



# Implementasi Metode Association Rule untuk Menganalisis Data Twitter tentang Badan Penyelenggara Jaminan Sosial dengan Algoritma Frequent Pattern-Growth

Jemaictry Tamaela<sup>a</sup>, Eko Sedyono<sup>b</sup>, Adi Setiawan<sup>c</sup>

<sup>ab</sup> Program Studi Magister Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana Salatiga

<sup>c</sup> Fakultas Sains dan Matematika, Universitas Kristen Satya Wacana Salatiga

Naskah Diterima : 25 Januari 2018; Diterima Publikasi : 29 Maret 2018

DOI : 10.21456/vol8iss1pp25-33

## Abstract

BPJS services cannot be separated from criticism and complaints of the people in Indonesia. Twitter is one of the social media choose to share experiences related to things about BPJS. The information that is shared can be processed to gain new knowledge (knowledge discovery), which is related to public opinion about BPJS. Tweets collected from the national BPJS twitter are divided into words, then, specified words can be used as items to form the itemset. The association rule technique with the FP-Growth algorithm that is implemented in the application can process text data from Twitter to form the item set. Each item set contains a collection of tweets that are responses and the opinion of the community about an event or phenomenon related to BPJS services. The tree structure of FP-Growth simplifies the process of the validation because it can track and display the frequency of occurrence of each word and itemset, before and after branch pruning which is not included in the support value. The OSM API integration with the application in this study provides visual information about where the tweet comes from, so it can be used to generate itemset from a collection of tweets from a particular region.

**Keywords:** Association rule; Data Mining; Knowledge Discover; FP-Growth; Twitter; BPJS.

## Abstrak

Layanan BPJS tidak lepas dari kritik dan keluhan masyarakat Indonesia. Twitter merupakan salah satu media sosial yang digunakan untuk membagikan pengalaman terkait hal-hal yang berhubungan dengan BPJS. Informasi yang dibagikan dapat diolah untuk mendapatkan pengetahuan baru (*knowledge discovery*), yang berhubungan dengan pendapat masyarakat tentang BPJS. *Tweet* yang dikumpulkan dari twitter BPJS nasional dipecah menjadi kumpulan kata, kemudian ditentukan kata-kata tertentu yang dapat dijadikan sebagai *item* untuk membentuk *itemset*. Teknik *association rule* dengan algoritme FP-Growth yang diimplementasikan pada aplikasi dapat mengolah data teks dari twitter untuk membentuk *itemset*. Setiap *itemset* berisi kumpulan *tweet* yang merupakan respon dan pendapat masyarakat tentang suatu kejadian atau fenomena terkait pelayanan BPJS. Struktur *tree* pada FP-Growth dapat mempermudah proses validasi karena dapat menelusuri dan menampilkan frekuensi kemunculan tiap kata dan *itemset*, sebelum dan sesudah pemangkasan cabang yang tidak masuk dalam nilai *support*. Integrasi OSM API dengan aplikasi pada penelitian ini memberikan informasi visual tentang lokasi asal *tweet* sehingga dapat dilakukan pembangkitan *itemset* dari kumpulan *tweet* yang berasal dari suatu daerah tertentu.

**Kata kunci:** Association rule; Data Mining; Knowledge Discover; FP-Growth; Twitter; BPJS.

## 1. Pendahuluan

Teknologi informasi memberikan kemudahan bagi manusia untuk memperoleh dan membagikan informasi. Twitter merupakan salah satu media sosial yang digunakan untuk membagikan informasi dari satu pengguna kepada pengguna-pengguna twitter yang lain. Informasi ini dapat berupa fakta maupun pendapat. Informasi yang dibagikan lewat Tweeter dikenal dengan sebutan *tweet*. *Tweet* merupakan pesan pendek yang hanya berjumlah 280 karakter. Masyarakat memanfaatkan twitter ini untuk

menyampaikan pengalaman, gagasan, keluhan, pendapat, maupun fakta yang dialami.

Badan Penyelenggara Jaminan Sosial (BPJS) merupakan lembaga penyelenggara jaminan sosial di Indonesia. BPJS bertanggung jawab memberikan jaminan sosial kepada setiap warga negara Indonesia meliputi usia tidak produktif, sakit, kecelakaan, dan bahkan kematian. BPJS dibentuk untuk menyelenggarakan Program Jaminan Sosial di Indonesia menurut Undang-undang Nomor 40 Tahun 2004 berisi tentang Sistem Jaminan Sosial Nasional. Berdasarkan Undang Undang Nomor 24 Tahun

\*) Penulis korespondensi: jeyzone77@gmail.com

2011, BPJS dibagi menjadi dua yaitu BPJS Kesehatan dan BPJS Ketenagakerjaan. BPJS berwenang menagih iuran, menempatkan dana, melakukan pengawasan dan pemeriksaan atas kepatuhan Peserta dan pemberi kerja (Ulinuha, 2014).

Layanan BPJS tidak lepas dari kritik dan keluhan masyarakat Indonesia. Masyarakat menyampaikan keluhan, pendapat, ataupun pengalaman menggunakan BPJS pada media yang dapat mereka akses. Twitter merupakan salah satu media sosial yang dipilih untuk membagikan pengalaman atau pendapat terkait hal yang berhubungan dengan BPJS. Pengguna aktif twitter di Indonesia pada tahun 2017 berjumlah 24,37 juta dari total 328 juta pengguna secara global (Herman dan Mononimbar, 2017). Informasi yang dibagikan oleh masyarakat melalui media twitter ini dapat dikumpulkan, kemudian digali untuk diperoleh informasi baru tentang suatu kejadian atau fenomena yang sedang terjadi. Teknik data *mining* merupakan teknik yang dapat digunakan untuk tujuan tersebut. Salah satu teknik *mining* adalah *association rule*. Teknik ini umum digunakan untuk mengenali pola transaksi pembelian barang. Teknik ini bekerja untuk menemukan kombinasi *item* (*itemset*) yang sering muncul.

