

The Implementation Of Apriori Algorithm And Chi-Square Test In Determining Pattern Of Relationship Among The Rawi Hadis

Rahmadi Yotenka

Department of Statistics Universitas Islam Indonesia, Jalan Kaliurang Km 14,5 Sleman
Yogyakarta
rahmadi.yotenka@uii.ac.id

ABSTRACT

In this research apriori algorithm was implemented on rawi data of hadits to find out pattern of relationship among the rawi hadits in shahih bukhori book. The analogy that can be compared is by assuming that series or chain of rawi in every sanad of hadis may be regarded as transaction, while rawi hadis were items in transaction. Data mining is the method that was used to analyze with association rule technique. Association rule technique is used to find a pattern rule between a combination of items. To find out association rule by using association rule application, it is used apriori algorithm by observing three important measurement, namely support, confidence, and lift values. The analysis result of apriori algorithm showed that for minimum support 0.03 and minimum confidence 0.9 having 9 strong association pattern based on the sequence of rawi hadis of its sanad. Every rule of association which was strong, then tested by chi-square to prove that the rawis that were in the rule were truly connected or statisticly significance.

Key Words : *Data Mining, Association Rule Mining, Apriori Algorithm, Chi square Test, Shahih Hadis of Bukhari, Rawi Hadis.*

ABSTRAK

Dalam penelitian ini algoritma apriori diimplementasikan terhadap data rawi hadis untuk menemukan pola hubungan yang terbentuk antar rawi hadis dalam Kitab Shahih Imam Bukhari. Analogi yang dapat diperbandingkan adalah dengan mengasumsikan bahwa rangkaian atau rantai rawi dalam tiap *sanad* hadis dapat dianggap sebagai transaksi, sedangkan rawi hadis adalah *item-item* yang berada dalam transaksi. Metode analisis yang digunakan adalah *data mining* dengan teknik *association rule mining*. Teknik *association rule mining* digunakan untuk menemukan aturan asosiasi antara suatu kombinasi item. Untuk menemukan aturan asosiasi di dalam aplikasi *association rule* digunakan algoritma *apriori* dengan memperhatikan tiga ukuran penting yaitu *support*, *confidence*, dan *lift*. Hasil analisis dari penerapan algoritma *apriori* menunjukkan bahwa untuk *minimum support* 0,03 dan *minimum confidence* 0,9 menghasilkan 9 aturan asosiasi *strong* berdasarkan urutan rawi hadis dalam *sanadnya*. Setiap aturan asosiasi *strong* ini selanjutnya dilakukan pengujian *chi-square* untuk membuktikan bahwa rawi-rawi yang berada dalam aturan benar-benar berhubungan atau signifikan secara statistik.

Kata Kunci : *Data Mining, Association Rule Mining, Algoritma Apriori, Uji Chi-Square, Hadis Shahih Imam Bukhari, Rawi Hadis.*

Pendahuluan

Menurut Al-Maliki (2005), salah satu aspek penting yang dibahas dalam ilmu *hadis* adalah tentang kaidah-kaidah untuk mengetahui keadaan *rawi (sanad)* dan redaksi (*matan*) suatu hadis, dari segi diterima dan ditolaknya. *Rawi* adalah orang yang meriwayatkan hadis. Hadis dapat diterima jika *sanad* dari *matan* hadis ini rawi-rawinya tidak terputus atau dengan kata lain bersambung dari permulaannya sampai pada akhir *sanad*. Artinya, menurut Imam Bukhari rawi-rawi hadis ini pernah bertemu atau berhubungan walaupun hanya satu kali. Dari hubungan ini membentuk suatu pola “*if-then*” antar rawi. Pola ini ditemukan tidak hanya pada satu *sanad* hadis saja, tetapi juga pada beberapa *sanad* hadis yang lain. Sehingga dibutuhkan metode yang tepat untuk melakukan analisis terhadap pola hubungan rawi hadis tersebut. Teknik data mining dibutuhkan untuk menganalisis kumpulan data *sanad* hadis beserta rawinya yang tersimpan dalam basis data hadis.

Salah satu teknik dalam data mining untuk mencari pola dalam kelompok data yang besar adalah *association rule mining* (Agrawal dan Srikant, 1994). Teknik *association rule*

mining ini menjelaskan kaitan aturan hubungan diantara rawi hadis dalam kombinasi hadis yang diriwayatkan. Kemudian, algoritma yang digunakan untuk menghasilkan aturan asosiasi dalam penelitian ini adalah algoritma apriori. Algoritma apriori merupakan salah satu algoritma yang banyak digunakan untuk proses analisis data-data yang selalu berkelompok dalam suatu transaksi besar. Misalnya, analisis terhadap keranjang belanja konsumen yang mengunjungi swalayan; tujuannya adalah menemukan aturan dari pembelian barang-barang yang cenderung muncul bersamaan dalam sebuah transaksi.

