

Nama = Shaka Rizky Ramadhan

Nim = A11.2019.12214

**Judul jurnal : Using Bayesian Network for Determining The Recipient of Zakat in BAZNAS Pekanbaru**

**Penulis : Akbarizan, Rahmad Kurniawan, Mohd Zakree Ahmad Nazri, Siti Norul Huda Sheikh Abdullah, Sri Murhayati, Nurcahaya.**

**Publikasi : International Conference on Electrical Engineering and Informatics (ICon EEI 2018)**

**Tahun : 2018**

**Reviewer : Siti Qomariyah**

### **Latar belakang :**

Zakat adalah pilar ketiga dari rukun Islam. Muzakki adalah seorang Muslim yang memiliki kualifikasi dan berkewajiban secara finansial untuk memberi zakat. Sedangkan Mustahik adalah orang yang berhak menerima zakat. Tugas pokok Baznas dalam menyalurkan zakat adalah untuk mengutamakan penyaluran zakat berdasarkan data yang akurat. Beberapa program zakat adalah untuk usaha produktif, beasiswa dan pelatihan, konstruksi sarana dan prasarana. Baznas di Pekanbaru pasti segera salurkan zakatnya ke mustahik sesuai ke skala prioritas yang telah disusun dalam program. Islam telah mengatur Mustahik dalam Quran, ayat 60 Surat At-Taubah. Zakat berpotensi mengurangi kesenjangan sosial di antara keduanya Muzakki dan Mustahik . Banyaknya Muzakki di Riau provinsi lebih dari Mustahik. Rasio Mustahik dan Muzakki mencapai 0,56. Khususnya, 100 Muzakki dapat menanggung 56 Mustahik. Secara prosedural, Muzakki melakukan pembayaran zakat masuk Kantor Baznas. Baznas menerima zakat dari Muzakki dan mendistribusikan ke Mustahik . Baznas menentukan Mustahik kandidat secara manual, yaitu mewawancarai kandidat dan melakukan observasi lapangan. Perasaan, emosi, sentimen saling terkait dengan atmosfer dan suasana hati yang dapat menyebabkan salah penilaian mempengaruhi distribusi zakat. Teknik klasifikasi dalam berpikir kecerdasan buatan lebih cepat dari yang manusia pikirkan. Salah satu metode klasifikasi yang populer adalah Bayesian network. Bayesian network dapat memfasilitasi pembelajaran tentang hubungan kausal antara variabel. Bayesian network mudah diubah menjadi alat untuk pendukung keputusan seperti membantu pengelolaan alam sumber daya. Oleh karena itu, penelitian ini menggunakan Bayesian Network sebagai metode klasifikasi dalam menentukan mustahik.

### **Metode**

Metode yang digunakan sesuai dengan proses data mining, mulai dari pendataan dan preprocessing. Data diperoleh dari Baznas Pekanbaru tahun 2017 sebanyak 1000 dengan 14 atribut (nomor urut, distrik, nomor IC, tanggal, nama, alamat, nomor telepon, usia, status

perkawinan, tanggungan, pekerjaan, program, pendapatan dan keputusan ). Gambar berikut menunjukkan Teknik klasifikasi menggunakan Bayesian network :

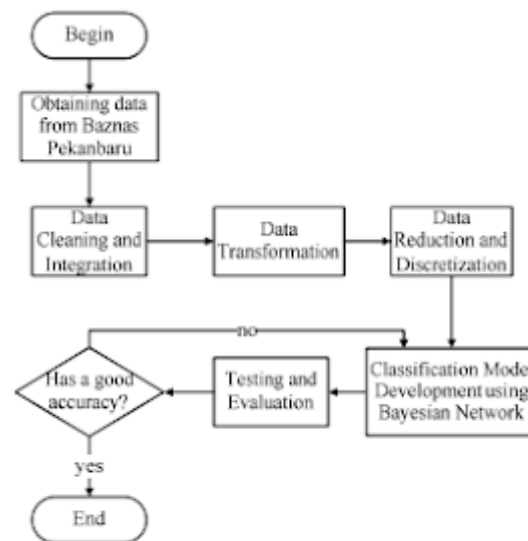


Fig. 1. Methodology of Bayesian Network classification technique

#### a. Data preprocessing

Teknik preprocessing data meliputi pembersihan data, integrasi data, transformasi data dan reduksi data. Penulis melakukan integrasi data dan menggabungkan data dari berbagai sumber. Penulis harus berintegrasi skema, mengidentifikasi entitas, mendeteksi dan menyelesaikan konflik nilai data. tidak menemukan nilai yang hilang dalam data ini. Namun, kami telah menghapus atribut yang tidak berpengaruh pada klasifikasi hasil seperti nomor urut, distrik, program, IC number, tanggal, alamat dan nomor telepon. Langkah terakhir dalam data preprocessing adalah data transformasi dan reduksi data. Penulis menerapkan reduksi data teknik dan diskritisasi untuk mengurangi penyajian Himpunan data. Diskritisasi melibatkan pengurangan jumlah nilai atribut kontinu dengan membagi rentang interval atribut seperti usia dan pendapatan.

