DATA MINING UNTUK MENDIAGNOSA PENYAKIT INFEKSI SALURAN PERNAFASAN (ISPA) MENGGUNAKAN METODE NAIVE BAYES

Pindan Jati Kusuma A12.2009.03424

Program Studi Sistem Informasi – S1 , Fakultas Ilmu Komputer

Universitas Dian Nuswantoro, Jl.Nakula I No.5-11 Semarang 50131

Pindanjati01@gmail.com

ABSTRAK

Kemajuan peradaban manusia berkembang pesat di segala bidang kehidupan sehingga ilmu pengetahuan dan teknologi menjadi bagian yang tidak terpisahkan. Dampak penggunaan teknologi dan modernisasi melahirkan industri yang berpengaruh besar terhadap penyebab penyakit Infeksi Saluran Pernafasan (ISPA). Berdasarkan data UNICEF/WHO pada tahun 2009 ISPA merupakan pembunuh balita pertama di dunia. Penelitian dibidang kesehatan untuk memprediksi pasien penderita ISPA, berdasarkan gejala-gejala penyakit, perlu dilakukan untuk pengobatan lebih dini, guna mencegah kematian akibat terlambatnya penanganan. Beberapa penelitian terkait prediksi penyakit menggunakan teknik *data mining* klasifikasi sudah secara luas digunakan. Penelitian ini menggunakan algoritma Naïve Bayes Classifier sebagai salah satu algoritma klasifikasi *data mining*. Algoritma naïve bayes diterapkan untuk menghitung probabilitas kemungkinan seseorang pasien dengan gejala-gejala tertentu apakah mengidap penyakit ISPA atau tidak. Obyek penelitan dilakukan pada Puskesmas Toroh 1 Kabupaten Grobogan untuk mengambil dataset pasien. Dataset memuat 39 atribut, 32 diantaranya merupakan atribut gejala-gejala penyakit, dengan total data berjumlah 1010 baris data. Hasil pemodelan diukur menggunakan table confusion matrix untuk menghitung akurasi. Pada penelitian ini terbukti naïve bayes classifier mampu menghasilkan akurasi yang tepat. Hasil dari penelitan ini dapat digunakan untuk memberikan referensi kepada pihak petugas kesehatan dan bagi pasien dalam penyimpulan hasil analisa penyakit ISPA.

Kata Kunci : *Data mining,* Klasifikasi, *Naive Bayes,* ISPA, Analisa

ABSTRACT

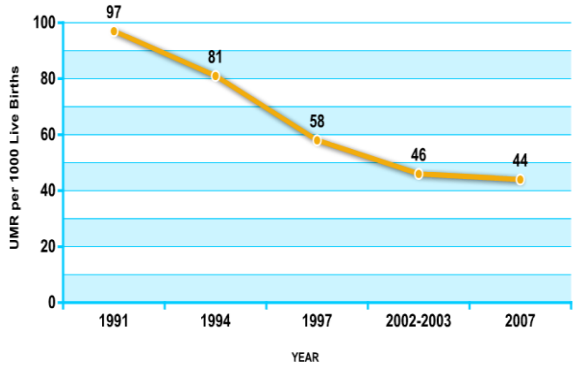
The progress of human civilization grew in all areas of life, that science and technology become the inseparable part. The impact from using of technology and modernizing creates industry that could greatly result in the cause of the disease Infection of lower Respiratory Tract (ISPA). Based on the data UNICEF/WHO in the year 2009 ISPA is the first murderer's children under five years in the world. Research in health to predict patients with ISPA, based on any symptoms of disease, needs to be done for the earlier treatment to prevent deaths from late handling. Some researches related to the prediction disease using technical *data mining* classification has been used widely. This research uses Naïve Bayes Classifier algorithm as one of the algorithm types of *data mining*. Naïve Bayes Algorithm applied to calculate probability likely whether a patient has certain symptoms of ISPA disease or not. Objects of laboratory will be done in a health clinic at Toroh – Grobogan Regency to take data-set patients. It has 39 attributes, including 32 symptoms of the disease attributes, with a total 1010s data lines. Modeling result was measured by using table confusion matrix to calculate accuracy. In this research has proven that Naïve Bayes Classifier is capable of producing high accuracy in the right direction. Results of laboratory can be used to give references to the health officials and for patients in result a logical deduction analysis of ISPA disease.