Algoritma apriori juga dapat diterapkan di bidang lain, misalnya dalam bidang komunikasi untuk analisis penggunaan dan pengembangan handphone pada masa depan (Kumar dan Muneeswaran, 2013), dalam bidang kesehatan untuk analisis pola hubungan antara penggunaan obat dan karakteristik pasien dalam mendeteksi pengaruh obat yang berbahaya (Kuo, *et al.*, 2009), dan lain sebagainya.

Kemudian dalam penelitian ini, algoritma apriori diimplementasikan untuk menemukan aturan asosiasi antar rawi hadis. Analogi yang dapat diperbandingkan adalah dengan mengasumsikan bahwa

rangkaian atau rantai rawi dalam tiap *sanad* hadis dapat dianggap sebagai transaksi, sedangkan rawi hadis adalah item-item yang berada dalam transaksi tersebut. Hasil akhir dari analisis *association rule mining* dengan penerapan algoritma apriori ini diharapkan untuk memperoleh aturan-aturan asosiasi antar rawi hadis yang bersifat *strong*. Dengan demikian aturan-aturan yang *strong* ini dapat digunakan untuk memperkuat argumen keshahihan suatu hadis. Kemudian aturan-aturan *strong* tersebut dilakukan pengujian *chi-square* untuk membuktikan bahwa rawi-rawi yang berada dalam aturan benar-benar berhubungan secara statistik dalam aturan yang terbentuk dan derajat keeratan hubungannya kuat.

Tujuan Penelitian

Beberapa tujuan yang ingin diperoleh pada penelitian ini adalah sebagai berikut:

1. Menerapkan algoritma apriori dalam menentukan pola hubungan antar rawi hadis.
2. Menemukan pola dan grafik asosiasi yang terbentuk sehingga dapat diketahui rawi-rawi mana saja yang banyak berhubungan dalam meriwayatkan hadis.

3. Membuktikan bahwa antar rawi hadis pada aturan asosiasi yang terbentuk memiliki hubungan secara statistik dalam proses periwayatan hadis.

Metode Penelitian

Sumber data dalam penelitian berasal dari kitab hadis shahih Imam Bukhari. Pada penelitian ini hadis yang digunakan sebanyak 500 hadis (hadis nomor 491 sampai nomor 991) dari Kitab Waktu–Waktu Sholat, Adzan, dan Jum’at.

Dalam penelitian ini variabel yang akan diteliti adalah rawi hadis. Yang dimaksud rawi hadis dalam penelitian ini adalah orang yang meriwayatkan hadis dari zaman Nabi Muhammad saw, *shahabat*, *tabi'in*, *tabi'ut tabi'in*, dan *tabi'ul atba'* sehingga *lafadz* hadisnya sampai kepada Imam Bukhari.

Penelitian ini bertujuan untuk menemukan pola hubungan antar rawi berdasarkan basis data hadis dalam kitab hadis shahih Imam Bukhari. Berdasarkan tujuan tersebut, digunakan metode *data mining* dengan teknik *association rule mining*. *Data mining* merupakan sebuah proses analisa dari sekumpulan (terkadang sangat besar) data, seperti data rawi hadis yang jumlahnya sangat banyak. *Association rule mining* merupakan salah satu teknik *data mining* untuk menemukan

aturan asosiasi antar variabel (rawi hadis) di dalam basis data transaksi. Yang dimaksud transaksi dalam penelitian ini adalah persambungan rawi hadis pada setiap hadis yang diriwayatkan, atau biasa disebut sebagai *sanad*. Contoh : *Sanad* hadis ke-1 = Umar bin al khaththab – Alqamah bin waqash – Muhammad bin Ibrahim – Yahya bin sa'id bin qais – Sufyan bin 'uyainah – Abdullah bin azzubair → disebut Transaksi 1, dan seterusnya sampai hadis ke-500 (transaksi 500). Selanjutnya, untuk menghasilkan aturan asosiasi di dalam program R 3.3.0, maka digunakan *algoritma apriori* dengan bantuan *package arules*. Kemudian juga digunakan *software* RapidMiner untuk membuat grafik pola aturan asosiasi. Aturan asosiasi yang bersifat *strong* kemudian diuji *chi-square* untuk mengetahui apakah *item-item* dalam aturan bersifat independen atau tidak.

Association Rule Mining

Association rule mining merupakan salah satu teknik data mining untuk menemukan aturan asosiasi antara suatu kombinasi item dengan item lainnya yang sering muncul bersamaan dengan jumlah atau frekuensi tertentu dari sekumpulan data. $I = \{i_1, i_2, \dots, i_m\}$ merupakan himpunan item. Misalnya, item-item seperti mentega,

roti, susu yang dibeli pada suatu swalayan; dan $A_i = v$ adalah item, dimana v adalah nilai domain dari atribut A_i , dalam relasi $R(A_1, \dots, A_n)$. Kumpulan dari item mengacu pada istilah itemset. X disebut *itemset* jika merupakan subset dari I .

