## Sequential Pattern Mining of Hadith Narrator with Sequential Pattern Discovery Using Equivalent Classes (SPADE) Algorithm

(Case Study : the Hadith Narrator in the book of Shahih Imam Al-Bukhari (the book of from *Software* Ensiklopedi Hadis Kitab 9 Imam for Beginning of revelation, Faith, Science, Wudlu, Bath, Haidl, Tayamum, Pray, Pray Times, Adzan, and Friday Books)

Ayu Septiani, Baiq Siska Febriani Astuti, and RB. Fajriya Hakim

Department of Statistics, Faculty of Mathematics and Natural Sciences, Islamic University of Indonesia

**Abstract:** This research is about narrator data of hadith shahih Imam Bukhari for Beginning of revelation, Faith, Science, Wudlu, Bath, Haidl, Tayamum, Pray, Pray Times, Adzan, and Friday Books. The purpose of the research is to find out the association pattern that is formed between narrator data in shahih Bukhari book (Beginning of revelation, Faith, Science, Wudlu, Bath, Haidl, Tayamum, Pray, Pray Times, Adzan, and Friday Books). The data used is secondary data from software Ensiklopedia Kitab 9 Imam was created by Lidwa Pusaka. Sequential pattern mining technique is the method that used to analysis. Transaction data generated from this narrator data of hadith shahih imam bukhari be used to search for patterns of narrator data. One pattern that can be searched is sequential pattern, which the narrator data spending patterns. Sequential Pattern Discovery Algorithms Using Equivalent Classes (SPADE) is one of the algorithm in sequential pattern mining, this algorithm find frequent sequences of narrator data transaction, using vertical database and sequence join process. Results of the SPADE algorithm is frequent sequences that then used to form a rule. It technique is used to find the association patern between items combination. Based on association rules sequential analysis with SPADE algorithm for minimum support 0,03 and minimum confidence 0,5 is gotten 7 association sequential pattern based on the sequence of narrator data in the sanad.

**Keywords:** The Hadith Narrator in the book of Shahih Imam Al-Bukhari, Data Mining, Sequential Pattern Mining, SPADE Algorithm

## **1. Introduction**

Mukri and Ismail [1,2] argued that Al-Qur'an and Hadith Nabi Muhammad S.A.W is a very reliable source of law for muslim. The Hadith is theory about liveable occupy second level after Al-Qur'an. Where one of the function Hadith of Al-Qur'an is clarify anything in the Al-Qur'an is still global and vague. The Hadith of Prophet which collected in the books of Hadith, such shahih Al-Bukhari and Shahih Muslim.

In the books of Al-Maliki [3] argued that the book of Hadith relate Hadith Shahih according to most of the scholar are the book of Hadith by Imam Bukhari because many muslim used as the hujjah (guidance) to the present. Imam Bukhari in his the book of Hadith publish entire only Hadith shahih so called Shahih Al-Bukhari. The books of Shahih Bukhari has obtained high appreciation from the scholars. The scholars has given statements thah the book of Hadith shahih bukhari is one of the book which very authentic after Al-Qur'an. It is the set of Hadith Shahih which at once became the object of discussion.

Imam Bukhari in his book of Hadith given criteria and conditions to narrator, namely its sanad to be mutthasil (to be continued), the narrator is fair (his attitude always shows a personal piety and distance themselves from characteristic or behavior is not good to do), and finally it is dhabith (he was able to memorize every Hadith that he heard and any time he can present or submit it). Then, Imam Bukhari and the scholars of Hadith argue that, the means of muttashil is it's has meet (autobiography), although only one time, and it's required the people who tsiqah (trusted), and not tadlis (lie). The narrator are selected and researched as a lot of peoples accompany his teacher and it's very understood of Hadith his teacher. If the narrator and his teacher

obvious has ever meet, then the Hadith narrated that rather he has hear from his teachers. After that, Imam Bukhari argue that conditionce muttasil was stronger, because he requires the existence of meet and one period. Overall it can be said that in the narrated Shahih of Hadith, Imam Bukhari is concerned the state or relationships betweet his narrator so formed Shahih of Hadith with sanad wich articulate A-Maliki [3].

