

SISTEM PENDUKUNG KEPUTUSAN (DSS) PENYELEKSIAN PEMILIHAN PENERIMA BEASISWA SMA N 1 ULUBELU TANGGAMUS MENGGUNAKAN DATA MINING

Eka Sugiyarti , Andino Maselena

Prodi Manajemen informatika STMIK Pringsewu Lampung

^{1,2}Jl. Wisma Rini No. 09 pringsewu Lampung

E-mail : ekasugiyarti14@yahoo.com , andimaselena@gmail.com

ABSTRAK

Education in the life of a country plays a very important role to ensure the survival of the state and the nation .. External factors affect the failure of students in completing the field of study, especially the field of study in the bench sma. . The problems faced in obtaining the scholarship are some of the criteria that have been set in this course such as majors, parent income, parental responsibility, academic achievement, non academic achievement Algorithm C4.5 is one method of data mining to predict the selection of recipients scholarships to students sma seen from external factors and internal students. Algorithm C4.5 .. The test results and analysis show that the Decision Tree C4.5 algorithm is accurately applied for the prediction of the final grades of high school students with an accuracy of 94.7368%.

Keywords: Data mining, Algorithm C4.5, Scholarship Reception

ABSTRAK

Pendidikan dalam kehidupan suatu negara memegang peranan yang sangat penting untuk menjamin kelangsungan hidup negara dan bangsa..Faktor eksternal berpengaruh pada kegagalan siswa dalam menyelesaikan bidang studi khususnya bidang studi dalam bangku sma. . Permasalahan yang dihadapi untuk mendapatkan beasiswa ada beberapa kriteria-kriteria yang sudah di tetapkan pada program studi ini seperti jurusan ,penghasilan orang tua ,tanggungan orang tua, prestasi akademik, ,prestasi non akademik Algoritma C4.5 merupakan salah satu metode data mining untuk memprediksi pemilihan penerima beasiswa kepada siswa sma dilihat dari faktor eksternal dan internal siswa. Algoritma C4.5.. Hasil pengujian dan analisis menunjukkan bahwa Algoritma Decision Tree C4.5 akurat diterapkan untuk prediksi nilai akhir siswa sekolah menengah dengan tingkat akurasi 94.7368%.

Kata Kunci : Data mining, Algoritma C4.5,Penerimaan Beasiswa

LPENDAHULUAN

1.1 Latar Belakang Masalah

Pada perkembangan teknologi saat ini berkembang sangat pesat dibidang teknologi maupun pendidikan ,banyak fasilitas –fasilitas yang kita dapatkan beasiswa merupakan program pengembangan agar siswa terus dapat melanjutkan studinya ,bantuan beasiswa bisa berupa biaya pendidikan ,bantuan belajar siswa dan suatu penghargaan untuk siswa yang berprestasi atau untuk siswa yang kurang mampudengan adanya sistem pendukung keputusan penyeleksian penerima beasiswa ini dengan data mining dapat mempermudah mendapatkan kriteria calon penerima beasiswa.

Menurut penelitian jurnal milik Kustanto (2016) sistem pendukung keputusan (DSS) penyeleksian penerima beasiswa berbasis web ini memakai data mining karena aplikasi *data mining* bisa dipakai dengan *database* berskala besar. Variabel-variabel yang diolah memiliki pembobotan yang dinamis .sehingga aplikasi dapat menentukan variabel yang menjadi prioritas dalam pengambilan derajat perkemendasian misalnya untuk beasiswa *peningkatan prestasi akademik*

(*ppa*),keputusan yang diambil lebih ditekankan kepada prestasi akademik dan untuk beasiswa bantuan belajar mahasiswa (BBM),lebih ditekankan pada kondisi ekonomi calon penerima beasiswa .[1]

Menurut penelitian jurnal milik Yoga Aditya Agassi (2014) Sistem Pendukung Keputusan (DSS) beasiswa merupakan salah satu program pengembangan yang didapatkan mahasiswa untuk terus melanjutkan studi ,baik berupa bantuan biaya, dengan berbagai kegiatan seminar maupun pelatihan .selai itu menunjang biaya pendidikan ,biaya siswa juga merupakan penghargaan untuk mahasiswa karena terdapat golongan beasiswa yang diperoleh atas prestasi yang diperoleh seseorang .[2]

Dari kedua penelitian tersebut peneliti ingin membedakan penelitian Sistem Pendukung Keputusan (DSS) penyeleksian penerima beasiswa salah satu program berupa bantuan biaya yang ditekankan pada ekonomi calon penerima lebih diutamakan kepada pihak yang kurang mampu agar siswa dapat terus melanjutkan studi penyeleksian penerima beasiswa ini akan menggunakan data

mining karena variabel-variabel yang diolah memiliki pembobotan yang dinamis.

