

2005. Tersedia dalam <http://www.digilib.depkes.go.id>
- [3]. Survey Kesehatan Nasional (Surkesnas) RI. Profil Kesehatan Indonesia, Jakarta, 2009.
  - [4]. Situasi Diare di Indonesia, Kementerian Kesehatan RI, 2011
  - [5]. Fajar Astuti Hermawati. *Data mining*. Yogyakarta, Indonesia : Andi Offset, 2013
  - [6]. Eko Prasetyo, *Data mining : Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta, Indonesia : Andi Offset, 2014
  - [7]. Kusrini & Emha Taufik Lutfi. 2009. *Algoritma Data mining*. Yogyakarta. Andi Offset.

c. Mencari kandidat item set

- Gabungkan item pada urutan langkah ke-2  
{ A B, A C, A D, B C, B D, C D }
- Hapus yang tidak ada pada item set  
Item set tidak ada yang dihapus karena tidak ada dalam item set

d. Hitung nilai support untuk masing-masing item set.

| Item Set | Support |
|----------|---------|
| A C      | 25%     |
| AB       | 50%     |
| A D      | 50%     |
| B C      | 25%     |
| B D      | 25%     |
| C D      | 75%     |

e. Tentukan item set yang memenuhi minimum support

| Item Set | Support |
|----------|---------|
| AC       | 50%     |
| AD       | 50%     |
| CD       | 75%     |

f. Ulangi langkah 3 dan 4

- Gabungkan item set pada tabel

| Item Set | Hasil gabungan 3 item set |
|----------|---------------------------|
| AC+AD    | ACD                       |
| AC+CD    | ACD                       |
| AD+CD    | ACD                       |

- Hapus yang tidak ada dalam item set  
Item set tidak ada yang dihapus karena ada semua dalam item set
- Hitung nilai support untuk masing-masing item set.

| Item Set | Support |
|----------|---------|
| B C D    | 50%     |

g. Stop , karena tidak ada lagi kandidat untuk 4 item set

h. Selanjutnya dilakukan asosiasi

| Aturan (X->Y) | Support (X U Y) | Support (X) | Confidence |
|---------------|-----------------|-------------|------------|
| AC→D          | 50%             | 75%         | 67%        |
| AD→C          | 50%             | 75%         | 67%        |
| CD→A          | 75%             | 75%         | 100%       |
| A→C           | 75%             | 75%         | 100%       |
| A→D           | 75%             | 75%         | 100%       |
| C→A           | 75%             | 75%         | 100%       |
| C→D           | 75%             | 75%         | 100%       |
| D→A           | 75%             | 75%         | 100%       |
| D→C           | 75%             | 75%         | 100%       |

Nilai confidence terbesar adalah

| Aturan (X->Y) | Support (X U Y) | Support (X) | Confidence |
|---------------|-----------------|-------------|------------|
| CD→A          | 75%             | 75%         | 100%       |
| A→C           | 75%             | 75%         | 100%       |
| A→D           | 75%             | 75%         | 100%       |
| C→A           | 75%             | 75%         | 100%       |
| C→D           | 75%             | 75%         | 100%       |
| D→A           | 75%             | 75%         | 100%       |
| D→C           | 75%             | 75%         | 100%       |

Artinya Item CD→A (alergi obat dan infeksi virus diare yang berakibat gangguan lambung), item A→C (gangguan lambung yang berakibat alergi obat), A→D (gangguan lambung yang berakibat infeksi virus diare), C→A (alergi obat yang berakibat gangguan lambung), C→D (alergi obat yang berakibat infeksi virus diare), D→C (infeksi virus diare yang berakibat gangguan lambung), D→A (infeksi virus diare yang berakibat alergi obat) menjadi faktor utama penyebab diare.

