

Implementasi *Association Rule Mining* Untuk Menganalisa Data Twitter Tentang Asuransi di Indonesia Menggunakan *Algoritma Frequent Pattern-Growth (FP-Growth)*

Hanifah Rizqy Mufidah, R.Mohamad Atok

Departemen Aktuaria, Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember

email: hanifahrizqy30@gmail.com, moh_atok@statistika.its.ac.id

Abstrak— Selama beberapa tahun belakangan ini, berdasarkan statistik yang dirilis Otoritas Jasa Keuangan (OJK) perkembangan asuransi di Indonesia menunjukkan trend yang positif pada lini asuransi sosial sebesar 5,65 T, *life insurance* sebesar 196,59 T, dan *non-life insurance* sebesar 49,35 T. Hal ini membuat perusahaan asuransi pun makin giat menunjukkan performa perusahaan yang mereka jalani sehingga semakin hari semakin banyak nasabah mulai menggunakan asuransi diperusahaannya untuk bisa menutupi risiko mereka dalam kesehatan, jiwa, dan produk asuransi lainnya. Twitter merupakan salah satu dari banyak sosial media yang digunakan masyarakat Indonesia untuk menyebarkan informasi yang sedang beredar kepada pengguna twitter lainnya dan juga dijadikan sebagai tempat berdiskusi yang bisa mendiskusikan banyak hal dari apa yang sedang *tranding* saat itu, hal yang disukai atau menarik untuk dibicarakan, dan terkadang ada pula yang menceritakan keluh kesah tentang keseharian mereka di twitter. Informasi yang disebarkan oleh pengguna twitter kemudian diolah menggunakan *software Rapidminer* untuk mendapatkan pengetahuan baru (*knowledge discover*) mengenai pendapat asuransi di Indonesia menggunakan proses *data mining* dengan metode *association rule* algoritma *FP-Growth*. Pada proses *association rule* digunakan nilai *support* dan nilai *confidence* hasil uji coba beberapa percobaan dengan hasil terbaik yang digunakan sebesar 0.03 dan 0.96. Serta hanya memasukan aturan yang memiliki *lift ratio* lebih dari 1. Hasil yang diperoleh dari proses *data mining* atau kata yang berhubungan dengan asuransi yaitu investasi, milik, dana, prudential, biaya, manfaat, keluarga, syariah, sehat, premi, rumah, klaim, sakit, jiwa, dan bpjs. Kata bpjs, jiwa, dan sakit menjadi 3 kata yang memiliki hubungan terkuat dengan asuransi. Jenis asuransi yang sering dibicarakan yaitu asuransi jiwa dan asuransi syariah sedangkan perusahaan yang sering dibicarakan yaitu perusahaan prudential.

Kata Kunci—Asuransi, Twitter, *Association Rule*, *FP-Growth*, *Data Mining*.

I. PENDAHULUAN

DEFINISI dari Asuransi itu sendiri ialah persetujuan dimana penanggung mengikatkan diri kepada tertanggung dengan mendapatkan premi, untuk mengganti kerugian, atau tidak diperolehnya keuntungan yang diharapkan, yang dapat diderita karena peristiwa yang tidak diketahui lebih dahulu. Selama beberapa tahun belakangan ini, perkembangan asuransi di Indonesia menunjukkan *trend* positif berdasarkan pada statistik yang dirilis oleh Otoritas Jasa Keuangan (OJK), pendapatan operasional dari asuransi sosial di Indonesia. Melihat perkembangan asuransi yang baik di Indonesia dan juga perkembangan pesat dari sisi teknologi yang dapat memunculkan ide atau pengetahuan-pengetahuan baru. Dalam hal ini peneliti ingin mencoba memaksimalkan peran teknologi informasi yakni sosial media twitter untuk mendapatkan penemuan pengetahuan mengenai karakter, layanan, dan perkembangan asuransi di Indonesia dari pendapatan masyarakat yang menggunakan twitter melalui teknik *data mining* dari data yang diperoleh dari *tweet* tiap pengguna twitter yang dikumpulkan di satu tempat dan kemudian digali untuk mendapatkan informasi baru. Hal ini bisa dilakukan disamping teknik ini biasa digunakan pada transaksi pembelian barang karena cara bekerjanya sama-sama dengan menemukan kombinasi item yang sering muncul. Pada teknik *Association Rule Mining* ini menggunakan algoritma *frequent pattern-growth* atau disingkat dengan *FP-Growth* karena pada algoritma ini menggunakan waktu yang lebih cepat dalam proses pencarian banyaknya itemset dibandingkan dengan algoritma yang umum dilakukan seperti algoritma *apriori* dan *tree projection*. Berdasarkan latar belakang tersebut, pada tugas akhir ini akan dilakukan penelitian menggunakan teknik *association rule* dengan algoritma *FP-Growth* pada *tweet* yang mengandung kata “Asuransi” atau “asuransi” dengan menggunakan bahasa Indonesia sebanyak 19.926 *tweets*. Atribut pada data penelitian ini yaitu hanya text dengan melalui proses data hingga mendapatkan format sesuai dengan aturan *software Rapidminer*. Hasil atau Output yang diharapkan pada penelitian ini adalah pengetahuan baru dari

hubungan atau asosiasi pada kata asuransi berdasarkan pendapat masyarakat Indonesia melalui sosial media twitter melalui metode *association rule* dengan algoritma *FP-Growth*.

II. TINJAUAN PUSTAKA

A. Asuransi

Menurut Undang-undang No. 40 Tahun 2014 pasal 1 Asuransi adalah perjanjian antara dua pihak, yaitu perusahaan asuransi dan pemegang polis, yang menjadi dasar bagi penerimaan premi oleh perusahaan asuransi sebagai imbalan untuk memberikan penggantian kepada tertanggung atau pemegang polis dan memberikan pembayaran yang didasarkan pada meninggalnya tertanggung.

