

## DESCRIPTIVE MODELLING UNTUK KETERJANGKITAN PENYAKIT DI KABUPATEN LAMONGAN DENGAN MENGUNAKAN MULTIDIMENSIONAL FUZZY ASSOCIATION RULE MINING

Retno Wardhani<sup>1</sup> dan Gunawan<sup>2</sup>

Teknik Informatika Universitas Islam Lamongan<sup>1</sup>

Teknik Informatika Sekolah Tinggi Teknik Surabaya<sup>2</sup>

retzno@yahoo.com dan gunawan@stts.edu

### ABSTRAK

Penelitian ini menitikberatkan pada masalah untuk mengetahui bagaimana hubungan antara data-data yang ada pada pasien dengan penyakit yang dideritanya. Data berasal dari data poliklinik RSUD Dr. Soegiri Kabupaten Lamongan dan data cuaca dari BMKG Surabaya untuk wilayah Lamongan dan sekitarnya. *Data preprocessing* yang dilakukan pada data asli mengurangi data hingga 60% dari jumlah awal. Selain itu juga merubah bentuk database yang awalnya masih menggunakan Microsoft Excel ke dalam database MySQL. Sedangkan data transformation akan menggabungkan database relational ke dalam satu pusat data yang berupa tabel, dimana menghasilkan lima dimensi pada data modelnya. Proses proses data mining menggunakan algoritma *multidimensional fuzzy association rule mining*. Algoritma ini akan mencari nilai *fuzzy mapping* dan menghitung nilai *support* itemset-nya.

Dari hasil implementasi pada sistem, didapatkan pemetaan beberapa penyakit dengan nilai confidence dan nilai support tertentu. Tiga penyakit dengan kemunculan tertinggi pada rule(s) yaitu, penyakit Osteo Arthristis muncul dengan nilai confidence berkisar antara 27% - 40% dan nilai support berkisar antara 6% - 19%, penyakit TB Paru muncul dengan nilai confidence berkisar antara 20% - 25% dan nilai support berkisar antara 4% - 21%, dan penyakit DM Tak Tergantung Insulin muncul dengan nilai confidence berkisar antara 20% - 23% dan nilai support berkisar antara 3% - 9%.

Kata kunci: *data preprocessing, multidimensional fuzzy association rule mining*

### ABSTRACT

*This study focuses on the problem of how to know the relationship between the data available in patients with the disease. The data comes from Polyclinic Hospital of Dr Soegiri Lamongan and weather data from BMKG Surabaya for Lamongan and the surrounding area. Data preprocessing reduces the original data up to 60% of the initial amount. It also changed the form of a database that initially still use Microsoft Excel to MySQL database. While the data transformation will incorporate relational database into a data center in the form of tables, which resulted in five dimensions in the data model. The process of data mining process using algorithms of multidimensional fuzzy association rule mining. This algorithm will find the value of fuzzy mapping and calculate its value itemset support.*

*From the results of the implementation of the system, obtained mapping several diseases with a value of confidence and support values specified. Three diseases with the highest occurrence of the rule(s) that are, the disease Osteo Arthristis appear with confidence values ranging between 27% - 40% and the value of the support ranged between 6% - 19%, pulmonary TB disease appears with confidence values ranging between 20% - 25 % and the value of support ranged from 4% - 21%, and DM Not Depending Insulin appears with confidence values ranging between 20% - 23% and the value of the support ranges from 3% - 9%.*

*Keywords: data preprocessing, multidimensional fuzzy association rule mining*

## I. PENDAHULUAN

Perubahan iklim sekarang ini berdampak luas pada keadaan air, ketinggian permukaan laut, dan ekosistem tanaman dan hewan. Selain dampak iklim pada kesehatan, letak geografis suatu tempat tinggal dengan berbagai keadaan kultur dan gaya hidup seseorang juga mempunyai pengaruh terhadap beberapa keterjangkitan penyakit.

Penelitian yang ingin dilakukan disini adalah apakah iklim khususnya cuaca, letak, dan beberapa faktor lain mempunyai hubungan terhadap keterjangkitan suatu penyakit. Dengan memanfaatkan proses data mining, diharapkan data yang ada dapat digali semaksimal mungkin untuk menghasilkan informasi-informasi dan pengetahuan yang berguna dan bermanfaat bagi pengguna. Di sini, akan diambil daerah Lamongan dan sekitarnya dengan mengambil sample data dari RS Umum Soegiri Kabupaten Lamongan dan data cuaca dari BMKG Surabaya untuk daerah Lamongan dan sekitarnya.