This study used data in the population was data Shahih Imam bukhari in the book Hadith Softeware Ensiklopedia Hadith 9 Imam work Lidwa Pusaka. In this study used Hadith as many as 1000 Hadith shahih of Imam Bukhari (Hadith number from number 1 to 1000) from book the Beginning of revelation, Faith, Science, Wudlu, Bath, Haidl, Tayamum, Pray, Pray Times, Adzan, and Friday Books. Based on the above issues, the authors wished to try fine other benefits from the hadith narrator of data gathered in such a vast sum. For example, the authors would like to know the association between the narrator X and the narrator Y from overall hadiths are examined. The authors wanted to find such rules, the rules that obtained later in accordance with the rules of Hadith sanad rules pertaining to the sequence of the narrator in the sanad Hadith (an extension of narrators from companions of the Prophet Muhammad SAW saw came to the one beneath that is tabi'in, tabi'ut tabi'in, and tabi'ul atba). Therefore, this research uses a Sequential Pattern Mining with Sequential Pattern Discovery using Equivalent classes algorithm (SPADE). Then retrieved the association pattern is formed between the Hadith Narrator in the book of Shahih Imam Al-Bukhari (the book of Beginning of revelation, Faith, Science, Wudlu, Bath, Haidl, Tayamum, Pray, Pray Times, Adzan, and Friday Books). So as to provide information, especially for Muslim in historical accounts of The Sahih Hadith in term of his narrator.

## 2. Concepts

#### 2.1 Association Rules

Mining association rules over transaction data was introduced in and many establish techniques has been recognized such as Apriori, FP-Tree, and also hash-based association mining [25,26,29]. Some of their advantages and disadvantages have been uncovered by Han [26]. Their application has been succeeded when it is implemented in clustering on sales day, in very large clustered domains and many other applications. The uncovered relationships then could be described in the form of rules of association between sets of frequent items. For example, some rules can be extracted from the transaction data set shown:

#### {Bread}-> {Diaper}, {Diapers}->{Beer} or {Bread} -> {Milk}

Those rules shows there are strong relationships exist between some of items. The sale of bread and diapers, diapers and beer or bread and milk occurred because many customers who buy bread also buy diapers, who buy diapers also buy beer and who buy bread also buy milk. The rule of bread and milk maybe is common behavior of customer's needs and does not more interesting than bread and diapers or diapers and beer. Those interesting rules uncovered from transcations data could help retailers as a new opportunity to cross selling their products for their customers.

For a given rule, say  $A \rightarrow B$ , and let N is number of transaction in which the rule is satisfied. How strong an association rule between items could be measured in terms of its support and confidence. The support for a rule  $A \rightarrow B$  is obtained by dividing the number of transactions by the total number of transactions, N. The support of rule  $A \rightarrow B$  is a relative frequency that indicates the proportion of A and B present together in data transactions, where A and B are disjoint itemsets. Bring this notation to probability rule, when a large sample is considered, the support will approximates the probability of occurrence [29]:

### Support $(A \rightarrow B) = Probability (A and B occur)$

Support usually performed to filter out any rules which are obtained that are less frequent. Support is a simple and very useful measure of interestingness association between items. The confidence of the rule  $A \rightarrow B$  is obtained by dividing the number of transactions which item A and item B present in the transactions by the number of transactions which contain item A:

Confidence 
$$(A \to B) = \frac{N_{A \to B}}{N_A} = \frac{N_{A \to B/N}}{N_{A/N}} = \frac{Support (A \to B)}{Support (A)}$$

The confidence expresses how frequently items B appear in transactions that contain A.

#### 2.2. Sequential Pattern Mining

Sequential pattern mining plays an important role in data mining research and has a broad range of application, including customer purchase behavior analysis [4,5], guidance systems [6], web usage behavior analysis [7], and so on. Agrawal and Srikant (1995) first proposed the sequential pattern mining in 1995. The same author later developed a generalized algorithm based on the apriori property, called GSP (8). Many other algorithms, have also been proposed to improve the effect of mining sequential patterns.

However, when the support value is set low, many sequential patterns including irrelevant or spurious patterns, may be obtained. Thus, designing an efficient sequential rule mining process to remove these spurious patterns is important. Sequential rules express the temporal relationships between sequential patterns from a sequence database [9]. Sequential rules can be considered natural extension of original sequential patterns, just as association rules are natural extension of frequent itemsets [8].

