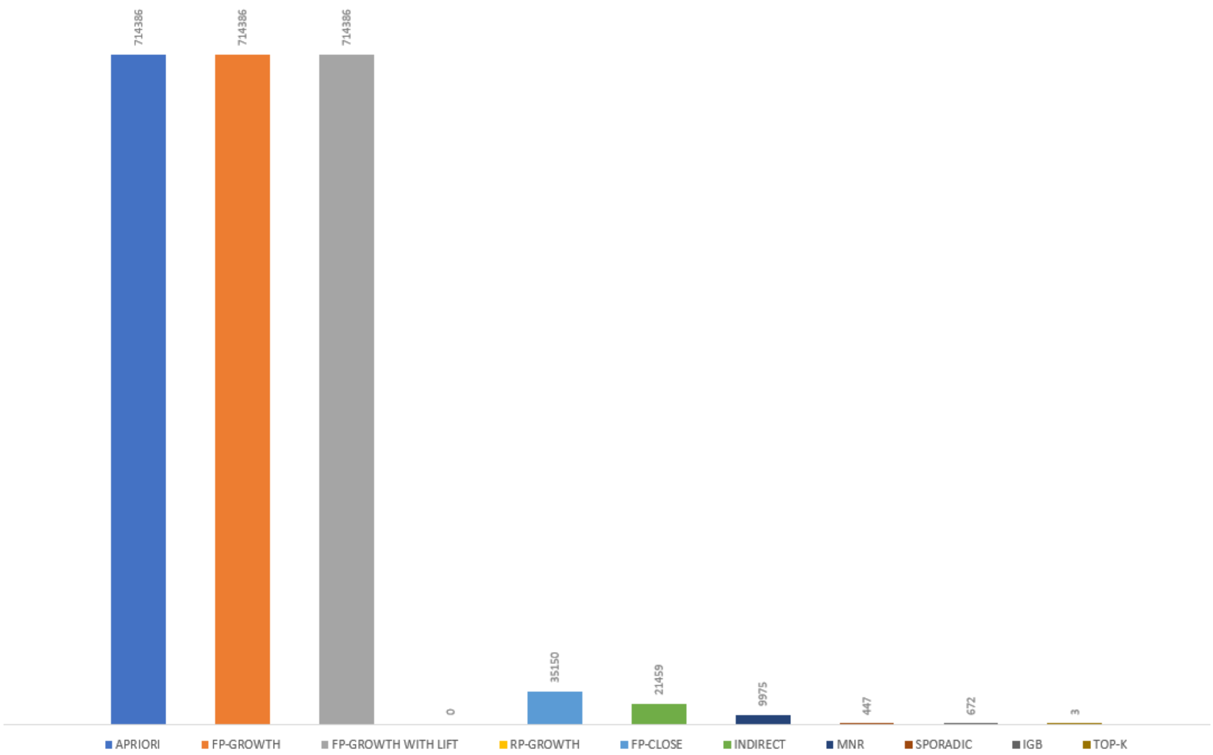


Figure 4.25: Association Rules generated in Zoo Dataset after keeping the confidence value fixed

generating the same amount of association rules. So, for this we have taken into the account the execution time.

In Figure 4.26, we can see that among these three algorithm the FP-Growth is performing really well. Moreover, in this dataset, with confidence value 0.7, RP-growth ,TOP-K and Sporadic Association rule mining algorithm are doing well also.

4.6 Conclusion

So, in this chapter we tried to show our performance of the algorithms. We also compared them in terms of association rule generated and execution time taken to generate these rule.Moreover, we discussed about our datasets and how we collect them. Moreover, we discussed about the social impact of our project and also showed the ethical impact of this project too.

