

ANALISA ALGORITMA NAÏVE BAYES UNTUK KLASIFIKASI KARYAWAN TELADAN PADA PT. TOYOSEAL INDONESIA

Candra Naya¹⁾, Muhlisin²⁾

Program Studi Teknik Informatika Fakultas Teknik
Universitas Pelita Bangsa
candranya@pelitabangsa.ac.id

Disetujui, 30 Desember 2019

Abstraksi

Karyawan adalah orang yang bekerja pada suatu lembaga (kantor, perusahaan dan sebagainya) dengan mendapat gaji (upah) atau orang penjual jasa (pikiran dan tenaga) dan mendapat kompensasi yang besarnya telah ditetapkan terlebih dahulu. Tujuan dalam pengujian yaitu agar memiliki hasil yang baik dan keunggulan pemilihan data dalam sistem pengenalan pola dengan kemampuan generalisasi yang baik dalam menentukan atau klasifikasi data karyawan teladan yang baik dalam meskipun dilatih dengan himpunan 5 atribut dan 160 data. Analisa dan hasil penelitian ini merupakan rangkuman hasil dari percobaan *algoritma naïve bayes*. Dengan melakukan pengujian untuk mengetahui hasil *accuracy*, *preccission*, dan *recall*. Data sampel yang diutarakn yaitu sebanyak 160 data karyawan teladan, kemudian hasil menyatakan tingkat *accuracy* 81.25 %, *Recall* 87.50 % dan *Percision* 77.78 % diputuskan dalam klasifikasi *naïve bayes* dalam klasifikasi data karyawan teladan menyatakan hasil teladan dengan tingkat *accuracy* 81.25%. Dalam menentukan karyawan teladan dan bukan teladan atau klasifikasi karyawan teladan dan bukan teladan, berdasarkan hasil yang telah didapat dalam klasifikasi karyawan teladan sehingga menghasilkan nilai akurasi *recall* diatas 50% dengan hasil optimal. Dari hasil yang sudah didapat dari semua *data training* dan *testing* dengan melakukan pengujian bahwa *algoritma naïve bayes* menghasilkan hasil yang baik .

Kata Kunci : *Data Mining*, Klasifikasi Karyawan Teladan, *Naïve Bayes*

Abstract

Employees are people who work at an institution (office, company and so on) with a salary (wages) or a service seller (mind and energy) and receive compensation, the amount of which has been determined in advance. The purpose of testing is to have good results and superior data selection in a pattern recognition system with good generalizability in determining or classifying good model employee data even though they are trained with the set of 5 attributes and 160 data. The analysis and results of this study are a summary of the results of the naïve Bayes algorithm experiment. By doing tests to find out the accuracy, preccission, and recall results. The sample data used were 160 exemplary employee data, then the results stated the accuracy level of 81.25%, Recall 87.50% and Percision 77.78% were decided in the naïve Bayes classification in the classification of exemplary employee data stating exemplary results with an accuracy level of 81.25%. In determining exemplary and non-exemplary employees or the classification of exemplary and non-exemplary employees, based on the results obtained in the classification of exemplary employees so as to produce a recall accuracy value above 50% with optimal results. From the results that have been obtained from all training and testing data by testing that the Naïve Bayes algorithm produces good results.

Keywords: *Data Mining*, *Classification of Exemplary Employees*, *Naïve Bayes*

1. Pendahuluan

Sebuah perusahaan tidak akan berkembang tanpa adanya dukungan dari karyawan sebagai salah satu aset perusahaan, sehingga perusahaan dituntut untuk memiliki karyawan yang berpotensi dan berbakat untuk mendukung keberhasilan suatu perusahaan. Tenaga kerja yang terampil yang dapat membawa perusahaan berkembang dan bersaing dengan perkembangan zaman. Berhasil atau tidaknya suatu perusahaan dalam menjalankan kegiatannya tidak terlepas dari karyawan yang melakukan pekerjaan diperusahaan tersebut.

