

KLASIFIKASI PENERIMAAN BEASISWA MENGGUNAKAN ALGORITMA C 4.5 (STUDI KASUS SMA PGRI CLURING KABUPATEN BANYUWANGI)

Bayu Maeky Nugroho, Triawan Adi Cahyanto M.Kom
Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah
Jember

maekybayu43@gmail.com, triawanac@unmuhjember.ac.id

ABSTRAK

SMA PGRI Cluring kabupaten Banyuwangi merupakan salah satu yang mendapat Progam Indonesia Pintar (PIP). Sekolah memberikan prioritas penerimaan beasiswa berdasarkan parameter yang telah ditentukan, yaitu penghasilan orang tua, penerima Kartu Penjamin Sosial (KPS), Layak Progam Indonesia Pintar (PIP), jumlah saudara kandung dan Penerima Kartu Indonesia Pintar (KIP). Penelitian ini bertujuan untuk membuat aplikasi data mining klasifikasi penerimaan beasiswa SMA PGRI Cluring Banyuwangi. Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Dari penelitian didapatkan hasil yaitu dalam implementasi algoritma C4.5 terhadap klasifikasi penerima KIP siswa SMA PGRI Cluring Banyuwangi dengan 163 data memiliki tingkat akurasi sebesar sebesar 76.74 % dan tingkat presisi sebesar 78.94 %. Dengan jumlah data 163, model yang dihasilkan belum cukup bagus atau metode ini belum bisa digunakan untuk mendukung kinerja pihak sekolah.

Kata kunci : Klasifikasi, siswa, beasiswa, Algoritma C4.5

1. Latar Belakang

Beasiswa adalah pemberian berupa bantuan keuangan yang diberikan kepada perorangan yang digunakan demi keberlangsungan pendidikan yang ditempuh. Beasiswa sangat membantu bagi siswa yang kurang mampu namun mempunyai prestasi tertentu. Setiap lembaga pendidikan khususnya sekolah pada umumnya

memiliki suatu program pendidikan, yaitu pemberian beasiswa kepada siswa yang berprestasi, maupun kepada siswa yang tidak mampu. Salah satu program beasiswa yaitu PIP. PIP adalah Progam Indonesia Pintar, pemberian bantuan tunai pendidikan kepada siswa yang berasal dari keluarga miskin.

SMA PGRI Cluring kabupaten Banyuwangi merupakan salah satu yang mendapat Progam Indonesia Pintar (PIP). Algoritma C 4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode pohon keputusan mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan.

2. Kajian Pustaka

a. Data mining

Data Mining merupakan sebuah analisis dari observasi data dalam jumlah besar untuk menemukan hubungan yang tidak diketahui sebelumnya dan metode baru

untuk meringkas data agar mudah dipahami serta kegunaannya untuk pemilik data, sebagai satu set teknik yang digunakan secara otomatis untuk mengeksplorasi secara menyeluruh dan membawa ke permukaan relasi-relasi yang kompleks pada set data yang sangat besar.

b. Algoritma decision tree

Decision tree adalah struktur flowchart yang menyerupai tree (pohon). Setiap simpul internal menandakan suatu tes pada atribut. Setiap cabang merepresentasikan hasil tes, dan simpul daun merepresentasikan kelas atau distribusi kelas. Alur pada decision tree di telusuri dari simpul akar ke simpul daun yang memegang prediksi kelas. Node yang paling atas dari decision tree disebut sebagai root. Decision Tree merupakan metode klasifikasi yang paling populer digunakan.

c. Algoritma C4.5

Algoritma C4.5 merupakan pengembangan dari ID3, beberapa pengembangan yang dilakukan pada C4.5 antara lain bisa mengatasi missing value, bisa mengatasi continue data, dan pruning. Proses pada pohon keputusan adalah mengubah bentuk data(tabel) menjadi model pohon, mengubah model pohon menjadi rule, dan menyederhanakan rule.

Persamaan :

$$Entropy(S) = \sum_{i=1}^k \frac{|S_i|}{|S|} * Entropy(S_i)$$

$$Gain = Entropy(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} * Entropy(S_i)$$

3. Metode Penelitian

1. Pengumpulan Data

Pada tahap ini ditentukan data yang akan diproses. Data yang akan diproses yaitu data peserta didik SMA PGRI Cluring Banyuwangi sebanyak 20 data. Mencari data yang tersedia, memperoleh data tambahan yang dibutuhkan, mengintegrasikan semua data ke dalam data set, termasuk variabel yang diperlukan dalam proses.