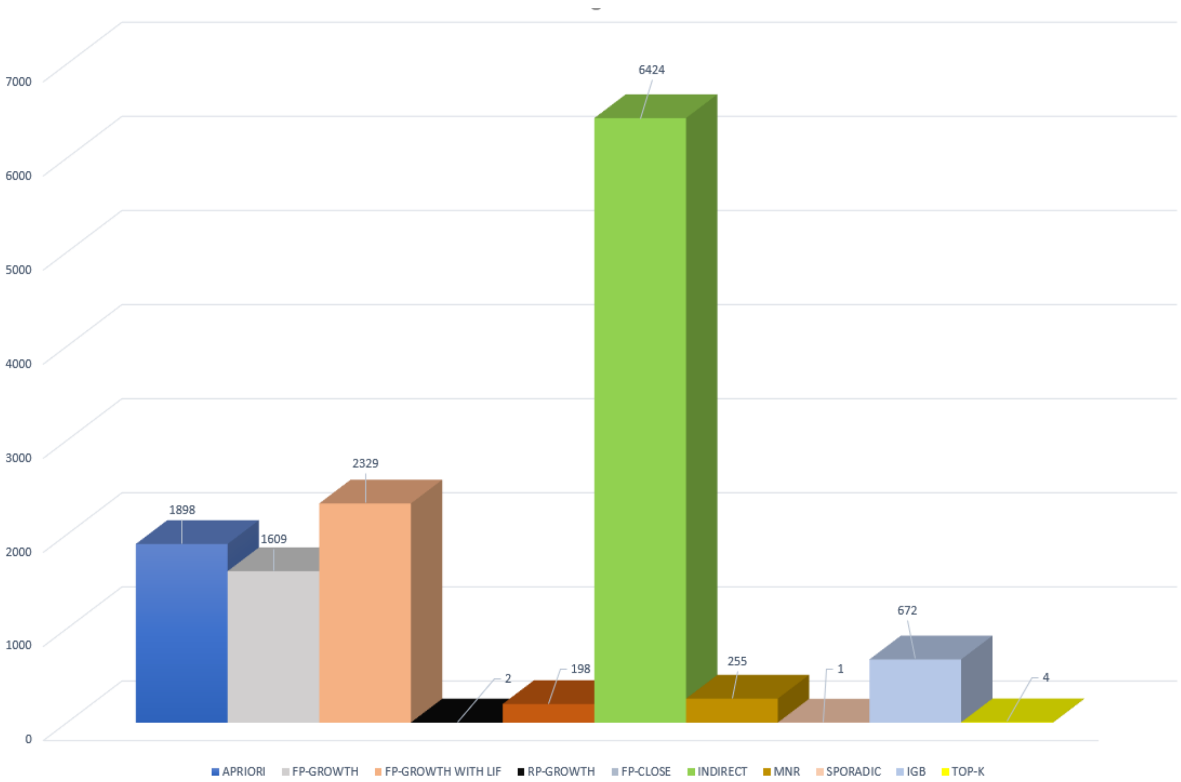


Figure 4.26: Execution time for ten algorithms in Zoo dataset with the confidence value 0.7

Chapter 5

Conclusion

5.1 Conclusion

In this project, our aim is to compare the association rule mining algorithms and show their performance. We have worked with ten algorithms and five datasets. This work is done before, but it is not done yet with this amount of data. Moreover, we create the variety of the algorithm so that our work can be unique. We Compare these algorithms based on association rules generated and the time it takes to generate the association rules. After the comparison, we have seen that the FP-Growth algorithm is working well in every kind of dataset and the execution time is better than the apriori and FP-Growth with lift measure.

Moreover, we can find the hidden sporadic and rare association rules in transaction datasets with the Sporadic and RP-Growth association rule mining algorithm. Moreover, we can find lots of closed associations in the social network datasets with the FP-close. Moreover, we also found the tremendous performance of the MNR algorithm in the facebook dataset.

5.2 Future Work

For the development of data mining, the data scientist must work with the algorithms because data mining is the future in our digital world. So, future work should be done on these things :

- The scientist should work on the availability of the datasets.
- The researchers should work to increase the efficiency of the algorithm.
- They have to think out of the box when they are developing the algorithms. Like, researchers should invent the new term like "K" instead of "Min sup" to

increase efficiency.

- The invention of the new data structure for making these algorithms will decrease the execution time.

- They should invent a new pruning technique for increasing the performance of the association rule mining algorithms.

References

[1] R. Agrawal, R. Srikant *et al.*, 'Fast algorithms for mining association rules,' in *Proc. 20th int. conf. very large data bases, VLDB*, Citeseer, vol. 1215, 1994, pp. 487–499 (cit. on pp. 1, 10).

[2] A. Rajak and M. K. Gupta, 'Association rule mining: Applications in various areas,' in *Proceedings of international conference on data management, Ghaziabad, India*, 2008, pp. 3–7 (cit. on p. 4).

[3] D. Gamberger, N. Lavrac and V. Jovanoski, 'High confidence association rules for medical diagnosis,' 1999 (cit. on p. 4).

[4] C. I. Branden and J. Tooze, *Introduction to protein structure.* Garland Science, 2012 (cit. on p. 5).

[5] A. Rajak and M. K. Gupta, 'Association rule mining: Applications in various areas,' in *Proceedings of international conference on data management, Ghaziabad, India*, 2008, pp. 3–7 (cit. on p. 5).