$D = \{t_i, t_{i+1}, \dots, t_n\}$ adalah himpunan transaksi, disebut basis data transaksi, dimana setiap transaksi t mempunyai *tid* dan t -itemset $t = (tid, t\text{-itemset})$. Sebagai contoh, keranjang belanja konsumen yang telah diperiksa kasir disebut sebagai satu transaksi; tupel (v_1, \dots, v_n) dari relasi $R(A_1, \dots, A_n)$ sebagai sebuah transaksi. Transaksi t mengandung itemset X untuk semua item, dimana $i \in X$, dan i merupakan t -itemset (Zhang dan Zhang, 2002).

Association rule merupakan implikasi dari bentuk $X \rightarrow Y$, dimana $X \subset I$, $Y \subset I$, dan $X \cap Y = \emptyset$. X (atau Y) bisa terdiri dari item tunggal ataupun seluruh kumpulan item. X dapat dikatakan sebagai *antecedent* dan Y sebagai *consequent*. Setiap *association rule* mempunyai dua ukuran utama yang berkualitas, yakni *support* dan *confidence*.

a) *Support* aturan $X \rightarrow Y$ adalah $support(X \cap Y) = P(X \cap Y)$ didefinisikan sebagai (Brin, Motwani, dan Silverstein, 1997):

$$\text{supp}(X \rightarrow Y) = P(X \cap Y) = \frac{n(X \cap Y)}{n(D)} \dots (1)$$

b) *Confidence* aturan $X \rightarrow Y$ didefinisikan sebagai (Zhao, 2013):

$$P(Y|X) = \frac{P(X \cap Y)}{P(X)} \dots (2)$$

Selain kedua ukuran tersebut, salah satu cara yang lebih baik untuk mengetahui kekuatan suatu aturan asosiasi adalah dengan melihat nilai *lift*. Menurut (Brin, Motwani, dan Silverstein, 1997) *lift* didefinisikan sebagai:

$$\text{Lift}(X \rightarrow Y) = \frac{P(X \cap Y)}{P(X)P(Y)} \dots (3)$$

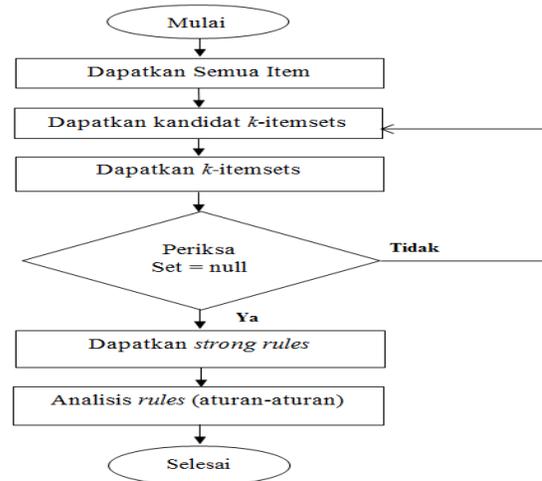
Association rule mining dapat dibagi menjadi dua sub permasalahan berikut:

1. Membangkitkan semua *itemset* yang mempunyai support lebih dari, atau sama dengan minimum *support* (*minsupp*) yang ditentukan peneliti. Artinya, membangkitkan semua *frequent itemset*.
2. Membangkitkan semua aturan yang mempunyai minimum *confidence* (*minconf*) sehingga dapat dijadikan sebagai aturan yang berlaku.

Algoritma Apriori

Apriori adalah algoritma populer dan banyak digunakan untuk identifikasi

mining frequent itemsets (Agrawal, Imielinski, dan Swami, 1993). Adapun tahapan-tahapan algoritma apriori digambarkan dalam diagram alir berikut:



Gambar 1. Diagram Alir Algoritma Apriori

Langkah pertama dalam *association rule mining* adalah menemukan *frequent itemsets*. Didefinisikan *k-itemset* sebagai sebuah *itemset* sebanyak *k* item, L_k sebagai himpunan dari *frequent k-itemsets*, dan C_k sebagai himpunan dari kandidat *k-itemset*. *Pseude-code* algoritma apriori berikut digunakan untuk menghasilkan semua *frequent itemset* dan melakukan pemangkasan *frequent itemset* yang tidak menarik dalam suatu basis data transaksi (Pang-Ning, Steinbach, Kumar, dan Vipin, 2005).

```

1: k=1.
2: Lk = {i | i ∈ I ∧ σ({i}) ≥ N × minsup}. {Dapatkan semua frequent 1-itemset}
3: repeat
4:   k=k+1
5:   Ck = apriori-gen (Lk-1). {Dihasilkan kandidat-kandidat itemset}
6:   for each transaction t ∈ D do
7:     Ct = subset (Ck, t). {Tentukan semua kandidat itemset Ck pada tiap t}
8:     for each candidate itemset c ∈ Ct do
9:       p(c) = p(c) + 1. {Menambahkan support count}
10:    end for
11:  end for
12:  Lk = {c | c ∈ Ck ∧ σ(c) ≥ N × minsup}. {Hilangkan kandidat kurang dari minsup}
13: until Lk = ∅
14: Result = ∪ Lk.
    
```

