**Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)** DOI: 10.25126/jtiik.201854803
Vol. 5, No. 4, September 2018, hlm. 455-464 p-ISSN: 2355-7699
Akreditasi KEMENRISTEKDIKTI, No. 51/E/KPT/2017 e-ISSN: 2528-6579

# PERBANDINGAN ALGORITMA NAIVE BAYES DAN C.45 DALAM KLASIFIKASI DATA MINING

**Yogiek Indra Kurniawan**

Informatika, Universitas Muhammadiyah Surakarta
Email: yogiek@ums.ac.id

(Naskah masuk: 09 Mei 2018, diterima untuk diterbitkan: 05 September 2018)

**Abstrak**

Pada paper ini, telah diterapkan metode *Naive Bayes* serta *C.45* ke dalam 4 buah studi kasus, yaitu kasus penerimaan "Kartu Indonesia Sehat", penentuan pengajuan kartu kredit di sebuah bank, penentuan usia kelahiran, serta penentuan kelayakan calon anggota kredit pada koperasi untuk mengetahui algoritma terbaik di setiap kasus. Setelah itu, dilakukan perbandingan dalam hal *Precision*, *Recall* serta *Accuracy* untuk setiap data training dan data testing yang telah diberikan. Dari hasil implementasi yang dilakukan, telah dibangun sebuah aplikasi yang dapat menerapkan algoritma *Naive Bayes* dan *C.45* di 4 buah kasus tersebut. Aplikasi telah diuji dengan blackbox dan algoritma dengan hasil valid dan dapat mengimplementasikan kedua buah algoritma dengan benar. Berdasarkan hasil pengujian, semakin banyaknya data training yang digunakan, maka nilai *precision, recall* dan *accuracy* akan semakin meningkat. Selain itu, hasil klasifikasi pada algoritma *Naive Bayes* dan *C.45* tidak dapat memberikan nilai yang absolut atau mutlak di setiap kasus. Pada kasus penentuan penerimaan Kartu Indonesia Sehat, kedua buah algoritma tersebut sama-sama efektif untuk digunakan. Untuk kasus pengajuan kartu kredit di sebuah bank, C.45 lebih baik daripada Naive Bayes. Pada kasus penentuan usia kelahiran, Naive Bayes lebih baik daripada C.45. Sedangkan pada kasus penentuan kelayakan calon anggota kredit di koperasi, Naive Bayes memberikan nilai yang lebih baik pada precision, tapi untuk recall dan accuracy, C.45 memberikan hasil yang lebih baik. Sehingga untuk menentukan algoritma terbaik yang akan dipakai di sebuah kasus, harus melihat kriteria, variable maupun jumlah data di kasus tersebut.

**Kata kunci**: *accuracy, C.45, Klasifikasi, Naive Bayes, Precision, Recall*

# *COMPARISON OF NAIVE BAYES AND C.45 ALGORITHM IN DATA MINING CLASSIFICATION*

***Abstract***

*In this paper, applied Naive Bayes and C.45 into 4 case studies, namely the case of acceptance of "Kartu Indonesia Sehat", determination of credit card application in a bank, determination of birth age, and determination of eligibility of prospective members of credit to Koperasi to find out the best algorithm in each case. After that, the comparison in Precision, Recall and Accuracy for each training data and data testing has been given. From the results of the implementation, has built an application that can apply the Naive Bayes and C.45 algorithm in 4 cases. Applications have been tested in blackbox and algorithms with valid results and can implement both algorithms correctly. Based on the test results, the more training data used, the value of precision, recall and accuracy will increase. The classification results of Naive Bayes and C.45 algorithms can not provide absolute value in each case. In the case of determining the acceptance of the Kartu Indonesia Indonesia, the two algorithms are equally effective to use. For credit card submission cases at a bank, C.45 is better than Naive Bayes. In the case of determining the age of birth, Naive Bayes is better than C.45. Whereas in the case of determining the eligibility of prospective credit members in the cooperative, Naive Bayes provides better value in precision, but for recall and accuracy, C.45 gives better results. So, to determine the best algorithm to be used in a case, it must look at the criteria, variables and amount of data in the case.*

**Keywords**: *accuracy, classification, C.45, Naive Bayes, Precision, Recall*

## 1. PENDAHULUAN

Dalam perkembangan teknologi masa kini, banyaknya data menjadi sebuah permasalahan sekaligus kesempatan bagi sebuah instansi. Data menjadi permasalahan apabila tidak dapat disimpan, dikelola, maupun diproses dengan baik. Data yang selalu bermunculan setiap waktu akan terus menumpuk dan bila tidak didokumentasikan dengan baik, maka data tersebut akan menjadi tidak berguna untuk perusahaan. Sedangkan data menjadi sebuah kesempatan apabila dapat disimpan, dikelola dan diproses menjadi lebih berarti untuk instansi tersebut. Dengan adanya data, maka dapat ditemukan sebuah trend maupun struktur yang nantinya dapat dipergunakan untuk mendapatkan informasi di masa mendatang.

Pengelolaan data yang sangat besar akan melibatkan proses *data mining*. Roiger (2017) menyatakan *data mining* adalah proses maupun tahapan dalam menemukan sebuah struktur data. Struktur data tersebut dapat mengambil banyak bentuk, termasuk aturan, grafik atau jaringan, pohon (*tree*) maupun persamaan, serta beberapa yang lain. Dengan menggunakan data mining, maka sebuah kasus dapat dilihat *trend*, struktur maupun prediksinya di masa mendatang. *Data mining* sendiri memiliki banyak tahapan dan teknik yang dapat diimplementasikan dalam kehidupan nyata.