B. Knowledge Discovery in Database

Knowledge Discovery in Database (KDD) adalah proses mengesktrak pola-pola yang menarik dari data yang jumlah yang sangat banyak berupa pengetahuan baru yang sulit diketahui secara manual. Kemudian untuk proses mendapatkan pengetahuan ini dilakukan dengan melalui beberapa tahap sebagai berikut [4].

1. Data Selection
2. Data Preprocessing
3. Data Transformation
4. Data Mining
5. Evaluasi
6. Knowledge

C. Text Processing

Text preprocessing adalah bagian penting dari sistem *Natural Language Processing* (NLP) apa pun, karena karakter, kata, dan kalimat yang diidentifikasi pada tahap ini adalah unit dasar yang diteruskan ke semua tahap pemrosesan lebih lanjut, mulai dari komponen analisis dan penandaan, seperti analisis morfologis dan *tagger part-of-speech*, melalui aplikasi, seperti pengambilan informasi dan sistem terjemahan mesin.

1. Tokenizing

Teks adalah data *unstructured* yang harus dirubah dahulu menjadi terstruktur sebelum dianalisis lebih lanjut. Teks *email* dimasukan kedalam aplikasi yang disimpan kedalam *array* 1 dimensi. Kata - kata dalam kalimat dibagi berdasarkan spasi.

2. Stopword removal

Kata-kata yang tidak memiliki arti yang relevan tersebut disebut *stopword*. Kumpulan dari *stopword* disebut *stop list* dan proses untuk menghapus *stopword* dalam dokumen disebut *stopword removal*.

3. Stemming

Stemming adalah proses pemetaan dan penguraian berbagai bentuk (variants) dari suatu kata menjadi bentuk kata dasarnya. Pada penelitian ini algoritma stemmer yang digunakan yaitu algoritma porter stemmer Bahasa Indonesia.

D. Data Mining

Istilah data mining memiliki beberapa padanan, seperti *knowledge discovery* ataupun *pattern recognition*. Kedua istilah tersebut sebenarnya memiliki ketepatanannya masing-masing. Istilah *knowledge discovery* atau penemuan pengetahuan tepat digunakan karena tujuan utama dari data mining memang untuk mendapatkan pengetahuan yang masih tersembunyi di dalam bongkahan data. Istilah *pattern recognition* atau pengenalan pola pun tepat untuk digunakan karena pengetahuan yang hendak digali memang berbentuk pola-pola yang mungkin juga masih perlu digali dari dalam bongkahan data yang tengah dihadapi.

E. Frequent Pattern – Growth

FP-Growth adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data. *FP growth* menggunakan pendekatan yang berbeda dari paradigma yang digunakan pada algoritma Apriori (Gunadi & Senses, 2012). Untuk menemukan *frequent itemset* dilakukan 2 tahapan, pertama pembuatan *FP-tree* kemudian penerapan algoritma *FP-Growth*. Pada penerapan *FP-Growth* dibagi menjadi 3 langkah: Tahap pembangkitan *conditional pattern base*, pembangkitan *conditional FP-Tree*, dan terakhir pencarian *frequent itemset* (*k-itemset*).

F. Association Rule

Association rule adalah suatu metode yang digunakan untuk mencari hubungan antar item suatu dataset yang telah ditentukan. *Association rule* mencari dan menemukan hubungan antar item yang ada pada suatu dataset [1]. Parameter yang digunakan pada metode ini adalah nilai *support* dan nilai *confidence*. Pengertian dari nilai *support* yaitu suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item atau itemset dari keseluruhan transaksi. Sedangkan pengertian nilai *confidence* adalah suatu ukuran yang menunjukkan hubungan antar dua item secara *conditional* (berdasarkan suatu kondisi tertentu).

Untuk menentukan nilai *support* digunakan rumus sebagai berikut.

$$\text{Support}(A) = \frac{\text{jumlah transaksi mengandung } A}{\text{Total Transaksi}} \quad (1)$$

Sementara itu, nilai *support* dari 2 item diperoleh dari rumus 2 berikut: $\text{Support}(A, B) = P(A \cap B)$

$$\text{Support}(A \cap B) = \frac{\sum \text{Transaksi yang Mengandung } A \text{ dan } B}{\sum \text{Transaksi}} \quad (2)$$

Sedangkan untuk mendapatkan nilai *confidence* digunakan rumus sebagai berikut.

$$\text{Confidence}(A \cap B) = P(B|A) = \frac{\text{Support}(A \cap B)}{\text{Support}(A)} \quad (3)$$

G. Lift Ratio

Cara lain untuk melihat kuat tidaknya aturan asosiasi yang terbentuk yaitu dengan menghitung nilai *lift ratio*. Rumus yang digunakan untuk mendapatkan nilai *lift ratio* sebagai berikut.

$$\text{Lift Ratio} = \frac{\text{Confidence}(A \rightarrow B)}{\text{Benchmark Confidence}(A, B)} \quad (4)$$

Sedangkan untuk menghitung *benchmark confidence* digunakan rumus sebagai berikut.

$$\text{Benchmark Confidence}(A, B) = \frac{N_c}{N} \quad (5)$$

Keterangan:

- N_c = Jumlah item dengan item yang menjadi *consequent*
- N = Jumlah transaksi di basis data

Apabila nilai *lift ratio* lebih besar dari 1 maka menunjukkan adanya manfaat atau hubungan yang kuat antar item. Jika nilai *lift ratio* kurang dari 1 maka hubungan asosiasi antar item berkorelasi negatif atau artinya kemunculan salah satu item mempengaruhi hal yang sebaliknya pada kemunculan item lainnya [3].