Penelitian ini menggunakan data *tweet* sebanyak 44.791 *tweet* yang diambil dari akun twitter BPJS Nasional. Untuk menggali pengetahuan baru dari tumpukan data *tweet* yang besar, diperlukan algoritme yang memiliki kemampuan proses yang cepat. Frequent Pattern (FP)-Growth merupakan salah satu algoritme *association rule*, yang membentuk *frequency pattern tree* untuk mencari pola asosiasi. FP-Growth bekerja lebih unggul dari metode populer lainnya, seperti algoritme Apriori dan Tree Projection (Han *et al.*, 2004).

Pada penelitian ini, teknik *association rule* menggunakan algoritme FP-Growth diimplementasikan pada data *tweet* yang mengandung kata "BPJS". *Output* dari proses *mining* ini adalah pengetahuan baru (*knowledge discovery*). Tiap kata dianggap sebagai *item* dan kemunculan beberapa kata disebut sebagai *itemset*. yang berhubungan dengan pendapat masyarakat dan fenomena yang terjadi yang berkaitan dengan pelayanan BPJS. Pola kemuculan ini merupakan "rule", yang dapat dianggap sebagai tren.

## 2. Kerangka Teori

### 2.1 Penelitian Terkait

Salah satu penelitian yang membahas *text mining* dilakukan oleh Godfrey *et al.* (2014) yang bertujuan untuk melakukan analisis *cluster* dalam *text mining* dengan kumpulan teks yang besar untuk menemukan kesamaan antar dokumen dengan menggunakan 30.000 *tweet* yang diekstraksi dari twitter sesaat sebelum Piala Dunia dimulai. Penelitiannya kemudian

menggunakan analisis *cluster* untuk menemukan topik yang *tweet* deskripsikan menggunakan k-means, algoritme *clustering* yang umum digunakan, dan Non-Negative Matrix Factorization (NMF) dan membandingkan hasilnya. Kedua algoritme tersebut memberikan hasil yang serupa, namun NMF terbukti lebih cepat dan memberikan hasil yang lebih mudah ditafsirkan. Penelitian ini menggunakan dua alat visualisasi, Gephi dan Wordle.

Studi lain dilakukan oleh Ashktorab *et al.* (2014), dengan mengumpulkan *tweet* dari 12 krisis yang berbeda di Amerika Serikat sejak tahun 2006. Dalam penelitiannya, diperkenalkan *tweedr* yaitu alat twitter *mining* yang berfungsi mengekstrak informasi yang dapat ditindaklanjuti untuk para pekerja bantuan bencana selama bencana alam. *tweedr* terdiri dari tiga bagian utama: klasifikasi, pengelompokan dan ekstraksi. Pada tahap klasifikasi, berbagai metode klasifikasi (sLDA, SVM, dan regresi logistik) digunakan untuk mengidentifikasi *tweet* yang melaporkan kerusakan atau korban. Pada tahap pengelompokan, filter digunakan untuk menggabungkan *tweet* yang serupa satu sama lain, dan tahap terakhir yaitu ekstraksi, token dan frasa diekstraksi untuk menghasilkan laporan informasi spesifik tentang berbagai jenis kerusakan infrastruktur, jenis kerusakan, dan korban.

Penelitian lain yang membahas tentang *association rule* salah satunya dilakukan oleh Apriliyanti (2015). Penelitian tersebut dilakukan untuk mengklasifikasikan opini atau sentimen masyarakat terhadap batik melalui twitter dengan *hashtag* #batik. Metode yang dipakai adalah metode naive bayes, yang dalam hal ini klasifikasi dikategorikan dalam *supervised learning* yang berakibat opini tersebut terbagi ke dalam kutub sentimen positif dan negatif. Hasil penelitian menunjukkan bahwa sentimen positif terbentuk paling banyak yaitu 494 *tweet* dibandingkan dengan sentimen negatif yang hanya 5 *tweet*. Jadi secara keseluruhan, persepsi masyarakat mengenai batik yang tertuang pada jejaring sosial twitter memberikan penilaian yang positif.

### 2.2 Data Mining

*Data Mining* merupakan solusi yang mampu menemukan kandungan informasi yang tersembunyi berupa pola dan aturan sekumpulan data yang besar agar mudah dipahami (Chakrabarti *et al.*, 2015). *Data Mining* didefinisikan sebagai suatu proses untuk menemukan hubungan, pola, dan tren baru yang bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan. menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. *Data Mining* dan *knowledge discovery in database* (KDD) merupakan istilah yang memiliki konsep yang berbeda akan tetapi saling berkaitan karena *data mining* adalah bagian dalam proses KDD (Fayyad *et al.*, 1996).

### 2.3 Text Mining

*Text mining* juga dikenal sebagai KDD tekstual (Kushima *et al.*, 2017). *Text mining* adalah perkembangan teknologi yang membahas masalah pengelolaan informasi melalui penggunaan teknik data *mining*, pembelajaran mesin, pengolahan bahasa alami, pencarian informasi, dan manajemen pengetahuan. Lebih khusus lagi, *text mining* melibatkan pemrosesan kumpulan dokumen, atau korpus, di mana dokumen diubah menjadi data terstruktur, sehingga masing-masing dokumen digambarkan menggunakan seperangkat fitur yang disebut konsep untuk memberikan perspektif holisme dari informasi tekstual dan nontekstual (Ordenes *et al.*, 2014). *Text mining* adalah pengembangan teknologi dengan potensi komersial yang tinggi. Sebagai contoh, penelitian menunjukkan bahwa 80% informasi perusahaan terdapat dalam dokumen teks (Ur-Rahman dan Harding, 2012). Manfaat otomatisasi analisis sejumlah besar data umpan balik pelanggan yang terkait dengan layanan pariwisata, keuangan, dan manufaktur telah menunjukkan bagaimana mengelola dan mengubah ulasan *online* pelanggan secara efektif untuk membantu menginformasikan strategi bisnis (Ludwig *et al.*, 2013).

