

KOMPARASI KINERJA ALGORITMA C.45 DAN NAIVE BAYES UNTUK PREDIKSI KEGIATAN PENERIMAANMAHASISWA BARU (STUDI KASUS : UNIVERSITAS STIKUBANK SEMARANG)

Norzam Yahya¹, Arief Jananto²

^{1,2}Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Stikubank
E-mail: ¹noerzam.yahya11@gmail.com, ²ajananto09@edu.unisbank.ac.id

ABSTRAK

Universitas Stikubank merupakan salah satu Perguruan Tinggi Swasta di Semarang yang telah berdiri cukup lama dengan jumlah mahasiswa yang diterima di tiap tahun cukup banyak. Setiap tahun ajaran baru, kegiatan penerimaan mahasiswa baru dikampus ini terus menghasilkan data mahasiswa yang semakin banyak.

Namun masih terdapat sebuah masalah yaitu jumlah mahasiswa baru yang mendaftar belum dapat menggambarkan mahasiswa yang akan melanjutkan ke proses registrasi ulang. Hal tersebut disebabkan paradig masyarakat yang masih memprioritaskan kuliah di Perguruan Tinggi Negeri dibandingkan Perguruan Tinggi Swasta. Maka diperlukan suatu upaya bagi Unisbanku ntuk mengidentifikasi pola data PMB yang ada menjadi sebuah informasi yang bernilai penting khususnya dalam memprediksi jumlah mahasiswa baru yang akan mendaftar.

Dalam penelitian ini penulis menerapkan salah satu metode dalam teknik data mining yaitu klasifikasi. Kemudian membandingkan kinerja antara algoritma C45 dan Naive Bayes untuk mencari algoritma dengan tingkat akurasi terbaik dalam memprediksi masuknya calon mahasiswa baru di Unisbank. Selanjutnya didapatkan hasil tertinggi pada pengujian dengan data training sebanyak 2133 dan data testing sebanyak 533 yaitu algoritma C45 memiliki nilai akurasi lebih tinggi sebesar 88.74% dibandingkan algoritma Naive Bayes yang memiliki akurasi sebesar 87.24%.

Kata Kunci: Penerimaan Mahasiswa Baru, Data Mining, Model Klasifikasi, Naive Bayes, Decision Tree C45.

1. PENDAHULUAN

Universitas Stikubank merupakan salah satu PerguruanTinggi Swasta di Semarang yang telah berdiri cukup lama dengan jumlah mahasiswa yang diterimadi tiap tahun cukup banyak Setia p pembukaan tahun ajaran baru, perguruan tinggi ini rutin menyelenggarakan kegiatan penerimaan mahasiswa baru (PMB). Kegiatan PMB yang sudah rutin dilaksanakan secara tidak langsung akan menghasilkan banyak data mahasiswa baru yang terkumpul. Suatu hal yang perlu dilakukan oleh perguruan tinggi adalah mengelola basis data tersebut sehingga nantinya dapat menjadi informasi yang bernilai penting. Informasi yang dapa tdihasilkan salahsatunya mengenai target jumlah penerimaan mahasiswa baru.

Masalah yang masih menjadi kendala adalah banyaknya jumlah mahasiswa baru yang mendaftar belum dapat menggambarkan jumlah mahasiswa yang akan lanjutr egistrasi ulang. Hal tersebut biasa terjadi karena perguruan tinggi swasta bukan menjadi pilihan pertama sehingga apabila calon mahasiswa diterima diperguruan tinggi negeri, maka biasanya perguruan tinggi swasta tersebut ditinggalkan tanpa pemberitahuan meskipun sudah dinyatakan diterima.

Dari permasalahan di atas, kiranya diperlukan sebuah upaya dalam mengolah data-data PMB yang sudah ada kemudian mengidentifikasi pola dari data tersebut tmenjadi sebuah informasi yang dapat memprediksi kemungkinan jumlah calon mahasiswa baru yang akan melanjutkan registrasi ulang. Proses identifikasi pola data tersebut dapat dilakukan dengan menerapkan metode yang terdapat di dalam ilmu *data mining* salah satunya adalah klasifikasi.