Key words : Analysis,*Data mining*, Classification, Naïve Bayes, ISPA

1. Pendahuluan

1.1 Latar Belakang

Kemajuan peradaban manusia sudah semakin berkembang pesat di segala bidang kehidupan. Ilmu pengetahuan dan teknologi dewasa ini menjadi bagian yang tidak terpisahkan dari kehidupan masyarakat modern. Tidak bisa dipungkiri bahwa hasil modernisasi melahirkan industri yang berpengaruhbesar terhadap penyebab penyakit Infeksi Saluran Pernafasan (ISPA). . ISPA adalah penyakit yang menyerang salah satu bagian atau lebih dari saluran pernafasan, mulai dari hidung (saluran atas) hingga alveoli (saluran bawah) termasuk jaringan adneksanya seperti sinus, rongga telinga tengah, dan pleura, Berikut adalah gambaran perkembangan Angka Kematian Balita atau AKABA pada tahun 1991 – 2007 :



Secara anatomik, ISPA dikelompokkan menjadi ISPA atas misalnya batuk, pilek, faringitis, dan ISPA bawah seperti bronkitis, bronkiolitis, pnemonia. ISPA atas jarang menyebabkan kematian walaupun insidennya jauh lebih tinggi daripada ISPA bawah. Menurut data di *United Nations International Children's Emergency Fund* (UNICEF) dan *World Health Organization* (WHO) pada tahun 2009 ISPA merupakan pembunuh balita pertama di dunia, lebih banyak dibandingkan dengan penyakit lain seperti AIDS, malaria dan campak. Di dunia setiap tahun diperkirakan lebih dari 2 juta meninggal karena ISPA (1 balita/15 detik) dari 9 juta total kematian balita. Di antara 5 kematian balita, 1 diantaranya di sebabkan oleh pneumonia. Bahkan karena besarnya kematian ISPA ini, ISPA/pneumonia disebut sebagai *pandemic* yang terlupakan atau *forgetten pandemic.* Berdasarkan latar belakang di atas maka peneliti tertarik untuk melakukan penelitian di Puskesmas Toroh 1 Kabupaten Grobogan dimana pada Puskesmas tersebut jumlah data untuk pasien yang terjangkit penyakit Infeksi Saluran Pernafasan (ISPA) belum spesifik.

Untuk melakukan analisa data dalam jumlah besar yang tersimpan pada *database*, menggunakan teknik *data mining.* Potensi *data mining* dalam bidang kesehatan sudah diakui secara luas. Banyak studi yang dilakukan menggunakan teknik *data mining* modern, antara lain *classification* dan *predictive* yang diterapkan pada rekam medis elektronis. Dalam hal ini penulis menggunakan metode algoritma *naïve bayes* untuk mendiagnosa penyakit ISPA. Kelebihan metode *naïve bayes* sendiri adalah mudah di implementasi serta memberikan hasil yang baik untuk banyak kasus. Teorema Bayes adalah teorema yang digunakan dalam statistika untuk menghitung peluang untuk suatu hipotesis, Bayes Optimal Classifier menghitung peluang dari suatu kelas dari masing-masing kelompok atribut yang ada, dan menentukan kelas mana yang paling optimal.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah diatas dapat dirumuskan suatu masalahnya adalah bagaimana prediksi klasifikasi gejala penyakit ISPA dan bagaimana akurasi teknik klasifikasi *data mining* menggunakan algoritma *naive bayes.*

1.3 Tujuan Penelitian

Tujuan dalam penelitian ini adalah untuk memprediksi gejala penyakit ISPA dan mendapatkan akurasi yang tepat untuk prediksi gejala penyakit ISPA menggunakan metode *naive bayes*.

2. Tinjauan Pustaka

2.1 Data Mining

Menurut Turba *data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstrasi, mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

2.2 Langkah-langkah Data Mining

Untuk melakukan penggalian data, ada beberapa tahapan. Tahap-tahap tersebut bersifat interaktif di mana pemakai terlibat langsung atau dengan perantaraan *knowledge base*. Langkah-langkah *data mining* adalah :

1. *Data cleaning* (untuk menghilangkan noise data yang tidak konsisten) Data integration (di mana sumber data yang terpecah dapat disatukan).
2. *Data selection* (di mana data yang relevan dengan tugas analisis dikembalikan ke dalam *database*).
3. *Data transformation* (di mana data berubah atau bersatu menjadi bentuk yang tepat untuk menambang dengan ringkasan performa atau operasi agresi).
4. *Data mining* (proses esensial di mana metode yang intelejen digunakan untuk mengekstrak pola data).
5. *Pattern evolution* (untuk mengidentifikasi pola yang benar-benar menarik yang

mewakili pengetahuan berdasarkan atas beberapa tindakan yang menarik).

6. *Knowledge presentation* (di mana gambaran teknik visualisasi dan pengetahuan digunakan untuk memberikan pengetahuan yang telah diberikan kepada *user*).

2.3 Algoritma Teorema Bayes

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Bayesian classification* didasarkan pada teorema bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neutral networ* . *Bayesian classification* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar.