### 2.4 Association Rule

*Association rule* merupakan salah satu metode yang bertujuan mencari pola yang sering muncul di antara banyak transaksi, dimana setiap transaksi terdiri dari beberapa *item* (Zhang *et al.*, 2010). Korelasi antar *item* pada transaksi-transaksi seperti berikut: Pada transaksi yang terdapat *item X* terdapat kemungkinan ada *item Y* juga didalamnya kemudian dinotasikan, dimana *X* dan *Y* adalah *disjoint itemset*, dinotasikan. Kumpulan dari transaksi-transaksi ini disebut dengan *itemset*, yang dinotasikan dengan himpunan  $k$  ( $I_k$ ) ( $k=1, 2, \dots, m$ ). Jika terdapat *itemset* yang mempunyai *item* sebanyak  $k$ , maka disebut dengan *k-itemset* (Zaki, 2000). *Association rule* ini nantinya akan menghasilkan *rules* yang menentukan seberapa besar hubungan antar *X* dan *Y*, dan diperlukan dua ukuran untuk rules ini, yakni *support* dan *confidence*. *Support* merupakan kemungkinan *X* dan *Y* muncul bersamaan. *Confidence* merupakan kemungkinan munculnya *Y* ketika *X* muncul.

### 2.5 FP-Growth

Han *et al.* (2012) mengatakan bahwa Algoritme FP-Growth menggunakan struktur data yang disebut FP-tree (*Frequent Pattern tree*) dalam melakukan pencarian frekuensi *Pattern* atau *itemset*. Sesuai dengan namanya, FP-tree merupakan struktur pohon dimana setiap cabang berisi informasi frekuensi *itemset*, dan setiap *node* menyimpan informasi *item* dan frekuensinya. Urutan *parent* dan *child* dari pohon ini ditentukan oleh frekuensi masing-masing *item*. *Item* dengan frekuensi lebih besar akan menjadi

*parent* bagi *item* dengan frekuensi yang lebih kecil. Begitu pula sebaliknya *item* dengan frekuensi yang lebih kecil akan menjadi *child* bagi *item* yang lebih besar. Dengan demikian, yang menjadi daun dari pohon ini adalah *item* dengan frekuensi yang paling kecil. FP-tree memiliki sebuah *header table* yang berisi informasi *item* dan frekuensinya dan memiliki penunjuk ke *item* yang ada pada *tree* tersebut. Daftar *item* pada tabel ini disusun berdasarkan frekuensinya, dimulai dari *item* dengan frekuensi terbesar lalu disusul oleh *item* dengan frekuensi yang lebih kecil dan seterusnya. Setelah *header table* dan FP-tree dibuat, maka pencarian frekuensi *itemset* hanya dilakukan melalui struktur ini, sehingga sampai pada tahap ini pengecekan terhadap *database* tidak diperlukan lagi. Frekuensi *itemset* tingkat berikutnya, diperoleh dengan membuat kondisional FP-tree dengan algoritme FP-Growth. Algoritme ini bekerja secara rekursif dalam membuat kondisional FP-tree.

## 3. Metode

Data yang digunakan dalam penelitian ini diperoleh dari data twitter BPJS nasional pada periode Juli – Desember 2017 yang berjumlah 44.791 *tweet*. *Tweet* yang diunduh, memiliki kriteria yaitu setidaknya-tidaknyanya menyebutkan kata BPJS satu kali, dengan mengabaikan huruf besar atau kecil. Pada tiap *tweet* terdapat atribut ditunjukkan pada Tabel 1.

Table 1. Struktur data tweet

Atribut	Tipe Data	Contoh
StatusID	bigint	8934748538979
Text	varchar(200)	Sosialisasi Program BPJS. #indonesia #14like #14likes #like4like #like4likes #like4follow...
Screen Name	varchar(200)	Ansacut
Response		
CreatedAt	datetime	2017-08-04 14:12:59.000
Longitude	float	124.8395041
Latitude	float	1.49095862
Location	varchar(200)	indonesia, manado
UserDescription	varchar(200)	I am just a normal person
UserName	varchar(200)	ansa

Penelitian yang dilakukan, diselesaikan melalui tahapan KDD yaitu proses menentukan informasi yang berguna serta pola-pola yang ada dalam data. Informasi ini terkandung dalam basis data yang berukuran besar yang sebelumnya tidak diketahui dan potensial bermanfaat. *Data mining* merupakan salah satu langkah dari serangkaian proses iteratif KDD. Penelitian ini menggunakan langkah KDD yang dirancang oleh Fayyad *et al.* (1996). Tahapan proses KDD dapat dilihat pada Gambar 1.

Pada tahapan KDD yang dilakukan adalah sebagai berikut :

#### 1. Tahap *Data selection*.

Pada proses ini dilakukan pemilihan himpunan data, menciptakan himpunan data target, atau memfokuskan pada sampel data dimana penemuan

(*discovery*) akan dilakukan. Hasil seleksi disimpan dalam suatu berkas yang terpisah dari basis data operasional. Pada tahap ini dilakukan pengumpulan data *tweet* dari twitter BPJS.

#### 2. Tahap *processing*.

Proses ini dilakukan untuk membuang data yang tidak konsisten atau tidak diperlukan. Pada tahap ini *tweet* yang tidak sesuai tidak akan digunakan.

#### 3. Tahap *transformation*.

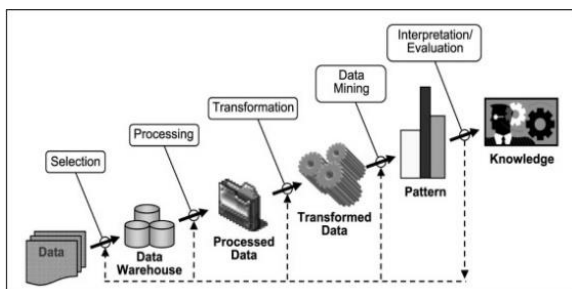
Proses ini mentransformasikan atau menggabungkan data ke dalam yang lebih tepat untuk melakukan proses *mining* dengan cara melakukan peringkasan (agregasi). Kata-kata yang akan digunakan sebagai *item* akan diringkas dengan menggabungkan kata-kata yang memiliki arti atau makna yang sama menjadi satu

#### 4. Tahap *data mining*.

Proses *data mining* yaitu proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik, metode atau algoritme tertentu sesuai dengan tujuan dari proses KDD secara keseluruhan. Pada tahap ini digunakan algoritme FP-Growth.