Karyawan merupakan aset terpenting yang memiliki pengaruh sangat besar terhadap kesuksesan sebuah perusahaan. Tanpa mesin canggih perusahaan dapat terus beroperasi dengan mesin manual, akan tetapi tanpa adanya karyawan perusahaan tidak akan dapat beroperasi. Untuk itu perusahaan dalam merekrut karyawan harus yang berkualitas dan mempunyai etos kerja yang baik agar dalam pelaksanaan proses produksi juga bisa menghasilkan produk yang berkualitas tinggi sehingga perusahaan dapat memuaskan pelanggan. Atas dasar ini maka perlunya menjaga salah satu aset dari perusahaan diluar dari aset tetap seperti bangunan dan tanah yakni karyawan. Pentingnya penghargaan pada karyawan merupakan salah satu apresiasi dari perusahaan untuk menjaga keberlangsungan suatu usaha.

Berdasarkan kondisi tersebut maka melalui penelitian ini dilakukan penerapan kemajuan teknologi informasi menggunakan *data mining* yang dapat membantu untuk mengklasifikasi karyawan terbaik pada PT. Toyoseal Indonesia. *Data mining* disebut sebagai proses ekstraksi pengetahuan dari data yang besar, sesuai fungsinya. *Data mining* adalah proses pengambilan keputusan dari *volume* data yang besar yang disimpan dalam basis data. Metode *data mining* yang digunakan untuk klasifikasi ini untuk mengetahui status karyawan, salah satunya yaitu menggunakan metode *naïve bayes*, sehingga data tersebut dapat digunakan secara maksimal. Dengan adanya masalah diatas, maka peneliti mengambil penelitian ini dengan judul “**Analisa Algoritma Naïve Bayes Untuk Klasifikasi Karyawan Teladan pada PT. Toyoseal Indonesia**”

2. Tinjauan Studi

2.1 Karyawan

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. (Pauziah, 2017).

2.2 Proses Tahapan *Data Mining*

Data mining adalah proses yang mempekerjakan satu atau lebih Teknik pembelajaran computer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi-definisi konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh spesifik dari konsep-konsep yang akan dipelajari. *Knowledge Discovery in Databases* (KDD) adalah penerapan metode saintifik pada data mining. Dalam konteks ini *data mining* merupakan satu langkah dari proses KDD. (Algoritma, 2017).

Data Mining adalah sebagai proses untuk mendapatkan informasi yang berguna dari gudang basis data yang besar. *Data mining* juga dapat diartikan sebagai pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. *Data mining* sering juga disebut dengan *Knowledge Discovery in Databases* atau disingkat KDD, adalah kegiatan yang

meliputi pengumpulan, pemakaian data histori untuk menulisi data yang ada untuk membangun sebuah model agar dapat mengenali pola data yang berukuran besar (Yanto & Kesuma, 2017).

2.3 Algoritma *Naïve Bayes*

Algoritma *Naïve Bayes* Pengklasifikasi bayesian adalah pengklasifikasi statistik dan didasarkan pada teorema bayes. Teori keputusan bayes adalah pendekatan statistik yang fundamental dalam pengenalan pola (pattern recognition), penggunaan algoritma ini dalam hal klasifikasi harus mempunyai masalah yang bisa dilihat statistiknya. Misalkan X adalah set atribut data dan h kelas variabel dan jika kelas memiliki hubungan dengan atribut maka diperlukan X dan h sebagai variabel acak dan menangkap hubungan peluang $P(h|X)$ ini peluang posterior untuk h dan sebaliknya perior $P(h)$. (Suyanto, 2017). Naive Bayes mengestimasi peluang kelas bersyarat dengan mengasumsikan bahwa atribut adalah independen secara bersyarat yang diberikan dengan label kelas label kelas label kelas y dengan tiap set atribut $X = \{X_1, X_2, \dots, X_d\}$ terdiri dari d atribut.

3. Desain Penelitian/Metodologi

3.1 Objek Penelitian

Objek penelitian adalah topik permasalahan yang dikaji pada suatu penelitian. Adapun objek penelitian ini yaitu menggunakan data karyawan pada PT. Toyoseal Indonesia, yang mana PT. Toyoseal Indonesia adalah sebuah perusahaan PMA Jepang yang didalamnya memproduksi sparepart otomotif. PT. Toyoseal Indonesia Mempunyai banyak karyawan yang saat ini masih bekerja. Sukses atau tidaknya sebuah usaha adalah

tergantung dari kinerja para karyawan yang ada di dalamnya, untuk itu sebuah perusahaan untuk bisa berkembang dengan baik maka perusahaan tersebut harus memiliki pekerja yang berprestasi atau pekerja yang memiliki etos kerja yang tinggi serta para karyawan harus memiliki rasa tanggung jawab yang penuh terhadap pekerjaannya. Peneliti melakukan identifikasi terhadap masalah yang ada, kemudian peneliti akan menetapkan penyelesaian dari suatu masalah yang ditemukan pada penelitian ini.

