SKRIPSI

ANALISIS DATA MINING PENYAKIT DEMAM BERDARAH MENGGUNAKAN TEKNIK KLASTERING DAN ASOSIASI

Disusun dan diajukan oleh:

MUHAMMAD WAHYUDI R SUMARA
D121 17 1502

PROGRAM STUDI SARJANA TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
GOWA
2024

Optimized using
trial version
www.balesio.com

LEMBAR PENGESAHAN SKRIPSI

ANALISIS DATA MINING PENYAKIT DEMAM BERDARAH MENGGUNAKAN TEKNIK KLASTERING DAN ASOSIASI

Disusun dan diajukan oleh

Muhammad Wahyudi R Sumara
D121171502

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka
Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika
Fakultas Teknik Universitas Hasanuddin
Pada tanggal 08 Maret 2024
dan dinyatakan telah memenuhi syarat kelulusan

Menyetujui,

Pembimbing Utama,	Pembimbing Pendamping,
Prof.Dr. Ir. Indrabayu, ST., MT., M.Bus.Sys., IPM, ASEAN. Eng.	Iqra Aswad, ST.,MT
NIP 197507162002121004	NIP 19830510201404001

Ketua Program Studi,

Dr. Ir. Indrabayu, ST., MT., M.Bus.Sys., IPM, ASEAN. Eng.
NIP 197507162002121004

Optimized using
trial version
www.balesio.com

PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini ;

Nama	:	Muhammad Wahyudi R Sumara
NIM	:	D121171502
Program Studi	:	Teknik Informatika
Jenjang	:	S1

Menyatakan dengan ini bahwa karya tulisan saya berjudul

Analisis Data Mining Penyakit Demam Berdarah Menggunakan Teknik Klastering dan Asosiasi

Adalah karya tulisan saya sendiri dan bukan merupakan pengambilan alihan tulisan orang lain dan bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Semua informasi yang ditulis dalam skripsi yang berasal dari penulis lain telah diberi penghargaan, yakni dengan mengutip sumber dan tahun penerbitannya. Oleh karena itu semua tulisan dalam skripsi ini sepenuhnya menjadi tanggung jawab penulis. Apabila ada pihak manapun yang merasa ada kesamaan judul dan atau hasil temuan dalam skripsi ini, maka penulis siap untuk diklarifikasi dan mempertanggungjawabkan segala resiko.

Segala data dan informasi yang diperoleh selama proses pembuatan skripsi, yang akan dipublikasi oleh Penulis di masa depan harus mendapat persetujuan dari Dosen Pembimbing.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan isi skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Gowa, 08 Maret 2024

Yang Menyatakan

Muhammad Wahyudi R Sumara

Optimized using
trial version
www.balesio.com

ABSTRAK

MUHAMMAD WAHYUDI R SUMARA. Analisis Data Mining Penyakit Demam Berdarah Menggunakan Teknik Klastering dan Asosiasi (dibimbing oleh Indrabayu dan Iqra Aswad)

Pemanfaatan *data mining* dengan teknik *clustering* dan asosiasi menjadi solusi dalam mengelompokkan data dan mencari karakteristik berdasarkan wilayah atau alamat hasil *cluster* data penyakit demam berdarah di Kota Makassar, serta pada pemanfaatan asosiasi menjadi solusi dalam membantu analisis pola-pola variabel terhadap penyakit demam berdarah untuk mengurangi,menanggulangi kenaikan penyakit demam berdarah. Penelitian ini mengimplementasikan dua metode yaitu metode asosiasi dengan algoritma FP-Growth dan metode *clustering* dengan algoritma *K-Means*. Set data yang digunakan adalah data pasien Demam Berdarah (DBD) pada Rumah Sakit Bhayangkara dari tahun 2019 hingga tahun 2022. Penerapan algoritma FP-Growth dilakukan dengan beberapa tahapan yaitu dalam peng-input-an dataset, menentukan *minimum support* dan *minimum confidence*, tahap Pembangunan *FP-Tree*, pembangkitan *conditional pattern base*, pembangkitan *conditional FP-Tree*, pencarian *frequent itemset*, aturan asosiasi dan perhitungan nilai *Lift ratio*. Digunakan nilai *minimum support* sebesar 0,09 dan nilai *minimum confidence* sebesar 0,8 dimana menghasilkan aturan asosiasi untuk beberapa Kecamatan yang ada pada dataset. Dengan variabel yang berkaitan diantaranya, Pasien tidak pernah vaksin, Pasien mengalami gejala DBD, Pasien pernah Opname, Pasien pernah mengalami gejala DBD, Keluarga pasien tidak pernah DBD, Kondisi IMT pasien normal, Jendela/Ventilasi berkawat, dan jenis kelamin laki-laki. Penggunaan algoritma *k-means clustering* dalam mengelompokkan data DBD dari 23 Kecamatan di Kota Makassar dan Kabupaten Gowa dengan 22 parameter menghasilkan sebanyak 6 *clustering* dengan parameter yang berbeda-beda diantaranya. Adapun nilai *Silhouette Coefficient*, yaitu sebesar 0,5228, 0,6231, 0,6595, 0,3207, 0,5874, dan 0,7606. Dan nilai SSE nya yaitu sebesar 9,1974, 12,049, 13,9711, 21,2452, 7,4372, dan 8,5803. Dimana tiap kecamatan dapat termasuk dalam 1 *cluster* atau lebih yang menunjukkan *cluster* yang paling berpengaruh terhadap DBD di kecamatan tersebut.

Kata Kunci: *DBD, FP-Growth, K-Means, data mining*

Optimized using
trial version
www.balesio.com

ABSTRACT

MUHAMMAD WAHYUDI R SUMARA. *Data Mining Analysis of Dengue Fever Using Clustering and Association Techniques* (supervised by Indrabayu dan Iqra Aswad)

The utilization of data mining with clustering and association techniques provides a solution for categorizing data and identifying characteristics based on regions or addresses in Makassar. This approach is applied to cluster data related to dengue fever, and the association technique is employed to aid in the analysis of patterns among variables associated with dengue fever, aiming to reduce and counteract the rise of the disease. This research implements two methods: association method using the FP-Growth algorithm and clustering method using the K-Means algorithm. The dataset used consists of Dengue Fever (DF) patient data from Bhayangkara Hospital spanning from 2019 to 2022. The FP-Growth algorithm is implemented in several stages, including dataset input, determination of minimum support and minimum confidence, FP-Tree construction, generation of conditional pattern base, conditional FP-Tree generation, frequent itemset search, association rule generation, and Lift ratio calculation. The minimum support value is set at 0,09, and the minimum confidence value is set at 0,8, resulting in association rules for several districts in the dataset. The associated variables include patients who have never been vaccinated, patients experiencing DF symptoms, patients who have been hospitalized, patients who have experienced DF symptoms, families with no history of DF, patients with normal BMI conditions, wired windows/ventilation, and male gender. The K-Means clustering algorithm is used to group DF data from 23 districts in Makassar and Gowa Regency with 22 parameters. The clustering process involves 6 clusters with different parameters. The Silhouette Coefficient values are 0,5228, 0,6231, 0,6595, 0,3207, 0,5874, and 0,7606, and the Sum of Squared Errors (SSE) values are 9,1974, 12,049, 13,9711, 21,2452, 7,4372, and 8,5803. Each district can be included in one or more clusters, indicating the clusters that most significantly influence DF in that district.