#### 5. Tahap *interpretation/evaluation*.

Proses interpretasi dilakukan untuk menterjemahkan pola-pola yang dihasilkan dari hasil *mining*. Dari hasil interpretasi dilakukan evaluasi apakah pola atau informasi yang ditemukan bersesuaian atau bertentangan. Interpretasi juga dilakukan secara visual dengan menghubungkan *tweet* dengan data geospasial menggunakan *OpenStreetMap* (OSM) API. Visualisasi ini bertujuan melihat informasi secara spesifik pada lokasi tertentu.



Gambar 1. Tahapan proses KDD (Fayyad *et al.*, 1996)

## 4. Hasil dan Pembahasan

Proses *data mining* dilakukan setelah dilakukan pemilihan data, pembersihan data dari *noise*, dan transformasi data. Proses pemilihan data dilakukan dengan cara menarik data dari *server* Twitter, dengan menggunakan Twitter API. Pencarian dilakukan dengan kata kunci “BPJS” dan disimpan dalam *database*.

Data twitter yang diperoleh diproses dengan cara dipecah menjadi kumpulan kata dan diambil kata-kata yang memiliki makna dan berkaitan dengan BPJS. Kata-kata yang diperoleh dalam proses ini

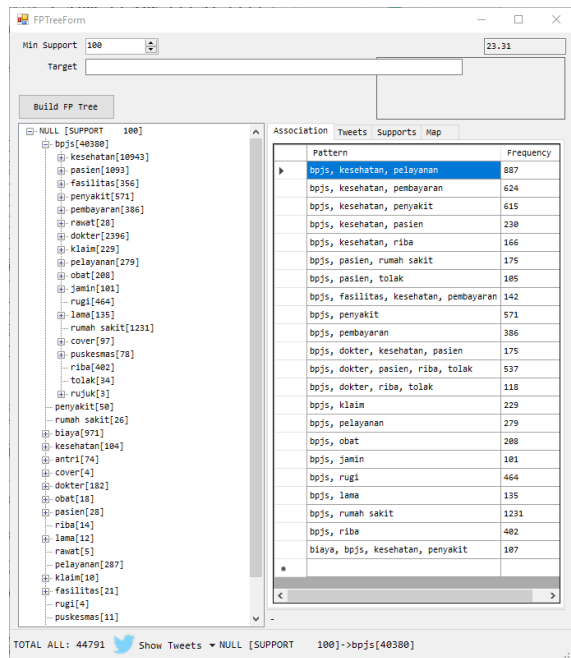
dijadikan sebagai *item* untuk membentuk *itemset* dalam proses *mining*. Kata-kata yang tidak sesuai atau tidak memenuhi syarat tidak akan digunakan. Penulisan kata-kata yang tidak baku, singkatan atau memiliki arti yang sama akan dijadikan satu sesuai dengan penulisan yang baku. Sebagai contoh kata “rumah sakit” yang ditulis “rs” atau “rmh skt” akan dianggap sama dan didaftarkan sebagai rumah sakit. Kata yang diperoleh kemudian disimpan dalam *database item list*. Daftar kata yang akan digunakan sebagai *item* untuk membentuk *itemset* dan frekuensi kemunculannya ditunjukkan pada Tabel 2.

Tabel 2. Kata dan frekuensi kemunculan

Peringkat	Kata	Frekuensi
1	bpjs	41328
2	kesehatan	13152
3	rumah sakit	3652
4	dokter	3281
5	pasien	2954
6	tolak	2205
7	riba	2074
8	jamin	1602
9	penyakit	1533
10	pembayaran	1289
11	pelayanan	1260
12	biaya	974
13	rugi	529
14	klaim	425
15	fasilitas	421
16	cover	392
17	rawat	331
18	obat	310
19	rujuk	227
20	puskesmas	180
21	lama	178
22	antri	173

Selain kata-kata yang digunakan sebagai *item*, kata-kata yang tidak sesuai atau mengandung konten negatif juga didaftarkan dalam *database* tersendiri yang digunakan sebagai *ignoring list* dengan tujuan untuk memfilter *tweet* yang tidak sesuai atau tidak digunakan dalam proses *mining*.

Semua proses *mining* dan pembangkitan *tree* dilakukan dengan menggunakan aplikasi yang dikembangkan sendiri, khusus untuk tujuan penelitian ini dengan mengimplentasikan algoritme FP-Growth. Aplikasi dikembangkan menggunakan Microsoft .Net Framework 4.6, dengan Visual Studio 2017 sebagai *development environment*. DBMS yang digunakan adalah SQL Server 2012 *Express Edition*. Tampilan aplikasi ditunjukkan pada Gambar 2.



Gambar 2. Tampilan aplikasi hasil implementasi FP-Growth

Proses *mining* dengan *association rule* dimulai dengan menentukan nilai *support*. Nilai *support* yaitu angka minimal frekuensi kata. Angka ini dapat berupa persentase atau angka jumlah. Nilai *support* menentukan banyaknya asosiasi yang dapat dihasilkan. Langkah menentukan nilai *support* tersebut dicapai dengan melakukan uji coba beberapa nilai *support* melalui aplikasi yang dikembangkan. Hasil uji coba ini ditunjukkan pada Tabel 3.

Tabel 3. Uji coba nilai *support*

Nilai <i>Support</i> yang Diujikan		Jumlah <i>Association Rule</i> yang Dihasilkan
Frekuensi	Persentase	
4000	8.93%	1
2000	4.47%	3
1000	2.23%	5
500	1.12%	10
250	0.56%	18
125	0.28%	24
100	0.22%	26

Nilai *support* pada Tabel 3 juga menentukan batas minimal kemunculan kata dalam seluruh transaksi. Kata yang masuk ke dalam batas *support* ini (memiliki frekuensi kemunculan lebih dari atau sama dengan nilai *support*), digunakan untuk membuat FP-tree. Kata yang tidak memenuhi syarat akan diabaikan. Semakin kecil nilai *support* yang dipakai, maka akan semakin besar jumlah kata yang digunakan. Tabel 4 menunjukkan perbandingan antara nilai *support* dengan jumlah kata yang masuk ke dalam *support*.