Permasalahan yang dihadapi untuk mendapatkan beasiswa ada beberapa kriteria-kriteria yang sudah ditetapkan pada program studi ini seperti jurusan, penghasilan orang tua, tanggungan orang tua, prestasi akademik, prestasi non akademik dll. Oleh sebab itu tidak semua yang mendaftarkan diri sebagai calon penerima beasiswa dapat diterima, hanya yang memenuhi kriteria dan syarat tersebut yang dapat menerima beasiswa, maka perlu sistem pendukung keputusan untuk membantu penentuan siapa yang berhak menerima beasiswa tersebut.

Manfaat dan tujuan Sistem pendukung keputusan (DSS) menggunakan data mining bisa dipakai dengan database untuk mempermudah penyeleksian calon penerima beasiswa untuk kelayakan penerima beasiswa terhadap siswa yang berhak mendapatkan beasiswa terutama ditekankan kepada siswa yang kurang mampu dan siswa yang berprestasi, karena dengan aplikasi database berskala besar dapat membantu mengambil keputusan penerimaan beasiswa.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas dengan demikian ada beberapa kriteria dalam pendaftaran beasiswa

1. jurusan
2. prestasi akademik
3. tanggungan orang tua
4. penghasilan orang tua
5. prestasi non akademik

1.3 Tujuan dan Manfaat Penelitian

Tujuan untuk penelitian ini untuk mendapatkan Sistem Pendukung Keputusan (DSS) untuk kelayakan penerima beasiswa terhadap siswa yang berhak untuk menerima beasiswa.

II TINJAUAN PUSTAKA

2.1 Definisi Sistem Pendukung Keputusan

Menurut Raymond McLeod, Jr (1998) Sistem Pendukung Keputusan (DSS) merupakan sebuah sistem yang menyediakan kemampuan untuk penyelesaian masalah dan komunikasi untuk permasalahan yang bersifat semi-terstruktur. [3]

Menurut Bonczek (1980) Sistem Pendukung Keputusan (DSS) sebagai sebuah Sistem berbasis Komputer yang terdiri atas komponen-komponen antara lain komponen sistem bahasa (language), komponen sistem pengetahuan (knowledge) dan komponen sistem pemrosesan masalah (problem processing) yang saling berinteraksi satu sama lain. [3]

2.2 Definisi Sekolah

Untuk mengetahui definisi pendidikan dalam sudut pandang kebijakan, kita telah mempunyai

rumusan formal dan operasional, sebagaimana tercantum dalam UU No 20 tahun 2003 tentang SISDIKNAS yaitu :

Pendidikan merupakan usaha dengan sengaja dan terencana untuk menciptakan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk mempunyai kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia serta ketrampilan yang diperlukan dirinya, masyarakat bangsa dan negara. [4]

2.3 Definisi Beasiswa

Pengertian Beasiswa adalah pemberian berupa bantuan keuangan yang diberikan kepada perorangan yang bertujuan untuk digunakan demi keberlangsungan pendidikan yang ditempuh, beasiswa dapat diberikan oleh lembaga pemerintah, perusahaan ataupun yayasan. [5]

2.4 Pengertian Data mining

Menurut urban dalam bukunya yang berjudul "Decisions support systems and intelligent systems" data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam basis data, data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari basis data besar [6].

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database [7]

2.5 Metode Pelatihan

Secara garis besar metode pelatihan dalam teknik-teknik *data mining* dibedakan dalam dua pendekatan yaitu:

1. *Unsupervised learning*, metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*) guru disini adalah label dari data
2. *Supervised learning*, yaitu metode belajar dengan adanya latihan dan pelatihan. dalam pendekatan ini untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi, digunakan beberapa contoh data yang mempunyai output atau label selama proses *training*. [6]

III. Metode penelitian

3.1. Tahapan pengumpulan data

3.1. Observasi

Pengumpulan data ini dikatakan secara langsung melalui pengamatan dilapangan, dalam penelitian ini dilakukan pada penerimaan beasiswa beserta pengetahuan kebutuhan mengenai beasiswa, oleh karena ini dilakukan

wawancara kepada pengurus beasiswa setelah melakukan observasi dapat dibuat kerangka konsep berdasarkan data- data dan informasi yang diperoleh .