Keywords: *Dengue fever, FP-Growth, K-Means, data mining*

DAFTAR ISI

Optimized using
trial version
www.balesio.com

Optimized using
trial version
www.balesio.com

Optimized using
trial version
www.balesio.com

DAFTAR GAMBAR

Optimized using
trial version
www.balesio.com

Optimized using
trial version
www.balesio.com

Optimized using
trial version
www.balesio.com

DAFTAR TABEL

Optimized using
trial version
www.balesio.com

DAFTAR LAMPIRAN

Optimized using
trial version
www.balesio.com

DAFTAR SINGKATAN DAN ARTI SIMBOL

Lambang/Singkatan	Arti dan Keterangan
DBD	Demam Berdarah
CSV	*Comma separated value*
HI	*House Index*
BI	*Breteau Index*
CI	*Container Index*
ABJ	Angka Bebas Jentik
IR	*Incidence Rate*
CFR	*Case Fatality Rate*
CFD	*Cumulative Frequency Distribution*
KDD	*Knowledge Discovery in Database*
Minsup	*Minimum support*
Minconf	*Minimum confident*
TID	*Transaction ID*
Prefix path	Lintasan awal
Suffic pattern	Pola akhiran
SSE	*Sum of squared error*
LAB	*Laboratory*
CYD	*Chimeric yellow fever virus-vi dengue virus- tertravalent dengue vaccine*
IMT	Indeks Massa Tubuh
PNS	Pegawai negeri sipil

Optimized using
trial version
www.balesio.com

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT yang senantiasa memberikan rahmat serta hidayahnya serta diberi kelancaran dan kesehatan sehingga penulis dapat menyelesaikan karya tulisnya yang berjudul **"Analisis Data Mining Penyakit Demam Berdarah Menggunakan Teknik Klastering dan Asosiasi"** guna memenuhi salah satu persyaratan dalam menyelesaikan jenjang Strata-1 di Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin.

Penulis menyadari sepenuhnya bahwa skripsi ini masih jauh dari kesempurnaan karena menyadari segala keterbatasan yang ada. Dalam penulisan skripsi ini penulis menghadapi berbagai kendala dan masalah, namun karena usaha yang maksimal dan kemampuan yang Tuhan berikan kepada penulis serta bantuan dan dukungan dari berbagai pihak, maka penulisan skripsi ini dapat selesai. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Allah SWT ,Tuhan pencipta alam semesta yang senantiasa memberikan rahmat serta hidayahnya kepada penulis.
2. Kedua orang tua penulis, Bapak Ir. Muh Rusli Sumara M. Ikom dan Ibu Erny Sriwahyuni yang selalu memberikan kasih sayang, nasehat, motivasi, dukungan, dan doa kepada penulis.
3. Bapak Prof. Dr. Ir. Indrabayu, ST., MT., M.Bus.Sys., IPM, ASEAN. Eng., selaku pembimbing utama dan Bapak A.Iqra Aswad, S.T., M.T. selaku pembimbing pendamping yang senantiasa menyediakan waktu, tenaga, pikiran, dan perhatian yang luar biasa dalam mengarahkan penulis dalam penyusunan tugas akhir ini.
4. Bapak Robert, Bapak Zainuddin dan Ibu Yuanita serta segenap staf Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin yang telah membantu kelancaran penyelesaian tugas akhir penulis.
5. Segenap keluarga AIMP *Research Group* Universitas Hasanuddin yang telah memberikan begitu banyak bantuan selama penelitian, pengambilan data dan diskusi *progress* penyusunan tugas akhir serta memberikan semangat di masa-masa sulit.
6. Muhammad Fadhil Bahrunnida, Alfarabi Alif Putra, Muhammad Irzam Kasyafillah, Moch Wahyu Faisal, Taslinda, Ilmi, Fitri, Devy, Irma, Jumriani, Herlina yang telah membantu penulis sejak awal perkuliahan dan selalu membantu dalam penyelesaian tugas akhir.
7. Seluruh pihak yang tidak sempat disebutkan satu persatu yang telah banyak meluangkan tenaga, waktu, dan pikiran selama penyusunan tugas akhir ini.

Optimized using
trial version
www.balesio.com

BAB I
PENDAHULUAN

1.1 Latar Belakang

Penyakit Demam Berdarah (DBD) atau biasa disebut dengan *Dengue Fever* yang dimana penyakit infeksi oleh *dengue* yang ditularkan melalui gigitan nyamuk *Aedes aegypti*, dengan ciri demam tinggi mendadak. Faktor utama terjadinya penyakit DBD di Indonesia adalah nyamuk *Aedes aegypti*. Sehingga saat ini masih merupakan masalah Kesehatan Masyarakat yang belum dapat diatasi sepenuhnya oleh karena sulitnya memutuskan mata rantai penularannya. Tempat yang disukai sebagai tempat perindukannya adalah tempat-tempat kotor, genangan air yang terdapat di dalam wadah contohnya misal bak mandi, gentong, dan ember (Dania, 2016).

Insiden demam berdarah sendiri telah meningkat secara drastis di seluruh dunia dalam beberapa dekade terakhir. Sebagian besar kasus tidak menunjukkan gejala atau ringan dan dikelola sendiri, dan jumlah sebenarnya dari kasus demam berdarah tidak dilaporkan di beberapa wilayah. Banyak kasus juga salah didiagnosis sebagai penyakit demam lainnya. Prediksi menunjukkan 390 juta yang ter-Infeksi virus *dengue* per tahunnya (95% interval kredibel 284–528 juta). Meskipun resiko terjadinya infeksi terdapat pada 129 Negara, tetapi 70% dari data ada pada di benua Asia. Jumlah kasus demam berdarah yang dilaporkan ke WHO meningkat lebih dari 8 kali lipat selama 2 dekade terakhir, dari 505.430 kasus pada tahun 2000, menjadi lebih dari 2,4 juta pada tahun 2010, dan 5,2 juta pada tahun 2019. Angka Kematian yang dilaporkan antara tahun 2000 dan 2015 meningkat dari 960 menjadi 4032 (WHO,2021).

Kota Makassar, Sulawesi Selatan termasuk menjadi daerah dengan kasus Demam Berdarah *Dengue* (DBD) tertinggi. Menurut Kepala Seksi Pencegahan dan Pengendalian Penyakit menular di Dinas Kesehatan Makassar telah melaporkan ngkatan kasus yang terjadi pada tahun 2021 terdapat 583 kasus DBD ʒ tahun dengan 1 kematian. Angka ini meningkat tajam dibandingkan 20 dimana tercatat ada 175 kasus dengan 0 kasus kematian. Angka kasus

Optimized using
trial version
www.balesio.com

di tahun 2021 juga menjadi yang tertinggi dalam sejak 7 tahun terakhir. Dimana pada tahun 2015 tercatat 142 kasus dengan 5 kasus kematian, Tahun 2016 tercatat ada 250 kasus dengan 2 kasus kematian, di tahun 2017 tercatat 135 kasus dengan 1 kasus kematian, dan di tahun 2019 tercatat 268 kasus tanpa adanya kasus kematian, lalu yang terakhir pada tahun 2020 tercatat ada 175 kasus tanpa kasus kematian. Dan di bulan Mei 2021 ada 227 kasus dengan 3 angka kematian yang diakibatkan penyakit DBD (Dinas Kesehatan, Jumlah Kasus DBD, 2021).

Berdasarkan uraian data yang diatas, untuk dapat mengetahui tingkat keparahan DBD di wilayah tertentu yang grafiknya naik berdasarkan data-data yang ada pada Rumah Sakit Bhayangkara. Maka tujuan dari klasterisasi dan asosiasi adalah dimana data dikelompokkan dengan karakteristik yang sama ke suatu *cluster* yang sama serta data dengan karakteristik yang berbeda ke *cluster* yang lain, serta asosiasi menemukan pola keterkaitan antara variabel penyakit demam berdarah.

Oleh sebab itu, dengan meningkatnya kasus DBD di Kota Makassar dan permasalahan yang sering terulang terjadi pada masyarakat Makassar, maka Pemanfaatan *data mining* dengan teknik *clustering* dan asosiasi dapat menjadi solusi dalam mengelompokkan data dan mencari karakteristik berdasarkan wilayah atau alamat hasil *cluster* data penyakit demam berdarah di kota Makassar, serta pada pemanfaatan asosiasi dapat menjadi solusi dalam membantu analisis pola-pola variabel terhadap penyakit demam berdarah untuk mengurangi, menanggulangi kenaikan penyakit demam berdarah.

