

TUGAS AKHIR - KA 184801

**IMPLEMENTASI ASSOCIATION RULE MINING UNTUK
MENGANALISIS DATA TWITTER TENTANG ASURANSI
DI INDONESIA MENGGUNAKAN ALGORITMA
FREQUENT PATTERN-GROWTH (FP-GROWTH)**

HANIFAH RIZQY MUFIDAH

NRP 06311840000042

Dosen Pembimbing

R. Mohamad Atok, S.Si, M.Si, Ph.D

NIP 197109151997021001

PROGRAM STUDI SARJANA SAINS AKTUARIA

DEPARTEMEN AKTUARIA

FAKULTAS SAINS DAN ANALITIKA DATA

INSTITUT TEKNOLOGI SEPULUH NOPEMBER

SURABAYA

2022



TUGAS AKHIR - KA 184801

**IMPLEMENTASI ASSOCIATION RULE MINING UNTUK
MENGANALISIS DATA TWITTER TENTANG ASURANSI
DI INDONESIA MENGGUNAKAN ALGORITMA
*FREQUENT PATTERN-GROWTH (FP-GROWTH)***

HANIFAH RIZQY MUFIDAH

NRP 06311840000042

Dosen Pembimbing

R. Mohamad Atok, S. Si, M. Si, Ph. D

NIP 197109151997021001

PROGRAM STUDI SARJANA SAINS AKTUARIA

DEPARTEMEN AKTUARIA

FAKULTAS SAINS DAN ANALITIKA DATA

INSTITUT TEKNOLOGI SEPULUH NOPEMBER

SURABAYA

2022



FINAL PROJECT - KA 184801

**IMPLEMENTATION OF ASSOCIATION RULE MINING
TO ANALYZE INSURANCE IN INDONESIA WITH
TWITTER DATA USING FREQUENT PATTERN-
GROWTH ALGORITHM (FP-GROWTH)**

HANIFAH RIZQY MUFIDAH

NRP 06311840000042

Advisor

R. Mohamad Atok, S. Si, M. Si, Ph. D

NIP 197109151997021001

UNDERGRADUATE PROGRAMME OF ACTUARIAL SCIENCE

DEPARTEMENT OF ACTUARIAL SCIENCE

FACULTY OF SCIENCE AND DATA ANALYTIC

INSTITUT TEKNOLOGI SEPULUH NOPEMBER

SURABAYA

2022

LEMBAR PENGESAHAN

IMPLEMENTASI ASSOCIATION RULE MINING UNTUK MENGANALISIS DATA TWITTER TENTANG ASURANSI DI INDONESIA MENGGUNAKAN ALGORITMA FREQUENT PATTERN-GROWTH (FP-GROWTH)




TUGAS AKHIR

Diajukan untuk memenuhi salah satu syarat
memperoleh gelar Sarjana Ilmu Aktuaria pada
Program Studi Sarjana Sains Aktuaria
Departemen Aktuaria
Fakultas Sains dan Analitika Data
Institut Teknologi Sepuluh Nopember

Oleh: **HANIFAH RIZQY MUFIDAH**

NRP. 0631184000042

Disetujui oleh Tim Penguji Tugas Akhir:

- | | | |
|------------------------------------|------------|---|
| 1. R. Mohamad Atok, M.Si, Ph. D. | Pembimbing | () |
| 2. Pratnya Paramitha O, S.Si, M.Si | Penguji | () |
| 3. Imam Safawi Ahmad, S.Si, M.Si | Penguji | () |

SURABAYA

Juli, 2022

APPROVAL SHEET

IMPLEMENTATION OF ASSOCIATION RULE MINING TO ANALYZE INSURANCE IN INDONESIA WITH TWITTER DATA USING FREQUENT PATTERN-GROWTH ALGORITHM (FP-GROWTH)




FINAL PROJECT

Submitted to fulfill one of the requirements
for obtaining a degree Bachelor of Actuarial Science at
Undergraduate Study Program of Actuarial Science
Departement of Actuarial Science
Faculty of Science and Data Analytics
Institut Teknologi Sepuluh Nopember

By: **HANIFAH RIZQY MUFIDAH**

NRP. 0631184000042

Approved by Final Project Examiner Team:

- | | | |
|-------------------------------------|----------|---|
| 1. R. Mohammad Atok, M.Si, Ph. D. | Advisor | () |
| 2. Pratnya Paramitha O., S.Si, M.Si | Examiner | () |
| 3. Imam Safawi Ahmad, S.Si, M.Si | Examiner | () |

SURABAYA

July, 2022

PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini:

Nama mahasiswa / NRP : Hanifah Rizqy Mufidah / 06311840000042
Program studi : Aktuaria
Dosen Pembimbing / NIP : R. Mohammad Atok, S.Si, M.Si, Ph.D /
197109151997021001

dengan ini menyatakan bahwa Tugas Akhir dengan judul “Implementasi *association rule mining* untuk menganalisis data twitter tentang asuransi di Indonesia menggunakan algoritma *frequent pattern-growth (FP-Growth)*” adalah hasil karya sendiri, bersifat orisinal, dan ditulis dengan mengikuti kaidah penulisan ilmiah.

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia menerima sanksi sesuai dengan ketentuan yang berlaku di Institut Teknologi Sepuluh Nopember.

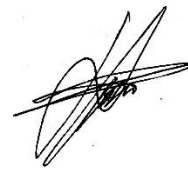
Surabaya, 20 Juli 2022

Mengetahui
Dosen Pembimbing



R. Mohamad Atok, S.Si, M.Si, Ph.D
NIP. 197109151997021001

Mahasiswa



Hanifah Rizqy Mufidah
NRP. 06311840000042

STATEMENT OF ORIGINALITY

The undersigned below:

Name of student/ NRP : Hanifah Rizqy Mufidah / 06311840000042
Department : Aktuaria
Advisor / NIP : R. Mohammad Atok, S.Si, M.Si, Ph.D /
197109151997021001

hereby declare that the Final Project with the title of “Implementation of Association Rule Mining To Analyze Insurance in Indonesia With Twitter Data Using Frequent Pattern-Growth Algorithm (*FP-Growth*)” is the result of my own work, is original, and is written by following the rules of scientific writing.

If in the future there is a discrepancy with this statement, then I am willing accept sanctions in accordance with the provisions that apply at Institut Teknologi Sepuluh Nopember.

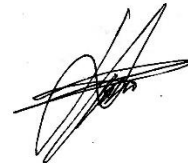
Surabaya, 20 Juli 2022

Acknowledged
Advisor



R. Mohamad Atok, S.Si, M.Si, Ph.D
NIP. 197109151997021001

Student



Hanifah Rizqy Mufidah
NRP 06311840000042

ABSTRAK

IMPLEMENTASI ASSOCIATION RULE MINING MENGGUNAKAN DATA TWITTER MENGENAI ASURANSI DI INDONESIA MENGGUNAKAN ALGORITMA FREQUENT PATTERN-GROWTH (FP-GROWTH)

Nama Mahasiswa / NRP : Hanifah Rizqy Mufidah / 06311840000042
Departemen : Aktuaria FSAD - ITS
Dosen Pembimbing : R. Mohammad Atok, S.Si, M.Si, Ph.D

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.

ABSTRACT

IMPLEMENTATION OF ASSOCIATION RULE MINING TO ANALYZE INSURANCE IN INDONESIA WITH TWITTER DATA USING FREQUENT PATTERN-GROWTH ALGORITHM (FP-GROWTH)

Student Name / NRP : Hanifah Rizqy Mufidah / 06311840000042
Department : Actuarial Science FSAD - ITS
Advisor : R. Mohammad Atok, S.Si, M.Si, Ph.D

Abstract

Over the past few years, insurance developments in Indonesia have shown a positive trend based on statistics released by the OJK in the social insurance line of 5.65 T, life insurance at 196.59 T, and non-life insurance at 49.35 T. Insurance companies are also increasingly active in showing the performance of the companies they live, which more and more customers are starting to use insurance in their companies to be able to cover their risks in health, life, and other insurance products. Twitter is one of the many social media used by Indonesian people to disseminate information that is currently circulating to other Twitter users and is also used as a place for discussion that can discuss many things from what is presently trending, things that are like or interesting to talk about, and sometimes some talk about their daily lives on Twitter. Twitter users can then process information to obtain new knowledge regarding insurance opinions in Indonesia using software application named *Rapidminer* for data mining process using the FP-Growth algorithm association rule method to form rules. In the association rule process, the support value and the confidence value of the results of several trials are 0.03 and 0.96. And include regulations that have a lift ratio of more than 1. The results obtained from the data mining process or related to the insurance obtained in 15 words are investment, property, fund, prudential, cost, benefit, family, sharia, healthy, premium, house, claim, sick, soul, and bpjs. The words “bpjs”, “life”, and “illness” are the three words that have the strongest relationship with insurance. The types of insurance that are often discussed are life insurance and sharia insurance, while the companies that are often discussed are prudential.

Keywords: Insurance, Twitter, Association rule, FP-Growth, data mining.

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat-Nya sehingga penulis dapat menyelesaikan laporan penelitian tugas akhir yang berjudul “Implementasi *Association Rule Mining* Menggunakan Data Twitter Mengenai Asuransi di Indonesia Menggunakan Algoritma *Frequent Pattern-Growth (FP-Growth)*”. Maksud dan tujuan penulisan laporan penelitian pada tugas akhir ini adalah sebagai salah satu syarat untuk memperoleh gelar Sarjana Ilmu Aktuaria pada program studi Sains Aktuaria Institut Teknologi Sepuluh Nopember Surabaya.

Penulis menyadari bahwa tugas akhir ini tidak akan dapat terselesaikan dengan baik tanpa bantuan dan kerjasama dari banyak pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak berikut:

1. Allah SWT yang telah memberikan rahmat, kemudahan dan kesehatan sehingga penulis dapat menyelesaikan skripsi ini.
2. Bapak Dr. Drs Soehardjoepri, M.Si selaku kepala departemen aktuaria yang selama ini selalu mendukung kegiatan positif di jurusan sehingga bisa sampai tahap terkahir ini.
3. Bapak R. Mohammad Atok, M.Si, Ph.D. sebagai pembimbing penulis yang selalu membimbing penulis sejak awal perkuliahan serta memberikan arahan dan nasehat selama proses penelitian skripsi ini. Ibu Pratnya Paramitha Oktaviana, M.Si sebagai dosen penguji pertama yang selalu memberikan masukan terhadap penulisan ini sampai tahap akhir.
4. Bapak Imam Safawi Ahmad, M.Si sebagai dosen penguji kedua yang selalu memberikan masukan terhadap penulisan ini sampai tahap akhirdan juga dosen wali saya selama kuliah di Aktuaria ITS yang selalu memantau perkembangan akademik saya.
5. Kepada dosen dan tendik aktuaria lainnya yang tidak bisa saya tuliskan satu persatu untuk ilmu yang sangat berharga.
6. Keluarga tercinta ayah saya Eko Suherwanto, ibu saya Komalasari, kakak saya Zhafirah Fauziyyah, adik saya Ali Zulhazmi, dan saudara saya Nur yang selalu memberikan dukungan moran dan juga materil dari awal sampai akhir penulisan ini.
7. Kepada sahabat terdekat saya dari SD, SMP, SMA, Kuliah, dan Vieri Donny yang sudah memberikan dukungan dan do'a agar penulis dapat menyelesaikan tugas akhir.
8. Dan yang tidak kalah pentingnya, saya ingin berterimakasih kepada diri saya sendiri karena sudah yakin untuk bisa menyelesaikan penulisan ini dan tidak menyerah.

Penulis menyadari bahwa dirinya memiliki keterbatasan dan kemampuan yang beragam, sehingga penulisan skripsi ini masih jauh dari sempurna dan membutuhkan kritik maupun saran untuk perbaikan. Penulis berharap para pembaca dapat mengambil manfaat dan terinspirasi dari skripsi ini.

Surabaya, 20 Juli 2022

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	iii
PERNYATAAN ORISINALITAS	v
STATEMENT OF ORIGINALITY	vi
ABSTRAK	vii
ABSTRACT	viii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
DAFTAR LAMPIRAN	xvi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Penelitian.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu.....	5
2.2 Asuransi.....	6
2.3 <i>Knowledge Discovery in Database</i>	7
2.4 <i>Text Processing</i>	7
2.5 <i>Data Mining</i>	8
2.6 <i>Frequent Pattern – Growth</i>	10
2.6.1 Pembuatan <i>FP-Tree</i>	10
2.6.2 Penerapan Algoritma FP-Growth.....	11
2.6.3 Implementasi Algoritma FP-Growth pada kalimat (tweet).....	12
2.7 <i>Association Rule</i>	15
2.8 <i>Lift Ratio</i>	17
2.9 <i>Word Cloud</i>	17
BAB III METODOLOGI PENELITIAN	19
3.1 Sumber Data.....	19
3.2 Variabel Penelitian.....	19
3.3 Tahapan Penelitian.....	19
3.3.1 <i>Preprocessing Data</i>	19
3.3.2 Penentuan Nilai <i>Support</i> Dan <i>Confidence</i>	21

3.3.3 Implementasi <i>FP-Growth</i> pada <i>RapidMiner</i> ver 9.10	21
3.3.4 Pencarian Aturan/ <i>rule</i> Terbaik	25
3.3.5 Interpretasi Hasil	25
3.4 Diagram Alir Penelitian	25
BAB IV HASIL DAN PEMBAHASAN	27
4.1 Karakteristik Data	27
4.1.1 Persebaran lokasi pengguna twitter yang berpendapat di Indonesia	27
4.1.2 Persebaran akun yang berpendapat mengenai asuransi di Indonesia	28
4.1.3 Word cloud kata asuransi	29
4.2 <i>Data Preprocessing</i>	30
4.2.1 <i>Data Selection</i>	30
4.2.2 <i>Cleaning Data</i>	31
4.2.3 <i>Stemming</i>	32
4.2.4 <i>Stopword Removal</i>	33
4.2.5 <i>Tokenizing</i>	33
4.2.6 <i>Data Transformation</i>	34
4.3 Menentukan Minimal <i>Support</i> Dan <i>Confidence</i>	34
4.4 Implementasi <i>FP-Growth</i>	36
4.5 Pencarian Aturan/ <i>rule</i> Terbaik	37
4.6 Intrepretasi/evaluasi Hasil Asosiasi	39
BAB V KESIMPULAN DAN SARAN	43
5.1 Kesimpulan	43
5.2 Saran	44
DAFTAR PUSTAKA	45
LAMPIRAN	47
BIODATA PENULIS	1

DAFTAR GAMBAR

Gambar 2.1 Tahapan <i>Knowledge Discovery in Database</i> (sumber: <i>sis.binus.ac.id</i>)	7
Gambar 2.2 Contoh <i>FP-Tree</i> dan Tabel <i>Header</i> (Sumber: Devina, 2017)	10
Gambar 2.3 Hasil <i>FP-Tree</i> Dari <i>Tweet</i> 1	13
Gambar 2.4 Hasil <i>FP-Tree</i> Dari <i>Tweet</i> 1 dan 2	14
Gambar 2.5 Hasil <i>FP-Tree</i> Dari <i>Tweet</i> 1, 2 dan 3	14
Gambar 2.6 Contoh <i>Word Cloud</i> (Sumber: https://zamharisyafii.medium.com/).....	17
Gambar 3.1 <i>Import Data</i>	21
Gambar 3.2 Halaman Proses <i>Rapidminer</i>	22
Gambar 3.3 Proses <i>FP-Growth</i> pada <i>RapidMiner</i> 9.10.....	22
Gambar 3.4 Mencantumkan Nilai Minimum <i>Support</i>	23
Gambar 3.5 Mencantumkan Nilai Minimum <i>Confidence</i>	23
Gambar 3.6 Hasil <i>Running FP-Growth</i> pada <i>RapidMiner</i> untuk 1-itemset	24
Gambar 3.7 Hasil <i>Running FP-Growth</i> pada <i>RapidMiner</i> untuk 2-itemset dan 3 itemset	24
Gambar 3.8 Diagram Alir Penelitian	24
Gambar 3.8 Diagram Alir Penelitian (Lanjutan)	24
Gambar 4.1 a). Identifikasi Lokasi Pengguna Twitter Yang Berpendapat Asuransi. b). Persebaran Lokasi yang Teridentifikasi	27
Gambar 4.2 <i>Word Cloud</i> Kata Asuransi	29
Gambar 4.3 Hasil Asosiasi Nilai <i>Support</i> 0,03 dan Nilai <i>Confidence</i> 0,96.....	38

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu.....	5
Tabel 2.2 <i>Conditional Pattern Base</i>	11
Tabel 2.3 <i>Conditional FP-Tree</i>	11
Tabel 2.4 <i>Frequent Itemset</i> untuk I_5	12
Tabel 2.5 Daftar Kata Secara <i>Descending Order</i>	12
Tabel 2.6 Contoh <i>Tweet</i>	13
Tabel 2.7 <i>Conditional Pattern Base</i> Pada Kalimat.....	14
Tabel 2.8 <i>Conditional FP-Tree</i> Pada Kalimat.....	15
Tabel 2.9 <i>Frequent Itemset</i> Untuk Kata "rumah sakit"	15
Tabel 2.10 Contoh Transaksi <i>Association Rule</i>	16
Tabel 3.1 Variabel Penelitian	19
Tabel 3.2 Struktur Data	19
Tabel 3.3 Contoh Proses <i>Stopword Removal</i>	20
Tabel 3.4 Hasil <i>Data Transformation</i>	21
Tabel 4.1 Proporsi Jumlah Penduduk Di Tiap Pulau Dengan <i>Tweets</i> Mengenai Asuransi.....	28
Tabel 4.2 Jumlah Akun <i>Verified</i> dan <i>Unverified</i> Yang Berpendapat Mengenai Asuransi di Indonesia.....	28
Tabel 4.3 Contoh Kalimat Mentah Dari Twitter	30
Tabel 4.4 Hasil Proses <i>Data Cleaning</i>	31
Tabel 4.5 Hasil proses <i>Data Stemming</i>	32
Tabel 4.6 Hasil Proses <i>Stopword Removal</i>	33
Tabel 4.7 Hasil proses <i>Data Transformation</i>	34
Tabel 4.8 Format Data Akhir Berdasarkan Data pada Tabel 4.7	34
Tabel 4.9 Nilai <i>Support</i> Tiap Kata.....	35
Tabel 4.10 Jumlah Kata Pada Nilai <i>Support</i> Tertentu	35
Tabel 4.11 Hasil FP-Growth Minimal <i>Support</i> 0,02 Dan <i>Confidence</i> 0,92	36
Tabel 4.12 Hasil Pencarian Nilai Minimal <i>Support</i> dan <i>Confidence</i> Terbaik.....	37
Tabel 4.13 Hasil Nilai <i>Support</i> 0,03 dan Nilai <i>Confidence</i> 0,96.....	38
Tabel 4.14 Interpretasi Hasil Asosiasi	39

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

Lampiran 1. Nilai <i>Confidence</i> Berdasarkan Nilai <i>Support</i> 0,02	47
Lampiran 2. Data Hasil Pencarian Nilai <i>Support</i> Dan Nilai <i>Confidence</i> Terbaik.....	47
Lampiran 3. Data Twitter	53
Lampiran 4. <i>Syntax</i> Pengolahan Data.....	54
Lampiran 5. List <i>Stopword</i>	57
Lampiran 6. List Kata Yang Digunakan Pada <i>FP-Growth</i>	61

(Halaman ini sengaja dikosongkan)

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

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. Subekti, dalam bukunya memberikan definisi mengenai asuransi yaitu, Asuransi atau pertanggungan sebagai suatu perjanjian yang termasuk dalam golongan perjanjian untung-untungan (*kansovereenkomst*) (Ajib, 2019).

Terminologi Asuransi (*Insurance*) barangkali bagi sebagian masyarakat bukanlah sesuatu yang asing, sebab keberadaan lembaga ini sudah merupakan bagian hidup sehari-hari dalam melakukan aktivitas. Namun sebaliknya bisa terjadi juga, bagi sebagian besar masyarakat lainnya. Asuransi bisa jadi sesuatu yang baru. Belum diketahui apa manfaatnya (Sembiring, 2014).

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 atau dikenal dengan Badan Penyelenggara Jaminan Sosial (BPJS) yang awal pada tahun 2015 sebesar 3,14 Triliun pada tahun 2019 menjadi 8,79 Triliun. Pada data *life insurance*, total aset pada tahun 2015 sebesar 378 Triliun menjadi 574,59 Triliun pada tahun 2019 dan pada data *non-life insurance*, total aset pada tahun 2015 sebesar 124,01 Triliun menjadi 173,36 Triliun pada tahun 2019. Data tersebut juga menunjukkan perkembangan industri perasuransian di Indonesia yang memiliki peran cukup signifikan dalam mendukung terjadinya proses pembangunan nasional sehingga perusahaan asuransi pun makin giat menunjukkan performa perusahaan yang mereka jalani semakin hari sehingga semakin banyak nasabah yang mulai menggunakan asuransi diperusahaannya guna bisa menutupi resiko mereka dalam kesehatan, jiwa, dan produk asuransi lainnya.

Risiko-risiko kesehatan yang dihadapi oleh masyarakat menyadarkan bahwa asuransi bisa menjadi sebuah payung atau perlindungan terhadap musibah yang bisa terjadi dan menimpa diri mereka kapan pun dan dimana pun dengan waktu yang tidak bisa kita prediksi. Apapun bisa terjadi walaupun mereka tidak menginginkan musibah menimpa dirinya. Hal ini menjadi salah satu penyebab meningkatnya jumlah pengguna asuransi di Indonesia dan juga sebuah keuntungan tersendiri bagi perusahaan asuransi untuk bisa melebarkan target pasar perusahaan mereka di Indonesia.