3.2 Interview (wawancara)

Yaitu pengumpulan data dengan melakukan wawancara langsung dengan para guru, staf atau murid untuk mendapatkan bahan masukan untuk penulis penelitian ini .sehingga penulis dapat mengetahui informasi –informasi tentang SMA N 1 ulubelu sehingga dapat membantu penulis untuk mendapatkan solusi dari iterview (wawancara) tersebut.

3.3 Studi pustaka

study pustaka ini adalah suatu penelitian yang dilakukan dengan mencari beberapa sumber-sumber atau referensi baik dengan buku,artikel atau dari sumber lain untuk referensi perancangan sistem dan pembuatan laporan yang akan disusun.[]

3.4 Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan ,pohon keputusan merupakan metodeklafikasi dan prediksi yang sangat terkuat dan terkenal metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang merepresentasikan aturan ,aturan dapat dengan mudah dipahami dengan bahasa alami .dan mereka juga dapat diekspresikan aturan .aturan dapat dengan mudah dipahami dengan bahasa alami ,dan mereka juga dapat diekspresikan dalam bentuk bahasa basis data seperti *Structured Query language* untuk mencari record pada kategori tertentu.

Ada beberapa tahap untukmembuat sebuah pohon keputusan dengan Algoritma C4.5

- 1.menyiapkan data training ,data training biasa diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokan sebelumnya ke kelas-kelas tertentu.
- 2.menentukan akar di pohon akar akan diambil dari atribut yang dipilih, dengan cara menghitung nilai gain dari,masing – masinatribut ,nilai gain yang paling tinggi yang akan menjadi akar pertama ,sebelum menghitung nilai gain dari atribut ,hitung dulu nilai entropy digunakan rumus

$$entropy (S) = \sum_{i=1}^n - p_i \cdot \log_2 p_i$$

Keterangan

S = himpunan khusus

n = jumlah partisi S

pi = proporsi Si terhadap S

3.kemudian hitung nilai gain menggunakan rumus :

Gain (S.A)

= Entropy (s)

$$\sum_{i=1}^n \frac{|S_i|}{|S|}$$

*Entropy (Si)

Keterangan:

S = himpunan kasus

A = fitur

N = jumlah partisi atribut A

|Si| = proposi Si terhadap

|S| = jumlah kasus dalam S

4. ulangi langkah ke-2 hingga semua record terpartisi
5. proses partisi pohon keputusan akan berhenti saat:
 - a.semua record dalam simpul N mendapat kelas yang sama
 - b.tidak ada atribut didalam record yang dipartisi lagi
 - c.tidak ada record didalam cabang yang yang kosong [8]

3.5 Confusion matrix

Confusion matrix adalah suatu metode yang biasanya digunakan sebagai perhitungan akurasi pada kosep data mining .informasi dalam confusion matrix diperlukan untuk menentukan kinerja model klasifikasi ,ringkasan informasi ini ke dalam sebuah nilai digunakan untukmembandingkan kinerja dari model-model yang berbeda hal ini dapat dilakukan dengan menggunakan performace metric

classificati on	Predicted class	
	Class=Yes	Class=No
Class=Yes	a(true positif TP)	B(false negatif TN)
1.4 Clas s=N o	1.5 c(fal se posi tif- FP)	1.6 d(true negati f-TN)

1. Akurasi

Perhitungan akurasi ini dengan cara membagi jumlah data dan klasifikasi secara benar dengan total sample data testing yang diuji .

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

2.Precision

Membagi nilai data yang benar yang bernilai positif (*true positif*) dibagi dengan jumlah data benar yang bernilai positif (*true positif*) dan data salah yang bernilai positif (*false negative*)

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

3. Recall

Dihitung dengan cara membagi data benar yang bernilai positif (*true positif*) dengan hasil penjumlahan dari data benar yang bernilai positif (*true positif*) dan data salah yang bernilai negatif (*false negative*)