1.2 Rumusan Masalah

Berdasarkan latar belakang, maka rumusan masalah pada tugas akhir ini adalah sebagai berikut:

1. Bagaimana mengetahui asosiasi keterkaitan antar variabel dan karakteristik pada hasil asosiasi dan hasil *clustering* serta mengetahui pengelompokkan akit demam berdarah?
   imana mengelompokkan data pasien DBD per-kecamatan dengan *k-*
   s clustering dalam *data mining?*

Optimized using
trial version
www.balesio.com

1.3.Tujuan Penelitian

Tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut:

1. Agar mengetahui keterkaitan antar variabel dan juga karakteristik pada hasil asosiasi dan serta mengetahui hasil *clustering* terhadap penyakit DBD
2. Untuk menerapkan *data mining* dengan teknik *clustering* untuk mengelompokkan data pasien DBD per-kecamatan berdasarkan kesamaan/kedekatan kasusnya.

1.4.Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Memberikan informasi dasar untuk mengambil tindakan dan juga pencegahan dalam mengurangi kasus Demam Berdarah di Kota Makassar.
2. Memberikan solusi dalam mengelompokkan daerah penyebaran demam berdarah dan membantu dalam melakukan analisis keterkaitannya pada demam berdarah.

1.5. Ruang Lingkup

Ruang lingkup dari penelitian ini diuraikan sebagai berikut

1. Pengambilan data dilakukan di Rumah Sakit Bhayangkara, yang berupa data rekam medis pasien dari tahun 2019 hingga tahun 2021, yang kemudian akan dikirimkan kuesioner/data primer ke responden untuk diisi.
2. Algoritma yang digunakan untuk *Clustering K-Means* dan *Association Fp-Growth*.
3. Pengambilan data yang didapatkan dari responden Rumah Sakit Bhayangkara hanya dilakukan di Kota Makassar dan Kabupaten Gowa.

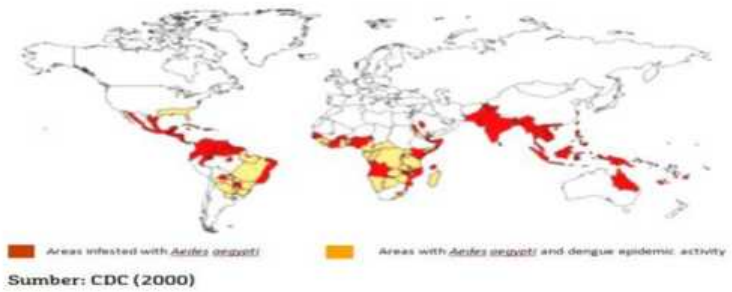
Optimized using
trial version
www.balesio.com

BAB II
TINJAUAN PUSTAKA

2.1 Demam Berdarah

Demam berdarah *dengue* (DBD) merupakan penyakit yang banyak ditemukan di sebagian besar wilayah tropis dan subtropis, terutama di asia tenggara, Amerika tengah, dan Karibia. Terjadinya wabah secara simultan di 3 benua menunjukkan bahwa virus melalui vektor nyamuk yang mempengaruhi distribusi penyakit demam dengue di seluruh dunia dalam kurun waktu 200 tahun. Selama ini, demam berdarah dianggap sebagai penyakit yang jinak atau *non-fatal* bagi masyarakat yang beriklim tropis (Arsin, 2013).

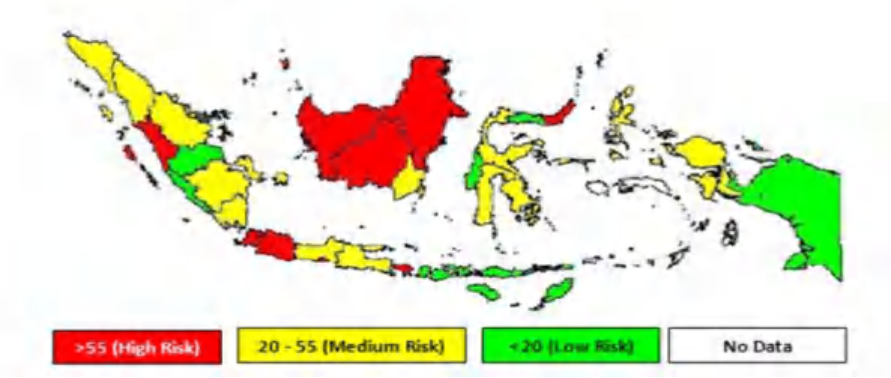
Umumnya, epidemi terjadi pada interval waktu yang panjang yaitu 10 minggu 40 tahun. Hal ini memicu penyebaran *virus dengue* melalui vektor nyamuk *Aedes aegypti* melalui sektor transportasi kapal yang transit di beberapa belahan dunia. Untuk itu, pada Gambar 1 menunjukkan distribusi DBD yaitu daerah infeksi virus *dengue* dan area epidemi (Arsin, 2013).



Sumber : Ditjen PP & PL Kemenkes RI
Gambar 1 Distribusi DBD di dunia (Arsin, 2013)

Berdasarkan Gambar 1 menunjukkan infeksi virus *dengue* telah ada di Indonesia sejak tahun 1779, seperti yang dilaporkan oleh David Bylon seorang dokter berkebangsaan Belanda. Saat itu infeksi yang diberikan oleh virus *dengue* menimbulkan demam tinggi yang terjadi selama 2-7 hari, disertai dengan nyeri sendi, nyeri otot, dan nyeri kepala.

Optimized using
trial version
www.balesio.com



Gambar 2 Kondisi demam berdarah di Indonesia (Arsin, 2013)

Pada Gambar 2 diatas menunjukkan peta kondisi demam berdarah di indonesia pada tahun 1968 penyakit Demam Berdarah *dengue* dilaporkan di Surabaya dan Jakarta sebanyak 58 kasus, dengan jumlah kematian yang sangat tinggi 24 orang (*Case fatality rate* 41,3%). Epidemi penyakit DBD di luar Jawa pertama kali dilaporkan di Sumatera Barat dan Lampung tahun 1972. Sejak itu, penyakit ini semakin menyebar luas ke berbagai wilayah di Indonesia. Penularan DBD hanya dapat terjadi melalui gigitan nyamuk yang di dalam tubuhnya mengandung virus *Dengue* (Arsin, 2013).

Masa inkubasi virus *dengue* dalam manusia (inkubasi intrinsik) berkisar antara 3 sampai 14 hari sebelum gejala muncul, gejala klinis rata-rata muncul pada hari keempat sampai hari ketujuh, sedangkan masa inkubasi ekstrinsik (di dalam tubuh nyamuk) berlangsung sekitar 8-10 hari. Kelembapan yang tinggi dengan suhu berkisar antara 28-32˚C membantu nyamuk *Aedes aegypti* bertahan hidup untuk jangka waktu yang lama. Pola penyakit di Indonesia sangat berbeda antara satu wilayah dengan wilayah lainnya. Tingginya angka kejadian DBD juga dapat dipengaruhi oleh kepadatan penduduk meningkat. Semakin banyak manusia maka peluang tergigitnya nyamuk *Aedes aegypti* juga akan semakin lebih tinggi (Trovancia et al., 2016).

Indikator kepadatan vektor DBD antara lain : *House Index* (HI), *Breteau* I), *Container Index* (CI) dan Angka Bebas Jentik (ABJ) merupakan ı di mana dapat ditentukan apakah daerah tersebut memiliki ungan setiap tahun akan terjadi kejadian DBD atau tidak. Sampai dengan

Optimized using
trial version
www.balesio.com

saat ini jumlah kabupaten/kota terjangkit DBD di Indonesia sebanyak 477 Kabupaten/Kota atau sebesar 92,8% dari seluruh kabupaten/kota yang ada di Indonesia. Jumlah ini cenderung meningkat sejak tahun 2010 sampai 2019 (Afsahyana et al., 2022)

Tahun 2017 dilaporkan Kasus DBD di Provinsi Sulawesi Selatan sebanyak 1,724 kasus, dengan *Incidens Rate* (IR) per 100,000 penduduk sebesar 19,84% dan *Case Fatality Rate* (CFR) sebesar 0,58% (Kemenkes RI, 2018). Kota Makassar merupakan salah satu daerah endemis DBD dengan IR sebesar 28 per 100,000 penduduk. CFD sebesar 0,50%. Tahun 2017, IR sebesar 17 per 100,000 penduduk dengan CFR sebesar 0,4% pada tahun 2018 dan IR 18 per 100,000 penduduk dengan CFR 0,0 pada Tahun 2019 (Afsahyana et al., 2022)

2.2 *Aedes Aegypti*

Aedes aegypti adalah jenis nyamuk penyebab penyakit DBD sebagai pembawa utama (*primary vector*) virus dengue (WHO, 2009). Nyamuk jenis *Aedes aegypti* yang sudah menghisap virus dengue sebagai penular penyakit demam berdarah. Adanya penularan itu karena setiap nyamuk itu menggigit, nyamuk tersebut menghisap darah yang akan menghasilkan air liur dengan bantuan alat tusuknya supaya darahnya yang telah dihisap tidak dapat membeku. Nyamuk *Aedes aegypti* mempunyai persebaran dengue yang sangat luas hampir semua mencakup daerah yang tropis maupun subtropis di seluruh dunia. Hal ini membawa siklus persebarannya baik di desa, kota maupun di sekitar daerah penduduk yang padat. Beberapa penularan penyakit DBD yang disebabkan nyamuk *Aedes aegypti* yaitu mulai dari perilaku menggigit, perilaku istirahat dan juga jangkauan terbang untuk disebarkannya virus dengue (Susanti & Suharyo, 2017).