Gambar 2. Pseude-code Algoritma Apriori

Apabila *frequent itemsets* telah diperoleh dari basis data transaksi, maka dapat langsung dibangkitkan aturan-aturan asosiasi kuat (*strong*) dari *frequent itemsets* tersebut (dimana aturan-aturan asosiasi kuat memenuhi nilai *minimum support* dan *minimum confidence*). Hal ini dapat dilakukan menggunakan persamaan dari *confidence*.

Pengujian Chi-Square Pada Association Rule

Metode ini mengukur signifikansi asosiasi dalam menentukan korelasi seperti pada uji statistik klasik. Sehingga penggunaan uji *chi-square* dapat memberikan informasi apakah asosiasi antara item *A* dan *B* dalam aturan $A \rightarrow B$

dapat dinyatakan signifikan atau tidak secara statistik.

Uji *chi-square* untuk kombinasi kejadian item *A* dan *B* dalam basis data transaksi digunakan dalam Tabel 1 (Alvarez, 2003).

Tabel 1. Tabel Kontingensi Observasi

	<i>B</i>	\bar{B}
<i>A</i>	$nP(A \cap B)$	$nP(A \cap \bar{B})$
\bar{A}	$nP(\bar{A} \cap B)$	$nP(\bar{A} \cap \bar{B})$

Kemudian apabila item *A* dan *B* kejadian independen, maka Tabel 1, kontingensi observasi dibandingkan dengan Tabel 2, kontingensi ekspektasi.

Tabel 2. Tabel Kontingensi Ekspektasi

	<i>B</i>	\bar{B}
<i>A</i>	$nP(A)P(B)$	$nP(A)(1 - P(B))$
\bar{A}	$n(1 - P(A))P(B)$	$n(1 - P(A))(1 - P(B))$

Menurut (Bayardo dan Agrawal, 1999) nilai *support*, *confidence*, dan *lift* pada aturan $A \rightarrow B$ memenuhi karakteristik berikut (ketika penyebutnya tidak nol):

$$P(A \cap B) = \text{supp}$$

$$P(A) = \frac{\text{supp}}{\text{conf}} \dots (4)$$

$$P(B) = \frac{\text{conf}}{\text{lift}}$$

Kemudian tabel kontingensi untuk aturan asosiasi $A \rightarrow B$ yang diberikan dengan istilah *support*, *confidence*, dan *lift* dapat dilihat pada Tabel 3 dan Tabel 4 (diasumsikan nilai *conf* tidak nol):

Tabel 3. Tabel Kontingensi Observasi untuk $A \rightarrow B$

	B	\bar{B}
A	$n \text{ supp}$	$n \frac{\text{supp}}{\text{conf}} (1 - \text{conf})$
\bar{A}	$n \left(\frac{\text{conf}}{\text{lift}} - \text{supp} \right)$	$n \left(1 - \frac{\text{supp}}{\text{conf}} (1 - \text{conf}) - \frac{\text{conf}}{\text{lift}} \right)$

Tabel 4. Tabel Kontingensi Ekspektasi untuk $A \rightarrow B$

	B	\bar{B}
A	$n \frac{\text{supp}}{\text{lift}}$	$n \frac{\text{supp}}{\text{conf}} \left(1 - \frac{\text{conf}}{\text{lift}} \right)$
\bar{A}	$n \left(1 - \frac{\text{supp}}{\text{conf}} \right) \frac{\text{conf}}{\text{lift}}$	$n \left(1 - \frac{\text{supp}}{\text{conf}} \right) \left(1 - \frac{\text{conf}}{\text{lift}} \right)$

Berdasarkan kombinasi kejadian A_i dan B_j pada tabel kontingensi 2×2 , formula uji *chi-square*; dengan *double sum* untuk baris i dan kolom j adalah (Bain dan Engelhardt, 1992):

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \dots (5)$$

Kemudian pada (Alvarez, 2003) statistik *chi-square* pada persamaan (5) memenuhi persamaan berikut.

$$\chi^2 = n(\text{lift} - 1)^2 \frac{\text{supp} \times \text{conf}}{(\text{conf} - \text{supp})(\text{lift} - \text{conf})} \dots (6)$$

Hipotesis *chi-square* adalah:

$H_0: P_{ij} = P_i.P_j$ (hubungan independen)

$H_1: P_{ij} \neq P_i.P_j$ (hubungan dependen)