Dalam kasus di dunia nyata, banyak teknik dalam *data mining* yang dapat digunakan, salah satunya adalah teknik klasifikasi. Klasifikasi sendiri merupakan bentuk dasar dari analisis data. Bansar, Sharma & Goel (2017) menyatakan klasifikasi adalah sebuah teknik untuk menentukan keanggotaan kelompok berdasarkan data-data yang sudah ada. Konsep dasar dari klasifikasi adalah beberapa data yang memiliki struktur data yang mirip akan memiliki klasifikasi yang mirip pula. Sebagai contoh penerapan klasifikasi adalah untuk membagi-bagi sebuah data pada perkreditan, dengan memilih beberapa *variable* yang tepat, lalu menentukan dan mengklasifikasi apakah data seorang nasabah itu nantinya akan menjadi nasabah yang membayar tepat waktu, membayar terlambat atau bahkan tidak membayar. Klasifikasi ini dapat diterapkan dalam berbagai kasus untuk membentuk sebuah aturan (*rule*) terhadap sebuah data.

Dalam beberapa penelitian sebelumnya, telah banyak metode klasifikasi yang diimplementasikan dalam kehidupan nyata. Beberapa algoritma yang sangat populer saat ini adalah *Naive Bayes* dan *C.45*. *Naive Bayes* merupakan algoritma klasifikasi dengan rumus yang sederhana dan mudah untuk di aplikasikan sebagaimana disampaikan oleh Jadhav, Pandita, Pawar, & Singh (2016) serta Asikin et all (2016), sedangkan algoritma *C.45* dalam beberapa penelitian yang menggunakan *decision tree classification*, seperti penelitian Purushottam, Saxena, & Sharma (2016) serta Kumar & Umatejaswi (2017) memberikan tingkat akurasi yang tinggi.

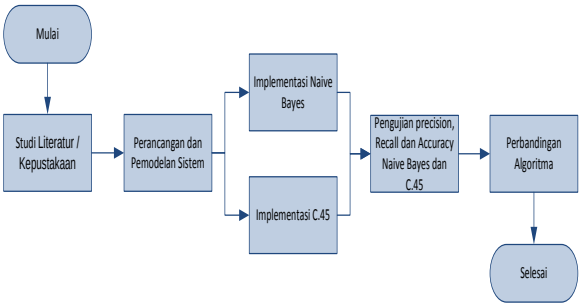
Pada penelitian ini dibangun 2 buah sistem, dengan metode *Naive Bayes* serta *C.45* ke dalam 4 buah studi kasus, yaitu :

1. Penentuan penerimaan Kartu Indonesia Sehat
2. Penentuan pengajuan kartu kredit di sebuah bank
3. Penentuan usia kelahiran
4. Penentuan kelayakan calon anggota kredit pada koperasi

Setelah itu, pada masing-masing sistem, dilakukan perhitungan *Precision*, *Recall* serta *Accuracy* untuk setiap kasus dengan perbandingan data training dan data testing yang ada. Hasil dari perhitungan *Precision*, *Recall* dan *Accuracy* diperbandingkan, sehingga dapat diambil kesimpulan mengenai algoritma terbaik untuk klasifikasi pada 4 buah kasus tersebut.

## 2. METODOLOGI PENELITIAN

Pada penelitian ini, menggunakan langkah metodologi yang terdiri dari beberapa tahap dan dapat ditunjukkan oleh gambar 1.




Gambar 1. Metodologi Penelitian

Langkah yang dilakukan di metodologi penelitian pada gambar 1 dapat dijelaskan sebagai berikut :

### 2.1 Studi Literatur atau Kepustakaan

Pada tahap ini, dilakukan pencarian studi literatur mengenai materi penelitian yang meliputi *data mining*, klasifikasi, algoritma *Naive Bayes*, algoritma *C.45* serta cara-cara pengujian menggunakan *precision*, *recall* dan *accuracy*. Pencarian didasarkan pada penelitian-penelitian terdahulu mengenai teori yang berkaitan dengan penelitian yang dilakukan serta teori-teori yang sedang berkembang saat ini.

### 2.2 Perancangan dan Pemodelan Sistem

Pada tahapan ini, dilakukan perancangan terhadap 2 buah sistem yang dibangun, yaitu sistem yang menerapkan algoritma *Naive Bayes* serta sistem yang menerapkan algoritma *C.45*. Selain itu

dibuat pula model dari 2 buah sistem tersebut, sehingga dihasilkan desain yang dapat diimplementasikan menjadi sistem yang utuh.

Sistem yang dibangun menerapkan 2 buah algoritma klasifikasi data mining pada 4 buah kasus sebagai berikut :

i. Penentuan penerimaan Kartu Indonesia Sehat (Rahman dan Kurniawan, 2018)
Data pada kasus ini merupakan data penerimaan Kartu Indonesia Sehat dari Dinas Sosial Kabupaten Sukoharjo, Jawa Tengah. Data yang diperoleh dari sumber data sebanyak 650 data. Data yang digunakan sebagai bahan pertimbangan penentuan penerima Kartu Indonesia Sehat adalah :
   a. *Variable dependent* (terikat) / *variable* Y : Penerimaan Kartu Indonesia Sehat (Diterima / Tidak Diterima)
   b. *Variable independent* (tidak terikat) / *variable* X : Usia, Pendidikan Terakhir, Pekerjaan, Pendapatan per Bulan, dan Tanggungan Anak

ii. Penentuan pengajuan kartu kredit di sebuah bank (Antaristi dan Kurniawan, 2017)
Data pada kasus ini merupakan data pengajuan kartu kredit di Bank BNI Syariah Surabaya. Data yang diperoleh dari sumber data sebanyak 290 data. Data yang digunakan sebagai bahan pertimbangan penentuan pengajuan kartu kredit adalah :
   a. *Variable dependent* (terikat) / *variable* Y : Pengajuan Kartu Kredit (Tidak Diterima, *Classic*, *Gold*, dan *Platinum*)
   b. *Variable independent* (tidak terikat) / *variable* X : Jenis Kelamin, Status Rumah, Status Menikah, Jumlah Tanggungan, Profesi, dan Penghasilan Per Bulan