Nyamuk *Aedes aegypti* siklus hidupnya mempunyai 4 fase yaitu dari mulai telur, jentik, pupa, sampai menjadi nyamuk dewasa.nyamuk jenis ini mempunyai siklus hidup sempurna. Spesies ini meletakkan telurnya pada kondisi permukaan air
sih secara individual. Telur yang memiliki bentuk elips warnanya hitam
terpisah satu dengan yang lain. Telurnya dapat menetas dalam waktu 1-2
ıdian akan berubah jentik (Susanti & Suharyo, 2017).

Optimized using
trial version
www.balesio.com

Terdiri dari 4 tahap didalam perkembangannya jentik yang dikenal sebagai instar. Perkembangan instar 1 ke instar 4 membutuhkan waktu kira-kira 5 hari. Selanjutnya untuk sampai instar ke 4, larva ini berubah menjadi pupa yang dimana jentik tersebut telah memasuki masa dorman. Pupa dapat bertahan selama 2 hari sebelum nyamuk dewasa keluar dari pupa. Perkembangan mulai dari telur hingga menjadi nyamuk dewasa membutuhkan waktu selama 8 hingga 10 hari, namun juga bisa lebih lama jika kondisi lingkungan yang tidak mendukung (Susanti & Suharyo, 2017).

Berikut merupakan morfologi nyamuk *Aedes aegypti* yaitu yang pertama telur *Aedes aegypti* setiap kali bertelur nyamuk betina dapat mengeluarkan kurang lebih 100 butir telur dengan berukuran 0,7 mm per butir. Ketika pertama kali dikeluarkan oleh induk nyamuk, telur *Aedes aegypti* berwarna putih dan juga lunak. Kemudian telur tersebut menjadi warna hitam dan keras. Telur tersebut dengan bentuk ovoid meruncing dan sering diletakkan satu per satu. Induk nyamuk biasanya meletakkan telurnya pada dinding tempat penampungan air seperti lubang batu, gentong, lubang pohon, dan bisa jadi dipelepah pohon pisang diatas garis air (WHO, 2009). Kedua jentik *Aedes aegypti* memiliki sifon yang besar dan pendek serta hanya terdapat sepasang sisik subsentral dengan jarak lebih dari seperempat bagian dari pangkal sifon. Dapat dibedakan jentik *Aedes aegypti* dengan genus yang lain yaitu dengan ciri-ciri tambahan seperti sekurang-kurangnya ada tiga pasang yang satu pada sirip ventral, antenna tidak melekat penuh dan tidak ada yang besar pada toraks (Susanti & Suharyo, 2017).

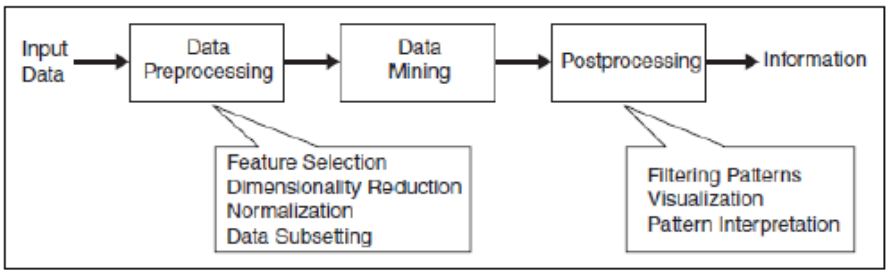
2.3 *Data Mining*

Data mining adalah proses untuk menemukan informasi yang berguna dalam penyimpanan data yang besar. Teknik *data mining* digunakan untuk menjelajahi kumpulan data dalam skala yang besar untuk menemukan pola baru, yang sebelumnya tidak diketahui dan berguna (Han et al., 2011).

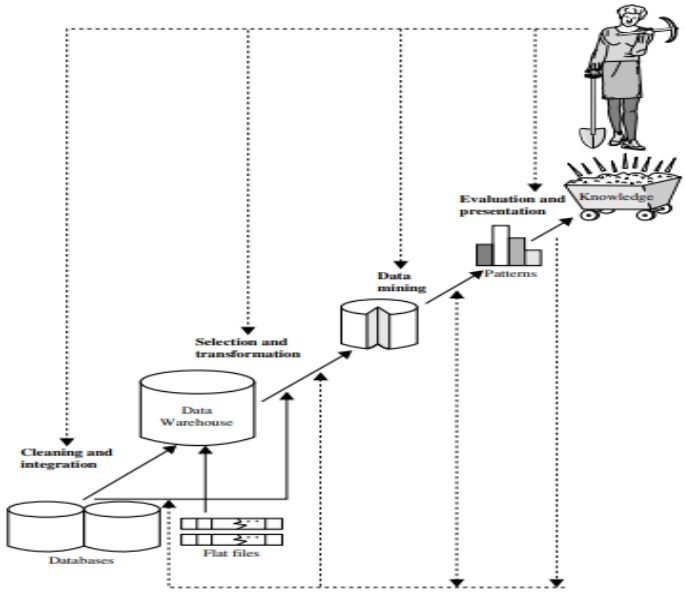
ı *et* al (2011) menyebutkan bahwa *data mining* merupakan bagian penting roses *Knowledge Discovery in Database* (KDD), yang merupakan ıan proses untuk mengubah data mentah menjadi informasi yang berguna.

Optimized using
trial version
www.balesio.com

Proses KDD ini terdiri dari beberapa langkah, mulai dari *preprocessing* data hingga *postprocessing* hasil dari *data mining*, seperti yang ditampilkan pada Gambar 3 (Han et al., 2011).



Gambar 3 Proses Knowledge Discovery in Database (Han et al., 2011)



Gambar 4 Tahap Data Mining (Han et al., 2011)

Gambar 4 menampilkan tahapan *data mining* yang Jiawei Han menyebutkan proses dari *Data Mining* atau *Knowledge Discovery from Database* (KDD) yang terbagi ke dalam 7 tahap, yaitu sebagai berikut (Han et al., 2011).

a. *Data cleaning*: tahap untuk membersihkan data yang hilang, *noise* dan tidak konsisten.

a integration: tahap dimana beberapa sumber data dapat digabungkan.

a selection: tahap untuk data yang relevan dengan analisis diambil dari *abase*.

Optimized using
trial version
www.balesio.com

d. *Data transformation*: tahap untuk data ditransformasikan dan dikonsolidasikan ke dalam bentuk yang sesuai untuk penambangan dengan melakukan operasi *summary* atau *aggregation*.
e. *Data mining*: tahap penting dimana metode cerdas diterapkan untuk mengekstrak pola data.
f. *Pattern evaluation*: tahap untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan ukuran keterkaitannya (*distance/interestingness measure*).
g. *Knowledge presentation*: tahap dimana teknik gambaran visualisasi yang digunakan untuk menyajikan pengetahuan kepada pengguna.

