CESS (Journal of Computer Engineering System and Science) **p-ISSN :2502-7131**
Vol. 4 No. 2 Juli 2019 **e-ISSN :2502-714x**

PERBANDINGAN ALGORITMA K-NN DAN CART PADA DATA MINING PENERIMAAN BEASISWA

Rahman Rosyidi
STMIK AMIKOM Purwokerto
jl. Let. Jend. Pol. Soemarto, Watumas, Purwokerto
amang@amikompurwokerto.ac.id

Abstrak—**Beasiswa merupakan salah satu bantuan yang diberikan untuk seseorang dalam menunjang keberhasilan pendidikannya. Oleh sebab itu, sewajarnya beasiswa hanya diberikan kepada yang layak dan pantas mendapat beasiswa. Penulis akan membandingkan algoritma K-Nearest Neighbor (KNN) dan Classification and Regression Tress (CART) untuk memprediksi penerima beasiswa. Data yang diperlukan merupakan para mahasiswa pendaftar beasiswa STMIK AMIKOM Purwokerto pada 2015-2016 dengan jumlah data sekitar 150. Jenis kelamin, semester, IPK, pekerjaan orang tua, jumlah anggota keluarga, penghasilan orang tua, prestasi, dan status merupakan atribut dalam proses analisa. Dari perhitungan yang dilakukan, didapatkan hasil akurasi dari masing-masing algoritma yaitu 99.2958 % dengan nilai precision 0.993, recall 0.993, dan F- measure 0.993 pada algoritma KNN. Sementara pada algoritma CART didapatkan nilai akurasi sebesar 71.1268% dengan nilai precision 0.506, recall 0.711, dan F- measure 0.591.**

Keywords— **Beasiswa, Data Mining, Klasifikasi, KNN, CART**

Abstract—**Scholarship is one of the assistance given to someone in supporting the success of their education. Therefore, naturally the scholarship is only given to the decent and deserves a scholarship. The author will compare the K-Nearest Neighbor (KNN) algorithm and Classification and Regression Tress (CART) to predict scholarship recipients. The required data are students of STMIK AMIKOM Purwokerto scholarship applicants in 2015-2016 with a total data of around 150. Gender, semester, GPA, parental work, family members, parents' income, achievements, and status are attributes in the analysis process. From the calculation, the accuracy of each algorithm is 99.2958% with precision 0.993, recall 0.993, and F-measure 0.993 on the KNN algorithm. While the CART algorithm obtained an accuracy value of 71.1268% with a value of precision 0.506, recall 0.711, and F- measure 0.591.**

Keywords— **Scholarships, Data Mining, Classification, KNN, CART**

I. PENDAHULUAN

Penambangan data adalah proses pengumpulan informasi penting dari sejumlah data besar yang tersimpan di basis data, gudang data, atau tempat penyimpanan lainnya (Han & Kamber, 2006). Data mining sangat diperlukan terutama dalam mengelola jumlah data yang besar agar dapat memberikan informasi yang akurat untuk penggunanya. Dalam menambang data terdapat beberapa algoritma klasifikasi untuk memproses data dalam jumlah besar tersebut. Proses penilaian terhadap objek data lalu memilah dan mengelompokkan ke dalam suatu kelas tertentu yang telah tersedia disebut klasifikasi (Prasetyo, 2012).

Teknik klasifikasi dapat diterapkan untuk menyelesaikan suatu kasus yang berhubungan dengan objek. Bidang kesehatan misalnya, tingkat penyakit pasien dapat terdeteksi dan membantu petugas medis ketika memberikan solusi terapi bagi pasien dengan tepat. Pada bidang ekonomi, aplikasi klasifikasi juga dapat digunakan oleh sebuah bank yang ingin mengetahui apakah customer yang mengajukan kredit termasuk dalam kategori customer yang menguntungkan atau tidak. Metode klasifikasi yang umum digunakan antara lain *decision tree, K-Nearest Neighbor, naive bayes, neural network* dan *support vector machine*.

Berdasarkan jurnal penelitian yang dilakukan oleh Arief Jananto berjudul "Perbandingan Performansi Algoritma Nearest Neighbor Dan SLIQ Untuk Prediksi Kinerja Akademik Mahasiswa Baru (Studi Kasus : Data Akademik Fakultas Teknologi Informasi UNISBANK)" peneliti mencoba membandingkan dua algoritma, yaitu *Nearest Neighbor* dan SLIQ yang menghasilkan bahwa metode *Nearest Neighbor* bekerja lebih baik dibandingkan dengan SLIQ dilihat dari tingginya nilai akurasi pada data akademik mahasiswa tahun 2005 sebesar 78,5% dan data tahun 2006 tingkat akurasinya sebesar 86,5% dibandingkan algoritma SLIQ yang menghasilkan nilai akurasi 41,67% pada data akademik mahasiswa tahun 2005 dan data akademik mahasiswa tahun 2006 tingkat akurasinya sebesar 63,11%. Dalam jurnal penelitian berbeda yang dilakukan oleh Aquarahma Margasari (2014) berjudul "Penerapan Metode CART