- 1) Adult (aged 20 to 30 years)
- 2) Adult (aged 31 to 59 years)
- 3) Adult (aged >60 years)

Berdasarkan peraturan yang berlaku untuk Baznas Pekanbaru, penulis telah mengkategorikan pendapatan sebagai berikut:

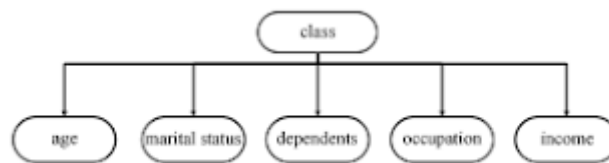
- 1) Income I = Rp. 0 – 500,000
- 2) Income II = Rp. 500,001 – 1,000,000
- 3) Income III = Rp. 1,000,001 – 1,500,000

Sedangkan untuk kelas, otomatis ditentukan oleh mesin menggunakan algoritma Bayesian Network.

- 1) Zakat I = Rp. 1,000,000 – 1,500,000
- 2) Zakat II = Rp. 1,500,001 – 2,000,000
- 3) Zakat III = Rp. 2,000,001 – 2,500,000

b. Classification model

Bayesian network adalah sebuah metode berdasarkan teorema Bayes yang dibuat oleh Thomas Bayes pada tahun 1763. Metode Bayesian network menjadi sangat populer selama dekade terakhir seperti yang digunakan untuk berbagai aplikasi cerdas seperti pembelajaran mesin, teks processing, pemrosesan bioinformatika, diagnostik medis, prakiraan cuaca, dan aplikasi sistem cerdas lainnya. Misalkan  $U = \{\text{umur, status, tanggungan, pekerjaan, pendapatan}\}$  menjadisaatu set variabel. lassifier telah belajar dari kumpulan data zakat yang terdiri darisampel atas ( atribut, kelas zakat ). Gambar berikut menunjukkan struktur Bayesian Network untuk masing-masing node berisi probabilitas.



c. Pengujian dan Evaluasi

Setiap percobaan dilakukan menggunakan pemisahan data set faktor (data pelatihan: data pengujian) dari 10 hingga 90 untuk pemodelan klasifikasi. Parameter yang akan dievaluasikinerja dari masing-masing pemodelan klasifikasi sebagai berikut:

- a.  $Accuracy = \frac{TN+TP}{P+N} = \frac{TP+TN}{TP+TN+FP+FN}$
- b.  $Precision = \frac{TP}{TP+FP}$
- c.  $Recall = \frac{TP}{TP+FN}$
- d.  $F-Measure = 2 \cdot \frac{Precision \cdot Recall}{Precision+ Recall}$

Dimana:

- Kondisi Positif (P): bilangan positif nyatakasus dalam data
- Kondisi Negatif (N): banyaknya negatif nyatakasus dalam data
- True Positive (TP): Ini merujuk pada kasus positif didata dan diberi label dengan benar oleh pengklasifikasi.

- True Negative (TN): Ini adalah kasus negatif diberi label dengan benar oleh pengklasifikasi. Biarkan TN menjadi jumlah negatif sebenarnya.
- False Positive (FP): Ini adalah kasus negatif salah diberi label sebagai positif
- False Negative (FN): Ini adalah kasus positif salah diberi label sebagai negative

Misalkan ada tiga kelas pada set data zakat C1, C2, dan C3, sehingga:

- "TP dari C1" adalah semua instans C1 diklasifikasikan sebagai C1.
- "TN C1" adalah semua instans non-C1 tidak diklasifikasikan sebagai C1.
- "FP dari C1" adalah semua instans non-C1 diklasifikasikan sebagai C1.
- "FN dari C1" adalah semua instans C1 tidak diklasifikasikan sebagai C1.

## Hasil

Dalam jurnal penulis berhasil melakukan preprocessing data dan data zakat set siap untuk diklasifikasikan menggunakan algoritma Bayesian network. Penulis menggunakan Weka versi 3.8.2 sebagai alat untuk pembelajaran mesin yang memiliki kemampuan dalam mengklasifikasikan data menggunakan Bayesian network. Pada percobaan yang telah dilakukan penulis menemukan bahwa Bayesian network menghasilkan akurasi rata-rata yang baik 93.24%. Bayesian network akurat dalam pengumpulan data zakat, bahkan memiliki rata-rata presisi 0,91. Semakin rendah nilai alpha yang digunakan maka tingkat akurasi akan semakin meningkat.

## Kesimpulan

Penelitian yang telah dilakukan penulis telah berhasil mengklasifikasikan siapa yang berhak menerima zakat ( Mustahik ). Penentuan mustahik sangat penting mendistribusikan zakat dengan benar dengan menggunakan metode Bayesian Network. Bayesian network menghasilkan rata-rata klasifikasi yang memiliki akurasi 93,24%. F-Measures telah menunjukkan kemampuan Bayesian network untuk mencakup distribusi kelas yang tidak seimbang.

on [Desember 25, 2020](#) [Tidak ada komentar:](#)

[Kirimkan Ini lewat Email](#) [BlogThis!](#) [Berbagi ke Twitter](#) [Berbagi ke Facebook](#) [Bagikan ke Pinterest](#)

**Jumat, 11 Desember 2020**

Judul jurnal : Automatic Rule Generator via FP-Growth for Eye Diseases Diagnosis

Penulis : Rahmad Kurniawan, Mohd Zakree Ahmad Nazri, Siti Norul Huda Sheikh Abdullah, Jemaima Che Hamzah, Rado Yendra, Westi Oktaviana

Publikasi : international journal on advanced science engineering information technology

