

**IMPLEMENTASI ALGORITMA APRIORI PADA PENYEBAB  
KEMATIAN BAYI**



**TUGAS AKHIR**

*Disusun dalam rangka memenuhi salah satu persyaratan  
Untuk menyelesaikan program Strata-1 Departemen Teknik Informatika  
Fakultas Teknik Universitas Hasanuddin  
Makassar*

**Disusun Oleh :**

**RYA DITA PURNAMA**

**D42116002**

**DEPARTEMEN TEKNIK INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2021**

**SKRIPSI**  
**IMPLEMENTASI ALGORITMA APRIORI PADA PENYEBAB**  
**KEMATIAN BAYI**

**Disusun dan diajukan oleh**

**RYA DITA PURNAMA**

**D421 16 002**



**DEPARTEMEN TEKNIK INFORMATIKA**  
**FAKULTAS TEKNIK**  
**UNIVERSITAS HASANUDDIN**  
**MAKASSAR**  
**2021**

**LEMBAR PENGESAHAN SKRIPSI**  
**IMPLEMENTASI ALGORITMA APRIORI PADA PENYEBAB**  
**KEMATIAN BAYI**

**Disusun dan diajukan oleh**

**RYA DITA PURNAMA**

**D42116002**

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Teknik Informatika Fakultas Teknik Universitas Hasanuddin pada tanggal 22 November 2021 dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,

Pembimbing Pendamping,



Dr. Ir. Ingrid Nurtanio, MT  
Nip. 196108131988112001

Elly Warni, ST., MT.  
Nip. 198202162008122001

Ketua Program Studi,



Dr. Anis Ahmad Iham, S.T., M.IT.  
Nip. 197310101998021002

## PERNYATAAN KEASLIAN

Yang bertanda tangan di bahwa ini :

Name : Rya Dita Purnama  
NIM : D42116002  
Program Studi : Teknik Informatika  
Jenjang : S1

Menyatakan dengan ini bahwa hasil karya tulisan saya berjudul :

### **Implementasi Algoritma Apriori pada Penyebab Kematian Bayi**

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

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.

Makassar, 22 November 2021

Yang menyatakan,



**RYA DITA PURNAMA**

## KATA PENGANTAR

*Assalamu Alaikum Wr. Wb.*

Puji dan syukur penulis panjatkan atas kehadiran Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini yang berjudul **“Implementasi Algoritma Apriori Pada Penyebab Kematian Bayi”** sebagai salah satu syarat dalam menyelesaikan jenjang Strata Satu (S1) pada Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin.

Penulis menyadari bahwa dalam penyusunan dan penulisan skripsi ini tidak lepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak, dari masa perkuliahan sampai dengan masa penyusunan Tugas Akhir. Oleh karena itu, dengan segala kerendahan hati, penulis mengucapkan rasa terima kasih sebesar-besarnya kepada:

1. Tuhan Yang Maha Esa atas semua berkat, karunia serta pertolongan-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini.
2. Orang tua penulis, Bapak M. Tahir dan Ibu Rostina S. Pd yang selalu memberikan motivasi, dukungan, doa, semangat, dan kasih sayang serta selalu sabar dalam mendidik penulis sejak kecil.
3. Ibu Dr. Ir. Ingrid Nurtanio, M.T selaku dosen pembimbing I dan Ibu Elly Warni., S.T., M.T selaku dosen pembimbing II yang selalu menyediakan waktu, tenaga dan pikirannya yang luar biasa untuk mengarahkan penulis dalam penyusunan Tugas Akhir ini.
4. Bapak Dr. Amil Ahmad Ilham, S.T., M.I.T., selaku Ketua Departemen Teknik Informatika Fakultas Teknik Univeristas Hasanuddin atas bimbingannya selama masa perkuliahan penulis.
5. Segenap Dosen dan Staff Departemen Teknik Informatika Fakultas Teknik Universitas Hasanuddin yang telah banyak membantu penulis selama masa perkuliahan.
6. Sahabat-sahabat TRICHE, Fabyola Larasati Masyita, Afifah Ilham, Ismayanti, Nishrina Nurul Amirah, dan Patricia Viola Palada yang selalu menjadi tempat berbagi cerita dan memberikan dukungan dan bantuan kepada penulis.
7. Teman-teman IGNITER16 atas dukungan dan semangat yang telah diberikan

8. Orang-orang berpengaruh lainnya yang tidak sempat penulis sebutkan satu persatu yang telah memberikan dukungan, semangat dan doanya selama penyusunan Tugas Akhir ini.

Akhir kata, penulis menyadari masih terdapat kekurangan dalam penyusunan laporan ini baik isi maupun cara penyajian. Oleh karena itu penyusun mengharapkan adanya saran dan kritik yang bersifat membangun demi kesempurnaan Tugas Akhir ini. Penulis berharap semoga tugas akhir ini dapat memberikan hal yang bermanfaat dan menambah wawasan bagi pembaca dan khususnya bagi penulis juga.