(Classification And Regression Trees) Dan Analisis Regresi Logistik Biner Pada Klasifikasi Profil Mahasiswa Fmipa Universitas Brawijaya" menghasilkan perbandingan dari dua analisis yang menunjukkan bahwa CART (94.2%) menghasilkan keakuratan prediksi lebih besar dari analisis regresi logistik biner (86.7%). Dari kedua jurnal itu diketahui metode *K-Nearest Neighbor* (KNN) dan CART lebih baik dari metode pembandingnya. Keunggulan K-NN sangat efektif untuk *training* data yang memiliki *noise* dan jumlah besar. Metode CART bisa digunakan untuk data berjumlah besar, data dengan variabel banyak atau data dengan variabel campuran berdasarkan pemilihan biner. Teknik eksplorasi umumnya menggunakan teknik pohon keputusan. Metode CART terbilang sederhana namun hasilnya lebih mudah diintrepetasikan, akurat dan perhitungannya lebih cepat dibandingkan dengan metode klasifikasi lainnya. Oleh karena itu penulis mencoba melakukan penelitian dengan membandingkan metode *K-Nearest Neighbor* (K-NN) dan *Classification and Regression Trees (CART)*. Perbandingan dilakukan untuk mengetahui algoritma mana yang tingkat akurasinya lebih tinggi.

Berdasarkan uraian dari latar belakang, maka masalah artikel ini adalah antara *K-Nearest Neighbor* (K-NN) dan *Classification and Regression Trees* (CART) mana yang akurasinya lebih tinggi ?

Tujuan yang akan dicapai untuk melihat nilai akurasi yang lebih tinggi antara *K-Nearest Neighbor* (K-NN) dan *Classification and Regression Trees* (CART) pada pendaftar beasiswa di STMIK AMIKOM Purwokerto periode 2015-2016.

II. METODOLOGI PENELITIAN

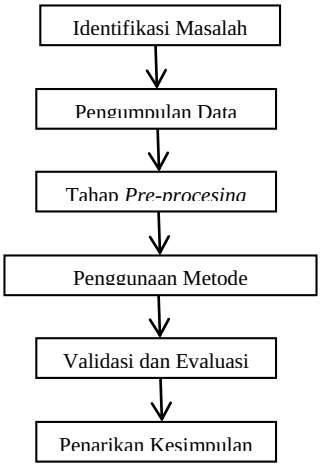
a. Analisa Permasalahan

Pemberian beasiswa ditujukan kepada mahasiswa kurang mampu dan mahasiswa yang memiliki prestasi akademik bagus. Penentuan mahasiswa penerima beasiswa harus memenuhi kriteria-kriteria. Yang paling diperhatikan adalah IPK, prestasi, dan penghasilan orang tua. Mula-mula mahasiswa yang ingin mengajukan beasiswa mengumpulkan berkas administrasi ke pihak kemahasiswaan. Setelah semua berkas administrasi masuk, diseleksi manual oleh sekretaris kemahasiswaan, dan membutuhkan waktu yang lama untuk menyeleksi karena hanya satu orang yang menyeleksi. Setelah berkas administrasi yang lengkap terpilih, nantinya diseleksi lagi, sebelum diserahkan ke atasan untuk dirapatkan dan menghasilkan keputusan mahasiswa yang menerima beasiswa.

Permasalahan yang sering muncul dalam penyaluran beasiswa terhadap mahasiswa. Penyaluran masih tidak tepat sasaran, karena ada mahasiswa yang berhak namun tidak mendapatkan beasiswa. Ada yang tidak berhak, ternyata mendapatkan beasiswa. Untuk itu diperlukan data pendaftar beasiswa di STMIK AMIKOM Purwokerto periode 2015-2016, karena tahun-tahun sebelumnya tidak adanya data prestasi, padahal data prestasi salah satu yang perlu dipertimbangkan mahasiswa itu berhak mendapatkan beasiswa atau tidak. Algoritma yang dibandingkan adalah *K-Nearest Neighbor* (K-NN) dan *Classification and Regression Trees* (CART) karena dalam jurnal diatas, terbukti algoritma itu memiliki akurasi tinggi. Sehingga dapat diketahui algoritma mana yang tingkat akurasinya lebih tinggi dalam memprediksi penerimaan beasiswa.

b. Metode Yang Digunakan

Perbandingan algoritma K-NN dengan CART akan dilakukan melalui beberapa tahap seperti ditunjukkan oleh gambar 21 sebagai berikut:



Gbr 1. Diagram Alur Penelitian (Kurniawati, 2016)

1. Identifikasi Masalah

Suatu upaya mengetahui permasalahan serta metode yang sesuai sehingga dapat ditentukan kriteria-kriteria untuk penentuan penerima beasiswa.

2. Pengumpulan Data

Dalam penelitian ini data diambil dari data pendaftar beasiswa di STMIK AMIKOM Purwokerto.