Tahun : 2019

Reviewer : Siti Qomariyah

## REVIEW

### **Latar belakang penelitian :**

Algoritma FP-Growth dapat memperbaiki kekurangan dari algoritma Apriori. Proses pencarian set item yang sering mereka lakukan adalah unik dan beragam. Saat ini, FP-Growth adalah salah satu algoritma tercepat di antara aturan asosiasi. Pertumbuhan Pola yang Sering adalah salah satu metode alternatif untuk menentukan kumpulan item yang paling sering dalam suatu kumpulan data. Fitur utama dari algoritma FP-Growth adalah FP-Tree diperlakukan sebagai struktur datanya. FP-Growth adalah teknik pembelajaran mesin yang digunakan untuk memperoleh aturan dari catatan diagnosis penyakit mata yang dikumpulkan dari Rumah Sakit Sumatera Eye Center (SMEC) di Pekanbaru, Riau, Indonesia. Sistem yang dikembangkan diuji dengan 17 kasus. Dokter mata memeriksa hasil dari generator aturan otomatis untuk diagnosis penyakit mata. Penulis menemukan bahwa pengenalan aturan asosiasi FP-Growth ke dalam sistem berbasis pengetahuan penyakit mata, mampu menghasilkan hasil diagnosa mata yang dapat diterima dan menjanjikan sekitar 88% dari tingkat akurasi rata-rata.

### **Tujuan penelitian :**

Penelitian dalam jurnal bertujuan untuk mensubstitusi metode perolehan pengetahuan tradisional menjadi akuisisi pengetahuan otomatis berdasarkan metode FP Growth dengan menggunakan rekam medis historis.

### **Materials :**

Penulis memperoleh data yang penyakit mata dari Rekam Medis di Pusat Mata Sumatera (SMEC) Rumah Sakit Pekanbaru. Rekam medis dikumpulkan hanya untuk tahun 2016 yang berisi 1.600 pasien dengan enam atribut (No Rekam Medis, Tanggal, Jenis Kelamin, Usia, Gejala, dan Diagnosis).

### **Metode :**

Ada 4 langkah dalam mendesain percobaan menggunakan algoritma FP-Growth

1. Seleksi

Penulis melakukan pemilihan atribut dengan menghapus atribut yang tidak perlu. Kami menemukan bahwa data rekam medis di rumah sakit tersebut disimpan dalam bentuk manual. Data medis penyakit mata manual tersebut kemudian kami digitalkan dengan menggunakan alat pengolah data. Penulis mengecualikan *nomor rekam medis*, *nama*, dan *tanggal*, dari enam atribut formulir laporan medis pasien seperti *jenis kelamin*, *usia*, *gejala* dan *diagnosis* penyakit. Selanjutnya menemukan hubungan antara usia, jenis kelamin, gejala dan penyakit.

## 2. Pre-processing

Proses selanjutnya adalah pre-processing data dimana kita menghilangkan noise atau membersihkan nilai data yang hilang, data yang tidak konsisten dan outlier. Penulis menggunakan beberapa teknik untuk mengatasi nilai yang hilang, seperti mengganti nilai yang hilang dengan nilai rata-ratanya dan menghapus baris dengan lebih dari satu nilai yang hilang. Dalam kasus atribut usia, penulis berasumsi bahwa nilai yang hilang didistribusikan secara mirip dengan nilai yang ada. Teknik ini disebut Missing Completely at Random (MCAR). Karena nilai yang hilang dalam kumpulan data penyakit mata, kami menghapus catatan dengan nilai yang hilang.

## 3. Transformasi data

Tahap selanjutnya adalah transformasi data dengan mengubah dan menyimpan data penyakit mata ke dalam format khusus, yaitu format file relasi-atribut. Pada tahap ini, kita menginisialisasi atribut dan memodifikasi tipe datanya, seperti konversi tipe data numerik menjadi binomial.

## 4. Menghasilkan aturan menggunakan algoritma FP-Growth

Penulis mengilustrasikan pada kasus 20 sampel data rekam medis penyakit mata untuk mengetahui hubungan antara umur, jenis kelamin, gejala dan jenis penyakit mata. Dalam sampel ini, penulis mengasumsikan bahwa jumlah dukungan  $\geq 10\%$ . Nilai jumlah dukungan ini mempengaruhi konstruksi pohon FP. Dukungan menghitung 10% sebagai ambang batas untuk memfilter set item, yaitu hanya set item yang memiliki jumlah dukungan sama, atau lebih dari 10% muncul. Langkah selanjutnya adalah memindai data dan menata ulang data dengan dukungan minimum. Nilai kepercayaan dan dukungan adalah parameter untuk mengukur aturan terbaik. Data tersebut tidak memenuhi minimum support dan tidak mengandung anteseden dan akibatnya akan dihapus dari transaksi. Atribut yang menjadi anteseden adalah *umur*, *jenis kelamin*, *gejala*, sedangkan akibatnya adalah *penyakit*. Langkah selanjutnya adalah pembuatan FP-Tree berdasarkan transaksi yang dipindai. Berdasarkan data sampel yang diberikan, P1, P2, P3, P4, P5, P6, P7, P8 dan P9 adalah konsekuensi, tetapi hanya P3 dan P4

yang memenuhi dukungan minimum. Langkah selanjutnya adalah mendapatkan frequent item set dengan membuat basis pola kondisional kemudian membuat aturan dengan menghitung kepercayaan dari setiap kombinasi aturan. Setelah aturan didapat dari frequent itemset, maka langkah selanjutnya adalah mengevaluasi aturan asosiasi kuat yang diperoleh dengan menggunakan lift ratio. Rasio peningkatan adalah rasio antara keyakinan suatu aturan dan nilai keyakinan tolok ukur. Benchmark Confidence merupakan perbandingan antara jumlah konsekuensi dengan total transaksi.