Data Mining adalah rangkaian dari beberapa proses untuk mencari sebuah nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [1]. Sedangkan klasifikasi merupakan salah satu tugas utama dalam *data mining* yang memiliki tujuan untuk memprediksi sebuah kategori dari masukkan data[2].

Berdasarkan uraian latar belakang masalah di atas maka pada penelitian ini akan diterapkan salah satu metode yang terdapat pada ilmu *data mining* yaitu klasifikasi dengan membandingkan kinerja antara algoritma C45 dan*Naive Bayes*. Diharapkan dengan dilakukannya model klasifikasi ini akan didapatkan algoritma dengan tingkat akurasi tertinggi yang bisa digunakan pihak perguruan tinggi dalam memprediksi jumlah mahasiswa yang registrasiulang dan menentukan kebijakan bagi PMB yang akan datang.

2. TINJAUAN PUSTAKA

Penelitian pertama yang dilakukan Gian Fiastantyo (2014) [3] dengan judul “Perbandingan Kinerja Metode Klasifikasi Data Mining Menggunakan *Naive Bayes* dan Algoritma C4.5 Untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa”, ia mencoba melakukan penelitian dengan membangun sebuah sistem prediksi ketepatan kelulusan mahasiswa dengan menggunakan algoritma C4.5 dan *Naive Bayes*. Selanjutnya diketahui

bahwa C4.5 memiliki nilai akurasi yang lebih baik yakni sebesar 77,34% sedangkan *naive bayes* memiliki nilai akurasi mencapai 74,094%. Kemudian berdasarkan hasil perbandingan kinerja dua algoritma tersebut, algoritma dengan pencapaian nilai akurasi terbaik akan diimplementasikan dalam bentuk sebuah *Decision Support System*. Penelitian kedua oleh Derick Iskandar dan Yoyon Suprpto (2015) [4] dengan judul “Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan Antar Algoritma C4.5 dan *Naive Bayes*”, ia mencoba memanfaatkan basis data terpadu yang dikeluarkan oleh TNP2K untuk diolah menggunakan metode *data mining* klasifikasi dan membandingkan kinerja algoritma C4.5 serta *Naive Bayes*. Hasil penelitian diperoleh bahwa algoritma *decision tree* C4.5 memiliki tingkat akurasi lebih baik 3% dibandingkan dengan algoritma *naive bayes*. Penelitian ketiga oleh Wiwit Supriyanti, dkk (2016) [5] dengan judul “Perbandingan Kinerja Algoritma C.45 Dan *Naive Bayes* Untuk ketepatan Pemilihan Konsentrasi Mahasiswa”, ia mencoba mengimplementasikan algoritma C4.5 dan *Naive Bayes* pada 539 data mahasiswa alumni Program Studi S1 Informatika Fakultas Komunikasi dan Informatika Universitas Muhammadiyah Surakarta. Kedua algoritma tersebut diuji coba menggunakan metode CRISP-DM dan *forward selection*, kemudian didapatkan hasil bahwa penggunaan metode *forward selection* meningkatkan akurasi algoritma C.45 mencapai 84,98% yang sebelumnya hanya sebesar 84,43% dan algoritma *Naive Bayes* mencapai 82,01% yang sebelumnya hanya sebesar 78,47%.

3. METODE PENELITIAN

3.1. Metode Analisis Data

Berikut ini adalah penjelasan mengenai enam tahap siklus hidup pengembangan *data mining* berdasarkan CRISP-DM (*CRoss-Industry Standard Process for Data Mining*):

1. Fase Pemahaman Tujuan (*Bussiness Understanding*)

Penelitian ini bertujuan untuk mengidentifikasi pola-pola dataset penerimaan mahasiswa baru terdahulu dengan membandingkan kinerja antara dua algoritma klasifikasi yaitu C4.5 dan *Naive Bayes* yang dapat ditujukan sebagai referensi atau acuan algoritma klasifikasi mana yang paling baik dan akurat untuk memprediksi masuknya calon mahasiswa baru dan menentukan kebijakan bagi PMB Unisbank yang akan datang.

2. Fase Pemahaman Data (*Data Understanding*)

Sumber data yang digunakan pada penelitian ini adalah data penerimaan mahasiswa baru Unisbank dari tahun 2015 hingga 2016 yang peneliti dapat dari Tim PMB Unisbank.