Using sequential rules, we can know the series of events that will usually occur after a series of previous ones. Thus, they can help users better understand the chronological order of the sequences present in the sequence database. Like a sequential pattern, a sequential rule is also applied in many application areas, including the stock market [10,11,12], weather observation [13], e-learning [14], trade [15], and software engineering [16,17]. In addition, sequential rule mining is also applied to address the prediction problem [18,19,20,21,22,23]. In the prediction problem, a sequence of events that appear frequently in a database is not sufficient to predict events, while sequential rules allow a better understanding of the prediction. An appropriate sequential rule mining process, instead of mining only sequential patterns, is also desired.

#### 2.3 Sequential Pattern Discovery using Equivalence classes (SPADE) Algorithm

SPADE, an algorithm for fast discovery of Sequential Patterns. The existing solutions to this problem make repeated database scans, and use complex hash structures which have poor locality. SPADE utilizes combinatorial properties to decompose the original problem into smaller sub-problems, that can be independently solved in main-memory using efficient lattice search techniques, and using simple join operations. All sequences are discovered in only three database scans. Experiments showthat SPADE outperforms the best previous algorithm by a factor of two, and by an order of magnitude with some pre-processed data. It also has linear scalability with respect to the number of input-sequences, and a number of other database parameters. Finally, we discuss how the results of sequence mining can be applied in a real application domain [27,28].

This algorithm is proposed in and it includes the features of a search space partitioning where the search space includes vertical database layout. The search space in SPADE is represented as a lattice structure and it use the notion of equivalence classes to partition it. It decomposes the original lattice into slighter sub-lattices, so that each sub-lattice can be entirely processed using either a breadth-first or depth-first search method (SPADE is also DFSbased method). The SPADE support counting of the candidate sequence method includes bitwise or logical operations. A conducted experimental results show that SPADE is about twice as fast as GSP. The reason behind this is that SPADE uses a more efficient support counting method based on the idlist structure. Additionally, SPADE shows a linear scalability with respect to the number of sequences [27,28].

#### **3. Results and Discussion**

# **3.1** Sequential Pattern Mining with Sequential Pattern Discovery using Equivalent classes Algorithm (SPADE)

The pattern of association between the narrator of Hadiths in Hadith Shahih of Imam Bukhari (the book of Beginning of revelation, Faith, Science, Wudlu, Bath, Haidl, Tayamum, Pray, Pray Times, Adzan, and Friday Books) by using technique Sequential Pattern Mining with Sequential Pattern Discovery using Equivalent classes algorithm (SPADE) with minimum limits support 0.03 and minimum limit confidence 0.5, it's analysis of results obtained using the assistance software R 3.03 is as follow.

No				Lift
	Aturan (Asosiasi)	Support	Confidence	Rasio
1	<{Aisyah binti Abi Bakar Ash Shiddiq}>=>			
	<{Urwah bin Az Zubair}>	0.067	0.503759	6.296992
2	${\rm \leq} {\rm Abdullahbin'Umar} > => {\rm \leq} {\rm Nafi'maulaibnu}$			
	'Umar}>	0.053	0.514563	9.528946
3	<{Urwah bin Az Zubair}> => <{Hisyam bin			
	'Urwah bin Az Zubair bin Al 'Awwam}>	0.042	0.525	10.9375
4	<{Aisyah binti Abi Bakar Ash Shiddiq},{Urwah			
	bin Az Zubair}> => <{Hisyam bin 'Urwah bin Az			
	Zubair bin Al 'Awwam}>	0.034	0.507463	10.57214
5	<{Malik bin Anas}> => <{Abdullah bin Yusuf}>	0.076	0.550725	6.716154
6	<{Syu'aib bin Abi Hamzah Dinar}>=><{Al			
	Hakam bin Nafi'}>	0.037	0.860465	23.25581
7	<{Muhammad bin Muslim},{Syu'aib bin Abi			
	Hamzah Dinar}> => <{Al Hakam bin Nafi'}>	0.032	0.914286	24.71043

TABLE I: The association Sequential Pattern Mining of The Hadith Narrator (minSup=0.03 dan minCof=0.5)

Based on the limitation specified by minimum support 0.03 and minimum confidence 0.5 retrieved 7 rules of the association formed, with regard the information obtained from these rules as follows:

#### • The rules {Aisyah binti Abi Bakar Ash Shiddiq} => {Urwah bin Az Zubair}

The rules with the support value of 0.067, confidence of 0.503759 and lift ratio of 6.296992. the meaning of the support value of 0.067 was 6.7% or 67 hadiths of the entire hadiths which examined (1000 hadiths) was narrated by Aisyah binti Abi Bakar Ash Shiddiq and Urwah bin Az Zubair as narrator (in one sanad). Confidence value of 0.503759 meaning if a particular hadith narrated by Aisyah binti Abi Bakar Ash Shiddiq also narrated by Aisyah binti Abi Bakar Ash Shiddiq also narated by Urwah bin Az Zubair. While the lift ratio value of 6.296992 indicates how strong of the rules or the rules that are formed of the sequential pattern mining algorithm. The value of lift ratio ranged between 0 to infinity. The value of the lift ratio, the recommended rules is if lift ratio of more than 1. So base on the lift ratio value of 6.296992 more than one, then these rules are recommended. That applies also to the six others rules.