Perkembangan juga terjadi pada bagian teknologi dan informasi yang selalu bisa memberikan kemudahan bagi manusia atau penggunanya untuk memperoleh informasi dari mana saja. Twitter merupakan salah satu dari banyak sosial media yang digunakan masyarakat Indonesia untuk menyebarkan informasi yang sedang beredar kepada pengguna twitter lainnya. Informasi yang disebarkan oleh pengguna twitter ini biasa disebut sebagai *tweet* yang hanya memuat 280 karakter atau kata pada setiap pesan. Pengguna twitter (organisasi atau perorangan) menjadikan twitter sebagai tempat berdiskusi yang bisa membicarakan banyak hal dari apa yang sedang *tranding* saat itu, hal yang disukai, hal yang menurut mereka menarik untuk dibicarakan walaupun tidak penting atau biasa disebut dengan gosip, dan terkadang ada pula yang menceritakan keluh kesah tentang keseharian mereka di twitter.

Melihat perkembangan asuransi yang baik di Indonesia dan juga perkembangan pesat dari sisi teknologi peneliti ingin mencoba memaksimalkan peran teknologi informasi yakni media sosial twitter untuk mendapatkan penemuan pengetahuan atau KDD (*knowledge discovery in database*) mengenai apa yang berhubungan dengan asuransi di Indonesia berdasarkan

pendapat-pendapat masyarakat yang menggunakan twitter dengan teknik *data mining* dari data yang diperoleh dari *tweet* tiap pengguna twitter yang dikumpulkan di satu berkas dan kemudian diproses menggunakan *software RapidMiner*. Untuk proses pengolahan *data mining* digunakan metode *association rule* algoritma *FP-Growth* seperti pada jurnal SAINTIKOM ditulis oleh (Ikhwan et al., 2015) yang berjudul “Penerapan Data Mining dengan Algoritma *FP-Growth* untuk Mendukung Strategi Promosi Pendidikan (Studi Kasus Kampus STMIK Triguna Dharma)” menjelaskan pengimplementasian algoritma *FP-Growth* dapat dilakukan menggunakan *software RapidMiner*.

Banyak metode *data mining* yang sudah beredar dan diketahui, akan tetapi pada penelitian kali ini menggunakan fungsi asosiasi yaitu metode *Association Rule Mining* yang biasa digunakan untuk mengenali pola transaksi pembelian barang. Pada penelitian ini *association rule* digunakan untuk mengenali pola pada kata yang beredar di *twitter* mengenai asuransi. Hal ini dapat dilakukan karena cara kerja metode *association rule* yang dilakukan pada transaksi pembelian barang maupun pada *tweets* sama yaitu menemukan kombinasi *item* yang sering muncul. Seperti pada penelitian yang dilakukan oleh Ami Gates pada tahun 2019 dengan judul penelitian “*Association Rule Mining with Tweets: Thinking Outside the Basket*” yang berhasil menghasilkan beberapa kata yang berhubungan dengan kata *chocolate* yaitu *national, happy, dessert, foodporn, giftbox, yummy, sweets, dan weekend*. Kelebihan pada penelitian ini ada pada visualisasi yang dihasilkan sangat mudah dibaca oleh orang awam berbentuk *plot* dari program R studio. Sedangkan kekurangan yang terdapat pada penelitian ini adalah perbedaan menggunakan algoritma, pada penelitian Ami Gates menggunakan algoritma *apriori* sedangkan pada penelitian ini menggunakan *FP-Growth*.

Alasan pada penelitian ini menggunakan algoritma *frequent pattern-growth* karena proses pencarian banyaknya *itemset* pada algoritma ini membutuhkan waktu yang lebih cepat dibandingkan dengan algoritma yang umum dilakukan seperti algoritma *apriori* atau *tree projection*. Dibuktikan pada penelitian Erwin (2009) yang mendapat kesimpulan apabila menggunakan algoritma *FP-Growth* lebih cepat 3 jam 18 menit dibandingkan algoritma *apriori*, dan masih banyak lagi hasil penelitian yang menunjukkan bahwa *FP-Growth* lebih cepat daripada algoritma *apriori*.

Studi literatur terdahulu lainya yang menjadi referensi penulisan pada penelitian yang dilakukan oleh Eko Sedyono dan Adi Setiawan (2018) dengan judul “Implementasi Metode *Association Rule Mining* untuk Menganalisis data BPJS dengan Algoritma *FP-Growth*” dimana data yang mereka gunakan melauai *tweets* yang berhubungan dengan akun twitter BPJS Nasional untuk menggali pengetahuan baru dari tumpukan *tweet* yang sangat banyak. Namun kekurangan dari penelitian ini, mereka hanya mengambil data dari akun twitter nya BPJS saja, untuk masyarakat umum yang tidak menandai akun BPJS saat mengunggah *tweet* nya tidak ikut masuk ke data yang diteliti.

Selain itu juga terdapat studi yang dilakukan oleh (Aditiya et al., 2020) dengan judul penelitian “Prediksi Tingkat Ketersediaan Stock Sembako Menggunakan Algoritma *FP-Growth* dalam Meningkatkan Penjualan”. Penelitian ini memiliki tujuan bisa memberi rekomendasi kepada pemilik toko sebagai informasi untuk menyediakan stock sembako untuk meningkatkan penjualan toko dengan menggunakan metode *data mining sssociation rule* algoritma *FP-Growth* yang sama dengan penelitian ini. Namun kekurangan pada penelitian ini yaitu tujuan dari penggunaan teknik *Association Rule* masih dalam lingkup transaksi jual beli belum kedalam lingkup kalimat pada sosial media.

Berdasarkan beberapa penelitian yang telah dipaparkan diatas, pada tugas akhir ini akan dilakukan penelitian yang menggunakan teknik *association rule mining* menggunakan algoritma *FP-Growth* pada *tweet* yang mengandung kata “Asuransi” atau “asuransi” dengan bahasa Indonesia sebanyak 19.926 *tweet*. Atribut pada data penelitian ini yaitu *text*. Hasil yang

diharapkan pada penelitian ini adalah pengetahuan baru dari hubungan atau asosiasi pada kata asuransi berdasarkan data twitter yang digunakan melalui metode *association rule* dengan algoritma *FP-Growth*.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang dijelaskan sebelumnya, permasalahan yang akan dibahas pada penelitian ini sebagai berikut:

1. Bagaimana penerapan *association rule* dengan pendapat-pendapat yang ada di twitter mengenai asuransi di Indonesia?
2. Apa saja pengetahuan baru yang didapatkan dari metode *association rule* pada twitter mengenai asuransi di Indonesia?

1.3 Batasan Penelitian

Batasan masalah dari penelitian ini adalah sebagai berikut:

1. Data yang diambil dari *twitter* dari 01 Januari 2017 s/d 1 Mei 2022 berjumlah 19.926 *tweets*.
2. Kata yang digunakan dalam *mining* data pada twitter yaitu “Asuransi” dan “asuransi”.
3. Metode yang digunakan adalah *association rule mining* dengan algoritma *FP-Growth*.
4. Proses *association rule* menggunakan *software data science* yang bernama *RapidMiner*. dan proses *mining* data pada twitter menggunakan *jupyter notebook*.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang terbentuk maka tujuan penelitian ini adalah:

1. Menerapkan *association rule* dengan pendapat-pendapat yang ada di twitter mengenai asuransi di Indonesia.
2. Mendapatkan pengetahuan baru dari metode *association rule* pada twitter mengenai asuransi di Indonesia.

1.5 Manfaat Penelitian

Manfaat yang diharapkan pada penelitian ini adalah sebagai berikut:

1. Bagi bidang keilmuan, menambah wawasan mengenai ilmu praktik *data mining* yang pada zaman sekarang sangat banyak digunakan terutama pada teknik *association rule mining* menggunakan algoritma *frequent pattern-growth*.
2. Bagi bidang keilmuan, memberika
3. Bagi perusahaan asuransi di Indonesia, memberikan saran ataupun rekomendasi untuk pengembangan produk berdasarkan hasil penelitian *data mining* yang ditemukan pada penelitian ini.
4. Menambah kajian ilmiah mengenai bagaimana data twitter bisa bermanfaat untuk asuransi di Indonesia ataupun hal yang lainnya dengan menggunakan *association rule*.

(Halaman ini sengaja dikosongkan)

BAB II TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Pada sub bab ini di uraikan beberapa penelitian terdahulu yang menjadi referensi penulis untuk bahan kajian dalam penyusunan penelitian ataupun jurnal ini. Referensi yang dikumpulkan membahas mengenai metode *association rule* pada berbagai kasus.

Pada Tabel 2.1 merupakan jurnal dan penelitian lain yang dijadikan referensi dalam penyusunan penelitian ini:

Tabel 2.1 Penelitian Terdahulu

No	Penulis	Judul Jurnal	Tahun	Deskripsi	Metode
1.	Erwin	Analisis <i>Market Basket</i> dengan Algoritma <i>Apriori</i> dan <i>FP-Growth</i>	2009	Analisis penerapan <i>apriori</i> dan <i>fp-growth</i> dalam proses pencarian <i>frequent itemset</i> . Hasil yang didapatkan bahwa kompleksitas waktu <i>fp-growth</i> lebih cepat dibandingkan dengan <i>apriori</i> karena <i>fp-growth</i> menggunakan metode <i>fp-tree</i> sedangkan <i>apriori</i> menggunakan proses <i>candidate generation</i> yang memakan waktu lebih lama.	<i>Association rule</i> algoritma <i>apriori</i> dan <i>fp-frowth</i>
2.	Ali Ikhwan, Dicky Nofriansyah, dan Sriani	Penerapan <i>Data Mining</i> dengan Algoritma <i>FP-Growth</i> untuk Mendukung Strategi Promosi (Studi Kasus Kampus STMIK Triguna Dharma)	2015	Algoritma <i>FP-Growth</i> dapat diimplementasikan menggunakan <i>software RapidMiner</i> .	<i>Association rule</i> algoritma <i>fp-frowth</i> dengan <i>RapidMiner</i>
3.	Adi Setiawan, Eko Sedyono, Jemaictry Tamaela	Implementasi Metode <i>Association Rule</i> untuk Menganalisis Data Twitter tentang Badan Penyelenggara Jaminan Sosial dengan Algoritma <i>Frequent Pattern - Growth</i>	2018	Teknik <i>association rule</i> dengan algoritma <i>FP-Growth</i> dapat diimplementasikan untuk mengolah data teks dari twitter mengenai BPJS Nasional dari pendapat-pendapat yang ada di twitter.	<i>Association rule</i> algoritma <i>fp-frowth</i>
4.	Rahmad Aditya, Sarjon Defit, dan Gunandi Widi N	Prediksi Tingkat Ketersediaan Stock Sembako Menggunakan Algoritma <i>FP-Growth</i> dalam meningkatkan Penjualan	2020	Metode <i>association rule</i> dengan algoritma <i>FP-Growth</i> digunakan kepada pemilik toko sebagai informasi untuk menyediakan stock sembako berikutnya guna meningkatkan penjualan toko.	<i>Association rule</i> algoritma <i>fp-frowth</i>

Tabel 2.1 Penelitian Terdahulu (Lanjutan)

No	Penulis	Judul Jurnal	Tahun	Deskripsi	Metode
5.	Ami Gates	<i>Association Rule Mining with Tweets: Thinking Outside the Basket</i>	2019	Analisis data mining dari twitter mengenai kata “chocolate” menggunakan <i>association rule</i> apriori. Menghasilkan kata yang berkaitan dengan kata <i>chocolate</i> yaitu <i>national, happy, dessert, foodporn, giftbox, yummy, sweets, dan weekend</i> .	<i>Association rule</i> algoritma <i>apriori</i>

2.2 Asuransi

Istilah asuransi berasal dari bahasa Belanda, *assurantie*. Dalam hukum Belanda sering dipakai kata ini dengan kata *verzekering* yang diterjemahkan ke dalam bahasa Indonesia dengan kata “pertanggungan”. Dari kata *assurantie* ini muncul istilah *assuradeur* bagi penanggung, dan *geassureerde* bagi tertanggung, atau dengan istilah lain disebut penjamin dan terjamin. Dari istilah *verzekering* itu juga timbullah istilah *verzekeraar* bagi penanggung dan *verzekerde* bagi tertanggung (Prodjodikoro, 1996).

Pihak penanggung atau penjamin adalah perusahaan asuransi, sedangkan tertanggung atau yang dijamin adalah peserta asuransi. Jadi dalam suatu asuransi, terdapat perjanjian antara kedua belah pihak dimana pihak yang dijamin diwajibkan membayar uang premi dalam jumlah tertentu dalam suatu masa tertentu pula, kemudian pihak yang menjamin akan mengganti kerugian jika terjadi sesuatu pada diri si penjamin (Mannan, 1997).

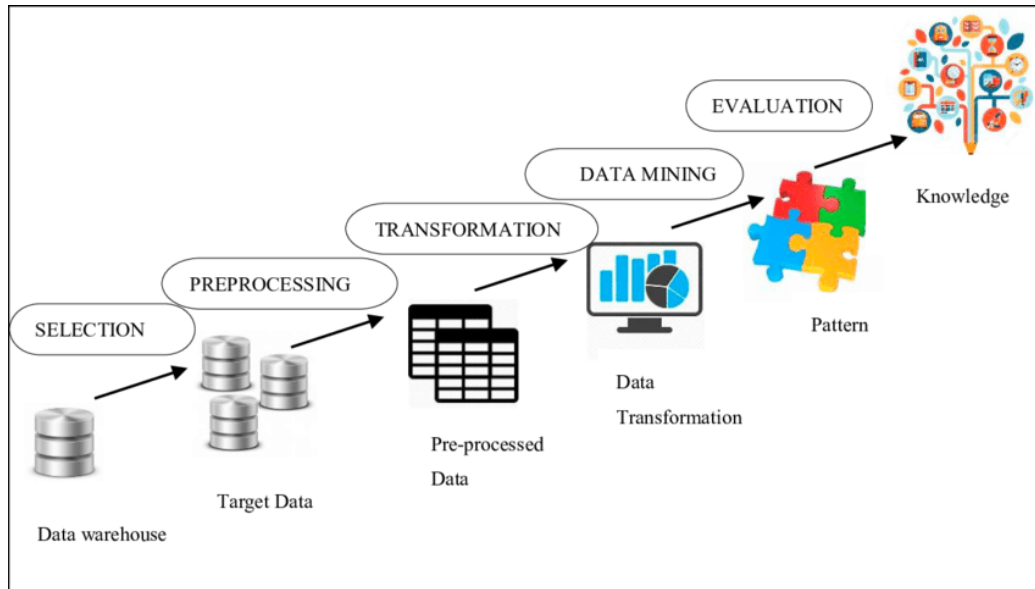
Menurut Undang-Undang No. 40 Tahun 2014 Pasal 1 Asuransi adalah perjanjian antara dua pihak, yaitu perusahaan asuransi dan pemegang polis, yang menjadi dasar bagi penerimaan premi oleh perusahaan asuransi sebagai imbalan untuk:

- a. Memberikan penggantian kepada tertanggung atau pemegang polis karena kerugian, kerusakan, biaya yang timbul, kehilangan keuntungan, atau tanggung jawab hukum kepada pihak ketiga yang mungkin diderita tertanggung atau pemegang polis karena terjadinya suatu peristiwa yang tidak pasti; atau
- b. Memberikan pembayaran yang didasarkan pada meninggalnya tertanggung atau pembayaran yang didasarkan pada hidupnya tertanggung dengan manfaat yang besarnya telah ditetapkan dan/atau didasarkan pada hasil pengelolaan dana. Berdasarkan pengertian diatas dapat ditarik kesimpulan bahwa terdapat dua jenis asuransi yaitu asuransi umum dan asuransi jiwa. Asuransi umum adalah perusahaan asuransi yang memberikan perlindungan pada harta benda atau kerusakan fisik, sedangkan asuransi jiwa adalah asuransi yang melindungi jiwa manusia atau memberikan manfaat pada pengelolaan dana (Susanto dkk, 2021).

Berdasarkan pengertian di atas suatu perjanjian asuransi minimal terdapat tiga unsur. Pertama, pihak yang sanggup menanggung atau menjamin bahwa pihak lain akan menadapat penggantian dari satu kerugian yang mungkin akan diderita sebagai akibat dari suatu peristiwa yang semula belum tentu akan terjadi. Kedua, pihak yang ditanggung diwajibkan membayar sejumlah uang kepada pihak yang ditanggung, Ketiga, apabila peristiwa yang dimaksud telah terjadi (Ridwan, 2016).

2.3 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 yang sesuai dengan Gambar 2.1 (Kusrini & Lutfhi, 2009).



Gambar 2.1 Tahapan *Knowledge Discovery in Database* (sumber: sis.binus.ac.id)

1. *Data Selection*. Pada tahap ini dilakukan pemilihan rangkaian data yang akan digunakan untuk mendapatkan data yang sesuai dengan tujuan penelitian yang akan dilakukan. Hasil pemilihan data ini kemudian disimpan dalam berkas terpisah untuk basis data operasional.
2. *Data Preprocessing*. *Data preprocessing* ini dilakukan pembuangan data yang tidak diperlukan, *noise* atau data pengganggu, duplikasi data, memperbaiki kesalahan data dan bisa juga memperkaya data dari eksternal agar data lebih relevan. Nama lain dari data *cleaning* bisa juga dikenal dengan *text processing* dimana suatu proses untuk menyeleksi data yang berjenis kalimat supaya lebih terstruktur dengan melalui beberapa tahap lainnya.
3. *Data Transformation*. Proses ini mempunyai tujuan untuk menggabungkan kedalam bentuk yang sesuai proses pengolahan.
4. *Data Mining*. Tahap ini memiliki arti pencarian pola atau informasi yang menarik pada data yang terpilih menggunakan berbagai macam teknik, metode, atau algoritma yang sesuai dengan tujuan penelitian.
5. *Evaluasi / Interpretation*. Proses menerjemahkan pola-pola yang dihasilkan dari tahap data *mining* dan juga mengevaluasi apakah pola yang dihasilkan sesuai dengan hipotesa.
6. *Knowledge*. Proses terakhir dalam tahapan KDD. Data yang sudah diproses kemudian divisualisasikan agar hasil yang didapatkan lebih mudah dipahami oleh pembaca.

2.4 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. Ini adalah Kumpulan kegiatan di mana Dokumen Teks diproses sebelumnya. Karena data teks sering berisi beberapa format

husus seperti format angka, format tanggal dan kata-kata yang paling umum yang tidak mungkin membantu *text mining* seperti preposisi, artikel, dan pro-kata benda dapat dihilangkan (Kannan, S. dkk,2015).

Dalam melakukan *text preprocessing* terdapat beberapa tahap yang dapat dilakukan. Tahapan-tahapan *text preprocessing* secara umum adalah *Tokenizing*, *Stopword Removal*, dan *Stemming*.

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*

Setelah proses *tokenizing* setiap kata menjadi berdiri sendiri / tidak terikat dengan kata yang lain. Akibat dari pemisahan kata tersebut, akan ada kata yang tidak memiliki arti yang relevan untuk menentukan ciri dari dokumen yang di *tokenizing* seperti “*ini, itu, adalah, dan, atau*” dan banyak lagi kata-kata sejenis. Kata-kata yang tidak memiliki arti yang relevan tersebut disebut *stopword*. Kumpulan dari *stop word* disebut *stop list* dan proses untuk menghapus *stop word* 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. Algoritma *porter stemmer* Bahasa Indonesia merupakan algoritma *stemmer* yang tidak menggunakan kamus kata dasar. Teknik algoritma *porter* Bahasa Indonesia dalam mencari kata dasar yaitu dengan melihat dan mengapus *Affixes* dari Bahasa Indonesia yang terdiri dari *Sufiks, prefix* dan *konfiks*. Sehingga algoritma *porter stemmer* memiliki kecepatan yang lebih dibandingkan algoritma *stemmer* yang menggunakan kamus kata dasar seperti yang dikembangkan oleh Nazief dan Adriani dari Universitas Indonesia pada tahun 1996 dan 2007.

Dalam dokumen Bahasa Indonesia proses *stemming* sangat diperlukan sebelum proses *text mining* karena Bahasa Indonesia memiliki *prefixes, suffixes, infixes* dan *confixes* yang membuat suatu kata dasar dapat berubah menjadi banyak bentuk dan akibatnya membuat pencarian kata dasar menjadi sulit. Berikut adalah arti dan contoh dari imbuhan dalam Bahasa Indonesia:

- a. *Sufiks* (Akhiran) adalah afiks yang ditambahkan pada bagian belakang kata dasar, misal “*-an, -kan,*” dan “*-i*”;
- b. *Prefiks* (Awalan) adalah imbuhan yang ditambahkan pada bagian awal sebuah kata dasar atau bentuk dasar; awalan: “*per-*” adalah yang paling produktif dalam bahasa Indonesia
- c. *Konfiks* (sifiks dan prefiks) afiks tunggal yang terjadi dari dua unsur yang terpisah (misal “*ke-...-an*” dalam kata “*kemerdekaan*”) (Ngrah dkk, 2017).

2.5 *Data Mining*

Istilah *data mining* memiliki beberapa padanan, seperti *knowledge discovery* ataupun *pattern recognition*. Kedua istilah tersebut sebenarnya memiliki ketepatannya 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. Bila dalam tulisan ini digunakan

istilah *data mining*, hal ini lebih didasarkan pada lebih populernya istilah tersebut dalam kegiatan penggalian pengetahuan data.