[6] D. Malerba, F. Esposito, F. A. Lisi and A. Appice, 'Mining spatial association rules in census data,' *Research in Official Statistics. v5 i1*, pp. 19–44, 2003 (cit. on p. 5).

[7] G. Saporta, 'Data mining and official statistics,' in *Quinta Conferenza Nationale di Statistica*, 2000, pp. 41–46 (cit. on p. 5).

[8] R.-C. Wu, R.-S. Chen and C.-C. Chen, 'Data mining application in customer relationship management of credit card business,' in *29th Annual International Computer Software and Applications Conference (COMPSAC'05)*, IEEE, vol. 2, 2005, pp. 39–40 (cit. on p. 5).

[9] C. Rygielski, J.-C. Wang and D. C. Yen, 'Data mining techniques for customer relationship management,' *Technology in society*, vol. 24, no. 4, pp. 483–502, 2002 (cit. on p. 5).

[10] J. Han, J. Pei and Y. Yin, 'Mining frequent patterns without candidate generation,' *ACM sigmod record*, vol. 29, no. 2, pp. 1–12, 2000 (cit. on p. 11).

[11] J. Han, J. Pei, Y. Yin and R. Mao, 'Mining frequent patterns without candidate generation: A frequent-pattern tree approach,' *Data mining and knowledge discovery*, vol. 8, no. 1, pp. 53–87, 2004 (cit. on p. 11).

[12] L. Szathmary, 'Symbolic data mining methods with the coron platform,' Ph.D. dissertation, Université Henri Poincaré-Nancy 1, 2006 (cit. on p. 11).

[13] Y. S. Koh and N. Rountree, 'Finding sporadic rules using apriori-inverse,' in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, 2005, pp. 97–106 (cit. on p. 11).

[14] G. Gasmi, S. B. Yahia, E. M. Nguifo and Y. Slimani, '\mathcal {igb} : $Anewinformativegenericbaseofassociationrules$,' in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, 2005, pp. 81–90 (cit. on p. 11).

[15] M. Kryszkiewicz, 'Representative association rules and minimum condition maximum consequence association rules,' in *European Symposium on Principles of Data Mining and Knowledge Discovery*, Springer, 1998, pp. 361–369 (cit. on p. 11).

[16] P.-N. Tan, V. Kumar and J. Srivastava, 'Indirect association: Mining higher order dependencies in data,' in *European Conference on Principles of Data Mining and Knowledge Discovery*, Springer, 2000, pp. 632–637 (cit. on p. 12).

[17] P. Fournier-Viger, C.-W. Wu and V. S. Tseng, 'Mining top-k association rules,' in *Canadian Conference on Artificial Intelligence*, Springer, 2012, pp. 61–73 (cit. on p. 12).

[18] Y. Kameya and T. Sato, 'Rp-growth: Top-k mining of relevant patterns with minimum support raising,' in *Proceedings of the 2012 SIAM international conference on data mining*, SIAM, 2012, pp. 816–827 (cit. on p. 12).

[19] K. Vani, 'Comparative analysis of association rule mining algorithms based on performance survey,' *International Journal of Computer Science and Information Technologies*, vol. 6, no. 4, pp. 3980–3985, 2015 (cit. on p. 12).

[20] S. Vijayarani and S. Sharmila, 'Comparative analysis of association rule mining algorithms,' in *2016 International Conference on Inventive Computation Technologies (ICICT)*, IEEE, vol. 3, 2016, pp. 1–6 (cit. on p. 13).

[21] K. Khurana and S. Sharma, 'A comparative analysis of association rule mining algorithms,' *International Journal of Scientific and Research Publications*, vol. 3, no. 5, p. 0, 2013 (cit. on p. 13).

[22] S. Ziauddin, E. Khan and N. M. Imran, 'Fet twin model,' *2012 7th International Conference on Electrical and Computer Engineering*, 2012. DOI: `10.1109/icece.2012.6471638` (cit. on p. 13).

[23] M. Girotra, K. Nagpal, S. Minocha and N. Sharma, 'Comparative survey on association rule mining algorithms,' *International Journal of Computer Applications*, vol. 84, no. 10, 2013 (cit. on p. 13).

[24] A. Saxena and V. Rajpoot, 'A comparative analysis of association rule mining algorithms,' in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, vol. 1099, 2021, p. 012 032 (cit. on p. 13).

[25] P. Prithiviraj and P. Dr.R, ‘A comparative analysis of association rule mining algorithms in data mining: A study,’ Dec. 2015 (cit. on p. 14).