Vol.1 (1) (2020) 14-23

**Information Science and Library**
http://journals.usm.ac.id/index.php/jils

# Perbandingan Kerangka Model Klasifikasi untuk Pemilihan Metode Kontrasepsi dengan Pendekatan CRIPS-DM

**Saeful Amri** [1]
saefulamri@usm.ac.id
Perpustakaan, Universitas Semarang, Indonesia

DOI: http://dx.doi.org/10.26623/jisl.v1i1

## Info Artikel

*Sejarah Artikel:*
Disubmit 25 Juni 2020
Direvisi 30 Juni 2020
Disetujui 06 Juli 2020

*Keywords:*
*Data Mining, Classification, Decision Tree, Naive Bayes, K-NN, Random Forest, and Deep Learning.;*

## Abstrak

Rumah Sakit Ibu dan Anak (RSIA) Kusuma Pradja Semarang dalam kesehariannya memberikan pelayanan reproduksi terpadu di Semarang. Dari banyaknya kepesertaan keluarga berencana (KB) perlu diterapkan pengambilan keputusan menggunakan alat kontrasepsi, dalam hal ini perlu dilakukan pendekatan data mining dengan melakukan komparasi 05 kerangka model algoritma klasifikasi yaitu: *Decision Tree*, *Naive Bayes*, *K-NN*, *Random Forest*, dan *Deep Learning* demi mendapatkan algoritma terbaik dalam menentukan metode kontrasepsi yang tepat untuk pasien RSIA Kusuma Pradja Semarang.
Hasil penelitian menunjukkan bahwa *Naive Bayes* (NB) merupakan model terbaik dalam menentukan metode kontrasepsi.

## *Abstract*

*Kusuma Pradja Semarang Mother and Child Hospital (RSIA) in its daily life provides integrated reproductive services in Semarang. Of the many members of family planning (KB) it is necessary to apply decision making using contraception, in this case a data mining approach needs to be done by comparing 05 framework classification algorithm models namely: Decision Tree, Naive Bayes, K-NN, Random Forest, and Deep Learning in order to get the best algorithm in determining the right method of contraception for RSIA Kusuma Pradja Semarang patients.*
*The results showed that Naive Bayes (NB) is the best model in determining contraceptive methods.*

✉ Alamat Korespondensi:
E-mail: saefulamri@usm.ac.id

e-ISSN:

## PENDAHULUAN

Latar Belakang

Laju pertumbuhan yang semakin meningkat akan mempengaruhi beberapa faktor diantaranya faktor sosial politik yang dapat membawa perubahan sistem pemerintahan yang cukup besar (Bandyopadhyay dan Chattopadhyay, 2008)

penentu kesuburan dan prediktor yang paling penting dari transisi fertilitas. Pemilihan metode kontrasepsi juga dipengaruhi oleh sejumlah faktor demografi yang saling bergantung diantaranya faktor budaya, ekonomi, dan sosial yang berarti bahwa pendekatan multidimensional perlu diadopsi untuk menganalisis pola penggunaan kontrasepsi. Setiap analisis berdasarkan indikator tunggal tidak mungkin untuk menangkap semua dimensi dari pemilihan metode kontrasepsi (Chaurasia, 2014).

RSIA Kusuma Pradja merupakan salah satu rumah sakit yang dalam kesehariannya memberikan pelayanan reproduksi terpadu di Semarang. Dari banyaknya data pasien yang memilih dalam penggunaan alat kontrasepsi, dalam hal ini peneliti melakukan mendekatan data mining dalam penggalian informasi untuk mendapatkan pengetahuan yang terkandung dalam data tersebut.

Peningkatan kebutuhan terhadap analisa dan pengolahan data yang sangat pesat, perlu adanya dukungan sebuah metode yang dapat mengambil pengetahuan dari data tersebut. Komparasi algoritma termasuk dalam kategori metode data mining yang bertujuan untuk mengetahui pola informasi yang terkandung dalam sebuah dataset.

Dalam hal ini peran data mining digunakan untuk mengatasi masalah penentuan pasien untuk pemilihan alat kontrasepsi di RSIA Kusuma Pradja. Dengan metode komparasi 05 model algoritma klasifikasi yang bertujuan mendapatkan pengetahuan dari data yang diperoleh guna menja dibahan rekomendasi pasien dalam pengambilan keputusan menggunakan alat kontrasepsi yang tepat.