H. Word Cloud

Word cloud merupakan salah satu metode visualisasi data ber tipe teks yang sering digunakan atau diartikan dalam hal lain untuk representasi grafis dari sebuah dokumen yang dilakukan dengan *plotting* kata-kata yang sering muncul pada sebuah dokumen pada ruang dua dimensi. Kata yang muncul berdasarkan frekuensi pada data yang ditunjukkan melalui ukuran huruf kata tersebut. Semakin besar ukuran kata pada visualisasi *word cloud* menunjukkan semakin besar frekuensi kata tersebut muncul dalam data. Berikut merupakan contoh dari visualisasi data teks dengan *word cloud* [2].

III. METODOLOGI PENELITIAN

A. Sumber Data dan Variabel Data

Dalam penelitian ini digunakan data yang diperoleh dari sosial media twitter pada tiap tweet yang mengandung kata “Asuransi” atau “asuransi” dengan bahasa Indonesia pada periode November 2017 – Mei 2022 yang berjumlah 19.926 tweets. Variabel pada penelitian ini adalah variabel X yaitu kalimat (*tweet*) dan variabel Y yaitu kata yang terdapat pada *tweet*.

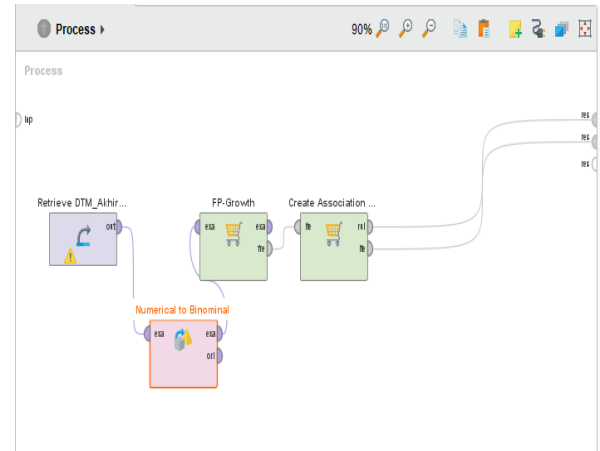
B. Tahapan Penelitian

1. Preprocessing Data

Setelah data dari twitter sudah terkumpul, dilakukan pengolahan *data mining* dengan mengikuti beberapa tahapan dalam KDD dan *text processing* untuk mendapatkan data yang dapat diolah oleh *software RapidMiner*. Tahapan untuk pengolahan dari data mentah atau data yang terkumpul dari twitter dilakukan melalui program *jupyter notebook* menjadi data akhir yang sesuai format *software Rapidminer* melalui tahapan *Data Selection, Cleaning Data, Stemming, Stopword Removal, Tokenizing, dan Data Transformation*.

- Menentukan minimal nilai *support* dan *confidence* yang akan digunakan pada tahap algoritma *frequent pattern-growth* dengan bantuan *software RapidMiner*.
- Implementasi FP – Growth pada *RapidMiner* ver. 9.10

RapidMiner memiliki 500 lebih metode *data science* dan salah satunya terdapat algoritma *FP-Growth* dalam *software* ini. Setelah data sudah berbentuk binomial untuk algoritma *FP-Growth* maka dapat dilanjutkan ke tahap implementasi data pada *RapidMiner*. Pada Gambar 1 merupakan contoh proses pengimplementasian *FP-Growth* pada *Rapidminer*.



Gambar 1.

Proses *FP-Growth* pada *Rapidminer*

4. Pencarian Aturan/rule Terbaik

Setelah mendapatkan hasil *running* dari dilakukan uji coba nilai *support* dan *confidence* menggunakan *software RapidMiner* untuk mendapatkan aturan yang terbaik. Nilai *support* maupun nilai *confidence* dapat berupa angka maupun persentase. Semakin rendah nilai *support* maka semakin banyak pula asosiasi yang terbentuk atau dapat diartikan data semakin bias. Dan juga perlu dipastikan nilai dari *lift ratio* pada asosiasi atau *rule* yang terbentuk melebihi 1.

5. Interpretasi Hasil

IV. HASIL DAN PEMBAHASAN

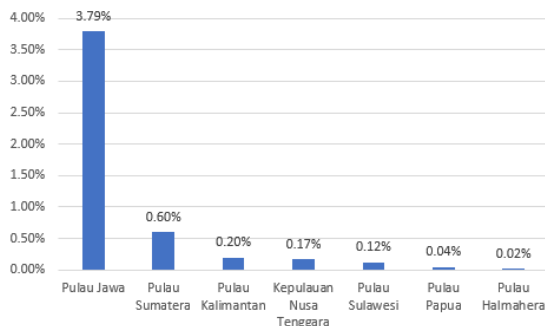
1. Karakteristik Data

Informasi yang dihasilkan dari data akan dipergunakan sebagai landasan pengetahuan bagi banyak orang dalam mengambil keputusan strategis. Dengan mengetahui karakteristik ini, diharapkan informasi yang dihasilkan dapat benar-benar relevan dan memiliki nilai tinggi, sehingga selanjutnya data dapat dipergunakan untuk mencapai tujuannya.

Pada data yang digunakan terdapat bermacam-macam pendapat yang tersebar di twitter mengenai asuransi di Indonesia. Untuk mengetahui persebaran lokasi akun yang berpendapat mengenai asuransi di Indonesia dibagi per pulau di Indonesia yang disajikan pada Gambar 2.