Makassar, November 2021

Penulis

**(Rya Dita Purnama)**

## ABSTRAK

Angka Kematian Bayi (AKB) merupakan salah satu indikator kesejahteraan suatu bangsa yang mencerminkan tingkat masalah kesehatan masyarakat dan dapat menggambarkan status kesehatan penduduk secara umum. Angka Kematian Bayi (AKB) adalah kematian anak usia kurang dari satu tahun.

Berdasarkan data dari RSIA Sitti Khadijah 1 Muhammadiyah Cabang Makassar, RSIA Sitti Fatimah Makassar, RSIA Bahagia, dan RSUD Labuang Baji tahun 2014 s.d 2021, terdapat beberapa indikator penyebab yang menyebabkan kematian pada bayi. Dengan beberapa indikator penyebab kematian tersebut, maka akan dilakukan asosiasi untuk mengetahui keterkaitan antar faktor penyebab pada kematian bayi. Salah satu teknologi yang dapat digunakan adalah *Data Mining* dengan metode *Association Rule*.

Pada penelitian ini algoritma dari metode *Association Rule* yang diterapkan adalah algoritma apriori. Faktor penyebab yang menjadi fitur dalam penelitian ini yaitu sianosis, kelainan kongenital, sepsis, hygroma colli, asfiksia, BBLR (Berat Badan Lahir Rendah), prematur, placenta pravia, RDS (*Respiratory Distress Syndrom*), dan pendarahan.

Hasil dari penelitian ini menunjukkan bahwa penggunaan algoritma apriori yang diujicobakan pada *dataset* penyebab kematian pada bayi dari RSIA Sitti Khadijah 1 Muhammadiyah Cabang Makassar, RSIA Sitti Fatimah Makassar, RSIA Bahagia, dan RSUD Labuang Baji tahun 2014 s.d 2021, dengan jumlah data yang diolah sebanyak 969 kejadian, dengan menggunakan nilai *minimum support* = 0.75 diperoleh panjang kombinasi item yang terbentuk sebanyak *2-itemset*. Dari *large itemset* yang diperoleh proses penemuan aturan asosiasi dilakukan dengan menerapkan *minimum confidence* = 0.95 maka didapatkan 1 aturan asosiasi. Aturan asosiasi yang didapatkan yaitu jika terjadi kematian bayi akibat Sianosis maka berhubungan dengan penyebab kematian Kelainan Kongenital dengan keyakinan sebesar 95%. Nilai *lift ratio* pada hasil penelitian ini lebih dari 1, yaitu 1.17 yang menunjukkan bahwa hubungan dari aturan asosiasi bisa dikatakan kuat.

**Kata Kunci :** *data mining*, algoritma *apriori*, pola keterkaitan, penyebab kematian bayi

## DAFTAR ISI

KATA PENGANTAR.....	i
ABSTRAK .....	iii
DAFTAR ISI .....	iv
DAFTAR GAMBAR.....	vii
DAFTAR TABEL .....	ix
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	2
1.3. Tujuan Penelitian .....	2
1.4. Manfaat Penelitian .....	2
1.5. Batasan Masalah .....	3
1.6. Sistematika Penulisan .....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1. <i>Data Mining</i> .....	5
2.2. Assosiasi .....	7
2.3. Algoritma Apriori .....	8
2.4. <i>Lift Ratio</i> .....	10
2.5. Penelitian Terkait.....	11
BAB III METODOLOGI PENELITIAN .....	13
3.1. Studi Literatur .....	13
3.2. Pengumpulan Data .....	14
3.3. Perancangan Sistem .....	14
3.3.1. <i>Data Collection</i> .....	14
3.3.2. <i>Data Cleaning</i> .....	16



3.3.3. <i>Data Integration</i> .....	16
3.3.4. <i>Data Selection &amp; Transformation</i> .....	17
3.3.5. <i>Data Mining</i> .....	18
3.3.6. <i>Pattern Evaluation</i> .....	18
3.3.7. <i>Knowledge Presentation</i> .....	18
3.4. Implementasi Sistem.....	19
3.5. Evaluasi Sistem.....	21
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....	<b>22</b>
4.1. Penerapan Algoritma Apriori dengan Menggunakan Program Python .....	22
4.1.1. <i>Dataset</i> Penyebab Kematian pada Bayi dari Dua RSIA di Makassar.....	22
4.1.2. Pembentukan Kandidat <i>1-Item</i> (C1).....	24
4.1.3. Menentukan <i>Large 1-Itemset</i> (L1) .....	28
4.1.4. Menentukan Kandidat <i>2-Itemset</i> (C2).....	29
4.1.5. Menentukan <i>Large 2-Itemset</i> (L2) .....	31
4.1.6. Aturan Asosiasi .....	34
4.1.7. Lift Ratio .....	36
4.2. Penerapan Algoritma Apriori dengan Menggunakan <i>Library</i> Python.....	36
4.2.1. <i>Library</i> Python .....	36
4.2.2. Aturan Asosiasi .....	36
4.2.3. Lift Ratio .....	37
4.3. Pembahasan .....	37
4.4. Pengembangan Aplikasi berbasis Website .....	40
<b>BAB V PENUTUP</b> .....	<b>42</b>
5.1. Kesimpulan .....	42
5.2. Saran .....	42