Tabel 4. Nilai *support* dan jumlah kata yang masuk dalam *support*

Nilai <i>Support</i>	Jumlah Kata masuk Dalam <i>Support</i>
4000	2
2000	8
1000	11
500	13
250	18
125	22
100	22

Penelitian ini menggunakan nilai *support* 100 (0, 22%), sehingga jumlah kata yang masuk dalam FP-tree adalah kata atau *item* yang memiliki frekuensi kemunculan lebih besar sama dengan 100 kali. Kata-kata yang muncul kurang dari 100 akan diabaikan. Nilai *support* 100 digunakan karena menghasilkan jumlah kata dan asosiasi lebih banyak dibandingkan nilai *support* lain pada tabel 3 dan tabel 4, akan tetapi nilai *support* lain juga dapat digunakan dengan menyesuaikan pada kebutuhan.

FP-tree dibangun dengan cara membaca tiap *tweet*, kemudian membaca kata-kata di dalam *tweet* tersebut dalam urutan peringkat kemunculan (lihat tabel 2). Jika terdapat kata yang berada di dalam daftar tabel 2, maka kata tersebut ditambahkan ke dalam *tree*. *Tree* dimulai dengan simbol “*null*” sebagai permulaan (*root*). Contoh *tweet* yang diproses dapat dilihat pada Tabel 5.

Tabel 5. Contoh *tweet*

No	Isi <i>tweet</i>
1	ada juga sih dokter yg tolak pasien bpjs, bukan karena riba sih, tapi karena nominal klaim bpjs-nya kecil
2	Ah iya dekat kost ada dokter yang bisa pake bpjs. Jangan2 tar disuruh ke rumah sakit besar
3	Apakah dokter boleh tolak pasien BPJS? Boleh. Syarat dan ketentuan berlaku

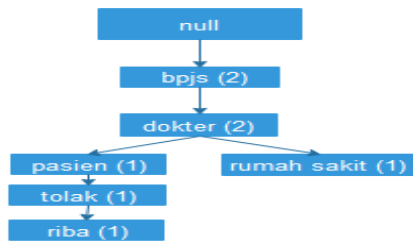
Berdasarkan *tweet* pada tabel 5, sebelum dapat dibentuk aturan asosiasi, *tweet* tersebut diuraikan kata-kata di dalamnya, kemudian diurutkan sesuai dengan peringkat kemunculannya (lihat Tabel 2). Hasil pengurutan kata pada *tweet* 1: bpjs, dokter, pasien, tolak, riba, *tweet* 2: bpjs, dokter, sakit rumah, dan *tweet* 3: bpjs, dokter, pasien, tolak. Setelah semua kata disusun berdasarkan peringkat kemunculannya, maka akan dibentuk FP-tree untuk menghasilkan pola asosiasi. Hasil pembentukan FP-tree ditunjukkan pada Gambar 3, Gambar 4, dan Gambar 5.



Gambar 3. Hasil *tree* dari *tweet* 1

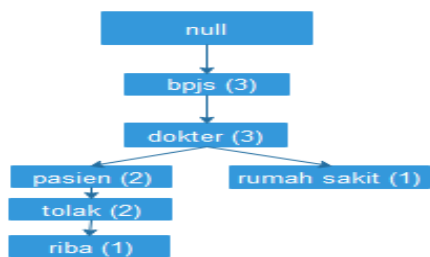
Pada Gambar 3, tiap kata pada *tree* memiliki nilai masing-masing satu. Proses pembentukan dimulai dari *null* kemudian diikuti kata yang memiliki frekuensi kemunculan terbanyak sesuai dengan Tabel 2.

Pada Gambar 4 menunjukkan *tree* setelah ditambahkan dengan *tweet* ke-2. Karena pada proses sebelumnya kata *bpjs* dan *dokter* telah terdaftar pada *tree*, maka nilai kata *bpjs* dan *dokter* bertambah menjadi dua, karena pada *tweet* ke-2 terdapat satu kata *bpjs* dan *dokter*. Cabang baru dibentuk pada kata *dokter* untuk menampung kata rumah sakit karena pada *tweet* 2 terdapat satu kata rumah sakit.



Gambar 4. Hasil *tree* dari *tweet* 1 dan 2

Pada Gambar 5 menunjukkan *tweet* ke-3 yang telah ditambahkan ke dalam *tree*, kata *bpjs* dan *dokter* bertambah nilainya menjadi tiga, dan juga kata *pasien* dan *tolak* bertambah menjadi dua.



Gambar 5. Hasil *tree* dari *tweet* 1, 2 dan 3

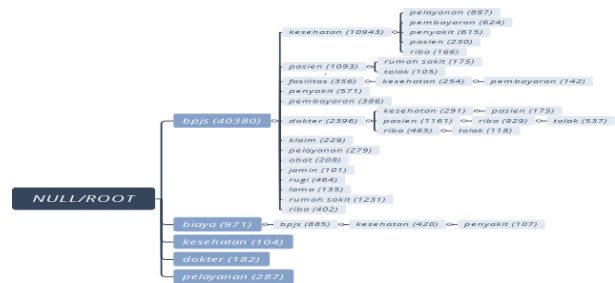
Setelah semua *tweet* diproses, maka akan terbentuk FP-tree dengan jumlah cabang yang semakin membesar (*growth*). Pada hasil akhir *tree*, dilakukan proses pemangkasan cabang, yaitu cabang yang memiliki kata dengan nilai frekuensi kemunculan kurang dari 100 akan dihapus. Pemangkasan cabang dilakukan setelah dilakukan proses validasi terhadap tiap kata yang muncul pada *tree*. Proses validasi pada FP-Growth dilakukan dengan cara menyesuaikan total nilai masing-masing kata yang muncul pada *tree* dengan total nilai frekuensi kemunculannya (lihat Tabel 2). Untuk mengetahui pola dan nilai kemunculan tiap kata pada *tree* dilakukan dengan menggunakan aplikasi yang dikembangkan. Gambar 6 menunjukkan frekuensi kata *support* dan hasil proses validasi.