## II. PENELITIAN SEBELUMNYA

Penelitian tentang teknik *Association Rule Mining* (ARM), khususnya yang multidimensi menjadi objek khusus dengan banyaknya data yang dijumpai yang mempunyai banyak faktor penentu. Multidimensional ARM didasari oleh adanya *multidimensional database* pada *data warehouse*. Jika *relational database* hanya menyimpan data pada tabel dua dimensi, maka *multidimensional database* dapat menyimpan lebih dari dua dimensi data, yang sering disebut juga *hypercube*.

Penelitian Seda Unal Calargun<sup>1</sup>, yang mengambil dataset berupa *spatio-temporal database* tentang meteorologi di daerah Turki, dengan menggunakan dua pendekatan dalam teknik data miningnya. Yaitu, dengan menggunakan *fuzzy association rule mining* yang dijalankan berdasarkan pembentukan *data cube* dan dengan menggunakan algoritma apriori. Dari hasil penelitian ditunjukkan bagaimana perbedaan hasil dari kedua pendekatan diatas bekerja. Dengan *data cube*, ada sebuah ukuran (*measure*) yang harus diisi sebagai acuan dari proses *data cube* selain

---

<sup>1</sup> Calargun, Unal, Seda; "Fuzzy Association Rule Mining from Spatio-Temporal Data : An Analysis of Meteorological Data In Turkey"; Middle East Technical University; 2008.

menambahkan dimensi yang diinginkan. Sedangkan pada algoritma Apriori, user bisa memilih dimensi-dimensi yang diinginkan tanpa menggunakan *measure*.

Sedangkan Intan, Yenti, Handojo<sup>2</sup>, dalam penelitiannya dengan menggunakan dataset medical record. Pada penelitiannya ini, digunakan dua langkah proses, yaitu menggabungkan tabel-tabel yang akan digunakan atau peneliti menyebutnya *denormalization process*, dan langkah selanjutnya *men-generate fuzzy rules* dengan menggunakan algoritma *multidimensional association rule mining*. Dengan menggunakan tiga variabel yaitu, nilai *support*, *confidence*, dan *correlation*, diharapkan rules yang dihasilkan lebih akurat.

### III. ALGORITMA FUZZY MULTIDIMENSIONAL ASSOCIATION RULE MINING

Ada dua jenis *Multidimensional Association Rule*, yaitu *Inter-dimensional Association Rule* dan *Hybrid-dimensional Association Rule*. *Inter-dimensional Association Rule* merupakan aturan asosiasi multidimensi atau multiatribut dimana tanpa ada predikat yang diulang. Sedangkan *Hybrid-dimensional Association Rule* adalah multidimensional ARM yang di dalam rule-nya terdapat predikat yang berulang. Berikut adalah contoh rule untuk *Inter-dimensional Association Rule*:

$$\text{age}(X, "20..29") \wedge \text{occupation}(X, "student") \Rightarrow \text{buys}(X, "laptop")$$

Sedangkan rule yang menggunakan predikat berulang pada *Hybrid-dimensional Association Rule*, bisa dilihat seperti contoh dibawah:

$$\text{age}(X, "20..29") \wedge \text{buys}(X, "printer") \Rightarrow \text{buys}(X, "laptop")$$

Pada penelitian ini, digunakan *Inter-dimensional association rule* pada prosesnya dengan tidak adanya dimensi atau predikat yang berulang pada rule-nya. Secara garis besar jalannya algoritma multidimensional fuzzy association rule mining ini, dapat dibagi menjadi 9 langkah, yaitu:

- Langkah 1: penetapan  $\lambda$  atau batas (*threshold*) untuk menentukan jumlah maksimum kategori/nilai dalam sebuah dimensi.
- Langkah 2: Set  $k = 1$ , di mana  $k$  adalah indeks variabel untuk menentukan nomor item kombinasi di itemsets, disebut  $k$ -itemsets.
- Langkah 3: Menentukan minimum support untuk itemset ke- $k$ .
- Langkah 4: Membangun setiap kandidat itemset ke- $k$ , dinotasikan dengan  $I_k$ , sebagai himpunan fuzzy set pada transaksi yang memenuhi syarat.
- Langkah 5: Menghitung total penjumlahan nilai fuzzy set tiap-tiap itemsetnya. Jumlah tersebut adalah mewakili nilai support masing-masing itemset.
- Langkah 6: Menyimpan itemset-itemset yang memiliki nilai sesuai *minimum support*-nya masing-masing.
- Langkah 7: Set  $k=k+1$ , jika  $k > \lambda$ .