Berikut ini keterangan tiap kolom atribut pada data set penerimaan mahasiswa baru Unisbank.

Tabel 1. Penjelasan Atribut Dataset PMB

No	Atribut	Keterangan
1	Tgl Register	Tanggal mahasiswa registrasi ulang
2	Tgl Daftar	Tanggal dimana calon mahasiswa mendaftar di Unisbank
3	Periode	Tahun mahasiswa mendaftar kuliah
4	Prodi	Program studi yang dipilih mahasiswa
5	NIM	Nomor induk mahasiswa yang didapatkan ketika sudah registrasi ulang
6	Nama	Nama lengkap calon mahasiswa
7	Alamat	Alamat lengkap mahasiswa
8	Kota Asal	Kota domisili mahasiswa
9	Telepon	Nomor telepon mahasiswa
10	Handphone	Nomor ponsel mahasiswa
11	E-mail	Alamat e-mail mahasiswa
12	Jenis Kelamin	Jenis kelamin mahasiswa
13	Agama	Agama mahasiswa
14	Asal Sekolah	Nama sekolah asal mahasiswa

15	Asal PT	Nama perguruan tinggi asal mahasiswa
16	Kelas Kuliah	Kelas kuliah yang dipilih mahasiswa

3. Fase Pengolahan Data (*Data Preparation*)

Pada tahap ini meliputi proses pengolahan data yaitu data Penerimaan Mahasiswa Baru tahun 2015-2016 sebanyak 4117 data dengan 16 atribut. Adapun tahap ini mencakup beberapa proses antara lain :

a. Pemilihan Data

Pada proses ini dilakukan seleksi terhadap beberapa atribut yang akan digunakan untuk pemodelan.

b. Pembersihan Data

Data awal sebanyak 4117 *record* tidak semuanya berisi data yang utuh/lengkap. Maka dilakukan proses *cleaning*, dimana data yang kosong tersebut dihilangkan sehingga tersisa data sebanyak 2666 *record*.

c. Transformasi Data

Beberapa atribut pada data dapat disederhanakan menjadi atribut baru seperti atribut Asal Sekolah dan Asal PT yang disederhanakan menjadi atribut Asal Pendidikan. Kemudian dilakukan pula konversi nilai pada atribut TglRegister menjadi atribut Registrasi dengan nilai Ya dan Tidak, atribut Kota Asal yang nilainya berisi kota di luar provinsi Jawa Tengah dikonversi menjadi satu kategori yaitu Luar Jateng, atribut Asal Pendidikan yang awalnya berisi nama sekolah dan universitas maka dikonversi menjadi beberapa kategori yaitu SMA, SMK, MA, PKBM, dan Transfer.

Adapun hasil dari fase pengolahan data di atas dapat ditunjukkan pada tabel 2.

Tabel 2. Potongan Hasil Fase Pengolahan Data

Prodi	KotaAsal	JenisKelamin	Agama	AsalPendidikan	KelasKuliah	Registrasi
Akuntansi	Kota Semarang	Pria	Islam	SMK	Sore	Ya
Akuntansi	Kota Semarang	Wanita	Katholik	SMK	Sore	Ya
Akuntansi	Boyolali	Pria	Islam	SMA	Sore	Tidak
Akuntansi	Kota Semarang	Pria	Islam	Transfer	Sore	Ya
Teknik Informatika	Kota Semarang	Pria	Katholik	SMK	Sore	Ya
Teknik Informatika	Kota Semarang	Pria	Kristen	SMA	Pagi	Tidak
Teknik Informatika	Purwodadi	Pria	Islam	SMK	Sore	Ya
Sastra Inggris	Grobogan	Pria	Islam	SMK	Sore	Ya

4. Fase Pemodelan (*Modelling*)

Untuk tahapan permodelan ini akan digunakan teknik *Data Mining* dengan metode Klasifikasi menggunakan algoritma C45 dan *Naive Bayes* dengan tools RStudio, yang nantinya akan menghasilkan aturan klasifikasi dan variabel paling berpengaruh dalam memprediksi masuknya calon mahasiswa baru.