3. Tahap Pre-procesing

Proses seleksi data dengan tujuan untuk mendapatkan data yang lebih bersih dan lebih siap digunakan sebagai bahan penelitian. Terdiri dari identifikasi dan pemilihan atribut, penanganan atribut yang tidak lengkap serta diskritisasi nilai.

4. Penggunaan Metode Klasifikasi

Perhitungan dalam algoritma K-NN:

a. Algoritma K-NN (IBk) dalam weka dalam penentuan dataset pendaftar beasiswa.
b. Nilai k yang digunakan berbeda-beda pada setiap proses pengujian.
c. Menghasilkan *classifier* dan *confusion matrix*.
d. Menghitung *precision*, *recall*, dan *F-measure*.

e. Melakukan perhitungan CART.

5. Validasi dan Evaluasi

Validasi dan keakuratan hasil dihitung dengan *confusion matrix* dan *cross-validation* yang terdapat pada aplikasi weka.

a) Confusion Matrix

Nilai *accuracy*, *precision*, *recall*, dan *F-measure* dihasilkan dengan rumus (Han & Kamber, 2006) :

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

$$Recall = \frac{TP}{TP + FN}$$

$$F - measure = \frac{2\ x\ precision\ x\ recall}{precision + recall}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

b) K-Fold Cross Validation

Himpunan contoh dibagi ke dalam k himpunan bagian secara acak. Pengulangan dilakukan sebanyak k kali dan pada setiap ulangan disisakan satu subset untuk pengujian dan subset-subset lainnya untuk pelatihan (Sulistyo, Kustiyo dan Buono, 2008). Pada penelitian ini, penulis menggunakan metode 10-cross validation. Cara kerja metode ini, dimana subset dibagi menjadi 10 subsets (9 subsets sebagai training sets dan 1 subset sebagai testing set) dengan jumlah 10 kali iterasi. Hasil pengukurannya yaitu berdasarkan nilai rata-rata dari 10 kali pengujian.

c) Pengujian Hipotesis

Perhitungan dengan *paired sample t test* menggunakan *software* SPSS versi 19. Langkah pengujiannya adalah sebagai berikut (Sugiyono, 2012) :

1) Perumusan hipotesis statistik

$H_0 : X_1 = X_2$ Tidak terdapat perbedaan antara nilai kinerja Algoritma KNN dan CART

$H_1 : X_1 \neq X_2$ Terdapat perbedaan antara nilai kinerja Algoritma KNN dan CART

2) Penentuan tingkat keyakinan (95 persen)

3) Kriteria penerimaan dan penolakan H_0

Jika nilai t tabel $\leq$ t hitung $\leq$ t tabel atau nilai sig $\geq \propto$ (0,05) maka H_0diterima.

Jika nilai t hitung $>$ t tabel atau t hitung $<$ - t tabel atau nilai sig $< \propto$ (0,05) maka H_0 ditolak

4) Menerima atau menolak H_0

6. Penarikan Kesimpulan

Melihat nilai *precision*, *recall*, *F-measure* dari masing-masing algoritma dengan tingkat diagnosanya yaitu : *excellent classification* = 0.90 - 1.00, *good classification* = 0.80 – 0.90, *fair classification* = 0.70 – 0.80, *poor classification* = 0.60 – 0.70, dan *failure* = 0.50 – 0.60 (Gorunescu, 2011).

III. HASIL DAN PEMBAHASAN

a. Pengumpulan Data

Sumber data utama adalah bagian kemahasiswaan STMIK AMIKOM Purwokerto yang menyediakan data penerima beasiswa. Terdapat 8 atribut / fitur dalam data tersebut, yang terdiri dari 7 atribut kriteria dan 1 atribut keputusan. Adapun atribut-atribut yang digunakan adalah sebagai berikut : jenis kelamin, semester, IPK, pekerjaan orang tua, jumlah anggota keluarga, penghasilan orang tua, prestasi dan status. Data ini harus diolah terlebih dahulu melalui tahap *pre-procesing*, dimana tahapan ini untuk menyesuaikan atribut-atribut yang akan digunakan dalam mengolah *dataset* tersebut. Berikut ini adalah tabel 1 *dataset* pendaftar beasiswa STMIK AMIKOM Purwokerto yang belum dilakukan penyesuaian.