Banyak definisi bagi istilah ini dan belum ada yang dibakukan atau disepakati semua pihak. Namun demikian, istilah ini memiliki hakikat (*notion*) sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari kata atau informasi yang kita miliki. Kegiatan inilah yang menjadi garapan atau perhatian utama dari disiplin ilmu *data mining*.

Terdapat beberapa teknik *data mining* yang dibagi sebagai berikut (Kusrini & Lutfhi, 2009):

1. Deskripsi
Deskripsi dari pola yang cenderung sering memberikan penjelasan untuk pola yang bersifat tentatif. Contoh apabila peneliti ingin menganalisis secara sederhana untuk mencari pola yang menggambarkan kecenderungan pada data.
2. Estimasi
Estimasi mempunyai kemiripan dengan teknik klasifikasi, kecuali pada teknik estimasi variabelnya lebih kearah numerik daripada kategori. Pembentukan model dilengkapi dengan nilai dari variabel target sebagai prediksi. Selanjutnya, peninjauan estimasi nilai dari variabel yang dibuat berdasarkan variabel prediksi.
3. Prediksi
Teknik ini memiliki kesamaan pada estimasi ataupun klasifikasi karena beberapa metode dan teknik yang digunakan sama dengan yang ada di teknik klasifikasi dan estimasi. Perbedaannya terletak pada prediksi termasuk dalam nilai dari hasil dimasa mendatang.
4. Klasifikasi
Proses ini digunakan untuk mendapatkan model atau fungsi yang menjelaskan konsep data dengan tujuan khusus untuk dapat memperkirakan kelas dari suatu objek yang masih tidak diketahui labelnya. Contoh dari model teknik klasifikasi berupa *decision tree*, formula matematis atau *neural network*.
5. Pengelompokan (*clustering*)
Clustering merupakan pengelompokkan data dengan membentuk kelas objek-objek yang memiliki kemiripan.
6. Asosiasi
Teknik ini memiliki tujuan untuk menemukan aturan asosiatif antara suatu kombinasi item. Contoh pada pembelian di suatu *supermarket* adalah bisa diketahui berapa besar kemungkinan pelanggan membeli roti bersamaan dengan selai coklat. Dari contoh ini menjadi pengetahuan baru bagi pihak *supermarket* untuk mengatur penempatan barangnya dalam ranga pemasaran barang agar banyak dibeli dan membuat kampanye diskon dengan menggabungkan kombinasi barang yang didapatkan. Parameter yang digunakan pada teknik asosiasi ini yaitu *support* dan *confidence*.

Mengacu kepada Berry dan Browne (2006), keenam fungsi *data mining* tersebut dapat dipilah menjadi:

1. Fungsi minor atau fungsi tambahan, yang meliputi ketiga fungsi yang pertama yaitu fungsi deskripsi, estimasi, dan prediksi;
2. Fungsi mayor ataupun fungsi utama, yang meliputi ketiga fungsi berikutnya, yaitu klasifikasi, pengelompokan, dan asosiasi (Susanto & Suryani, 2010)

2.6 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 & Sensuse, 2012).

Sebelum lebih lanjut kepada penjelasan *FP-Growth*, ada beberapa definisi yang akan digunakan kedepannya pada sub bab ini. Beberapa diantaranya ada **item** untuk himpunan kata seperti $I = \{a_1, a_2 \dots, a_n\}$ dan **itemset** adalah subset dari I. Sebuah *item* dapat muncul paling banyak satu kali dalam sebuah transaksi, tapi dapat muncul lagi dalam jumlah yang berbeda. Pada suatu transaksi memiliki **support count** atau jumlah transaksi dalam basis data yang digunakan untuk memuat suatu item. Misalkan, itemset $X = \{a\}$ yang muncul 4 kali transaksi maka *support count*-nya adalah 4. Kemudian ada batas **minimum support** yang digunakan untuk membatasi nilai support untuk digunakan supaya banyak data yang tidak terlalu banyak transaksi digunakan dan bisa menjadikan data menjadi bervariasi.

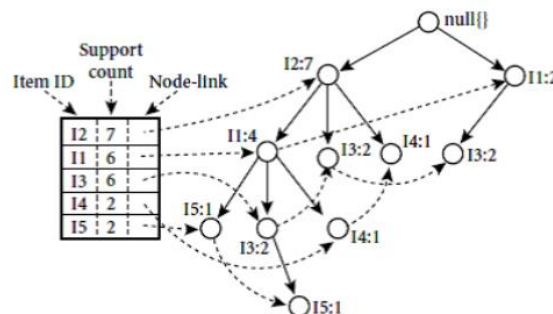
Pada penentuan *frequent itemset* terdapat 2 tahap proses yang dilakukan yaitu: pembuatan *FP-tree* dan penerapan algoritma *FP-Growth* untuk menemukan *frequent itemset*. Struktur data yang digunakan untuk mencari *frequent itemset* dengan algoritma *FP-Growth* adalah perluasan dari penggunaan sebuah pohon *prefix*, yang biasa disebut adalah *FP-tree*. Dengan menggunakan *FP-Tree*, algoritma *FP-Growth* dapat langsung mengekstrak *frequent itemset* dari *FP-tree* yang telah terbentuk dengan menggunakan prinsip *divide and conquer*.

2.6.1 Pembuatan FP-Tree

FP-tree merupakan struktur penyimpanan data yang dimampatkan. *FP-tree* dibangun dengan memetakan setiap data transaksi ke dalam setiap lintasan tertentu dalam *FP-tree*. Karena dalam setiap transaksi yang dipetakan, mungkin ada transaksi yang memiliki item yang sama, maka lintasannya memungkinkan untuk saling menimpa. Semakin banyak data transaksi yang memiliki item yang sama, maka proses pemampatan dengan struktur data *FP-tree* semakin efektif.

Adapun *FP-tree* adalah sebuah pohon dengan definisi sebagai berikut:

- FP-tree* dibentuk oleh sebuah akar yang diberi label null, sekumpulan *sub tree* yang beranggotakan item-item tertentu, dan sebuah tabel *frequent header*.
- Setiap simpul dalam *FP-tree* mengandung tiga informasi penting, yaitu label item, menginformasikan jenis item yang direpresentasikan simpul tersebut, support count, merepresentasikan jumlah lintasan transaksi yang melalui simpul tersebut, dan pointer penghubung yang menghubungkan simpul-simpul dengan label item sama antar-lintasan, ditandai dengan garis panah putus-putus (Gunadi & Sensuse, 2012). Berikut contoh dari *FP-Tree*.



Gambar 2.2 Contoh *FP-Tree* dan Tabel *Header* (Sumber: Devina, 2017)

Berdasarkan Gambar 2.2 dijelaskan bahwa Item ID menjelaskan tentang kode dari setiap item yang terdapat pada dataset yang digunakan. *Support count* adalah seberapa banyak item yang ada pada transaksi yang diurutkan secara *descending order* atau dari nilai terbesar ke nilai terkecil. *Node-link* berfungsi untuk menghubungkan titik selanjutnya pada *FP-Tree* yang membawa nama item yang sama ditandai dengan garis panah putus-putus (Qomariyah, 2017).

2.6.2 Penerapan Algoritma FP-Growth

Setelah tahap pembuatan *FP-tree* dari sekumpulan data transaksi, akan diterapkan algoritma *FP-Growth* untuk mencari *frequent itemset*. Algoritma *FP-Growth* dibagi menjadi tiga langkah utama, yaitu:

a. Tahap Pembangkitan *Conditional Pattern Base*

Conditional Pattern Base merupakan subdatabase yang berisi *prefix path* (lintasan prefix) dan *suffix pattern* (pola akhiran). Pembangkitan *conditional pattern base* didapatkan melalui *FP-tree* yang telah dibangun sebelumnya.

Contoh: sebuah itemset yang sudah diurutkan berdasarkan banyak muncul itemsetnya secara *descending* seperti pada Gambar 2.2 untuk lintasan $\{I_2, I_1, I_3, I_5\}$ maka I_2 adalah *prefix path* dan (I_1, I_3, I_5) adalah *suffix pattern*. Berikut pada Tabel 2.2 merupakan hasil pembangkitan *conditional pattern base* berdasarkan contoh *FP-Tree* yang ada pada Gambar 2.2.

Item	<i>Conditional Pattern Base</i>
I_3	$\{I_1 : 2\}, \{I_2 : 2\}, \{I_1, I_2 : 2\}$
I_4	$\{I_1, I_2 : 1\}, \{I_2 : 1\}$
I_5	$\{I_1, I_2 : 1\}, \{I_3, I_1, I_2 : 1\}$

Item yang terdapat pada Tabel 2.2 adalah **selain** item yang terdapat pada Gambar 2.2 yang mempunyai *prefix* adalah null atau *root* pada *FP-Tree* yang terbentuk. Seperti contoh pada item I_3 mempunyai *suffix pattern* $\{I_1, I_2\}, \{I_2\}$ dan $\{I_1\}$. Kemudian jumlah pada setiap *suffix pattern* nya mengikuti jumlah yang tertera pada item I_3 dari *FP-Tree* yang terbentuk yaitu 2 maka terbentuk lah *conditional pattern base* untuk item I_3 adalah $\{I_1, I_2 : 2\}, \{I_2 : 2\}$ dan $\{I_1 : 2\}$ begitu seterusnya untuk item yang lainnya.

b. Tahap Pembangkitan *Conditional FP-tree*

Pada tahap ini setelah *conditional pattern base* sudah terbentuk kemudian dilakukan penggabungan item yang sama pada *conditional pattern base* dan dijumlahkan. Setelah digabungkan maka dilakukan pemangkasan berdasarkan nilai *minimal support* yang sudah ditentukan sehingga item yang mempunyai nilai dibawah nilai *minimal support* tidak masuk kedalam *conditional FP-Tree*. Tabel 2.3 menggambarkan contoh untuk *conditional FP-Tree* dari Tabel 2.2 dengan asumsi *minimal support* adalah 2.

Item	<i>Conditional Pattern Base</i>	<i>Conditional FP-Tree</i>
I_3	$\{I_1 : 2\}, \{I_2 : 2\}, \{I_1, I_2 : 2\}$	$\{I_1: 4, I_2: 4\}$
I_4	$\{I_1, I_2 : 1\}, \{I_2 : 1\}$	$\{I_1: 1, I_2: 2\}$
I_5	$\{I_1, I_2 : 1\}, \{I_3, I_1, I_2 : 1\}$	$\{I_1: 2, I_2: 2\}$

Pada Tabel 2.3 hasil dari *conditional FP-Tree* pada item I_5 terdapat I_1 berjumlah 2 dan I_2 berjumlah 2 sedangkan I_3 bernilai 1 tidak masuk kedalam *conditional FP-Tree* karena tidak memenuhi minimal nilai *support* yaitu 2.

c. Tahap Pencarian *frequent itemset* (k-itemset)

Apabila *conditional FP-tree* merupakan lintasan tunggal (*single path*), maka didapatkan *frequent itemset* dengan melakukan kombinasi item untuk setiap *conditional FP-tree*. Jika bukan lintasan tunggal, maka dilakukan pembangkitan *FP-growth* secara rekursif (Aditiya dkk, 2020). Tabel 2.4 merupakan hasil *frequent itemset* untuk item I_5 pada *conditional FP-Tree* yang ada di Tabel 2.3.

Tabel 2.4 *Frequent Itemset* untuk I_5

k	<i>Frequent Itemset</i>
1	I_5
2	I_5I_1, I_5I_2
3	$I_5I_1I_2$

2.6.3 Implementasi Algoritma *FP-Growth* pada kalimat (*tweet*)

Pada kali ini contoh yang digunakan dalam algoritma *FP-Growth* adalah dengan menggunakan kalimat dari twitter atau disebut *tweet* melalui penelitian Jemaictry Tameli, Eko Sedyono, dan Adi Setiawan pada tahun 2018.

Data twitter yang diperoleh kemudian diproses dengan cara dipecah menjadi kumpulan kata dan diambil kata-kata yang memiliki makna dan berkaitan dengan topik penelitian. Contoh topik yang digunakan kali ini adalah BPJS yang berjumlah 44.791 *tweets*. 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 dimulai dari yang tertinggi ke terendah atau *descending order* yang ditunjukkan pada Tabel 2.5 (Tamaela et al., 2018).

Tabel 2.5 Daftar Kata Secara *Descending Order*

Peringkat	Kata	Frekuensi
1	bpjs	41.328
2	kesehatan	13.152
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

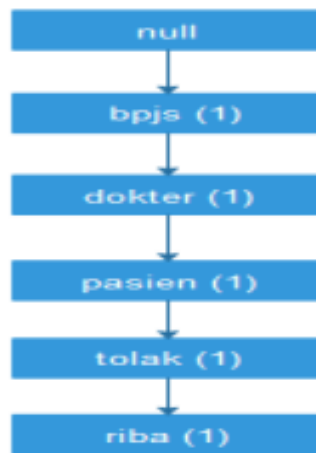
a. Pembuatan *FP-Tree* Dengan Contoh Kalimat

FP-Tree dibangun dengan cara membaca tiap *tweet*, kemudian membaca kata-kata di dalam *tweet* tersebut dalam urutan peringkat kemunculan (lihat Tabel 2.5). Jika terdapat kata yang ada di dalam daftar Tabel 2.5, maka kata tersebut ditambahkan ke dalam *tree*. *Tree* dimulai dengan simbol “*null*” sebagai permulaan (*root*). Contoh *tweet* yang akan digunakan untuk proses pengimplementasian algoritma *FP-Growth* dari pembentukan *FP-Tree* sampai dengan mendapatkan *frequent itemset* pada kalimat hanya menggunakan 3 *tweet* saja yang dapat dilihat pada Tabel 2.6 (Tamaela et al., 2018).

Tabel 2.6 Contoh *Tweet*

No	Isi <i>Tweet</i>
1	ada juga sih dokter yg tolak pasien bpjs, bukan karena riba sih, tapi karena nominal klaim bpjsnya kecil
2	Ah iya deket 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

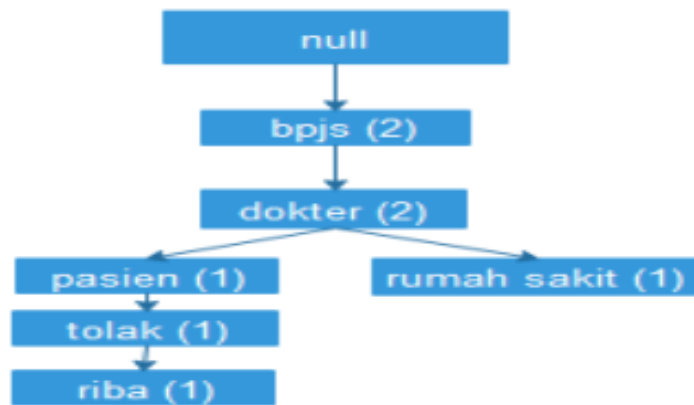
Berdasarkan *tweet* pada Tabel 2.6, sebelum dapat dibentuk aturan asosiasi, *tweet* tersebut diuraikan kata-kata di dalamnya, kemudian diurutkan sesuai dengan peringkat kemunculannya (lihat Tabel 2.5). Hasil pengurutan kata pada *tweet* 1: bpjs, dokter, pasien, tolak, riba, *tweet* 2: bpjs, dokter, sakit 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 2.3 untuk *tweet* 1, Gambar 2.4 untuk *tweet* 1 dan 2, dan Gambar 2.5 untuk *tweet* 1,2 dan 3.



Gambar 2.3 Hasil *FP-Tree* dari *Tweet* 1

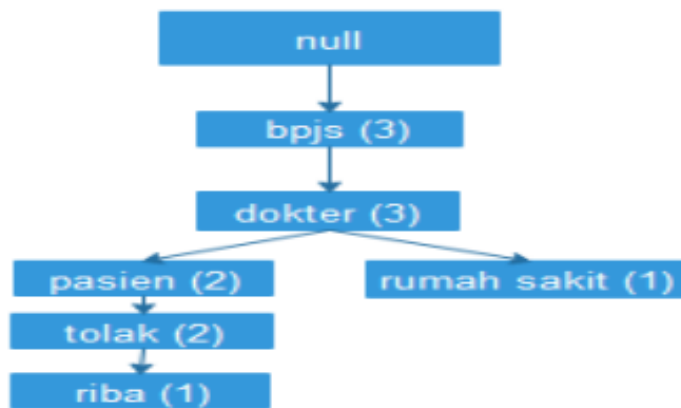
Tiap kata pada *tree* di Gambar 2.3 memiliki nilai masing-masing satu. Proses pembentukan dimulai dari *null* kemudian diikuti kata yang memiliki frekuensi kemunculan terbanyak sesuai dengan Tabel 2.5 yaitu “bpjs” dilanjutkan dengan “dokter”, “pasien”, “tolak”, dan yang terkecil frekuensinya “riba”.

Pada Gambar 2.4 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”.



Gambar2.4 Hasil *FP-Tree* dari *tweet* 1 dan 2

Pada Gambar 2.5 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 2.5 Hasil *FP-Tree* dari *tweet* 1, 2 dan 3

Setelah kalimat pada Tabel 2.6 terbentuk *FP-Tree*, kemudian dilanjutkan penerapan algoritma *FP-Growth* untuk mencari *frequent itemset* dengan melakuakn tiga langkah utama, yaitu: tahap pembangkitan *conditional pattern base*, tahap pembangkitan *conditional FP-Tree*, dan pencarian *frequent itemset*.

b. Tahap Pembangkitan *Conditional Pattern Base* Dengan Contoh Kalimat

Tahap selanjutnya yaitu pembangkitan *conditional pattern base* dari ketiga kalimat tersebut yang mendapatkan hasil seperti pada Tabel 2.7.

Tabel 2.7 *Conditional Pattern Base* Pada Kalimat

Item	Conditional Pattern Base
riba	{tolak, pasien, dokter, bpjs :1}
rumah sakit	{dokter, bpjs :1}

Pada Tabel 2.7 yang menjadi item adalah kata yang tidak mempunyai *prefix null* dan yang menjadi kata terakhir pada batang pohon atau *tree*. Riba dan rumah sakit menjadi item karena menunjukkan kata diakhir *tree* tdan tidak mempunya *prefix null*.

c. Tahap Pembangkitan *Conditional FP-Tree* Dengan Contoh Kalimat

Dalam pengimplementasian algoritma *FP-Growth*, langkah selanjutnya sebelum mendapatkan *frequent itemset* adalah membangkitkan *conditional FP-Tree*. Dari hasil *conditional pattern base* pada Tabel 2.7 dilakukan penggabungan antar kata menggunakan minimal *support* sebesar 1. Nilai *support* yang digunakan kali ini disesuaikan dengan jumlah kata dihasilkan. Karena pada contoh kali ini memiliki frekuensi kata yang sedikit maka digunakan minimal *support* yang kecil juga. Sehingga menghasilkan *conditional FP-Tree* yang dijelaskan pada Tabel 2.8.

Tabel 2.8 *Conditional FP-Tree* Pada Kalimat

Item	<i>Conditional Pattern Base</i>	<i>Conditional FP-Tree</i>
riba	{tolak, pasien, dokter, bpjs :1}	{bpjs: 1}, {dokter: 1}, {pasien: 1}, {tolak: 1}
rumah sakit	{dokter, bpjs :1}	{bpjs: 1}, {dokter: 1}

Pada Tabel 2.8 tidak terdapat kata yang sama sehingga hanya dipisahkan per kata yang ada pada *conditional patter base*. Apabila ingin melihat contoh penggabungan kata yang sama dapat melihat contoh pada Tabel 2.3.

d. Tahap Pencarian *Frequent Itemset* Dengan Contoh Kalimat

Tahap terakhir pada proses algoritma *FP-Growth* yaitu mendapatkan *frequent itemset* dari *conditional FP-Tree* yang terbentuk seperti pada Tabel 2.8. Pada Tabel 2.9 ditunjukkan hasil *frequent itemset* untuk kata “rumah sakit” sebagai berikut.

Tabel 2.9 *Frequent Itemset* Untuk Kata "rumah sakit"

k	<i>Frequent Itemset</i>
1	{rumah sakit}
2	{rumah sakit, bpjs}, {rumah sakit, dokter}
3	{rumah sakit, bpjs, dokter}

Pada Tabel 2.9 kata “rumah sakit” memiliki 3-itemset dengan kata yang berhubungan dengan “rumah sakit” yaitu “bpjs” dan “dokter”.

2.7 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. Penerapan data mining dengan aturan asosiasi bertujuan menemukan informasi item-item yang saling berhubungan dalam bentuk aturan/*rule*. Aturan asosiasi adalah teknik data mining untuk menemukan aturan asosiasi antar suatu kombinasi item (Ardani & Fitriana, 2016).

Tugas dari *association* adalah mencari aturan yang tidak mengcover untuk mengukur hubungan antara dua atau lebih atribut (Gunadi & Sensuse, 2012). *Interestingness measure* yang dapat digunakan dalam data mining adalah:

a. *Support* / Penunjang

adalah suatu ukuran yang menunjukkan seberapa besar tingkat dominasi suatu item atau itemset dari keseluruhan transaksi.

b. *Confidence* / Kepastian

adalah suatu ukuran yang menunjukkan hubungan antar dua item secara conditional (berdasarkan suatu kondisi tertentu).