Tujuan

Tujuan penelitian ini adalah menerapkan kerangka model klasifikasi terbaik dalam menentukan alat kontrasepsi yang tepat pada pasien di RSIA Kusuma Pradja Semarang.

Batasan Masalah

Dalam penelitian ini, pembahasan terbatas pada:

a. Mengetahui hubungan antar faktor.
b. Metode perbandingan 05 model algoritma klasifikasi.
c. Dataset yang diolah adalah dataset pasien penggunaan alat kontrasepsi di RSIA Kusuma Pradja Semarang.
d. Model dibuat dengan mengunakan RapidMiner 7.2.

## METODE

Data mining atau disebut juga Knowledge Discovery in Database (KDD) adalah ekstraksi pola secara otomatis mewakili pengetahuan yang disimpan atau ditangkap secara tersembunyi di dalam sebuah database besar (Han dan Kamber, 2012)

Berdasarkan tugasnya data mining dikelompokkan menjadi (Larose, 2005): Estimasi, Prediksi, Klasifikasi, Clustering, dan Asosiasi. Berikut adalah penjabaran dari tugas data mining:

a. Estimasi

   Estimasi adalah permodelan yang hampir sama dengan klasifikasi, target variabel lebih mengarah pada numerik daraipada kearah kategori.

b. Prediksi

   Permodelan prediksi akan menampilkan sebuah hasil yang dapat dilihat dimasa mendatang.

c. Klasifikasi

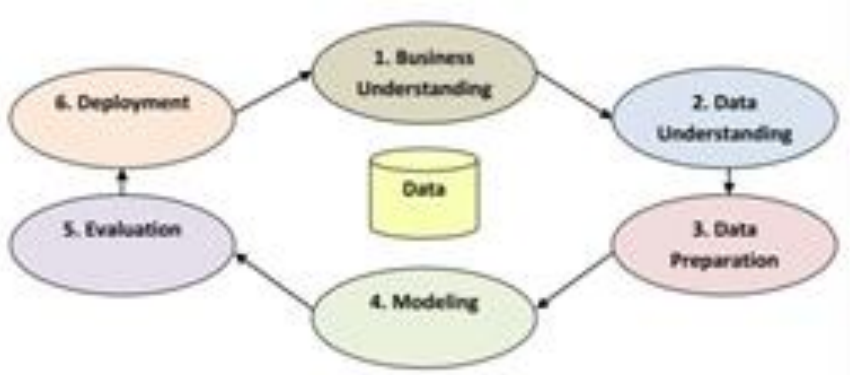
   Permodelan klasifikasi lebih ditargetkan pada variabel kategori.

d. Clustering
   Merupakan metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (similarity) antara satu data dengan data yang lain. Clustering dalam data mining juga bersifat tanpa arahan (unsupervised).
e. Asosiasi
   Metode asosiasi digunakan untuk menemukan atribut yang muncul dalam suatu waktu.

Dalam data mining terdapat tiga metodologi yang bisa diterapkan untuk industri maupun penulisan ilmiah. Ketiga metodologi tersebut antara lain KDD (Knowlege Discovery from Data) (Han and Kamber, 2015) SEMMA (Sample, Explore, Modify, Model and Assess) (Makhabel, 2015) dan CRIPS-DM (Cross Industry Standard Process for Data Mining) (North, 2012). Dalam penelitian ini, metodologi yang digunakan adalah CRIPS-DM. Proses standar data mining dari portal industri (CRISP-DM) sudah dikembangkan pada tahun 1996 (Chapman, 2000). Proses tersebut dilakukan untuk mendapatkan strategi memecahan masalah dari sebuah industri dan juga sebagai wadah pengetahuan bagi para peneliti .

Proses standar data mining terdiri dari enam fase yaitu:

a. Fase Pemahaman Bisnis
b. Fase Pemahaman Data
c. Fase Persiapan Data
d. Fase Permodelan
e. Fase Evaluasi
f. Fase *Deployment*




Gambar 1. Alur Proses Standar Data Mining

Pemahaman bisnis merupakan fase awal dari CRISP-DM, yaitu untuk mengetahui masalah yang akan diselesaikan untuk mencapai tujuan yang di inginkan.