Berdasarkan Gambar 2 dapat dilihat bahwa total persebaran pengguna twitter yang berpendapat mengenai asuransi di Indonesia hanya 5% dari total data yang digunakan pada penelitian ini. Dari 5% terbagi ke beberapa lokasi, untuk lokasi tertinggi berada pada pulau Jawa sebesar 76,73%, yang kedua ada pada pulau Sumatera sebesar 12,2%, yang ketiga pada pulau Kalimantan 4,07%, berikutnya pada Kepulauan Nusa Tenggara 3,46%, dilanjut

pada pulau Sulawesi 2,44%, kemudian pada pulau Papua 0,81%, dan yang terkecil ada pada pulau Halmahera 0,3%. Karena nilai lokasi yang teridentifikasi terbilang cukup kecil hanya 5% saja maka sulit untuk dikatakan bahwa pada pulau Jawa lokasi yang paling sering berpendapat mengenai asuransi di Indonesia.



Gambar 2.
Persebaran Lokasi Yang Teridentifikasi

Data jumlah penduduk dari BPS Indonesia bahwa jumlah penduduk di pulau Jawa sebanyak 151,6 juta, pulau Sumatera 59,1 juta, pulau Sulawesi 19,7 juta, pulau Kalimantan 16,4 juta, Kepulauan Nusa Tenggara 15,1 juta, pulau Papua 4,3 juta, dan pulau Halmahera sebanyak 3 juta jiwa. Sehingga didapatkan proporsi persebaran *tweets* yang berpendapat mengenai asuransi di Indonesia per jumlah penduduk tiap pulau sebagai berikut pada Tabel 1.

Tabel 1

Proporsi Jumlah Penduduk Di Tiap Pulau Dengan *Tweets* Mengenai Asuransi

	Penduduk Indonesia	<i>Tweets</i>	Proporsi <i>tweets</i> dengan penduduk
Sumatera	21,92%	0,60%	0,03
Jawa	56,23%	3,79%	0,07
Kepulauan Nusa	5,60%	0,17%	0,03
Kalimantan	6,08%	0,20%	0,03
Sulawesi	7,30%	0,12%	0,02
Halmahera	1,12%	0,02%	0,01
Papua	1,59%	0,04%	0,02

Dari hasil proporsi persebaran *tweets* pada tiap pulau per jumlah penduduk yang ada di pulau tersebut pada Tabel 4.1, dihasilkan bahwa pulau Jawa menjadi pulau yang dominan atau kategori paling tinggi dengan penduduk yang membicarakan asuransi di twitter yaitu sebesar 0,07. Sehingga mendapatkan hasil bahwa akun yang berpendapat mengenai asuransi Indonesia di Twitter tidak merata pada tiap pulau di Indonesia.



Gambar 3.
Word Cloud Kata Asuransi

Berdasarkan *word cloud* atau awan kata yang diolah menggunakan *jupyter notebook* dari semua kalimat tweet yang mengandung kata asuransi menampilkan kata “sehat”, “bpjs”, “beli” yang memiliki jumlah frekuensi tertinggi dari ukuran kata yang paling terbesar diantara kata lainnya. Yang berarti ketiga kata tersebut merupakan kata yang paling sering muncul pada saat pengguna twitter berpendapat mengenai asuransi.

2. Data Preprocessing

Data yang diperoleh dari twitter kemudian di proses menjadi data yang sesuai format agar dapat digunakan untuk analisis berikutnya yaitu algoritma *FP-Growth* melalui beberapa tahapan *data processing* sebagai berikut.

a. Data Selection

Pada penelitian ini, digunakan hanya kalimat/*tweets* yang berhubungan dengan asuransi yang masih mengandung atribut seperti *link*, *hashtag*, *emoji*, angka, spasi berlebih, *username*, *hashtag*, huruf kapital, atau kata yang tidak mengandung arti.

b. Data Cleaning

Dilakukan penghilangan atribut yang tidak digunakan pada kalimat. Sehingga hasil dari *data cleaning* ini adalah kalimat yang teridiri dari kata dan tidak terdapat atribut seperti pada Tabel 2.

Tabel 2
Hasil Data *Cleaning*

Text	Text setelah <i>cleaning data</i>
@rahasia_ihhh	ya adalah kemarin ada
@Emak2Dasteran21 ya adalah,	yang nelson aku cewe
kemarin ada yang nelson aku	minta kenalan sama aku eh
cewe, minta kenalan sama aku,	tau nya nawarin asuransi
eh tau nya nawarin	asuransi...😭😭
Berkah Lebaran, Premi Asuransi	berkah lebaran premi
Perjalanan Ikut Terkerek	asuransi perjalan ikut
https://t.co/SOMIsh32ll	terkerek
#Asuransi	
@jaehyunkriuk yang punya	yang punya asuransi kan
asuransi kan bapakmu, bukan	bapakmu bukan kamu
kamu	

c. Stemming

Di tahap *stemming*, menghilangkan kata yang berimbuhan dengan menggunakan *library* di program *jupyter notebook* yaitu Sastrawi. Disajikan beberapa data dari hasil proses *stemming* pada Tabel 3.

Tabel 3
Hasil Data *Stemming*.

Text	Text setelah <i>stemming</i>
ya adalah kemarin ada yang	ya adalah kemarin ada yang
nelson aku cewe minta	nelson aku cewe minta
kenalan sama aku eh tau nya	kenalan sama aku eh tau
nawarin asuransi	nya nawarin asuransi
berkah lebaran premi asuransi	berkah lebaran premi
perjalan ikut terkerek	asuransi jalan ikut kerek
yang punya asuransi kan	yang punya asuransi kan
bapakmu bukan kamu	bapak bukan kamu

d. Stopword Removal

Pada proses ini dibuang kata yang tidak memiliki arti penting. Contoh *stopwords* dalam bahasa Indonesia

seperti yang, juga, dari, kamu, aku, saya, ini, itu, pada, jika, maka, pun, saja, dan lainnya. Ditunjukkan beberapa hasil data setelah proses *stopword removal* pada Tabel 4.