Teorema Bayes memiliki bentuk umum sebagai berikut :

$$P(H|X) = = \frac{P(X\backslash H) * P(H)}{P(X)}$$

Dalam hal ini :

X: data dengan *class* yang belum diketahui

H :hipotesis data X merupakan suatu *class* spesifik

P(H|X) :probabilitas hipotesis H berdasar kondisi X (*posteriori probability*)

P(H) :probabilitas hipotesis H (*prior probabilty*)

P(X|H) :probabilitas X berdasar kondisi pada hipotesis H

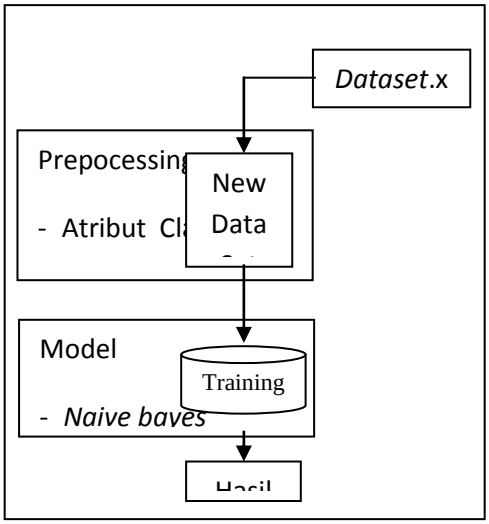
P(X) : probabilitas dari X

2.4 Cross Validation

Cross validation adalah teknik pengambilan sampel secara random yang menjamin setiap jumlah kemunculan data yang diamati sama dengan jumlah data *training* dan hanya sekali pada data *testing.*

2.5 Pemodellan

Metode yang digunakan yaitu algoritma *Naïve Bayes*. Untuk menghitung data dalam penelitian ini akan menggunakan *framework* RapidMiner versi 5.3 sehingga akan ditemukan nilai akurat. Di bawah ini model yang diusulkan :



3. Analisa Data Mining

Adapun sumber data utama yang digunakan dalam penelitian ini adalah *dataset* pasien dari bulan Januari-Maret 2013. Data tersebut terdiri dari beberapa tabel (*class*) antara lain tabel register pendaftaran dan tabel jenis penyakit. Tabel register pendaftaran berisi tentang informasi pendaftaran pasien yang terdiri dari 5 atribut antara lain nama pasien, umur, jenis kelamin, alamat dan kode penyakit sedangkan tabel jenis penyakit terdiri dari 2 atribut antara lain kode penyakit dan gejala-gejala penyakit. Dari kedua tabel tersebut di relasikan dan di sederhanakan menjadi 5 atribut antara lain :

1. Nama Pasien ⟶ type data "Polynominal"
2. Umur ⟶ type data "Integer"
3. Jenis Kelamin ⟶ type data "Binominal"
4. Gejala Penyakit ⟶ type data "Binominal"
5. Kode Penyakit ⟶ type data "Integer"

3.1 Penyeleksian Data

Pada data pasien selama 3 bulan terdapat beraneka ragam kategori penyakit yang tercatat. Dalam penelitian ini, penulis mengambil 15 kategori penyakit dalam bahan penelitian, antara lain : Myalgia, Herpes, Dermatitis, Katarak, Dispepsia, Hipertensi, ISPA, Diabetes Mellitus, Typus, TBC, Diare , Bronkhitis, Anemia, Influenza, Alergi.

3.2 Processing / Cleaning

Pada tabel registrasi pasien dan tabel jenis penyakit terdapat banyak atribut, atribut-atribut tersebut tidak semua diperlukan dalam proses mining, maka dari itu perlu dilakukan pembersihan atau cleaning yang bertujuan memilih atribut data yang menjadi focus penelitian dan

menghapus atribut yang tidak dipakai. Dari kedua tabel tersebut kemudian direlasikan dan atribut yang nanti akan dipakai dalam penelitian yaitu Umur,Jenis Kelamin,Gejala Penyakit. Dataset hasil relasi dari kedua tabel kemudian di ubah menjadi bilangan binominal agar mudah perhitungannya menggunakan rapidminer.

3.3 Perhitungan Data Mining

Berikut perhitungan manual *naïve bayes* dengan menggunakan data set pada tabel 4.5 jika data terakhir di jadikan data training :

1. Menghitung jumlah kelas dari klasifikasi yang terbentuk (prior probability) :

   - C1 (Class ISPA = “yes”) = jumlah “yes” pada kolom AM class ISPA = 4/10 = 0.4
   - C2 (Class ISPA = “no”) = jumlah ‘no” pada kolom AM class ISPA = 6/10 = 0.6

2. Menghitung jumlah kasus yang sama dari kelas yang sama

C	AM
UMUR	CLASS ISPA
32	NO
25	NO
6	YES
48	NO
58	NO
40	YES
12	YES
34	NO
47	NO
35	YES
40	NO

- P(umur=”40” | Class ISPA = “yes”) = ¼ = 0,25
- P(umur=”40” | Class ISPA = “no”) = 1/6 = 0.167