2.4 Metode *Association Rules*

Asosiasi dikenal juga sebagai salah satu teknik *Data Mining* yang menjadi dasar dari berbagai teknik *Data Mining* lainnya. Khususnya salah satu tahap dari analisis asosiasi yang disebut analisis pola *frequent* tinggi (*frequent pattern mining)* menarik perhatian banyak peneliti untuk menarik perhatian banyak peneliti untuk menghasilkan algoritma yang efisien. Penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* (nilai penunjang) yaitu persentase kombinasi item tersebut dalam *database* dan *confidence* (Ikhwan et al., 2021).

Association rule merupakan suatu proses pada data mining untuk menentukan semua aturan asosiatif yang memenuhi syarat minimum untuk *support (minsup)* dan *confidence (minconf)* pada sebuah database. Kedua syarat tersebut digunakan untuk *interesting association rules* dengan dibandingkan dengan batasan yang telah ditentukan, yaitu *minsup* dan *mincof* (Lestari, 2015).

Association Rule Mining adalah suatu prosedur untuk mencari hubungan antar *item* dalam suatu *dataset.* dimulai dengan mencari *frequent itemset*, yaitu kombinasi yang paling sering terjadi dalam suatu *itemset* dan harus memenuhi *minsup*. Dalam tahap ini akan dilakukan pencarian kombinasi *item* yang memenuhi inimum dari nilai *support* dalam *database*. Untuk mendapatkan nilai lari suatu *item* A dapat diperoleh dengan persamaan di bawah berikut 2015).

Optimized using
trial version
www.balesio.com

1. Pembentukan Pola Frekuensi Tinggi

Tahap ini mencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam suatu *database*. Nilai *support* adalah nilai penunjang atau persentase kombinasi sebuah *item* bersamaan dalam suatu *database*. Semakin besar nilai *support* menandakan semakin banyak data pendukung yang ditemukan dalam *database*. Nilai *support* sebuah *item* diperoleh dari persamaan:

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

Kemudian, untuk mendapatkan nilai support dari dua *item* diperoleh dengan persamaan berikut:

$$\text{Support (A,B)} = P(A \cap B) = \frac{Jumlah\ transaksi\ yang\ mengandung\ A\ dan\ B}{Total\ Transaksi} \quad (2)$$

2. Pembentukan Aturan Asosiasi

Setelah seluruh pola frekuensi tinggi ditemukan, maka tahap selanjutnya adalah membentuk aturan asosiasi dengan melihat kombinasi *item* yang memenuhi syarat minimum dari nilai *confidence*. Nilai *confidence* adalah nilai keyakinan berupa kuatnya hubungan antar *item* yang didapatkan. Semakin besar nilai *confidence* menandakan semakin besar kemungkinan kombinasi *item* muncul secara bersamaan, Persamaan dari nilai *confidence* dengan menggunakan kondisi "jika A maka B" adalah sebagai berikut.

$$\text{Confidence } (A \rightarrow B) = P(A|B) = \frac{Jumlah\ Transaksi\ yang\ mengandung\ A\ dan\ B}{Jumlah\ Transaksi\ yang\ mengandung\ A} \quad (3)$$

3. Rasio Peningkatan (*Lift Ratio*)

Lift ratio merupakan nilai yang menunjukkan keabsahan aturan yang terbentuk dalam proses transaksi dan memberikan informasi apakah benar produk A dibeli bersamaan dengan produk B. *Lift ratio* mengukur seberapa ting aturan yang telah terbentuk berdasarkan nilai *support* dan *confidence* g telah didapatkan sebelumnya. Jika nilai *lift ratio* kurang dari atau sama gan ($\leq$) 1, maka hubungan sebab-akibat yang terjadi bersifat saling lepas

Optimized using
trial version
www.balesio.com

satu sama lain. Sedangkan, jika nilai *lift ratio* lebih dari (>)1, maka hubungan sebab-akibat yang terjadi bersifat saling berhubungan satu sama lain dan dapat dikatakan kejadian tersebut bukan kebetulan dan akan berulang. Nilai *lift ratio* diperoleh dari persamaan :

$$Lift\ ratio = \frac{Confidence\ Antecedent}{Support\ Consequent} \quad (4)$$

2.5 Algoritma *FP-Growth*

Algoritma FP-Growth merupakan pengembangan dari algoritma apriori. Algoritma *Frequent Pattern Growth* adalah salah satu alternatif algoritma yang dapat digunakan untuk menentukan himpunan data yang paling sering muncul (*frequent itemset*) dalam sebuah kumpulan data (Lestari, 2015).

Pada algoritma FP-Growth menggunakan konsep pembangunan *tree*, yang biasa disebut *FP-Tree*, dalam pencarian *frequent itemsets* bukan menggunakan *generate candidate* seperti yang dilakukan pada algoritma apriori. Dengan menggunakan konsep tersebut, algoritma FP-Growth menjadi lebih cepat daripada algoritma apriori (Lestari, 2015).

Algoritma lainnya yang sering digunakan untuk menentukan *frequent itemsets* adalah paradigma pendekatan algoritma apriori. Paradigma apriori yang dikembangkan oleh Agrawal dan Srikan (1994) yaitu *anti-monotone Apriori Heuristic*, yaitu dimana setiap pola *itemsets* dengan panjang pola *k* yang tidak sering muncul (tidak *frequent*) dalam sebuah dataset, maka pola dengan panjang (k+1) yang mengandung panjang sub pola *k* tersebut tidak akan sering muncul (tidak frequent). Ide dasar dari paradigma apriori ini adalah dengan mencari himpunan dengan panjang (k+1) dari semua kumpulan pola *frequent* dengan panjang k, lalu sehingga dapat mencocokkan jumlah kemunculan pola tersebut dengan yang dalam *database*. Adapun hal ini mengakibatkan paradigma algoritma apriori akan melakukan *scanning database* yang berulang-ulang, apalagi jika lata cukup besar. Berbeda dengan algoritma FP-Growth yang hanya kan dua kali *scanning database* untuk menentukan *frequent itemset* , 2016).

Optimized using
trial version
www.balesio.com

Struktur data yang digunakan untuk mencari *frequent itemset* dalam algoritma FP-Growth yaitu perluasan dari penggunaan sebuah pohon *prefix*, yang biasa disebut dengan *FP-Tree*. Dengan menggunakan *FP-Tree*, algoritma FP-Growth dapat langsung mengekstrak *frequent itemset* dari *FP-Tree* yang telah terbentuk dengan menggunakan metode *divide-conquer* (Fitriyani, 2016).

2.5.1 Pembangunan *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* adalah hanya memerlukan 2 kali pemindaian data transaksi yang terbukti sangat efisien (Samuel, 2008).

Misal I = {a1,a2,..., an} adalah kumpulan dari *item*. Dan basis data transaksi DB = {T1,T2,....Tn}, dimana Ti (i € [1..n] adalah sekumpulan transaksi yang mengandung *item* di I. Sedangkan *support* adalah penghitung (*counter*) frekuensi kemunculan transaksi yang mengandung suatu pola. Suatu pola dikatakan sering muncul (*frequent pattern*) apabila *support* dari pola tersebut tidak kurang dari suatu konstanta ξ (batas ambang *minimum support*) yang telah didefinisikan sebelumnya. Permasalahan mencari pola *frequent* dengan batas ambang *minimum support count* ξ inilah yang dicoba untuk dipecahkan oleh FP-Growth dengan bantuan struktur *FP-Tree* (Samuel, 2008).

Kekurangan dalam algoritma apriori diperbaiki oleh algoritma FP-Growth dengan menghilangkan *candidate generation*, karena algoritma FP-Growth menggunakan konsep pembangunan *FP-Tree* dalam pencarian *frequent itemsets*. Hal tersebut yang membuat algoritma FP-Growth lebih cepat daripada algoritma apriori dalam pencarian *frequent itemset* (Samuel, 2008).

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

- *FP-Tree* dibentuk oleh sebuah akar yang diberi label *null*, sekumpulan upapohon yang beranggotakan banyak *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 (Samuel, 2008).

Misalkan diberikan pada Tabel 1 data transaksi sebagai berikut, dengan minimum *support count* $\xi = 2$

Tabel 1. Tabel data transaksi mentah (Samuel, 2008)

No	Transaksi
1	a,b
2	b,c,d,g,h
3	a,c,d,e,f
4	a,d,e
5	a,b,z,c
6	a,b,c,d
7	a,r
8	a,b,c
9	a,b,d
10	b,c,e

kuensi kemunculan tiap *item* dapat dilihat pada Tabel 2 sebagai berikut.

