PENERAPAN ALGORITMA C4.5 UNTUK PENENTUAN PENERIMA BEASISWA (Studi Kasus: SDN Karawang Kulon III)

[1] Jajam Haerul Jaman, [2]Novia Indriyani Puji Astuti
[1]jajam.haeruljaman@unsika.ac.id, [2]indriyaninovia27@gmail.com
Program Studi Teknik Informatika
Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang

ABSTRAK
Pendidikan merupakan hal yang sangat penting untuk meningkatkan kualitas seseorang. Memperoleh pendidikan tentunya tidak terlepas dari masalah pendidikan yang kerap dihadapi oleh seseorang yaitu, biaya pendidikan. Maka dari itu pemerintah mengadakan program beasiswa yaitu beasiswa PIP dimaksudkan untuk siswa kurang mampu dan seleksi pemberian beasiswa PIP dilakukan oleh pihak sekolah. penelitian ini memanfaatkan teknik data mining dengan Algoritma C4.5 sebagai prediksi penerima beasiswa di SDN Karawang Kulon III dan kriteria yang digunakan: Pekerjaan orang tua, jumlah penghasilan orangtua, jumlah tanggungan orang tua, jarak tempuh, Penerima KIP, Penerima KPS, nilai rata-rata rapot. Penelitian tanpa fitur seleksi dan menggunakan fitur seleksi Information Gain. Kemudian diaplikasikan menggunakan tools WEKA 3.9.2 dengan Algoritma C4.5. Performa algoritma dilihat dari acuracy, precision, dan error rate, sehingga memberikan nilai performa yang berbeda-beda pada setiap hasilnya. Hasil acuracy, tertinggi precision, dan error rate terendah terdapat pada skenario I dengan 7 kriteria, yaitu nilai acuracy sebesar 93,58%, nilai precision 95,5%, dan nilai error rate terendah 6,14%.

Kata Kunci: Beasiswa, *Data mining*, Algoritma C4.5, *information gain, CRISP-DM.*

1. PENDAHULUAN

Pendidikan merupakan hal yang sangat penting untuk meningkatkan kualitas seseorang. Memperoleh pendidikan tentunya tidak terlepas dari masalah pendidikan yang kerap dihadapi oleh seseorang yaitu, biaya pendidikan. Biaya pendidikan merupakan salah satu faktor yang penting untuk menentukan keberhasilan tujuan pendidikan.

Usaha pemerintah untuk meningkatkan kualitas lulusan siswa salah satunya melalui beasiswa Program Indonesia Pintar (PIP) yang merupakan bantuan berupa uang tunai dari pemerintah yang diberikan kepada peserta didik yang orang tuanya kurang mampu membiayai pendidikannya. Beasiswa PIP dimaksudkan untuk siswa kurang mampu dan kebijakan baru bahwa pada tahun ajaran 2017/2018 seleksi pemberian beasiswa PIP dilakukan oleh pihak sekolah. Hal itu tentu menjadi tugas bagi pihak sekolah untuk melakukan penyeleksian penerima beasiswa agar beasiswa yang diberikan tepat sasaran.

Penelitian ini di lakukan SDN Karawang Kulon III yang sekarang ini mempunyai siswa sebanyak 732 orang yang terbagi dalam 6 kelas. Penelitian ini bertujuam untuk memprediksi penerima beasiswa dengan memanfaatkan teknik *data mining*, menggunakan algoritma C4.5, sehingga membantu pihak sekolah dalam mencari solusi dan dapat mengetahui tingkat prestasi siswa juga meningkatkan lagi mutu pendidikan sekolah dengan adanya siswa-siswa yang berprestasi.

2. PENELITIAN SEBELUMNYA

Klasifikasi rekomendasi beasiswa dilakukan untuk mengklasifikasi apakah siswa akan mendapatkan rekomendasi beasiswa sesuai dengan bobot yang akan di nilai. Klasifikasi dilakukan menggunakan data mining algoritma C4.5. Data yang digunakan yaitu data jurusan, kelas, jumlah nilai, penghasilan orangtua, dan jumlah saudara kandung. Proses data mining pada data training akan menghasilkan pohon keputusan atau rule. Metode evaluasi yang dilakukan dalam penelitian ini yaitu menggunakan confusion matrix dan nilai akurasi, untuk sekali pengujian tingkat akurasi yang dihasilkan yaitu 77%. hal ini membuktikan bahwa algoritma C4.5 cukup akurat dalam menentukan rekomendasi beasiswa pada SMA Muhammadiyah Gubug.

ISSN: 2503-054X
Vol. 3 No: 1, April 2018

3. METODOLOGI PENELITIAN

Pada penelitian ini dalam penentuan penerima beasiswa menggunakan algoritma C4.5 menggunkan metode *CRIPS-DM*.