Pembahasan

Diketahui keseluruhan item adalah $I = \{r1, r2, \dots, r500\}$ dan keseluruhan transaksi adalah $TID = \{1, 2, 3, \dots, 500\}$. Data yang digunakan pada penelitian ini bersumber dari basis data hadis shahih Imam Bukhari. Dari basis data ini, peneliti mengumpulkan data sebanyak 500 hadis beserta *sanad* dan *matan*. Data-data tersebut digunakan untuk menemukan aturan asosiasi antar rawi hadis pada basis data. Ada dua parameter yang akan digunakan dalam proses analisis untuk mencapai hasil yang optimal. Kedua parameter tersebut adalah data *sanad* dan data rawi dari *sanad* hadis tersebut. Data *sanad* adalah transaksi sedangkan data rawi hadis adalah item dalam transaksi. Untuk mempermudah proses analisis, maka kedua parameter diringkas dan dikodekan dalam Tabel 5.

Tabel 5. Basis Data Hadis Dalam Penelitian

ID_Transaksi	Items
1	r1, r2, r3, r4, r5, r6
2	r7, r8, r9, r10, r11
3	r7, r8, r12, r13, r14, r15
4	r16, r17, r18, r19, r20
5	r16, r21, r12, r22, r23, r24
:	:
:	:
500	r74, r283, r284, r5, r6
Keterangan:	r = rawi

Berdasarkan data seperti Tabel 5, tahapan-tahapan dari algoritma apriori untuk membangkitkan *frequent itemset* adalah:

1. Setiap item dianggap sebagai kandidat 1-itemset yang disimbolkan dengan C_1 . Algoritma akan membaca semua transaksi secara berurutan untuk menentukan jumlah dari setiap item. *Support* dari tiap kandidat kemudian dihitung dan diuji terhadap batas *minsupp* yaitu 3%. Dengan kata lain, dalam kasus ini akan dicari rawi-rawi yang muncul sebanyak minimal 15 kali pada basis data hadis Imam Bukhari. Kandidat 1-itemset yang *support* nya lebih dari *minsupp* akan dijadikan *frequent 1-itemset*, L_1 .
2. Pada tahapan kedua, L_1 atau *frequent 1-itemsets* digunakan untuk mencari L_2 ,

frequent 2-itemsets. Kandidat 2-itemset (C_2) dihasilkan dengan cara menggabungkan (*join*) L_{k-1} dengan dirinya sendiri, dimana k merupakan urutan iterasi. Pada tahapan ini, L_1 digabung dengan L_1 , $L_1 \circ L_1$. Kandidat 2-itemset yang *support* nya lebih dari *minsupp* akan dijadikan *frequent 2-itemset*, L_2 .

3. Menggunakan prosedur yang sama, maka kandidat 3-itemsets, C_3 dapat dihitung. L_2 digabung L_2 , $L_2 \circ L_2$. Hasilnya jika kandidat 3-itemset yang *support* nya lebih dari *minsupp* akan dijadikan *frequent 3-itemset*, L_3 .
4. Kemudian pada tahapan berikutnya algoritma menggunakan $L_3 \circ L_3$ untuk menghasilkan *candidate 4-itemset*, C_4 . Hasilnya $C_4 = \Phi$ dan proses algoritma selesai sampai tahapan ini.
5. *Frequent itemset* telah diperoleh dari basis data transaksi, maka dapat langsung dibangkitkan aturan-aturan asosiasi kuat (*strong*) dari *frequent itemset* tersebut, dimana aturan-aturan asosiasi kuat memenuhi nilai *minsupp* = 3% dan *minconf* = 90% dihasilkan aturan asosiasi menarik dan bersifat *strong* yang ditampilkan pada Tabel 6.

Tabel 6. Aturan Asosiasi Antar Rawi Dengan *Minsupp* 3% dan *Minconf* 90%

ID Rules	Aturan Rawi	Support	Confidence	Lift
r83→r5	{Abdullah bin yusuf} => {Malik bin anas bin malik bin abi 'amir}	0,066	0,943	7,143
r85→r84	{Nafi' maula Ibnu 'Umar} => {Abdullah bin 'Umar bin Al khatthab}	0,064	0,970	8,081
r64→r63	{Al Hakam bin Nafi'} => {Syu'aib bin abi hamzah dinar}	0,052	1,000	17,857
r63→r64	{Syu'aib bin abi hamzah dinar} => {Al Hakam bin Nafi'}	0,052	0,929	17,857
r64, r4→r63	{Al Hakam bin Nafi', Muhammad bin muslim bin 'ubaidillah} => {Syu'aib bin abi hamzah dinar}	0,044	1,000	17,857
r4,r63→r64	{Muhammad bin muslim bin 'ubaidillah, Syu'aib bin abi hamzah dinar} => {Al Hakam bin Nafi'}	0,044	1,000	19,231
r6→r5	{Abdullah bin Maslamah} => {Malik bin anas bin malik bin abi 'amir}	0,042	0,955	7,231
r94→r84	{Salim bin 'Abdullah bin 'Umar} => {Abdullah bin 'Umar bin Al khatthab}	0,030	1,000	8,333
r74,r4→r3	{Aisyah binti abi bakar ash shiddiq, Muhammad bin muslim bin 'ubaidillah} => {Urwah bin az zubair bin al 'awwam}	0,030	0,938	14,648