Metodologi dasar analisis asosiasi terbagi menjadi dua tahap:

1. Analisis Pola Frekuensi Tinggi (*Support*)

Tahap ini mencari item yang memenuhi syarat minimum dari nilai *support* dalam database. Nilai *support* sebuah item diperoleh dengan rumus berikut: (Pandiangan, 2019).

$$Support(A) = \frac{\text{jumlah transaksi mengandung } A}{\text{Total Transaksi}} \quad (2.2)$$

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

$$Support(A \cap B) = \frac{\sum \text{Transaksi yang Mengandung } A \text{ dan } B}{\sum \text{Transaksi}} \quad (2.3)$$

2. Pembentukan Aturan Asosiasi

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat minimum untuk *confidence* dengan menghitung *confidence* aturan asosiatif. Pada contoh Tabel 2.10, *itemset* adalah barang yang dibeli pada toko dan transaksi adalah kode transaksinya. Dapat dilihat bahwa pada transaksi 1 yang dibeli adalah barang A, B, dan E. Transaksi 2 yang dibeli adalah barang B dan D, dan demikian seterusnya. *Association analysis itemset* adalah kumpulan nol atau lebih item yang akan melakukan analisis hubungan dengan aturan tertentu disebut sebagai *Association Rule*. Berikut contoh tabel transaksi *Association Rule* pada Tabel 2.10.

Tabel 2.10 Contoh Transaksi *Association Rule*

Transaksi	Itemset
1	A,B,E
2	B,D
3	B,C
4	A,B,D
5.	A,C
6	B,C
7	A,C
8	A,B,C,E
9	A,B,C

Contoh pada data transaksi diatas, apabila pembeli membeli barang A (*Indiquent*) maka akan membeli barang B (*consequent/kata kunci*). Aturan ini dapat ditulis $A \Rightarrow B$. Rumus untuk menghitung nilai *confidence* dari dua item dapat dilihat pada persamaan 2.4.

$$Confidence(A \Rightarrow B) = P(B|A) = \frac{Support(A \cap B)}{Support(A)} \quad (2.4)$$

Dengan rumus tersebut maka nilai *support* yang didapat dari hubungan ($A \Rightarrow B$) adalah $4/9 = 0,44$. Sedangkan nilai *confidence* yang diperoleh adalah ($A \Rightarrow B$) $0,44/0,6 = 0,66$. Nilai 4 pada *support* diperoleh dari *itemset* A dan B berjumlah 4 yaitu pada transaksi 1, 2, 8 dan 9 dengan total transaksi pada Tabel 2.10 adalah 9. Sedangkan nilai 0,44 pada *confidence* diperoleh dari nilai *support* yang sudah diperoleh sebelumnya dan nilai 0,6 didapatkan dari nilai *support* item A pada transaksi pada Tabel 2.10. Perolehan yang didapatkan selanjutnya dapat digunakan untuk menentukan persentasi yang diinginkan untuk menentukan kepastian asosiasinya. Aturan yang terbaik apabila nilai *support* ataupun *confidence* nya tinggi (Firmananda, 2020).

2.8 Lift Ratio

Cara lain untuk melihat kuat tidaknya aturan asosiasi yang terbentuk yaitu dengan menghitung nilai *lift ratio*. Nilai *lift ratio* biasanya digunakan sebagai penentu apakah aturan asosiasi yang terbentuk valid atau tidak. Untuk mendapatkan nilai *lift ratio* digunakan rumus sebagai berikut (Devina, 2017).

$$\text{Lift Ratio} = \frac{\text{Confidence}(A \cap B)}{\text{Benchmark Confidence}(A,B)} \quad (2.5)$$

Sedangkan untuk menghitung expected confidence digunakan rumus sebagai berikut.

$$\text{Benchmark Confidence}(A,B) = \frac{N_c}{N} \quad (2.6)$$

Keterangan:

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

Aturan yang ada pada nilai *lift ratio* yaitu apabila nilai nya 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 berkolerasi negatif atau artinya kemunculan salah satu item mempengaruhi hal yang sebaliknya pada kemuncula item lainnya. Contoh dari korelasi negatif adalah jika penjualan item A dan item B memiliki rasio 0,96 berarti jika penjualan item A naik maka mempengaruhi penjualan item B menjadi menurun. Sedangkan jika nilai *lift ratio* sama dengan 1 maka diartikan bahwa item A dan B *independent* atau tidak ada korelasi diantara keduanya (Fitriyanto, 2017).

2.9 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. (Castella & Sutton, 2014)



Gambar 2.6 Contoh Word Cloud (Sumber: <https://zamharisyafii.medium.com/>)

(Halaman ini sengaja dikosongkan)

BAB III METODOLOGI PENELITIAN

3.1 Sumber 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 berjumlah 19.926 *tweets* atau data.

3.2 Variabel Penelitian

Pada penelitian ini menggunakan atribut pada data yaitu kalimat *tweet*.

Tabel 3.1 Variabel Penelitian

Variabel	Keterangan	Contoh
X	Kalimat (<i>tweet</i>)	perawatan gigi mahal banget untung asuransi kantor coba
Y	Kata yang terdapat pada kalimat	asuransi, barang, beli

Definisi variabel X adalah rangkaian kata sebagai bahan utama yang akan digunakan pada penelitian kali ini, berisi pikiran-pikiran yang dituangkan dalam kalimat ke *twitter*. Sedangkan variabel Y adalah satuan kata dari variabel X yang akan digunakan untuk penelitian ini. Struktur data pada penelitian ini tersusun menjadi kata yang dapat dilihat pada Tabel 3.2.

Tabel 3.2 Struktur Data

	Y_1	Y_2	Y_3	...	Y_j
X_1	$Z_{1(1)}$	$Z_{1(2)}$	$Z_{1(3)}$	⋮	$Z_{1(j)}$
X_2	$Z_{2(1)}$	$Z_{2(2)}$	$Z_{2(3)}$	⋮	$Z_{2(j)}$
X_3	$Z_{3(1)}$	$Z_{3(2)}$	$Z_{3(3)}$	⋮	$Z_{3(j)}$
⋮	⋮	⋮	⋮	⋮	⋮
X_n	$Z_{n(1)}$	$Z_{n(2)}$	$Z_{n(3)}$	⋮	$Z_{i(j)}$

Keterangan:

- n = Jumlah Kalimat (*tweet*)
- j = Jumlah kata pada kalimat ke- i , $i = 1, 2, \dots, n$
- Z_{ij} = Nilai data yang bersifat binomial

3.3 Tahapan Penelitian

Pada penelitian ini disusun dalam berbagai tahapan sebagai berikut:

3.3.1 Preprocessing Data

Setelah data dari *twitter* sudah terkumpul, dilakukan pengolahan data *mining* dengan mengikuti beberapa tahapan dalam *Knowledge Discovery in Database* (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 beberapa tahapan sebagai berikut:

a. *Data Selection*

Data yang tersusun pada tabel di program *jupyter notebook* terdapat beberapa kolom yang tidak digunakan seperti *text_id* dan *location*, kemudian dihapus sehingga tersisa hanya kolom *text* atau kalimat.

b. *Cleaning data*

Merapihkan data pada kolom *text* sehingga menjadi kalimat yang layak untuk diteliti dengan menghapus data yang mengandung *link*, *retweet*, *username*, baris baru, angka, *hashtag*, tanda baca, spasi yang berlebih, dan mengubah semua huruf pada atribut *text* menjadi huruf kecil atau yang umum disebut dengan *case folding* agar tidak terjadi salah baca oleh program karena perbedaan pada huruf kecil dan huruf kapital.

c. *Stemming*

Pada proses ini, data yang sudah bersih dari *link*, *retweet*, *username*, baris baru, angka, *hashtag*, tanda baca, spasi yang berlebih, dan sudah huruf kecil semua di ubah menjadi kata dasarnya dengan meng-*install library Sastrawi* yang tersedia di program *jupyter notebook*. Misalnya kata “pengulangan” menjadi kata “ulang” atau contoh lain pada kata “menggunakan” menjadi “guna”. Selain itu tahap ini digunakan agar tidak membuat banyak variasi kata yang nanti mengakibatkan hasil akhir yang kurang efektif apabila ada beberapa variasi dalam satu arti kata.

d. *Stopword Removal*

Stopword adalah kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Pada program *jupyter notebook* sudah terdapat kumpulan kata yang termasuk pada *stopword* dengan meng-*install library nltk*. Adapun contoh *stopword* dalam bahasa Indonesia adalah “yang”, “dan”, “di”, “dari”, dll. Makna di balik penggunaan *stopword* yaitu dengan menghapus kata-kata yang memiliki informasi rendah dari sebuah teks, kita dapat fokus pada kata-kata penting sebagai gantinya (Nugroho, 2019) pada Tabel 3.3.

Tabel 3.3 Contoh Proses *Stopword Removal*

<i>Text</i>	<i>Stopword Removal</i>
Butuh waktu cepat untuk kirim barang ke luar pulau jawa	butuh waktu cepat kirim barang luar Pulau Jawa
Barang yang sudah pakai lebih baik kirim ke tempat awal beli barang	barang pakai baik kirim tempat beli barang
rekomendasi banget ya buat kamu ikut asuransi unit link dari axa mandiri pokoknya ngga usah khawatir lagi bagaimana masa depan kamu semua sudah cover dengan baik lah ya	rekomendasi asuransi unit link axa mandiri pokok khawatir cover
mungkin ada yang mau tau asuransi kesehatan dari produk Allianz bisa DM yaa thankyou hehe	asuransi sehat produk allianz dm thankyou hehe

e. *Tokenizing*

Tahap selanjutnya yaitu dengan memisahkan kalimat menjadi kata-kata agar memudahkan proses analisis. Proses ini digunakan *library NLTK* pada program *jupyter notebook* yang biasa digunakan untuk memudahkan saat proses teks seperti melakukan *classification*, *stemming*, *tagging*, dsb.

f. *Data Transformasion*

Untuk bisa dilanjutkan ke tahap *FP-Growth* melalui *tool RapidMiner*, diperlukan format data yang berbentuk binomial yaitu 0 dan 1. Oleh karena itu data di ubah atau di transformasi menjadi 0 untuk nilai yang menunjukkan tidak ada kata dalam kalimat tersebut sedangkan nilai 1 menunjukkan terdapat kata pada kalimat tersebut. Berikut pada Tabel 3.4 contoh data setelah melewati proses data transformation menggunakan library *counvectorizer*.

Tabel 3.4 Hasil *Data Tranformation*

	axa	allianz	beli	barang	cepat	cover	dm	hehe	...	waktu
1.	0	0	0	1	1	0	0	0	...	1
2.	0	0	1	2	0	0	0	0	...	0
3.	1	0	0	0	0	1	0	0	...	0
4.	0	1	0	0	0	0	1	1	...	0

Seperti pada Tabel 3.4 terdapat data yang mempunyai nilai lebih dari 1 seperti pada kata “barang” di kalimat nomer 3 sedangkan format yang digunakan hanya 0 untuk tidak ada kata pada data dan 1 untuk ada kata pada data maka dari itu di transformasi lagi menggunakan *jupyter notebook* atau bisa juga dengan bantuan *Microsoft Excel* agar data berisi binomial.

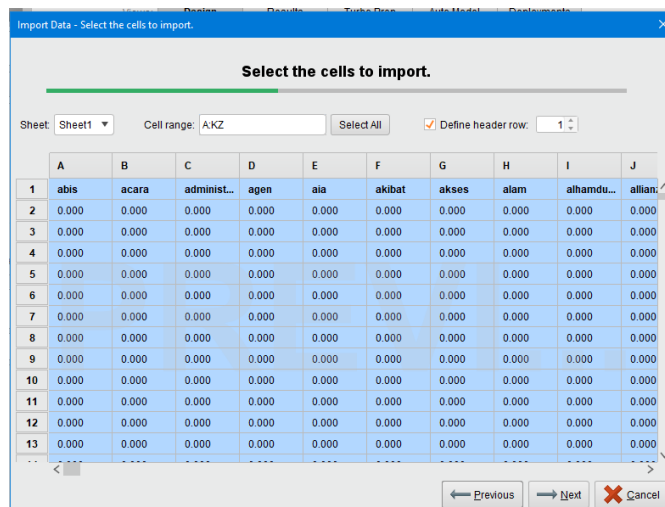
3.3.2 *Penentuan Nilai Support Dan Confidence*

Sebelum melalui tahap algoritma *FP-Growth* ditentukan nilai *support* dan nilai *confidence* sebagai pembatas hasil yang terbentuk agar tidak menyebabkan hasil yang bias. Untuk mendapatkan nilai *support* dapat menggunakan rumus pada Persamaan 2.2 dan untuk mendapatkan nilai *confidence* dapat menggunakan rumus pada Persamaan 2.4.

3.3.3 *Implementasi FP-Growth pada RapidMiner ver 9.10*

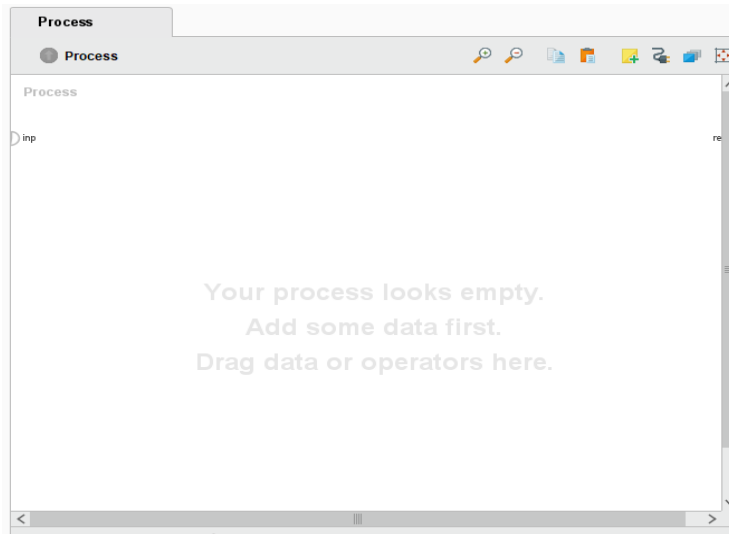
RapidMiner memiliki 500 lebih metode *data science* dan salah satunya terdapat algoritma *FP-Growth* dalam *software* ini. Setelah data sudah berbentuk binomial untuk algoritma *FP-Growth* maka dapat dilanjutkan ke tahap implementasi data pada *RapidMiner*.

1. Hal pertama yang dilakukan saat memproses data yaitu dengan memindahkan data yang akan digunakan dari komputer ke software atau dalam bahasa inggris dikatakan *import data*. Pada Gambar 3.1 terdapat tampilan untuk tahap awal yaitu *import data*.



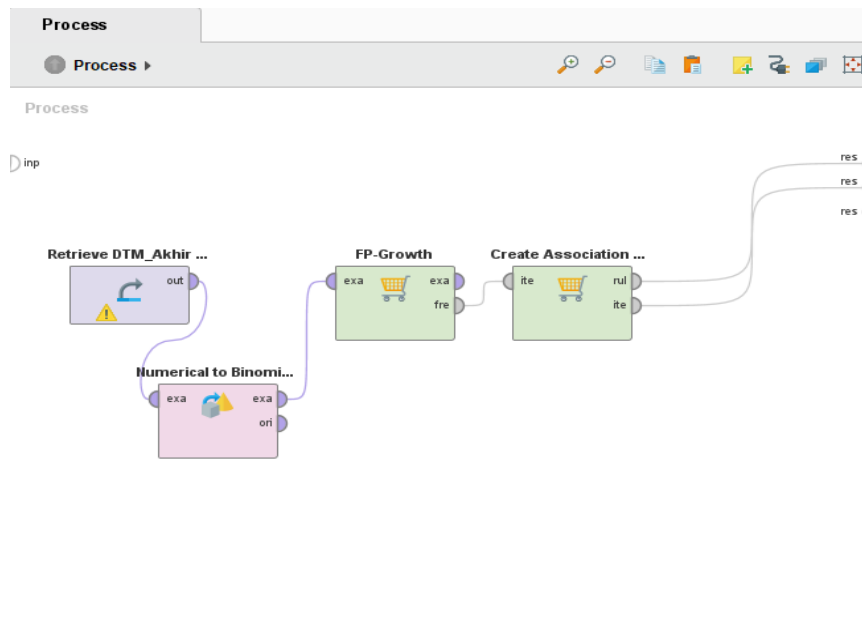
Gambar 3.1 *Import Data*

2. Apabila data sudah tercatat oleh *RapidMiner*, kemudian terdapat halaman proses yang kosong seperti pada Gambar 3.2 untuk meletakkan kotak-kotak proses yang akan digunakan untuk melakukan *FP-Growth*.



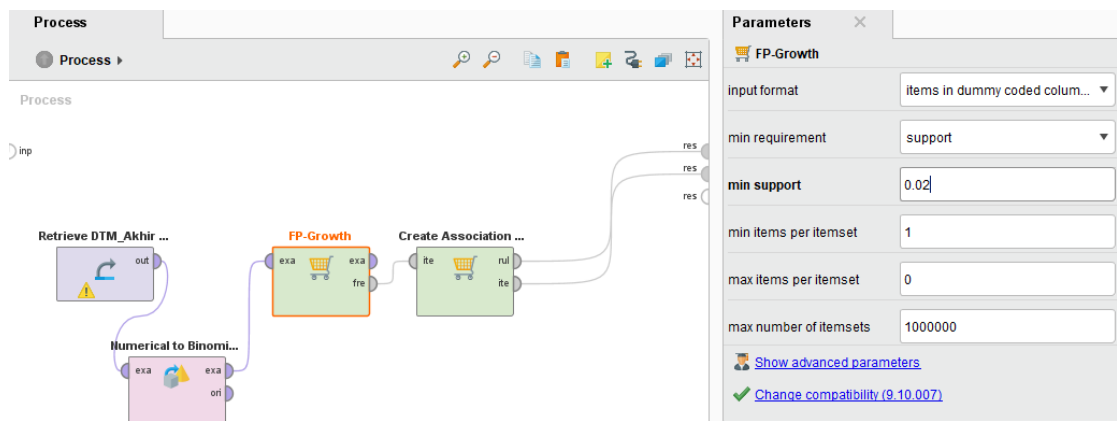
Gambar 3.2 Halaman Proses *Rapidminer*

3. Pada Gambar 3.3 terdapat 4 kotak proses yang diletakkan pada halaman proses. Kotak pertama dari sisi kiri ke kanan berisi data akhir yang digunakan, kotak kedua yaitu *numerical to binomial* yang berfungsi untuk merubah seluruh tipe data akhir dari *integer* menuju binomial. Karena syarat penggunaan *FP-Growth* di *RapidMiner* data harus bertipe binomial agar bisa dilanjutkan ke kotak ketiga yaitu proses *FP-Growth*. Kemudian kotak yang terakhir yaitu kotak untuk metode *association rule*.



Gambar 3.3 Proses *FP-Growth* pada *RapidMiner* 9.10

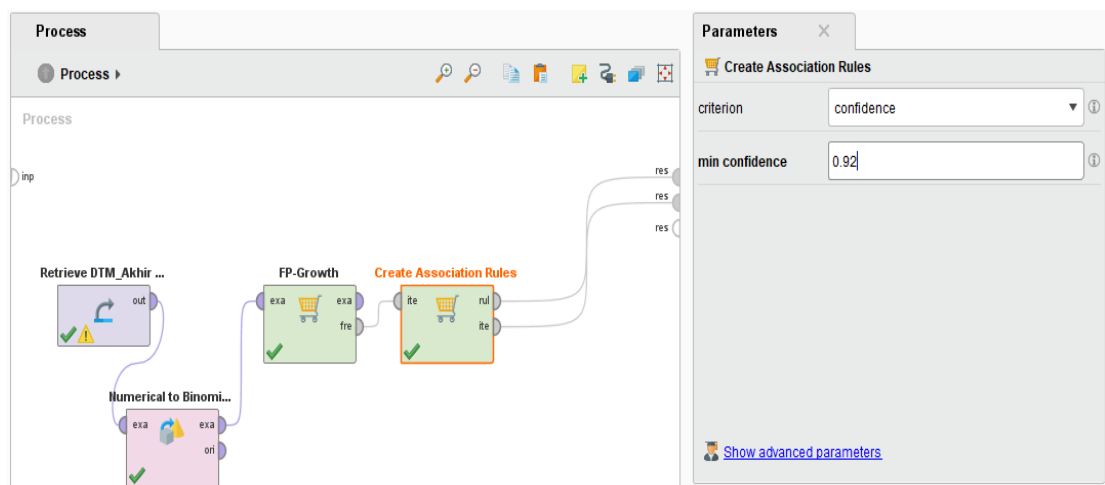
4. Sebelum proses *FP-Growth* di jalankan, pada proses awal *FP-Growth* harus mencantumkan minimum nilai *support* dan nilai *confidence*. Untuk dapat menentukan minimum nilai *support* dapat menggunakan Persamaan 2.2 dan untuk menentukan minimum nilai *confidence* digunakan rumus pada Persamaan 2.4. Pada Gambar 3.4 mencantumkan minimal *support* pada kotak ketiga dari sisi kiri yang berjudul *FP-Growth* sebesar 0,022 sebagai hasil awal untuk melihat nilai *support* paling kecil yang muncul agar bisa menjadi acuan untuk uji coba nilai *support* di tahap *association rule* berikutnya.



Gambar 3.4 Mencantumkan Nilai Minimum *Support*

Selanjutnya mencantumkan pada kotak terakhir dari sisi kiri atau kotak yang bernama “*Create Association Rules*” seperti pada Gambar 3.5 sebesar 0,92 sebagai batasan minimal nilai kepastian antar kata yang terdapat di itemset. Setelah kedua nilai *support* dan *confidence* sudah dimasukkan maka proses algoritma *FP-Growth* dengan bantuan *software Rapidminer* dapat dijalankan.