Association		Tweets		Supports		Map	
Run Tree Validation							
Kata	Freq	Kata	Path				
antri	173	biaya	NULL [SUPPORT 100]\biaya[970]				
biaya	974	biaya	NULL [SUPPORT 100]\antri[169]\biaya[4]				
bpjs	41328	bpjs	NULL [SUPPORT 100]\bpjs[40293]				

Gambar 6. Frekuensi kata *support* dan hasil proses validasi *tree*

Sebagai contoh pada gambar 6, kata “biaya” memiliki frekuensi kemunculan 974. *Path* diperoleh dengan cara menelusuri tiap *note* pada *tree* dimulai dari *node* root (*null*) sampai ditemukan *node* yang mengandung kata “biaya”. Tiap *node* ”biaya” memiliki nilai frekuensi yang jika dijumlahkan, nilainya akan sama dengan jumlah frekuensi kemunculan kata “biaya” pada Tabel 2 yaitu 974. Karena total nilai kata pada *tree* sama dengan frekuensi kemunculan, maka *tree* ini memiliki nilai yang valid.

Setelah semua *tree* divalidasi nilai total masing-masing kata nya, dilakukan pemangkasan cabang untuk memangkas kata yang frekuensi kemunculan kurang dari 100. Hasil akhir *tree* ditunjukkan pada Gambar 7.



Gambar 7. Hasil akhir pembentukan *tree*

Hasil akhir *tree* setelah proses validasi dan pemangkasan cabang pada gambar 7 menunjukkan pola frekuensi (*frequency pattern*) kata-kata yang sering muncul secara bersamaan dalam satu *tweet*. Pola yang muncul dari *tree* akhir tersebut merupakan *itemset* yang disebut *association rule*. Contoh *Association rules* yang muncul setelah semua data twitter diproses menggunakan aplikasi yang dikembangkan dapat dilihat pada Tabel 6.

*Itemset* pada Tabel 6 merupakan *Association rules* yang dihasilkan dengan nilai frekuensi minimal 100. Setiap *itemset* yang terbentuk berisi kumpulan *tweet*, dimana setiap *tweet* mengandung semua kata yang membentuk *itemset*. Kumpulan *tweet* dalam setiap *itemset* merupakan respon atau pendapat masyarakat yang bersifat positif atau negatif terkait dengan permasalahan BPJS.

Tabel 6. Hasil asosiasi

No	Itemset	Frek	Pendapat
1	bpjs, rumah sakit	1231	positif, negatif
2	bpjs, kesehatan, pelayanan	887	positif, negatif
3	bpjs, kesehatan, pembayaran	624	positif
4	bpjs, kesehatan, penyakit	615	positif
5	bpjs, dokter, pasien, riba, tolak	537	negatif
6	bpjs, dokter, kesehatan, pasien	175	negatif
7	bpjs, fasilitas, kesehatan, pembayaran	142	positif
8	biaya, bpjs, kesehatan, penyakit	107	positif

Berdasarkan respon atau pendapat masyarakat ini juga dapat ditemukan adanya fenomena atau kejadian yang terjadi. Untuk validasi tiap *itemset* yang terbentuk, dapat dilakukan dengan memunculkan kumpulan *tweet* dari masing-masing *itemset* pada aplikasi yang dikembangkan, sehingga dapat diketahui apakah tiap *tweet* mengandung kata yang ada didalam kombinasi *itemset*. Hasil interpretasi dari tiap *itemset* yang diperoleh adalah sebagai berikut:

1. *Itemset* BPJS, Kesehatan, Pelayanan (887).

Kumpulan *tweet* pada *Itemset* ini berisi dua pendapat atau respon masyarakat yang berkaitan dengan BPJS. Pertama berisi pendapat atau respon positif masyarakat mengenai kemudahan pelayanan yang diberikan oleh BPJS untuk para pemudik.

*Itemset* ini diperoleh dari *tweet* yang sebagian besar memiliki atribut waktu yaitu sampai dengan Juni 2017. Sebagian *tweet* yang masuk ke dalam *itemset* ini, menunjukkan fenomena positif layanan BPJS.

Kedua pendapat atau respon negatif masyarakat terhadap keputusan Direksi BPJS terhadap pencabutan pelayanan kepada pekerja yang terkena PHK dan menurunnya kualitas layanan BPJS di daerah tertentu. *Tweet* pada *itemset* ini juga menunjukkan adanya fenomena negatif.

2. *Itemset* BPJS, Kesehatan, Pembayaran (624).

Kumpulan *tweet* pada *itemset* ini berisi pendapat positif masyarakat tentang kemudahan proses pembayaran iuran BPJS yang dapat dilakukan melalui mesin ATM, virtual account, dan layanan gojek. *Itemset* ini berisi pendapat terhadap fenomena positif.

3. *Itemset* BPJS, fasilitas, kesehatan, pembayaran (142).

Kumpulan *tweet* pada *Itemset* ini berisi pendapat positif masyarakat tentang proses pembayaran BPJS menggunakan fasilitas perbankan berupa *e-banking*. *Tweet* pada *itemset* ini menunjukkan fenomena positif.

4. *Itemset* BPJS, Kesehatan, Penyakit (615).

Kumpulan *tweet* pada *Itemset* ini berisi pendapat masyarakat dan klarifikasi pihak BPJS tentang informasi *hoax* dihapusnya delapan penyakit katastropik yang ditanggung oleh BPJS. *Tweet* pada *itemset* ini menunjukkan fenomena positif.

5. *Itemset* BPJS, dokter, pasien, riba, tolak (537).