5. Fase Evaluasi (*Evaluation*)

Pada tahap ini dilakukan evaluasi terhadap keefektifan dan kualitas model yang digunakan, selanjutnya didapatkan hasil prediksi untuk masing-masing algoritma klasifikasi. Hasil prediksi kemudian diuji tingkat akurasi dengan bantuan metode *confusion matrix*. Setelah itu dilakukan proses perbandingan tingkat akurasi di tiap algoritma untuk menentukan algoritma mana yang memiliki akurasi tertinggi.

6. Fase Penyebaran (Deployment)

Pada tahap ini program yang telah dibuat akan dipresentasikan dalam bentuk laporan dan dapat direkomendasikan ke pihak perguruan tinggi sebagai acuan dalam memprediksi masuknya calon mahasiswa baru di masa yang akan datang.

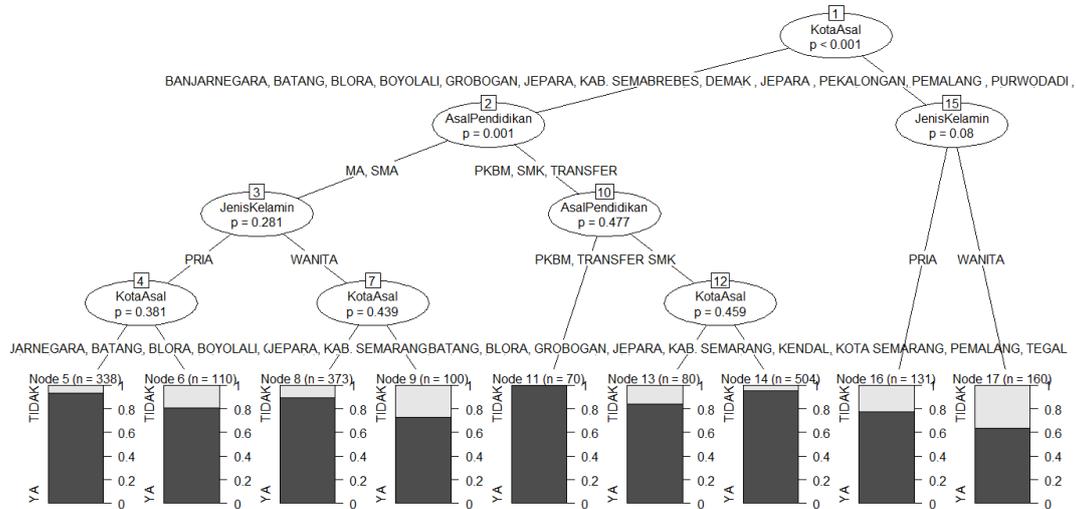
3.2. Metode Klasifikasi

Klasifikasi merupakan proses penemuan model yang menggambarkan dan membedakan kelas data dengan tujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya belum diketahui. Metode ini terdiri dari 2 proses yaitu fase *training*, dimana algoritma klasifikasi menganalisa data *training* lalu hasilnya dipresentasikan dalam bentuk *rule* klasifikasi. Proses kedua yaitu klasifikasi, dimana data *testing* digunakan untuk menguji tingkat akurasi dari *rule* klasifikasi tersebut[6].

4. HASIL DAN PEMBAHASAN

4.1. Implementasi Algoritma C45

Proses *data mining* diawali dengan membuat model klasifikasi, dimana pemodelan menggunakan data *training* sebanyak 1866record. Berikut adalah hasil *model decision tree* C45 :