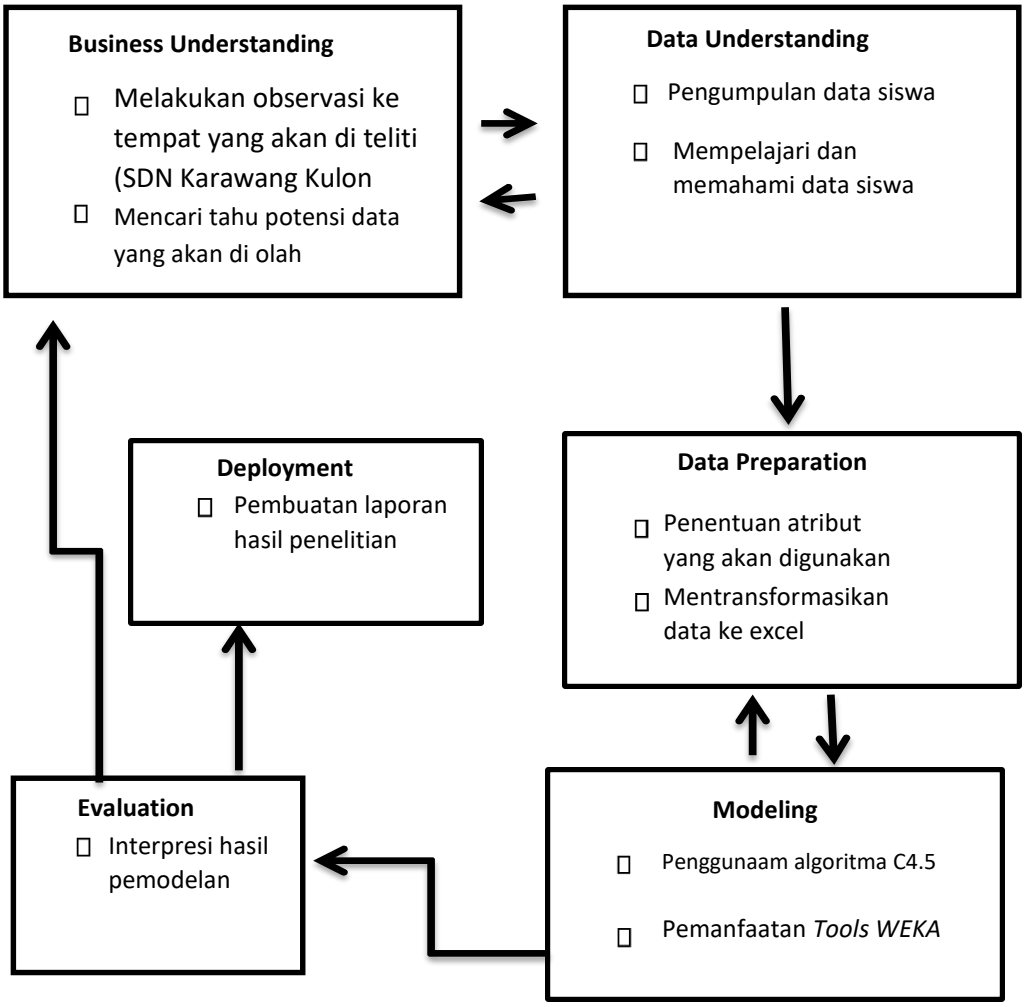
3.1. Dataset

Penelitian untuk penentuan penerima beasiswa di SDN Karawang Kulon III ditinjau berdasarkan Pekerjaan orang tua, jumlah penghasilan orangtua, jumlah tanggungan orangtua, jarak tempuh, Penerima KIP, Penerima KPS, nilai rata-rata rapot. Kriteria yang dipilih sesuai dengan kombinasi dari berbagai jurnal dan informasi dari pihak Sekolah. Pada tahun ajaran 2017/2018 tercatat jumlah siswa 732 siswa dengan jumlah penerima beasiswa 142 siswa.

3.2. Metode

Pada pemodelan *CRIPS-DM*, berikut tahapan-tahapannya:

1. *Business Understanding* (Pemahaman Bisnis)
2. *Data Understanding* (Pemahaman Data)
3. *Data Perparation* (Persiapan Data)
4. *Modeling* (Pemodelan)
5. *Evaluation* (Evaluasi)
6. *Deployment* (Penyebaran)



ISSN: 2503-054X
Vol. 3 No: 1, April 2018

3.3. Seleksi Atribut

Dalam menyelesaikan kasus prediksi penerima beasiswa ini terdapat beberapa data yang diubah dan ditransformasikan ke dalam bentuk lain:

1. Jumlah Penghasilan Orangtua Transformasi dapat dilihat pada gambar 4.2

Wiraswasta	Rp. 500,000 - Rp. 999,999
Buruh	0 - Rp. 500,000
Karyawan Swasta	Rp. 1,000,000 - Rp. 1,999,999
Buruh	Rp. 500,000 - Rp. 999,999
Buruh	Rp. 500,000 - Rp. 999,999
PNS/TNI/Polri	Rp. 2,000,000 - Rp. 4,999,999

menjadi

Wiraswasta	750000
Buruh	250000
Karyawan Swasta	1500000
Buruh	750000
Buruh	750000
PNS/TNI/Polri	3500000

Gambar 4.1 Transformasi Numerik Jumlah Penghasilan Orang Tua

1. Nilai rata-rata

Dilakukan transformasi dengan mengubah bilangan desimal ke bilangan bulat. Dengan menggunakan fungsi *Roundup* di *Excel* Dapat dilihat pada gambar 4.3

RATA-RATA
80,6
86,5
76,5
82,5
80,5
85,3

menjadi

Rata-rata
81
87
77
83
81
86

Gambar 4.2 Perubahan Bilangan Desimal ke Bilangan Bulat

2. Jarak Tempuh

Transformasi data hasil perhitungan Google Maps dapat dilihat pada gambar 4.4

alamat
ANJUN KANOMAN
Jl. Suriadipati
Bunut Kertayasa
R. Suriadipati
Bunut Kertayasa
R. Suriadipati

menjadi

jarak tempuh
1600
800
500
800
500
800

Gambar 4.3 Transformasi Jarak Tempuh Menggunakan *Google Maps*

Dari hasil *construct data* maka didapat *dartaset* dengan tipe data kategorikal. *Dataset* dapat dilihat pada tabel 4.7.

Tabel 4.1 Hasil *Construct* Data

No	Nama	JTO	nilai	jarak	Pekerjaan	JPO	Penerima KIP	Penerima KPS
1	AAB ABDUL JABBAR	1	81	1600	buruh	1500000	Tidak	Tidak
2	Abdul Aziz Muttaqin	2	87	800	Wiraswasta	1500000	Ya	Ya
3	ABDUL FATHIR	2	77	500	Buruh	1500000	Tidak	Tidak
4	ABDUL ROHIM SURYANA	2	83	800	Buruh	1500000	Tidak	Tidak
5	ABI RODIANSAH	2	81	500	Tidak bekerja	750000	Ya	Ya
6	ABIL SYAHRUL FADHILAH	1	86	800	Buruh	1500000	Tidak	Tidak
7	ADAM TIRADI	2	82	500	Buruh	750000	Tidak	Tidak

ISSN: 2503-054X
Vol. 3 No: 1, April 2018

3.4. Pengelompokan data

Dari hasil *construct* data selanjutnya memproduksi *dataset* akhir yang siap ditambang atau diolah dalam *tools* pemodelan *data mining*. Format *dataset* akhir berupa tabel atau *query* (*flat file*) untuk pemodelan *data mining*. Format tabel yang berupa *excel* akan diganti menjadi format *csv* agar tabel siap untuk diproses menggunakan *Weka*. Pada Tabel 4.8 adalah data set akhir.

Tabel 4.2 Dataset Akhir

Kriteria	Tipe Data	Keterangan
JTO	Nominal	Merupakan kriteria yang akan diakumulasikan
Nilai Rata-rata	Nominal	Merupakan kriteria yang akan diakumulasikan
Jarak tempuh	Nominal	Merupakan kriteria yang akan diakumulasikan
Pekerjaan orang tua	Nominal	Merupakan kriteria yang akan diakumulasikan
JPO	Nominal	Merupakan kriteria yang akan diakumulasikan
Penerima KIP	Nominal	Merupakan kriteria yang akan diakumulasikan
Penerima KPS	Nominal	Merupakan kriteria yang akan diakumulasikan

4. HASIL DAN PEMBAHASAN

Hasil yang diperoleh dari keseluruhan tes yang dilakukan baik dengan *use training set,* dari algoritma *C4.5*
dapat dilihat pada Tabel di bawah ini:
Evaluasi menggunakan *Accuracy* dan *Error Rate* dengan pengujian *Use Training Set* untuk *dataset* Skenario 1 sampai skenario 6

Tabel 4. Error! No text of specified style in document.**.3** *Accuracy, Error Rate* dan *Precision Use Training Set*

Klasifikasi Algoritma C4.5	Data berhasil di klasifikasikan	Data gagal di klasifikasikan	Accuracy (%)	Error Rate (%)	*Precision (%)*
Skenario 1	687	45	93,85%	6,14%	95,5 %
Skenario 2	686	46	93,715%	6,28%	95,2 %
Skenario 3	682	50	93,169%	6,83%	95,2 %
Skenario 4	667	65	91,12%	8,879%	98,4 %
Skenario 5	664	68	90,71%	9,28%	97,8 %
Skenario 6	661	71	90,30%	9,699%	98,5 %