Kumpulan *tweet* pada *itemset* ini berisi pendapat masyarakat tentang kasus-kasus dokter yang menolak pasien dengan alasan riba. *itemset* ini menunjukkan fenomena negatif.

6. *Itemset* BPJS, dokter, kesehatan, pasien (175).

Kumpulan *tweet* pada *itemset* ini berisi pendapat masyarakat tentang dokter yang terkesan kurang peduli dengan pasien yang menggunakan layanan BPJS dan adanya masalah riba. *Tweet* pada *itemset* ini menunjukkan fenomena negatif.

7. *Itemset* Biaya, BPJS, kesehatan, penyakit (107).

Kumpulan *tweet* pada *itemset* ini berisi pendapat masyarakat mengenai pembiayaan delapan penyakit katastropik yang tetap di tanggung BPJS. *Tweet* pada *itemset* ini menunjukkan fenomena positif.

8. *Itemset* BPJS, rumah sakit (1231).

Kumpulan *tweet* pada *itemset* ini berisi pendapat masyarakat terhadap hal yang bersifat negatif dan positif. Respon pertama merupakan respon negatif masyarakat mengenai masalah warga bekasi yang ditolak oleh 7 rumah sakit. *Tweet* pada *itemset* ini spesifik untuk lokasi Bekasi pada Juni 2017. Sebagian *tweet* pada *itemset* ini menunjukkan fenomena negatif.

*Tweet* pada *itemset* ini juga berisi pendapat atau respon masyarakat terhadap adanya fenomena negatif yaitu rumah sakit yang mempersulit pasien pengguna program BPJS dalam hal pelayanan.

*Tweet* dalam *Itemset* ini juga menunjukkan respon atau pendapat positif masyarakat yaitu tentang pernyataan Ridwan Kamil tentang akan dibangunnya rumah sakit baru

yang dikhususkan bagi warga kurang mampu, spesifik pada lokasi Bandung, Juni 2017. Sebagian *tweet* pada *itemset* ini menunjukkan fenomena positif.

Berdasarkan hasil pengolahan data twitter BPJS menggunakan aplikasi yang dikembangkan, teknik *association rule* dengan algoritme FP-Growth dapat membentuk *itemset*, yang berisi informasi tentang respon dan pendapat masyarakat terhadap suatu kejadian atau fenomena terkait dengan pelayanan BPJS. Setiap *itemset* yang terbentuk berisi kumpulan *tweet*, yang setiap *tweet* mengandung kombinasi kata dalam *itemset*. Teknik *association rule* dengan algoritme FP-Growth yang diimplementasikan mempermudah proses validasi setiap kata dan *itemset* yang terbentuk. Hal ini karena struktur *tree* pada FP-Growth memudahkan aplikasi dalam menelusuri dan menampilkan frekuensi kemunculan setiap kata dan *itemset* yang masuk dalam nilai *support* yang ditetapkan. Dengan demikian, setiap kombinasi kata dalam *itemset* dapat dipastikan jumlah total frekuensi kemunculannya secara bersamaan dalam setiap *tweet*.

Beberapa *tweet* memiliki atribut *latitude* dan *longitude*. Atribut ini dapat digunakan untuk menyaring twitter berdasarkan lokasi geospasialnya. Aplikasi diintegrasikan dengan *OpenStreetMap* (OSM) API. Layanan OSM ini digunakan untuk

menerjemahkan koordinat *latitude* dan *longitude*, menjadi nama kota/propinsi. Visualisasi data geospasial *tweet* diintegrasikan dengan OSM API diakses melalui menu *map* pada aplikasi seperti pada Gambar 9.



Gambar 9. Visualisasi lokasi pengiriman

Warna biru pada peta merupakan daerah atau lokasi dari mana *tweet* berasal. Proses *mining* juga dapat dilakukan untuk mencari asosiasi pada *tweet* berdasarkan lokasi geospasial suatu daerah tertentu, dengan tujuan untuk melihat pendapat masyarakat terhadap pelayanan BPJS di daerah tersebut, karena setiap daerah memiliki potensi kejadian atau fenomena yang berbeda terkait dengan pelayanan BPJS. Berdasarkan lokasinya, *tweet* paling banyak berasal dari pulau Jawa sebanyak 26304 *tweet*.

Dari hasil penelitian ini dapat diketahui bahwa data teks dari media *online* seperti twitter, dapat digunakan untuk menggali informasi terkait dengan pelayanan BPJS. Berbeda dengan pengumpulan data seperti yang dilakukan oleh Sari (2015), melalui observasi, wawancara, dan dokumentasi dengan pemilihan *sample* sebanyak 25 orang, pengumpulan data dari media twitter tidak memerlukan observasi, wawancara atau kuesioner karena dapat diatasi dengan melakukan *mining* melalui API twitter. Selain itu, tidak dibutuhkan pemilihan *sample* karena data yang diambil merupakan seluruh populasi pengguna twitter yang menyampaikan respon dan pendapat terkait BPJS. Widada *et al.* (2017) dalam penelitiannya menekankan pentingnya BPJS dalam peningkatan ketahanan masyarakat. Penelitian ini terfokus pada daerah Bengkulu dengan satu rumah sakit sebagai objek. Dengan pengumpulan data dari media *online*, memungkinkan cakupan daerah penelitian dapat diperluas, tidak hanya terfokus pada satu daerah tertentu dan informasi yang diperoleh juga lebih beragam, sehingga temuan yang dihasilkan dapat menjadi masukan kepada BPJS secara nasional sebagai bagian dari ketahanan masyarakat.

## 5. Kesimpulan

Berdasarkan hasil implementasi teknik *association rule* dengan algoritme FP-Growth dalam penelitian ini, dapat disimpulkan bahwa teknik *association rule* dengan algoritme FP-Growth dapat diimplementasikan pada data teks seperti *tweet*,

yang diperoleh dari akun twitter BPJS Nasional dengan *hashtag* BPJS.

Kata-kata tertentu dalam *tweet* dapat dijadikan *item* untuk membentuk *itemset*. Setiap *itemset* berisi kumpulan *tweet* yang memberikan informasi mengenai respon dan pendapat masyarakat terhadap suatu kejadian atau fenomena yang terjadi terkait dengan pelayanan BPJS.