Kemudian pemahaman data untuk mengenali data yang akan diolah dengan mengumpulkan mendeskripsikan dan mengevalusi kualitas data.

Persiapan data diperlukan untuk mempersiapkan data yang akan diolah agar dapat dimodelkan dengan memeriksa kenormalan, kelengkapan, dan konsistensi data, sehingga data dapat dimodelkan sesuai dengan metode data mining yang akan digunakan.

Permodelan merupakan fase untuk memodelkan sesuai dengan metode yang digunakan untuk mengolah dataset, dalam paper ini digunakan peran data mining, yaitu klasifikasi. Pada klasifikasi dibandingkan 05 algoritma yaitu : *Decision tree*, *Naive Bayes*, *K-NN*, *Random Forest*, dan *Deep Learning*.

Evaluasi digunakan untuk mengevaluasi hasil pengolahan data sehingga dapat diketahui pola/ pengetahuan dari hasil permodelan tersebut.

Deployment merupakan tindak lanjut setelah diketahui pola/ pengetahuan, dari dataset yang sudah diolah..

## HASIL DAN PEMBAHASAN

### Pemahaman Bisnis

Pemahaman bisnis merupakan fase untuk mengetahui sebuah masalah, dalam hal ini pasien di RSIA Kusuma Pradja Semarang banyak yang menggunakan alat kontrasepsi. Dari data tersebut dapat dilakukan pengolahan data untuk mengetahui faktor-faktor yang berpengaruh pada penggunaan alat kontrasepsi dan diperoleh model algoritma klasifikasi terbaik untuk memberikan rekomendasi pasien dalam pengambilan keputusan menggunakan alat kontrasepsi yang tepat.

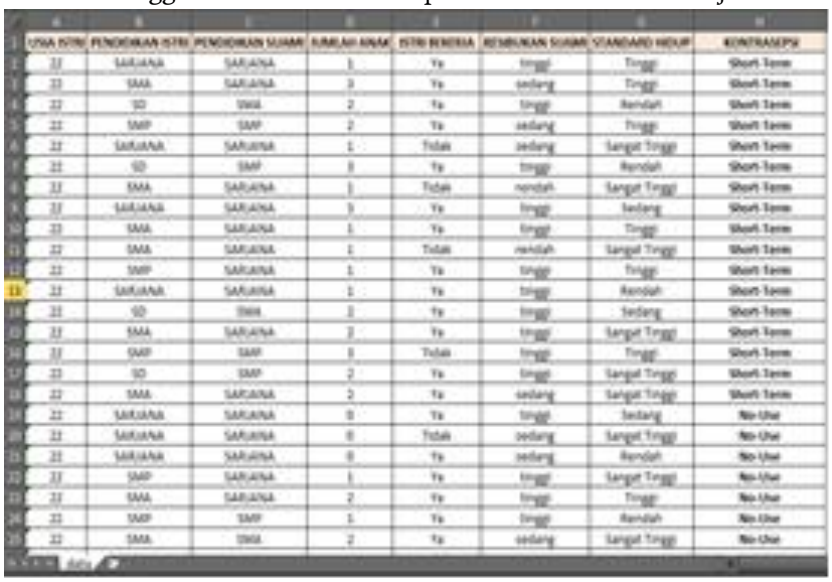
### Pemahaman Data

Tahapan pemahaman data dimulai dengan pengumpulan dari database pasien yang menggunakan alat kontrasepsi di RSIA Kusuma Pradja Semarang, dari data tersebut dapat di identifikasi beberapa hal baik dalam masalah data, mendeteksi subyek menarik dari data, dan memperoleh informasi yang tersembunyi.