DAFTAR PUSTAKA.....	43
LAMPIRAN .....	45

## DAFTAR GAMBAR

Gambar 2. 1. Proses <i>data mining</i> untuk menemukan pengetahuan.....	7
Gambar 3.1. Tahapan penelitian.....	13
Gambar 3. 2. Contoh Data RSIA Sitti Khadijah 1 Muhammadiyah Cabang Makassar.....	15
Gambar 3. 3. Contoh Data RSIA Sitti Fatimah Makassar.....	15
Gambar 3. 4. Contoh Data RSIA Bahagia.....	15
Gambar 3. 5. Contoh Data RSUD Labuang Baji.....	16
Gambar 3. 6. Penggabungan Data .....	17
Gambar 4. 1. Contoh <i>Dataset</i> Penyebab Kematian pada Bayi.....	23
Gambar 4. 2. <i>Kode Program input dataset</i> .....	23
Gambar 4. 3. <i>Dataset</i> setelah dilakukan <i>cleaning data</i> .....	24
Gambar 4. 4. <i>Kode Program cleaning data</i> .....	24
Gambar 4. 5. Hasil Kandidat 1- <i>Item</i> (C1) .....	25
Gambar 4. 6. Kode Python untuk Pembentukan Kandidat 1- <i>Item</i> (C1).....	25
Gambar 4. 7. Frekuensi <i>Item</i> .....	26
Gambar 4. 8. Kode Python untuk menghitung Frekuensi <i>Item</i> .....	26
Gambar 4. 9. Nilai <i>Support Itemset</i> .....	27
Gambar 4. 10. Kode python untuk menghitung nilai <i>support item</i> .....	28
Gambar 4. 11. <i>Large 1-Itemset</i> (L1).....	28
Gambar 4. 12. <i>Kode python</i> untuk mendapatkan <i>Large 1-Itemset</i> (L1).....	29
Gambar 4. 13. Kandidat 2- <i>Itemset</i> (C2) .....	29
Gambar 4. 14. Frekuensi Kandidat 2- <i>Itemset</i> (C2).....	30
Gambar 4. 15. Nilai <i>support</i> Kandidat 2- <i>Itemset</i> (C2) .....	30
Gambar 4. 16. Kode python untuk menghasilkan Kandidat k- <i>itemset</i> (Ck).....	30
Gambar 4. 17. Kode python untuk menghitung Frekuensi <i>Itemset</i> .....	31
Gambar 4. 18. Kode python untuk menghitung Nilai <i>suppport itemset</i> .....	31
Gambar 4. 19. Hasil <i>Large 1-itemset</i> (L2) .....	32
Gambar 4. 20. Kode python untuk menentukan <i>Large Itemset</i> (Lk).....	32
Gambar 4. 21. Kode python untuk menghitung frekuensi <i>Large Itemset</i> (Lk) .....	33
Gambar 4. 22. Kode python untuk menghitung nilai <i>support Large Itemset</i> (Lk).....	33

Gambar 4. 23. Kode python untuk menampilkan Kandidat <i>k-itemset</i> (Ck), frekuensi Ck, nilai <i>support</i> Ck, <i>Large k-itemset</i> (Lk), frekuensi Lk, dan nilai <i>support</i> Lk.....	34
Gambar 4. 24. Aturan Asosiasi yang dihasilkan .....	35
Gambar 4. 25. Kode program python untuk menghasilkan aturan asosiasi .....	35
Gambar 4. 26. Nilai Lift Ratio.....	36
Gambar 4. 27. Kode program python menggunakan <i>library efficient_apriori</i> .....	36
Gambar 4. 28. <i>Large itemset</i> dari <i>library apriori</i> .....	37
Gambar 4. 29. Aturan Asosiasi dari <i>library apriori</i> .....	37
Gambar 4. 30. <i>Lift Ratio</i> .....	37
Gambar 4. 31. Kode program python untuk menghitung <i>confidence</i> dan <i>lift ratio</i> .....	37
Gambar 4. 32. Algoritma <i>Fp-Growth</i> .....	39
Gambar 4. 33. Tampilan Aplikasi Website .....	41

## DAFTAR TABEL

Tabel 4. 1. Aturan Asosiasi dari program <i>non-library apriori</i> dan <i>library apriori</i> .....	39
Tabel 4. 2. Perbandingan waktu komputasi antara algoritma Apriori dan algoritma <i>FP-Growth</i> .....	40

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Angka Kematian Bayi (AKB) merupakan salah satu indikator kesejahteraan suatu bangsa yang mencerminkan tingkat masalah kesehatan masyarakat dan dapat menggambarkan status kesehatan penduduk secara umum (Budiati, 2016). Bayi juga merupakan investasi Sumber Daya Manusia sebab mereka adalah generasi penerus yang akan meneruskan pembangunan bangsa.