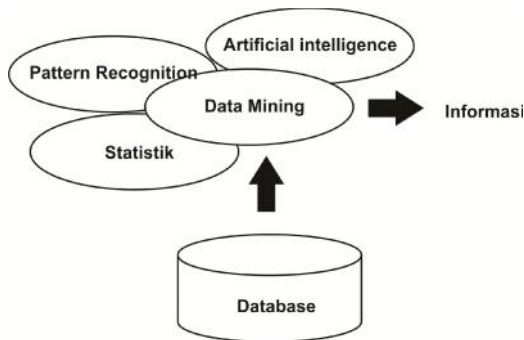
#### 4. KESIMPULAN DAN SARAN

Algoritma *A-priori* mampu memberikan solusi dalam penentuan pola data penyebaran penyakit diare di Puskesmas Tambakboyo. Dengan menyediakan seleksi atau pengelompokan terhadap data keseluruhan penyebaran penyakit diare berdasarkan jenis kelamin, ataupun pengelompokan penyebaran per wilayah. Hal ini akan menunjang kemampuan pengambil keputusan dalam pengambilan keputusan. Penyediaan laporan akan penyebaran penyakit diare di wilayah Tambakboyo juga semakin cepat.

#### DAFTAR PUSTAKA

- [1]. Permenkes no.75 tahun 2014 tentang Pusat Kesehatan Masyarakat.
- [2]. Luza, YB : Hubungan kualitas bakteriologis peralatan makan balita dengan kejadian diare pada balita di wilayah kerja Puskesmas Sukaresmi kabupaten Cianjur provinsi Jawa Barat tahun 2005,

yang dapat berguna.



**Gambar 1.** Akar ilmu *data mining*

Berdasarkan akar keilmuannya, terdapat 4 (empat) akar bidang ilmu yang terdapat dalam *data mining* yaitu statistik, database system, artificial intelligence, pattern recognition. [6]

### 2.3. Association Rule Mining

Analisis asosiasi atau association rule mining adalah teknik data mining untuk menemukan aturan asosiatif antara kombinasi item [7]. Aturan ini biasanya digunakan pada tempat-tempat yang melakukan transaksi dengan minimal dua item transaksi yang dilibatkan.

Ada dua parameter penting dalam *association rules mining* yang berfungsi dalam pembentukan *rules*, yaitu *Support* (nilai penunjang) dan *confidence* (nilai kepastian).

### 2.4. Algoritma Apriori

Algoritma A-priori adalah salah satu algoritma yang digunakan pada market basket analysis. Terdapat dua metodologi dasar dalam algoritma A-priori yaitu [7]:

#### a. Analisis pola frekuensi tinggi

Adalah tahap pembentukan kombinasi item yang memenuhi syarat minimum support yang ditentukan. Nilai support pada suatu itemset diperoleh dengan rumus :

$$\text{support (A)} = \frac{\text{jumlah transaksi mengandung A}}{\text{total transaksi}}$$

Sedangkan pada dua itemset atau lebih diperoleh dengan rumus :

$$\text{support (A, B)} = \frac{\Sigma \text{Transaksi mengandung A dan B}}{\Sigma \text{Total transaksi}}$$

#### b. Pembentukan aturan asosiasi

Setelah menentukan nilai support pada itemset frekuensi tinggi selanjutnya adalah membentuk aturan asosiasi untuk menyatakan kuatnya hubungan kombinasi itemset pada transaksi. Minimal terdapat dua kandidat itemset A dan itemset B dalam menentukan aturan asosiasi dan pada rules yang telah terbentuk berlaku hukum asosiatif  $A \rightarrow B$  tidak berlaku  $B \rightarrow A$ . Untuk menentukan aturan  $A \rightarrow B$  digunakan rumus :

*Confidance*

$$= P(B|A) = \frac{\Sigma \text{Transaksi mengandung A dan B}}{\Sigma \text{Transaksi mengandung A}} \frac{an B}{A}$$

## 3. HASIL PEMBAHASAN

### 3.1. Perhitungan A-priori

#### a. Data Item Transaksi

Item A = Gangguan Usus

Item B = Gangguan Lambung

Item C = Alergi Obat

Item D = Infeksi Virus

| Id.Transaksi | Item Set Penyebab      |
|--------------|------------------------|
| 1            | Item A, item C, item D |
| 2            | Item A, item C, item D |
| 3            | Item A, item B         |
| 4            | Item B, item C, item D |