Dari dataset pasien yang menggunakan alat kontrasepsi diperoleh beberapa atribut antara lain:

a. Usia Istri
   Berisi data usia istri, berkisar antara umur 22 tahun sampai dengan 50 tahun.
b. Pendidikan Istri
   Berisi pendidikan formal istri, mulai dari pendidikan dasar, SMP, SMA, dan Sarjana.
c. Pendidikan Suami
   Berisi pendidikan formal istri, mulai dari pendidikan dasar, SMP, SMA, dan Sarjana.
d. Jumlah Anak
   Berisi jumlah anak yang dilahirkan oleh istri.
e. Istri Bekerja
   Berisi jawaban istri, jika bekerja dijawab ya, jika tidak bekerja dijawab tidak.
f. Kesibukan Suami
   Berisi aktifitas pekerjaan suami yaitu kesibukan sangat tinggi, tinggi, sedang, dan rendah.
g. Standar Hidup
   Berisi pola hidup dari keluarga pasien, yaitu pola hidup sangat tinggi, tinggi, sedang, dan rendah.
h. Metode Kontrasepsi
   Berisi label yaitu *no-use*, *short-term* dan *long-term*.

Tabel 1. Dataset Menggunakan Alat Kontrasepsi di RSIA Kusuma Pradja Semarang



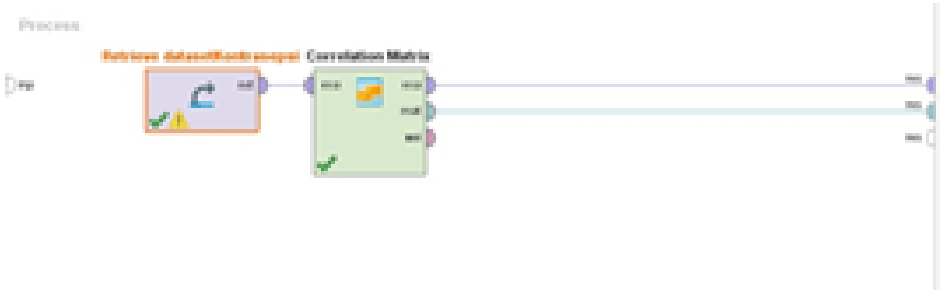
**Persiapan Data**

Tahapan persiapan data, dalam tahap ini digunakan semua atribut-atribut yang ada pada dataset penggunaan alat kontrasepsi, kemudian atribut tersebut disesuaikan dengan metode data mining yaitu untuk klasifikasi dalam bentuk label *polynominal* yaitu *no-use*, *short-term* dan *long-term*. Apabila atribut belum sesuai maka perlu di transformasi data agar dapat dimodelkan sesuai dengan metode algoritma data mining.

**Permodelan**

Merupakan fase permodelan data mining dengan dari hubungan antar faktor dan menentukan algoritma yang akan digunakan. tool berupa RapidMiner 7.2

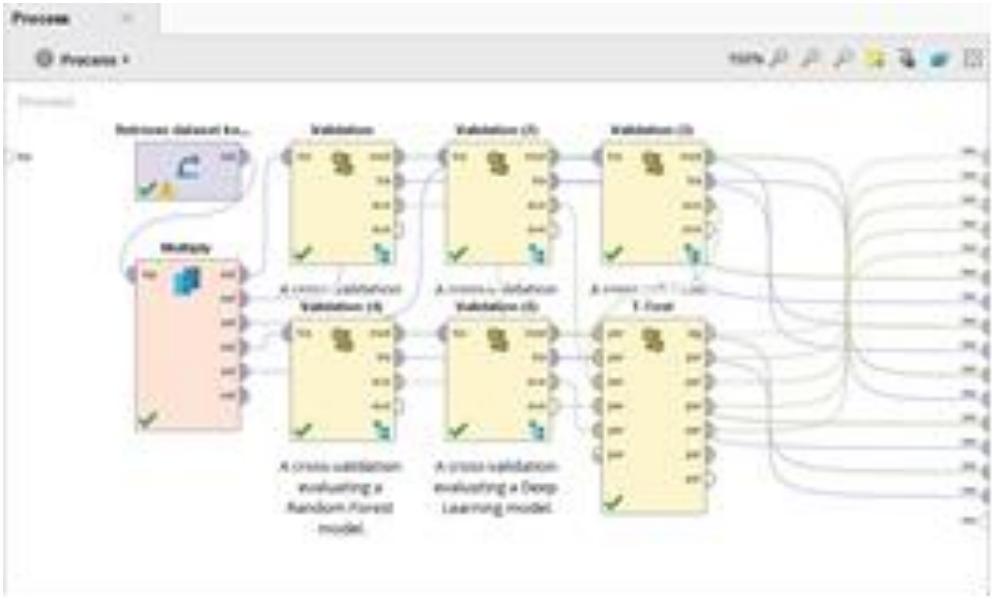
Hubungan antar faktor dilakukan untuk mengetahui faktor-faktor mana yang berpengaruh pada metode kontrasepsi.