Hasil dari Tabel 4.16 klasifikasi pada *dataset* dengan menggunakan algoritma *C4.5* menunjukan bahwa *classifier* yang terbentuk dengan pengukuran *use training set*. Terlihat bahwa dengan *dataset* Skenario 1 menghasilkan tingkat *Accuracy* sebesar 93,855% dengan *Error rate* paling rendah 6,14% yang artinya bahwa *classifier* tersebut membentuk model yang baik.
Pada penelitian ini skenario terbaik adalah skenario 1 (7 kriteria: Jumlah tanggungan orang tua, nilai rata- rata, jarak tempuh, pekerjaan orang tua, jumlah penghasilan orang tua, penerima KIP dan penerima KPS).

Selain terjadi kestabilan performa pada skenario 1, performa algoritma C4.5 mengalami penurunan. Setelah dianalisi penyebab dari menurunnya performa algoritma C4.5 pada skenario 2 sampai 6 disebabkan oleh atribut yang relevan dengan kelas target dikurangi atau dihilangkan. Hal ini dapat membuktikan bahwa algoritma C4.5 mengalami kendala dalam mendesain pohon keputusan yang optimal. Hasil kualitas keputusan yang

didapatkan dengan metode pohon keputusan sangat tergantung pada bagaimana pohon tersebut didesain.

5. KESIMPULAN

Pengujian Algoritma C4.5 diukur berdasarkan nilai *accuracy*, *precision*, dan *error rate*. Pengujian kinerja menerapkan *information gain* dengan 7 kriteria yaitu: JTO, nilai, jarak, pekerjaan, JPO, penerima KIP, penerima KPS menghasilkan nilai akurasi sebesar 93,58%, *precision* sebesar 95,5%, *error rate* 6,14% dikarenakan memiliki nilai ketepatan prediksi secara benar tertinggi dan persentase dari total data yang diprediksi secara salah terendah. Hasil ini merupakan model terbaik jika dibandingkan dengan skenario yang lain untuk memprediksi penerima beasiswa di SDN Karawang Kulon III karena Semua algoritma klasifikasi berusaha membentuk model yang memiliki akurasi tinggi (*Error Rate* yang rendah).

DAFTARPUSTKA

Apriliawan, Y. E. (2015). *Teknik Imputasi Missing Values pada Data Mining : Studi Kasus pada Data Hepatitis.*SKRIPSI

Fanani, A. Z., & Maurina, D. (2015). SemarangPenerapan data mining untuk rekomendasi beasiswa pada SMA Muhammadyah Gubug menggunakan algoritma c4.5. *Jurnal Informatika , 11*(5), 1-5.

Farida, S. (2017). *Penentuan model terbaik pada algoritma navie bayes untuk prediksi penerima beasiswa (Studi kasus : SMA Negri 2 Karawang).* SKRIPSI, Karawang.

Guterres, J. A. (2015). Kelayakan algoritma c4.5 sebagai pendukung keputusan dalam pengajuan penerima beasiswa. *Jurnal Informtika , A23*, 142-147.

Hastuti, N. F. (2013). *Pemanfaatan metode k-means clustering dalam penentuan penerima beasiswa.* SKRIPSI.

Hijriana, N., & Muttaqin, R. (2016). *Penerapan metode decision tree algoritma c4.5 untuk klasifikasi siswa berprestasi.* SKRIPSI.

Meilina, P. (2015). Penerapan data mining dengan metode kalsifikasi menggunakan decision tree dan regresi.

Jurnal Teknologi, 7(1), 12-20.

Muhardi, & Nisar. (2015). Penentuan penerima beasiswa dengan algoritma fuzzy c-means di universitas megow pak tulang bawang. *Jurnal TIM Darmajaya , 1*(2), 158.

Musthafa, A., Suyono, H., & Sarosa, M. (2015). Perbandingan kinerja algoritma c.45 dan ahp topsis. *Jurnal EECCIS , 9*(2), 109-114.

Restuputri , B. A., & Putri, A. M. (2014). *Penentuan kategori beasiswa mahasiswa menggunakan metode fuzzy tsukamoto.* SKRIPSI.

Shella, P. (2015). *Sistem pendukung keputusan dengan menggunakan decission tree dalam pemberian beasiswa di sekolah menengah pertama (studi kasus di SMP N 2 Rembang).* SKRIPSI.

Supriyanti, W., Kusrini, & Amborowati, A. (2016). perbandingan algoritma c45 dan navie bayes untuk ketepatan pemilihian konsentrasi mahasiswa. *Jurnal INFORMA Politeknik Indonusa Surakarta, 1*(3), 62-67.

Vulandari, R. T. (2017). *Teori dan aplikasi raoiminer.* Surakarta: Gava Media.