<sup>2</sup> Intan, Rolly; Yenti Yuliana, Oviliani; Handojo, Andreas; "Mining Multidimensional Fuzzy Association Rules From A Database of Medical Record Patients"; Jurnal Informatika Vol. 9; 2008.

Langkah 8: Mencari kemungkinan/kandidat itemsets ke-k dari  $L_{k-1}$ . Jika tidak ada kandidat lagi dari itemset ke-k kemudian dilanjutkan ke Langkah-9.

Langkah 9: Menghitung nilai *confidence* untuk masing-masing asosiasi aturan yang mungkin.

Hasil dari algoritma ini adalah berupa rule(s) yang nilainya sesuai dengan minimum support dan minimum confidence yang diberikan.

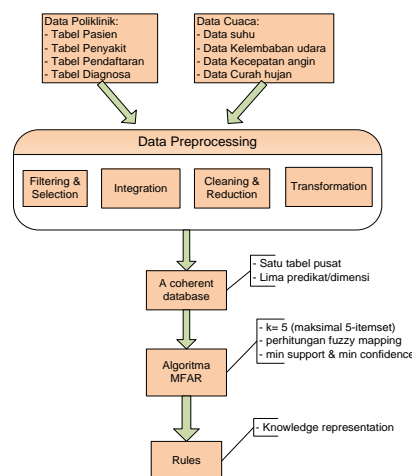
#### IV. PROSES DATA MINING

Data berasal dari data poliklinik RSUD Dr. Soegiri Kabupaten Lamongan dan data cuaca dari BMKG Surabaya untuk wilayah Lamongan dan sekitarnya. Dalam kenyataannya, pada data asli banyak yang perlu dikaji ulang. Mulai dari banyaknya file yang kosong, data yang *redundant*, nilai data yang kuantitatif sehingga sulit untuk terklasifikasi, semua ditemukan dalam kedua sumber data. Jumlah data asli yang berasal dari kedua data ditunjukkan pada Tabel 1.

**Tabel 1. Jumlah Record Data Asli**

Nama data	Jumlah record
Data Pasien	107.335
Data Jenis Penyakit	1.093
Data Ruang Poli	29
Data Pendaftaran	256.749
Data Diagnosa	256.749
Data Cuaca	730

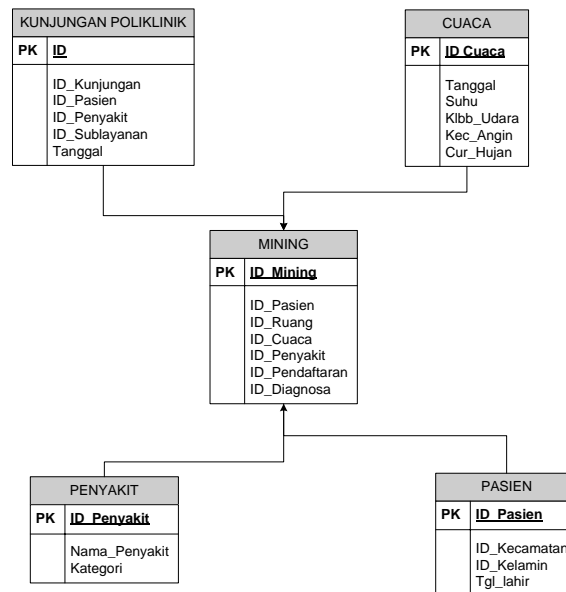
Kedua data tersebut akan diolah dalam data preprocessing terlebih dahulu kemudian disimpan pada satu tabel pusat pada proses data integration dan transformation dilanjutkan ke proses data mining. Alur kerja sistem dapat digambarkan pada Gambar 1.



**Gambar 1. Alur Kerja Sistem**

Pada proses data preprocessing, akan mengurangi data asli sebesar 60%, hal tersebut disebabkan banyaknya data asli yang records terdapat fields yang kosong. Sedangkan pada data integration akan menggabungkan semua tabel-tabel yang ada menjadi satu tabel pusat yang siap untuk di-mining.

Gambaran umum pada data integration digambarkan pada Gambar 2. Semua tabel yang diperlukan akan digabung menjadi satu di tabel pusat. Informasi-informasi yang diperlukan saja yang akan disimpan dalam tabel pusat.



**Gambar 2. Relasi dalam data pusat**

Implementasi yang dihasilkan dalam penelitian ini akan menghitung tiap iterasi itemset ke-k, dengan memberikan minimum support di tiap iterasinya. Pada implementasi akan menyajikan semua data yang sudah siap di-mining.