Gambar 2. Correlation Matrix

Klasifikasi dibandingkan 05 algoritma yaitu : *Decision tree, Naive Bayes, K-NN, Random Forest, dan Deep Learning*. Untuk diketahui algoritma yang terbaik.

Uji beda (T-Test) juga digunakan untuk membandingkan kinerja (performa) dari 05 algoritma, dengan uji beda (T-Test) dapat diketahui akurasi dari kelima model algoritma.



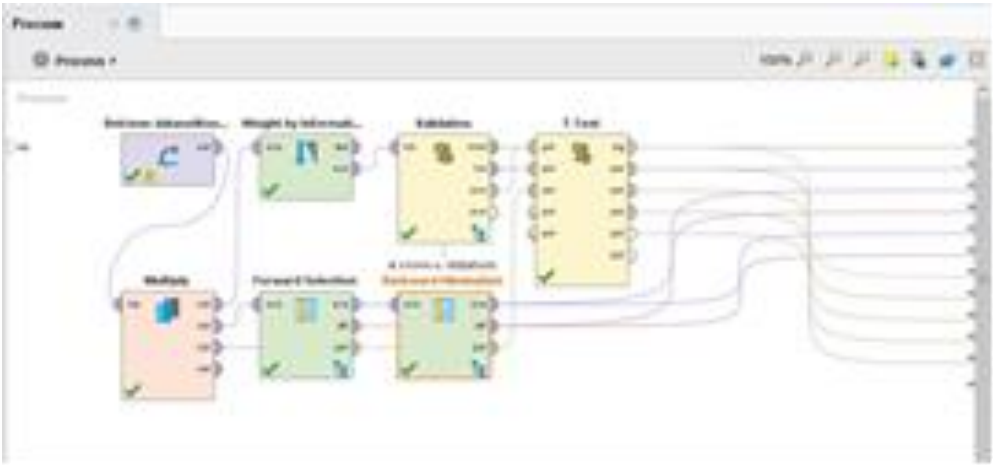
Gambar 3. 05 Model Algoritma Klasifikasi

Untuk memperbaiki kinerja (performa) dari masing-masing metode dapat digunakan metode pengurangan dimensi yang terdiri dari beberapa jenis, sedangkan yang digunakan dalam penelitian ini adalah *Featur Selection* dan *wrapper* (*Forward Selection* dan *Backward Elimination*).

*Feature selection* adalah masalah yang berterkaitan erat dengan pengurangan dimensi. Tujuan *feature selection* adalah untuk mengidentifikasi fitur dalam kumpulan data yang sama pentingnya, dan

membuang semua fitur lain seperti informasi yang tidak relevan dan berlebihan. Karena *feature selection* mengurangi dimensi dari data, sehingga memungkinan operasi yang lebih efektif (Matatov, Rokach, & Maimon, 2010) *Feature selection* adalah salah satu faktor yang paling penting yang dapat mempengaruhi tingkat akurasi klasifikasi (Liu et al., 2011).

Uji beda (T-Test) juga digunakan untuk membandingkan kinerja (performa) pengurangan dimensi untuk mengetahui dimensi-dimensi yang terbaik dalam metode pemilihan kontrasepsi.