### **Hasil :**

Pengujian aturan asosiasi dilakukan dengan dua cara: pengujian berdasarkan dukungan minimum, keyakinan minimum, dan rasio peningkatan. Penulis telah menguji aturan yang diperoleh secara otomatis oleh FP-Growth dengan pengetahuan ahli. Berdasarkan tes penerimaan dokter mata, akurasi rata-rata algoritma aturan asosiasi FP-Growth sebagai aturan generator otomatis adalah 88%.

### **Kesimpulan :**

Penulis telah berhasil melakukan penelitian menggunakan aturan asosiasi sebagai dasar pengetahuan dalam sistem pakar penyakit mata dengan menggunakan data rekam medis dari Rumah Sakit Sumatera Eye Center (SMEC) di Pekanbaru, Riau, Indonesia. Berdasarkan eksperimen yang dilakukan, penentuan dukungan minimum dan keyakinan minimum dalam FP-Growth harus sesuai dengan hasil aturan. Kami menetapkan dukungan minimum 3% dan keyakinan minimum 90% untuk mencapai 2208 aturan. Kami telah berhasil menggunakan algoritma FP-Growth sebagai pembuat aturan dengan akurasi 88%. Diagnosis penyakit mata harus menggunakan aturan yang akurat dan dapat diterima. Dokter mata memeriksa hasil dari generator aturan otomatis. Pada penelitian ini didapatkan aturan 6, 9, 16 dan 17 yang memiliki kesesuaian antara Confidence Level of optalmologists dan nilai Lift Ratio. Penulis menetapkan dukungan minimum 3% dan keyakinan minimum 90% untuk mencapai 2208 aturan. Algoritma FP-Growth dengan cepat membuat aturan dalam data rekam medis yang hanya melakukan scan database dua kali. FP-growth membangun FP-tree yang sangat terkompresi, yang secara substansial mengurangi database aslinya. Berdasarkan hasil pengujian didapatkan bahwa *Konjungtivitis* dan *Presbiopia* penyakit paling dominan diderita dari 225 dan 167 kasus. Pria atau wanita berusia 31-59 tahun sangat *Presbiopia*. Ada tiga gejala kritis pada presbiopia: kabur penglihatan (dekat), pusing, dan rasa tidak nyaman pada mata, sedangkan untuk konjungtivitis ada empat gejala yang mempengaruhi: berair mata, mata gatal, mata lengket, dan mata merah, tetapi paling gejala yang penting adalah mata lengket. FP-Growth dapat menghasilkan aturan sebagai basis pengetahuan dalam pakar berbasis aturan sistem.

Algoritma ini dianggap sebagai teknik yang cepat dalam akuisisi pengetahuan, bahkan menggantikan tradisional akuisisi pengetahuan.

on [Desember 11, 2020](#) [Tidak ada komentar:](#)

[Kirimkan Ini lewat Email](#)[BlogThis!](#)[Berbagi ke Twitter](#)[Berbagi ke Facebook](#)[Bagikan ke Pinterest](#)

### **Tugas Data Mining**

**Judul jurnal : Automatic Rule Generator via FP-Growth for Eye Diseases Diagnosis**

**Penulis : Rahmad Kurniawan, Mohd Zakree Ahmad Nazri, Siti Norul Huda Sheikh Abdullah, Jemaima Che Hamzah, Rado Yendra, Westi Oktaviana**

**Publikasi : International Journal on Advanced Science Engineering Information Technology**

**Tahun : 2019**

**ISSN : 2088-5334**

**Reviewer : Siti Qomariyah**

---

### **Latar belakang penelitian**

Algoritma FP-Growth dapat memperbaiki kekurangan dari algoritma Apriori. Proses pencarian set item yang sering mereka lakukan adalah unik dan beragam. Saat ini, FP-Growth adalah salah satu algoritma tercepat di antara aturan asosiasi. Pertumbuhan Pola yang Sering adalah salah satu metode alternatif untuk menentukan kumpulan item yang paling sering dalam suatu kumpulan data. Fitur utama dari algoritma FP-Growth adalah FP-Tree diperlakukan sebagai struktur datanya. FP-Growth adalah teknik pembelajaran mesin yang digunakan untuk memperoleh aturan dari catatan diagnosis penyakit mata yang dikumpulkan dari Rumah Sakit Sumatera Eye Center (SMEC) di Pekanbaru, Riau, Indonesia. Sistem yang dikembangkan diuji dengan 17 kasus. Dokter mata memeriksa hasil dari generator aturan otomatis untuk diagnosis penyakit mata. Penulis menemukan bahwa pengenalan aturan asosiasi FP-Growth ke dalam sistem berbasis pengetahuan penyakit mata, mampu menghasilkan hasil diagnosa mata yang dapat diterima dan menjanjikan sekitar 88% dari tingkat akurasi rata-rata.

### **Tujuan penelitian**



Penelitian dalam jurnal bertujuan untuk mensubstitusi metode perolehan pengetahuan tradisional menjadi akuisisi pengetahuan otomatis berdasarkan metode FP Growth dengan menggunakan rekam medis historis.