3. Kalikan semua hasil variable

   - Untuk semua atribut Class ISPA =”yes”
   - P (X | Class ISPA = “yes”)

     = 0,25 x 0,5 x 0 = 0
   - Untuk semua atribut Class ISPA = “no”
   - P (X | Class ISPA = “no”)

     = 0,167 x 1 x 0,333 = 0,0556
   - Perkalian prior probability dengan semua atribut yang Class ISPA =”yes”

- P (X | Class ISPA = “yes”) P(X| Class ISPA = “yes”)

  = 0,4 x 0

  = 0

➢ Perkalian prior probability dengan semua atribut yang Class ISPA =”no”

- P (X | Class ISPA = “no”) P(X| Class ISPA = “no”)

  = 0,6 x 0,0556

  = 0,03336

4. Bandingkan hasil kelas

- P (X | Class ISPA = “yes”) P(X| Class ISPA = “yes”) < P (X | Class ISPA = “no”) P(X| Class ISPA = “no”)

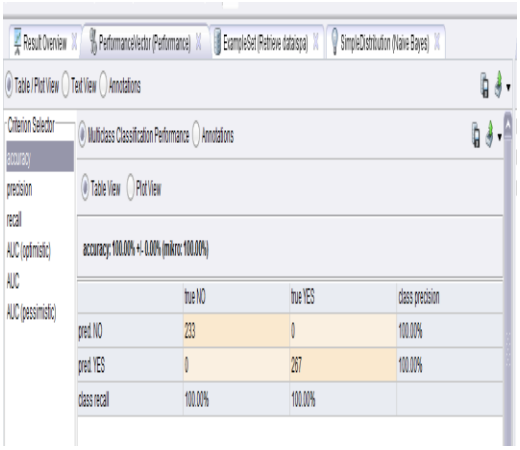
Kesimpulan =

Class ISPA = “NO”

(Perhitungan antara perkalian class ISPA = “yes” dengan class ISPA = ”no” menunjukkan bahwa nilai lebih besar class ISPA = “no”)

3.4 Hasil Percobaan dan Pengujian

Pada percobaan dengan algoritma *naïve bayes* di tools Rapidminer diperoleh waktu komputasi adalah 0 second. 0 second disini artinya komputasi menggunakan *naïve bayes* berjalan cukup cepat. Hal ini sesuai dengan kelebihan *naïve bayes* dibandingkan beberapa algoritma lain seperti *neural network* yang membutuhkan waktu berjam-jam untuk melakukan komputasi data.



Hasil akurasi model *naïve bayes* menunjukkan tingkat akurasinya 100% artinya model prediksi penyakit ISPA dengan *naïve bayes* terbukti baik hal ini dilihat dari tingkat akurasinya yang mencapai 100% akan tetapi hal ini perlu di tinjau ulang dari sudut pandang kompleksitas datasetnya.

4. Kesimpulan dan Saran

4.1 Kesimpulan

Dari analisa data dapat di tarik kesimpulan bahwa pasien yang mempunyai gejala penyakit sesak Nafas, nafas lemah, sakit kepala, hidung tersumbat, batuk, panas memiliki potensi tinggi mengidap

penyakit ISPA. Dari hasil observasi terhadap sejumlah *dataseet* penyakit ISPA pada Puskesmas Toroh 1 yang diambil dari bulan Januari-Maret 2013 dan mengalami proses perhitungan menggunakan metode *Naïve Bayes* dengan atribut yang telah dijelaskan di pembahasan sebelumnya, didapatkan sebuah hasil bahwa nilai akurasi terhadap penyakit ISPA mencapai 100%. Dimana 100% bisa juga disebabkan oleh kurang kompleksitas data yang mengakibatkan model dapat memprediksi dengan sangat akurat.

4.2 Saran

Diharapkan dalam penelitian selanjutnya dapat dibandingkan penelitian tersebut dengan memanfaatkan metode klasifikasi lainnya seperti metode C.4.5 ,metode *nearest neighbor* guna menentukan kelas berdasarkan atribut-atribut yang telah ditentukan sehingga dengan menggunakan banyak metode dapat lebih mengetahui kelebihan masing-masing metode dan metode mana yang menghasilkan nilai akurasi yang lebih baik.Metode Naïve Bayes dalam penelitian Diagnosa Penyakit ISPA mendapati kekurangan dalam hasil perhitungan akurasi terhadap penyakit ISPA, dikarenakan kompleksitas data yang digunakan kurang kompleks , maka saran untuk penelitian selanjutnya data yang digunakan lebih kompleks dan lebih detail.