#### b. Mencari support masing-masing item

Support (A) = Jumlah transaksi yang sesuai (A) / Total Transaksi

Contoh = Dari transaksi diatas terdapat 4 transaksi dan 3 transaksi yang sesuai item A yaitu transaksi 1, 2 dan 3 maka support = 3 / 4, dihasilkan 75%.

| Item Set | Support |
|----------|---------|
| A        | 75%     |
| B        | 50%     |
| C        | 75%     |
| D        | 75%     |

lingkungan fisik maupun rendahnya perilaku masyarakat untuk hidup bersih dan sehat[2].

Hasil Survey Kesehatan Nasional (Surkesnas) tahun 2009 Kejadian Luar Biasa (KLB) Diare terjadi di 15 provinsi dengan jumlah penderita sebanyak 5.756 orang, jumlah kematian sebanyak 100 orang atau CFR sebesar 1,74%. CFR tahun 2009 tersebut lebih rendah dibandingkan tahun sebelumnya, pada tahun 2006 sebesar 2,52% kemudian turun 1,26% pada tahun 2007. Angka ini naik 2,48% pada tahun 2008 dan turun 1,74 pada tahun 2009[3]. Salah satu langkah dalam pencapaian target MDG's (Goal ke-4) adalah menurunkan kematian anak menjadi 2/3 bagian dari tahun 1990 sampai pada 2015. Berdasarkan Survei Kesehatan Rumah Tangga (SKRT), Studi Mortalitas dan Riset Kesehatan Dasar dari tahun ke tahun diketahui bahwa diare masih menjadi penyebab utama kematian balita di Indonesia. Penyebab utama kematian akibat diare adalah tata laksana yang tidak tepat baik di rumah maupun di sarana kesehatan. Untuk menurunkan kematian karena diare perlu tata laksana yang cepat dan tepat[4].

Kemajuan teknologi informasi pada masa kini bergerak dengan cepatnya, kebutuhan akan penyajian data sangat diperlukan untuk mendapatkan hasil analisa berbagai bidang pekerjaan yang membuat informasi menjadi hal penting dalam perkembangan masyarakat pada saat ini maupun dimasa yang akan datang. Namun dengan besarnya data informasi yang telah dikelola sekian lama terkadang tidak diimbangi dengan penyajian informasi yang memadai. Didalam pengambilan keputusan tidak hanya mengandalkan data oprasional saja akan tetapi data – data tersebut perlu dianalisa untuk mendapatkan informasi mengenai potensi – potensi

apa saja yang dapat dimanfaatkan oleh pengambil keputusan. Hal ini mamicu lahirnya suatu cabang keilmuan baru yang berkaitan dengan penggalian informasi dengan metode yang dapat mengolah data dalam jumlah besar yang dinamakan *data mining*.

Berdasarkan permasalahan di atas maka penulis perlunya suatu solusi untuk mengetahui penyebaran penyakit diare. Misalkan mencari pasien yang terkena diare berdasarkan jenis kelamin dan juga pasien terkena diare dari wilayah kelurahan tertentu. Maka dari itu dalam penulisan tugas akhir ini mengambil judul “Aplikasi *Data mining* Distribusi Penyakit Diare Dengan Metode Association Rule Menggunakan Algoritma Apriori (Studi Kasus Puskesmas Tambakboyo)”

## 2. METODE

### 2.1. Diare

### 2.2. Data Mining

*Data mining* adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara observasi contoh-contoh spesifik dari kensep-konseps yang akan dipelajari[5].