iii. Penentuan usia kelahiran (Indraswari dan Kurniawan, 2018)
Data pada kasus ini merupakan data yang diperoleh dari data rekam medik pasien melahirkan di RSUD. Dr. Moewardi Provinsi Jawa Tengah dan Klinik Pratama An-Nisa Surakarta. Data yang diperoleh dari sumber data sebanyak 550 data. Data yang digunakan sebagai bahan pertimbangan penentuan usia kelahiran adalah :
   a. *Variable dependent* (terikat) / *variable* Y : Usia Kelahiran (*Premature* / Kurang dari 37 Minggu, Normal / 38-42 Minggu, *Postmature* / Lebih dari 42 Minggu)
   b. *Variable independent* (tidak terikat) / *variable* X : Usia Ibu, Tekanan Darah, Jumlah Bayi, Riwayat Persalinan, Riwayat Abortus, Malnutrisi/tidak, Penyakit Lain, dan Masalah Saat Kehamilan

iv. Penentuan kelayakan calon anggota kredit pada koperasi (Kurniawan dan Kurniawan, 2018)
Data pada kasus ini merupakan data kelayakan calon anggota kredit dari kantor pusat KSPPS BMT "Arta Jiwa Mandiri" Wonogiri. Data yang diperoleh dari sumber data sebanyak 450 data. Data yang digunakan sebagai bahan pertimbangan kelayakan calon anggota kredit adalah :
   a. *Variable dependent* (terikat) / *variable* Y : Kategori Kelayakan Kredit (Lancar, Kurang Lancar, dan Macet)
   b. *Variable independent* (tidak terikat) / *variable* X : Jenis Kelamin, Umur, Jenis Pekerjaan, Jumlah Pinjaman, Jangka Waktu Pengembalian, Jaminan, dan Penghasilan

### 2.3 Implementasi *Naive Bayes*

Langkah yang dilakukan pada tahapan ini adalah menerjemahkan desain yang telah terbentuk menjadi sebuah sistem yang menerapkan algoritma *Naive Bayes.*

Jadhav et all (2016) menyatakan bahwa *Naïve Bayes Classifier* adalah suatu model independen yang membahas mengenai klasifikasi sederhana berdasarkan teorema *Bayes*. *Naïve Bayes* merupakan suatu algoritma yang dapat mengklasifikasikan suatu variable tertentu dengan menggunakan metode probabilitas dan statistik. Secara garis besar algoritma *Naïve Bayes* dapat dijelaskan seperti persamaan (1).

$$P(R|S) = \frac{P(R)P(S|R)}{\mathrm{P(S)}} \quad (1)$$

Keterangan:

R : Data yang belum diketahui kelasnya

S : Hipotesis pada data R yang merupakan *class* khusus

*P(R|S)* : Nilai probabilitas pada hipotesis *R* yang berdasarkan kondisi *S*

*P(R)* : Nilai probabilitas pada hipotesis *R*

*P(S|R)* : Nilai probabilitas *S* yang berdasarkan dengan kondisi hipotesis *R*

*P(S)* : Nilai probabilitas *S*

Dengan menggunakan persamaan diatas, data yang telah diperoleh dapat diproses dengan algoritma *Naive Bayes* untuk penilaian data yang akan diklasifikasikan.

### 2.4 Implementasi *C.45*

Langkah yang dilakukan pada tahapan ini adalah menerjemahkan desain yang telah terbentuk menjadi sebuah sistem yang menerapkan algoritma

*C.45* pada 4 buah kasus yang telah ditentukan sebelumnya.

Purushottam, et al (2016) menyatakan algoritma C4.5 merupakan algoritma yang dipergunakan dalam membentuk *decision tree* (pengambilan keputusan). Algoritma C4.5 adalah salah satu algoritma dalam induksi *decision tree* yaitu ID3 (*Iterative Dichotomiser* 3) yang dikembangan oleh J. Ross Quinlan. Dalam prosedur algoritma ID3, input berupa sampel *training*, label *training* dan atribut. Algoritma C4.5 ini merupakan pengembangan dari ID3. Ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai *gain* tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai poros atribut klasifikasi. Kemudian secara rekursif cabang-cabang pohon diperluas sehingga seluruh pohon terbentuk. Terdapat empat langkah dalam proses pembuatan pohon keputusan pada algoritma C4.5, yaitu:

a. Memilih atribut sebagai akar.
b. Membuat cabang untuk masing-masing nilai.
c. Membagi setiap kasus dalam cabang.
d. Mengulangi proses dalam setiap cabang sehingga semua kasus dalam cabang memiliki kelas yang sama.

Kemudian dilakukan perhitungan untuk mencari nilai *entropy* dan *gain*. Berikut ini rumus untuk mencari nilai *entropy* dan *gain*.

$$Entropy(S) = \sum_{j=1}^{k} - pj \log_2 pj \quad (2)$$

Persamaan (2) adalah persamaan yang digunakan dalam perhitungan *entropy* untuk menentukan *heterogenity* dari sebuah kumpulan *data sample* (Amin et al, 2015). Berikut keterangannya :

$S$ : Himpunan kasus
$k$ : Jumlah partisi $S$
$pj$ : Jumlah kasus pada partisi ke-$j$

$$Gain(A) = Entropy(S) - \sum_{i=1}^{k} \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (3)$$

Rumus (3) merupakan rumus yang digunakan dalam perhitungan *gain* setelah melakukan perhitungan *entropy*. Berikut keterangannya :

$A$ : Atribut dari dataset
$k$ : Jumlah partisi $S$
$S$ : Himpunan kasus

Dengan mengetahui rumus–rumus diatas, data yang telah diperoleh dapat dimasukkan dan diproses dengan algoritma C4.5 untuk proses pembuatan *decision tree*.

