

## **ANALISIS PENENTUAN KARYAWAN TERBAIK MENGGUNAKAN METODE ALGORITMA NAIVE BAYES (STUDI KASUS PT. XYZ)**

**Ulfa Pauziah**

Teknik Informatika, Universitas Indraprasta PGRI  
[pelangi\\_ulfa@yahoo.com](mailto:pelangi_ulfa@yahoo.com)

**Abstrak.** Di dalam dunia pekerjaan adanya karyawan terbaik menjadi tolak ukur kemajuan dari perusahaan itu sendiri. Dalam penentuan biasanya dengan melihat kinerja karyawan tersebut misal dari kerajinan, kedisiplinan dan juga prestasi lainnya. Dengan cara seperti ini agak kurang efektif dan akurat, oleh karena itu dalam penelitian menganalisa penentuan karyawan terbaik dengan menggunakan metode algoritma naive bayes, dengan bantuan datamining Penelitian ini bertujuan untuk mengukur tingkat akurasi dari kajian algoritma naive bayes dalam penentuan karyawan terbaik di PT. XYZ. Pengujian dilakukan dengan mengukur kinerja algoritma tersebut menggunakan bantuan aplikasi weka, dimana dilakukannya dua pengujian. Pertama pengujian dengan Cross Validation dan yang kedua dengan Confusion Matrix serta Kurva ROC.

Kata kunci: *algoritma Naïve Bayes, Datamining, karyawan*

### **PENDAHULUAN**

Didalam dunia pekerjaan adanya karyawan terbaik menjadi tolak ukur kemajuan perusahaan itu sendiri. Dalam penentuan biasanya dengan melihat kinerja karyawan tersebut misal dari kerajinan dan kedisiplinan dan ada juga dengan prestasi lainnya. Dengan cara penentuan seperti ini agak kurang efektif. Agar para karyawan di pt ini dapat terpacu lagi dalam meningkatkan kinerjanya dalam bekerja jadi untuk hal ini peneliti akan melakukan penentuan karyawan terbaik dengan menggunakan metode naive bayes agar hasil yang didapat akurat dan memudahkan pihak perusahaan tersebut dalam melakukan penentuan karyawan tersebut.

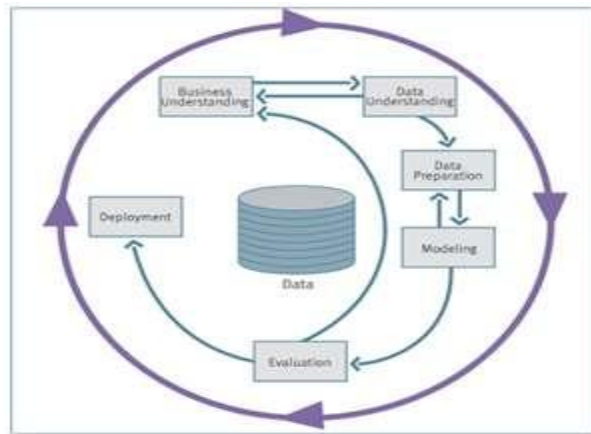
### **PEMBAHASAN**

Menurut kamus besar bahasa Indonesia, karyawan adalah orang yang bekerja pada suatu lembaga (kantor, perusahaan dan sebagainya) dengan mendapat gaji (upah). Menurut Hasibuan (dalam Manulang, 2002), karyawan adalah orang penjual jasa (pikiran dan tenaga) dan mendapat kompensasi yang besarnya telah ditetapkan terlebih dahulu.

Gartner group menyebutkan bahwa data mining adalah proses menelusuri pengetahuan yang baru, pola, dan tren yang dipilah dari jumlah data yang besar yang disimpan dalam repositori atau tempat penyimpanan dengan menggunakan teknik pengenalan pola serta statistik dan teknik matematika[2].

Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan didalam database. Data Mining adalah proses yang menggunakan teknik statistic, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai database besar (Turban, 2005).