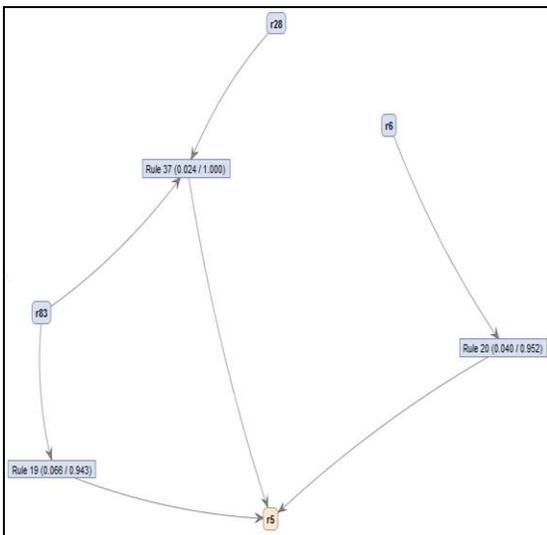
Menarik tidaknya suatu pola hubungan antara kombinasi *item* dapat diketahui dengan dua parameter yaitu *support* dan *confidence*. Semakin tinggi nilai *confidence* dan *support*, maka semakin kuat nilai hubungan antar atribut. Terdapat 9 aturan asosiasi yang terbentuk untuk *minimum confidence* 0,90 dan *minimum support* 0,03 dengan iterasi 2 kali (*Large 2-itemset*) dan 3 kali (*Large 3-itemset*). Untuk aturan asosiasi dengan iterasi 1 kali (*Large 1-itemset*) dan iterasi 4 kali (*Large 4-itemset*) tidak dipertahankan lagi (dikeluarkan) dari *output*. Hal ini dikarenakan nilai *confidencenya* tidak

mencapai 0,90 (90%) serta nilai *supportnya* juga tidak mencapai 0,03. Jumlah aturan asosiasi dengan iterasi 2 kali yaitu 6 aturan dan 3 kali iterasi yaitu 3 aturan.

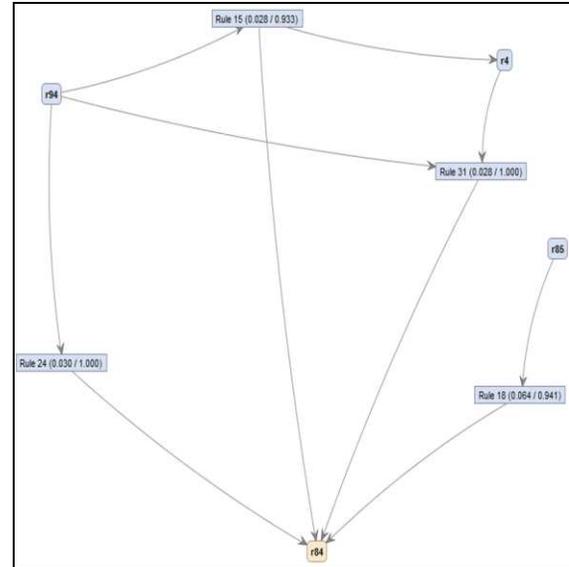
Aturan yang mempunyai hubungan antar atribut tertinggi yaitu aturan pertama (Abdullah bin yusuf => Malik bin anas bin malik bin abi 'amir) dengan nilai *support* 0,065 dan nilai *confidence* 0,943. Artinya, jika hadis tertentu diriwayatkan oleh Abdullah bin Yusuf, maka Malik bin anas bin malik bin abi 'amir juga meriwayatkan hadis tersebut, dengan tingkat kepercayaan 94,3 % atau dengan

kata lain 94,3 % dari hadis yang diriwayatkan oleh Abdullah bin Yusuf juga diriwayatkan oleh Malik bin anas bin malik bin abi 'amir.

Contoh visualisasi aturan asosiasi dengan suatu rawi sebagai *consequence* beserta nilai *support* dan *confidencenya* ditampilkan pada Gambar 3 dan Gambar 4. Kemudian, aturan asosiasi keseluruhan dengan sifat *strong* divisualisasikan dalam bentuk grafik *association rules* pada Gambar 5.