Based on the explanation of the association sequence above, look all te rules can be used because it is accordance with the rules of the science of hadith related to the sequence of the narrator in hadith sanad (an extension of narrators from companions of the Prophet Muhammad SAW saw came to the one beneath that is tabi'in, tabi'ut tabi'in, and tabi'ul atba), then the rules of association used in this research was adapted to the rule. So, a Sequential Pattern Mining with Sequential Pattern Discovery using Equivalent classes algorithm (SPADE) acquired the association rules in accordance with the order of the hadith of narrator based on his category in sanad. The Authors try to summarized information from the association pattern is formed between the Hadith of narrator in the books of Shahih Imam Bukhari by raising the value of confidence be 0.6, so the association pattern obtained as follows:

No	Rule	Support	Confidence	Lift Ratio		
1	<{Urwah bin Az Zubair}> =><{Hisyam bin 'Urwah bin					
	Az Zubair bin Al 'Awwam}>	0.042	0.525	10.9375		
2	<{Syu'aib bin Abi Hamzah Dinar}> =><{Al Hakam bin					
	Nafi'}>	0.037	0.860465	23.25581		
3	<{Muhammad bin Muslim},{Syu'aib bin Abi Hamzah					
	Dinar}>=><{Al Hakam bin Nafi'}>	0.032	0.914286	24.71043		

Table II. the association sequential pattern mining of the hadiths narrator (minSup=0.03 and minCof=0.6)

Based on the above table, acquired three rules that formed, where from three rules have higher levels confidence compared with other rules. The purpose of the summary is to explained to explain these rules in detail sequence pattern mining algorithm with SPADE, where that the rules are in accordance with the rules of hadith sanad. As for the interesting information is as follows:

• The rule {Urwah bin Az Zubair} => {Hisyam bin 'Urwah bin Az Zubair bin Al 'Awwam}



Fig. 1: association {Urwah bin Az Zubair} => {Hisyam bin 'Urwah bin Az Zubair bin Al 'Awwam}

Based on the association rules have been adapted to te rules of the science of hadith, it is evident that Urwah bin Az Zubair and Hisyam bin 'Urwah bin Az Zubair bin Al 'Awwam is the narrator whow is fair (always indicate te personal taqwa) and tsiqah (credibility) is good. Additionally, it is note that this narrator is surely ever met each other, consider both have relationship father and son so it would have been in accordance with the criteria set forth by Imam Bukhri about the narrators of shahih hadiths. So, the pattern rule {Urwah bin Az Zubair} => {Hisyam bin 'Urwah bin Az Zubair bin Al 'Awwam} formed or valid for the beginning of the revelation, faith, science, wudlu, bath, Haidl, Tayamum, Pray, pray times, adzan, and Friday Books were in accordance with criteria of narrator shahih hadith by Imam Bukhari [24].

• The rule {Syu'aib bin Abi Hamzah Dinar} => {Al Hakam bin Nafi'}



Fig. 2: Association {Syu'aib bin Abi Hamzah Dinar} => {Al Hakam bin Nafi'}

Based on the association rules have been adapted to te rules of the science of hadith, it is evident that Syu'aib bin Abi Hamzah and Al Hakam bin Nafi' is the narrator whow is fair (always indicate te personal taqwa) and tsiqah (credibility) is good. Additionally, it is note that this narrator is surely ever met each other, consider both have relationship teacher and student so it would have been in accordance with the criteria set forth by Imam Bukhri about the narrators of shahih hadiths. So, the pattern rule {Syu'aib bin Abi Hamzah Dinar} => {Al Hakam bin Nafi'} formed or valid for the beginning of the revelation, faith, science, wudlu, bath, Haidl, Tayamum, Pray, pray times, adzan, and Friday Books were in accordance with criteria of narrator shahih hadith by Imam Bukhari. [24]