TABEL I
DATASET SEBELUM PENYESUAIAN

Jenis Kelamin	Semester	IPK	Pekerjaan Orang Tua	Jumlah Anggota Keluarga	Penghasilan (Rp)	Prestasi	STATUS
P	4	3,82	Wiraswasta	3	1.300.000	-	Ditolak
P	4	3,53	Wiraswasta	4	1.500.000	-	Ditolak
L	6	3,51	Wiraswasta	4	2.000.000	-	Ditolak
P	6	3,66	Petani	4	1.500.000	-	Ditolak
L	6	3,6	Swasta	7	650.000	Finalis IT Competition "Elinfo" Tingkat Nasional Tahun 2013	Diterima
L	6	3,46	Buruh_Tani	4	300.000	Finalis "Elinfo" Competition Tingkat Nasional Tahun 2014	Diterima
L	6	3,36	Pedagang	4	200.000	-	Diterima
L	2	3,25	Pedagang	6	1.500.000	-	Ditolak
L	6	3,73	PNS	4	1.500.000	Finalis ITCC Udayana 2013, Semifinalis Software Development Competition Technocorner 2014	Ditolak
P	4	3,66	Buruh	5	1.000.000	Finalis Elinfo Tahun 2014	Ditolak

P	6	3,1	Pedagang	4	1.500.000	-	Ditolak
P	6	3,28	Wiraswasta	4	1.750.000	-	Ditolak
L	6	3,51	Buruh	4	1.500.000	Finalis ITCC Udayana 2013, Semifinalis Software Development Competition Technocorner 2014	Ditolak
L	6	3,72	Supir	3	1.000.000	-	Diterima
L	4	3,59	PNS	4	303.6000	-	Ditolak
P	2	3,38	Buruh	3	1.000.000	-	Diterima
P	2	3,38	Polri	4	2.220.400	_	Ditolak
P	2	3,46	Buruh_Tani	9		_	Ditolak
L	2	3,75	Wiraswasta	8	1.500.000	_	Ditolak
L	2	4	TNI	7	3.762.228	_	Ditolak
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
L	6	2,92	Swasta	4	2.500.000	_	Ditolak
L	6	3,48	Karyawan_BUMN	5	3.000.000	Juara II Kompetisi kelas Prima	Ditolak
P	6	3,57	Swasta	4	1.300.000	Forum Asisten	Ditolak
L	2	3,21	Penjahit		2.500.000	_	Ditolak

b. Tahap Pre-processing

Awal *dataset* terdiri dari 150 data dengan 8 atribut yaitu jenis kelamin, semester, IPK, pekerjaan orang tua, jumlah anggota keluarga, penghasilan, prestasi, dan status. Jumlah mahasiswa yang diterima 44 dan yang ditolak berjumlah 106. Terdapat data yang tidak lengkap yaitu 1 data berasal dari atribut jumlah anggota keluarga dan 7 data dari atribut penghasilan orang tua. Atribut jenis kelamin, semester, IPK, pekerjaan orang tua dan status memiliki nilai yang lengkap. Diasumsikan pekerjaan orang tua dibedakan menjadi dua yaitu bekerja dan tidak bekerja. Untuk prestasi, data yang kosong diasumsikan tidak mempunyai prestasi.

Setelah melakukan penanganan *missing value*, data menjadi 142 dengan jumlah kasus yang diterima berjumlah 41, dan yang ditolak berjumlah 101. Proses dikritisasi akan dilakukan untuk mempermudah pengelompokan nilai dan mempersempit permasalahan serta meningkatkan keakurasian (Lesmana, 2012). Berikut ini adalah penyesuaian atribut yang digunakan untuk mengolah *datast* pada tabel 2.

TABEL II
ATRIBUT/FITUR PENGHASILAN

Atribut	Keterangan	Nilai	Nilai Baru
Penghasilan orang tua	Berisikan besarnya penghasilan orang tua mahasiswa.	<=1.500.000	Rendah
		1.500.001 - 2.500.000	Sedang
		2.500.001 - 3.500.000	Tinggi
		>=3.500.001	Sangat Tinggi

(Sumber : Badan Pusat Statistik)

Setelah dilakukan proses *pre-processing*, data berjumlah 142 dengan 41 mahasiswa yang diterima dan 101 yang ditolak. Berikut adalah *dataset* yang siap digunakan dalam aplikasi Weka pada tabel 3.