Angka Kematian Bayi (AKB) adalah jumlah kematian anak berusia di bawah dari 1 tahun per 1000 kelahiran. Berdasarkan evaluasi *Millennium Development Goals* (MDGs) pada tahun 2015, kasus kematian bayi di Indonesia adalah 305 per 100.000 kelahiran. Sedangkan target yang dicanangkan Perserikatan Bangsa-Bangsa (PBB) ialah 102 per 100.000 kelahiran. Dari jumlah kasus tersebut, Indonesia ditempatkan sebagai negara dengan angka kematian tertinggi kedua di Asia Tenggara.

Untuk mengurangi jumlah kasus kematian bayi, maka pemerintah ikut andil dalam melakukan upaya untuk meminimalisir jumlah kasus kematian bayi. Berdasarkan Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2015-2019 menargetkan penurunan angka kematian bayi 32 per 1.000 kelahiran hidup menjadi 24 di tahun 2019. Selain itu, sehubungan dengan salah satu target *Sustainable Development Goals* (SDGs) di bidang kesehatan dan kesejahteraan diantaranya adalah mengakhiri kematian bayi.

Di Sulawesi Selatan sendiri tercatat 664 kasus kematian bayi pada tahun 2018 dari 8.6 juta penduduk berdasarkan data dari Dinas Kesehatan Sulawesi Selatan. Makassar sebagai ibu kota dari Provinsi Sulawesi Selatan ikut menambahkan jumlah kasus kematian bayi. Angka kematian bayi di Kota Makassar pada tahun 2013 sebesar 6,71 per 1000 kelahiran hidup, dengan jumlah kasus 165 kematian dari 24.576 kelahiran hidup (M & Thaha, n.d.).

Secara garis besar, kematian bayi terdiri dari dua macam yaitu Endogen dan Eksogen. Kematian bayi endogen atau biasa disebut dengan kematian neonatal yaitu kematian bayi yang terjadi pada bulan pertama setelah dilahirkan yang disebabkan oleh faktor-faktor yang dibawa anak sejak lahir, yang diperoleh dari orang tuanya pada saat konsepsi atau didapat selama kehamilan. Sedangkan kematian bayi eksogen atau kematian *post* neonatal, adalah kematian

bayi yang terjadi setelah usia satu bulan sampai menjelang usia satu tahun yang disebabkan oleh faktor-faktor yang bertalian dengan pengaruh lingkungan luar (Budiati, 2016).

Berdasarkan data dari RSIA Sitti Khadijah 1 Muhammadiyah Cabang Makassar, RSIA Sitti Fatimah Makassar, RSIA Bahagia, dan RSUD Labuang Baji tahun 2014 s.d 2021, terdapat beberapa penyebab-penyebab kematian yang dapat menyebabkan kematian pada bayi. Dari beberapa penyebab tersebut, dapat dilakukan assosiasi atau pola hubungan untuk mengetahui keterkaitan antara penyebab kematian sehingga dapat dilakukan pencegahan terhadap faktor penyebab kematian itu sendiri. Faktor penyebab kematian antara lain sianosis, kelainan kongenital, sepsis, hygroma colli, asfiksia, BBLR (Berat Badan Lahir Rendah), prematur, placenta pravia, RDS (*Respiratory Distress Syndrom*), pendarahan.

*Data mining* dilihat dari sisi teknik pengolahan data menyediakan sejumlah algoritma yang dapat digunakan untuk menggali informasi tersembunyi dari kumpulan data yang multidimensi. Dengan menerapkan metode *Association Rule* (aturan asosiasi) dengan Algoritma Apriori yang akan melakukan penelusuran pada data historis untuk mengidentifikasi pola data yang didasarkan pada sifat-sifat yang teridentifikasi sebelumnya (Istiasih, 2017).

Informasi dari hasil penelitian diharapkan dapat membantu dalam pengambilan keputusan untuk meminimalisir kematian bayi berdasarkan pada pola keterkaitan dari faktor penyebab kematian bayi.

## **1.2. Rumusan Masalah**

Berdasarkan latar belakang masalah, dapat dirumuskan permasalahan adalah “Bagaimana mengetahui keterkaitan pada penyebab Kematian Bayi menggunakan Algoritma Apriori?”

## **1.3. Tujuan Penelitian**

Adapun tujuan dari penelitian ini ialah Mengetahui keterkaitan pada penyebab Kematian Bayi menggunakan Algoritma Apriori.