Data *Mining* adalah sebuah proses, yang mana dalam melakukan prosesnya harus sesuai dengan prosedur dari proses tersebut, yaitu CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang terdiri dari keseluruhan proses, preprosesing data, pembentukan model, model evaluasi, dan tahap akhir penyebaran model [2]. Gambar 1 merupakan proses *Data Mining* CRISP-DM sebagai berikut :



Gambar 1. Proses CRISP-DM [2]

1. *Business/Research Understanding Phase* (Fase Pemahaman Bisnis)
  - a. penentuan tujuan proyek dan kebutuhan secara detail dalam lingkup bisnis atau unit penelitian secara keseluruhan.
  - b. Menerjemahkan tujuan dan batasan menjadi formula dari permasalahan *data mining*.
  - c. Menyiapkan strategi awal untuk mencapai tujuan.
2. *Data Understanding Phase* (Fase Pemahaman Data)
  - a. Mengumpulkan data
  - b. Menggunakan analisis penyelidikan data untuk mengenali lebih lanjut data dan pencarian pengetahuan awal.
  - c. Mengevaluasi kualitas data.
  - d. Jika diinginkan, pilih sebagian kecil group data yang mungkin mengandung pola dari permasalahan.
3. *Data Preparation Phase* (Fase Pengolahan Data)
  - a. Siapkan dari data awal, kumpulan data yang akan digunakan untuk keseluruhan fase berikutnya.
  - b. Pilih kasus dan variable yang ingin dianalisis dan yang sesuai analisis yang akan dilakukan.
  - c. Lakukan perubahan pada beberapa variabel jika dibutuhkan.
  - d. Siapkan data awal sehingga siap untuk perangkat pemodelan.
4. *Modeling Phase* (Fase Pemodelan)
  - a. Pilih dan aplikasikan teknik pemodelan yang sesuai.
  - b. Kalibrasi atur model untuk mengoptimalkan hasil.
  - c. Perlu diperhatikan bahwa beberapa teknik mungkin untuk digunakan pada permasalahan *data mining* yang sama.
  - d. Jika diperlukan, proses dapat kembali ke fase pengolahan data untuk menjadikan data kedalam bentuk yang sesuai dengan spesifikasi kebutuhan *data mining* tertentu.
5. *Evaluation Phase* (Fase Evaluasi)
  - a. Mengevaluasi satu atau lebih model yang digunakan dalam fase pemodelan untuk mendapatkan kualitas dan eektivitas sebelum disebarkan untuk digunakan.
  - b. Menetapkan apakah terdapat model yang memenuhi tujuan pada fase awal.
  - c. Menentukan apakah terdapat permasalahan penting dari bisnis atau penelitian yang tidak tertangani dengan baik.
  - d. Mengambil keputusan berkaitan dengan penggunaan hasil dari *data mining*.

6. *Deployment Phase* (Fase Penyebaran)
  - a. Menggunakan model yang dihasilkan. Terbentuknya model tidak menandakan telah terselesaikannya proyek.
  - b. Contoh sederhana penyebaran : Pembuatan laporan
  - c. Contoh kompleks penyebaran : Penerapan proses *data mining* secara parallel pada departemen lain.

Kemajuan luar biasa yang terus berlanjut dalam bidang Data mining didorong oleh beberapa faktor, antara lain (Larose, 2005):

1. Pertumbuhan yang cepat dalam pengumpulan data
2. Penyimpanan data dalam data warehouse, sehingga seluruh perusahaan memiliki akses kedalam database yang andal
3. Adanya peningkatan akses data melalui navigasi web dan intranet
4. Tekanan kompetisi bisnis untuk meningkatkan penguasaan pasar dalam globalisasi ekonomi
5. Perkembangan teknologi perangkat lunak untuk Data mining (ketersediaan teknologi)
6. Perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.

Mengacu pada Larose (2005), fungsi-fungsi *data mining* terdapat enam fungsi yaitu (Susanto dan Suryadi, 2010) :

1. Fungsi dekripsi  
Fungsi dekripsi adalah cara yang digunakan untuk menggambarkan sekumpulan data secara ringkas. Banyak cara yang digunakan dalam memberikan gambaran secara ringkas bagi sekumpulan data yang besar jumlahnya dan banyak macamnya yaitu deskripsi grafis, deskripsi lokasi, dan deskripsi keragaman.
2. Fungsi Estimasi  
Fungsi estimasi adalah fungsi untuk memperkirakan suatu hal yang sudah ada datanya. Fungsi estimasi terdiri dari dua cara yaitu estimasi titik dan estimasi selang kepercayaan
3. Fungsi Prediksi  
Fungsi prediksi adalah memperkirakan hasil dari hal yang belum diketahui, untuk mendapatkan hal baru yang akan muncul selanjutnya. Cara memprediksi dalam fungsi ini adalah regresi linier.
4. Fungsi Klasifikasi  
Fungsi klasifikasi atau menggolongkan suatu data, cara yang digunakan terdiri dari algoritma *mean vector*, *algoritme K-nearest Neighbor*, algoritma ID3, algoritma C4.5, algoritma C5.0.
5. Fungsi Pengelompokan  
Pada fungsi pengelompokan data yang dikelompokkan disebut objek atau catatan yang memiliki kemiripan atribut kemudian dikelompokkan kepada kelompok yang berbeda.
6. Fungsi Asosiasi  
Fungsi asosiasi adalah untuk menemukan aturan asosiasi yang mampu mengidentifikasi item-item yang menjadi objek.