### 2.5 Pengujian *Precision, Recall* dan *Accuracy Naive Bayes* dan *C.45*

Kurniawan, et al (2018) menyatakan pengujian sebuah algoritma membutuhkan standar dan alat uji. Membandingkan 2 buah algoritma harus memiliki standar yang sama sehingga dapat diketahui algoritma yang terbaik dari perbandingan tersebut.

Pada tahap ini, dilakukan pengujian dengan menghitung nilai *precision*, *recall* serta *accuracy* dari *Naive Bayes* dan *C.45*. Langkah awal dalam tahap ini adalah dengan membagi data di setiap kasus menjadi 2, yaitu data training atau data latih dan data testing atau data uji. Data training digunakan sebagai data rujukan dalam perhitungan setiap algoritma, sedangkan data testing digunakan untuk menilai prediksi maupun penentuan yang dilakukan oleh setiap algoritma sudah tepat atau tidak. Dalam pembagian data menjadi data training dan data testing, dilakukan dengan beberapa perbandingan. Sebagai contoh, pada data untuk penentuan pengajuan kartu kredit di bank yang berjumlah 290 data, dibagi menjadi beberapa bagian seperti yang diperlihatkan oleh tabel 1.

Tabel 1. Contoh Pembagian Data Training dan Data Testing

| Pengujian | Data Training | Data Testing |
|---|---|---|
| Pengujian 1 | 50 | 240 |
| Pengujian 2 | 100 | 190 |
| Pengujian 3 | 150 | 140 |
| Pengujian 4 | 200 | 90 |
| Pengujian 5 | 250 | 40 |

Dalam setiap *dataset* di atas, dilakukan pengujian untuk data testing sebanyak 3 kali di setiap pengujian, setelah itu dihitung nilai *precision*, *recall* dan *accuracy* untuk masing-masing algoritma. Untuk setiap nilai *precision*, *recall* dan *accuracy* dari 3 kali percobaan tersebut akan dihitung nilai rata-rata dan ditulis sebagai nilai akhir *precision*, *recall* dan *accuracy*.

*Precision* merupakan perhitungan terhadap perkiraan proporsi kasus positif yang benar dan dirumuskan dalam persamaan 4: (Vafeiadis, Diamantaras, Sarigiannidis, & Chatzisavvas, 2015)

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

*Recall* merupakan perhitungan terhadap perkiraan proporsi kasus positif yang diidentifikasi benar dan dirumuskan dalam persamaan 5:

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

*Accuracy* merupakan perhitungan terhadap proporsi dari jumlah total prediksi yang benar dan dirumuskan dalam persamaan 6:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (6)$$

Keterangan :
TP : *True Positive*
TN : *True Negative*

FP : *False Positive*
FN : *False Negative*

### 2.6 Perbandingan Algoritma

Pada tahapan ini dilakukan perbandingan nilai *precision*, *recall* dan *accuracy* pada masing-masing algoritma di setiap kasus. Setelah itu, dilakukan rekapitulasi hasil dari masing-masing algoritma sehingga dapat diambil kesimpulan mengenai algoritma terbaik untuk setiap kasus.

## 3. HASIL DAN PEMBAHASAN

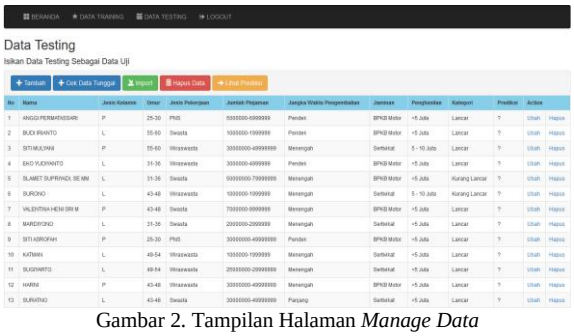
Hasil dari penelitian ini adalah sebuah aplikasi beserta dengan pengujian yang dapat memperlihatkan nilai *precision*, *recall* dan *accuracy* dari algoritma *Naive Bayes* dan *C.45* dalam 4 kasus yang ada.

Berikut ini merupakan hasil dan pembahasan dari penelitian yang telah dilakukan :

### 3.1 Implementasi Aplikasi

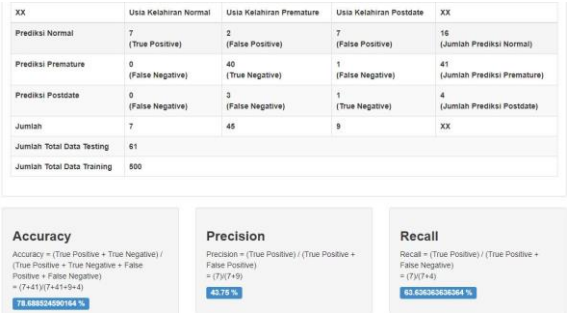
Implementasi dari aplikasi pada perancangan sebelumnya, dibangun dengan bahasa pemrograman *PHP* dan *database MySQL*. Aplikasi ini memiliki 1 hak akses, yaitu Administrator yang dimulai dari *login* terlebih dahulu, dengan fitur sebagai berikut :

a. *Manage Data*, fitur untuk mengelola data training maupun data testing, yang berisi *insert* (tambah data), *update* (ubah data), *delete* (hapus data), *search* (pencarian data), serta *import* (tambah banyak data sekaligus dengan menggunakan file *excel*). Contoh tampilan untuk *manage data* dapat ditunjukkan oleh gambar 2.




Gambar 2. Tampilan Halaman *Manage Data*

b. Pengujian, fitur untuk melakukan pengujian terhadap data testing yang ada dengan menggunakan algoritma *Naive Bayes* dan *C.45* sekaligus melakukan perhitungan nilai *precision*, *recall* dan *accuracy*. Contoh tampilan untuk prediksi di salah satu studi kasus, yaitu pada kasus prediksi usia kelahiran dapat ditunjukkan oleh gambar 3.