Struktur *tree* dalam algoritme FP-Growth mempermudah aplikasi yang dikembangkan dalam proses validasi setiap *itemset*, yang menjadi *association rule*. Selain itu, implementasi teknik *association rule* dengan algoritme FP-Growth pada aplikasi dapat menampilkan secara keseluruhan frekuensi total kemunculan tiap kata dan *itemset* dalam *tweet*, sebelum dan sesudah pemangkasan cabang yang tidak masuk dalam batas minimal nilai *support*.

Integrasi OSM API dengan aplikasi pada penelitian ini, memberikan informasi visual tentang lokasi asal *tweet* sehingga dapat dilakukan pembangkitan *itemset* dari kumpulan *tweet* yang berasal dari suatu daerah tertentu. Berdasarkan lokasinya, *tweet* paling banyak berasal dari pulau Jawa sebanyak 26304 *tweet*.

Saran yang dapat dijadikan sebagai rekomendasi pengembangan adalah karena data yang digunakan dalam penelitian ini adalah data teks dari twitter, maka perlu ditentukan terlebih dahulu kata-kata yang dapat dijadikan sebagai *item* pada Tabel 2. Jumlah kata yang digunakan sebagai *item* di Tabel 2 masih sangat terbatas yaitu 22 kata. Oleh karena itu masih diperlukan penambahan kata untuk dapat menghasilkan kombinasi *itemset* yang lebih banyak dan mempermudah proses interpretasi, atau dengan mengembangkan teknik lain yang dapat digunakan untuk mengatasi hal ini. Semakin banyak kata yang dapat dijadikan sebagai *item*, kombinasi *itemset* yang dihasilkan akan lebih bervariasi, sehingga semakin banyak juga informasi yang dapat diperoleh.

## Daftar Pustaka

- Apriliyanti, 2015. Sentiment Analysis Dengan Naive Bayes untuk Melihat Persepsi Masyarakat Terhadap Batik Pada Jejaring Sosial Twitter, *Pros. Semin. Nas. Mat. dan Pendidik. Mat. UMS*.
- Ashktorab, Z., Brown, C., Nandi, M., Culotta, A. 2014. Tweedr: Mining Twitter to Inform Disaster Response. *International System for Crisis Response and Management*.
- Chakrabarti, S., Ester, M., Fayyad, U., Gehrke, J., Han, J., Morishita, S., Gregory Piatetsky-Shapiro, W. W. 2015. Data Mining Curriculum. <http://www.kdd.org/curriculum/index.html>. Diakses 5 Agustus 2017.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. 1996. From Data Mining to Knowledge Discovery in Databases. *AI magazine* 17, 37



- Godfrey, D., Johns, C., Meyer, C., Race, S., Sadek, C. 2014. A Case Study in Text Mining: Interpreting Twitter Data From World Cup Tweet. *Cornell University Library*, arXiv preprint rXiv:1408.5427
- Han, J., Pei, J., Yin, Y., Mao, R. 2004. Mining Frequent Patterns Without Candidate Generation: A Frequent-Pattern Tree Approach. *Data Mining and Knowledge Discovery* 8, 53–87. (doi:10.1023/B:DAMI.0000005258.31418.83)
- Han, J., Pei, J., Kamber, M. 2012. *Data Mining: Concepts and Technique 3rd Edition*, Morgan Kaufmann Publisher (doi:10.1017/CBO9781107415324.004)
- Herman., Mononimbar, D, A. 2017. Indonesia Fifth-largest-country-in-terms-twitter-users. Diakses 20 Februari 2018 <http://jakartaglobe.id/news/indonesia-fifth-largest-country-in-terms-twitter-users>.
- Kushima, M., Araki, K., Yamazaki, T., Araki, S., Ogawa, T., Sonehara, N. 2017. Text Data Mining of Care Life Log by the Level of Care Required Using KeyGraph. *In Proceedings of the International Multi Conference of Engineers and Computer Scientists*
- Ludwig, S., De Ruyter, K., Friedman, M., Brüggem, E.C., Wetzels, M. & Pfann, G. 2013. More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates. *Journal of Marketing* 77, 87–103.
- Ordenes, F.V., Theodoulidis, B., Burton, J., Gruber, T., Zaki, M. 2014. Analyzing Customer Experience Feedback Using Text Mining: A Llinguistics-based Approach. *Journal of Service Research* 17, 278–295.
- Sari, F.P., Syafrizal. 2015. Persepsi Masyarakat Pengguna Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan Mandiri Dalam Pelayanan RSUD Lubuk Basung Kabupaten Agam. Fakultas Ilmu Sosial dan Ilmu Politik, *Jurnal Online Mahasiswa* 2, No.2.
- Ulinuha, F.E. 2014. Kepuasan Pasien BPJS (Badan Penyelenggara Jaminan Sosial) Terhadap Pelayanan di Unit Rawat Jalan (URJ) Rumah Sakit Permata Medika Semarang Tahun 2014. Skripsi, Fakultas Kesehatan, Universitas Dian Nuswantoro
- Ur-Rahman, N., Harding, J. A. 2012. Textual Data Mining for Industrial Knowledge Management and Text Classification: A Business Oriented Approach. *Expert Systems with Applications* 39, 4729-4739.
- Widada, T., Pramusinto, A., Lazuadi, L. 2017. Peran Badan Penyelenggara Jaminan Sosial (BPJS) Penyelenggara Jaminan Sosial Terhadap Ketahanan Masyarakat (Studi di RSUD Damrah Mana, Kabupaten Bengkulu Selatan, Provinsi Bengkulu). *Jurnal Ketahanan Nasional* 23, No.2. (doi:http://dx.doi.org/10.22146/jkn.26388)
- Zaki, M. J. 2000. Generating Non-Redundant Association Rules. *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* , 34-43.
- Zhang, S., Zhang, C., Yang, Q. 2010. Data Preparation For Data Mining. *Applied Artificial Intelligence* 17, 2003.