Algoritma Naive bayes merupakan salah satu metode pengklasifikasi berpeluang sederhana yang berdasarkan pada penerapan Teorema Bayes dengan asumsi antar variabel penjelas saling bebas (independen). Dua kelompok peneliti, satu oleh Pantel dan Lin, dan yang lain oleh Microsoft Research memperkenalkan metode statistik Bayesian ini pada teknologi anti spam filter. Tetapi yang membuat algoritma Bayesian filtering ini populer adalah pendekatan yang dilakukan oleh Paul Graham Weka (*Waikato Environment for Knowledge Analysis*) adalah sebuah alat (*tool*) yang merupakan aplikasidata mining berbasis *open source (GPL)* yang ditulis dengan Java. *Software* ini mulai dikembangkan sejak tahun 1994 oleh University of Waikato, *New Zealand*. Kelebihan dari Weka yaitu, mudah digunakan, selalu *up to dated* dengan algoritma-algoritma baru, Teknik klasifikasi dan algoritma yang digunakan di Weka disebut *classifier*. Weka berbasis *GUI (Graphical Interface User)* dan dapat digunakan untuk mengintegrasikan metode baru yang dibuat sendiri dengan beberapa

ketentuan, selain itu Weka tidak saja digunakan untuk akademik namun banyak juga dipakai untuk memprediksi suatu bisnis perusahaan.

Untuk menguji model, pada penelitian ini, digunakan metode *Cross Validation*, *Confusion Matrix*, dan kurva ROC (*Receiver Operating Characteristic*).

### 1. Confusion Matrix

Metode ini hanya menggunakan table matriks seperti pada Tabel 1, jika data set hanya terdiri dari dua kelas, kelas yang satu dianggap sebagai positif dan yang lainnya negatif [3]

Tabel 1 Tabel Model Confision Matrix

Correct classification	Classified as	
	+	-
+	true positives	false negatives
-	false positives	true negatives

*True Positive* adalah jumlah record positif yang diklasifikasikan sebagai positif, *false positive* adalah jumlah record negative yang diklasifikasikan sebagai positif, *false negative* adalah jumlah record positif yang diklasifikasikan sebagai negative, *true negative* adalah jumlah record negative yang diklasifikasikan sebagai negative, kemudian masukkan data uji. Setelah data uji dimasukkan ke dalam *confusion matrix*, hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity (recall)*, *Specifity*, *precision*, dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah TP terhadap jumlah record yang positif sedangkan *Specifity*, *precision* adalah perbandingan jumlah TN terhadap jumlah record yang negative. Untuk menghitung digunakan persamaan dibawah ini [4]:

$$\text{Sensitifity} = \frac{TP}{P}$$

$$\text{Specifity} = \frac{TN}{N}$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$\text{accuracy} = \text{Sensitivity} \frac{P}{(P+N)} + \text{Specifity} \frac{N}{(P+N)}$$

Keterangan :

TP = Jumlah *true positives*

TN = Jumlah *true negative*

P = Jumlah *record positives*

N = Jumlah *tupel negatives*

FP = Jumlah *false positives*

### 2. Kurva ROC

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horizontal dan *true positive* sebagai garis vertical [5]. *The area under curve (AUC)* dihitung untuk mengukur perbedaan performansi metode yang digunakan. AUC digunakan dengan menggunakan rumus [6]:

$$\theta^r = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m \psi(xt^r, xj^r)$$

Dimana :

$$(x, y) = \begin{cases} 1 & y < x \\ \frac{1}{2} & y = x \\ 0 & y > x \end{cases}$$

Keterangan :

K = Jumlah algoritma klasifikasi yang dikomparasi

X = Output positif

Y = Output negatif

dikembangkan sejak tahun 1994 oleh University of Waikato, *New Zealand*.

### METODE

Jenis penelitian ini adalah penelitian eksperimental merupakan penelitian yang bersifat uji coba, memanipulasi dan mempengaruhi hal-hal yang terkait dengan seluruh variabel atau atribut.