Catatan untuk menjalankan algoritma *FP-Growth* dengan *RapidMiner* untuk mencantumkan minimal *support* dan *confidence* tidak terlalu tinggi karena hasil akhir/output yang dihasilkan *RapidMiner* akan mengeluarkan semua item yang digunakan. Karena itu alangkah baiknya jika melakukan uji coba nilai *support* dan nilai *confidence* terbaik.



Gambar 3.5 Mencantumkan Nilai Minimum *Confidence*

5. Gambar 3.6 adalah hasil dari data dari proses algoritma *FP-Growth* setelah di *run*.

Size	Support	Item 1
1	0.978	asuransi
1	0.109	sehat
1	0.071	jiwa
1	0.066	beli
1	0.064	bayar
1	0.062	bpjs
1	0.057	uang
1	0.053	syariah
1	0.045	biaya
1	0.045	kerja
1	0.042	sakit
1	0.042	usaha
1	0.036	keluarga
1	0.036	prudential
1	0.035	masuk
1	0.034	investasi

Gambar 3.6 Hasil *Running FP-Growth* pada *RapidMiner* untuk 1-itemset

Size	Support	Item 1	Item 2	Item 3
2	0.025	asuransi	mandiri	
2	0.025	asuransi	rs	
2	0.024	asuransi	barang	
2	0.024	asuransi	mobil	
2	0.023	asuransi	bank	
2	0.023	asuransi	agen	
2	0.022	asuransi	link	
2	0.023	asuransi	polis	
2	0.022	asuransi	yuk	
2	0.023	asuransi	wajib	
2	0.022	asuransi	kaya	
2	0.022	asuransi	nasabah	
2	0.022	asuransi	pakai	
2	0.022	asuransi	cari	
2	0.025	syariah	prudential	
3	0.025	asuransi	syariah	prudential

Gambar 3.7 Hasil *Running FP-Growth* pada *RapidMiner* untuk 2-itemset dan 3 itemset

Berdasarkan Gambar 3.7 menjelaskan hasil dari proses algoritma *FP-Growth* dengan *RapidMiner* untuk 2-itemset yang ber label 2 pada kolom *size* kemudian 3-itemset untuk

ber label 3 pada kolom *size*, begitu seterusnya apabila didapatkan pada data yang digunakan lebih dari 3-itemset.

3.3.4 Pencarian Aturan/rule Terbaik

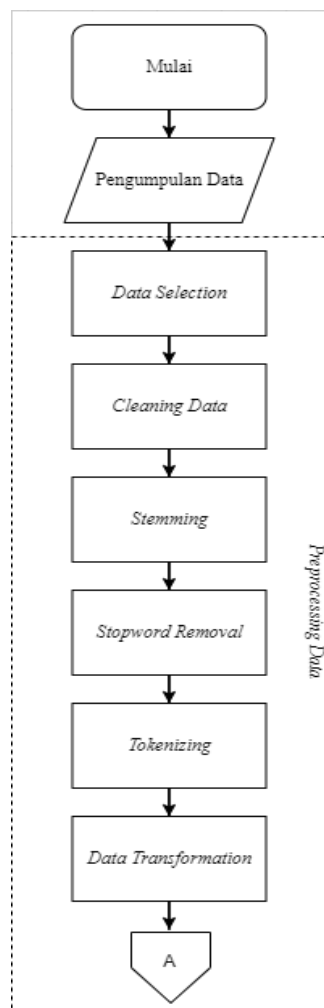
Setelah mendapatkan hasil dari algoritma *FP-Growth*, kemudian aturan yang didapatkan masih terlalu banyak. Oleh sebab itu dilakukan pencarian aturan/*rule* yang terbaik dengan mengulang algoritma *FP-Growth* melalui *RapidMiner* dengan menggunakan beberapa nilai *support* dan *confidence* dan tidak memasukkan kata yang memiliki nilai *lift ratio* kurang dari 1 yang berarti tidak memiliki hubungan antar kata sehingga lebih mudah untuk menginterpretasikan dan mendapatkan *insight* dari kata yang berasosiasi dengan asuransi.

3.3.5 Interpretasi Hasil

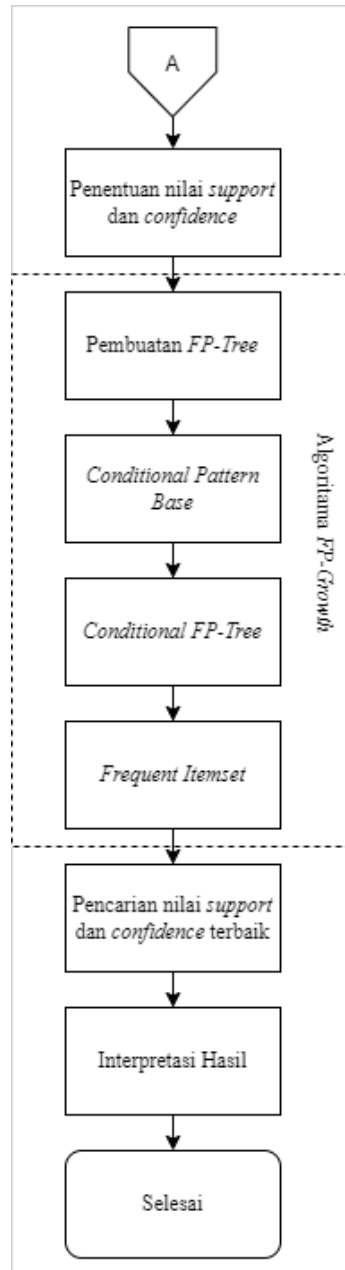
Pada tahap ini dilakukan interpretasi hasil dari aturan yang didapat sesuai dengan tujuan penelitian yang sudah ditentukan.

3.4 Diagram Alir Penelitian

Diagram alir penelitian “Implementasi *Association Rule Mining* Untuk Menganalisis Data Twitter Tentang Asuransi di Indonesia Menggunakan Algoritma *Frequent Pattern-Growth*” disajikan pada Gambar 3.8.



Gambar 3.8 Diagram Alir Penelitian



Gambar 3.8 Diagram Alir Penelitian (Lanjutan)

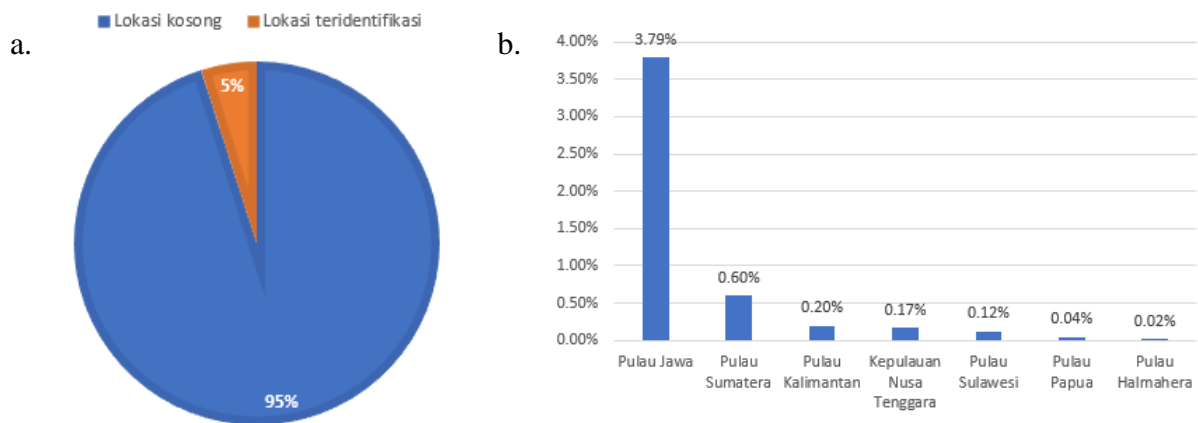
BAB IV HASIL DAN PEMBAHASAN

4.1 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 relevan dan memiliki nilai tinggi, sehingga selanjutnya data dapat dipergunakan untuk mencapai tujuannya.

4.1.1 Persebaran lokasi pengguna twitter yang berpendapat di Indonesia

Pada data yang digunakan terdapat bermacam-macam pendapat yang tersebar di twitter mengenai asuransi di Indonesia. Untuk mengetahui persebaran lokasi pendapat mengenai asuransi di Indonesia dibagi per pulau di Indonesia dengan menggunakan *pie chart* untuk visualisasi persebaran pengguna twitter yang berpendapat mengenai asuransi di Indonesia yang disajikan pada Gambar 4.1.a dan Gambar 4.1.b.



Gambar 4.1 a). Identifikasi Lokasi Pengguna Twitter Yang Berpendapat Asuransi. b). Persebaran Lokasi yang Teridentifikasi

Berdasarkan Gambar 4.1.a dapat dilihat bahwa persebaran pengguna twitter yang berpendapat mengenai asuransi di Indonesia hanya terisi 5% dari total data yang digunakan pada penelitian ini. Pada Gambar 4.1.b dijelaskan lebih dalam isi lokasi yang teridentifikasi 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.

Pengguna twitter yang berpendapat mengenai asuransi di Indonesia pada pulau tertinggi yaitu pulau Jawa sebesar 76,73% tersebar di provinsi Jawa Barat 207 orang, DKI Jakarta sebesar 197 orang, Jawa Tengah 150 orang, Jawa Timur 122 orang, Banten 50 orang, dan DIY 34 orang. Kemudian pulau tertinggi kedua yang ada pada pulau Sumatera sebesar 12,20% atau 103 orang. Pada pulau Kalimantan terdapat pengguna twitter yang berpendapat mengenai asuransi di Indonesia sebesar sebesar 4,07% atau 40 orang. Dilanjut pada Kepulauan Nusa Tenggara sebesar 3,46% yang terdiri dari provinsi Bali 22 orang, NTB sebanyak 10 orang dan NTT hanya 2 orang. Pada pulau Sulawesi terdapat pengguna twitter yang berpendapat mengenai asuransi di Indonesia sebesar sebesar 2,44% atau 24 orang. Pada pulau Papua terdapat

pengguna twitter yang berpendapat mengenai asuransi di Indonesia sebesar sebesar 0,81% atau 8 orang. Dan pada pulau terkecil yaitu pulau Halmahera sebesar 0,3% atau 3 orang.

Data jumlah penduduk dari BPS Indonesia bahwa total penduduk Indonesia yaitu sebanyak 269,6 juta jiwa yang tersebar ke berbagai pulau yaitu, 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 4.1.

Tabel 4.1 Proporsi Jumlah Penduduk Di Tiap Pulau Dengan Tweets Mengenai Asuransi

	Penduduk Indonesia	Tweets	Proporsi tweets dengan penduduk
Sumatera	21,92%	0,60%	0,03
Jawa	56,23%	3,79%	0,07
Kepulauan Nusa Tenggara	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. kemudian pulau Sumatera, Kepulauan Nusa Tenggara, dan pulau Kalimantan menjadi pulau dengan kategori tinggi dengan proporsi sebesar 0,03. Pulau Sulawesi dan pulau Papua menjadi kategori sedang dengan proporsi penduduk yang membicarakan asuransi di twitter sebesar 0,02. Dan pulau yang masuk kategori rendah dalam membicarakan mengenai asuransi di Twitter dengan proporsi 0,01 yaitu pulau Halmahera. Sehingga mendapatkan hasil bahwa akun yang berpendapat mengenai asuransi Indonesia di Twitter tidak merata pada tiap pulau di Indonesia.

4.1.2 Persebaran akun yang berpendapat mengenai asuransi di Indonesia

Berdasarkan website databoks.co.id, total pengguna twitter di Indonesia kurang lebih ada di angka 15,7 juta pengguna dan pengguna twitter yang berpendapat tentang asuransi di Indonesia sejak 2017 sekitar 0,0007%. Angka ini terbilang tidak sedikit dikarekanakan berjuta-juta topik pembicaraan yang ada di forum twitter. Informasi pada Tabel 4.2 disajikan untuk mengetahui siapa saja yang berpendapat mengenai asuransi di Indonesia dari 0,0007% pengguna twitter yang tercatat berpendapat mengenai asuransi di Indonesia dengan membaginya ke dalam akun twitter yang *verified* (centang biru) atau *unverified*. Karena dari pihak twitter, semenjak tahun 2017 diadakan kembali sistem *verified* hanya untuk kategori akun khusus seperti akun lembaga pemerintahan, perusahaan, merk, organisasi besar, media, hiburan, olahraga, aktivis, dan *influencer* (Stephanie, 2021).

Tabel 4.2 Jumlah Akun *Verified* dan *Unverified* Yang Berpendapat Mengenai Asuransi di Indonesia

	Total (%)	Verified (%)	Unverified (%)
Total Akun Twitter	100	0,18	99,82
Total Akun Yang Berpendapat Mengenai Asuransi di Indonesia	100	1,33	98,67

Pada Tabel 4.2 didapatkan bahwa ternyata pengguna twitter yang berbicara atau berpendapat mengenai asuransi di Indonesia pada akun *verified* seperti akun lembaga pemerintahan, perusahaan, merk, organisasi besar, media, hiburan, olahraga, aktivis, dan *influencer* memiliki nilai perbandingan yang lebih besar daripada seluruh akun *verified* yang ada di Twitter, berarti banyak akun *verified* yang berpendapat mengenai asuransi di Indonesia di Twitter jika dilihat dari perbandingan total akun yang *verified* di Indonesia sebesar 1,33% dari keseluruhan data yang digunakan dengan total akun *verified* yang berpendapat mengenai asuransi di Indonesia sebesar 0,18%. Hal ini juga dapat menunjukkan apabila akun-akun *verified* yang berpendapat mereka kemungkinan besar tidak mengeluarkan kata-kata yang sembarang atau kemungkinan besar mengeluarkan kalimat yang senonoh bahkan bijak karena pada setiap akun *verified* mereka mempunyai popularitas yang harus dijaga karena sering menjadi sorotan banyak pengguna twitter lainnya atau masyarakat di seluruh dunia dan juga akun *verified* hanya akun yang terpilih dan memenuhi syarat dari pihak twitter.

Mayoritas akun yang berpendapat mengenai asuransi di Indonesia yaitu akun yang *unverified* yaitu sebesar 98,67%. Dapat dikatakan bahwa pengguna yang berpendapat mengenai asuransi di twitter kebanyakan mengeluarkan pendapat sesuai isi hati dan kondisi dari pengguna twitter itu sendiri sehingga menyebabkan banyak kalimat/*tweets* yang berhubungan dengan asuransi menjadi tidak formal atau tidak sesuai dengan KBBI dan kalimat harus dibersihkan/*cleaning data* sebelum digunakan pada algoritma *FP-Growth*.

4.1.3 Word cloud kata asuransi

Penggunaan *word cloud* pada kata tiap tweet tentang asuransi di Indonesia membantu untuk mendapatkan pemahaman tentang suatu ide atau gagasan yang sedang diteliti. Visualisasi *word cloud* ini memberikan kemudahan untuk melakukan interpretasi terhadap kata asuransi dan juga untuk dapat mengetahui kata apa saja yang sering muncul pada tiap kalimat (*tweet*) yang berpendapat mengenai asuransi secara cepat dan mudah dipahami.

Untuk melihat dengan mudah kata apa saja yang sering bermunculan pada kalimat tweet yang ada di twitter yang berpendapat mengenai asuransi di Indonesia dapat dilihat dari *word cloud* pada Gambar 4.2 dibawah ini.



Gambar 4.2 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. Semakin besar ukuran kata pada Gambar 4.2

menunjukkan semakin mendominasi kata tersebut pada dataset yang berhubungan dengan asuransi.

Kata sehat yang muncul merujuk kepada kata kesehatan yang berarti banyak orang pada saat berpendapat atau menulis di sosial media twitter mengenai asuransi, membicarakan mengenai kesehatan pula. Untuk kata bpjs yang sering muncul berkaitan dengan program pemerintah untuk membantu masyarakat Indonesia yang kesusahan untuk mendapatkan asuransi. Kata beli yang muncul pada kumpulan awan yang terbentuk menjelaskan bahwa banyak orang yang membicarakan beli saat berpendapat mengenai asuransi di Indonesia. Dari *word cloud* ini dapat dikatakan bahwa kata “bpjs”, “sehat”, dan “beli” merupakan kata yang paling sering muncul pada saat pengguna twitter berpendapat mengenai asuransi.

4.2 Data Preprocessing

Data yang digunakan dalam penelitian ini berasal dari twitter mengenai asuransi di Indonesia yang terdapat pada Lampiran 3. Data tersebut di proses menjadi data yang sesuai format agar dapat digunakan untuk analisis berikutnya yaitu algoritma *FP-Growth* melalui beberapa tahapan *data processing* sebagai berikut.

4.2.1 Data Selection

Dari data yang ada sering ditemukan bahwa tidak semua data dapat digunakan pada saat proses algoritma. Oleh karena itu, data ini di seleksi agar dapat digunakan saat proses *association rule*. Setelah data diambil dari twitter melalui program *jupyter notebook* kolom yang digunakan untuk data penelitian selanjutnya hanya kolom *text*/kalimat tweet.

Dari kolom *text* yang berjumlah 19.926 terdapat 602 kata berbeda yang akan digunakan untuk mencari hubungan kata “asuransi” dengan hasil *association rule* yang akan didapatkan. Dikarenakan kalimat yang di dapatkan dari twitter masih berupa data mentah/*rawdata* yang mengandung *link*, *hashtag*, *emoji*, angka, spasi berlebih, *username*, *hashtag*, huruf kapital, kata yang tidak mengandung arti. Sehingga data masih harus diberishkan dari atribut tersebut hingga tersisa hanya kalimat yang memiliki arti pada setiap kata. Oleh karena itu dilakukan pada tahap *processing data* yaitu *cleaning data* untuk tahap pertama dengan hasil pada Tabel 4.3.

Tabel 4.3 Contoh Kalimat Mentah Dari Twitter

No.	Tweet
1.	bisa banget si kamu memilih asuransi unit link dari AXA Mandiri ini dan nantinya bisa menambah asuransi lainnya sesuai kebutuhan kamu selain dari asuransi jiwa nih. @AXA_Mandiri https://t.co/T9R0vjgpBB
2.	@sudahdipakai @PartaiSocmed Lalu apa hubungannya sanksi gabisa urus SIM/STNK klo gapunya BPJS? Atau klo semua WNI diwajibkan punya BPJS, terus perusahaan asuransi swasta jualan apa dong?
3.	Bayangkan anak-anak Anda tidak memiliki akses ke pendidikan apa pun, orang tua Anda tidak memiliki akses ke asuransi kesehatan apa pun. Beginilah kehidupan pengungsi Afghanistan di Indonesia. #RescueAfgRefugees_Indonesia https://t.co/f7VUBpZK1g
4.	@KehangatanAI ANDIN TIDAK BUTUH ASURANSI JIWA DARI SEORANG "PENGACARA YANG SAKIT JIWA" Cuma Aldebaran yang bisa menyembuhkan jiwa nya. Titik! No Debat!
5.	Lalu terkena penyakit siapa yang akan membayarkan tagihan Rumah sakit kamu? Sedangkan kalau kamu sakit otomatis akan lebih banyak pengeluaran dari pada punya asuransi.

Tabel 4.3 merupakan 5 contoh dari keseluruhan data yang digunakan yang mengandung kata asuransi di dalamnya. Dari kelima contoh kalimat mentah yang didapatkan dari twitter, terdapat kata yang hurufnya kapital semua seperti pada kalimat di no.4 yaitu “ANDIN TIDAK BUTUH ASURANSI JIWA DARI SEORANG "PENGACARA YANG SAKIT JIWA”, ada yang mengandung *link* seperti pada no.1 yaitu <https://t.co/T9R0vjgpBB>, ada yang terdapat *username twitter* seperti pada no.1 yaitu “@AXA_Mandiri”. Dari tahap data selection ini hanya menghasilkan kalimat pada twitter yang berhubungan dengan asuransi untuk dicari hubungan asosiasinya akan tetapi masih terdapat atribut seperti *link*, *hashtag*, *emoji*, angka, spasi berlebih, *username*, *hashtag*, huruf kapital, kata yang tidak mengandung arti.

4.2.2 Cleaning Data

Pada tahap *data cleaning* dilakukan penghilangan atribut yang tidak digunakan pada tahap analisis seperti *link*, *retweet*, *username*, baris baru, angka, *hashtag*, tanda baca, emoji, spasi yang berlebih, dan mengubah semua huruf menjadi huruf kecil. Untuk melihat hasil *data cleaning* terdapat beberapa data yang ditunjukkan pada Tabel 4.4.