TABEL III
HASIL *PREPROCESSING* DATA

Jenis Kelamin	Semester	IPK	Pekerjaan Orang Tua	Jumlah Anggota Keluarga	Penghasilan (Rp)	Prestasi	STATUS
P	4	3,82	Bekerja	3	Rendah	tidak	Ditolak
P	4	3,53	Bekerja	4	Rendah	tidak	Ditolak
L	6	3,51	Bekerja	4	Sedang	tidak	Ditolak
P	6	3,66	Bekerja	4	Rendah	tidak	Ditolak
L	6	3,6	Bekerja	7	Rendah	ada	Diterima
L	6	3,46	Bekerja	4	Rendah	ada	Diterima
L	6	3,36	Bekerja	4	Rendah	tidak	Diterima
L	2	3,25	Bekerja	6	Rendah	tidak	Ditolak
L	6	3,73	Bekerja	4	Rendah	ada	Ditolak
P	4	3,66	Bekerja	5	Rendah	ada	Ditolak
P	6	3,1	Bekerja	4	Rendah	tidak	Ditolak
P	6	3,28	Bekerja	4	Sedang	tidak	Ditolak
L	6	3,51	Bekerja	4	Rendah	ada	Ditolak
L	6	3,72	Bekerja	3	Rendah	tidak	Diterima
L	4	3,59	Bekerja	4	Tinggi	tidak	Ditolak
P	2	3,38	Bekerja	3	Rendah	tidak	Diterima
P	2	3,38	Bekerja	4	Sedang	tidak	Ditolak
L	2	3,75	Bekerja	8	Rendah	tidak	Ditolak
L	2	4	Bekerja	7	Sangat_tinggi	tidak	Ditolak
P	4	3,56	Bekerja	4	Sedang	tidak	Ditolak
P	2	3,42	Bekerja	0	Sangat_tinggi	tidak	Ditolak
P	2	3,83	Bekerja	4	Sangat_tinggi	tidak	Ditolak
P	4	3,61	Bekerja	4	Rendah	tidak	Ditolak
P	2	3,71	Bekerja	5	Tinggi	tidak	Ditolak
P	4	3,29	Bekerja	0	Rendah	tidak	Ditolak
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
L	6	2,92	Bekerja	4	Sedang	tidak	Ditolak
L	6	3,48	Bekerja	5	Tinggi	Ada	Ditolak
P	6	3,57	Bekerja	4	Rendah	Ada	Ditolak

c. Penggunaan Metode Klasifikasi

Setelah tahap *pre-procesing* selesai kemudian *dataset* tersebut mulai diolah dengan aplikasi Weka. Tahapan ini juga bertujuan menghasilkan *confusion matrix* dan melakukan 2 kali percobaan, percobaan pertama dengan metode evaluasi 10-*fold cross validation* , dimana *dataset* dibagi menjadi 10 *subsets* (9 *subsets* sebagai *training sets* dan 1 *subsets* sebagai *testing sets)* dengan jumlah 10 kali iterasi, dan yang kedua dengan *use training set*.

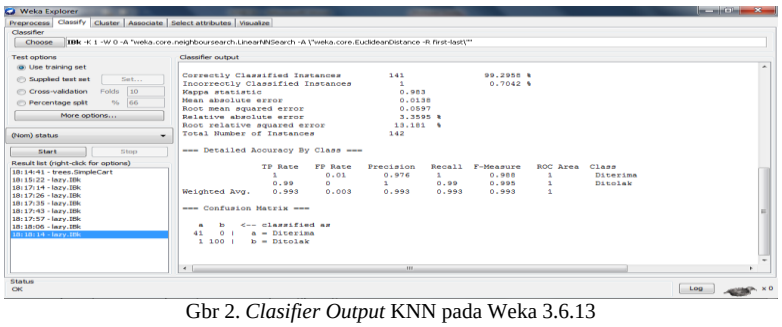
Motode K-NN akan dicoba dengan 142 *dataset* dengan 6 kali percobaan sehingga nilai k yang digunakan dari 1 sampai 6. Percobaan pertama menggunakan 10-*fold cross validation* dan percobaan kedua menggunakan *use training set.* Hasil akurasi ditunjukkan pada tabel 4 berikut ini :

TABEL IV
PERBANDINGAN AKURASI YANG DIPEROLEH DENGAN NILAI K YANG BERBEDA-BEDA.

Nilai k yang digunakan	Hasil Akurasi	
	Menggunakan 10-*fold cross validation*	Menggunakan *use training set*
k-1	64.7887 %	99.2958 %
k-2	46.4789 %	83.8028 %
k-3	57.0423 %	80.2817 %
k-4	54.2254 %	72.5352 %
k-5	64.7887 %	75.3521 %
k-6	59.8592 %	75.3521 %

Nilai akurasi berdasarkan uji coba *dataset* pendaftar beasiswa sangat dipengaruhi nilai k. Nilai k yang tinggi berakibat pada semakin banyak tetangga dalam proses klasifikasi dan *noise* semakin tinggi.

Diketahui hasil akurasi yang terbaik pada nilai k-1 yaitu sebesar 99.2958 % dengan menggunakan *use training set*. Berikut ditunjukkan hasil *output clasifier* pada weka secara rinci pada gambar2.



Gbr 2. *Clasifier Output* KNN pada Weka 3.6.13

Perhitungan hasil akurasi berdasarkan *precision, recall,* dan *F-measure tercantum dalam tabel* berikut :

TABEL V
OF CONFUSION KELAS "DITERIMA"

41*(True Positive)*	0*(False Negative)*
1 *(False Positive)*	100*(True Negative)*

dalam persamaan (1)

$$Precision = \frac{\text{TP}}{TP + FP} = \frac{41}{41 + 1} = 0.976$$

$$Recall = \frac{TP}{TP + FN} = \frac{41}{41 + 0} = 1$$

$$F - measure = \frac{2\ x\ precision\ x\ recall}{precision + recall}$$

$$= \frac{2\ x\ 0.976\ x\ 1}{0.976\ +\ 1} = 0.988$$

Kelas "Ditolak"