Tabel 4
Hasil *stopword removal*

Text	Text setelah <i>stopword removal</i>
ya adalah kemarin ada yang nelson cewe nelson aku cewe minta kenal nawarin asuransi kenalan sama aku eh tau nya nawarin asuransi	kemarin nelson cewe kenal nawarin asuransi
berkah lebaran premi asuransi jalan ikut kerek	berkah lebaran premi asuransi jalan kerek
yang punya asuransi kan bapak bukan kamu	asuransi

Setelah setiap kalimat sudah dihapus, dilakukan tahap *tokenize* atau pemisahan kata di dalam kalimat menjadi potongan kata tunggal atau token supaya *library countervectorizer* dapat menghitung frekuensi tiap kata dengan lebih efisien.

e. Data Transformation

Tahap terakhir yaitu mengubah data menjadi format yang diperlukan *Rapidminer* dengan bentuk biner 0 dan 1. Di Tabel 5 disajikan beberapa hasil perhitungan frekuensi tiap kata pada kalimat di twitter yang berpendapat mengenai asuransi.

Tabel 5
Format Data Yang Digunakan *Rapidminer*

	acak	asuransi	cewe	...	premi
Text 1	0	1	1	...	0
Text 2	0	1	0	...	1
Text 3	0	1	0	...	0

3. Menentukan Nilai *Support* Dan *Confidence*

Setelah data sudah sesuai format akhir untuk digunakan pada *software Rapidminer*, kemudian menentukan nilai *support* yang akan di implementasikan pada algoritma *FP-Growth* dengan membagi pada tiap minimal nilai *support* seperti pada Tabel 6.

Tabel 6
Jumlah Kata Pada Nilai *Support* Tertentu

Minimal Nilai Support	Jumlah Kata
0,1	1
0,09	1
0,08	1
0,07	2
0,06	5
0,05	7
0,04	11
0,03	23
0,02	60
0,01	144
0,009	166
0,008	186
0,007	215

Berdasarkan Tabel 6 digunakan minimal nilai *support* 0,02% pada nilai yang diberi warna kuning menghasilkan 60 kata untuk digunakan ke tahap analisis *FP-Growth* karena

menghasilkan kata yang tidak terlalu banyak dan juga tidak terlalu sedikit agar data tidak bias atau banyak kata mengandung arti yang sama.

Setelah didapatkan nilai *support* maka didapatkan nilai *confidence* yang akan digunakan. Pada penelitian ini digunakan nilai *confidence* 0,92 karena merupakan nilai *confidence* terendah dari nilai *support* 0,02 yang telah dipilih.

4. Implementasi *FP-Growth* pada *RapidMiner* ver.9.10

Software RapidMiner membantu berjalannya proses *FP-Growth* dari pembuatan *FP-Tree* hingga pencarian *frequent itemset*. Hasil akhir yang didapatkan dari seluruh data yang digunakan pada proses kali ini dengan menggunakan minimal *support* sebesar 0,02 dan *confidence* 0,92 dapat dilihat pada Tabel 7.

Tabel 7

Hasil *FP-Growth* minimal *support* 0,02 dan *confidence* 0,92

k	Frequent itemset
1	{asuransi}, {sehat}, {jiwa}, {beli}, {bayar}, {bpjs}, {uang}, {syariah}, {biaya}, {kerja}, {sakit}, {usaha}, {keluarga}, {prudential}, {masuk}, {investasi}, {dana}, {butuh}, {klaim}, {milik}, {salah}, {manfaat}, {pilih}, {indonesia}, {rumah}, {premi}, {produk}, {tanggung}, {lindung}, {lengkap}, {sesuai}, {kirim}, {aman}, {jalan}, {kantor}, {laku}, {langsung}, {swasta}, {mandiri}, {rs}, {barang}, {mobil}, {bank}, {agen}, {link}, {polis}, {yuk}, {wajib}, {hidup}, {kaya}, {nasabah}, {pakai}, {cari}, {sampe}, {tinggal}, {bikin}.
2	{asuransi, sehat}, {asuransi, jiwa}, {asuransi, beli}, {asuransi, bayar}, {asuransi, bpjs}, {asuransi, uang}, {asuransi, syariah}, {asuransi, biaya}, {asuransi, kerja}, {asuransi, sakit}, {asuransi, usaha}, {asuransi, keluarga}, {asuransi, prudential}, {asuransi, masuk}, {asuransi, investasi}, {asuransi, dana}, {asuransi, butuh}, {asuransi, klaim}, {asuransi, milik}, {asuransi, salah}, {asuransi, manfaat}, {asuransi, pilih}, {asuransi, indonesia}, {asuransi, rumah}, {asuransi, premi}, {asuransi, produk}, {asuransi, tanggung}, {asuransi, lindung}, {asuransi, lengkap}, {asuransi, sesuai}, {asuransi, kirim}, {asuransi, aman}, {asuransi, jalan}, {asuransi, kantor}, {asuransi, laku}, {asuransi, langsung}, {asuransi, swasta}, {asuransi, mandiri}, {asuransi, rs}, {asuransi, barang}, {asuransi, mobil}, {asuransi, bank}, {asuransi, agen}, {asuransi, link}, {asuransi, polis}, {asuransi, yuk}, {asuransi, wajib}, {asuransi, hidup}, {asuransi, kaya}, {asuransi, nasabah}, {asuransi, pakai}, {asuransi, cara}, {asuransi, sampe}, {asuransi, tinggal}, {asuransi, bikin}, {asuransi, prudential}.
3	{asuransi, syariah, prudential}.