Tabel 4.4 Hasil Proses *Data Cleaning*

<i>Text</i>	<i>Text setelah cleaning data</i>
bisa banget si kamu memilih asuransi unit link dari AXA Mandiri ini dan nantinya bisa menambah asuransi lainnya sesuai kebutuhan kamu selain dari asuransi jiwa nih. @AXA_Mandiri https://t.co/T9R0vjgpBB	bisa banget si kamu memilih asuransi unit link dari axa mandiri ini dan nantinya bisa menambah asuransi lainnya sesuai kebutuhan kamu selain dari asuransi jiwa nih
@sudahdipakai @PartaiSocmed Lalu apa hubungannya sanksi gabisa urus SIM/STNK klo gapunya BPJS? Atau klo semua WNI diwajibkan punya BPJS, terus perusahaan asuransi swasta jualan apa dong?	lalu apa hubungannya sanksi gabisa urus sim stnk klo gapunya bpjs atau klo semua wni diwajibkan punya bpjs terus perusahaan asuransi swasta jualan apa dong
Bayangkan anak-anak Anda tidak memiliki akses ke pendidikan apa pun, orang tua Anda tidak memiliki akses ke asuransi kesehatan apa pun. Beginilah kehidupan pengungsi Afghanistan di Indonesia. #RescueAfgRefugees_Indonesia https://t.co/f7VUBpZK1g	bayangkan anak anak anda tidak memiliki akses ke pendidikan apa pun orang tua anda tidak memiliki akses ke asuransi kesehatan apapun beginilah kehidupan pengungsi afghanistan di indonesia
@KehangatanAI ANDIN TIDAK BUTUH ASURANSI JIWA DARI SEORANG "PENGACARA YANG SAKIT JIWA" Cuma Aldebaran yang bisa menyembuhkan jiwa nya. Titik! No Debat!	andin tidak butuh asuransi jiwa dari seorang pengacara yang sakit jiwa cuma aldebaran yang bisa menyembuhkan jiwanya titik no debat
Lalu terkena penyakit siapa yang akan membayarkan tagihan Rumah sakit kamu? Sedangkan kalau kamu sakit otomatis akan lebih banyak pengeluaran dari pada punya asuransi.	lalu terkena penyakit siapa yang akan membayarkan tagihan rumah sakit kamu sedangkan kalau kamu sakit otomatis akan lebih banyak pengeluaran dari pada punya asuransi

Dari Tabel 4.4 terlihat bahwa hasil proses *cleaning data* menjadikan kalimat mentah dari twitter menjadi kalimat umum dengan menjadikan setiap huruf berukuran kecil. Dan pada tahap

ini menghasilkan kata yang berbeda menjadi 588 kata dari 602 kata pada tahap sebelumnya *data selection*. 14 kata yang hilang yaitu angka yang tercantum pada kalimat sebelumnya di tahap *data selection* dan kemudian dihapus karena angka tidak dibutuhkan untuk mendapatkan kata lain yang berhubungan dengan asuransi. Sehingga hasil dari *data cleaning* ini adalah kalimat yang terdiri dari kata dan tidak terdapat atribut seperti *link*, *retweet*, *username*, baris baru, angka, *hashtag*, tanda baca, emoji, spasi yang berlebih, dan huruf kapital.

4.2.3 Stemming

Setelah data yang digunakan sudah tidak ada *link*, *retweet*, *username*, baris baru, angka, *hashtag*, tanda baca, emoji, spasi yang berlebih, dan huruf kecil pada setiap kalimat. Hal tersebut dilakukan agar program tidak membaca atribut menjadi sesuatu yang berarti. Kemudian dilanjutkan ke tahap *stemming* dengan menggunakan *library* di program *jupyter notebook* yaitu Sastrawi. *Library* Sastrawi digunakan karena memiliki kelebihan pada proses *stemming* dengan bahasa Indonesia dengan syntax yang terdapat pada Lampiran 4. Pada Tabel 4.5 disajikan beberapa data dari hasil proses *stemming*.

Tabel 4.5 Hasil proses Data *Stemming*

<i>Text</i>	<i>Text setelah stemming</i>
bisa banget si kamu memilih asuransi unit link dari axa mandiri ini dan nantinya bisa menambah asuransi lainnya sesuai kebutuhan kamu selain dari asuransi jiwa nih	bisa banget si kamu pilih asuransi unit link dari axa mandiri ini dan nanti bisa tambah asuransi lain sesuai butuh kamu selain dari asuransi jiwa nih
Lalu apa hubungannya sanksi gabisa urus sim stnk klo gapunya bpjs atau klo semua wni diwajibkan punya bpjs terus perusahaan asuransi swasta jualan apa dong	lalu apa hubung sanksi gabisa urus sim stnk klo gapunya bpjs atau klo semua wni wajib punya bpjs terus usaha asuransi swasta jual apa dong
bayangkan anak anak anda tidak memiliki akses ke pendidikan apa pun orang tua anda tidak memiliki akses ke asuransi kesehatan apapun beginilah kehidupan pengungsi afghanistan di indonesia	bayang anak anak anda tidak milik akses ke pendidikan apa pun orang tua anda tidak milik akses ke asuransi sehat apapun begini hidup ungsi afghanistan di indonesia
andin tidak butuh asuransi jiwa dari seorang pengacara yang sakit jiwa cuma aldebaran yang bisa menyembuhkan jiwanya titik no debat	tidak butuh asuransi jiwa dari orang acara yang sakit jiwa cuma aldebaran yang bisa sembuh jiwa titik no debat
lalu terkena penyakit siapa yang akan membayarkan tagihan rumah sakit kamu sedangkan kalau kamu sakit otomatis akan lebih banyak pengeluaran dari pada punya asuransi	lalu kena sakit siapa yang akan bayar tagih rumah sakit kamu sedang kalau kamu sakit otomatis akan lebih banyak pengeluaran dari pada punya asuransi

Pada proses *stemming*, jumlah kata yang berbeda tidak berubah dikarenakan pada proses ini hanya menyederhanakan kata dari kata yang memiliki imbuhan menjadi kata dasarnya. Seperti pada contoh kalimat pertama pada Tabel 4.5, pada kata “memilih” yang apabila diuraikan menjadi imbuhan *me-* dan kata dasar *pilih*. Dengan imbuhan *me-* berarti sedang melakukan perbuatan pada kata dasar (Purwanto, 2021). Sehingga pada proses *stemming* hasil yang didapatkan yaitu kata “*pilih*” saja. Dari seluruh data hasil proses *stemming*, menyisakan kata pada tiap *tweets* yang tidak memiliki imbuhan agar program tidak membaca 2 kata dengan arti berbeda.

4.2.4 Stopword Removal

Setelah kalimat sudah melalui proses *stemming* atau tidak ada kata pada kalimat yang memiliki imbuhan kemudian dilakukan proses *stopword removal* untuk menghilangkan kata yang tidak memiliki arti agar kata yang digunakan untuk mendapatkan hubungan dengan asuransi hanya kata yang memiliki arti saja. Daftar kata yang akan digunakan pada proses *stopword removal* sebanyak 726 kata yang terlampir pada Lampiran 5. Hasil proses *stopword removal* dapat dilihat pada Tabel 4.6.

Tabel 4.6 Hasil Proses *Stopword Removal*

<i>Text</i>	<i>Text setelah stopwords removal</i>
bisa banget si kamu pilih asuransi unit link dari axa mandiri ini dan nanti bisa tambah asuransi lain sesuai butuh kamu selain dari asuransi jiwa nih	pilih asuransi unit link axa mandiri asuransi sesuai butuh asuransi jiwa
lalu apa hubung sanksi gabisa urus sim stnk klo gapunya bpjs atau klo semua wni wajib punya bpjs terus usaha asuransi swasta jual apa dong	hubung sanksi gabisa urus sim stnk gapunya bpjs wni wajib bpjs usaha asuransi swasta jual
bayang anak anak anda tidak milik akses ke pendidikan apa pun orang tua anda tidak milik akses ke asuransi sehat apapun begini hidup ungsi afghanistan di indonesia	bayang anak anak tidak milik akses didik orang tua tidak milik akses asuransi sehat hidup ungsi afghanistan indonesia
andin tidak butuh asuransi jiwa dari orang acara yang sakit jiwa cuma aldebaran yang bisa sembuh jiwa titik no debat	andin tidak butuh asuransi jiwa orang acara sakit jiwa aldebaran sembuh jiwa titik no debat
lalu kena sakit siapa yang akan bayar tagih rumah sakit kamu sedang kalau kamu sakit otomatis akan lebih banyak pengeluaran dari pada punya asuransi	kena sakit bayar tagih rumah sakit sakit otomatis asuransi

Ternyata pada *list stopwords removal* yang berisi 726 kata tersebut tidak semua digunakan, hanya dibentuk menjadi satu kamus *stopword removal*. Apabila pada data penelitian tertentu terdapat kata yang tidak memiliki arti yang masuk pada kamus *stopword removal* maka dapat langsung terhapus tanpa perlu menulis ulang *list* kata apa saja yang mau dihapus dan masuk ke kamus *stopword removal*.

Pada data ini berhasil menghapus kata yang tidak memiliki arti sebanyak 186 kata yang menyisakan 402 kata berbeda yang akan digunakan untuk ke analisis *FP-Growth*. Contoh beberapa hasil kata yang berhasil dihapus pada tahap *stopword removal* yaitu menggunakan baris pertama pada Tabel 4.6 sebagai berikut: “bisa”, “banget”, “si”, “kamu”, “dari”, “ini”, “dan”, “nanti”, “bisa”, “tambah”, “lain”, “dari”, “selain”, “nih”. Dari proses *stopword removal* menghasilkan kalimat yang berpendapat mengenai asuransi dan pada tiap kata memiliki arti atau sesuai dengan KBBI.

4.2.5 Tokenizing

Pada Tabel 4.6, setelah setiap kalimat sudah dihapus dari *list stopwords* kemudian dilakukan tahap *tokenize* atau pemisahan kata di dalam kalimat menjadi potongan kata tunggal atau token supaya program dapat membaca dan menghitung frekuensi tiap kata dengan lebih efisien. Pada proses *tokenizing* dengan *jupyter notebook* data yang digunakan diubah menjadi

tipe data *string* agar memudahkan program menghitung tiap kata untuk proses selanjutnya yaitu *data transformation*.

4.2.6 Data Transformation

Setelah setiap kata sudah terpisah dilanjutkan ke tahap terakhir yaitu menghitung frekuensi tiap kata menggunakan *library scikit-learn* yang bernama *CountVectorizer* dengan mengubah data menjadi format yang diperlukan *RapidMiner*. Contoh hasil perhitungan frekuensi tiap kata dari Tabel 4.7 dapat dilihat pada Tabel 4.8.

Tabel 4.7 Hasil proses *Data Transformation*

	acara	afghanistan	akses	anak	andin	asuransi	...	wni
1.	0	0	0	0	0	3	...	0
2.	0	0	0	0	0	1	...	1
3.	0	1	1	2	0	0	...	0
4.	1	0	0	0	1	0	...	0
5.	0	0	0	0	0	1	...	0

Selanjutnya dilakukan pengecekan kembali secara manual apakah masih terdapat kata yang tidak mempunyai arti ataupun salah tulis (*typo*) karena banyaknya kata yang ada agar memiliki arti yang berbeda antar kata. Beberapa contoh pada data ini untuk kata yang dihapus secara manual yaitu “wkwk”, “yah”, “tks”, “sampe”. Menghasilkan jumlah kata yang digunakan untuk proses algoritma *FP-Growth* sebanyak 312 kata yang terlampir pada Lampiran 6.

Karena pada proses algoritma *FP-Growth* dengan *RapidMiner* harus biner 0 atau 1 maka apabila terdapat kata memiliki nilai lebih dari 1 dilakukan perubahan nilai menjadi 1 menggunakan *Microsoft excel*. Seperti pada Tabel 4.7 baris ke 3 untuk kata “anak” bernilai 2 maka di transformasi menjadi 1 begitu juga untuk kata yang lainnya. Berikut pada Tabel 4.7 disajikan data yang sudah sesuai format pada *software RapidMiner* berdasarkan hasil dari *data transformation* pada Tabel 4.7.

Tabel 4.8 Format Data Akhir Berdasarkan Data pada Tabel 4.7

	acara	afghanistan	akses	anak	andin	asuransi	...	wni
1.	0	0	0	0	0	1	...	0
2.	0	0	0	0	0	1	...	1
3.	0	1	1	1	0	0	...	0
4.	1	0	0	0	1	0	...	0
5.	0	0	0	0	0	1	...	0

Sehingga data akhir yang sudah sesuai dengan format *software Rapidminer* yaitu biner 0 dan 1 seperti pada Tabel 4.8 kemudian dapat dilakukan langkah selanjutnya yaitu tahap *association rule*

4.3 Menentukan Minimal Support Dan Confidence

Setelah data sudah sesuai format akhir untuk digunakan pada *software RapidMiner*, kemudian ditentukan nilai batas minimal kemunculan kata atau nama lainnya nilai *support* dengan menggunakan rumus Persamaan 2.2. Contoh pada kata “indonesia” muncul 599 kata pada dataset dengan total data yang tercatat sebanyak 19.926 maka untuk menghitung nilai *support* yaitu dengan membagi total frekuensi kata “indonesia” dibagi total data atau $599/19926 = 0,0301$, begitu juga untuk hasil nilai *support* dari 311 kata lainnya yang kemudian diurutkan secara *decending* atau dari nilai tertinggi ke terendah yang tertera pada Tabel 4.9.

Tabel 4.9 Nilai *Support* Tiap Kata

No.	Kata	Jumlah Kata	Support
1	asuransi	19478	0,9775
2	sehat	2175	0,1092
3	jiwa	1413	0,0709
4	beli	1307	0,0656
5	bayar	1281	0,0643
6	bpjs	1226	0,0615
7	uang	1126	0,0565
8	syariah	1053	0,0528
9	biaya	897	0,045
10	kerja	887	0,0445
11	sakit	833	0,0418
12	usaha	832	0,0418
13	keluarga	715	0,0359
14	prudential	715	0,0359
15	masuk	697	0,035
16	investasi	684	0,0343
17	dana	628	0,0315
18	butuh	625	0,0314
19	klaim	619	0,0311
⋮	⋮	⋮	⋮
312	tugas	100	0,005

Setela hasil nilai *support* tiap kata sudah didapatkan dan sudah terurut dari nilai besar ke nilai terkecil seperti pada Tabel 4.9. Kemudian ditentukan batas minimal nilai *support* yang akan digunakan untuk proses algoritma *FP-Growth* dengan membagi kedalam 11 nilai yang menghasilkan jumlah kata berbeda. Semakin besar nilai *support* maka semakin sedikit jumlah kata yang masuk dalam analisis selanjutnya. Akan tetapi apabila jumlah kata yang diperoleh terlalu banyak maka menyebabkan sulit untuk mendapatkan kata kunci yang berhubungan dengan asuransi. Pada Tabel 4.10 yang akan menjelaskan hasil banyaknya kata pada tiap minimal nilai *support* yang ditentukan.

Tabel 4.10 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

Digunakan minimal *support* **0,02** pada nilai yang diberi warna kuning pada Tabel 4.10 menghasilkan 60 kata untuk digunakan ke tahap analisis *FP-Growth*. Nilai *support* 0,02 dipilih karena perbedaan antara 0,02 dan nilai atasnya yaitu 0,03 terbilang cukup banyak sebesar 47 kata. Dan perbedaan jumlah kata antara 0,02 dan nilai bawahnya yaitu 0,01 sebanyak 84 kata. Sehingga menghasilkan perbedaan kata dan juga jumlah kata yang cukup.

Dari minimal nilai *support* yang digunakan yaitu 0,02 menghasilkan 60 kata, kemudian menentukan nilai kepastian atau *confidence* menggunakan Persamaan 2.4 antar kata tersebut. Contoh perhitungan nilai *confidence* pada salah satu kata yaitu “indonesia” (A) dan “asuransi” (B), untuk dapat menghitung nilai *confidence* antara kata “indonesia” dan “asuransi” harus mendapatkan nilai nilai *support* ($A \cap B$) terlebih dahulu dengan menggunakan rumus Persamaan 2.3 yaitu jumlah data yang mengandung kata “indonesia” dan juga “asuransi” secara bersamaan sebanyak 557 data dan total data 19.926 sehingga didapatkan nilai *support* ($A \cap B$) yaitu $557/19926 = 0,028$. Setelah itu baru dapat menghitung nilai *confidence* ($A \cap B$) dengan membagi nilai *support* ($A \cap B$) dengan nilai *support* A yang menghasilkan $0,028/0,0301 = 0,927$. Berdasarkan hasil nilai *confidence* pada 59 kata lainnya yang terlampir pada Lampiran 1, digunakan minimal *confidence* terendah yaitu **0,92**.

4.4 Implementasi *FP-Growth*

Software RapidMiner membantu berjalannya proses *FP-Growth* dari pembuatan *FP-Tree* hingga mendapatkan *frequent itemset* pada data yang digunakan. Nilai *support* yang digunakan untuk menemukan kata yang berhubungan dengan asuransi pada data twitter yaitu sebesar 0,02 dan nilai *confidence* 0,92 berdasarkan hasil perhitungan diatas yang dapat dilihat hasil *FP-Growth* pada Tabel 4.11.

Tabel 4.11 Hasil *FP-Growth* Minimal *Support* 0,02 Dan *Confidence* 0,92

k	<i>Frequent itemset</i>
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 4.11 diperoleh pola kemunculan 3-itemset dengan menggunakan minimal nilai *support* 0,02 dan nilai *confidence* 0,92. Hasil dari tiap itemset tersebut berarti kata yang paling banyak muncul pada data twitter tentang asuransi di Indonesia.

4.5 Pencarian Aturan/*rule* Terbaik

Pada *association rule* digunakan minimal nilai *support* dan *confidence* untuk membatasi kemunculan kata pada hasil akan diperoleh. Sebelumnya pada tahap *FP-Growth* digunakan minimal nilai *support* sebesar 0,02 dan nilai *confidence* 0,92. Dilakukan kembali pencarian nilai *support* dan nilai *confidence* agar mendapatkan aturan yang terbaik sehingga tujuan dari penelitian ini dapat digunakan untuk memberikan saran mengenai asuransi di Indonesia ke perusahaan asuransi maupun masyarakat Indoneisa dengan hasil yang lebih valid. Berikut pada Tabel 4.12 disajikan jumlah aturan yang didapat dari setiap nilai *support* dan *confidence*.

Tabel 4.12 Hasil Pencarian Nilai Minimal *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	16
	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 minimal nilai *support* dan *confidence* terbaik pada Tabel 4.12 menggunakan 12 kombinasi dari nilai *support* 0,02, 0,03, 0,04 dan 0,05 dan nilai *confidence* 0,92, 0,96 dan 0,98 yang dicoba satu per satu pada setiap nilai *support* dan nilai *confidence*. Didapatkan pada baris pertama, ketika minimum *support* 0,02 dan minimum *confidence* 0,92 diperoleh aturan/*rule* sebanyak 48. Pada baris kedua, ketika minimum *support* 0,02 dan minimum *confidence* 0,96 diperoleh aturan/*rule* sebanyak 45 dan seterusnya sampai baris ke 12 yang terlampir hasil aturan pada Lampiran 2.

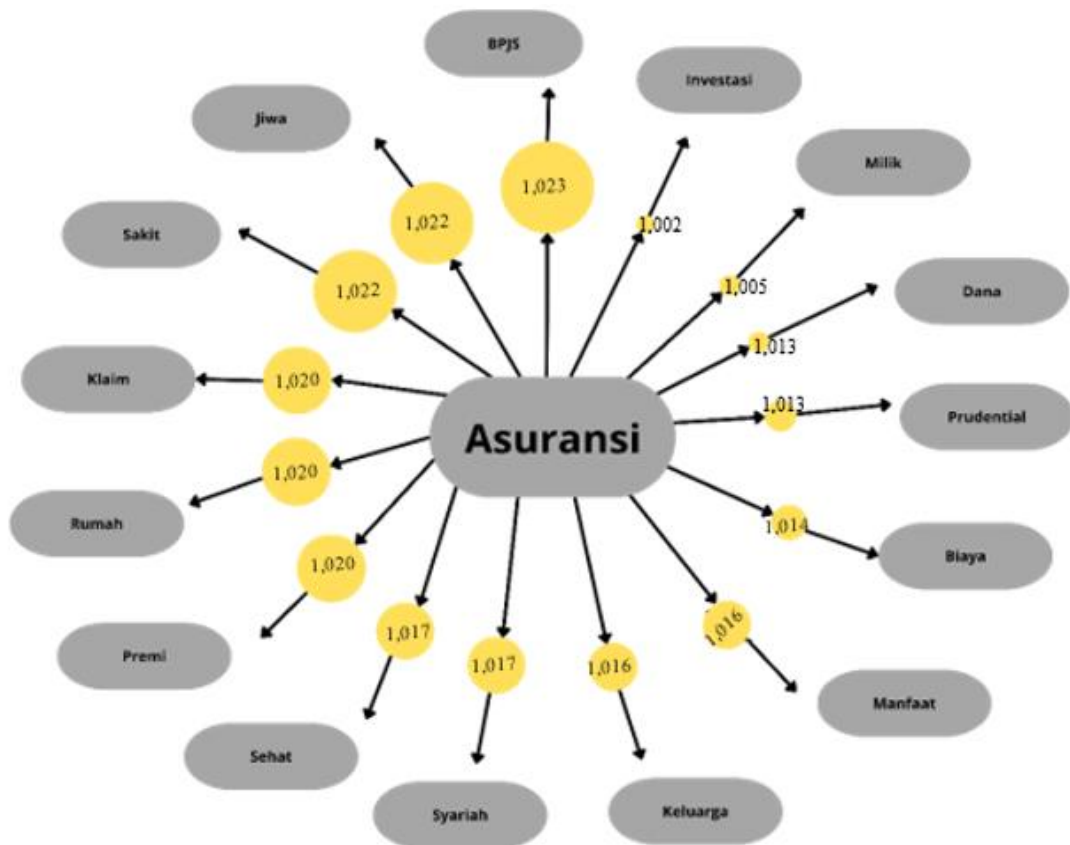
Dari pencarian minimal nilai *support* dan *confidence* terbaik yang telah dilakukan (Tabel 4.12), baris yang diberi warna kuning adalah kombinasi aturan/*rule* yang terbaik. Karena memiliki cukup jumlah kata yang berhubungan dengan asuransi sebanyak 15 kata dan hanya 2 kata yang memiliki arti sama yaitu pada kata “biaya” dan “dana” sehingga apabila akan membuat kesimpulan, menghasilkan informasi yang cukup dalam mendapatkan kata apa saja yang berhubungan dengan asuransi. Alasan lainnya memilih nilai *support* 0,03 karena apabila menggunakan nilai *support* 0,04 atau 0,05 hanya mendapatkan 5 kata yang berhubungan dengan asuransi sehingga tidak menghasilkan informasi yang cukup.