## **Materials**

Penulis memperoleh data yang penyakit mata dari Rekam Medis di Pusat Mata Sumatera (SMEC) Rumah Sakit Pekanbaru. Rekam medis dikumpulkan hanya untuk tahun 2016 yang berisi 1.600 pasien dengan enam atribut (No Rekam Medis, Tanggal, Jenis Kelamin, Usia, Gejala, dan Diagnosis).

## **Metode Penelitian**

Ada 4 langkah dalam mendesain percobaan menggunakan algoritma FP-Growth

### **1. Seleksi**

Penulis melakukan pemilihan atribut dengan menghapus atribut yang tidak perlu. Kami menemukan bahwa data rekam medis di rumah sakit tersebut disimpan dalam bentuk manual. Data medis penyakit mata manual tersebut kemudian kami digitalkan dengan menggunakan alat pengolah data. Penulis mengecualikan *nomor rekam medis*, *nama*, dan *tanggal*, dari enam atribut formulir laporan medis pasien seperti *jenis kelamin*, *usia*, *gejala* dan *diagnosis* penyakit. Selanjutnya menemukan hubungan antara usia, jenis kelamin, gejala dan penyakit.

### **2. Pre-processing**

Proses selanjutnya adalah pre-processing data dimana kita menghilangkan noise atau membersihkan nilai data yang hilang, data yang tidak konsisten dan outlier. Penulis menggunakan beberapa teknik untuk mengatasi nilai yang hilang, seperti mengganti nilai yang hilang dengan nilai rata-ratanya dan menghapus baris dengan lebih dari satu nilai yang hilang. Dalam kasus atribut usia, penulis berasumsi bahwa nilai yang hilang didistribusikan secara mirip dengan nilai yang ada. Teknik ini disebut Missing Completely at Random (MCAR). Karena nilai yang hilang dalam kumpulan data penyakit mata, kami menghapus catatan dengan nilai yang hilang.

### **3. Transormasi data**

Tahap selanjutnya adalah transformasi data dengan mengubah dan menyimpan data penyakit mata ke dalam format khusus, yaitu format file relasi-atribut. Pada tahap ini, kita

menginisialisasi atribut dan memodifikasi tipe datanya, seperti konversi tipe data numerik menjadi binominal.

#### 4. Menghasilkan aturan menggunakan algoritma FP-Growth

Penulis mengilustrasikan pada kasus 20 sampel data rekam medis penyakit mata untuk mengetahui hubungan antara umur, jenis kelamin, gejala dan jenis penyakit mata. Dalam sampel ini, penulis mengasumsikan bahwa jumlah dukungan  $\geq 10\%$ . Nilai jumlah dukungan ini mempengaruhi konstruksi pohon FP. Dukungan menghitung 10% sebagai ambang batas untuk memfilter set item, yaitu hanya set item yang memiliki jumlah dukungan sama, atau lebih dari 10% muncul. Langkah selanjutnya adalah memindai data dan menata ulang data dengan dukungan minimum. Nilai kepercayaan dan dukungan adalah parameter untuk mengukur aturan terbaik. Data tersebut tidak memenuhi minimum support dan tidak mengandung anteseden dan akibatnya akan dihapus dari transaksi. Atribut yang menjadi anteseden adalah *umur*, *jenis kelamin*, *gejala*, sedangkan akibatnya adalah *penyakit*. Langkah selanjutnya adalah pembuatan FP-Tree berdasarkan transaksi yang dipindai. Berdasarkan data sampel yang diberikan, P1, P2, P3, P4, P5, P6, P7, P8 dan P9 adalah konsekuensi, tetapi hanya P3 dan P4 yang memenuhi dukungan minimum. Langkah selanjutnya adalah mendapatkan frequent item set dengan membuat basis pola kondisional kemudian membuat aturan dengan menghitung kepercayaan dari setiap kombinasi aturan. Setelah aturan didapat dari frequent itemset, maka langkah selanjutnya adalah mengevaluasi aturan asosiasi kuat yang diperoleh dengan menggunakan lift ratio. Rasio peningkatan adalah rasio antara keyakinan suatu aturan dan nilai keyakinan tolok ukur. Benchmark Confidence merupakan perbandingan antara jumlah konsekuensi dengan total transaksi.

### Hasil Penelitian

Pengujian aturan asosiasi dilakukan dengan dua cara: pengujian berdasarkan dukungan minimum, keyakinan minimum, dan rasio peningkatan. Penulis telah menguji aturan yang diperoleh secara otomatis oleh FP-Growth dengan pengetahuan ahli. Berdasarkan tes penerimaan dokter mata, akurasi rata-rata algoritma aturan asosiasi FP-Growth sebagai aturan generator otomatis adalah 88%.

### Kesimpulan

Penulis telah berhasil melakukan penelitian menggunakan aturan asosiasi sebagai dasar pengetahuan dalam sistem pakar penyakit mata dengan menggunakan data rekam medis dari Rumah Sakit Sumatera Eye Center (SMEC) di Pekanbaru, Riau, Indonesia. Berdasarkan eksperimen yang dilakukan, penentuan dukungan minimum dan keyakinan minimum dalam FP-Growth harus sesuai dengan hasil aturan. Penulis menetapkan penentuan dukungan