Berikut adalah tabel 6 yang menggambarkan *of confusion* kelas "ditolak":

TABEL VI
OF CONFUSION KELAS "DITOLAK"

100*(True Positive)*	1 *(False Negative)*
0*(False Positive)*	41*(True Negative)*

dalam persamaan (2)

$$Precision = \frac{\text{TP}}{TP + FP} = \frac{100}{100 + 0} = 1$$

$$Recall = \frac{TP}{TP + FN} = \frac{100}{100 + 1} = 0.99$$

$$F - measure = \frac{2\ x\ precision\ x\ recall}{precision + recall}$$

$$= \frac{2\ x\ 1\ x\ 0.99}{1\ +\ 0.99} = 0.995$$

Dari hasil *precision, recall,* dan *F-measure* kelas Diterima dan Ditolak, dapat dihitung nilai rata-rata dari kelas kelas-kelas yang ada *(Weighted Avg)* dengan terlebih dulu menjumlahkan nilai A = (41 + 0) = 41 dan B = (100 + 1) = 101. Rumusnya sebagai berikut :

dalam persamaan (3)

$$Weighted\ Avg\ (precision)$$

$$= \frac{0.976\ x\ 41 + 1\ x\ 101}{142}$$

$$= 0.993$$

$$Weighted\ Avg\ (recall)$$

$$= \frac{1\ x\ 41 + 0.99\ x\ 101}{142}$$

$$= 0.993$$

$$Weighted\ Avg\ (F - measure)$$

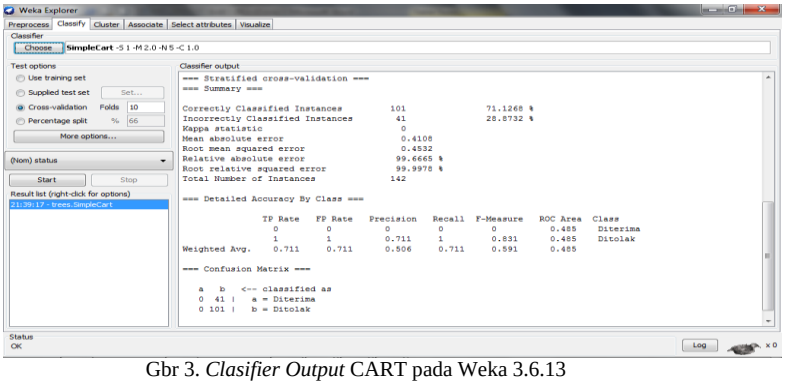
$$= \frac{0.988\ x\ 41 + 0.995\ x\ 101}{142} = 0.993$$

Nilai akurasi *confusion matrix* dapat dilihat pada tabel 7.

TABEL VII
NILAI AKURASI BERDASARKAN *CONFUSION MATRIX*

Class	*Precision*	*Recall*	*F-measure*
Diterima	0.976	1	0.988
Ditolak	1	0.99	0.995
Weighted Avg	0.993	0.993	0.993

Setelah itu, dilanjutkan dengan metode CART dengan *dataset* yang sama. Percobaan pertama menggunakan 10-*fold cross validation* dan percobaan kedua menggunakan *use training set*. Setelah dilakukan percobaan pertama dan kedua, tidak ada perbedaan nilai akurasi. Berikut ditunjukkan hasil *output clasifier* pada weka secara rinci pada gambar 3



Gbr 3. *Clasifier Output* CART pada Weka 3.6.13

Algoritma CART yang terlihat pada gambar 3 menunjukkan nilai akurasi 71,1268%. Nilai akurasi tersebut didapat dari perhitungan nilai precision, recall, dan F-Measure. Tabel 8 berikut menggambarkan proses perhitungannya:

TABEL VIII
OF CONFUSION KELAS "DITERIMA"

0(True Positive)	41 (False Negative)
0(False Positive)	101(True Negative)

Dalam persamaan (4)

$$Precision = \frac{\text{TP}}{TP + FP} = \frac{0}{0 + 0} = 0$$

$$Recall = \frac{TP}{TP + FN} = \frac{0}{0 + 0} = 0$$

$$F - measure = \frac{2\ x\ precision\ x\ recall}{precision + recall}$$

$$= \frac{2\ x\ 0\ x\ 0}{0\ +\ 0} = 0$$

Kelas "Ditolak"
Berikut adalah tabel 9 yang menggambarkan *of confusion* kelas "ditolak":

TABEL IX
OF CONFUSION KELAS "DITOLAK"

101*(True Positive)*	0*(False Negative)*
41*(False Positive)*	0*(True Negative)*

dalam persamaan (5)

$$\frac{\text{TP}}{TP + FP} = \frac{101}{101 + 41} = 0.711$$

$$Recall = \frac{TP}{TP + FN} = \frac{101}{101 + 0} = 1$$

$$F - measure = \frac{2\ x\ precision\ x\ recall}{precision + recall}$$

$$= \frac{2\ x\ 0.711\ x\ 1}{0.711\ +\ 1} = 0.831$$

Dari perhitungan *precision*, *recall*, dan *F-measure* dapat dihitung nilai rata-rata dari kelas kelas-kelas yang ada *(Weighted Avg)* dengan terlebih dulu menjumlahkan nilai A = (0 + 41) = 41 dan B = (0 + 101) = 101. Rumusnya sebagai berikut :