Perbedaan hasil aturan/*rule* dari nilai *support* 0,03 dan *confidence* 0,96 dengan nilai *support* 0,03 dan *confidence* 0,98 adalah pada satu kata yang cukup berhubungan dengan asuransi yaitu kata syariah yang menjadi salah satu jenis produk asuransi. Maka dari itu dipilih nilai *support* 0,03 dan nilai *confidence* 0,96 yang menghasilkan aturan terbaik dengan kata yang berhubungan dengan asuransi pada Tabel 4.13.

Tabel 4.13 Hasil Nilai *Support* 0,03 dan Nilai *Confidence* 0,96

No.	Kata	Support	Confidence	Lift Ratio
1.	syariah	0,053	0,975	1,017
2.	investasi	0,034	0,98	1,002
3.	Milik	0,03	0,982	1,005
4.	dana	0,031	0,99	1,013
5.	prudential	0,036	0,99	1,013
6.	manfaat	0,03	0,993	1,016
7.	keluarga	0,036	0,993	1,016
8.	biaya	0,045	0,994	1,014
9.	sehat	0,109	0,994	1,017
10.	premi	0,03	0,997	1,02
11.	rumah	0,03	0,997	1,02
12.	klaim	0,031	0,997	1,02
13.	sakit	0,042	0,999	1,022
14.	jiwa	0,071	0,999	1,022
15.	bpjs	0,062	1	1,023

Pada Tabel 4.13 merupakan hasil yang digunakan dari proses algoritma *FP-Growth* dengan jumlah *lift ratio* tidak ada yang dibawah 1 yang berarti tiap kata memiliki hubungan dengan asuransi. Nilai *lift ratio* dari 1,002 hingga 1,023 yang berarti apabila nilai *lift ratio* lebih dari 1 mengindikasikan adanya hubungan yang kuat antara kata dan semakin besar nilai *lift ratio* berarti semakin kuat hubungan antar kata tersebut. Hasil yang didapatkan juga divisualisasikan pada Gambar 4.3 agar mudah melihat kata yang berhubungan dengan kata asuransi.



Gambar 4.3 Hasil Asosiasi Nilai *Support* 0,03 dan Nilai *Confidence* 0,96

Berdasarkan hasil kata yang terdapat pada Tabel 4.13 kata yang berasosiasi dengan asuransi adalah “investasi”, “milik”, “dana”, “prudential”, “biaya”, “manfaat”, “keluarga”, “syariah”, “sehat”, “premi”, “rumah”, “klaim”, “sakit”, “jiwa”, dan “bpjs”. Keterangan lingkaran kuning pada Gambar 4.3 apabila nilai *lift ratio* semakin besar maka lingkaran kuning menjadi semakin besar atau dapat diartikan semakin kuat hubungan antar kata tersebut. Kata “bpjs”, “jiwa”, dan “sakit” menjadi 3 kata tertinggi pada nilai *lift ratio* dengan maksud yaitu “bpjs”, “jiwa” dan “sakit” menjadi kata yang paling berhubungan dengan asuransi.

4.6 Intrepretasi/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 Gambar 4.3 atau yang berhubungan dengan asuransi pada Tabel 4.14 sebagai berikut.

Tabel 4.14 Interpretasi Hasil Asosiasi

Kata Kunci	Kata Berhubungan	Frekuensi Kata (Support)	Interpretasi
Asuransi	Bpjs	1226 (0,062%)	Maksud “bpjs” pada asuransi yaitu program jaminan kesehatan dari pemerintah Indonesia agar seluruh rakyat Indonesia mendapatkan perlindungan dalam memenuhi kebutuhan dasar kesehatan.
	Jiwa	1413 (0,071%)	Menurut <i>website</i> 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.
	Sakit	833 (0,042%)	Hubungan kata “sakit” dengan asuransi yaitu 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.
	Klaim	619 (0,031%)	Hubungan antara kata “klaim” dengan asuransi adalah apabila seseorang memiliki asuransi kesehatan yang menanggung manfaat sakit <i>stroke</i> . Jika suatu saat penanggung sakit dan harus dirawat di rumah sakit karena penyakit <i>stroke</i> , 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.
	Rumah	595 (0,030%)	Kata 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 yang berperan penting dalam biaya pengobatannya.

Tabel 4.14 Interpretasi Hasil Asosiasi (Lanjutan)

Kata Kunci	Kata Berhubungan	Frekuensi Kata (Support)	Interpretasi
Asuransi	Premi	590 (0,030%)	Arti premi yaitu jumlah bayaran yang ditetapkan penanggung terhadap tertanggung sebagai biaya risiko yang ditanggung perusahaan asuransi apabila seseorang mempunyai asuransi. Kata “premi” sering dibicarakan oleh penggunanya yang menggunakan twitter dalam hal harga premi yang dimiliki setiap asuransi apakah terlalu mahal atau cukup
	Sehat	2175 (0,109%)	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.
	Syariah	1053 (0,053%)	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 <i>tabarru</i> 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.
	Keluarga	715 (0,036%)	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.
	Manfaat	605 (0,030%)	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.
	Biaya	897 (0,045%)	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

Tabel 4.14 Interpretasi Hasil Asosiasi (Lanjutan)

Kata Kunci	Kata Berhubungan	Frekuensi Kata (Support)	Interpretasi
Asuransi			yang mengeluh mengenai biaya produk asuransi yang terlalu jauh dari anggaran finansialnya.
	Prudential	715 (0,036%)	Maksud kata “prudential” pada asuransi adalah salah satu perusahaan asuransi yang ada di Indonesia. Berarti banyak masyarakat di Indonesia yang membicarakan tentang perusahaan prudential saat berpendapat di twitter mengenai asuransi di Indonesia.
	Dana	628 (0,031%)	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.
	Milik	615 (0,030%)	Terdapat 615 kali kata “milik” muncul pada pendapat di twitter yang berhubungan dengan asuransi, berarti banyak pengguna twitter yang membicarakan untuk setiap warga Indonesia bisa memiliki asuransi sebagai cadangan keuangan jika sewaktu-waktu sakit dan membutuhkan biaya yang besar.
	Investasi	684 (0,034%)	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.

Dari Tabel 4.14 dijelaskan pengertian tiap kata yang berhubungan dengan asuransi dari pendapat yang ada di twitter yang dapat dilihat di tiap baris sehingga bisa memberikan saran kepada perusahaan asuransi maupun masyarakat Indonesia dari pengetahuan baru yang didapat.

(Halaman ini sengaja dikosongkan)

BAB V KESIMPULAN DAN SARAN

5.1 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:
 - a. 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 hubunannya 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. Proses *association rule* menggunakan data twitter tentang asuransi menghasilkan beberapa pengetahuan baru 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. Berdasarkan 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.
 - 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 merasakan 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.

5.2 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

- Aditiya, R., Defit, S., & Nurcahyo, G. W. (2020). Prediksi Tingkat Ketersediaan Stock Sembako Menggunakan Algoritma FP-Growth dalam Meningkatkan Penjualan. *Jurnal Informatika Ekonomi Bisnis*, 2, 67–73. <https://doi.org/10.37034/infv2i3.44>
- Ajib, M. (2019). *Asuransi Syariah* (A. Husna (ed.)). Rumah Fiqih Publishing.
- Ardani, N. R., & Fitriana, 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.
- Castella, Q., & Sutton, C. (2014). *Word Storm: Multiples of Word Clouds for Visual Comparison of Documents*.
- Devina, I. E. (2017). Penggunaan Struktur FP-Tree dan Algoritma FP-Growth dalam Rekomendasi Promosi Produk Pada Situs Belanja Online. *Makalah IF2120 Matematika Diskrit*.
- Firmananda, F. I. (2020). Penerapan Algoritma FP-Growth Dan Economy Order Quantity Untuk Menganalisa Pola Belanja Konsumen Di Apotek. Universitas Islam Negeri Sultan Syarif Kasim Riau Prkanbaru.
- Fitriyanto, E. T. (2017). Penentuan Aturan Asosiasi Pada Transaksi Penjualan Obat Menggunakan Algoritma Apriori *Studi Kasus Pada RSUD Dr. Soetrasno Rembang*.
- Gates, A. (2019). *Association Rule Mining with Tweets: Thinking Outside the Basket*. Georgetown University.
- Gunadi, G., & Sensuse, D. I. (2012). Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth (Fp-Growth) : *Telematika*, 4(1), 118–132.
- Ikhwan, A., Nofriansyah, D., & Sriani. (2015). Penerapan Data Mining dengan Algoritma Fp-Growth untuk Mendukung Strategi Promosi Pendidikan (Studi Kasus Kampus STMIK Triguna Dharma). *Saintikom*, 14(3), 211–226.
- Kannan, S., Gurusamy, V., Vijayarani, S., Ilamathi, J. & Nithya, M. (2015). Preprocessing Techniques for Text Mining Preprocessing Techniques for Text Mining. *International Journal of Computer Science & Communication Networks*, 5(October 2014), 7–16.
- Kusrini, & 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>
- Mannan, M. A. (1997). *Teori dan Praktek Ekonomi Islam* (P. D. B. P. Yasa (ed.)).
- Ngurah, G., Nata, M., & Yudiastra, P. P. (2017). Preprocessing Text Mining Pada Email Box Berbahasa Indonesia. *Konferensi Nasional Sistem & Informatika 2017*, 479–483.
- Nugroho, K. S. (2019). *Dasar Text Preprocessing dengan Python*. Ksnugroho.Medium.Com/. <https://ksnugroho.medium.com/dasar-text-preprocessing-dengan-python-a4fa52608ffe>
- Prodjodikoro, W. (1996). *Hukum Asuransi di Indonesia* (1st ed.). PT Intermasa.
- Purwanto, C. A. (2021). *Bentuk, Fungsi, dan Makna Imbuhan Me(n)- Dalam Novel Kenanga Karya Oka Rusmini. 2*.
- 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
- Qomariyah, S. (2017). Perbandingan Algoritam FP-Growth, Apriori, dan Squezeer Pada Analisis Perilaku Konsumen Di Minimarket K1Mart ITS. *Final Project - SS 141501*.

- Ridwan, A. A. (2016). Asuransi Perspektif Hukum Islam. *Jurnal Hukum Dan Ekonomi Syariah*, 4, 78.
- Sembiring, S. (2014). *Hukum Asuransi* (1st ed.). Nuansa Aulia.
- Stephanie, C. (2021). *Twitter Kembali Buka Program Centang Biru, Begini Cara Mengajukannya*. Kompas. <https://tekno.kompas.com/read/2021/05/21/13070057/twitter-kembali-buka-program-centang-biru-begini-cara-mengajukannya?page=all>
- Susanto, Ambarwati, R., & Tumanggor, S. (2021). Analisis Altman Z-Score Untuk Memprediksi Kebangkrutan Pada Asuransi Umum di Indonesia Sebelum dan Saat Pandemi Covid-19. *HUMANIS (Humanities, Management and Science Proceedings)*, 1(2), 942–950.
- Susanto, S., & Suryani, D. (2010). Pengantar Data Mining Menggali Pengetahuan Dari Bongkahan Data. *Buku*, 1–3.
- Tamaela, J., Sedyono, E., & Setiawan, A. (2018). Implementasi Metode Association Rule untuk Menganalisis Data Twitter tentang Badan Penyelenggara Jaminan Sosial dengan Algoritma Frequent Pattern-Growth. *Jurnal Sistem Informasi Bisnis*, 8(1), 25. <https://doi.org/10.21456/vol8iss1pp25-33>

LAMPIRAN

Lampiran 1. Nilai *confidence* berdasarkan nilai *support* 0,02

No.	Kata	Confidence
1	indonesia	0,927
2	butuh	0,954
3	pilih	0,958
4	bank	0,966
5	hidup	0,967
6	lindung	0,967
7	uang	0,975
8	yuk	0,976
9	investasi	0,980
10	link	0,980
11	usaha	0,981
12	laku	0,981
13	milik	0,982
14	beli	0,982
15	jalan	0,983
16	tanggung	0,986
17	produk	0,988
18	daftar	0,988
19	kerja	0,989
20	pakai	0,989
21	aman	0,989
22	bayar	0,990
23	prudential	0,990
24	dana	0,990
25	biaya	0,991
26	salah	0,992
27	syariah, prudential	0,992
28	keluarga	0,993
29	nasabah	0,993
30	manfaat	0,993
31	polis	0,993
32	masuk	0,994
33	syariah	0,994
34	lengkap	0,994
35	sehat	0,994
36	tabung	0,995
37	bikin	0,995
38	mobil	0,996
⋮	⋮	⋮
60	tugas	1

Lampiran 2. Data Hasil Pencarian Nilai *Support* Dan Nilai *Confidence* Terbaik

- Nilai *support* 0,02 dan nilai *confidence* 0,92

No.	Item	Support	Confidence	Lift
1.	indonesia	0,028	0,927	0,981
2.	butuh	0,03	0,954	0,976
3.	pilih	0,029	0,958	0,976
4.	bank	0,023	0,966	1,023
5.	hidup	0,022	0,967	0,989
6.	lindung	0,027	0,967	1,019
7.	syariah	0,053	0,975	1,017
8.	yuk	0,022	0,976	0,998
9.	link	0,022	0,980	1,016
10.	investasi	0,034	0,980	1,002
11.	laku	0,026	0,981	1,021
12.	milik	0,03	0,982	1,005
13.	jalan	0,026	0,983	0,99
14.	tanggung	0,028	0,986	1,02
15.	produk	0,029	0,988	1,02
16.	pakai	0,022	0,989	1,012
17.	aman	0,027	0,989	1,017
18.	dana	0,031	0,99	1,013
19.	prudential	0,036	0,99	1,013
20.	syariah, prudential	0,025	0,992	1,012
21.	nasabah	0,022	0,993	0,998
22.	polis	0,023	0,993	1,003
23.	manfaat	0,03	0,993	1,016
24.	keluarga	0,036	0,993	1,016
25.	lengkap	0,027	0,994	0,948
26.	biaya	0,045	0,994	1,014
27.	sehat	0,109	0,994	1,017
28.	bikin	0,022	0,995	1,018
29.	mobil	0,024	0,996	0,989
30.	rs	0,025	0,996	1,019
31.	langsung	0,025	0,996	1,006
32.	sesuai	0,027	0,996	1,011
33.	premi	0,03	0,997	1,02
34.	rumah	0,03	0,997	1,02
35.	klaim	0,031	0,997	1,02
36.	kaya	0,022	0,998	1,012
37.	agen	0,023	0,998	1,016
38.	barang	0,024	0,998	1,021
39.	swasta	0,025	0,998	1,015
40.	kantor	0,027	0,998	1,021
41.	kirim	0,027	0,998	1,009
42.	sakit	0,042	0,999	1,022
43.	jiwa	0,071	0,999	1,022
44.	tinggal	0,022	1	1,023
45.	sampe	0,022	1	1,023

Lampiran 2. (Lanjutan)

No.	Item	Support	Confidence	Lift
46.	wajib	0,023	1	1,021
47.	mandiri	0,025	1	1,004
48.	bpjs	0,062	1	1,023

- Nilai *support* 0,02 dan nilai *confidence* 0,96

No.	Item	Support	Confidence	Lift
1.	bank	0,023	0,966	1,023
2.	hidup	0,022	0,967	0,989
3.	lindung	0,027	0,967	1,019
4.	syariah	0,053	0,975	1,017
5.	yuk	0,022	0,976	0,998
6.	link	0,022	0,98	1,016
7.	investasi	0,034	0,98	1,002
8.	laku	0,026	0,981	1,021
9.	milik	0,03	0,982	1,005
10.	jalan	0,026	0,983	0,990
11.	tanggung	0,028	0,986	1,02
12.	produk	0,029	0,988	1,02
13.	pakai	0,022	0,989	1,012
14.	aman	0,027	0,989	1,017
15.	dana	0,031	0,99	1,013
16.	prudential	0,036	0,99	1,013
17.	syariah, prudential	0,025	0,992	1,012
18.	nasabah	0,022	0,993	0,998
19.	polis	0,023	0,993	1,003
20.	manfaat	0,03	0,993	1,016
21.	keluarga	0,036	0,993	1,016
22.	lengkap	0,027	0,994	0,948
23.	biaya	0,045	0,994	1,014
24.	sehat	0,109	0,994	1,017
25.	bikin	0,022	0,995	1,018
26.	mobil	0,024	0,996	0,989
27.	rs	0,025	0,996	1,019
28.	langsung	0,025	0,996	1,006
29.	sesuai	0,027	0,996	1,011
30.	premi	0,03	0,997	1,02
31.	rumah	0,03	0,997	1,02
32.	klaim	0,031	0,997	1,02
33.	kaya	0,022	0,998	1,012
34.	agen	0,023	0,998	1,016
35.	barang	0,024	0,998	1,021
36.	swasta	0,025	0,998	1,015
37.	kantor	0,027	0,998	1,021
38.	kirim	0,027	0,998	1,009
39.	sakit	0,042	0,999	1,022
40.	jiwa	0,071	0,999	1,022

Lampiran 2. (Lanjutan)

No.	Item	Support	Confidence	Lift
41.	tinggal	0,022	1	1,023
42.	sampe	0,022	1	1,023
43.	wajib	0,023	1	1,021
44.	mandiri	0,025	1	1,004
45.	bpjs	0,062	1	1,023

- Nilai *support* 0,02 dan nilai *confidence* 0,98

No.	Item	Support	Confidence	Lift
1.	link	0,022	0,98	1,016
2.	investasi	0,034	0,98	1,002
3.	laku	0,026	0,981	1,021
4.	milik	0,03	0,982	1,005
5.	jalan	0,026	0,983	0,99
6.	tanggung	0,028	0,986	1,02
7.	produk	0,029	0,988	1,02
8.	pakai	0,022	0,989	1,012
9.	aman	0,027	0,989	1,017
10.	dana	0,031	0,99	1,013
11.	prudential	0,036	0,99	1,013
12.	syariah, prudential	0,025	0,992	1,012
13.	nasabah	0,022	0,993	0,998
14.	polis	0,023	0,993	1,003
15.	manfaat	0,03	0,993	1,016
16.	keluarga	0,036	0,993	1,016
17.	lengkap	0,027	0,994	0,948
18.	biaya	0,045	0,994	1,014
19.	sehat	0,109	0,994	1,017
20.	bikin	0,022	0,995	1,018
21.	mobil	0,024	0,996	0,989
22.	rs	0,025	0,996	1,019
23.	langsung	0,025	0,996	1,006
24.	sesuai	0,027	0,996	1,011
25.	premi	0,03	0,997	1,02
26.	rumah	0,03	0,997	1,02
27.	klaim	0,031	0,997	1,02
28.	kaya	0,022	0,998	1,012
29.	agen	0,023	0,998	1,016
30.	barang	0,024	0,998	1,021
31.	swasta	0,025	0,998	1,015
32.	kantor	0,027	0,998	1,021
33.	kirim	0,027	0,998	1,009
34.	sakit	0,042	0,999	1,022
35.	jiwa	0,071	0,999	1,022
36.	tinggal	0,022	1	1,023
37.	sampe	0,022	1	1,023
38.	wajib	0,023	1	1,021

Lampiran 2. (Lanjutan)

No.	Item	Support	Confidence	Lift
39.	mandiri	0,025	1	1,004
40.	bpjs	0,062	1	1,023

- Nilai *support* 0,03 dan nilai *confidence* 0,92

No.	Item	Support	Confidence	Lift
1.	butuh	0,03	0,954	0,976
2.	syariah	0,053	0,975	1,017
3.	investasi	0,034	0,980	1,002
4.	milik	0,03	0,982	1,005
5.	dana	0,031	0,99	1,013
6.	prudential	0,036	0,99	1,013
7.	manfaat	0,03	0,993	1,016
8.	keluarga	0,036	0,993	1,016
9.	biaya	0,045	0,994	1,014
10.	sehat	0,109	0,994	1,017
11.	premi	0,03	0,997	1,02
12.	rumah	0,03	0,997	1,02
13.	klaim	0,031	0,997	1,02
14.	sakit	0,042	0,999	1,022
15.	jiwa	0,071	0,999	1,022
16.	bpjs	0,062	1	1,023

- Uji Coba nilai *support* 0,03 dan nilai *confidence* 0,96

No.	Item	Support	Confidence	Lift
1.	syariah	0,053	0,975	1,017
2.	investasi	0,034	0,98	1,002
3.	milik	0,03	0,982	1,005
4.	dana	0,031	0,99	1,013
5.	prudential	0,036	0,99	1,013
6.	manfaat	0,03	0,993	1,016
7.	keluarga	0,036	0,993	1,016
8.	biaya	0,045	0,994	1,014
9.	sehat	0,109	0,994	1,017
10.	premi	0,03	0,997	1,02
11.	rumah	0,03	0,997	1,02
12.	klaim	0,031	0,997	1,02
13.	sakit	0,042	0,999	1,022
14.	jiwa	0,071	0,999	1,022
15.	bpjs	0,062	1	1,023