minimum dan keyakinan minimum dalam FP-Growth harus sesuai dengan hasil aturan. Penulis menetapkan dukungan minimum 3% dan keyakinan minimum 90% untuk mencapai 2208 aturan. Penulis telah berhasil menggunakan algoritma FP-Growth sebagai pembuat aturan dengan akurasi 88%. Diagnosis penyakit mata harus menggunakan aturan yang akurat dan dapat diterima. Dokter mata memeriksa hasil dari generator aturan otomatis. Pada penelitian ini didapatkan aturan 6, 9, 16 dan 17 yang memiliki kesesuaian antara Confidence Level of optalmologists dan nilai Lift Ratio. Penulis menetapkan dukungan minimum 3% dan keyakinan minimum 90% untuk mencapai 2208 aturan. Algoritma FP-Growth dengan cepat membuat aturan dalam data rekam medis yang hanya melakukan scan database dua kali. FP-growth membangun FP-tree yang sangat terkompresi, yang secara substansial mengurangi database aslinya. Berdasarkan hasil pengujian didapatkan bahwa *Konjungtivitis* dan *Presbiopia* penyakit paling dominan diderita dari 225 dan 167 kasus. Pria atau wanita berusia 31-59 tahun sangat *Presbiopia*. Ada tiga gejala kritis pada presbiopia: kabur penglihatan (dekat), pusing, dan rasa tidak nyaman pada mata, sedangkan untuk konjungtivitis ada empat gejala yang mempengaruhi: berair mata, mata gatal, mata lengket, dan mata merah, tetapi paling gejala yang penting adalah mata lengket. FP-Growth dapat menghasilkan aturan sebagai basis pengetahuan dalam pakar berbasis aturan sistem. Algoritma ini dianggap sebagai teknik yang cepat dalam akuisisi pengetahuan, bahkan menggantikan tradisional akuisisi pengetahuan.

JUDUL	<b>Aplikasi Data Mining untuk Menampilkan Tingkat Kelulusan Mahasiswa dengan Algoritma Apriori</b>
PENULIS	Benni R Siburian
TAHUN	2014
JURNAL & VOL	<i>ISSN (2301-9425) Volume VII – No. 2, Agustus 2014</i>
REVIEWER	Muhammad Andri Wardika (1441177004174)

### **Abstrak :**

Pertumbuhan yang pesat dari akumulasi data telah menciptakan kondisi kaya akan data tapi minim informasi. Data mining merupakan penambangan atau penemuan informasi baru dengan mencari pola aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut.

Dengan memanfaatkan data induk mahasiswa dan data kelulusan mahasiswa, diharapkan dapat menghasilkan informasi tentang tingkat kelulusan dengan data induk mahasiswa melalui teknik data mining.

Kategori tingkat kelulusan di ukur dari lama studi dan IPK. Algoritma yang digunakan adalah algoritma apriori, informasi yang ditampilkan berupa nilai support dan confidence dari masing-masing kategori tingkat kelulusan.

## 1. **INTRODUCTION**

Data Mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Patut diingat bahwa kata mining sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar.

Tujuan dari penelitian ini adalah untuk menerapkan konsep Data Mining untuk menampilkan tingkat kelulusan mahasiswa dan juga menerapkan Algoritma Apriori untuk menampilkan hal tersebut sekaligus merancang Aplikasi dengan Data Mining.

Tujuan dari penggunaan teknik Algoritma Apriori adalah algoritma paling terkenal untuk menemukan pola frekuensi tinggi. Pola frekuensi tinggi adalah polapola item di dalam suatu database yang memiliki frekuensi atau support di atas ambang batas tertentu yang disebut dengan istilah minimum support. Pola frekuensi tinggi ini digunakan untuk menyusun aturan asosiatif dan juga beberapa teknik data mining lainnya.

## 2. **LITERATURE REVIEW**

Han menjelaskan Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam database, data warehouse, atau penyimpanan informasi lainnya. Data mining berkaitan dengan bidang ilmu – ilmu lain, seperti database system, data warehousing, statistik, machine learning, information retrieval, dan komputasi tingkat tinggi. Selain itu, data mining didukung oleh ilmu lain seperti neural network, pengenalan pola, spatial data analysis, image database, signal processing.

Davies membagi karakteristik data mining sebagai berikut :

- a. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
- b. Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
- c. Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi.

Han juga menjelaskan beberapa tahapan pada data mining, tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau perantaraan knowledge base. Dan berikut adalah tahap-tahap dalam data mining : Pembersihan data (data cleaning), Integrasi data (data integration), Seleksi Data (Data Selection), Transformasi data (Data Transformation), Proses mining, Evaluasi pola (pattern evaluation), Presentasi pengetahuan (knowledge presentation).

## 3. **METODOLOGI**

Ada beberapa algoritma yang sudah dikembangkan mengenai aturan asosiasi, namun ada satu algoritma klasik yang sering dipakai yaitu algoritma apriori. Ide dasar dari algoritma ini adalah dengan mengembangkan frequent itemset. Dengan menggunakan satu item dan secara rekursif mengembangkan frequent itemset dengan dua item, tiga item dan seterusnya hingga frequent itemset dengan semua ukuran.

Untuk mengembangkan frequent set dengan dua item, dapat menggunakan frequent set item. Alasannya adalah bila set satu item tidak melebihi support minimum, maka sembarang ukuran itemset yang lebih besar tidak akan melebihi support minimum tersebut. Secara umum,

mengembangkan set dengan fc-item menggunakan frequent set dengan  $k - 1$  item yang dikembangkan dalam langkah sebelumnya. Setiap langkah memerlukan sekali pemeriksaan ke seluruh isi database.