Dalam persamaan (6)

$$Weighted\ Avg\ (precision)$$

$$= \frac{0\ x\ 41 + 0.711\ x\ 101}{142}$$

$$= 0.506$$

$$Weighted\ Avg\ (recall) = \frac{0x\ 41 + 1\ x\ 101}{142}$$

$$= 0.711$$

$$Weighted\ Avg\ (F - measure)$$

$$= \frac{0\ x\ 41 + 0.831\ x\ 101}{142}$$

$$= 0.591$$

Nilai akurasi *confusion matrix* dapat dilihat pada tabel 10:

TABEL X
NILAI AKURASI BERDASARKAN *CONFUSION MATRIX*

Class	*Precision*	*Recall*	*F-measure*
Diterima	0	0	0
Ditolak	0.711	1	0.831
Weighted Avg	0.506	0.711	0.591

Perbandingan hasil algoritma KNN dan CART yang difokuskan pada keakurasian terdapat pada tabel 11:

TABEL XI
PERBANDINGAN HASIL AKURASI KNN DAN CART

Algoritma	Menggunakan *use training set*				Menggunakan 10-*cross fold validation*			
	Hasil akurasi	Precision	Recall	F-measure	Hasil akurasi	Precision	Recall	F-measure
KNN	99.2958 %	0.993	0.993	0,993	64.7887 %	0.638	0.648	0.642
CART	71.1268 %	0.506	0.711	0,591	71.1268%	0.506	0.711	0,591

d. Validasi dan Evaluasi

Untuk mengukur tingkat akurasi dari metode klasifikasi yang digunakan yaitu dengan *confusion matrix* yang disajikan pada tabel 12 dan tabel 13. Tabel 12 merupakan tabel hasil *confusion matrix* dari pengujian *dataset* menggunakan algoritma KNN.

TABEL XII
CONFUSION MATRIX

	Diterima	Ditolak
Diterima	41	0
Ditolak	1	100
142	21	121

Nilai 41 merupakan nilai data bentukan *rule* diterima beasiswa dengan bentuk data *testing* yang diterima. Nilai 0 didapatkan dari bentukan *rule* data yang ditolak dengan data *testing.* Nilai 1 merupakan *rule* diterima dengan data *testing* ditolak. Nilai 100 merupakan data *rule* yang ditolak dengan data *testing* ditolak.

Tabel 13 merupakan hasil *confusion matrix* dari pengujian *dataset* menggunakan algoritma CART.

TABEL XIII
CONFUSION MATRIX

	Diterima	Ditolak
Diterima	0	41
Ditolak	0	101
142	0	142

Nilai 0 didapatkan pada (1) *rule* diterima beasiswa dengan *testing* diterima; (2) *rule* yang.

Perbedaan KNN dan CART dianalisis dengan uji t dengan menggunakan *use training set.* Data tersebut dapat dilihat pada tabel 14.

TABEL XIV
STATISTIK DESKRIPTIF DATA PENELITIAN

		Mean	N
Pair 1	KNN	,97525	4
	CART	,62975	4

Berdasarkan tabel 14 dapat diketahui dari tiga data yang diamati diperoleh rata-rata algoritma KNN sebesar 97,525 persen dan rata-rata CART sebesar 62,975 persen. Hasil uji t menujukkan rerata selisih sebesar 0,345500. Rerata nilai KNN tercatat lebih besar dibandingan dengan rerata CART dengan nilai 34,55 persen.

TABEL XV
STATISTIK DESKRIPTIF DATA PENELITIAN

		Paired Differences Mean	t	df	Sig. (2-tailed)
Pair 1	KNN - CART	,345500	5,621	3	,011

Tahapan uji t tersebut adalah sebagai berikut :

1. Perumusan hipotesis statistik
   $H_0 : X_1 = X_2$ Tidak terdapat perbedaan antara nilai kinerja Algoritma KNN dan CART
   $H_1 : X_1 \neq X_2$terdapat perbedaan antara nilai kinerja Algoritma KNN dan CART
2. Penentuan tingkat keyakinan (95 persen)
3. H_0 diterima jika $-$ t tabel $\leq$ t hitung $\leq$ t tabel atau nilai sig $\geq \propto$ (0,05)
   H_0 ditolak jika t hitung $>$ t tabel atau t hitung $<$ - t tabel atau nilai sig $< \propto$ (0,05).
4. Hasil uji t
   Berdasarkan hasil uji maka H_0 ditolak dan H_1 diterima karena nilai signifikansi lebih kecil dari $\propto$ (0,042 < 0,05). Terdapat perbedaan nilai kinerja KNN dengan CART. Kinerja KNN lebih besar dibandingkan kinerja CART.