Gambar 3. Tampilan Halaman Pengujian

### 3.2 Pengujian *Blackbox*

Pengujian *Blackbox* merupakan pengujian terhadap fitur-fitur serta fungsionalitas yang terdapat pada aplikasi dengan memberikan sejumlah *test case* terhadap aplikasi tersebut. Pada pengujian *blackbox* diberikan sejumlah input tertentu untuk menghasilkan *output* yang diinginkan. Pengujian dianggap berhasil jika *output* dari aplikasi sesuai dengan *output* yang diharapkan. Daftar *test case* yang diberikan ke aplikasi untuk pengujian *blackbox* dapat ditunjukkan oleh tabel 2.

Tabel 2. *Test Case* Pengujian *Blackbox*

| Modul | Test Case / Scenario / Input | Hasil / Output yang diharapkan | Hasil di Aplikasi |
|---|---|---|---|
| Login. | Input username dan password benar. | Masuk ke halaman admin | Valid |
| Manage Data. | Memasukkan data dan semua data telah terisi. | Data masuk kedalam database | Valid |
| Manage Data. | Mengubah data dengan beberapa inputan | Data di dalam database telah ter-update | Valid |
| Manage Data | Menghapus 1 buah data | Data di dalam database terhapus | Valid |
| Manage Data | Memasukkan beberapa buah data sekaligus menggunakan file excel | Data yang ada pada file excel masuk ke dalam database | Valid |
| Manage Data | Melakukan pencarian terhadap data-data tertentu | Data yang dicari dapat muncul dan diperlihatkan | Valid |
| Pengujian | Menekan tombol pengujian | Aplikasi menghitung hasil / prediksi dari algoritma Naive Bayes | Valid |
| Pengujian | Menekan tombol pengujian | Aplikasi menghitung hasil / prediksi dari algoritma C.45 | Valid |
| Pengujian | Menekan tombol pengujian | Aplikasi menghitung nilai precision, recall dan accuracy dari algoritma Naive Bayes dan C.45 | Valid |
| Logout. | Keluar dari aplikasi. | Keluar dari halaman admin. | Valid |

Tabel 3. Hasil Pengujian *Precision, Recall* dan *Accuracy* pada Penentuan Penerimaan Kartu Indonesia Sehat

| Jumlah Data Training | Jumlah Data Testing | *Precision Naive Bayes* (%) | *Recall Naive Bayes* (%) | *Accuracy Naive Bayes* (%) | *Precision C.45* (%) | *Recall C.45*(%) | *Accuracy C.45*(%) |
|---|---|---|---|---|---|---|---|
| 50 | 600 | 92.9 | 92.7 | 92.67 | 86.1 | 86 | 86 |
| 100 | 550 | 95 | 94.9 | 94.91 | 97.8 | 97.8 | 97.81 |
| 150 | 500 | 94.9 | 94.8 | 94.8 | 98 | 98 | 98 |
| 200 | 450 | 94.6 | 94.4 | 94.44 | 98 | 98 | 98 |
| 250 | 400 | 95 | 94.8 | 94.75 | 98.8 | 98.3 | 98.33 |
| 300 | 350 | 98.6 | 98.6 | 98.57 | 98.6 | 98.4 | 98.5 |
| 350 | 300 | 98.3 | 98.3 | 98.33 | 98.3 | 98.6 | 98.57 |
| 400 | 250 | 98.4 | 98.4 | 98.4 | 98.4 | 98.8 | 98.75 |
| 450 | 200 | 98 | 98 | 98 | 98 | 98 | 98 |
| 500 | 150 | 98 | 98 | 98 | 98 | 98 | 98 |
| 550 | 100 | 99 | 99 | 99 | 99 | 99 | 99 |
| 600 | 50 | 100 | 100 | 100 | 100 | 100 | 100 |

Tabel 4. Hasil Pengujian *Precision, Recall* dan *Accuracy* pada Penentuan Pengajuan Kartu Kredit di Sebuah Bank

| Jumlah Data Training | Jumlah Data Testing | *Precision Naive Bayes* (%) | *Recall Naive Bayes* (%) | *Accuracy Naive Bayes* (%) | *Precision C.45* (%) | *Recall C.45*(%) | *Accuracy C.45*(%) |
|---|---|---|---|---|---|---|---|
| 50 | 240 | 81.2 | 81.8 | 81.81 | 92.8 | 92 | 92.04 |
| 100 | 190 | 87 | 87.4 | 87.36 | 93.5 | 93.2 | 93.16 |
| 150 | 140 | 88.3 | 88.6 | 88.57 | 93.8 | 93.9 | 93.85 |
| 200 | 90 | 92.1 | 92.2 | 92.33 | 97.9 | 97.8 | 97.78 |
| 250 | 40 | 98.2 | 98.4 | 98.56 | 99.4 | 99.5 | 99.54 |

Tabel 5. Hasil Pengujian *Precision, Recall* dan *Accuracy* pada Penentuan Kelahiran

| Jumlah Data Training | Jumlah Data Testing | *Precision Naive Bayes* (%) | *Recall Naive Bayes* (%) | *Accuracy Naive Bayes* (%) | *Precision C.45* (%) | *Recall C.45*(%) | *Accuracy C.45*(%) |
|---|---|---|---|---|---|---|---|
| 50 | 500 | 51.9 | 58 | 58 | 45.9 | 52 | 52 |
| 100 | 450 | 58.5 | 63.6 | 63.55 | 56 | 62.4 | 62.44 |
| 150 | 400 | 57.9 | 63.3 | 63.25 | 57.7 | 62.5 | 62.5 |
| 200 | 350 | 63.3 | 67.1 | 67.14 | 59.6 | 64 | 64 |
| 250 | 300 | 66.2 | 67.7 | 67.67 | 59.1 | 63.7 | 63.67 |
| 300 | 250 | 68 | 68.4 | 68.4 | 57.5 | 64 | 64 |
| 350 | 200 | 67.9 | 68 | 68 | 62.3 | 66.5 | 66.5 |
| 400 | 150 | 62.9 | 64 | 64 | 57.8 | 62.7 | 62.67 |
| 450 | 100 | 66.9 | 68 | 68 | 59.1 | 62 | 62 |
| 500 | 50 | 79.5 | 80.1 | 80.13 | 77.4 | 76 | 76 |