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

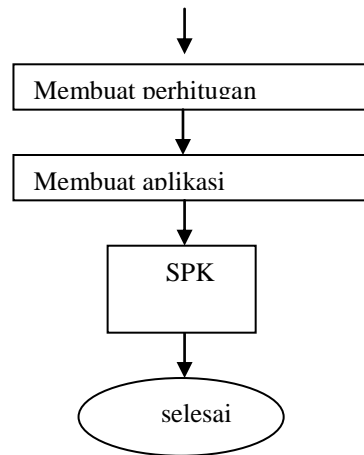
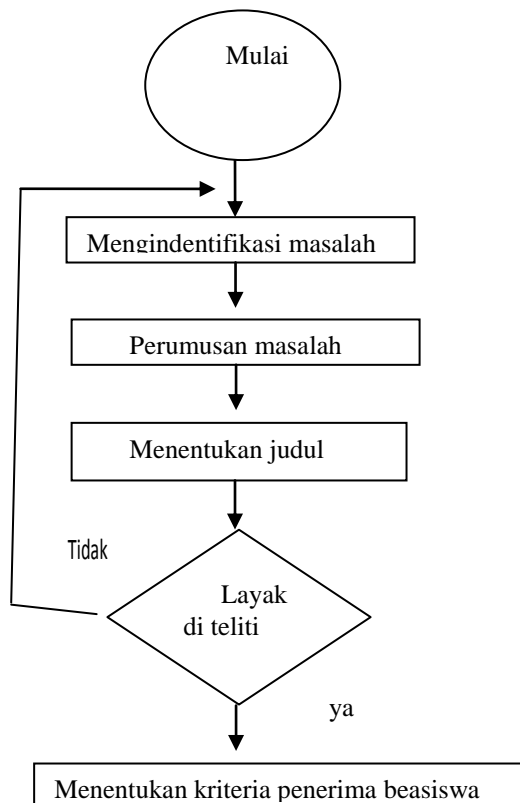
4. f-Measure

Nilai ini didapat dari perhitungan pembagian hasil dari perkalian *precision* dan *recall* dengan hasil penjumlahan *precision* dan *recall* kemudian dikalikan dua

$$f - Measure = 2 * \frac{precision * recall}{precision + recall}$$

3.6 Konsep Pemikiran Penelitian

dalam perancangan sistem pendukung keputusan ini melalui beberapa tahapan dan proses agar nantinya akan mendapatkan hasil dengan baik dan mencapai tujuan yang diinginkan. beberapa tahapan penelitian ini dapat digambarkan dalam diagram alir/flowchart yang dapat dilihat dibawah ini.



Keterangan

- 1.pengguna pertama kali harus mengidentifikasi masalah ,kemudian merumuskan masalah selanjutnya menentukan judul.
- 2.setelah menentukan judul kemudian memutuskan apakah data sudah lengkap layak untuk diteliti atau tidak,jika tidak akan dikembalikan ke indentifikasi masalah,jika ya maka sistem akan dilanjutkan untuk menentukan kriteria penerimaan calon mahasiswa.
- 3.setelah menentukan kriteria penerima baesiswa selanjutnya membuat perhitungan
- 4.setelah mendapatkan inputan data atau nilai selanjutnya membuat aplikasi
- 5.selanjutnya akan diambil keputusan untuk peserta penerimaan beasiswa .

IV. HASIL DAN PEMBAHASAN

4.1 Parameter yang diukur

Parameter yang diukur untuk pada data siswa sebagai atribut untuk klafikasi yaitu :

Tabel 1. Data jurusan.

No	jurusan
1	IPA
2	IPS

Berikut tabel data jurusan yang diukur dari data jurusan dengan pembobotan nilainya sebagai berikut

Tabel 2. Penghasilan orang tua

Golo ngan	Pendapatan	kategori
Gol 1	>=4000000	Mampu
Gol 2	2100000-3900000	Menengah
Gol 3	<=2000000	Bawah

Berikut tabel penghasilan orang tua yang diukur dari tabel penghasilan orang tua dengan pembobotan nilainya sebagai berikut

Tabel 3. Jumlah tanggungan orang tua

Golongan	tanggungan	kategori
Gol 1	≤ 2500000	Baik
Gol 2	> 2500000	Buruk

Berikut tabel jumlah tanggungan orang tua yang diukur dari jumlah tanggungan dengan pembobotan nilainya sebagai berikut

Tabel 4. Prestasi Akademik

Prestasi akademik	kategori
Juara 1,2,3	bagus
Juara 4,5,6,7	sedang
Juara > 7	kurang

Berikut tabel data jurusan yang diukur dari prestasi akademik dengan pembobotan nilainya sebagai berikut