Pengumpulan data dilakukan untuk mendapatkan informasi dan data yang diperoleh yang berhubungan dengan penelitian ini. Untuk mengumpulkan data dan informasi tersebut, dilakukan metode pengumpulan data sebagai berikut:

- Pengumpulan data primer  
Metode ini digunakan untuk mendapatkan data primer, yaitu dengan cara melakukan wawancara langsung dengan pihak pt tersebut.
- Pengumpulan data sekunder  
Data sekunder dikumpulkan dengan mengamati data, membaca, mempelajari dan mengutip dari buku literatur, serta sumber-sumber yang berhubungan erat dengan penelitian ini.

Teknik analisis data

Teknik Analisis data yang dipakai menggunakan Data Kuantitatif berupa kaidah-kaidah matematika terhadap data atau numerik. Analisa dilakukan melalui data nilai rata-rata hasil menilai kinerja yang dilihat dari atribut-atribut yang untuk menilai para karyawan. Dimana pengujian menggunakan pada masing-masing algoritma yaitu algoritma Naïve Bayes.

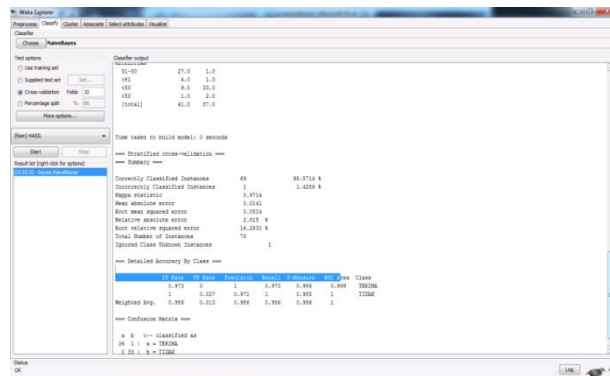
### HASIL DAN PEMBAHASAN

Dalam penerapan ada data yang diolah, dimana dalam mengolah data dibantu dengan aplikasi weka. Adapun data yang dipakai adalah data training yang berjumlah 70 data:

NO	KEBAPAHAN	LOYALITAS	KEPATIHAN	KEGOSPUNAN	PRODUKTIVITAS	KETELITAN	HASIL
1	1	<50	<50	<50	<50	<50	TERIMA
2	2	<50	<50	<50	<50	<50	TERIMA
3	3	<50	<50	<50	<50	<50	TERIMA
4	4	>81	<50	<50	<50	<50	TERIMA
5	5	<50	<50	<50	<50	<50	TERIMA
6	6	<50	<50	<50	<50	<50	TERIMA
7	7	<50	<50	<50	<50	<50	TIDAK
8	8	<50	<50	<50	<50	<50	TERIMA
9	9	>81	>81	<50	<50	<50	TERIMA
10	10	<50	<50	<50	<50	<50	TIDAK
11	11	>81	<50	<50	<50	<50	TERIMA
12	12	<50	<50	<50	<50	<50	TIDAK
13	13	<50	<50	<50	<50	<50	TIDAK
14	14	<50	<50	<50	<50	<50	TIDAK
15	15	<50	<50	<50	<50	<50	TIDAK
16	16	<50	<50	<50	<50	<50	TIDAK
17	17	<50	<50	<50	<50	<50	TIDAK
18	18	<50	<50	<50	<50	<50	TIDAK
19	19	<50	<50	<50	<50	<50	TIDAK
20	20	<50	<50	<50	<50	<50	TIDAK
21	21	<50	<50	<50	<50	<50	TIDAK
22	22	<50	<50	<50	<50	<50	TIDAK
23	23	<50	<50	<50	<50	<50	TIDAK
24	24	<50	<50	<50	<50	<50	TIDAK
25	25	<50	<50	<50	<50	<50	TIDAK
26	26	<50	<50	<50	<50	<50	TIDAK
27	27	<50	<50	<50	<50	<50	TIDAK

Gambar 3. Data Training





Gambar 7. Nilai akurasi algoritma Naïve bayes.

Dari hasil perhitungan algoritma naïve bayes diatas dengan menggunakan bantuan tools weka didapat hasil bahwa 98,5714 % algoritma naïve bayes dapat membantu dalam pengambilan keputusan pada penentuan karyawan terbaik, sedangkan 1,4286% tidak dapat membantu untuk pengambilan keputusan.