Dari Tabel 7 diperoleh pola kemunculan 3-itemset telah memenuhi syarat minimal nilai *support* 0,02 dan nilai *confidence* 0,92. Kemunculan 3 itemset tersebut berarti kata yang paling banyak muncul pada data twitter tentang asuransi di Indonesia.

5. Pencarian Aturan/*rule* Terbaik

Dilakukan aturan/*rule* terbaik dari beberapa nilai *support* dan *confidence* agar pada tujuan dari penelitian ini untuk bisa memberikan saran mengenai asuransi di Indonesia bisa lebih jelas hasilnya karena apabila semakin

banyak aturan yang terbentuk maka saran yang diberikan bisa semakin bias dan tidak valid. Pada Tabel 8 hasil pencarian aturan terbaik.

Tabel 8

Hasil pencarian minimal nilai *support* dan *confidence* Terbaik

<i>Support</i>	<i>Confidence</i>	Jumlah aturan/ <i>rule</i> yang terbentuk
0,02	0,92	48
	0,96	45
	0,98	40
0,03	0,92	17
	0,96	15
	0,98	14
0,04	0,92	0
	0,96	0
	0,98	5
0,05	0,92	0
	0,96	0
	0,98	3

Berdasarkan hasil pencarian aturan terbaik yang telah dilakukan pada Tabel 8, baris yang diberi warna kuning adalah kombinasi aturan/*rule* yang terbaik, karena memiliki jumlah kata yang cukup menggambarkan kata asuransi di Indonesia dan juga apabila digunakan minimal nilai *support* 0,02 terlalu banyak aturan yang terbentuk. Alasan kedua apabila menggunakan nilai *support* 0,04 atau 0,05 menghasilkan terlalu sedikit kata dan tidak menghasilkan banyak pengetahuan baru. Hasil yang digunakan dengan nilai *support* 0,03 dan *confidence* 0,96 tersebut disajikan pada Tabel 9 sebagai berikut.

Tabel 9

Hasil uji coba minimal nilai *support* dan *confidence*

Item	Support	Confidence	Lift
syariah	0,053	0,975	1,017
investasi	0,034	0,98	1,002
milik	0,03	0,982	1,005
dana	0,031	0,99	1,013
prudential	0,036	0,99	1,013
manfaat	0,03	0,993	1,016
keluarga	0,036	0,993	1,016
biaya	0,045	0,994	1,014
sehat	0,109	0,994	1,017
premi	0,03	0,997	1,02
rumah	0,03	0,997	1,02
klaim	0,031	0,997	1,02
sakit	0,042	0,999	1,022
jiwa	0,071	0,999	1,022
bpjs	0,062	1	1,023

6. Interpretasi/evaluasi Hasil Asosiasi

Berdasarkan hasil proses *data mining* mengenai pendapat pengguna twitter mengenai asuransi di Indonesia dengan menggunakan metode *association*

rule diinterpretasikan hasil kata pada Tabel 9 atau yang berhubungan dengan asuransi sebagai berikut.

1. Kata “bpjs” (1226).

Terdapat 1226 kali kata “bpjs” muncul pada pendapat di twitter yang berhubungan dengan asuransi. Maksud bpjs pada asuransi yaitu program jaminan kesehatan dari pemerintah Indonesia agar seluruh rakyat Indonesia mendapatkan perlindungan dalam memenuhi kebutuhan dasar kesehatan.

2. Kata “jiwa” (1413).

Terdapat 1413 kali kata “jiwa” muncul pada pendapat di twitter yang berasosiasi dengan asuransi. Interpretasi kata jiwa pada asuransi adalah salah satu jenis asuransi yaitu asuransi jiwa. Menurut website manulife.co.id, asuransi jiwa adalah program perlindungan bagi keluarga apabila terjadi hal-hal yang tak diinginkan, seperti kematian, terhadap pemegang polis asuransi. Seperti contoh, perlindungan jiwa memang ditunjukkan untuk memberikan keamanan finansial serta perlindungan yang pasti pada keluarga yang ditinggalkan apabila tertanggung meninggal. Hubungan lainnya yaitu asuransi jiwa menjadi salah satu asuransi yang banyak dibicarakan atau bahkan digunakan masyarakat Indonesia.

3. Kata “sakit” (833)

Terdapat 833 kali kata “sakit” muncul pada pendapat di twitter yang berhubungan dengan asuransi. Maksud sakit adalah asuransi dapat menanggung biaya rawat jalan dan rawat inap atas risiko sakit maupun kecelakaan apabila orang tersebut memiliki asuransi. Hubungan pada kata sakit juga mempunyai arti lain apabila digabungkan dengan kata rumah menjadi rumah sakit.

4. Kata “klaim” (619).

Terdapat 619 kali kata “klaim” muncul pada pendapat di twitter yang berhubungan dengan asuransi. Hubungan antara klaim dengan asuransi adalah apabila seseorang memiliki asuransi kesehatan yang menanggung manfaat sakit *stroke*. Jika suatu saat penanggung sakit dan harus dirawat di rumah sakit karena penyakit *stroke*, maka penanggung dapat mengajukan hak klaim-nya kepada perusahaan sehingga penanggung tidak perlu membayar biaya rawat inap dan biaya lainnya sesuai polis asuransi yang dimilikinya.