Tabel 5. Prestasi Non Akademik

Prestasi non akademik	kategori
Tidak Mendapat Prestasi Non Akademik	Buruk
Level Kecamatan	cukup
Level Kabupaten	Sedang
Level Provinsi	Baik
Level Nasional	Sangat baik

Berikut tabel prestasi non akademik yang diukur dari prestasi non akademik dengan pembobotan nilainya sebagai berikut

Perhitungan algoritma C4.5 awal

dapat beasiswa tidak dapat beasiswa (tidak dapat beasiswa)

jurusan ipa	5	5	0.5	0.5	0.5
jurusan ips	6	3	0.66666667	0.33333333	0.389975
penghasilan mampu	1	3	0.25	0.75	0.5
penghasilan menengah	5	3	0.625	0.375	0.42379494
penghasilan bawah	5	2	0.71428571	0.28571429	0.34673345
tanggungan baik	5	3	0.625	0.375	0.42379494
tanggungan buruk	6	5	0.54545455	0.45454546	0.47698316
prestasi akad bagus	6	4	0.6	0.4	0.44217936
prestasi akad sedang	4	0	1	0	0

Dengan gambar perhitungan awal untuk pengumpulan data untuk kelayakan penerimaan beasiswa

Perhitungan algoritma C 4.5

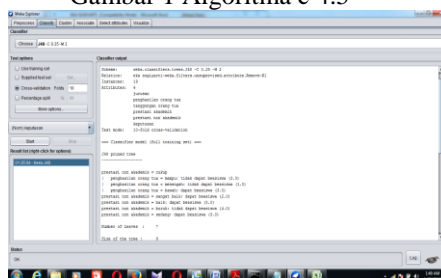
Totalp(juru	E(total*p(jurus	GAIN
0.5263158	0.961298027	0.020643
0.4349822		
0.1707954	0.890654132	0.091287
0.4018669		
0.3179918		
0.4018669	0.977358124	0.004583
0.5754912		
0.5110266	0.701007706	0.280933
0		
0.1899811		
0	0.362978787	0.618962
0		

Dengan perhitungan algoritma dapat menentukan gain Dengan gambar algoritma c 4.5 perhitungan gain ,untuk menghitung gain dan mendapatkan hasil gain

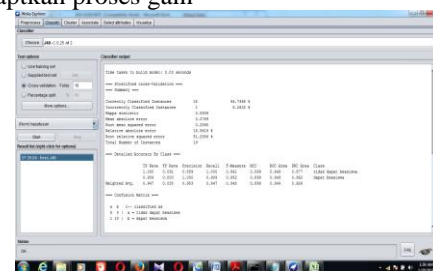
- Gain jurusan = 0,020643
- Gain penghasilan orang tua =0,091287
- Gain jumlah tanggungan orang tua=0,004583
- Gain prestasi akademik=0,280933
- Gain prestasi non akademik =0,618962

Dari proses perhitungan penentuan akar pohon keputusan di atas, maka dapat di gambarkan sebuah pohon keputusan sebagai berikut :

Gambar 1 Algoritma c 4.5

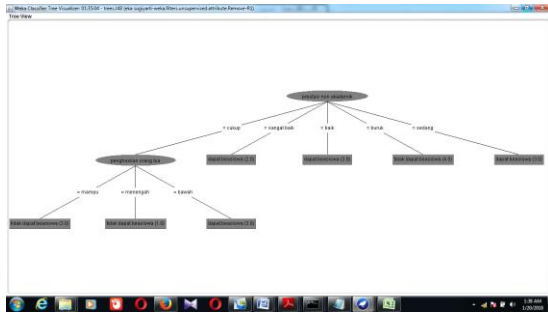


Dari hasil perhitungan gambar 1 akan mendapatkan proses gain



Gambar 2. Algoritma C 4.5

Dari hasil perhitungan gain akan mendapatkan gain untuk mendapatkan pohon keputusan



Gambar 3. Model Pohon Keputusan

Dari hasil perhitungan gain akan mendapatkan pohon keputusan . Dengan model pohon keputusan didapatkan hasil akhir dari perhitungan gain yang menentukan beasiswa yg dominan yaitu prestasi akademik

J48 pruned tree

-
- prestasi non akademik = cukup
penghasilan orang tua = mampu: tidak dapat beasiswa (3.0)
 - prestasi non akademik = cukup
penghasilan orang tua = menengah: tidak dapat beasiswa (1.0)
 - prestasi non akademik = cukup
penghasilan orang tua = bawah: dapat beasiswa (3.0)
 - prestasi non akademik = sangat baik: dapat beasiswa (2.0)
 - prestasi non akademik = baik: dapat beasiswa (3.0)
 - prestasi non akademik = buruk: tidak dapat beasiswa (4.0)
 - prestasi non akademik = sedang: dapat beasiswa (3.0).