## **1.4. Manfaat Penelitian**

Adapun manfaat dari penelitian ini ialah:

- 1.) Bagi pemerintah, penelitian ini dapat digunakan sebagai pendukung pengambilan keputusan untuk meminimalisir kematian bayi.
- 2.) Bagi ilmu pengetahuan, dapat memberikan suatu karya penelitian baru yang dapat mendukung pengembangan sistem informasi.

### **1.5. Batasan Masalah**

Untuk menghindari kesalahpahaman dan meluasnya pembahasan, maka penulis memberikan batasan masalah sebagai berikut:

- 1.) Data yang digunakan adalah data kematian bayi di RSIA Sitti Khadijah 1 Muhammadiyah Cabang Makassar, RSIA Sitti Fatimah Makassar, RSIA Bahagia, dan RSUD Labuang Baji tahun 2014 s.d 2021.
- 2.) Fitur yang digunakan yaitu sianosis, kelainan kongenital, sepsis, hygroma colli, asfiksia, BBLR, prematur, placenta previa. RDS, pendarahan.
- 3.) Menggunakan bahasa pemrograman Python.

### **1.6. Sistematika Penulisan**

Untuk memberikan gambaran singkat mengenai isi tulisan ini, maka akan diuraikan beberapa tahapan dari penulisan secara sistematis, yaitu :

#### **BAB I PENDAHULUAN**

Bab ini menguraikan latar belakang diangkatnya judul penelitian implementasi algoritma apriori pada penyebab kematian bayi, disertai dengan rumusan masalah, tujuan dan manfaat penelitian, batasan masalah, dan sistematika penulisan.

#### **BAB II TINJAUAN PUSTAKA**

Bab ini berisi teori-teori yang berhubungan dengan penelitian termasuk *data mining*, *association*, dan algoritma yang digunakan.

#### **BAB III METODOLOGI PENELITIAN**

Bab ini berisi tentang perencanaan dan penerapan algoritma sehingga menghasilkan aturan asosiasi.



#### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini berisi tentang hasil penelitian data dan pembahasan terkait pengolahan data yang telah dilakukan.

#### **BAB V PENUTUP**

Bab ini berisi tentang kesimpulan yang didapatkan berdasarkan hasil penelitian yang telah dilakukan serta saran-saran untuk pengembangan lebih lanjut.

## BAB II TINJAUAN PUSTAKA

### 2.1. *Data Mining*

*Data mining* merupakan aktivitas yang berkaitan dengan pengumpulan data, pemakaian data historis untuk menemukan pengetahuan, informasi, pola atau hubungan dalam data yang berukuran besar. *Output data mining* dapat digunakan sebagai alternatif dalam pengambilan keputusan atau untuk memperbaiki pengambilan keputusan di masa yang akan datang (Eferi Buulolo, 2020).

*Data mining* sebagai proses menemukan struktur data yang menarik menggunakan satu atau beberapa algoritma untuk mengidentifikasi *trend* atau pola menarik dalam suatu data. *Knowledge* yang diperoleh dari model *data mining* akan digenerelasikan yang tujuannya dapat diterapkan pada situasi yang baru (Roiger, 2017).

*Data mining* atau kadang disebut juga *Knowledge Discovery in Database (KDD)* menggunakan metode teknik data mining untuk mengekstraksi pengetahuan dan informasi potensial dari sekumpulan data besar. Pengetahuan dan informasi yang dihasilkan dari KDD bersifat baru dan mudah dimengerti serta bermanfaat.

Proses *data mining* merupakan serangkaian proses yang terdiri dari langkah-langkah sebagai berikut (Zamroni, Werdiningsih, & Purbandini, 2017):

#### 1) *Collection*

Proses *collection data* dilakukan dengan mengumpulkan data yang bertujuan untuk menghasilkan data mentah (*raw data*) yang akan diolah menggunakan teknik data mining. Pada tahap ini, informasi yang akan diperoleh seperti jumlah *data* yang digunakan, *raw data*, dan fitur-fitur dari *data*.

#### 2) *Cleaning / Selection*

Pada tahap ini data yang sudah dipilih akan dilakukan pembersihan. Proses *cleaning* bertujuan untuk menghilangkan *noise* dan *data-data* yang tidak konsisten. *Cleaning* juga

dapat dilakukan proses memperkaya data dengan menambah informasi lain yang relevan yang disebut dengan istilah *enrichment*.

Sedangkan, proses *selection* adalah dengan melakukan pemilihan data yang akan digunakan karena tidak semua data yang ada dapat dipergunakan. Aktivitas pemilihan data meliputi pembuatan kumpulan data target, penentuan fitur, pemilihan sampel data dan penyimpanan data pada sebuah berkas.