Dari jumlah besar aturan yang mungkin dikembangkan, perlu memiliki aturanaturan yang cukup kuat tingkat ketergantungan antar item dalam antecedent dan consequent. Untuk mengukur kekuatan aturan asosiasi ini, digunakan ukuran support dan confidence. Support adalah rasio antara jumlah transaksi yang memuat antecedent dan consequent dengan jumlah transaksi. Confidence adalah rasio antara jumlah transaksi yang meliputi semua item dalam antecedent dan consequent dengan jumlah transaksi yang meliputi semua item dalam antecedent.

$S =$

Keterangan :

$S =$  Support

$(+)$  = Jumlah transaksi yang mengandung antecedent dan consequent

$\Sigma()$  = Jumlah transaksi

$C =$

$C =$  Confidence

$(+)$  = Jumlah transaksi yang mengandung antecedent dan consequent

$\Sigma()$  = Jumlah transaksi yang mengandung antecedent.

## **KESIMPULAN**

Algoritma apriori ini dapat diimplementasikan dalam perancangan aplikasi menggunakan Data Mining. Metode yang diusulkan untuk membuat merancang aplikasi data mining menggunakan aplikasi Microsoft Visual Basic Net 2008 dapat mudah dipahami oleh programmer bahkan yang pemula. Dan metode ini dapat menampilkan informasi tingkat kelulusan secara efektif

LITERATUR REVIEW : PREDIKSI KELULUSAN MAHASISWA DENGAN MENGGUNAKAN ILMU DATA MINING Yobella Anjelika Jurusan Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya email: yobellaanjelika@mhs.eng.upr.ac.id Abstrak Pendidikan adalah pembelajaran pengetahuan, keterampilan, dan kebiasaan sekelompok orang yang diturunkan dari satu generasi ke generasi berikutnya melalui pengajaran, pelatihan, atau penelitian. Pendidikan sering terjadi di bawah bimbingan orang lain, tetapi juga memungkinkan secara otodidak. Perguruan tinggi memberikan pendidikan terbaik untuk mahasiswanya. Setiap perguruan tinggi memiliki fakultas dengan sistem pendidikan yang berbeda. Pemodelan prediksi kelulusan mahasiswa dibutuhkan demi mengetahui apakah sistem pembelajaran yang dilakukan sudah cukup baik atau tidak. Penelitian ini berkaitan dengan systematic literature review untuk mencari pemodelan prediksi menggunakan metode-

metode klasifikasi data mining. Hasil penelitian berupa hasil akurasi dari prediksi dengan menggunakan metode klasifikasi data mining dan didukung dengan pemilihan data yang relevan dengan faktor kelulusan mahasiswa. Kata kunci : Pendidikan, Prediksi kelulusan, Data mining

Abstract Education is the learning of knowledge, skills and habits of a group of people that are passed down from one generation to the next through teaching, training, or research. Education often takes place under the guidance of others, but is also possible on a self-taught basis. Higher education provides the best education for its students. Each college has a faculty with a different education system. Modeling the prediction of student graduation is needed in order to find out whether the learning system carried out is good enough or not. This research is related to systematic literature review to find predictive modeling using data mining classification methods. The results of the research are the results of the accuracy of predictions using the data mining classification method and are supported by the selection of data relevant to student graduation factors. Keywords : Education, Prediction of graduation, Data mining

1. Pendahuluan Pendidikan merupakan aspek penting dalam suatu kehidupan. Dalam dunia pendidikan akan banyak sekali ditemui ilmu-ilmu pengetahuan yang berguna nantinya untuk masa depan. Ilmu-ilmu tersebut dipelajari dan dipahami sehingga didapat suatu manfaat yang berguna dalam menyelesaikan suatu masalah. Ilmu-ilmu tersebut bisa didapatkan salah satunya dijenjang pendidikan perguruan tinggi seperti universitas, akademi, sekolah tinggi, institut, politeknik dan lain-lain. Setiap fakultas memiliki sistem kurikulum pendidikan yang berbeda dalam mengontrol ilmu pengetahuan, perilaku mahasiswa, dan prestasi mahasiswanya. Setiap sistem kurikulum yang diterapkan bertujuan untuk menciptakan mahasiswa dengan lulusan terbaik guna meningkatkan akreditasi fakultas maupun perguruan tinggi tersebut dan juga jumlah kelulusan menjadi salah satu tolak ukur penilaian kesuksesan dalam proses belajar mengajar dalam perguruan tinggi. Namun, mengingat salah satu faktor yang membuat mahasiswa tidak lulus tepat waktu ialah metode pembelajaran yang kurang tepat sehingga mahasiswa lebih sulit mencerna ilmu baru yang diberikan ditambah beban SKS yang diberikan kepada mahasiswa. Model prediksi yang diinginkan berdasarkan hasil penelitian sebelumnya dimana faktor apa saja yang membuat mahasiswa tersebut lulus tepat waktu atau tidak. Dan metode apa yang layak dan akurat untuk digunakan dalam memprediksi ketepatan kelulusan mahasiswa.