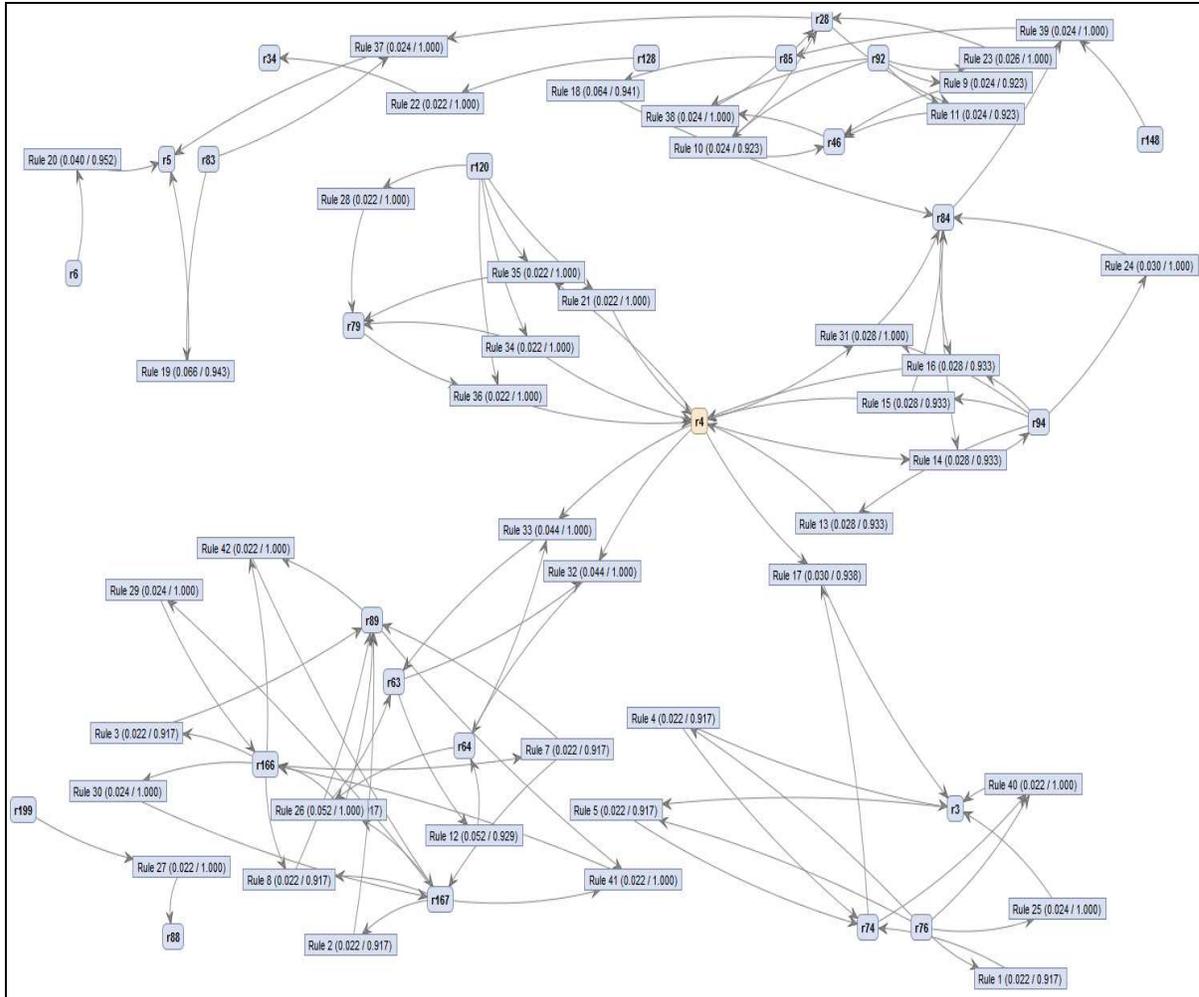


Gambar 3. Aturan Asosiasi Dengan r5 (Malik bin anas) Sebagai *Consequence*



Gambar 4. Aturan Asosiasi Dengan r84 (Abdullah bin 'Umar) Sebagai *Consequence*

Nilai signifikansi statistik pada suatu aturan asosiasi menggunakan uji *chi-square* untuk mengetahui dan membuktikan apakah antar rawi dalam aturan berhubungan atau tidak dalam proses periwayatan hadis. Hasil uji *chi-square* untuk *association rules* dengan $\alpha = 5\%$; $\chi^2_{tabel} = 3,84$ berdasarkan nilai *support*, *confidence*, dan *lift* ditampilkan pada Tabel 7.



Gambar 5. Visualisasi Aturan Asosiasi Rawi Hadis Dengan *Minsupp* 3% dan *Minconf* 90%

Tabel 7. Uji *Chi-Square* Pada Aturan Asosiasi

ID_Rules	χ^2_{hitung}	Keputusan H_0 /Kesimpulan
r83→r5	215,97	Ditolak/rawi 83 dan 5 berhubungan.
r85→r84	241,57	Ditolak/rawi 85 dan 84 berhubungan.
r63→r64	462,32	Ditolak/rawi 63 dan 64 berhubungan.
r6→r5	135,89	Ditolak/rawi 6 dan 5 berhubungan.