### 3) *Integration*

*Data integration* merupakan proses penggabungan data yang berasal dari beberapa sumber data. Integritas data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti nama, nomor pelanggan, dan lainnya. *Data* yang berbeda memiliki konvensi penamaan fitur yang berbeda, sehingga akan menyebabkan redundansi data. Pembersihan data dapat dilakukan untuk menghilangkan redundansi data.

### 4) *Transformation*

Pada tahap ini, *data-data* yang telah terpilih akan di-*transformasi* ke dalam bentuk yang sesuai untuk prosedur *data mining* dengan cara melakukan normalisasi atau agregasi data.

### 5) *Data Mining*

*Data mining* merupakan tahapan utama dalam *Knowledge Discovery in Database (KDD)*. *Data mining* adalah proses penggalian dan pencarian pengetahuan dan informasi yang bermanfaat dengan menggunakan algoritma/metode/teknik tertentu sesuai dengan pengetahuan atau informasi yang dicari.

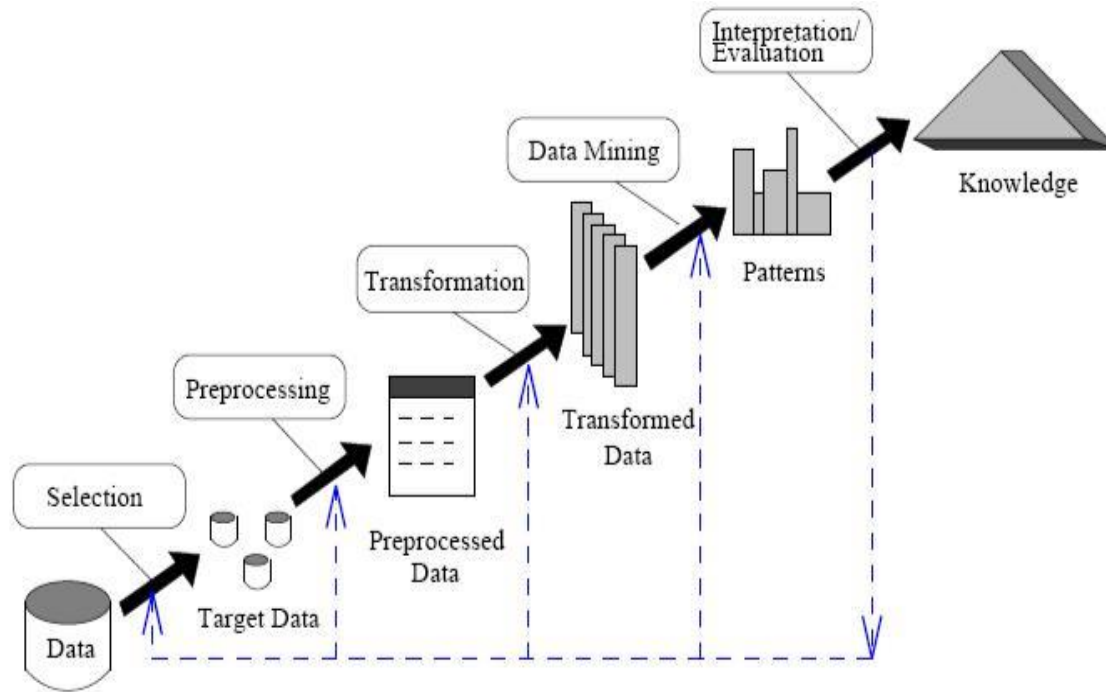
### 6) *Pattern Evaluation*

Pengetahuan atau informasi yang dihasilkan dari proses *data mining* akan dipresentasikan atau ditampilkan ke dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan seperti informasi ditampilkan dalam bentuk grafik, pohon keputusan, ataupun dalam bentuk *rule*.

### 7) *Knowledge Presentation*

Tujuan utama proses *Knowledge Discovery in Database (KDD)* adalah memperoleh pengetahuan atau informasi yang bermanfaat dan mudah dimengerti. *Knowledge presentation* merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir

dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining.



**Gambar 2. 1.** Proses *data mining* untuk menemukan pengetahuan

## 2.2. Asosiasi

Asosiasi berguna untuk mengungkap hubungan yang menarik yang tersembunyi dalam dataset besar. Hubungan yang terungkap tersebut dapat dipresentasikan dalam bentuk aturan asosiasi (*association rules*) atau himpunan *item* yang sering muncul (*sets of frequent items*) (Fricles Ariwisanto Santuri, 2019). Aturan asosiasi merupakan salah satu metode yang bertujuan mencari pola yang sering muncul di antara banyak transaksi, dimana setiap transaksi terdiri dari beberapa *item* (Azwar Anas, 2018). Aturan asosiasi sering digunakan dalam menganalisis keranjang belanja. Pengetahuan yang dihasilkan dari aturan asosiasi berbentuk

“*if... then...*” atau “jika...maka...”. Bisa di umpamakan jika seseorang membeli barang A maka akan membeli barang B pula dalam satu transaksi.