2. Metode Penelitian

Penulisan literature review ini berdasarkan 4 jurnal yang berbeda, salah satunya yaitu Dosen Fakultas Teknik Universitas Palangka Raya dan 3 jurnal lainnya yang didapat dari internet. Berdasarkan jurnal-jurnal tersebut maka dilakukan identifikasi dengan judul yaitu : Prediksi Kelulusan Mahasiswa Menggunakan Ilmu Data Mining yang dimana sumber data penelitian ini diambil dari beberapa jurnal yang diperoleh melalui internet yaitu berupa hasil penelitian mengenai metode prediksi ketepatan kelulusan dari beberapa universitas di Indonesia yang dipublikasikan di internet. Metodologi penelitian yang dapat dipakai ialah menggunakan metodologi data mining CRISP-DM. Pada tahap ini, akan dicari data mahasiswa apa saja yang relevan dalam menentukan kelulusan mahasiswa tepat waktu atau tidak. Di bawah ini pada tabel 1 menjelaskan atribut apa saja yang memungkinkan dipakai dalam penelitian sebelum melakukan sebuah pengujian diharapkan untuk mengumpulkan data mahasiswa yang akan di uji. Metode yang

dapat digunakan seperti metode klasifikasi yaitu naïve bayes, k-nearest neighbor, support vector machine, C4.5, C5.0, decision tree, bagging CART, regresi logistik biner, chaid regression tree dan lain sebagainya. Tujuan dari penelitian ini untuk melihat sejauh mana pemodelan prediksi terhadap ketepatan kelulusan mahasiswa di suatu Universitas. Data akan berisi sebuah atribut yang menentukan mahasiswa tersebut dapat lulus tepat waktu atau tidak.

3. Hasil dan Pembahasan Pada tahap ini menjelaskan atribut yang digunakan atau yang memungkinkan untuk digunakan dalam penelitian. Tabel 1. Atribut Data Mahasiswa

No	Atribut	Tipe Data
1	Nama	Nominal
2	NIM	Numerik
3	Jurusan	Nominal
4	Penghasilan orang tua	Numerik
5	Status pekerjaan orang tua	Nominal
6	IPS Semester 1-7	Numerik
7	SKS Semester 1-7	Numerik
8	IPK	Numerik
9	Jenis tugas akhir	Nominal
10	Lama skripsi	Nominal
11	Status kelulusan / prediksi	Nominal

Pada tabel 1 menjelaskan atribut apa saja yang relevan untuk mencari prediksi ketepatan kelulusan mahasiswa. Dari semua atribut yang berjumlah 11, dipilih menyesuaikan kebutuhan penelitian yang dilakukan pada perguruan tinggi negeri atau swasta, jurusan yang ingin diteliti, mahasiswa yang menjadi target penelitian apakah lelaki saja atau wanita saja atau keduanya. Jika sudah mendapatkan data dan juga sudah diketahui kebutuhannya maka tahap selanjutnya adalah menentukan metode prediksi yang akan dipakai dalam pengolahan data yang sudah didapatkan sebelumnya. Tabel 2. Hasil Akurasi Prediksi Dengan Berbagai Metode dan Atribut

No	Penulis	Judul	Atribut yang dipakai	Metode	Hasil Akurasi
1	Agus Sehatman Saragih, Ade Chandra Saputra	Rancang Bangun Sistem Prediksi Kelulusan Mahasiswa Jurusan Teknik Informatika Universitas Palangka Raya	Jenis kelamin, agama, SKS 1-3, IPS 1-3, Jenis Tugas Akhir, Lulus semester C4.5	Dengan hasil akurasi 60%	
2	Indah Puji Astuti	Prediksi Ketepatan Waktu Kelulusan Dengan Algoritma Data Mining C4.5	NIM, Jenis sekolah, asal, asal daerah, pekerjaan orang tua, kelas C4.5	Dengan hasil akurasi 82%	
3	Gusmelia Testiana	Perancangan Model Prediksi Kelulusan Mahasiswa Tepat Waktu pada UIN Raden Fatah IP Semester, IPK, Prediksi/status Naïve Bayes	Dengan hasil akurasi 82.08%		
4	Ketut Putu Suniantara, Muhammad Rusli	Ketepatan Klasifikasi Bagging CART Pada Klasifikasi Ketidaktepatan Waktu Kelulusan Mahasiswa STIKOM Bali	Status Kelulusan, Jenis Kelamin, Jurusan, Lama Skripsi IPK, IPS Semester 6, Nilai Ujian Masuk CART	Dengan hasil akurasi 86.18%	

Beberapa metode yang dipakai untuk mengolah data mahasiswa sehingga mampu memberikan suatu prediksi ketetapan kelulusan, memberikan nilai akurasi yang berbeda beda. Hal ini disebabkan pemilihan atribut yang dipakai yang menjadi faktor dari kelulusan mahasiswa.

4. Kesimpulan Hasil kesimpulan pada literatur review adalah, dalam menentukan prediksi ketepatan kelulusan dapat menggunakan metode klasifikasi, pada setiap metode akan memberikan hasil prediksi dan akurasi yang berbeda-beda. Semakin baik hasil nilai akurasi dengan pemilihan atribut yang benar maka dapat menentukan faktor yang menyebabkan seorang mahasiswa dapat lulus tepat waktu atau tidak.

5. Saran Artikel ini merupakan hasil dari telaah literatur yang didapat dari penelitian sebelumnya dan hasil literatur review ini adapun saran dari hasil penelitian, perancangan, pembuatan, dan pengujian sistem prediksi kelulusan siswa dengan menggunakan berbagai metode klasifikasi yang ada pada data mining naïve bayes, C4.5, dan bagging CART. Untuk penelitian selanjutnya diharapkan dalam dapat dilakukan perbandingan dengan

memanfaatkan metode klasifikasi lainnya. Dengan begitu dapat diketahui kelebihan masingmasing metode dan metode mana yang lebih baik dari keakuratan nilai y