Kesimpulan

Berdasarkan hasil analisis dan pembahasan, diperoleh kesimpulan sebagai berikut:

1. Aplikasi *data mining* dapat digunakan untuk menemukan pola aturan (asosiasi) antar rawi hadis dalam Kitab Shahih Imam Bukhari untuk Kitab Waktu-Waktu Shotat, Adzan dan Jum'at. Berdasarkan hasil proses *association rules* dengan algoritma *apriori* diketahui bahwa untuk *minimum*

support 0,03 dan *minimum confidence* 0,90 menghasilkan 9 aturan asosiasi.

2. Berdasarkan kaidah ilmu hadis tentang urutan rawi dalam *sanadnya*, maka aturan asosiasi yang tetap dipertahankan yaitu sebanyak 2 *rules* dari 9 *rules* (*minimum support* = 0,03 dan *minimum confidence* = 0,90). Dari *rules* yang terbentuk tersebut diketahui bahwa : Aturan (Syu'aib bin abi hamzah dinar => Al Hakam bin Nafi') memiliki nilai *support* tertinggi yakni sebesar 0,052 dan nilai *confidence* sebesar 0,929 yang divisualisasikan dengan *id_rules* r63→r64. Kemudian aturan (Muhammad bin muslim bin 'ubaidillah, Syu'aib bin abi hamzah dinar} => {Al Hakam bin Nafi') mempunyai nilai *confidence* tertinggi, yakni sebesar 1,000, serta nilai *lift rasio* tertinggi (asosiasi terkuat) yaitu sebesar 19,21 divisualisasikan dengan *id_rules* r4,r63→r64 serta memiliki nilai *support* sebesar 0,044
3. Pengujian *chi-square* terhadap aturan asosiasi yang terbentuk dari proses algoritma apriori menunjukkan angka yang signifikan secara statistik. Hal ini berarti bahwa rawi-rawi yang terdapat pada masing-masing aturan memiliki hubungan atau berhubungan dalam

meriwayatkan hadis. Sehingga jika di dalam *sanad* suatu hadis mengandung aturan asosiasi rawi tersebut, maka argumen akan keshahihan hadis akan semakin kuat. Karena menurut Imam Bukhari, salah satu kriteria hadis shahih adalah hubungan antar rawi dalam *sanad* tidak terputus dan pernah bertemu.

Pustaka

- Agrawal, R., dan Srikant, R., 1994, Fast algorithm of mining association rule. Pada: *Proceedings of the 20th International Conference on VLDB*, Santigao, hal. 487–499.
- Agrawal, R., Imielinski, T., dan Swami, A., 1993, Mining Association rule Between Sets of Items in Large Databases, Pada: *Proceedings of ACM SIGMOD Conference on Management of Data on VLDB*, hal. 207–216.
- Al-Maliki, M. A., 2005, *Ilmu Ushul Hadis*. Yogyakarta: Pustaka Pelajar.
- Alvarez, S., 2003, *Chi-square Computation For Association rule: Preliminary Results. Technical Report BC-CS-2003-01. Computer Science Department, Boston College. USA. www.cs.bc.edu/ChiSquare/chi2tr.pdf*. Diakses tanggal 10 November 2014.
- Bain, L.J., dan Engelhardt, M., 1992, *Introduction to Probability and Mathematical Statistics*, Duxbury Press, California.

- Bayardo, R., dan Agrawal, R., 1999, *Mining The Most Interesting Rules*. Pada: *Proceedings The Fifth International SIGKDD Conference Knowledge Discovery and Data mining (KDD1999)*, hal. 145-154.
- Brin, S., Motwani, R., dan Silverstein, 1997, Beyond Market Baskets: General-izing Association rule to Correlations, Dari: *Proceedings ACM SIGMOD Conference on Management of Data (SIGMOD1997)*, hal. 265 – 276.
- Kumar, G., dan Muneeswaran, K.2013. *Intelligent Apriori Algorithm For Complex Activity Mining In Supermarket Applications*. Journal of Computer Science, **9(4)**: 433-438. Science Publications.
- Kuo, M., Kushniruk, A., Boryck, E., dan Greig, D., 2009, *Application Of The Apriori Algorithm For Adverse Drug Reaction Detection*. Jurnal di School of Health Information Science, University of Victoria. Canada: IOS Press.
- Pang-Ning, T., Steinbach, M., Kumar, dan Vipin, 2005, *Association Analysis Basic Concepts and Algorithms. Introduction to Data mining*. Addison-Wesley, ISBN 0-321-32136-7.
- Zhang, C., dan Zhang, S., 2002, *Association rule mining: Models and Algorithms*, Sydney: Springer. ISBN 3-540-43533-6.