3.2 Data Yang Digunakan

Pada penelitian ini data yang digunakan adalah data sekunder. Data sekunder yaitu berupa data yang didapatkan dari pihak-pihak tertentu yang dapat diolah sebagai bahan analisis untuk melengkapi data yang ditemukan di lapangan yang terkait dengan objek

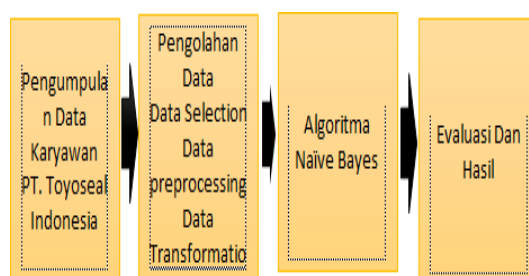
penelitian. Data sekunder tersebut merupakan alat pelengkap dari metode observasi dan wawancara yang berfungsi sebagai data awal penelitian.

Dari *dataset* ini didapatkan jumlah *record* atau data sebanyak 800 data, dengan atribut termasuk atribut kelas, yaitu:

1. Nama lengkap
Nama lengkap adalah nama-nama dari karyawan yang bekerja pada PT. Toyoseal Indonesia.
2. Divisi
Divisi merupakan bagian atau tempat karyawan melakukan pekerjaannya pada PT. Toyoseal Indonesia.
3. Jabatan
Jabatan adalah status suatu kepercayaan atau tugas yang diberikan kepada karyawan yang di dalamnya berfungsi untuk membagi pekerjaan yang akan dilakukan pada perusahaan tersebut.
4. *Attendance*
Attendance atau kehadiran adalah suatu informasi mengenai datang atau tidaknya suatu karyawan untuk bekerja pada perusahaan tersebut.
5. *Grade*
Grade merupakan nilai yang diberikan oleh atasan kepada bawahan dengan kriteria tertentu.
6. *Attitude*
Attitude atau sikap adalah suatu tindakan yang ada pada suatu karyawan berdasarkan baik atau tidaknya karyawan tersebut.
7. *Category* merupakan kelas yang menjadi target perhitungan.

3.3 Tahap Penelitian

Pada tahap ini, data yang digunakan adalah data karyawan pada PT. Toyoseal Indonesia. Data tersebut akan diolah dengan menggunakan *Knowledge Discovery in Database* (KDD), sebagai berikut:



Gambar 1. Tahap Penelitian

4. Hasil Penelitian dan Pembahasan

4.1 Split validation

Berdasarkan data yang ada, maka peneliti akan membagi data menjadi dua yaitu *data training* dan *data testing*. Pada penelitian ini peneliti mengambil 800 data untuk dijadikan *data training* dan *data testing*, *Split Validation* merupakan teknik validasi yang membagi data menjadi dua bagian, sebagian *data training* dan sebagian *data testing*. Data yang sudah disiapkan untuk klasifikasi dibagi menjadi dua menggunakan teknik *sampling random* untuk *data training* 640 (80%) dan *data testing* 160 (20%). Contoh perhitungan untuk pengambilan *data testing*. *Dataset* akan dibagi menjadi *data training* dan *data testing*. Proses *training* dan *testing* dilakukan sebanyak 5 kali secara berulang-ulang. Pada iterasi ke-1, partisi *D1* disajikan sebagai *data*

testing dan partisi sisanya digunakan secara bersamaan dan berurutan sebagai data *training*. Iterasi kedua, subset D_1, D_2, \dots, D_k akan dites pada D_2 , Iterasi ketiga, subset $D_1, D_2, D_3 \dots, D_k$ akan dites pada D_3 , dan selanjutnya hingga D_5).