Optimized using
trial version
www.balesio.com

Tabel 2. Frekuensi kemunculan tiap karakter (Samuel, 2008)

Item	Frekuensi
a	8
b	7
c	6
d	5
e	4
f	1
g	1
h	1

Setelah dilakukan pemindaian pertama didapat *item* yang memiliki frekuensi di atas *support count* ξ = 2 adalah a,b,c,d, dan e. Kelima *item* inilah yang akan berpengaruh dan akan dimasukkan ke dalam *FP-Tree*, selebihnya (r,z,g, dan h) dapat dibuang karena tidak berpengaruh signifikan. Tabel 3 berikut menampilkan data kemunculan *item* yang *frequent* dalam setiap transaksi, dan diurut berdasarkan frekuensinya yang paling tinggi.

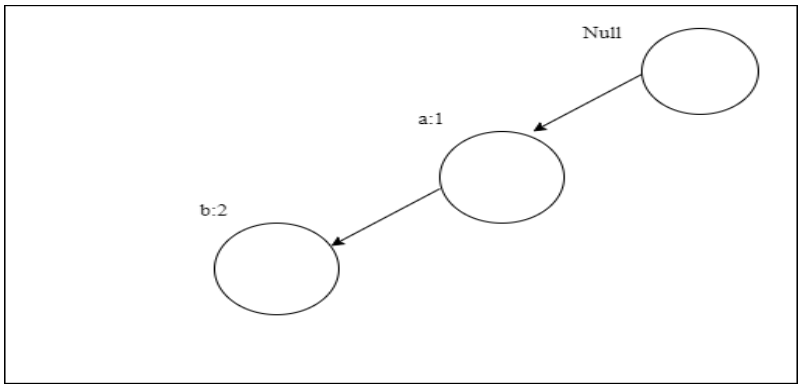
Tabel 3. Tabel data transaksi (Samuel, 2008)

TID	Item
1	{a,b}
2	{b,c,d}
3	{a,c,d,e}
4	{a,d,e}
5	{a,b,c}

Optimized using
trial version
www.balesio.com

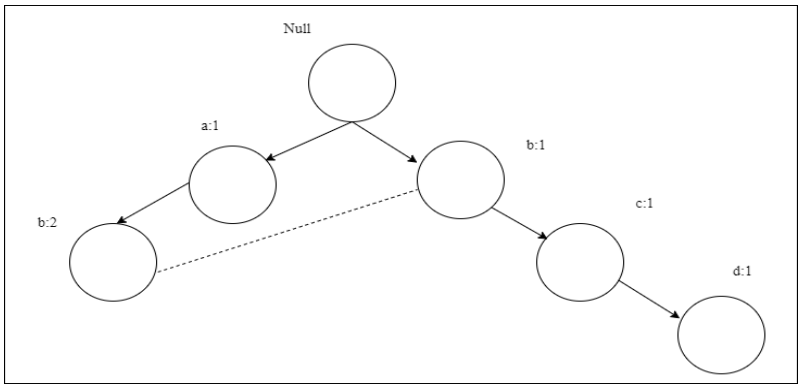
TID	Item
6	{a,b,c,d}
7	{a}
8	{a,b,c}
9	{a,b,d}
10	{b,c,e}

Ilustrasi mengenai pembentukan *FP-Tree* setelah pembacaan TID 1, dapat dilihat pada Gambar 5.



Gambar 5 Hasil pembentukan *FP-Tree* (Samuel, 2008)

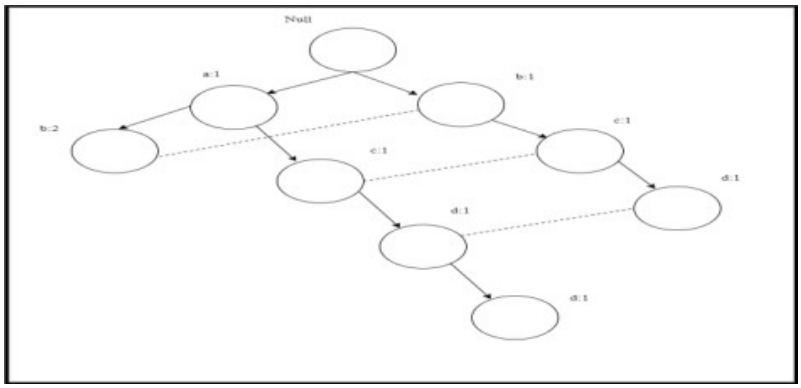
Gambar 6 menunjukkan ilustrasi hasil pembentukan *FP-Tree Transaction*-ID 2.



Gambar 6 *FP-Tree* TID 2 (Samuel, 2008)

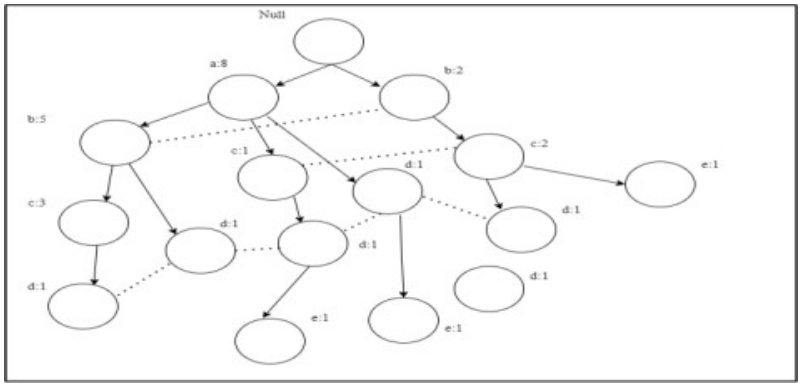
Optimized using
trial version
www.balesio.com

Gambar 7 menunjukkan ilustrasi hasil pembentukan dari *FP-Tree* TID 3



Gambar 7 Hasil pembentukan *FP-Tree* TID 3 (Samuel, 2008)

Gambar 8 menunjukkan ilustrasi hasil pembentukan dari *FP-Tree* TID 4



Gambar 8 Hasil pembentukan *FP-Tree* TID 10 (Samuel, 2008)

Diberikan 10 data transaksi dengan 5 jenis *item* seperti pada Tabel 2. Gambar 5, Gambar 6, Gambar 7, dan Gambar 8 menunjukkan proses terbentuknya *FP-Tree* setiap TID dibaca. Setiap simpul pada *FP-Tree* mengandung nama sebuah *item* dan *counter support* yang berfungsi untuk menghitung frekuensi kemunculan item tersebut dalam tiap lintasan transaksi. *FP-Tree* yang merepresentasikan data transaksi pada Tabel 1 dibentuk dengan cara sebagai berikut :

1. Kumpulan data dipindai pertama kali untuk menentukan *support count* dari setiap *item*. *Item* yang tidak *frequent* dibuang, sedangkan *frequent item* sukkan dan disusun dengan urutan menurun, seperti yang terlihat pada 1.

Optimized using
trial version
www.balesio.com

2. Pemindaian kedua, yaitu pembacaan TID pertama {a,b} akan membuat simpul a dan b, sehingga terbentuk lintasan transaksi Null → a → b. *Support count* dari setiap simpul bernilai awal 1
3. Setelah pembacaan transaksi kedua {b,c,d}, terbentuk lintasan kedua yaitu *Null* →b→c→d. *Support count* masing-masing count juga bernilai awal 1. Walaupun b ada pada transaksi pertama, namun karena *prefix* transaksinya tidak sama, maka transaksi kedua ini tidak bisa dimampatkan dalam satu lintasan.
4. Transaksi keempat memiliki *prefix* transaksi yang sama dengan transaksi pertama, yaitu a, maka lintasan transaksi ketiga dapat ditimpakan di a, sambil menambah *support count* dari a, dan selanjutnya membuat lintasan baru sesuai dengan transaksi ketiga.
5. Proses ini dilanjutkan sampai *FP-Tree* berhasil dibangun berdasarkan tabel data transaksi yang diberikan.