2. Pengolahan Data Awal (Data Pre-processing)

Di tahap ini peneliti melakukan penyelesaian data, data dibersihkan dan ditransformasikan ke bentuk yang diinginkan sehingga dapat dilakukan persiapan dalam pembuatan model.

3. Model/Metode yang Diusulkan (Proposed Model/Method)

Pada tahap ini data peneliti menganalisis data, mengelompokkan variabel mana yang berhubungan dengan satu sama lainnya. Setelah data dianalisis lalu menerapkan model-model yang sesuai dengan jenis data.

Pembagian data ke dalam data latihan (training data) dan data uji (testing data) juga diperlukan untuk pembuatan model.

4. Eksperimen dan Pengujian Metode (Method Test and Experiment) Pada tahap ini peneliti menguji metode untuk melihat hasil berupa rule yang akan dimanfaatkan dalam pengambilan keputusan.

5. Evaluasi dan Validasi Hasil (Result Evaluation and Validation)

Pada tahap ini peneliti melakukan evaluasi terhadap metode yang diterapkan untuk mengetahui tingkat keakuratan metode.

4. Implementasi

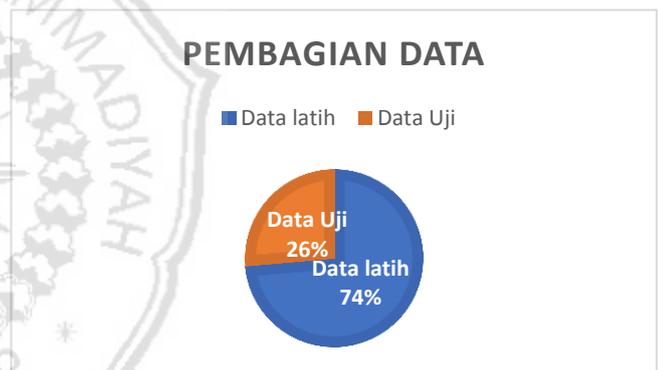
a. Gambaran data

Jumlah siswa yang terdaftar dalam dataset berjumlah 163 orang. Pada dataset ini dibagi menjadi data latihan dan data uji, data latihan berjumlah 120 dan data uji berjumlah 43. Dari 163 data jumlah penerima beasiswa 63 dan tidak mendapatkan beasiswa 100. Setelah dipecah menjadi data latihan dan data uji peta rasio menjadi, data latihan 120 dengan penerima beasiswa berjumlah 41 dan tidak menerima beasiswa 76. Pada data uji berjumlah 43, dengan penerima beasiswa 21 dan tidak menerima beasiswa 22.

b. Preprocessing

No.	Data sebelumnya	Data sesudah preprocessing
1	<ul style="list-style-type: none"> - Tidak berpenghasilan - Kurang dari 500.000 - 500.000 – 999.999 - Kurang dari 1 juta - 1 juta – 1.999.999 - 2.000.000 – 4.999.999 	<ul style="list-style-type: none"> - Tidak berpenghasilan - Kurang dari 500.000 - Menjadi rendah - 500.000 – 999.999 - Kurang dari 1 juta - Menjadi sedang - 1 juta – 1.999.999 - 2.000.000 – 4.999.999 - Menjadi tinggi <p>Jadi parameter penghasilan orang tua rendah, sedang dan tinggi</p>