Total jumlah data yang telah melalui data preprocessing sebanyak 99.384 records. Hasil perhitungan nilai confidence dan pembatasan rule dengan nilai minimum confidence = 30% menghasilkan 36 rules yang valid. Rules tersebut adalah kombinasi dari 5 predikat atau dimensi yang telah lolos di iterasi 5-itemset.

Selama proses iterasi k-itemset pada proses data mining, rule(s) yang dihasilkan tiap iterasi memiliki jumlah dan itemset yang berbeda sesuai dengan minimum support yang yang diberikan di tiap iterasinya. Sehingga di tiap uji coba, menghasilkan jumlah rule(s) yang berbeda pula.

## V. KNOWLEDGE REPRESENTATION

Dari hasil output yang dihasilkan oleh aplikasi, didapatkan rule(s) dengan disertai nilai support dan confidence-nya. Hasil akhir dari algoritma ini memang menunjukkan rules tersebut. Dari rule(s) tersebut akan didapatkan informasi-informasi yang diharapkan dapat bermanfaat atau istilah dalam KDD adalah *knowledge representation*.

Knowledge representation di sini juga bisa diartikan pengetahuan apa saja yang bisa diartikan atau didapat dari rule(s) yang dihasilkan, dan berapa besar kevalidan atau keabsahan dari knowledge tersebut yang dapat dilihat dari nilai support dan confidence-nya. Pengetahuan yang bisa didapat dapat diartikan sebagai berikut, yang kali ini diwakili dengan beberapa contoh rules yang dihasilkan:

**KELOMPOK\_UMUR (x, ORANG TUA) => WILAYAH (x, Lamongan Tengah) {sup: 29%, conf: 74%}**

Rule di atas dapat diartikan bahwa pasien orang tua yang berkunjung ke rumah sakit adalah berasal dari Lamongan bagian tengah dengan nilai support = 29% dan nilai confidence = 74%.

**CURAH\_HUJAN (x, HUJAN RENDAH) ^ GENDER (x, Perempuan) => PENYAKIT (x, Osteo arthritis) {sup: 18%, conf: 32%}**

Rule di atas akan diartikan bahwa pasien perempuan yang berkunjung ke rumah sakit pada saat terjadi hujan dengan curah hujan rendah, didiagnosa mempunyai penyakit Osteo Arthritis dengan nilai support = 18% dan nilai confidence = 32%.

**CURAH\_HUJAN (x, HUJAN RENDAH) ^ KELOMPOK\_UMUR (x, ORANG TUA) ^ GENDER (x, Perempuan) => WILAYAH (x, Lamongan Tengah) {sup: 15%, conf: 75%}**

Rule di atas dapat diartikan bahwa pasien perempuan yang berusia tua dan berkunjung pada saat hujan dengan curah hujan rendah berasal dari Lamongan bagian tengah dengan nilai support = 15% dan nilai confidence = 75%.

**CURAH\_HUJAN (x, HUJAN RENDAH) ^ KELOMPOK\_UMUR (x, ORANG TUA) ^ GENDER (x, Perempuan) ^ WILAYAH (x, Lamongan Tengah) => PENYAKIT (x, Osteo arthritis) {sup: 6%, conf: 40%}**

Sedangkan rule di atas dapat diartikan bahwa pasien perempuan yang berusia tua dan berasal dari Lamongan bagian tengah yang berkunjung pada saat hujan dengan curah hujan rendah, didiagnosa mempunyai penyakit Osteo Arthritis dengan nilai support = 6% dan nilai confidence = 40%.

Bisa diketahui bahwa nilai support menunjukkan sering munculnya rule tersebut muncul dalam seluruh data. Sedangkan nilai confidence menunjukkan kuat tidaknya rule yang terbentuk itu di probabilitas kemunculannya. Sehingga bisa dikatakan apabila nilai confidence tinggi, rule tersebut dianggap merupakan strong association rule.

## **VI. PENUTUP**

Dari hasil penelitian yang telah dilakukan dalam mengolah data poliklinik dan data cuaca menggunakan metode Multidimensional Fuzzy Association Rule mining ini, dapat diambil kesimpulan sebagai berikut:

1. Dengan dilakukan data preprocessing, mengurangi jumlah data hingga 60% dari jumlah data awal. Jumlah data awal pada data poliklinik sebesar 256.749 records,