2.5.2 Penerapan Algoritma FP-Growth

Setelah tahap *FP-Tree* terbentuk, maka langkah selanjutnya adalah tahap pembangkitan *conditional pattern base*, tahap pembangkitan *conditional FP-Tree*, dan tahap pencarian *frequent itemset*. Pada tahap ini dapat dilakukan dengan melihat kembali *FP-Tree* yang sudah dibuat sebelumnya.

a. Tahap Pembangkitan *Conditional Pattern Base*

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

b. Tahap Pembangkitan *Conditional FP-Tree*

   Pada tahap ini, *support count* dari setiap *item* pada setiap *conditional pattern base* dijumlahkan, lalu setiap *item* yang memiliki jumlah *support count* lebih besar atau sama dengan minimum *support count* akan dibangkitkan dengan *conditional FP-Tree*.

c. Tahap Pencarian *Frequent Itemset*

   ila *Conditional FP-Tree* merupakan lintasan tunggal (*single path*), maka atkan *frequent itemset* dengan melakukan kombinasi *item* untuk setiap *tional FP-Tree*. Jika bukan lintasan tunggal, maka dilakukan

Optimized using
trial version
www.balesio.com

pembangkitan *FP-Growth* secara *rekursif* (proses memanggil dirinya sendiri) (Lestari, 2015).

Ketiga tahap tersebut merupakan langkah yang akan dilakukan untuk mendapat *frequent itemset*, yang dapat dilihat pada algoritma yang dimuat pada Gambar 9.

Input : FP-tree Tree

Output : R_t sekumpulan lengkap pola frequent

Method : FP-Growth(Tree, null)

Procedure : FP-Growth(Tree, α)

{

1: IF Tree mengandung *single path* p;

2: Then untuk tiap kombinasi (dinotasikan β) dari node-node dalam path P D0

3: bangkitkan pola β α dengan support = minimum support dari node-node dalam β

4: ELSE untuk tiap a_1 dalam header dari Tree D0 {

5: bangkitkan pola

6: bangun β = a_1 α dengan

Support = a_1.support

7: IF Tree β =

8: THEN panggil FP-Growth(Tree, β)}

Gambar 9 *Pseudocode* Algoritma FP-Growth (Lestari, 2015)

2.6 *Clustering*

Clustering atau klasterisasi adalah suatu teknik atau metode untuk mengelompokkan data. Menurut Tan, (2006) *clustering* adalah sebuah proses untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum (Fatmawati & Windarto, 2018).

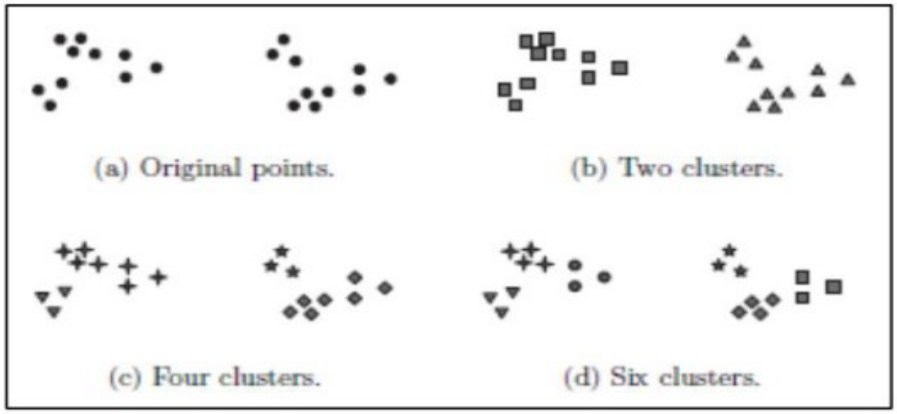
Clustering merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan *cluster*. Objek yang di dalam *cluster* memiliki ı karakteristik antar satu sama lainnya dan berbeda dengan *cluster* yang :isi tidak dilakukan secara manual melainkan dengan suatu algoritma ɟ. Oleh karena itu, *clustering* sangat berguna dan bisa menemukan *group*

Optimized using
trial version
www.balesio.com

atau kelompok yang tidak dikenal dalam data. *Clustering* banyak digunakan dalam berbagai aplikasi seperti misalnya pada *business intelligence*, pengenalan pola citra, *web search*, bidang ilmu biologi, dan untuk keamanan (*security*). Di dalam *business intelligence*, *clustering* bisa mengatur banyak *customer* ke dalam banyaknya kelompok. Contohnya mengelompokkan *customer* ke dalam beberapa *cluster* dengan kesamaan karakteristik yang kuat. *Clustering* juga dikenal sebagai data segmentasi karena *clustering* mempartisi banyak data set ke dalam banyak *group* berdasarkan kesamaannya. Selain itu *clustering* juga bisa sebagai *outlier detection* (Edy Irwansyah, 2015).

Clustering mengacu pada pengelompokan data, observasi, atau kasus ke dalam kelas-kelas objek serupa. *Cluster* adalah kumpulan data yang mirip satu sama lain dan berbeda dengan data di *cluster* lain. *Clustering* berbeda dari klasifikasi dimana tidak ada variabel target *class* untuk pengelompokan. Tugas *clustering* tidak mencoba untuk mengklasifikasikan, memperkirakan, atau memprediksi nilai variabel target. Sebagai gantinya, algoritma *clustering* berusaha untuk mengelompokkan seluruh kumpulan data ke dalam sub-kelompok atau kelompok yang relatif homogen, dimana kesamaan dalam *cluster* yang sama dimaksimalkan, dan kesamaan dalam *cluster* yang berbeda ini diminimalkan. Gambar 10 menunjukkan ilustrasi 3 cara berbeda untuk melakukan *clustering* (Larose & Larose, 2015).

Optimized using
trial version
www.balesio.com



Gambar 10 Ilustrasi 3 cara yang berbeda untuk melakukan *clustering* (Larose & Larose, 2015)

2.6.1 *K-Means Clustering*

Algoritma *K-means* merupakan salah satu metode data *clustering* non-hirarki(*non-hierarchical*) yang berusaha mempartisipasi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan dari data *clustering* ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu *cluster* dan memaksimalisasikan variasi antar *cluster* (Agusta, 2007)

Menurut (Wakhidah, 2010) data *clustering* menggunakan metode *K-Means* ini secara umum dilakukan dengan algoritma dasar sebagai berikut :

1. Tentukan jumlah *cluster*
2. Menentukan nilai *centroid*

   Dalam menentukan nilai *centroid* untuk awal iterasi, nilai awal *centroid* dilakukan secara acak. Sedangkan jika, menentukan nilai *centroid* yang upakan tahap dari iterasi, maka digunakan rumus persamaan 5 sebagai kut.

$$vi = \frac{1}{Ni} \sum_{k=0}^{Ni} Xkj \qquad (5)$$

Optimized using
trial version
www.balesio.com

Dimana:

V_{ij} adalah *centroid* I rata-rata *cluster* ke-I untuk variable ke-j

N_i adalah jumlah data yang menjadi anggota *cluster* ke-i

I *k* adalah indeks dari *cluster*

j adalah indeks dari variabel

x_{kj} adalah nilai data ke-k yang ada di dalam *cluster* tersebut untuk variabel ke-j

3. Menghitung jarak setiap objek ke masing-masing *centroid* dengan masing-masing *cluster*. Untuk menghitung jarak antara objek dengan *centroid* dapat menggunakan persamaan 6 *Euclidean Distance*, yaitu :

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (6)$$

dimana:

D_e adalah *Euclidean Distance*

i adalah banyaknya objek,

(x,y) merupakan koordinat objek dan

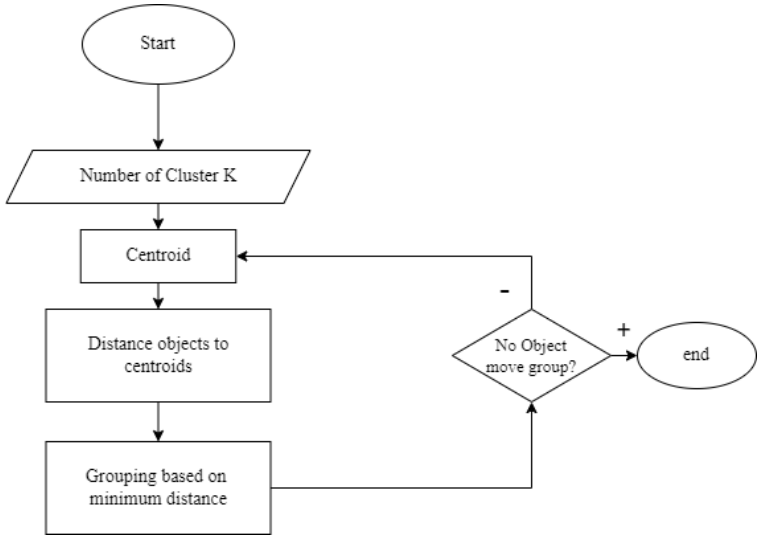
(s,t) merupakan koordinat *centroid*.