#### • The rule {Muhammad bin Muslim},{Syu'aib bin Abi Hamzah Dinar} =>{Al Hakam bin Nafi'}



Fig. 3: Association {Muhammad bin Muslim},{Syu'aib bin Abi Hamzah Dinar} =>{Al Hakam bin Nafi'}

Based on the association rules have been adapted to te rules of the science of hadith, it is evident Muhammad bin Muslim, Syu'aib bin Abi Hamzah Dinar, and Al Hakam bin Nafi' is the narrator whow is fair (always indicate te personal taqwa) and tsiqah (credibility) is good. Additionally, it is note that this narrator is surely ever met each other, consider both have third relationship teacher and student so it would have been in accordance with the criteria set forth by Imam Bukhri about the narrators of shahih hadiths. So, the pattern rule {Muhammad bin Muslim}, {Syu'aib bin Abi Hamzah Dinar} =>{Al Hakam bin Nafi'} formed or valid for the beginning of the revelation, faith, science, wudlu, bath, Haidl, Tayamum, Pray, pray times, adzan, and Friday Books were in accordance with criteria of narrator shahih hadith by Imam Bukhari. [24]

## 4. Conclusion

- Sequential Pattern Discovery Algorithms Using Equivalent Classes (SPADE) is one of the algorithm in sequential pattern mining, this algorithm find frequent sequences of narrator data transaction, using vertical database and sequence join process. Results of the SPADE algorithm is frequent sequences that then used to form a rule. It technique is used to find the association patern between items combination. Based on association rules sequential analysis with SPADE algorithm for minimum support 0,03 and minimum confidence 0,5 is gotten 7 association sequential pattern based on the sequence of narrator data in the sanad.
- So, a Sequential Pattern Mining with Sequential Pattern Discovery using Equivalent classes algorithm (SPADE) acquired the association rules in accordance with the order of the hadith of narrator based on his category in sanad. The Authors try to summarized information from the association pattern is formed between the Hadith of narrator in the books of Shahih Imam Bukhari by raising the value of confidence be 0.6, so as for the interesting information is as follows:
  - 1. The rule {Urwah bin Az Zubair} => {Hisyam bin 'Urwah bin Az Zubair bin Al 'Awwam}
  - 2. The rule {Syu'aib bin Abi Hamzah Dinar} => {Al Hakam bin Nafi'}
    3. The rule {Muhammad bin Muslim}, {Syu'aib bin Abi Hamzah Dinar} =>{Al Hakam bin Nafi'}

## 5. References

- [1] B. Mukri, Kontekstualisasi Hadis Rasulullah (Mengungkap Akar dan Implementasinya), Yogyakarta: Ideal Press, 2005
- [2] M.S Ismail, Kaedah Kesahihan Sanad Hadis (Telaah Kritis dan Tinjauan dengan Pendekatan Ilmu Sejarah), PT Bulan Bintang: Jakarta, 1998
- [3] M. A. Al-Maliki, Ilmu Ushul Hadis, Yogyakarta: Pustaka Pelajar, 2006
- [4] M. J. Berry and G. S. Linoff, Data Mining Techniques for Marketing, Sales and Customer Support, John Wiley & Sons, 1997.
- [5] S. Brin, R. Motwani, J. Ullman and S. Tsur, Dynamic itemset counting and implication rules for market basket data, Proc. of the 1997 ACMSIGMOD Int. Conf. on the Management of Data, pp.255264, 1997. http://dx.doi.org/10.1145/253260.253325
- [6] C.-Y. Tsai and P.-H. Lo, A sequential pattern based route suggestion system, International Journal of Innovative Computing, Information and Control, vol.6, no.10, pp.4389-4408, 2010.
- [7] G. T. Raju, P. S. Satyanarayana and L. M. Patnaik, Knowledge discovery from web usage data: Extraction and applications of sequential and clustering patterns { A survey, International Journal of Innovative Computing, Information and Control, vol.4, no.2, pp.381-389, 2008.
- [8] R. Srikant and R. Agrawal, Mining sequential patterns: Generalizations and performance improvements, Proc. of the 5th Int. Conf. Extending Database Technology, pp.3-17, 1996. http://dx.doi.org/10.1007/bfb0014140
- M. Spiliopoulou, Managing interesting rules in sequence mining, Proc. of European Conference on Principles of Data Mining and Knowledge Discovery, pp.554-560, 1999. http://dx.doi.org/10.1007/978-3-540-48247-5\_73
- [10] S. Brin, R. Motwani, J. Ullman and S. Tsur, Dynamic itemset counting and implication rules for market basket data, Proc. of the 1997 ACMSIGMOD Int. Conf. on the Management of Data, pp.255264, 1997.