Data yang sudah disiapkan untuk klasifikasi dibagi menjadi dua menggunakan Teknik *sampling random* untuk *data training* 80% dan *data testing* 20%. Contoh perhitungan untuk pengambilan data testing adalah sebagai berikut:

$$\begin{aligned} \text{Jumlah data keseluruhan (N)} &= 800 \\ \text{Jumlah data testing} &= 20\% \times 800 = 160 \\ \text{Internal sampling (k)} &= 800/160 = 5 \end{aligned}$$

Unsur pertama yang diambil untuk *data testing* adalah $(k)=1$

Tabel 1. Model *Split Validation*

<i>Split validation</i>	<i>Data training</i>	<i>Data Testing</i>
k1	640	160
k2	640	160
k3	640	160
k4	640	160
k5	640	160

Dari hasil diatas diperoleh data testing sebanyak 160 data, maka sisanya dijadikan data training sebanyak $800 - 160 = 640$ data. Pembahasan dilakukan guna untuk mendapatkan nilai accuracy, precision, untuk algoritma naïve bayes, memprediksi, klasifikasi dos , recall. Pengertian akurasi adalah tingkat kedekatan hasil prediksi dengan hasil fakta. Presisi adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Serta recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Perhitungan dari algoritma naïve bayes, akurasi dilakukan dengan cara jumlah TP + TN dibagi jumlah total data testing yang diuji.

accuracy: 81,25%			
	true Bukan Teladan	true Teladan	class precision
pred. Bukan Teladan	14	4	77,78%
pred. Teladan	2	12	85,71%
class recall	87,50%	75,00%	

Gambar 2. Accuracy

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} * 100\% \\ &= \frac{14 + 12}{14 + 12 + 4 + 2} * 100\% \\ &= \frac{26}{32} * 100\% \\ &= 0.8125 * 100\% \\ &= 81.25\% \end{aligned}$$

Nilai *precision* dihitung dengan cara membagi jumlah data benar yang bernilai positif (*True Positive*) dibagi dengan jumlah data benar yang bernilai positif (*True Positive*) dan data salah yang bernilai positif (*False Positive*).

accuracy: 81,25%

	true Bukan Teladan	true Teladan	class precision
pred. Bukan Teladan	14	4	77,78%
pred. Teladan	2	12	85,71%
class recall	87,50%	75,00%	

Gambar 3. Precision

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FP} * 100\% \\
 &= \frac{14}{14 + 4} * 100\% \\
 &= \frac{14}{18} * 100\% \\
 &= 0.7778 * 100\% \\
 &= 77.78\%
 \end{aligned}$$

Nilai *recall* dihitung dengan cara membagi data benar yang bernilai positive (*True Positive*) dengan hasil penjumlahan dari data benar yang bernilai positif (*True Positive*) dan data salah yang bernilai negatif (*False Negative*).

accuracy: 81,25%

	true Bukan Teladan	true Teladan	class precision
pred. Bukan Teladan	14	4	77,78%
pred. Teladan	2	12	85,71%
class recall	87,50%	75,00%	

Gambar 4. Recall

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FN} * 100\% \\
 &= \frac{14}{14 + 2} * 100\% \\
 &= \frac{14}{16} * 100\% \\
 &= 0.875 * 100\% \\
 &= 87.50\%
 \end{aligned}$$

Tabel 2. Prediksi

No	Algoritma	Accucry	Percision	Recall
1	Naïve Bayes	81.25%	77.78%	87.50%

4.2 Analisa Hasil

Analisa dan hasil penelitian ini merupakan rangkuman hasil dari percobaan *algoritma naïve bayes*. Dengan melakukan pengujian untuk mengetahui hasil *accuracy*, *precision*, dan *recall*. Dalam menentukan karyawan teladan dan bukan teladan atau klasifikasi karyawan teladan dan bukan teladan, berdasarkan hasil yang telah didapat dalam klasifikasi karyawan teladan sehingga menghasilkan nilai akurasi *recall* diatas 50% dengan hasil optimal. Dari hasil yang sudah didapat dari semua *data training* dan testing dengan melakukan pengujian bahwa algoritma naïve bayes menghasilkan hasil yang baik dalam mengklasifikasi data karyawan PT. Toyoseal Indonesia. Dari data testing yang diutarakan yaitu sebanyak 160 data, kemudian hasil dari data karyawan teladan menyatakan tingkat ke *accuracy*, *recall* dan *percision* naïve bayes, tingkat *accuracy* 81.25 %, *Recall* 87.50 % dan *Percision* 77.78 % putuskan dalam klasifikasi prediksi naïve bayes. Model model for label attribute Prediksi Naïve Bayes, hasil dari penelitian ini merupakan rangkuman hasil percobaan dari algoritma tersebut dalam menghasilkan hasil optimal dalam klasifikasi karyawan teladan