4. Pengelompokan object

Untuk menentukan anggota *cluster* adalah dengan memperhitungkan jarak minimum objek. Nilai yang diperoleh dalam keanggotaan data pada *distance matriks* adalah 0 atau 1, dimana nilai 1 untuk data yang dialokasikan ke *cluster* dan nilai 0 untuk data yang dialokasikan ke *cluster* yang lain.

5. Kembali ke tahap 2, lakukan perulangan hingga nilai *centroid* yang dihasilkan tetap dan anggota *cluster* tidak berpindah ke *cluster* lain, atau telah memenuhi nilai maksimum iterasi yang ditentukan (Nur Wakhidah, 2010).

Optimized using
trial version
www.balesio.com



Gambar 11 *Flowchart* algoritma *K-Means* (Wakhidah, 2010)

2.6.2 *Silhouette Coefficient*

Silhouette coefficient adalah metode yang digunakan untuk mengukur seberapa baik hasil *cluster* untuk sebuah titik tertentu. *Silhouette* yang bernilai positif menunjukkan bahwa hasil *cluster* tersebut baik, dengan semakin tinggi nilainya maka semakin bagus juga hasil *cluster*-nya. Sedangkan apabila nilai *silhouette* nya mendekati nol dianggap sebagai hasil *cluster* yang lemah. Nilai *silhouette* yang negatif dianggap salah klasifikasi/pengelompokan, dimana penempatan pada *cluster* terdekat berikutnya mungkin akan lebih baik (Larose & Larose, 2015).

Menghitung nilai *silhouette coefficient* terletak pada kisaran dari -1 sampai +1. Nilai +1 berarti titik yang bersangkutan berjarak sangat jauh dari *cluster* lain, nilai 0 mengindikasikan bahwa titik yang bersangkutan sama saja berada di *cluster* yang sekarang maupun di *cluster* yang lain. Sedangkan nilai -1 mengindikasikan bahwa titik yang bersangkutan berada pada *cluster* yang salah kelompok atau posisi. Nilai *silhouette* dari setiap data dapat didefinisikan pada persamaan 7 berikut ini.

$$s_i = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (7)$$

nana a_i adalah jarak antara titik data dengan titik pusat *cluster*-nya, dan b_i rak antara titik data dengan titik pusat *cluster* terdekat berikutnya. Jelas $_i > a_i$ untuk setiap sampel. Jadi, nilai s_i adalah bilangan positif, dan

Optimized using
trial version
www.balesio.com

diharapkan nilai s_i lebih besar dari 0,5 untuk sebagian besar sampel jika hasil pengelompokan berhasil. Untuk menghitung nilai rata-rata dari *silhouette* terhadap semua data N digunakan persamaan 8 berikut.

$$S = \frac{1}{N} \times \sum_{i=0}^{N} S_i \quad (8)$$

Mengambil nilai *silhouette* rata-rata di atas terhadap semua data dapat digunakan untuk mengukur seberapa baik solusi *cluster* yang dihasilkan. Menurut Daniel T (2015). Menyebutkan bahwa interpretasi dari nilai rata-rata *silhouette* adalah sebagai berikut:

Tabel 4. Interpretasi nilai *Silhouette Coefficient* (Larose & Larose, 2015)

Silhouette Coefficient	**Interpretasi**
0.71 – 1.00	Struktur yang dihasilkan kuat
0.51 – 0.70	Struktur yang dihasilkan baik
0.26 – 0.50	Struktur yang dihasilkan lemah
≤ 0.25	Tidak terstruktur

2.6.3 Sum of Squared Error (SSE)

Sum of Squared Error (SSE) merupakan salah satu cara untuk mengukur *clustering* dengan menggunakan teknik statistik yang mampu mencari apakah objek cocok pada satu *cluster*. Adapun nilai SSE dapat didefinisikan sebagai persamaan 9 berikut ini.

$$SSE = \sum_{i=1}^{n} (d_i - c_i)^2 \quad (9)$$

Dimana:

SSE adalah nilai kuadrat selisih antara koordinat *centroid* setiap data

n adalah jumlah data

d_i adalah nilai data ke-i

c_i adalah nilai *centroid cluster* ke-i

a objek sangat cocok dengan *cluster* tersebut maka nilai SSE adalah nol
g berarti tidak ada *error*. Namun hal itu jarang terjadi. Oleh karena itu,
g yang baik adalah yang memiliki nilai SSE yang serendah mungkin.

Optimized using
trial version
www.balesio.com

Semakin rendah nilai SSE maka semakin sama. SSE yang tinggi maka memiliki derajat perbedaan antara objek dan *cluster* yang dituju (Shofiani, 2017).

2.7 Pemetaan

Pemetaan adalah pengelompokkan suatu kumpulan wilayah yang berkaitan dengan beberapa letak geografis wilayah yang meliputi dataran tinggi, pegunungan, sumber daya dan potensi penduduk yang berpengaruh terhadap sosial kultural yang memiliki ciri khas khusus (Rusdiyanto, 2017). Pengertian lain tentang pemetaan yaitu sebuah tahapan yang harus dilakukan dalam pembuatan peta. Langkah awal yang dilakukan dalam pembuatan data, dilanjutkan dengan pengolahan data, dan penyajian dalam bentuk peta (Mudhari, 2018).

Jadi, dari dua definisi di atas dan disesuaikan dengan penelitian ini maka pemetaan merupakan proses pengumpulan data untuk dijadikan sebagai Langkah awal dalam pembuatan peta, dengan menggambarkan penyebaran kondisi alamiah tertentu secara meruang, memindahkan keadaan sesungguhnya kedalam peta dasar, yang dinyatakan dengan penggunaan skala peta.

2.8 *Microsoft Power Bi*

Microsoft Power BI merupakan salah satu *business intelligent software* atau seperangkat alat *business analytics* yang dapat meningkatkan wawasan terhadap instansi atau organisasi. *Microsoft Power BI* dapat terhubung hingga ratusan sumber data, menyederhanakan persiapan data, dan menggerakkan analisis *adhoc*. Laporan yang dihasilkan dapat ditampilkan di *web* maupun perangkat *mobile* serta mampu membuat *dashboard* yang dipersonalisasi dengan tampilan 360 derajat. Ada berbagai macam bentuk visualisasi grafik yang dapat digunakan di *Microsoft Power BI* diantaranya *stacked bar chart, stacked column chart, clustered bar chart, clustered column chart, line chart, area chart, stacked area chart, ribbon chart, pie chart, donut chart, treemap,* dan yang lainnya. *Microsoft Power BI* terintegrasi dengan Bing Maps untuk menyediakan koordinat peta *default* (sebuah proses yang *eo-coding*) sehingga dapat membuat tampilan peta dilengkapi dengan untuk mengidentifikasi lokasi yang benar. *Microsoft Power BI* dapat an bidang geografis berdasarkan *field* yang telah diberi *geo-code* dengan

Optimized using
trial version
www.balesio.com

menetapkan *data category* pada *data fields*, dengan cara : pilih tabel yang diinginkan, lalu pergi ke *Advanced Ribbon* dan *set data category*-nya ke Alamat, kota, benua, negara, kode pos, provinsi, kecamatan (Darman et al., 2018)

Optimized using
trial version
www.balesio.com