*Data mining* juga disebut sebagai *Knowledge Discovery in Database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pola atau hubungan dalam set data berukuran besar. Memang dalam *data mining* atau KDD bertujuan memanfaatkan data dalam basis data dengan pengolahnya sehingga menghasilkan informasi baru

patients receiving care at the health center Tambakboyo. with the amount of information data that has been maintained for so long sometimes not matched with adequate information presentation. In the decision-making not only rely on the data oprasional alone but the data - the data needs to be analyzed to obtain information regarding the potential - the potential of what can be used by decision makers. It spur the birth of a new branch of science concerned with extracting information with methods that can process large amounts of data, called data mining. Based on the above problems, the authors need for a solution to determine the spread of diarrheal diseases. Let's look for patients with diarrhea by sex and also patients with diarrhea of ??certain sub region. The method used is the method Association Rule Algorithm Using Apriori Apriori algorithm is able to provide solutions in determining the pattern of the data spread of diarrheal diseases in PHC Tambakboyo. By providing the selection or grouping of the overall data diarrheal diseases spread by sex, or grouping spread per region. This will support the ability of decision makers in decision making. Provision of reports about the spread of diarrheal diseases in the region Tambakboyo also faster.

**Keywords:** Distribution Data mining, Association Rule, Apriori Algorithm

## 1. PENDAHULUAN

Pusat Kesehatan Masyarakat yang selanjutnya disebut Puskesmas adalah fasilitas pelayanan kesehatan yang menyelenggarakan upaya kesehatan masyarakat dan upaya kesehatan perseorangan tingkat pertama, dengan lebih mengutamakan upaya promotif dan preventif, untuk mencapai derajat kesehatan masyarakat yang setinggi – tingginya di wilayah kerjanya[1].

Puskesmas Tambakboyo merupakan fasilitas pelayanan kesehatan yang memiliki area pelayanan kesehatan untuk wilayah Kecamatan Tambakboyo dibawah naungan Dinas Kesehatan Kabupaten Tuban. Dalam perkembangannya Puskesmas Tambakboyo senantiasa meningkatkan mutu pelayanan kesehatan kepada masyarakat. Hal ini dibuktikan dengan diperolehnya sertifikat ISO 9001:2008 yang merupakan wujud dari sistem pelayanan yang sesuai mutu standar pelayanan. Pada saat ini Puskesmas Tambakboyo telah memiliki beberapa fasilitas dan layanan diantaranya poli gigi, poli umum, UGD, KIA/KB, radiologi, persalinan, rawat inap, poli MTBM/MTBS dll.

Selama tahun 2014, Puskesmas Tambakboyo memiliki data pasien

sebanyak 12.399 orang. Di dalam gudang data tersebut terdapat beberapa penyakit yang mendominasi dialami oleh masyarakat diantaranya demam, ispa, diare, tukak lambung dll. Dari data tersebut penulis mendapatkan data tentang penyakit yang memiliki tingkat pertumbuhan yang signifikan dalam setiap bulannya adalah penyakit diare. Dari data tersebut terlihat bahwa tingkat pertumbuhan penyakit diare sebesar 332 kasus pada kuartal pertama 2014 mendapatkan peringkat ke 4(empat) dari 10 nominasi data penyakit. Kemudian pada kuartal kedua terjadi peningkatan tajam sebesar 553 kasus yang menempatkan diare sebagai pemuncak. Dilanjutkan pada kuartal ketiga antara bulan juli sampai september terjadi penurunan kasus diare menjadi 366 kasus dan pada kuartal ke-empat diare meningkat lagi menjadi 577 kasus dan bertengger di posisi ke tiga dari sepuluh nominasi. Diakhir tahun didapatkan secara keseluruhan 1808 kasus atau sekitar 14,7% dari total pasien yang mendapatkan pelayanan di Puskesmas Tambakboyo.