Gambar 2. Hasil Plot Decision Tree C45

Dari hasil plot pada gambar 2. menunjukkan bahwa atribut Kota Asal menjadi *root node* sehingga menjadi atribut paling berpengaruh terhadap keputusan calon mahasiswa saat mendaftar di Unisbank. Selain plot, pemodelan C45 ini juga menghasilkan beberapa aturan pohon keputusan sebagai berikut :

```

Model formula:
Registrasi ~ Prodi + KotaAsal + JenisKelamin + Agama + AsalPendidikan +
kelaskuliah
Fitted party:
[1] root
  [2] KotaAsal in BANJARNEGARA, BATANG, BLORA, BOYOLALI, GROBOGAN, JEPARA, KAB. SEMARANG, K
  ENDAL, KOTA SEMARANG, KUDUS, LUAR JATENG, PATI, PEMALANG, TEGAL
    [3] AsalPendidikan in MA, SMA
      [4] JenisKelamin in PRIA
        [5] KotaAsal in BANJARNEGARA, BATANG, BOYOLALI, JEPARA, KENDAL, KOTA SEMARANG
        TEGAL : YA (n = 338, err = 6.8%)
        [6] KotaAsal in BLORA, GROBOGAN, KAB. SEMARANG, KUDUS, LUAR JATENG, PATI, PE
        MALANG : YA (n = 110, err = 19.1%)
      [7] JenisKelamin in WANITA
        [8] KotaAsal in BANJARNEGARA, BATANG, BLORA, BOYOLALI, GROBOGAN, KOTA SEMARAN
        G, KUDUS, LUAR JATENG, PATI, PEMALANG : YA (n = 373, err = 10.5%)
        [9] KotaAsal in JEPARA, KAB. SEMARANG, KENDAL, TEGAL : YA (n = 100, err = 27.
        0%)
    [10] AsalPendidikan in PKBM, SMK, TRANSFER
      [11] AsalPendidikan in PKBM, TRANSFER: YA (n = 70, err = 0.0%)
      [12] AsalPendidikan in SMK
        [13] KotaAsal in BANJARNEGARA, BOYOLALI, KUDUS, LUAR JATENG, PATI: YA (n = 8
        0, err = 16.2%)
        [14] KotaAsal in BATANG, BLORA, GROBOGAN, JEPARA, KAB. SEMARANG, KENDAL, KOTA
        SEMARANG, PEMALANG, TEGAL : YA (n = 504, err = 4.6%)
    [15] KotaAsal in BREBES, DEMAK, JEPARA, PEKALONGAN, PEMALANG, PURWODADI, REMBANG
      [16] JenisKelamin in PRIA: YA (n = 131, err = 22.9%)
      [17] JenisKelamin in WANITA: YA (n = 160, err = 36.2%)
Number of inner nodes: 8
Number of terminal nodes: 9
    
```

Gambar 3. Hasil Aturan Pohon Keputusan C45

Setelah melakukan pemodelan maka model akan diuji pada data *testing* sebanyak 800record sehingga dihasilkan *class* prediksi sebagai berikut :

	prediksi	aktual
[1,]	"YA"	"YA"
[2,]	"YA"	"TIDAK"
[3,]	"YA"	"TIDAK"
[4,]	"YA"	"YA"
[5,]	"YA"	"YA"
[6,]	"YA"	"TIDAK"
[7,]	"YA"	"YA"
[8,]	"YA"	"TIDAK"
[9,]	"YA"	"TIDAK"
[10,]	"YA"	"TIDAK"
[11,]	"YA"	"YA"
[12,]	"YA"	"YA"
[13,]	"YA"	"YA"
[14,]	"YA"	"TIDAK"
[15,]	"YA"	"TIDAK"
[16,]	"YA"	"YA"
[17,]	"YA"	"TIDAK"
[18,]	"YA"	"TIDAK"
[19,]	"YA"	"TIDAK"
[20,]	"YA"	"YA"

Gambar 4. Potongan Hasil Prediksi C45

Pada gambar 4. menunjukkan prediksi yang dihasilkan menggunakan algoritma C45, dari hasil tersebut terlihat bahwa prediksi dan aktual menghasilkan nilai yang berbeda yaitu sebagian besar nilai aktual TIDAK

diprediksi YA oleh program. Selanjutnya hasil prediksi tersebut diuji tingkat akurasi dengan menggunakan metode *confusion matrix* dan dicari tingkat kesalahan prediksinya dengan hasil sebagai berikut :

```

Confusion Matrix and Statistics

      aktual
prediksi TIDAK YA
TIDAK      0    0
YA         100  700

      Accuracy : 0.875
      95% CI : (0.8501, 0.8971)
      No Information Rate : 0.875
      P-value [Acc > NIR] : 0.5266

      Kappa : 0

      Mcnemar's Test P-value : <2e-16

      Sensitivity : 0.000
      Specificity : 1.000
      Pos Pred Value : NaN
      Neg Pred Value : 0.875
      Prevalence : 0.125
      Detection Rate : 0.000
      Detection Prevalence : 0.000
      Balanced Accuracy : 0.500

      'Positive' Class : TIDAK

> #nilai akurasi
> akurasi1 <- percent(sum(diag(tabt1))/sum(tabt1))
> akurasi1
[1] "87.5%"
> #nilai missclassification
> error1 <- percent(1-sum(diag(tabt1))/sum(tabt1))
> error1
[1] "12.5%"
    
```

Gambar 5. Hasil Pengujian Akurasi dan Tingkat Kesalahan Prediksi

Pada gambar 5. menunjukkan hasil *confusion matrix* berupa tingkat akurasi prediksi dengan algoritma C45 sebesar 87.5% sedangkan tingkat kesalahan prediksinya sebesar 12.5%.