Sebagai contoh aturan asosiasi dari analisa pembelian di sebuah supermarket, dapat diketahui berapa besar kemungkinan seorang pelanggan membeli roti bersamaan dengan membeli susu. Dengan pengetahuan ini, pemilik supermarket dapat mengatur penempatan barang-barangnya dan merancang strategi pemasaran dengan pengetahuan yang diperoleh dari hasil asosiasi tersebut.

Menurut Daniel T. Larose (2005), aturan asosiasi dari *database* besar terdiri dari dua langkah, yaitu (Mandala, 2017):

- 1) Temukan semua *frequent itemset*, yaitu menemukan semua *itemset* dengan frekuensi  $\geq$  *minimum support*.
- 2) Dari *frequent itemset*, buat aturan asosiasi yang memenuhi kondisi *minimum support* dan *minimum confidence*.

Pencarian pola kaidah asosiasi menggunakan dua buah parameter nilai yaitu dukungan (*support*) dan keterpercayaan (*confidence*) yang memiliki rentang nilai antara 0%-100% (Putra, Raharjo, Sandi, Ridwan, & Prasetyo, 2019).

### 2.3. Algoritma Apriori

Algoritma apriori adalah algoritma pengembalian data dengan aturan asosiasi (*Association Rules*) untuk menentukan hubungan asosiasi suatu kombinasi *item*. Algoritma apriori merupakan algoritma yang paling populer dikenal sebagai dengan paradigma *generate and test*, yaitu pembuatan kandidat kombinasi *item* yang mungkin berdasar aturan tertentu lalu diuji apakah kombinasi item tersebut memenuhi syarat *support minimum* (Malik, Azis, Supianto, & Setiawan, 2019).

Menurut Tampubolon (2013), beberapa istilah yang digunakan dalam algoritma apriori antara lain (Tommy, 2017):

- 1) *Support* atau dukungan : probabilitas atribut muncul secara bersamaan dari seluruh transaksi. *Support* untuk aturan “X→Y” adalah probabilitas atribut atau kumpulan atribut X dan Y yang terjadi bersamaan.
- 2) *Confidence* atau tingkat kepercayaan : probabilitas kejadian atribut terjadi bersamaan dimana salah satu atribut sudah pasti terjadi.
- 3) *Minimum Support* : parameter yang digunakan sebagai batasan frekuensi kejadian atau *support count* yang harus dipenuhi suatu kelompok data untuk dapat dijadikan aturan.
- 4) *Minimum Confidence* : parameter yang mendefinisikan minimum level dari *confidence* yang harus dipenuhi oleh aturan yang berkualitas.
- 5) *Itemset* : kelompok atribut
- 6) Kandidat *Itemset* ( $C_k$ ) : *item-item* yang akan dihitung *support count*-nya.
- 7) *Frequent Itemset* ( $F_k$ ) : *itemset* yang sering terjadi, atau *itemset- itemset* yang sudah melewati batas *minimum support* yang telah ditentukan.

Algoritma apriori dinyatakan dengan algoritma paling terkenal untuk menemukan pola frekuensi tinggi. Pola frekuensi tinggi adalah pola-pola item di dalam suatu *database* yang memiliki frekuensi atau *support* diatas ambang batas tertentu atau dikenal dengan *Minimum Support* (Fricles Ariwisanto Santuri, 2019). Metode algoritma apriori terdiri dari dua tahap, yaitu :

1) Analisa Pola Frekuensi Tinggi

Tahap ini mencari kombinasi *item* yang memenuhi syarat minimum dari nilai *support* dalam *database*. Rumus untuk mendapatkan nilai *support* adalah sebagai berikut :

$$Support(A) = \frac{\text{jumlah kejadian mengandung } A}{\text{total kejadian}} \quad (2.1)$$

Sementara nilai *support* untuk 2 *itemset* diperoleh dengan menggunakan rumus sebagai berikut :

$$Support(A, B) = \frac{\text{jumlah kejadian mengandung } A \text{ dan } B}{\text{total kejadian}} \quad (2.2)$$

Untuk menghitung *support* untuk 3 *itemset*, dilakukan dengan rumus sebagai berikut :

$$Support(A, B, C) = \frac{\text{jumlah kejadian mengandung } A, B, \text{ dan } C}{\text{total kejadian}} \quad (2.3)$$

## 2) Pembentukan Aturan Asosiasi

Akurasi dari suatu aturan asosiasi disebut *confidence*. *Confidence* disebut dengan nilai kepastian adalah kuatnya hubungan antar *item* dalam aturan asosiasi. Setelah semua pola frekuensi tinggi telah ditemukan, barulah dicari aturan asosiasi yang memenuhi syarat *minimum confidence*, dengan menghitung *confidence* aturan asosiasi  $A \rightarrow B$ . Untuk memperoleh nilai *confidence*, dilakukan dengan rumus :

$$Confidence(A \rightarrow B) = \frac{\text{jumlah kejadian mengandung } A, \text{ dan } B}{\text{jumlah kejadian mengandung } A} \quad (2.4)$$

Aturan asosiasi  $A \rightarrow B$  dikatakan dengan A sebagai *anteseden* dengan B sebagai *konsekuen*. *Anteseden* merupakan sebab yang menjadi *item* *konsekuen*. Sedangkan *konsekuen* adalah sebuah akibat atau juga *item* yang akan terjadi setelah terjadi *anteseden* (Tommy, 2017).