Penyakit diare merupakan salah satu penyakit berbasis lingkungan. Penyakit diare masih merupakan masalah kesehatan terbesar di indonesia karena masih buruknya kondisi sanitasi dasar,

# **APLIKASI DATA MINING DISTRIBUSI PENYAKIT DIARE DENGAN METODE ASSOCIATION RULE MENGUNAKAN ALGORITMA A-PRIORI (STUDI KASUS PUSKESMAS TAMBAKBOYO)**

**Sandi Kelana Aji, Desi Purwanti K.**

Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, (024) 3517261

E-mail : sandikelana5@gmail.com, desi.purwanti@dsn.dinus.ac.id

---

## **Abstrak**

Selama tahun 2014, Puskesmas Tambakboyo memiliki data pasien sebanyak 12.399 orang. Di dalam gudang data tersebut terdapat beberapa penyakit yang mendominasi dialami oleh masyarakat diantaranya demam, ispa, diare, tukak lambung dll. Dari data tersebut penulis mendapatkan data tentang penyakit yang memiliki tingkat pertumbuhan yang signifikan dalam setiap bulannya adalah penyakit diare. Dari data tersebut terlihat bahwa tingkat pertumbuhan penyakit diare sebesar 332 kasus pada kuartal pertama 2014 mendapatkan peringkat ke 4(empat) dari 10 nominasi data penyakit. Kemudian pada kuartal kedua terjadi peningkatan tajam sebesar 553 kasus yang menempatkan diare sebagai pemuncak. Dilanjutkan pada kuartal ketiga antara bulan juli sampai september terjadi penurunan kasus diare menjadi 366 kasus dan pada kuartal ke-empat diare meningkat lagi menjadi 577 kasus dan bertengger di posisi ke tiga dari sepuluh nominasi. Diakhir tahun didapatkan secara keseluruhan 1828 kasus atau sekitar 14,7% dari total pasien yang mendapatkan pelayanan di Puskesmas Tambakboyo. dengan besarnya data informasi yang telah dikelola sekian lama terkadang tidak diimbangi dengan penyajian informasi yang memadai. Didalam pengambilan keputusan tidak hanya mengandalkan data oprasional saja akan tetapi data – data tersebut perlu dianalisa untuk mendapatkan informasi mengenai potensi – potensi apa saja yang dapat dimanfaatkan oleh pengambil keputusan. Hal ini mamicu lahirnya suatu cabang keilmuan baru yang berkaitan dengan penggalian informasi dengan metode yang dapat mengolah data dalam jumlah besar yang dinamakan data mining. Berdasarkan permasalahan di atas maka penulis perlunya suatu solusi untuk mengetahui penyebaran penyakit diare. Misalkan mencari pasien yang terkena diare berdasarkan jenis kelamin dan juga pasien terkena diare dari wilayah kelurahan tertentu. Metode yang digunakan yaitu Metode Association Rule Menggunakan Algoritma Apriori Algoritma Apriori mampu memberikan solusi dalam penentuan pola data penyebaran penyakit diare di Puskesmas Tambakboyo. Dengan menyediakan seleksi atau pengelompokan terhadap data keseluruhan penyebaran penyakit diare berdasarkan jenis kelamin, ataupun pengelompokan penyebaran per wilayah. Hal ini akan menunjang kemampuan pengambil keputusan dalam pengambilan keputusan. Penyediaan laporan akan penyebaran penyakit diare di wilayah Tambakboyo juga semakin cepat.

**Kata Kunci** : Data mining Distribusi, Association Rule, Algoritma Apriori

## **Abstract**

During 2014, Tambakboyo health centers have patient data as much as 12 399 people. In the data warehouse, there are several diseases that dominate experienced by people such as fever, ispa, diarrhea, stomach ulcers etc. From these data the authors obtain data on diseases that have significant growth rates in each month is a diarrheal disease. From these data it appears that the growth rate of 332 cases of diarrheal illness in the first quarter of 2014 to get ranked into four (4) out of 10 nominations disease data. Then in the second quarter a sharp increase of 553 cases were put diarrhea as standings. Continued in the third quarter between the months of July to September decreased to 366 cases of diarrhea cases and in the fourth quarter increased to 577 diarrheal cases and perched on position three of the ten nominations. Obtained at the end of the year as a whole in 1828 cases, or about 14.7% of