http://dx.doi.org/10.1145/253260.253325

- [11] C.-Y. Tsai and P.-H. Lo, A sequential pattern based route suggestion system, International Journal of Innovative Computing, Information and Control, vol.6, no.10, pp.4389-4408, 2010.
- [12] Y. L. Hsieh, D.-L. Yang and J. Wu, Using data mining to study upstream and downstream causal relationship in stock market, Proc. of 2006 Joint Conference on Information Sciences, Kaohsiung, Taiwan, 2006.
- [13] H. J. Hamilton and K. Karimi, The TIMERS II: Algorithm for the discovery of causality, Proc. of the 9th Paci\_c-Asia Conference on Knowledge Discovery and Data Mining, Hanoi, Vietnam, pp.744-750, 2005.
- [14] U. Faghihi, P. Fournier-Viger, R. Nkambou and P. Poirier, Generic episodic learning model imple- mented in a cognitive agent by means of temporal pattern mining, Proc. of IEA-AIE 2010, Cordoba, Spain, pp.438-449, 2010.
- [15] G. Dong and J. Pei, Sequence data mining, Springer Science + Business Media, LLC, 2007.
- [16] D. Lo and S.-C. Khoo, SMArTIC: Toward building an accurate, robust and scalable specication miner, Proc. of SIGSOFT Symposium on the Foundations of Software Engineering, pp.265-275, 2006 http://dx.doi.org/10.1145/1181775.1181808
- [17] J. Yang, D. Evans, D. Bhardwaj, T. Bhat and M. Das, Mining temporal API rules from imperfect traces, Proc. of International Conference on Software Engineering, pp.282-291, 2006.
- [18] J. S. Deogun and L. Jiang, Prediction mining { An approach to mining association rules for predic-tion, Proc. of RSFDGrC 2005 Conference, Regina, Canada, pp.98-108, 2005.
- [19] P. Fournier-Viger, U. Faghihi, R. Nkambou and E. M. Nguifo, CMRules: An efficient algorithm for mining sequential rules common to several sequences, Proc. of the 23th Int. Florida Artificial Intelligence Research Society Conference, Daytona, USA, pp.410-415, 2010.
- [20] P. Fournier-Viger, R. Nkambou and V. S. Tseng, RuleGrowth: Mining sequential rules common to several sequences by Pattern-Growth, SAC, TaiChung, Taiwan, 2011. http://dx.doi.org/10.1145/1982185.1982394
- [21] H. J. Hamilton and K. Karimi, The TIMERS II: Algorithm for the discovery of causality, Proc. of the 9th Paci\_c-Asia Conference on Knowledge Discovery and Data Mining, Hanoi, Vietnam, pp.744-750,2005.
- [22] S. K. Harms, J. Deogun and T. Tadesse, Discovering sequential association rules with constraints and time lags in multiple sequences, Proc. of the 13th Int. Symp. on Methodologies for Intelligent Systems, Lyon, France, pp.373-376, 2002.

http://dx.doi.org/10.1007/3-540-48050-1\_47

- [23] Y. L. Hsieh, D.-L. Yang and J. Wu, Using data mining to study upstream and downstream causal relationship in stock market, Proc. of 2006 Joint Conference on Information Sciences, Kaohsiung, Taiwan, 2006.
- [24] T.M. Hasbi Shiddieqy, Rijalul Hadis, Matahari Masa, 1970.
- [25] Tom Fomby, Association Rules (Affinity Analysis or Market Basket Analysis). Departemen of Economics Southern Methodist University Dallas. Texas, 2011.
- [26] J. and Kamber, M. Han, "Data Mining Concepts and Techniques Second Edition". Morgan Kauffman: San Francisco, 2006.
- [27] Mohammed J. Zaki and Wagner Meira JRData Mining and Analysis (Fundamental Concepts and Algorithms). New York. Cambridge University Press, .. 2014.
- [28] M.J. Zaki, SPADE: An Efficient Algorithm for Mining Frequent Sequences. Computer Science Department, Rensselaer Polytechnic Institute, Troy, 2001.
- [29] Yanchang Zhao and Yonghua Cen. Data Mining Applications with R. Academic Press : UK , Amsterdam , the Netherlands, 2014.