Dalam algoritma apriori, ada 2 proses utama yang dilakukan untuk mendapatkan *frequent itemset*, yaitu (Saputro, 2017) :

### 1) *Join* (penggabungan)

Dalam proses ini, setiap *item* dikombinasikan dengan *item* lain sampai tidak dapat terbentuk kombinasi lagi.

### 2) *Pruning* (pemangkasan)

Pada proses ini, hasil kombinasi *item* akan dipangkas berdasarkan *minimum support* yang telah ditentukan.

## 2.4. Lift Ratio

*Lift Ratio* menghasilkan ukuran untuk menguji ke-valid-an aturan yang sudah terbentuk dan mengetahui kekuatan pada *association rule* yang telah terbentuk (Riszky & Sadikin, 2019). Nilai *Lift* menunjukkan adanya kekuatan dari aturan atas kejadian acak dari *antecedent* dan *consequent* berdasarkan pada *support*-nya masing-masing. Hal ini akan memberikan informasi

tentang probabilitas dari *consequent* berdasarkan *antecedent* (Despitaria, Sujaini, & Tursina, 2016). Untuk menghitung nilai *lift ratio* digunakan rumus sebagai berikut :

$$Lift\ Ratio = \frac{Confidence\ Antecedent}{Support\ Consequent} \quad (2.5)$$

Aturan dikatakan kuat apabila nilai *lift ratio* > 1. Lebih tinggi nilai *lift ratio* menunjukkan bahwa lebih besar kekuatan dari asosiasi.

## 2.5. Penelitian Terkait

Dalam penyusunan skripsi ini, terdapat penelitian-penelitian terdahulu yang berkaitan dengan judul penelitian ini antara lain :

Citra Verawati Purba dan Efori Buulo (2020). Judul penelitian “Implementasi Algoritma Apriori Untuk Menentukan Pola Data Penyakit Pada Anak Usia Dini (Studi Kasus : RS. Estomihi)”. Penelitian tersebut terdapat 7 jenis penyakit yang digunakan sebagai atribut penelitian, diantaranya yaitu : (1) Thypoid, (2) Demam Dengue, (3) Bacterial Infection, (4) Prolenged Fever, (5) Infeksi Saluran Pernapasan Atas, (6) Demam Rematik, (7) Viral Infection. Algoritma apriori digunakan untuk mengetahui pola data penyakit yang sering diderita anak usia dini pada Rumah Sakit Estomihi. Terdapat 3 kesimpulan dari penelitian tersebut, yaitu : (1) Pola penyakit pasien anak usia dini dapat diketahui dengan menggunakan data rekam medis dari rumah sakit dan mengambil data-data yang mempengaruhi penyakit pasien tersebut, (2) Penerapan Algoritma Apriori pada teknik *data mining* sangat efisien dan dapat mempercepat proses pembentukan pola kombinasi *itemset* pada anak usia dini pada RS Estomihi, yaitu dengan *support* dan *confidence* yang tertinggi, (3) Perhitungan *data mining* dengan weka sangat mudah dilakukan dengan menginput data ke dalam sebuah *file* yang berformat .csv kemudian menyimpan *file* tersebut dengan format .arff yang terdapat di dalam weka untuk mendapatkan hasil prediksi dari aplikasi tersebut (Purba & Buulolo, 2020).

Ismul Zamroni, Indah Werdiningsih, dan Purbandini (2017). Judul penelitian “Identikasi Pola Penyakit Anak Di Bawah 5 Tahun (Balita) Dengan Menggunakan Algoritma Apriori”. Atribut yang digunakan dalam penelitian tersebut meliputi keluhan penyakit (1) Panas, (2) Batuk, (3) Pucat, (4) Pilek, (5) Muntah, (6) Pusing, (7) Diare, (8) Mencret, (9) Mual. Dengan



menggunakan Algoritma apriori, penulis mengharapkan untuk dapat membantu menemukan dan mengetahui pola penyakit pada balita. Kesimpulan dari penelitian tersebut adalah berdasarkan hasil analisa yang dilakukan dengan beberapa kombinasi minimum support dan *minimum confidence* diperoleh berupa hasil gejala penyakit yang paling kuat yaitu panas, batuk, pilek, dan pucat (Zamroni et al., 2017).