- dan setelah dilakukan data preprocessing berkurang menjadi 99.384 records. Hal ini terjadi karena proses penyimpanan atau model database pada data awal masih belum tertata dan banyaknya record dengan field kosong atau null.
2. Terdapat lima dimensi atau predikat yang diikutsertakan dalam proses mining. Lima dimensi ini kesemuanya adalah faktor-faktor yang dimungkinkan mempunyai asosiasi atau hubungan dengan keterjangkitan penyakit pasien. Dimensi-dimensi atau predikat tersebut adalah: curah hujan, usia, jenis kelamin, wilayah tempat tinggal, dan jenis penyakit.
  3. Maksimal iterasi adalah sampai pada iterasi 5-itemset. Dengan maksimal sampai di 5-itemset, maka akan mempengaruhi rules yang akan terbentuk yaitu maksimal juga akan mempunyai kombinasi predikat dalam rule sebanyak lima predikat.
  4. Rules yang dihasilkan mempunyai nilai support dan confidence sesuai dengan minsupp dan minconf. Dalam penyajiannya, rules yang dihasilkan akan langsung ditunjukkan nilai support dan confidence-nya, dan rule yang ditampilkan sudah merupakan rules yang sesuai dengan minimum confidence yang diberikan oleh user.
  5. Uji coba sistem dengan menggunakan beberapa variasi minimum support dan minimum confidence, aplikasi dapat berjalan dengan baik. Dilakukan tiga pengujian dengan berbagai variasi minimum support dan aplikasi menghasilkan rules sesuai dengan yang diharapkan dengan baik.
  6. Dari hasil uji coba dan evaluasi rule yang dihasilkan pada penelitian ini, diketahui bahwa terdapat tiga penyakit yang muncul. Pemetaan ketiga penyakit tersebut, yang pertama penyakit Osteo Arthristis adalah terjadi di Lamongan bagian tengah pada musim kemarau dimana terjangkitnya didominasi kelompok umur orang tua berjenis kelamin perempuan dengan nilai support berkisar antara 6% - 19%, dan nilai confidence berkisar antara 27% - 40%. Kedua, penyakit TB Paru adalah terjadi di Lamongan bagian tengah pada musim kemarau dimana terjangkitnya didominasi kelompok umur dewasa berjenis kelamin perempuan dengan nilai support berkisar antara 4% - 21%, dan nilai confidence berkisar antara 20% - 25%. Ketiga, penyakit DM tak tergantung Insulin adalah terjadi di Lamongan bagian tengah pada musim kemarau dimana terjangkitnya didominasi kelompok umur orang tua berjenis kelamin perempuan dengan nilai support berkisar antara 3% - 9%, dan nilai confidence berkisar antara 20% - 23%.

## VII. DAFTAR PUSTAKA

- [1] Anumalla, Kalyani; *Data Preprocessing Management System*; The Graduate Faculty of The University of Akron; 2007.
- [2] Bih, Joseph; *Paradigm Shift : An Introduction To Fuzzy Logic* ; IEEE Potentials; 2006.
- [3] Calargun, Unal, Seda; *Fuzzy Association Rule Mining from Spatio-Temporal Data : An Analysis of Meteorological Data In Turkey* ; Middle East Technical University; 2008.
- [4] Han, Jiawei; Kamber, Micheline; *Data Mining : Concept and Techniques*; Second Edition; 2006.
- [5] Intan, Rolly; *A Proposal of Fuzzy Multidimensional Association Rules*; 2006.
- [6] Intan, Rolly; *An Algorithm for Generating Single Dimensional Fuzzy Association Rule Mining*; 2006.
- [7] Intan, Rolly; Yenti Yuliana, Oviliani; Handojo, Andreas; *Mining Multidimensional Fuzzy Association Rules From A Database of Medical Record Patients*; Jurnal Informatika Vol. 9; 2008.

- [8] Kusumadewi, Sri; Purnomo, Hari; *Aplikasi Logika Fuzzy Untuk Pendukung Keputusan*; Graha Ilmu : Edisi 2; 2010.
- [9] Michael Case, Fitriani Ardiansyah, Emily Spector; *Climate Change in Indonesia Implications for Humans and Nature*; 2007.
- [10] *National Meteorological Library and Archive Fact sheet 6 : The Beaufort Scale*; Met Office London; 2010.
- [11] Neelu Khare, Neeru Adlakha, K.R. Pardasani; *An Algorithm for Mining Multidimensional Fuzzy Association Rule*; 2009.
- [12] Sankaradass, Veeramalai; Arputharaj, Kannan; *A Descriptive Framework for the Multidimensional Medical Data Mining and Representation*; Journal of Computer Science; 2011.
- [13] Swami, Shashank; Thakur, R.S.; Chandel, R.S.; *Multidimensional Association Rule Extraction in Smoking Habits Database*; Int. J. Advance Networking and Applications Vol. 3; 2011.