- Nilai *support* 0,03 dan nilai *confidence* 0,98

No.	Item	Support	Confidence	Lift
1.	investasi	0,034	0,98	1,002
2.	milik	0,03	0,982	1,005
3.	dana	0,031	0,99	1,013
4.	prudential	0,036	0,99	1,013
5.	manfaat	0,03	0,993	1,016

Lampiran 2. (Lanjutan)

No.	Item	Support	Confidence	Lift
6.	keluarga	0,036	0,993	1,016
7.	biaya	0,045	0,994	1,014
8.	sehat	0,109	0,994	1,017
9.	premi	0,03	0,997	1,02
10.	rumah	0,03	0,997	1,02
11.	klaim	0,031	0,997	1,02
12.	sakit	0,042	0,999	1,022
13.	jiwa	0,071	0,999	1,022
14.	bpjs	0,062	1	1,023

- Nilai *support* 0,04 dan nilai *confidence* 0,98

No.	Item	Support	Confidence	Lift
1.	biaya	0,045	0,994	1,014
2.	sehat	0,109	0,994	1,017
3.	sakit	0,042	0,999	1,022
4.	jiwa	0,071	0,999	1,022
5.	bpjs	0,062	1	1,023

- Nilai *support* 0,05 dan nilai *confidence* 0,98

No.	Item	Support	Confidence	Lift
1.	bpjs	0,062	1	1,023
2.	jiwa	0,071	0,999	1,022
3.	sehat	0,109	0,994	1,017

Lampiran 3. Data Twitter

No	Date	Tweet_id	Text	Username	Followers	Following	Location
1	2022-05-09 23:57:35 +00:00	15238144 70543640 000	@rahasia__ihhh @Emak2Dasteran21 ya adalah, kemarin ada yang nelpon aku cewe, minta kenalan sama aku, eh tau nya nawarin asuransi,,, 😞 🙄	b'si_cio'	1611	1710	
2	2022-05-09 23:44:55 +00:00	15238112 84659360 000	Berkah Lebaran, Premi Asuransi Perjalanan Ikut Terkerek https://t.co/SOMIsh321 l #Asuransi	b'KontanNews'	366031	496	Jakarta, Indonesia
3	2022-05-09 23:41:35 +00:00	15238104 42380210 000	@jaehyunkriuk yang punya asuransi kan bapakmu, bukan kamu	b'dylandidntsleep'	6	16	
4	2022-05-09 23:41:22 +00:00	15238103 89401960 000	Tau gitu dari kemaren beli obat pake asuransi kantor di halodoc	b'dittafloo'	225	248	Dreamland
5	2022-05-09 23:22:48 +00:00	15238057 15323710 000	gpp acak aku terus aku kuat bapakku punya asuransi	b'jaehyunkriuk'	202	452	pubrant, dnf w/o consent,
6	2022-05-09 22:41:23 +00:00	15237952 92436370 000	asuransi kesehatan 60 tahun keatas	b'AyuSarwiti'	0	102	Indonesia, Jakarta Barat
7	2022-05-09 22:18:35 +00:00	15237895 55018600 000	@calyatabinaa Gilak lu, jangan lupa siapin asuransi buat rsnya	b'novansy4h'	18	57	
8	2022-05-09 21:12:11 +00:00	15237728 47721970 000	@SPICYBOSS Aduh,,	b'BochorAlvs'	939	378	
9	2022-05-09 21:01:34 +00:00	15237701 75920570 000	@kleponwajik Asuransi tidak bisa dijadikan konten sedangkan tempat kerja yg bagus, unlimited snack, free coffee, dan lainnya mungkin lebih menjual	b'hydoni'	64	542	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
199 26	2022-03-13 12:20:27 +00:00	15029829 20683760 000	@collegemenfess Asuransi kendaraan?? Emang kendaraannya apa sampai asuransi juga diurusin????	b'dhnaath'	195	199	19926

Lampiran 4. *Syntax* Pengolahan Data

```
#data mining twitter
# Import libraries
import csv
import pandas as pd
import snsrape.modules.twitter as sntwitter
#Specify name and directory of csv file as the result of the scraping
csvFile = open('C:\\Users\\WINDOWS 10\\outpput_skripsi.csv'. 'a'. encoding = 'utf-8') #creates a
file in which you want to store the data.
csvWriter = csv.writer(csvFile)
# Customize the scraping based on variables needed. keywords of tweets. the amount of tweets.
and the date of tweets
maxTweets = 20000 # the number of tweets you require
#2021-09-22
for i,tweet in enumerate(sntwitter.TwitterSearchScaper('Asuransi OR asuransi' + 'since:2017-01-
01 until:2022-05-10').get_items()):
    if i > maxTweets :
        break
    print(tweet.date. i)
    csvWriter.writerow([tweet.date. tweet.id. tweet.content. tweet.user.username.encode('utf-8').
tweet.user.verified. tweet.user.followersCount. tweet.user.friendsCount. tweet.mentionedUsers.
tweet.replyCount. tweet.retweetCount. tweet.likeCount. tweet.media. tweet.lang.encode('utf-8').
tweet.user.location])
    #If you need more information. just provide the attributes
header_name = ['date'. 'tweet_id'. 'text'. 'username'. 'verified'. 'followers'. 'following'.
'mentioned users'. 'retweet'. 'like '. 'reply'. 'media'. 'language'. 'location']
df = pd.read_csv("dataskripsi_outpput.csv". names=header_name)
#preprocessing data
import re
import nltk
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
nltk.download('stopwords')
nltk.download('punkt')
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
text=df['text']
text
#Menghapus link
data_link=[]
for line in text:
    result=re.sub(r"https\S+"." ".line)
    data_link.append(result)
#Menghapus Retweet
data_rt=[]
for line in data link:
    result=re.sub(r"RT"." ".line)
    data_rt.append(result)
```


Lampiran 4. (Lanjutan)

```
#Menghapus Username
data_uname=[]
for line in data_rt:
    result=re.sub(r"@S+." ".line)
    data_uname.append(result)
#Menghapus Baris Baru
data_line=[]
for line in data_uname:
    result=re.sub("\n" ".line)
    data_line.append(result)
#Menghapus Angka
data_num=[]
for line in data_line:
    result=re.sub("\d" ".line)
    data_num.append(result)
#Menghapus Hashtag
data_hashtag=[]
for line in data_num:
    result=re.sub(r"#S+." ".line)
    data_hashtag.append(result)
#Menghapus Emoticon
data_emoticon=[]
for line in data_hashtag:
    result=re.sub(r'<.*?>' ".line)
    data_emoticon.append(result)
#Menghapus Punctuation
data_punc=[]
for line in data_emoticon:
    result=re.sub(r"^[w\s]" ".line)
    data_punc.append(result)
#Menghapus spasi berlebih
data_doubleSPACE=[]
for line in data_punc:
    result=re.sub(r"s+' '.line)
    data_doubleSPACE.append(result)
#Case Folding
data_casef=[]
for line in data_doubleSPACE:
    a=line.lower()
    data_casef.append(a)
#stemming
factory=StemmerFactory()
stemmer=factory.create_stemmer()
df_stemmed=map(lambda x: stemmer.stem(x). data_casef)
data_stemmed=list(df_stemmed)
```

Lampiran 4. (Lanjutan)

```
#Stopword
stopWords=set(stopwords.words('indonesian'))
stopword=open("stopwords_skrpisi.txt").read()
stopword=set(stopword.split())
not_stopword={ }
new_stopword=set([word for word in stopword if not word in not_stopword])
data_stop=[]
for line in data_stemmed:
    word_token=nlk.word_tokenize(line)
    word_token=[word for word in word_token if not word in stopword]
    data_stop.append(" ".join(word_token))
data_stop
#Wordcloud
!pip install wordcloud
import string
import matplotlib as mpl
import matplotlib.pyplot as plt
%matplotlib inline
from subprocess import check_output
from wordcloud import WordCloud, STOPWORDS
#Replace data_stop into string type
s=str(data_stop)
word=re.sub(r"'\s'","").s)
word
#Applying the Wordcloud
wordcloud=WordCloud(collocations=False.background_color='white'.
stopwords=stopword. max_words=25.
max_font_size=95.random_state=1.normalize_plurals=False).generate(word)
print(wordcloud)
fig=plt.figure(1)
plt.imshow(wordcloud. interpolation='bilinear')
plt.axis('off')
plt.show()
#TF-IDF
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
vectorizer=CountVectorizer(min_df=0.005)
x=vectorizer.fit_transform(data_stop)
DTM=pd.DataFrame(x.toarray().columns=vectorizer.get_feature_names())
DTM['kalimat tweet']=data_stop
DTM['sum_features']=DTM.sum(axis=1)
DTM=DTM.loc[DTM['sum_features'] !=0].drop(['sum_features'].axis=1)
DTM.to_csv('DTM_testing.csv')
```

Lampiran 5. List Stopword

<i>Stopword</i>					
ada	bagaimana	berapakah	biasa	dia	dimulailah
adalah	bagaimana	berapalah	biasanya	diakhiri	dimulainya
adanya	bagaimanakah	berapapun	bila	diakhirinya	dimungkinkan
adapun	bagaimanapun	berarti	bilakah	dialah	dini
agak	bagi	berawal	bisa	diantara	dipastikan
agaknya	bagian	berbagai	bisakah	diantaranya	diperbuat
agar	bahkan	berdatangan	boleh	diberi	diperbuatnya
aja	bahwa	beri	bolehkah	diberikan	dipergunakan
akan	bahwasanya	berikan	bolehlah	diberikannya	diperkirakan
akankah	baik	berikut	bu	dibuat	diperlihatkan
akhir	bakal	berikutnya	buat	dibuatnya	diperlukan
akhiri	bakalan	berjumlah	buk	didapat	diperlukannya
akhirnya	balik	berkali-kali	bukan	didatangkan	dipersoalkan
aku	banget	berkata	bukankah	digunakan	dipertanyakan
akulah	banyak	berkehendak	bukanlah	diibaratkan	dipunyai
amat	bapak	berkeinginan	bukannya	diibaratkannya	diri
amatlah	baru	berkenaan	bulan	diingat	dirinya
amp	bawah	berlainan	bung	diingatkan	disampaikan
anda	beberapa	berlalu	cara	diinginkan	disebut
andalah	begini	berlangsung	caranya	dijawab	disebutkan
antar	beginian	berlebihan	cukup	dijelaskan	disebutkannya
antara	beginikah	bermacam	cukupkah	dijelaskannya	disini
antaranya	beginilah	bermacam-macam	cukuplah	dikarenakan	disinilah
apa	begini	bermaksud	cuma	dikatakan	ditambahkan
apaan	beginitukah	bermula	cuman	dikatakannya	ditandaskan
apabila	beginitulah	bersama	dahulu	dikerjakan	ditanya
apakah	beginipun	bersama-sama	dalam	diketahui	ditanyai
apalagi	bekerja	bersiap	dan	diketuainya	ditanyakan
apatah	belakang	bersiap-siap	dapat	dikira	ditegaskan
artinya	belakangan	bertanya	dari	dilakukan	ditujukan
asal	belum	bertanya-tanya	daripada	dilalui	ditunjuk
asalkan	belumlah	berturut	datang	dilihat	ditunjuki
asuransi	benar	berturut-turut	deh	dimaksud	ditunjukkan
atas	benarkah	bertutur	dekat	dimaksudkan	ditunjukkannya
atau	benarlah	berujar	demi	dimaksudkannya	ditunjuknya
ataukah	berada	berupa	demikian	dimaksudnya	dituturkan
ataupun	berakhir	besar	demikianlah	diminta	dituturkannya
awal	berakhirilah	betul	dengan	dimintai	diucapkan
awalnya	berakhirnya	betulkah	depan	dimisalkan	diucapkannya
bagai	berapa	bgt	di	dimulai	Diungkapkan

Lampiran 5. (Lanjutan)

<i>Stopword</i>					
dong	inginkan	kapan	kita	melalui	menyangkut
dua	ini	kapankah	kitalah	melihat	menyatakan
dulu	inikah	kapanpun	klo	melihatnya	menyebutkan
emang	inilah	karena	kok	memang	menyeluruh
empat	itu	karenanya	kurang	memberi	menyiapkan
enggak	itukah	kasus	lagi	membuat	merasa
enggaknya	itulah	kata	lagian	memihak	mereka
entah	jadi	katakan	lah	meminta	merekalah
entahlah	jadilah	katakanlah	lain	memisalkan	merupakan
ga	jadinya	katanya	lainnya	memperbuat	meski
gak	jangan	ke	lalu	mempunyai	meskipun
gitu	janganakan	keadaan	lama	memulai	meyakini
gue	janganlah	kebetulan	lamanya	menaiki	minta
guna	jauh	kecil	lanjut	menanti	mirip
gunakan	jawab	kedua	lanjutnya	menanti-nanti	misal
gw	jawaban	keduanya	lebih	menanya	misalkan
hal	jawabnya	keinginan	lewat	menanyai	misalnya
hampir	jelas	kelamaan	lima	mendapat	mula
hanya	jelaskan	kelihatan	luar	mendatang	mulai
hanyalah	jelastah	kelihatannya	macam	mendatangi	mulailah
hari	jelasnya	kelima	maka	mengakhiri	mulanya
harus	jika	keluar	makanya	mengapa	mungkin
haruslah	jikalau	kembali	makin	mengenai	mungkinkah
harusnya	juga	kemudian	malah	mengetahui	nah
hendak	juga	kemungkinan	malahan	menggunakan	naik
hendaklah	jumlah	kemungkinannya	mampu	menghendaki	namun
hendaknya	jumlahnya	kenapa	mampukah	mengibaratkannya	nanti
hingga	justru	kepada	mana	mengingat	nantinya
ia	kak	kepadanya	manakala	menginginkan	nih
ialah	kala	kesampaian	manalagi	mengira	nya
ibarat	kalau	keseluruhan	masa	mengucapkan	nya
ibaratkan	kalaulah	keseluruhannya	masalah	mengucapkannya	nyaris
ibaratnya	kalaupun	keterlalu	masalahnya	menjadi	nyatanya
ibu	kalian	ketika	masih	menjawab	oleh
ikut	kalo	khususnya	masihkah	menuju	olehnya
in	kami	kini	masing	menunjuk	pada
ingat	kamilah	kinilah	masing-masing	menunjuki	padahal
ingat-ingat	kamu	kira	mau	menunjuknya	padanya
ingin	kamulah	kira-kira	maupun	menurut	pak
inginkah	kan	kiranya	melainkan	menyampaikan	Pake
tiga	tinggi	toh	tp	trus	tunjuk

Lampiran 5. (Lanjutan)

<i>Stopword</i>					
paling	sana	segalanya	semasih	setengah	tapi
panjang	sangat	segera	semata	seterusnya	tau
pantas	sangatlah	seharusnya	semata-mata	setiap	tegas
para	satu	sehingga	semaunya	setiba	tegasnya
pasti	saya	seingat	sementara	setibanya	telah
pastilah	sayalah	sejak	semisal	setidaknya	tempat
penting	se	sejauh	semisalnya	setidak-tidaknya	tengah
pentingnya	sebab	sejenak	sempat	setinggi	tentang
per	sebabnya	sejumlah	semua	seusai	tentu
percuma	sebagai	sekadar	semuanya	sewaktu	tentulah
perlu	sebagaimana	sekadarnya	semula	si	tentunya
perlukah	sebagainya	sekali	sendiri	siap	tepat
perlunya	sebagian	sekalian	sendirian	siapa	terakhir
pernah	sebaik	sekaligus	sendirinya	siapakah	terasa
persoalan	sebaik-baiknya	sekali-kali	seolah	siapapun	terbanyak
pertama	sebaiknya	sekalipun	seolah-olah	sih	terdahulu
pertama-tama	sebaliknya	sekarang	seorang	sini	terdapat
pertanyaan	sebanyak	sekarang	sepanjang	sinilah	terdiri
pertanyakan	sebegini	sekecil	sepantasnya	soal	terhadap
pihak	sebegini	seketika	sepantasnyalah	soalnya	terhadapnya
pihaknya	sebelum	sekiranya	seperlunya	suatu	teringat
pukul	sebelumnya	sekitar	seperti	sudah	teringat-ingat
pula	sebenarnya	sekitarnya	sepertinya	sudahkah	terjadi
pun	seberapa	sekurang-kurangnya	sepihak	sudahlah	terjadilah
punya	sebesar	sekurangnya	sering	supaya	terjadinya
rasa	sebetulnya	sela	seringnya	tadi	terkira
rasanya	sebisanya	selain	serta	tadinya	terlalu
rata	sebuah	selaku	serupa	tahu	terlebih
rupanya	sebut	selalu	sesaat	tahun	terlihat
saat	sebutlah	selama	sesama	tak	termasuk
saatnya	sebutnya	selama-lamanya	sesampai	tambah	ternyata
saja	secara	selamanya	sesegera	tambahnya	tersampaikan
sajalah	secukupnya	selanjutnya	sesekali	tampak	tersebut
saling	sedang	seluruh	seseorang	tampaknya	tersebutlah
sama	sedangkan	seluruhnya	sesuatu	tandas	tertentu
sama-sama	sedemikian	semacam	sesuatunya	tandasnya	tertuju
sambil	sedikit	semakin	sesudah	tanpa	terus
sampai	sedikitnya	semampu	sesudahnya	tanya	terutama
sampaikan	seenaknya	semampunya	setelah	tanyakan	tetap
sampai-sampai	segala	semasa	setempat	tanyanya	Tetapi
tiap	tiba	tiba-tiba	tidak	tidaklah	tidaklah

Lampiran 5. (Lanjutan)

<i>Stopword</i>				
turut	ujar	untuk	wahai	ya
tutur	ujarnya	usah	waktu	yaitu
tuturnya	umum	usai	waktunya	yakin
ucap	umumnya	utk	walau	yakni
ucapnya	ungkap	waduh	walaupun	yang
udah	ungkapnya	wah	wong	yg

Lampiran 6. List Kata Yang Digunakan Pada FP-Growth

abis	bisnis	full	jiwa	lengkap	ojk	rs
acara	bonus	fwd	jiwasraya	life	ongkir	rugi
administrasi	bpjs	gabung	jual	lihat	online	rumah
agen	bri	gagal	kadang	lindung	operasi	rupa
aia	bukti	gaji	kait	link	packing	rusak
akibat	bumn	gampang	kakak	lumayan	paham	sadar
akses	bunga	ganti	kantor	lupa	pajak	saham
alam	butuh	gantung	kartu	maaf	pakai	sakit
alhamdulillah	cair	gede	kasih	mahal	paket	salah
allianz	care	gimana	kaya	main	pandemi	sales
aman	cari	gratis	kecuali	makan	pasar	sampe
ambil	cek	habis	kelas	maksa	pasien	saran
anggota	celaka	hadap	kelola	maksud	pensiun	sudah
aplikasi	cepat	hak	keluarga	mandiri	penuh	sedia
apps	chat	harga	kemarin	manfaat	percaya	sehat
april	cicil	hasil	kembang	masuk	perintah	selamat
arti	ciri	hati	kena	masyarakat	pesawat	selesai
asuransi	claim	hidup	kenal	mati	pikir	seller
atur	coba	hilang	kendara	mending	pilih	sesuai
awat	cover	hospital	keren	milik	plus	sinar
axa	covid	hp	kerja	mobil	pokok	sistem
baca	daftar	hubung	ketemu	mohon	polis	sobat
bagus	dana	hukum	khusus	motor	potensi	sosial
bahas	dapet	ilang	kirim	muda	potong	suami
bakar	darurat	inap	klaim	mudah	premi	suka
bank	dasar	indonesia	klik	mudik	pribadi	surat
bantu	data	industri	kondisi	murah	produk	suruh
barang	discover	info	konsep	nabung	program	susah
bawa	didik	informasi	korban	nama	promo	swasta
bayar	digital	insurance	kredit	nanya	proses	syarat
bca	diskon	investasi	kritis	nasabah	protection	syariah
bebas	dokter	istri	kuat	nawarin	proteksi	tabung
beda	download	jaga	kuliah	negara	prudential	tagih
beli	dpt	jalan	kurir	negeri	pt	takut
benefit	duit	jam	lahir	ngurus	putus	tangan
bener	dunia	jamin	laku	nilai	rawat	tanggung
biaya	ekspedisi	jangka	langsung	nunggu	rencana	tani
bikin	enak	janji	lapor	nyaman	resiko	tarik
bilang	fasilitas	jasa	layan	nyata	ribet	tawar
bingung	finansial	jenis	lebaran	obat	risiko	Telpon

Lampiran 6. (Lanjutan)

temen	tenang	terima	tetep	tiket	tindak
tinggal	tingkat	tips	tipu	toko	tokped
tolak	tolong	tua	tugas	tulis	tumbuh
tunggu	tutup	uang	ubah	unit	untung
urus	usaha	usia	utama	vaksin	wajib
waris	yuk				

BIODATA PENULIS



Merdeka,

Penulis dilahirkan di Jakarta, 04 Juli 2000, merupakan anak kedua dari 3 bersaudara, Penulis telah menempuh pendidikan formal yaitu di TK Qatrunada, SDIT Meranti, SMPN 10 Jakarta dan SMAN 68 Jakarta, Setelah lulus dari SMAN 68 pada tahun 2018, Penulis mengikuti SBMPTN dan diterima di Departemen Aktuaria - ITS pada tahun 2018 dan terdaftar dengan NRP 0631184000042,

Di Departemen Aktuaria Penulis sempat aktif di beberapa kegiatan Himpunan Aktuaria (HIMASAKTA) sebagai kepala departemen dalam negeri dan juga mengikuti program pembelajaran yang diadakan ITS seperti *exchange program* dengan *Asia University* dan short course dari Gojek Kampus