Tabel 6. Hasil Pengujian *Precision, Recall* dan *Accuracy* pada Penentuan Kelayakan Calon Anggota Kredit pada Koperasi

| Jumlah Data Training | Jumlah Data Testing | *Precision Naive Bayes* (%) | *Recall Naive Bayes* (%) | *Accuracy Naive Bayes* (%) | *Precision C.45* (%) | *Recall C.45*(%) | *Accuracy C.45*(%) |
|---|---|---|---|---|---|---|---|
| 50 | 400 | 54.2 | 46.3 | 46.25 | 50.5 | 42.8 | 42.75 |
| 100 | 350 | 53.5 | 57.1 | 57.14 | 45.9 | 67.7 | 67.71 |
| 150 | 300 | 58.3 | 55.7 | 55.67 | 51.4 | 71.7 | 71.67 |
| 200 | 250 | 64.1 | 62.2 | 62 | 54.8 | 74 | 74 |
| 250 | 200 | 60.3 | 62.5 | 62.2 | 52.6 | 74.4 | 74.54 |
| 300 | 150 | 66.3 | 64.5 | 65.4 | 58.8 | 76.3 | 76.52 |
| 350 | 100 | 69.4 | 67.2 | 67.45 | 61.4 | 78.2 | 79.31 |
| 400 | 50 | 75.7 | 68.1 | 69.54 | 66.1 | 79.5 | 80.23 |

### 3.3 Pengujian Algoritma

Pada pengujian ini, dilakukan dengan memasukkan 10 data training dan 5 data testing pada aplikasi. Setelah aplikasi menghitung nilai hasil dan prediksi dari algoritma *Naive Bayes* dan *C.45*, lalu dilakukan perhitungan secara manual untuk menentukan hasil dan prediksi dari algoritma *Naive Bayes* dan *C.45* pada ke-5 data testing. Dari hasil perhitungan pada data testing dengan menggunakan aplikasi serta perhitungan manual, diperoleh hasil prediksi yang sama. Hal ini menunjukkan bahwa aplikasi telah mengimplementasikan algoritma *Naive Bayes* dan *C.45* dengan benar dan valid.

### 3.4 Pengujian *Precision, Recall* dan *Accuracy*

Pada pengujian ini, dilakukan percobaan untuk perhitungan *precision, recall* dan *accuracy* terhadap algoritma *Naive Bayes* dan *C.45* pada 4 buah studi kasus yang ada dengan beberapa perbandingan jumlah data training dan data testing. Setelah itu, dihitung nilai rata-rata dari *precision, recall* dan *accuracy* dari masing-masing algoritma. Pada setiap percobaan tersebut, dilakukan sebanyak 3 kali dengan data yang acak. Nilai *precision, recall* dan *accuracy* yang dituliskan adalah hasil dari rata-rata 3 kali percobaan tersebut.

#### 3.4.1 Pengujian *Precision, Recall* Dan *Accuracy* Pada Kasus Penentuan Penerimaan Kartu Indonesia Sehat

Hasil pengujian *precision, recall* dan *accuracy* pada kasus penentuan penerimaan Kartu Indonesia Sehat dapat ditunjukkan oleh tabel 3.

Dari data pada tabel 3 tersebut dapat dilihat bahwa semakin banyak jumlah data training membuat nilai *precision, recall* dan *accuracy* dari algoritma *Naive Bayes* dan *C.45* cenderung semakin meningkat. Pada beberapa bagian data, nilai *precision, recall* dan *accuracy* dari 2 buah algoritma tersebut menunjukkan nilai yang sama. Pada satu titik (data training sejumlah 600 dan data testing sejumlah 50), dua buah algoritma tersebut dapat mencapai tingkat *precision, recall* dan *accuracy* yang mencapai 100 persen. Hal ini menunjukkan 2 buah algoritma tersebut sama-sama efektif untuk kasus penentuan penerimaan Kartu Indonesia Sehat.

#### 3.4.2 Pengujian *Precision, Recall* Dan *Accuracy* Pada Kasus Penentuan Pengajuan Kartu Kredit Di Sebuah Bank

Hasil pengujian *precision, recall* dan *accuracy* pada kasus penentuan pengajuan kartu kredit di sebuah bank dapat ditunjukkan oleh tabel 4.

Dari data pada tabel 4 tersebut dapat dilihat bahwa semakin banyak jumlah data training membuat nilai *precision, recall* dan *accuracy* dari algoritma *Naive Bayes* dan *C.45* semakin meningkat. Dari data tersebut, dapat pula dilihat bahwa nilai *precision, recall* dan *accuracy* dari algoritma *C.45* lebih tinggi dari *Naive Bayes* di setiap pembagian data. Hal ini menunjukkan bahwa algoritma *C.45* lebih baik daripada algoritma *Naive Bayes* untuk kasus pengajuan kartu kredit di sebuah bank.

#### 3.4.3 Pengujian *Precision, Recall* Dan *Accuracy* Pada Kasus Penentuan Usia Kelahiran

Hasil pengujian *precision, recall* dan *accuracy* pada kasus penentuan usia kelahiran dapat ditunjukkan oleh tabel 5.