4.2. Implementasi Algoritma Naive Bayes

Proses awal implementasi ini dimulai dengan membuat model *naive bayes* menggunakan data *training* sebanyak 1866record. Berikut adalah hasil pemodelannya :

```

A-priori probabilities:
Y
  TIDAK    YA
0.1254019 0.8745981

Conditional probabilities:
Prodi
Y
  AKUNTANSI  ILMU HUKUM  KEUANGAN  PERBANKAN  MANAJEMEN  MANAJEMEN  INFORMATIKA  PERHOTELAN
TIDAK 0.324786325 0.068376068      0.034188034 0.316239316      0.025641026 0.055555556
YA    0.256740196 0.049019608      0.052083333 0.259803922      0.041666667 0.049019608

Prodi
Y
  SASTRA  INGGRIS  SISTEM  INFORMASI  TEKNIK  INDUSTRI  TEKNIK  INFORMATIKA
TIDAK 0.008547009      0.042735043      0.017094017      0.106837607
YA    0.051470588      0.068627451      0.023284314      0.148284314

KotaAsal
Y
  BANJARNEGARA  BATANG  BLORA  BOYOLALI  BREBES  DEMAK  GROBOGAN  JEPARA
TIDAK 0.012820513 0.008547009 0.017094017 0.008547009 0.047008547 0.136752137 0.021367521 0.021367521
YA    0.012867647 0.017769608 0.016544118 0.015318627 0.013480392 0.053308824 0.023284314 0.020220588

KotaAsal
Y
  JEPARA  KAB.  SEMARANG  KENDAL  KOTA  SEMARANG  KUDUS  LUAR  JATENG  PATI
TIDAK 0.004273504 0.029914530 0.085470085 0.260683761 0.017094017 0.047008547 0.064102564
YA    0.000000000 0.036151961 0.094362745 0.497549020 0.014705882 0.046568627 0.057598039

KotaAsal
Y
  PEKALONGAN  PEMALANG  PEMALANG  PURWODADI  REMBANG  TEGAL
TIDAK 0.034188034 0.004273504 0.081196581 0.029914530 0.042735043 0.025641026
YA    0.008578431 0.003063725 0.016544118 0.014705882 0.017769608 0.019607843

JenisKelamin
Y
  PRIA  WANITA
TIDAK 0.3974359 0.6025641
YA    0.4914216 0.5085784

Agama
Y
  BUDHA  HINDHU  ISLAM  KATHOLIK  KRISTEN
TIDAK 0.000000000 0.000000000 0.940170940 0.012820513 0.047008547
YA    0.002450980 0.001838235 0.911151961 0.029411765 0.055147059

AsalPendidikan
Y
  MA  PKBM  SMA  SMK  TRANSFER
TIDAK 0.055555556 0.004273504 0.705128205 0.222222222 0.012820513
YA    0.041666667 0.005514706 0.548406863 0.363357843 0.041053922

kelaskuliah
Y
  PAGI  SORE
TIDAK 0.8504274 0.1495726
YA    0.7824755 0.2175245
    
```