5. Kata “premi” (590).

Terdapat 590 kali “kata” premi muncul pada pendapat di twitter yang berhubungan dengan asuransi. Maksud premi pada asuransi yaitu jumlah bayaran yang ditetapkan penanggung terhadap tertanggung sebagai biaya risiko yang ditanggung perusahaan asuransi apabila seseorang mempunyai asuransi. Contohnya

6. Kata “rumah” (595).

Terdapat 595 kali kata “rumah” muncul pada pendapat di twitter yang berhubungan dengan asuransi. Maksud rumah pada penelitian ini bergabung dengan kata sakit menjadi rumah sakit.

Hubungan antara rumah sakit dengan asuransi yang dibicarakan oleh pengguna twitter tentang bagaimana tiap warga yang mengunjungi rumah sakit untuk berobat ataupun sedang mengalami sakit, asuransi berperan penting dalam biaya pengobatannya.

7. Kata “sehat” (2175).

Terdapat 2175 kali kata “sehat” muncul pada pendapat di twitter yang berhubungan dengan asuransi. Maksud sehat pada asuransi berawal dari kata kesehatan yang mana menjadi jaminan setiap masyarakat untuk masalah kesehatan.

Hubungan antara kesehatan dengan asuransi yang dibicarakan oleh pengguna twitter yaitu asuransi kesehatan digunakan sebagai jaminan saat pemilik asuransi sedang sakit sehingga pemilik asuransi tidak mengeluarkan biaya yang terlalu besar jika sewaktu-waktu mengalami sakit.

8. Kata “syariah” (1053).

Terdapat 1053 kali kata “syariah” muncul pada pendapat di twitter yang berhubungan dengan asuransi. Maksud syariah pada asuransi adalah salah satu jenis asuransi yaitu asuransi syariah. Menurut website manulife.co.id, asuransi syariah adalah usaha untuk saling membantu dan berbagi di antara sejumlah orang atau pihak melalui investasi dalam bentuk aset atau *tabarru* yang memberikan pola pengembalian untuk menghadapi risiko tertentu menggunakan akad yang sesuai dengan syariah. Hubungan lainnya yaitu asuransi syariah menjadi salah satu asuransi yang banyak dibicarakan atau bahkan digunakan masyarakat Indonesia.

9. Kata “keluarga” (715).

Terdapat 715 kali kata “keluarga” muncul pada pendapat di twitter yang berhubungan dengan asuransi. Maksud keluarga pada asuransi yaitu kelompok orang yang mempunyai hubungan darah dan perkawinan yang berjumlah dua orang atau lebih.

Hubungan antara keluarga dan asuransi yang dibicarakan oleh pengguna twitter yaitu sebagai persiapan bagi setiap keluarga dalam menghadapi biaya yang tak terduga jika salah satu anggota keluarganya mengalami sakit dan membutuhkan biaya yang besar, asuransi bisa menjadi salah satu jawabannya.

10. Kata “manfaat” (605).

Terdapat 605 kali kata “manfaat” muncul pada pendapat di twitter yang berhubungan dengan asuransi, berarti banyak manfaat yang diperoleh jika seseorang memiliki asuransi untuk biaya kesehatan mereka.

11. Kata “biaya” (897).

Terdapat 897 kali kata “biaya” muncul pada pendapat di twitter yang berhubungan dengan asuransi. Kata biaya yang dimaksud yaitu pengeluaran yang digunakan untuk menghasilkan sesuatu produk ataupun jasa.

Hubungan antara biaya dengan asuransi yang

dibicarakan oleh pengguna twitter adalah banyak faktor yang mempengaruhi biaya produk asuransi tergantung dengan latar belakang mereka. Dan juga banyak masyarakat yang mengeluh mengenai biaya produk asuransi yang terlalu jauh dari anggaran finansialnya.

12. Kata “prudential” (715).

Terdapat 715 kali kata “prudential” muncul pada pendapat di twitter yang berhubungan dengan asuransi. Maksud kata prudential pada asuransi adalah salah satu perusahaan asuransi yang ada di Indonesia. Berarti banyak masyarakat di Indonesia yang membicarakan tentang perusahaan prudential dalam hal produk perusahaannya saat berpendapat di twitter mengenai asuransi di Indonesia.

13. Kata “dana” (628).

Terdapat 628 kali kata “dana” muncul pada pendapat di twitter yang berhubungan dengan asuransi. Kata dana yang dimaksud sama dengan pengertian biaya yaitu pengeluaran yang digunakan untuk menghasilkan sesuatu produk ataupun jasa. Hubungannya pun sama dengan biaya yaitu pada faktor harga produk asuransi supaya setiap masyarakat dapat membeli asuransi agar kebutuhan dasar kesehatannya dapat tercukupi.

14. Kata “milik” (615).

Terdapat 615 kali kata “milik” muncul pada pendapat di twitter yang berhubungan dengan asuransi, berarti banyak pengguna twitter yang membicarakan untuk setiap warga bisa memiliki asuransi sebagai cadangan keuangan jika sewaktu-waktu sakit dan membutuhkan biaya yang besar.

15. Kata “investasi” (684).

Terdapat 684 kali kata “investasi” muncul pada pendapat di twitter yang berhubungan dengan asuransi. Maksud kata investasi pada asuransi yaitu melindungi masyarakat yang dari risiko keuangan di masa depan.

Hubungan antara kata investasi dan asuransi yang dibicarakan oleh pengguna twitter yaitu asuransi bisa menjadi salah satu bentuk investasi seseorang untuk risiko finansial di masa mendatang.