Dari data pada tabel 5 tersebut dapat dilihat bahwa semakin banyak jumlah data training membuat nilai *precision, recall* dan *accuracy* dari algoritma *Naive Bayes* dan *C.45* semakin meningkat. Dari data tersebut, dapat pula dilihat bahwa nilai *precision, recall* dan *accuracy* dari algoritma *Naive Bayes* selalu lebih tinggi dari algoritma *C.45* di setiap pembagian data. Hal ini menunjukkan bahwa algoritma *Naive Bayes* lebih baik daripada algoritma *C.45* untuk kasus penentuan usia kelahiran.

#### 3.4.4 Pengujian *Precision, Recall* Dan *Accuracy* Pada Kasus Penentuan Kelayakan Calon Anggota Kredit Pada Koperasi

Hasil pengujian *precision, recall* dan *accuracy* pada kasus penentuan kelayakan calon anggota kredit pada koperasi dapat ditunjukkan oleh tabel 6.

Dari data pada tabel 6 tersebut dapat dilihat bahwa semakin banyak jumlah data training membuat nilai *precision, recall* dan *accuracy* dari algoritma *Naive Bayes* dan *C.45* semakin meningkat. Dari data tersebut, dapat pula dilihat bahwa nilai *precision*, dari algoritma *Naive Bayes* selalu lebih tinggi dari *precision* algoritma *C.45* di setiap pembagian data. Sedangkan untuk *recall* dan *accuracy* menunjukkan hal yang berbeda, yaitu algoritma *C.45* selalu lebih tinggi dari algoritma *Naive Bayes.* Jika mengacu pada hasil akhir, yaitu nilai kebenaran dari klasifikasi (banyaknya *true/false*), maka parameter *accucary* yang dapat digunakan sebagai acuan. Oleh sebab itu, Hal ini menunjukkan bahwa algoritma *C.45* lebih baik daripada algoritma *Naive Bayes* untuk kasus penentuan kelayakan calon anggota kredit di koperasi.

### 3.5 Analisa Hasil

Berdasarkan hasil implementasi, telah dibangun sebuah aplikasi yang dapat melakukan *manage* data training dan data testing untuk ke-4 buah studi kasus, yaitu studi kasus penentuan penerimaan Kartu Indonesia Sehat, kasus penentuan pengajuan kartu kredit di sebuah bank, kasus penentuan usia kelahiran, serta kasus penentuan kelayakan calon anggota kredit pada koperasi. Selain itu, aplikasi juga dapat menerapkan algoritma *Naive Bayes* serta algoritma *C.45* ke dalam 4 buah studi kasus tersebut. Pada bagian akhir, aplikasi dapat memperlihatkan nilai dari *precision, recall* serta *accuracy* untuk setiap data yang telah dimasukkan dengan menggunakan 2 buah algoritma tersebut.

Berdasarkan hasil pengujian *blackbox*, maka aplikasi yang dibangun telah dapat menjalankan setiap fitur dengan baik. Hal ini dibuktikan dengan setiap *test case* yang diberikan memberikan hasil yang valid sesuai dengan yang diinginkan. Pada pengujian algoritma, dapat dibuktikan bahwa aplikasi tersebut dapat menerapkan dua buah algoritma dengan valid. Hal ini dibuktikan dengan pengujian data testing pada aplikasi dan penilaian secara manual menunjukkan hasil yang sama, sehingga diambil kesimpulan bahwa aplikasi telah menerapkan kedua buah algoritma tersebut secara valid dan benar.

Berdasarkan pengujian *precision*, *recall* dan *accuracy*, dapat diambil kesimpulan bahwa dengan semakin banyaknya data training yang digunakan, maka nilai *precision*, *recall* dan *accuracy* akan semakin meningkat. Hal ini dikarenakan probabilitas terhadap setiap data akan semakin besar dengan semakin meningkatnya jumlah data. Semakin besar nilai probabilitasnya, maka semakin tinggi pula nilai *precision*, *recall* dan *accuracy* dari kedua buah algoritma tersebut.

Rekapitulasi hasil pengujian berdasarkan pengujian precision, recall dan accuracy dapat ditunjukkan oleh tabel 7. Pada tabel 7 menunjukkan algoritma yang unggul pada setiap pengujian dan kasus.

Tabel 7. Rekapitulasi Algoritma Yang Unggul Di Setiap Kasus

| No | Kasus | Precision | Recall | Accuracy |
|---|---|---|---|---|
| 1 | Penerimaan "Kartu Indonesia Sehat" | Sama | Sama | Sama |
| 2 | Penentuan pengajuan kartu kredit di sebuah bank | C.45 | C.45 | C.5 |
| 3 | Penentuan usia kelahiran | Naive Bayes | Naive Bayes | Naive Bayes |
| 4 | Penentuan kelayakan calon anggota kredit pada koperasi | Naive Bayes | C.45 | C.45 |

Pada kasus penentuan penerimaan Kartu Indonesia Sehat dapat dilihat bahwa kedua buah algoritma tersebut sama-sama efektif untuk digunakan. Pada kasus penentuan pengajuan kartu kredit di sebuah bank, algoritma *C.45* memberikan hasil yang lebih baik daripada algoritma *Naive Bayes*. Sebaliknya, pada kasus penentuan usia kelahiran, algoritma *Naive Bayes* memberikan klasifikasi yang lebih baik daripada algoritma *C.45*. Pada kasus terakhir, yaitu kasus penentuan kelayakan calon anggota kredit pada koperasi, algoritma *Naive Bayes* memberikan nilai yang lebih baik pada *precision*, tapi untuk *recall dan accuracy*, algoritma *C.45* memberikan hasil yang lebih baik. Dari ke-4 buah kasus tersebut, dapat diambil kesimpulan bahwa hasil klasifikasi pada algoritma *Naive Bayes* dan *C.45* tidak dapat memberikan nilai yang absolut atau mutlak di setiap kasus. Sehingga untuk menentukan algoritma terbaik yang akan dipakai di sebuah kasus, harus melihat kriteria, *variable* maupun jumlah data di kasus tersebut.