Gambar 6. Hasil Model Naive Bayes

Pada gambar 6. menunjukkan hasil pemodelan *Naive Bayes* dimana tiap nilai pada atribut memiliki nilai probabilitas untuk kasus YA dan TIDAK. Setelah melakukan pemodelan maka model akan diuji pada data *testing* sebanyak 800 record sehingga dihasilkan *class* prediksi sebagai berikut :

	prediksi	aktual
[1,]	"YA"	"YA"
[2,]	"YA"	"TIDAK"
[3,]	"YA"	"TIDAK"
[4,]	"YA"	"YA"
[5,]	"TIDAK"	"YA"
[6,]	"YA"	"TIDAK"
[7,]	"YA"	"YA"
[8,]	"YA"	"TIDAK"
[9,]	"YA"	"TIDAK"
[10,]	"YA"	"TIDAK"
[11,]	"YA"	"YA"
[12,]	"YA"	"YA"
[13,]	"YA"	"YA"
[14,]	"YA"	"TIDAK"
[15,]	"TIDAK"	"TIDAK"
[16,]	"YA"	"YA"
[17,]	"TIDAK"	"TIDAK"
[18,]	"TIDAK"	"TIDAK"
[19,]	"TIDAK"	"TIDAK"
[20,]	"YA"	"YA"

Gambar 7. Potongan Hasil Prediksi *Naive Bayes*

Pada gambar 7. menunjukkan prediksi yang dihasilkan menggunakan algoritma *Naive Bayes*, dari hasil tersebut terlihat bahwa prediksi dan aktual menghasilkan nilai yang berbeda, artinya algoritma *Naive Bayes* tidak mutlak memprediksi kelas untuk semua kasus secara tepat. Selanjutnya hasil prediksi tersebut diuji tingkat akurasi dengan menggunakan metode *confusion matrix* dan dicari tingkat kesalahan prediksinya dengan hasil sebagai berikut :

```

Confusion Matrix and Statistics

      aktual
prediksi TIDAK  YA
TIDAK      8    15
YA         92   685

      Accuracy : 0.8662
      95% CI : (0.8407, 0.8891)
      No Information Rate : 0.875
      P-Value [Acc > NIR] : 0.7901

      Kappa : 0.0874

      Mcnemar's Test P-Value : 2.024e-13

      Sensitivity : 0.08000
      Specificity : 0.97857
      Pos Pred Value : 0.34783
      Neg Pred Value : 0.88160
      Prevalence : 0.12500
      Detection Rate : 0.01000
      Detection Prevalence : 0.02875
      Balanced Accuracy : 0.52929

      'Positive' Class : TIDAK

> #nilai akurasi
> akurasinb1 <- percent(sum(diag(tabnb1))/sum(tabnb1))
> akurasinb1
[1] "86.6%"
> #nilai missclassification
> errornb1 <- percent(1-sum(diag(tabnb1))/sum(tabnb1))
> errornb1
[1] "13.4%"
    
```

Gambar 8. Hasil Pengujian Akurasi dan Tingkat Kesalahan Prediksi

Pada gambar 8. menunjukkan hasil *confusion matrix* berupa tingkat akurasi prediksi dengan algoritma C45 sebesar 86.6% sedangkan tingkat kesalahan prediksinya sebesar 13.4%. Selain itu telah dilakukan juga dua kali pengujian dengan menggunakan data *training* sebanyak 1999 data dan data *testing* sebanyak 667 data, juga dengan data *training* sebanyak 2133 data serta data *testing* sebanyak 533 data. Hasilnya ditunjukkan pada tabel 3. di bawah ini :

Tabel 3. Hasil Pengujian

Kegiatan	Data Training	Data Testing	Akurasi %		Error Rate %	
			C45	NB	C45	NB
Uji coba 1	1866 data	800 data	87.5	86.6	12.5	13.4
Uji coba 2	1999 data	667 data	87.41	85.8	12.59	14.2
Uji coba 3	2133 data	533 data	87.43	86.5	12.57	13.5

Pada tabel 3. menunjukkan bahwa di tiga kali uji coba, algoritma C45 memiliki hasil akurasi yang lebih baik dibandingkan dengan algoritma *Naive Bayes*. Dari ketiga uji coba itu pada uji coba pertama memiliki hasil yang lebih baik yaitu untuk akurasi algoritma C45 sebesar 87.5% dan akurasi algoritma *Naive Bayes* sebesar 86.6%.