## 2. Evaluasi dan Validasi

Model yang telah dibentuk diuji tingkat akurasi dengan memasukan data uji yang berasal dari data *training*. Karena data yang didapat dalam penelitian ini setelah proses *preprocessing* hanya 70 data maka digunakan metode *cross validation* untuk menguji tingkat akurasi. Untuk nilai akurasi model untuk metode *naïve bayes* sebesar 98,5714 %, Selain itu dalam penelitian ini diuji juga dengan menggunakan *confusion matrix* dan kurva ROC

### a. Confusion matrix

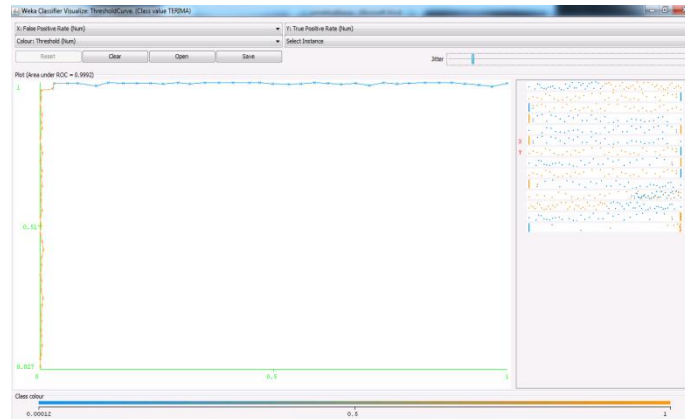
Tabel 5.1 adalah perhitungan berdasarkan data *training*, diketahui dari 70 data, 43 diklasifikasikan terima, lalu 32 data diprediksi tidak.

Tabel 2. Model *Confusion Matrix* untuk Metode Naïve Bayes

```

=== Confusion Matrix ===
      a  b  <-- classified as
 36  1  | a = TERIMA
  0 33  | b = TIDAK
    
```

a. Kurva ROC



Gambar 8. kurva ROC

Kurva ROC pada gambar 5.15 diatas mengekspresikan *confusion matrix* Garis X adalah *false positives* dan garis Y *true positives*. Sebesar 0.9992.

3. Impilkasi penelitian

Hasil dari penelitian yang telah dilakukan ini diharapkan dapat memberikan inspirasi dan masukan bagi pihak perusahaan tersebut untuk dapat memanfaatkan metode algoritma naïve bayes yang dapat membantu dalam pengambilan keputusan pada penentuan karyawan terbaik..

Implikasi dari temuan penelitian ini mencakup pada dua bagian aspek, yaitu manajerial dan sistem.

**1. Aspek Manajerial**

Dari hasil evaluasi ternyata Algoritma Naïve Bayes terbukti dapat dipakai dalam penentuan karyawan terbaik..Dengan hasil ini, algoritma Naïve Bayes merupakan algoritma yang dapat membantu dan dapat diterapkan untuk memberikan pemecahan untuk permasalahan penentuan karyawan terbaik.

**2. Aspek sistem**

a. Penelitian semacam ini dapat dikembangkan pada unit bisnis serupa atau yang lain. Penelitian ini dapat dikembangkan dengan algoritma yang lain misalkan saja dengan metode *Support Vector Machinne* (SVM).

b. Untuk mendukung pengambilan keputusan dan pengembangan system informasi model ini dapat diterapkan pada sekolah menggunakan *software* Weka.

**PENUTUP**

**Simpulan**

Dapat Berdasarkan penelitian yang dilakukan dapat diambil kesimpulan sebagai berikut:

1. Adanya bentuk dan fungsi aplikasi yang dapat dipakai untuk perusahaan- perusahaan, yang bisa digunakan untuk menghitung layak atau tidaknya seorang karyawan menjadi karyawan terbaik di perusahaan tersebut.
2. Mempermudah bagi pihak perusahaan dalam menentukan karyawan terbaik.

**Saran**

Penelitian ini dapat dikembangkan lagi, agar lebih banyak obyek yang bisa mengetahui aplikasi system keputusan ini.



**DAFTAR PUSTAKA**

- Larose, Daniel. T. (2005). **Discovering Knowledge in Data**. New Jersey: John Willey & Sons, Inc.
- Asda Bramer, Max. (2007). **Principles of Data Mining**. London : Springer
- Han, J. & Kamber, M. (2006). **Data Mining Concept and Tehniques**. San Fransisco: Morgan Kauffman.
- Vercellis, Carlo. (2009). **Business Intelegent: Data Mining and Optimization for Decision Making**. Southern Gate, Chichester, west Sussex: John Willey & Sons, Ltd, 2009.
- Liao. (2007). **Recent Advances in Data Mining of Enterprise Data: Algorithms and Application**. Singapore: WorldScientific Publishing
- <http://kbbi.web.id/karyawan>