## 4. Kesimpulan

Dari penjelasan dan pengujian yang telah dilakukan, dapat diambil beberapa kesimpulan sebagai berikut :

1. Telah dibangun sebuah aplikasi klasifikasi data mining menggunakan algoritma *Naive Bayes* dan *C.45* beserta dengan pengujian *precision*, *recall* dan *accuracy* untuk ke-4 buah studi kasus, yaitu studi kasus penentuan penerimaan Kartu Indonesia Sehat, kasus penentuan pengajuan kartu kredit di sebuah bank, kasus penentuan usia kelahiran, serta kasus penentuan kelayakan calon anggota kredit pada koperasi.
2. Berdasarkan hasil pengujian *blackbox*, maka aplikasi yang dibangun telah dapat menjalankan setiap fitur dengan baik.
3. Pada pengujian algoritma, dapat dibuktikan bahwa aplikasi tersebut dapat menerapkan dua buah algoritma dengan valid.
4. Semakin banyaknya data training yang digunakan, maka nilai *precision*, *recall* dan *accuracy* akan semakin meningkat.
5. Pada kasus penentuan penerimaan Kartu Indonesia Sehat, kedua buah algoritma tersebut sama-sama efektif untuk digunakan.
6. Pada kasus penentuan pengajuan kartu kredit di sebuah bank, algoritma *C.45* memberikan hasil yang lebih baik daripada algoritma *Naive Bayes*.
7. Pada kasus penentuan usia kelahiran, algoritma *Naive Bayes* memberikan klasifikasi yang lebih baik daripada algoritma *C.45*.
8. Pada kasus penentuan kelayakan calon anggota kredit pada koperasi, algoritma *Naive Bayes* memberikan nilai yang lebih baik pada *precision*, tapi untuk *recall* dan *accuracy*, algoritma *C.45* memberikan hasil yang lebih baik.
9. Hasil klasifikasi pada algoritma *Naive Bayes* dan *C.45* tidak dapat memberikan nilai yang absolut atau mutlak di setiap kasus. Sehingga untuk menentukan algoritma terbaik yang akan dipakai di sebuah kasus, harus melihat kriteria, variable maupun jumlah data di kasus tersebut.

## DAFTAR PUSTAKA

AMIN, R. K., INDWIARTI, & SIBARONI, Y. (2015). Implementation of Decision Tree Using C4.5 Algorithm in Decision Making of Loan Application by Debtor (Case Study: Bank Pasar of Yogyakarta Special Region). Information and Communication Technology (ICoICT), Nusa Dua, Bali, 27-29 May. 75-80. DOI: 10.1109/ICoICT.2015.7231400

ANTARISTI, M., & KURNIAWAN, Y. I. (2017). Aplikasi Klasifikasi Penentuan Pengajuan Kartu Kredit Menggunakan Metode Naive Bayes di Bank BNI Syariah Surabaya. Jurnal Teknik Elektro, 9(2), 45-52.

ASIKIN, M.F., KURNIAWATY, D., SARI, S.K. AND CHOLISSODIN, I., 2016. Implementasi Metode Naïve Bayes Classifier Untuk Seleksi Asisten Praktikum Pada Simulasi Hadoop

Multinode Cluster. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *3*(4), pp.273-278.

BANSAL, A., SHARMA, M., & GOEL, S. (2017). Improved k-mean clustering algorithm for prediction analysis using classification technique in data mining. International Journal of Computer Applications (0975–8887) Volume, 157, 33-40.

INDRASWARI, N. R., & KURNIAWAN, Y. I. (2018). Aplikasi Prediksi Usia Kelahiran dengan Metode Naive Bayes. Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer, 9(1), 129-138.

JADHAV, A., PANDITA, A., PAWAR, A., & SINGH, V. (2016). Classification of Unstructured Data using Naïve Bayes Classifier and Predictive Analysis for RTI Application. ABHIYANTRIKI: An International Journal of Engineering & Technology, 3(6), 1–6.

KUMAR, P. S., & UMATEJASWI, V. (2017). Diagnosing Diabetes using Data Mining Techniques. International Journal of Scientific and Research Publications, 7(6), 705-709.

KURNIAWAN, D. A., & KURNIAWAN, Y. I. (2018). Aplikasi Prediksi Kelayakan Calon Anggota Kredit Menggunakan Algoritma Naïve Bayes. Jurnal Teknologi dan Manajemen Informatika, 4(1).

KURNIAWAN, Y. I., SOVIANA, E., & YULIANA, I. (2018, June). Merging Pearson Correlation and TAN-ELR algorithm in recommender system. In AIP Conference Proceedings (Vol. 1977, No. 1, p. 040028). AIP Publishing.

PURUSHOTTAM, SAXENA, K., & SHARMA, R. (2016). Efficient Heart Disease Prediction System using Decision Tree. International Conference on Computing, Communication and Automation (ICCCA), Noida, India, 15-16 May. 72-77. DOI: 10.1109/CCAA.2015.7148346

RAHMAN, A. A., & KURNIAWAN, Y. I. (2018). APLIKASI KLASIFIKASI PENERIMA KARTU INDONESIA SEHAT MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER. Jurnal Teknologi dan Manajemen Informatika, 4(1).

ROIGER, R. J. (2017). Data mining: a tutorial-based primer. CRC Press.

VAFEIADIS, T., DIAMANTARAS, K., SARIGIANNIDIS, G., & CHATZISAVVAS, K. (2015). A Compariosn Of Machine Learning Techniques For Customer Chrun Prediction. Simulation Modelling Practice and Theory, 55, 1-9.

*Halaman ini sengaja dikosongkan*