5. KESIMPULAN

Berdasarkan hasil penelitian klasifikasi menggunakan algoritma *decision tree* C45 dan *Naive Bayes* pada kasus prediksi masuknya calon mahasiswa baru di Unisbank, dapat diambil kesimpulan sebagai berikut :

1. Dari tiga kali uji coba didapatkan hasil akurasi terbaik pada uji coba pertama yaitu dengan data *training* sebanyak 1866 data dan data *testing* sebanyak 800 data. Adapun hasilnya untuk algoritma C45 adalah akurasi sebesar 87.5 % dan tingkat kesalahan prediksi sebesar 12.5% sedangkan untuk algoritma *Naive Bayes* akurasi sebesar 86.6% dan tingkat kesalahan prediksinya sebesar 13.4%.
2. Dari hasil pengujian menunjukkan bahwa algoritma C45 memiliki tingkat akurasi dan kesalahan prediksi yang lebih baik dibandingkan dengan algoritma *Naive Bayes*. Hal tersebut membuktikan bahwa algoritma C45 lebih unggul dibandingkan algoritma *Naive Bayes* untuk kasus prediksi masuknya calon mahasiswa baru di Unisbank.
3. Dari hasil pengujian pada algoritma C45 telah didapatkan kumpulan aturan pohon keputusan (*rule*) yang menunjukkan bahwa atribut Kota Asal merupakan atribut yang paling berpengaruh dalam keputusan seorang calon mahasiswa baru mendaftar di Unisbank, kemudian diikuti dengan atribut Asal Pendidikan dan Jenis Kelamin.

6. SARAN

Adapun saran yang diberikan untuk penelitian lebih lanjut yaitu sebagai berikut:

1. Hasil penelitian berdasarkan analisa dengan Rstudio dapat digunakan sebagai *Decision Support System* bagi pihak Unisbank guna merancang strategi atau kebijakan dalam penerimaan mahasiswa baru khususnya memprediksi masuknya calon mahasiswa baru.
2. Sebaiknya untuk penelitian selanjutnya dapat menggunakan *atribut predictor* dan nilai data yang lebih banyak serta lebih bervariasi untuk menghasilkan nilai akurasi yang lebih baik lagi.
3. Untuk kedepannya bisa dilakukan implementasi terhadap algoritma klasifikasi selain C45 dan *Naive Bayes* seperti KNN, *Neural Network*, maupun *Support Vector Machine* yang hasilnya dapat dilakukan perbandingan dan analisa.

DAFTAR PUSTAKA

- [1] Retnosari, P. and Jananto, A. (2013) 'Implementasi Data Mining Untuk Menemukan Hubungan Antara Kota Kelahiran Mahasiswa Dengan Tingkat Kelulusan Mahasiswa Pada Fakultas Teknologi Informasi Unisbank', *Jurnal Dinamika Informatika*, 5(2), pp. 112–121.
- [2] Supriyatna, A. and Mustika, W. P. (2018) 'Komparasi Algoritma Naive Bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil', *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 2(2), pp. 152–161.
- [3] Fiastantyo, G. (2014) 'Perbandingan Kinerja Metode Klasifikasi Data Mining Menggunakan Naive Bayes dan Algoritma C4.5 untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa', *Skripsi*, Program Studi Teknik Informatika FILKOM Udinus, Semarang.
- [4] Iskandar, D. and Suprpto, Y. K. (2015) 'Perbandingan Akurasi Klasifikasi Tingkat Kemiskinan Antara Algoritma C4.5 dan Naive Bayes', *Jurnal Ilmiah NERO*, 2(1), pp. 37–43.
- [5] Supriyanti, W., Kusri and Amborowati, A. (2016) 'Perbandingan Kinerja Algoritma C4.5 Dan Naive Bayes Untuk Ketepatan Pemilihan Konsentrasi Mahasiswa', *Jurnal Informa Politeknik Indonesia Surakarta*, 1(2016), pp. 46–52.
- [6] Niswatin, R. K. (2015) 'Penempatan Jurusan Mahasiswa Baru Menggunakan Metode K-Nearest Neighbor', *Journal Cogito Smart*, 1(1), pp. 55–67.
- [7] Novianti, B., Rismawan, T. and Bahri, S. (2016) 'Implementasi Data Mining Dengan Algoritma C4.5 Untuk Penjurusan Siswa (Studi Kasus: SMA Negeri 1 Pontianak)', *Jurnal Coding Sistem Komputer Untan*, 4(3).
- [8] Jananto, A. (2013) 'Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa', *Jurnal Teknologi Informasi DINAMIK*, 18(1), pp. 9–16.