IV. KESIMPULAN DAN SARAN

Perhitungan dengan menggunakan dua algoritma yaitu KNN dan CART, serta melakukan percobaan dengan *use training set* dan dievaluasi dengan *confusion matrix* dan 10-*fold cross validation* mendapatkan nilai akurasi 99.2958 % dengan nilai *precision, recall* dan *F-measure* sebesar 0.993 pada algoritma KNN. Pada algoritma CART didapatkan nilai akurasi sebesar 71.1268% dengan nilai *precision* 0.506, *recall* 0.711, dan *F-measure* 0.591. Sehingga dapat diketahui bahwa akurasi algoritma KNN lebih tinggi dibanding algortima CART untuk memprediksi penerima beasiswa.

Adapun sarannya adalah dari analisis yang dilakukan penulis untuk memprediksi penerima beasiswa di STMIK AMIKOM Purwokerto maka hasil akurasi lebih tinggi pada algoritma KNN. Hasil ini bisa diterapkan kedalam sebuah sistem penunjang keputusan mengenai pemberian beasiswa di STMIK AMIKOM Purwokerto.

REFERENSI

[1] Breimen, L., Friedman, J.H., Olshen, R.A dan Stone, C.J., *Classification and Regression Trees(CART) Theory and Application*, Humboldt University, Berlin, 1984.

[2] Fiastantyo, Gian. "Perbandingan Kinerja Metode Klasifikasi Data Mining Menggunakan *Naive Bayes* Dan Algoritma C4.5 Untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa", 2014.

[3] Gorunescu, F. *Data Mining Concepts, Model and Techniques.* Verlag Berlin Heidelberg : Spinger, 2011.

[4] Han, J., & Kamber, M. Data Mining Concepts, Model and Techniques 2nd Edition. San Fransisco : Elsevier, 2006

[5] Hanik, Umi. Fuzzy Decision Tree dengan Algoritma C4.5 pada Data Diabetes Indian Pima. Tugas Akhir Periode Januari 2011. Institut Teknologi Sepuluh Nopember, 2011.

[6] Herlawati, dkk., *Penerapan Data Mining Dengan Matlab*. Bandung : Rekayasa Sains, 2013

[7] Hermawati, Fajar Astuti. *Data Mining*.Yogyakarta : Andi Offset, 2013

[8] Hssina, Badr, dkk. "A Comparative Study of Decision Tree ID3 and C4.5", 2014.

[9] Lesmana, I Putu Dody,. 2012. "Perbandingan Kinerja Decision Tree J48 dan ID3 dalam Pengklasifikasian Diagnosis Penyakit Diabetes Mellitus", Jurnal Teknologi dan Informatika, Vol. 2 No.2 Mei 2012.

[10] Lestari, Mei. 2014. Penerapan Algoritma Klasifikasi *Nearest Neighbor* (K-NN) Untuk Mendeteksi Penyakit Jantung. Faktor Exacta 7(4) :366-371, 2014 ISSN : 1979-276X.
[11] Ndaumanu, R.I, Kusrini, Arief, R. 2014. Analisis Prediksi Tingkat Pengunduran Diri Mahasiswa Dengan Metode *K-Nearest Neighbor*. Jatisi, Vol. 1 No.1 September 2014.
[12] Prasetyo, Eko.DATA MINING –Konsep dan Aplikasi Menggunakan Matlab.Yogyakarta : Andi Offset, 2012
[13] Setiyaji, Restu. "Perbandingan Algoritme C4.5 dengan algoritme RIPPER dalam Penentuan Beasiswa", 2016
[14] Sikki, Muhammad Ilyas. 2009. Pengenalan Wajah Menggunakan *K-Nearest Neighbor* Dengan Praproses Transformasi Wavelet. Jurnal Paradigma *Vol. X No.2 Desember 2009.*
[15] Sopiah, danSangadji, E.M., *Metodologi Penelitian-Pendekatan Praktis dalam Penelitian*. Yogyakarta :Andi Offset, 2010
[16] Susanto, S., dan Suryadi, D. *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta : Andi Offset, 2010
[17] Utomo, Hengky Setiawan. "Perbandingan Kinerja Algoritme C4.5 dan *Naive Bayes* dalam Mengklasifikasi Penyakit Diabetes", 2016
[18] WEKA, Machine Learning Group at University of Wakaito, diambil dari www.cs.waikato.ac.nz/ml/weka/, 25 Oktober 2016.
[19] Wicaksana, Paulus Dian. "Perbandingan Algoritma *K-Nearest Neighbor* dan *Naive Bayes* untuk Studi Data *Wincosin Diagnosis Breast Cancer*", 2015
[20] S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
[21] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
[22] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.
[23] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in *Proc. ECOC'00*, 2000, paper 11.3.4, p. 109.
[24] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.
[25] (2002) The IEEE website. [Online]. Available: http://www.ieee.org/
[26] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/