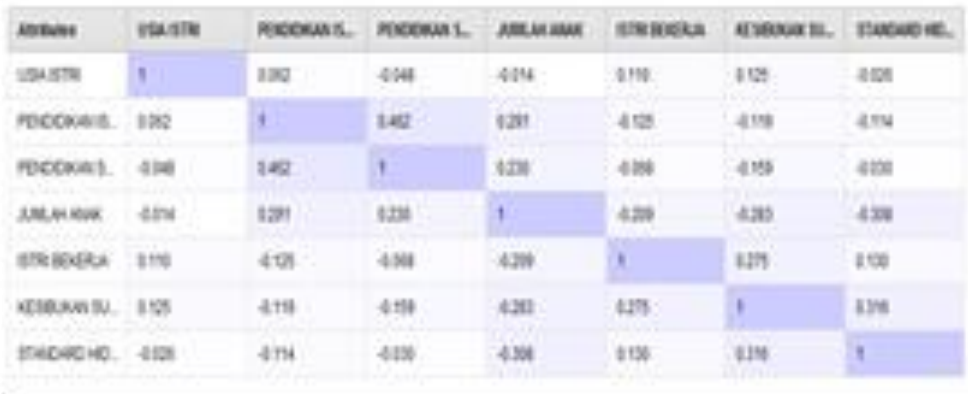


Gambar 4. Featur Selection dan wrapper (Naive Bayes)

**Evaluasi**

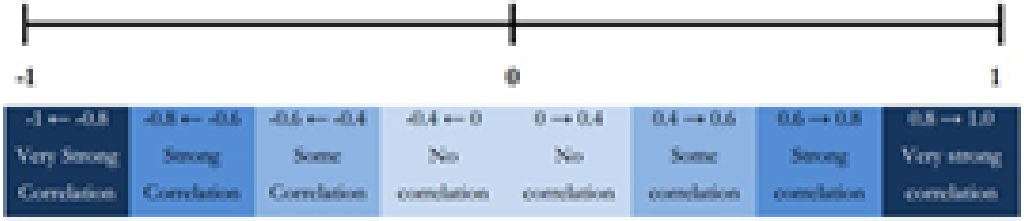
Dari hasil penelitian, dapat diketahui bahwa faktor jumlah anak paling berpengaruh dalam menentukan metode kontrasepsi.

Hasil *Correlation Matrix* memperlihatkan faktor jumlah anak memiliki hubungan negatif karena didapat nilai -0.014 dari faktor-faktor yang lain.



Gambar 5. Hasil Correlation Matrix

Kekuatan hubungan antar faktor didapatkan dari nilai korelasinya, semakin banyak nilai hubungan antar faktor maka dapat dikatakan kuat korelasinya, sebaliknya jika nilai hubungan antar faktor kecil maka lemah kekuatannya.




Gambar 6. Kekuatan antar Faktor

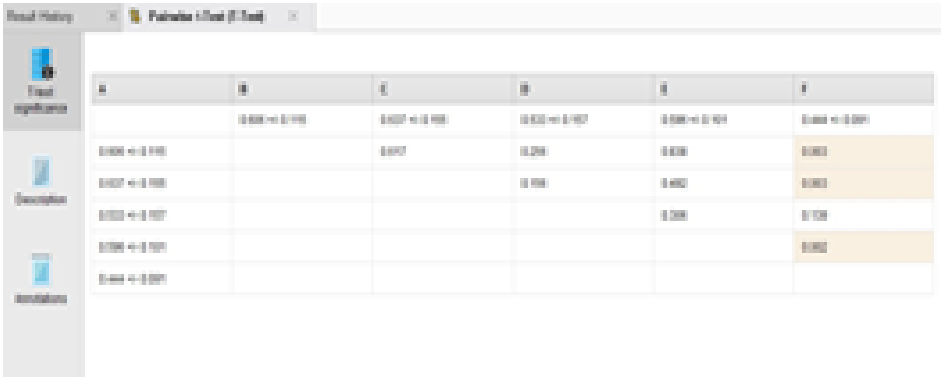
Dari hasil *Correlation Matrix* juga diketahui hubungan antar faktor diantaranya adalah hubungan positif (berbanding lurus) antara pendidikan istri dengan pendidikan suami, hal ini dapat diketahui dari nilai korelasinya jika pendidikan istri positif maka pendidikan suami positif, begitupula jika nilai pendidikan istri negatif maka pendidikan suami nilainya negatif.


Gambar 7. Hubungan Positif Antara Faktor Pendidikan Istri dengan Pendidikan Suami

Selain hasil hubungan positif juga didapat hubungan negatif (berbanding terbalik) antara faktor jumlah anak dengan usia istri, hal ini dapat diketahui dari nilai korelasinya jika jumlah anak positif maka usia istri negatif, sebaliknya jika nilai jumlah anak negatif maka usia istri nilainya positif.