V. KESIMPULAN DAN SARAN

A. Kesimpulan

1. Berdasarkan proses *data mining* dengan menggunakan metode *association rule* pada data twitter mengenai pendapat pengguna twitter tentang asuransi di Indonesia diperoleh kesimpulan sebagai berikut:

- Pada proses *association rule*, digunakan nilai *support* dan nilai *confidence* hasil uji coba terbaik yaitu nilai *support* sebesar 0.03 dan nilai *confidence* sebesar 0.96 dengan menghasilkan 15 kata yang berhubungan pada asuransi yaitu “bpjs”, “jiwa”, “sakit”, “klaim”, “premi”, “rumah”, “sehat”, “syariah”, “keluarga”, “manfaat”, “biaya”, “prudential”, “dana”, “milik”, dan “investasi”.

- b. Kata yang paling kuat hubungannya dengan asuransi adalah kata “bpjs” karena memiliki *lift ratio* tertinggi sebesar 1.023
 - c. Jenis asuransi yang banyak dibicarakan oleh masyarakat Indonesia adalah asuransi jiwa dan asuransi syariah.
 - d. Perusahaan yang paling banyak muncul pada *tweet* mengenai asuransi dan memiliki hubungan yang kuat dengan asuransi karena memiliki *lift ratio* 1.013 atau >1 yaitu perusahaan prudential yang bergerak di bidang asuransi.
 - e. Terdapat kata “manfaat” dan kata “milik” dari hasil yang didapatkan yang berarti pengguna twitter merasakan manfaat apabila memiliki asuransi.
2. Dari hasil yang didapatkan dari proses *association rule* menggunakan data twitter tentang asuransi dapat dihasilkan beberapa pengetahuan baru mengenai asuransi di Indonesia, yaitu:
- a. Berdasarkan aturan/*rule* yang terbentuk, jenis produk asuransi yang sering dibicarakan yaitu asuransi jiwa dan asuransi syariah sebesar 0,071% dan 0,053% dari pengguna twitter yang membicarakan asuransi di Indonesia. Oleh karena itu bagi perusahaan yang sudah memiliki jenis produk asuransi jiwa dan asuransi syariah sebaiknya lebih mengembangkan kedua jenis asuransi tersebut agar masyarakat tidak hanya membicarakan tetapi menggunakan jenis asuransi tersebut,
 - b. Dari hasil nilai *support* untuk kata “biaya” dan “dana” yang muncul sebesar 0,045% dan 0,031% dari seluruh pengguna twitter yang berpendapat mengenai asuransi di Indonesia. Apabila perusahaan asuransi ingin mengeluarkan produk mereka sebaiknya lebih mempertimbangkan biaya produk yang terjangkau banyak kalangan masyarakat.
 - c. Untuk perusahaan asuransi pemerintah yaitu bpjs yang menjadi kata paling berhubungan dengan asuransi untuk dapat terus bermanfaat bagi masyarakat dan memperbaiki kekurangannya supaya masyarakat Indonesia merasa terlindungi dari biaya kesehatan dan tidak merasa khawatir akan biaya pengobatan.
 - d. Dari hasil nilai *lift ratio* sebesar 1,002 pada kata “investasi” yang berarti ada hubungan kuat antara asuransi dengan investasi. Sehingga dapat dikatakan bahwa asuransi bisa menjadi alternatif seseorang dalam berinvestasi.
3. Pendapat mengenai asuransi di twitter oleh penduduk Pulau Jawa adalah paling tinggi proporsinya dibandingkan penduduk pulau lainnya di Indonesia.

B. Saran

1. Berdasarkan hasil *data mining* metode *association rule* dengan algoritma *FP-Growth* yang diperoleh dapat lebih diperkecil penelitiannya sampai kepada sentimen dari kata yang diperoleh, sehingga mendapatkan hasil apakah kata tersebut sering muncul dan berhubungan dari segi negatif atau positif.
2. Dapat melakukan sosialisasi mengenai asuransi selain pada pulau Jawa agar persebaran pengetahuan mengenai asuransi di Indonesia merata dan bisa mendapatkan manfaatnya.

DAFTAR PUSTAKA

- [1] Ardani, N. R., & Fitrina, N. (2016). Sistem Rekomendasi Pemesanan Sparepart Dengan Algoritma Fp-Growth (Studi Kasus Pt. Rosalia Surakarta). Seminar Nasional Teknologi Informasi Dan Multimedia 2016, 1–6. Rekomendasi, sparepart, Assosiation Rule, FP-Growth.%0A1.
- [2] Castella, Q., & Sutton, C. (2014). Word Storm: Multiples of Word Clouds for Visual Comparison of Documents.
- [3] Fitriyanto, E. T. (2017). PENENTUAN ATURAN ASOSIASI PADA TRANSAKSI PENJUALAN OBAT MENGGUNAKAN ALGORITMA APRIORI Studi Kasus Pada RSUD Dr. Soetrasno Rembang.
- [4] Kusriani, & Lutfhi, E. T. (2009). Algoritma Data Minign. In T. A. Prabawati (Ed.), Universitas Amikom. ANDI. <https://books.google.co.id/books?id=-Ojclag73O8C&printsec=frontcover#v=onepage&q&f=false>
- [5] Putra, A. A. A. (2016). Implementasi Text Summarization Menggunakan Metode Vector Space Model pada Artikel Berita Bahasa Indonesia. *Universitas Komputer Indonesia*.
- [6] Putri, R. E. (2021). Implementasi data mining untuk prediksi efektivitas pada mesin injection menggunakan algoritma c4. 5 studi kasus: pt. tridaya artaguna santara. *Repository.Uinjkt.Ac.Id*, 5. [https://repository.uinjkt.ac.id/dspace/handle/123456789/57846%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/57846/1/RIZKY EVITA PUTRI-FST.pdf](https://repository.uinjkt.ac.id/dspace/handle/123456789/57846%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/57846/1/RIZKY%20EVITA%20PUTRI-FST.pdf)