c. Pembagian data

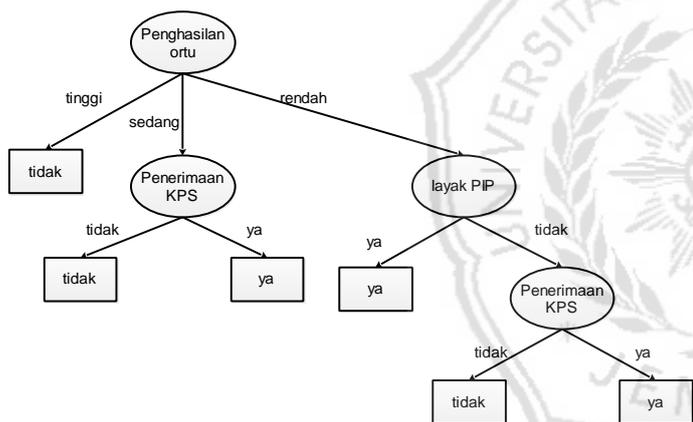


d. Pembuatan pohon keputusan

atribut	node	jumlah	Ya	Tidak	entropi	gain
total		120	49	71	0.97561	7
Penghasilan Ortu		120				0.1637878551742
	rendah	25	21	4	0.63431	
	sedang	92	28	64	0.88654	1
	tinggi	3	0	3	0	
Penerimaan KPS		120				0.1225084933273
	Ya	25	20	5	0.72192	8
	Tidak	95	29	66	0.88763	

layak PIP (usulan dari sekolah)		120				0.0068095897162
	Ya	51	18	33	0.936667	
	Tidak	69	31	38	0.992563	
Jumlah Saudara Kandung		120				0.0096470085435
	1	101	43	58	0.98403	
	2	18	6	12	0.91829	
	3	1	0	1	0	

e. Pohon keputusan



f. Syntax pohon keputusan

Id	Aturan
1	IF penghasilan orangtua = tinggi THEN keputusan = tidak
2	IF penghasilan orangtua = sedang AND penerimaan KPS = tidak THEN keputusan = tidak
3	IF penghasilan orangtua = sedang AND penerimaan KPS = ya THEN keputusan = ya
4	IF penghasilan orangtua = rendah AND layak PIP = ya THEN keputusan = ya
5	IF penghasilan orangtua = rendah AND layak PIP = tidak AND penerimaan KPS = tidak THEN keputusan = tidak
6	IF penghasilan orangtua = rendah AND layak PIP = tidak AND penerimaan KPS = ya THEN keputusan = ya

5. Kesimpulan

a. Hasil

Dari penelitian ini didapatkan beberapa penemuan terhadap hasil dari implementasi algoritma C4.5 dalam mengklasifikasi pemberian KIP terhadap siswa SMA PGRI Cluring Banyuwangi dengan jumlah data latih 120 dan data uji 43 siswa. Beberapa temuan itu antara lain:

1. Algoritma ini dapat diaplikasikan terhadap data siswa dalam mengklasifikasi penerima beasiswa pada siswa SMA PGRI Cluring Banyuwangi.
2. Confusion matrix menunjukkan hasil, tingkat akurasi sebesar 76.74 % dan tingkat presisi sebesar 78.94 %.

b. Saran

Penulis menyadari bahwa penelitian ini jauh dari kesempurnaan. Perhatian dari pembaca yang ingin mengembangkan penelitian sehingga mendekati kesempurnaan akan penulis dukung. Untuk itu beberapa poin dapat penulis kemukakan untuk dapat diperbaiki, antara lain:

1. Dataset yang terlalu sedikit, untuk selanjutnya ditambah agar mendapatkan pola yang lebih baik dan akurat.
2. Pengembang dapat menggunakan data baru, mengingat penelitian ini belum terintegrasi dengan data baru.

Daftar Pustaka

Anam, K., Widharna, R. M., & Kusriani, D. (2009). α -glucosidase inhibitor activity of Terminalia species. *Int J Pharmacol*, 5(4), 277-280.

Basuki, A., & Syarif, I. (2003). Decision Tree. *Surabaya: Politeknik Elektronika Negeri Surabaya ITS*.

Busyrah, Hamidah. 2012. "Analisis Struktural Model Aktansial dan Fungsional Greimas Pada Sepuluh Cerkak Dalam Antologi Geguritan Lan Cerkak Pisungsung". Skripsi. Universitas Indonesia, Depok.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37.

Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining (adaptive computation and machine learning)* (pp. 361-452). Cambridge, MA: MIT Press.

Kusnawi. 2007. Pengantar Solusi Data Mining. Seminar Nasional Teknologi 2007 (SNT). Yogyakarta: STMIK AMIKOM Yogyakarta.

Larose, D. T. (2005). An introduction to data

mining. *Discovering Knowledge in Data, John Wiley & Sons Publication, Hoboken, New Jersey, USA*, 1-25.

Murniasih, E. (2009). *Buku Pintar Beasiswa*. GagasMedia.

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106.

Suyanto, M. (2007). *Marketing Strategy Top Brand Indonesia*. Penerbit Andi.