Gambar 8. Hubungan Negatif Antara Faktor Usia Istri dengan Faktor Jumlah Anak

Dalam permodelan 05 algoritma klasifikasi menggunakan uji beda (T-Test) diketahui bahwa model klasifikasi terbaik dalam menentukan metode kontrasepsi adalah *Naive Bayes* (NB).



Gambar 9. Hasil Uji Beda (T-Test) 05 Algoritma Klasifikasi

Metode *Naive Bayes* merupakan model klasifikasi statistik yang digunakan untuk memprediksi probalitas keanggotaan suatu *class*.

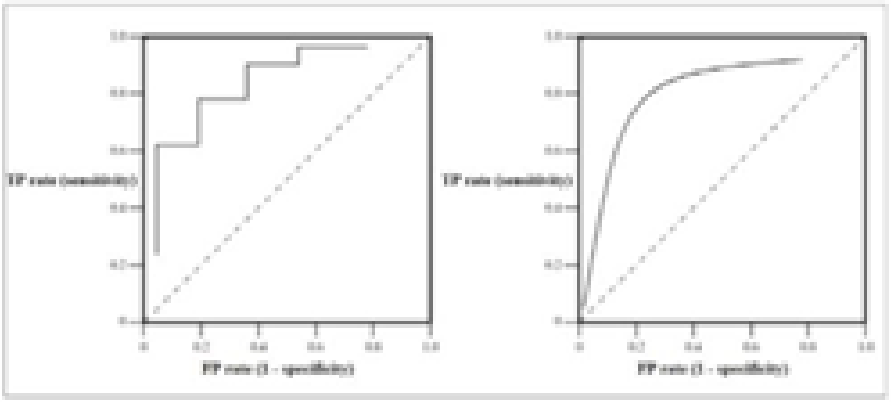
Metode tersebut terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan kedalam database dengan data besar. Berikut adalah formula umum *Naive Bayes*:

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

Untuk melihat akurasi penggukuran Metode klasifikasi digunakan *Confution Matrix* dan ROC *Curve/ Area Under Curve* (AUC).

Penentuan kurva ROC-AUC (*Area Under Curve*) adalah sebagai berikut:

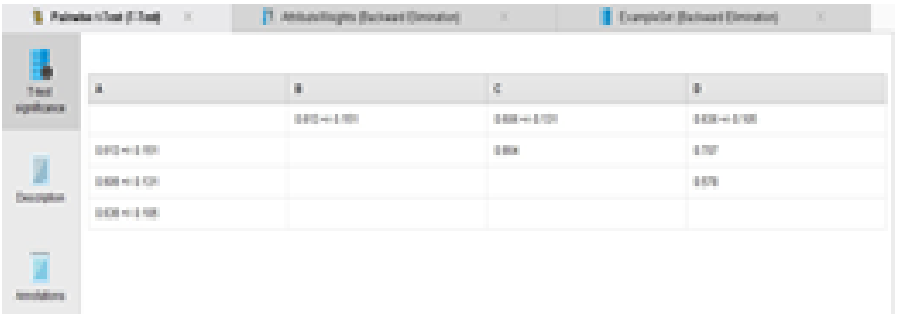
a. Kurva ROC adalah grafik dua dimensi dimana laju TP di plot pada sumbu Y dan laju FP di plot pada sumbu X.
b. Kurva ROC menggambarkan hubungan relatif timbal balik antara manfaat (*True Positives*) dan kerugian (*False Positives*)



Gambar 10. Kurva ROC-AUC (Area Under Curve)

Kategori klasifikasi kurva ROC-AUC (*Area Under Curve*) sebagai berikut :

a. 0.90 – 1.00 = *Excellent Classification*
b. 0.80 – 0.90 = *Good Classification*
c. 0.70 – 0.80 = *Fair Classification*
d. 0.60 – 0.70 = *Poor Classification*
e. 0.50 – 0.60 = *Failure*



Gambar 11. Confusion Matrix Algoritma Naive Bayes

Hasil *Confusion Matrix* dari algoritma *Naive Bayes* dengan uji beda (T-Test) didapatkan akurasi untuk *feature selection* sebesar 0.61, untuk *Forward Selection* sebesar 0.60 dan *Backward Elimination* sebesar 0.64. hal tersebut menunjukkan performa terbaik dari *Backward Elimination*. Tetapi dalam penelitian ini tidak perlu menggunakan *Backward Elimination* dikarenakan hasil dari metode tersebut membuat penurunan hasil akurasi dari dataset awal. Dimungkinkan pemilihan metode akurasi dapat meningkat jika *record* data pemilihan metode kontrasepsi ditambah.