Number of Leaves : 7

Size of the tree : 9.

Dari hasil perhitungan untuk pemilihan beasiswa untuk siswa dengan menggunakan algoritma C 4.5 dapat di simpulkan sebagai berikut:

1. Jika Prestasi Non Akademik = cukup Dan Pendapatan Orang Tua = Mampu Maka Tidak Dapat Beasiswa (3)
2. Jika Prestasi Non Akademik = cukup Dan Pendapatan Orang Tua = Menengah Maka Tidak Dapat Beasiswa (1)
3. Jika Prestasi Non Akademik = cukup Dan Pendapatan Orang Tua = Bawah Maka Dapat Beasiswa (3)
4. Jika Prestasi Non Akademik = Sangat Baik

Maka Dapat Beasiswa (2)

5. Jika prestasi Non Akademik = Baik Maka Dapat Beasiswa (3)
6. Jika prestasi Non Akademik = Buruk Maka Tidak Dapat Beasiswa (4)
7. Jika prestasi Non Akademik = Sedang Maka Dapat Beasiswa (3)

V. Kesimpulan Dan Saran

5.1 Kesimpulan

Kesimpulan yang dapat diambil dari jurnal ini adalah dapat dihasilkan sistem pendukung keputusan (DSS) untuk kelayakan penerimaan beasiswa terhadap siswa yang berhak mendapatkan beasiswa sistem pendukung keputusan (DSS) dengan data mining metode algoritma C45 dapat membantu mengefisienkan dan mengefektifkan yang lebih baik dari sebelumnya . serta dalam pendapatan laporan yang didapat dapat diminimalkan dalam kesalahannya.

5.2 Saran

Berdasarkan proses pengujian dan kesimpulan yang telah dilakukan, maka ada beberapa saran dalam penelitian ini yaitu:

- a. Jumlah data yang digunakan hanya 19 dengan jumlah atribut 9, sehingga untuk hasil pengukuran yang lebih baik lagi disarankan untuk menambah jumlah data yang lebih besar dan jumlah atribut yang lebih banyak.
- b. Gunakan metode optimasi lain seperti, Genetic Algorithm (GA), Ant Colony Optimization dan lainnya.

Aspek Sistem.

- a. Meningkatkan sistem analisa beasiswa untuk penentuan kelayakan pemberian beasiswa bagi calon penerima.
- b. Mengukur apakah model yang telah dikembangkan berhasil atau tidak, evaluasi digunakan untuk mengukur keakuratan hasil yang dicapai oleh model.

DAFTAR PUSTAKA

- [1] Shella,pradega.(2015). Sistem Penduung Keputusan Dengan menggunakan *decisson tree* dalam perberian beasiswa di sekolah menengah pertamaK (Studi Kasus di SMP N 2 Rembang). Teknik Informatika dan Komputer FT. UNNES.Semarang
- [2] Aras, Zainul., Sarjono. (2016).” Analisis Data Mining untuk menentukan kelompok penerima bantuan bedah rumah menggunakan *K-means*, Vol 1 No.2 ISSN 2540-8011.

- [3] Iriadi, nandang dan Nuraeni nia. (2016). Kajian Penerapan Metode Data Mining Algoritma c4.5 untuk prediksi kelayakan kredit pada bank mayapa jakarta .Jurnal Teknik Komputer AMIK BSI. Vol II. No.1. ISSN 2442-2436
- [4] Y. C. G. Tomy Salim, "Data Mining Identifikasi Website Phising Menggunakan Algoritma C4.5," *J. TAM (Technol. Accept. Model)*, vol. 8, no. 2, pp. 130–135, 2017.
- [5] Haqmanullah, Rizky. Pambudi1. Setiawan Budi Darma dan Indriati. (2017). Penerapan Algoritma C4.5 Dalam Program Untuk Memprediksi Kinerja Siswa Sekolah Menengah .Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN: 2548-964X Vol. 2, No. 7. Universitas Brawijaya