5. Kesimpulan

5.1 Kesimpulan

Berdasarkan hasil pengujian dengan menggunakan *algoritma Naïve Bayes* dapat diambil suatu kesimpulan sebagai berikut :

1. Hasil dari penelitian data karyawan menyatakan tingkat *accuracy* 81.25 %, *Recall* 87.50 % dan *Percision* 77.78 % putuskan dalam klasifikasi *naïve bayes* dalam klasifikasi data karyawan teladan menyatakan hasil teladan dengan tingkat *accuracy* 81.25%.
2. Hasil yang baik dan keunggulan pemilihan data dalam sistem pengenalan pola dengan kemampuan generalisasi yang baik dalam menentukan data karyawan teladan yang baik meskipun dilatih dengan himpunan, pemodelan label atau atribut hasil klasifikasi dari *naïve bayes* atau pengujian tersebut menunjukkan, menyimpulkan *algoritma naïve bayes* lebih optimal dalam penentuan karyawan teladan dan bukan teladan (kinerja).

Daftar Pustaka

- Algoritma, C. (2017). *DATA MINING UNTUK MEMREDIKSI JENIS TRANSAKSI NASABAH PADA KOPERASI SIMPAN PINJAM DENGAN*. 1(2), 32–37.
- Ardiansyah Sembiring, M., Fitri Larasati Sibuea, M., Sapta, A., Studi Sistem Informasi, P., & Royal, S. (2018). Analisa Kinerja Algoritma C.45 Dalam Memprediksi Hasil Belajar. *Journal of Science and Social Research*, 1(February), 73–79. Retrieved from <http://jurnal.goretanpena.com/index.php/JSSR>. Data Mining, Suyanto : jogyakarta. 2017.
- Gunadi, G., & Senses, D. I. (2012). Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth (Fp-Growth) : *Telematika*, 4(1), 118–132.
- Kusumo, D. S., Bijaksana, M. A., & Darmantoro, D. (2016). Data Mining Dengan Algoritma Apriori Pada Rdbms Oracle. *TEKTRIKA - Jurnal Penelitian Dan Pengembangan Telekomunikasi, Kendali, Komputer, Elektrik, Dan Elektronika*, 8(1), 1–5. <https://doi.org/10.25124/tektrika.v8i1.215>
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 2(2), 213–219.
- Pauziah, U. (2017). Analisis Penentuan Karyawan Terbaik Menggunakan Metode Algoritma Naïve Bayes (Studi Kasus PT. XYZ). *Prosiding Diskusi Panel Pendidikan “Menjadi Guru Pembelajar,”* 1(1), 94–102.
- Penerapan Data Mining Retno Tri Vulandari, jogyakarta 2017.

- Purwanto, E., & Kunci, K. (2019). *PREDIKSI KELULUSAN TEPAT WAKTU MENGGUNAKAN METODE C4 . 5 DAN K - NN (Studi Kasus : Mahasiswa Program Studi S1 Ilmu Farmasi , Fakultas Universitas Muhammadiyah Purwokerto)*. 20(2).
- Septiani, W. D. (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis. *None*, 13(1), 76–84.
- Silalahi, Kristiani Desri., Murfi, Hendri., Satria, Yudi. (2017)..Studi Perbandingan Pemilihan Fitur untuk *Support Vector Machine* pada Klasifikasi Penilaian Risiko Kredit.
- Yanto, R., & Kesuma, H. Di. (2017). Pemanfaatan Data Mining Untuk Penempatan Buku Di Perpustakaan Menggunakan Metode Association Rule. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 4(1), 1–10. <https://doi.org/10.35957/jatisi.v4i1.83>