Dalam penelitian ini juga tidak menampilkan kurva ROC-AUC (*Area Under Curve*), hal ini dikarenakan dataset yang digunakan/ diolah mempunyai label *polynominal* yang pada keluaran permodelannya hanya memunculkan akurasi dari algoritma yang dipilih.

***Deployment***

Pola pengetahuan yang didapatkan dalam proses standar data mining perlu adanya tindak lanjut. Dalam penelitian ini didapatkan beberapa tindak lanjut yang harus dilakukan sebagai berikut:

a. Menghilangkan Atribut
   Hasil hubungan antar faktor dapat diketahui bahwa faktor istri bekerja dan kesibukan suami memiliki pengaruh sangat kecil pada penggunaan metode kontrasepsi, sehingga faktor tersebut dikatakan tidak penting.
b. Menambahkan Atribut
   Faktor agama dapat dijadikan alternatif baru untuk menghasilakn pola/informasi pada penggunaan metode kontrasepsi.
c. Fokus pada Atribut Jumlah Anak
   Atribut jumlah anak memiliki pengaruh tinggi pada pemilihan metode kontrasepsi, hal ini dapat dijadikan standar utama dalam penentuan metode kontrasepsi di RSIA Kusuma Pradja.
d. Pemilihan metode kontrasepsi menggunakan Algoritma *Naive Bayes*

## KESIMPULAN

Sebuah kerangka perbandingan diusulkan untuk memperoleh kinerja algoritma terbaik dalam pemilihan alat kontrasepsi dengan dataset pasien yang menggunakan alat kontrasepsi di RSIA Kusuma Pradja Semarang.

Hasil penelitian menunjukkan bahwa Naive Bayes merupakan algoritma terbaik dengan akurasi 63.67%. Sedangkan algoritma yang lain cenderung menghasilkan akurasi kecil, sehingga penentuan metode kontrasepsi pada RSIA Kusuma Pradja Semarang dapat digunakan Algoritma Naive Bayes.

## DAFTAR PUSTAKA

Bandyopadhyay, G & Chattopadhyay. (2008). An Artificial Neural Net Approach to Forecast The Population of India. India.

BKKBN. Nd. Cara-Cara Kontrasepsi yang Digunakan Dewasa Ini. Diambil dari: http://www.bkkbn-jatim.go.id/bkkbnjatim/html/cara.htm. (3 Desember 2014).

Badan Pusat Statistik. nd. Laju Pertumbuhan Penduduk Menurut Provinsi. Diambil dari: http://bps.go.id/tab_sub/view.php?tab el=1&daftar=1&id_subyek=12&nota b=2. (3 November 2014).

H. Jiawei, M. Kamber, J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012

Larose, D. (2005). Discovering Knowledge in Data. New Jersey, John Willey & Sons.Inc.

Liao, Warren. T. & Triantaphyllou. Evangelos. (2007). Recent Advances in Data Mining of Enterprise Data: Algorithms and Applications. Series: Computer and Operation Research. 6. 190.

Lim TS, Loh WY, Shih YS. (1999). A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. Kluwer Academic Publishers: Boston.

Liu, Huan, Yu, Lei.(2005). Toward Integrating Feature Selection Algorithms for Classification and Clustering. Department of Computer Science and Engineering. Arizona State University.

Liu, Yuaning, Wang G., Chen, M., Dong, M., Zhu, X., Wang, S. (2011). An Improved Particle Swarm Optimization for Feature Selection. College of Computer Science and Technology. China.

Matatov, N., Rokach, L., & Maimon, O. (2010). Privacy-preserving data mining: A feature set partitioning approach. *Information Sciences*, *180*(14), 2696–2720.

Makhabel, B. (2015), *Learning Data Mining with R. Packt Publishing. Birmingham*: *Packt Publishing Ltd*.

North, M. (2012). *Data Mining for the Masses. Computer Global Text Project*.