Penerapan Data Mining Untuk Memprediksi Tingkat Kemampuan Anak Dalam Mengikuti Mata Pelajaran Dengan Metode C4.5 Pada SDN 105351 Bakaran Batu

Aprianto Tumanggor^{1*}, Penda Sudiarto Hasugian²

^{1,2} STMIK Pelita Nusantara

Jl. Iskandar Muda No. 1 Medan 20154 Indonesia

Corresponding author's e-mail: apriantotumanggor@gmail.com

Abstrak— Perkembangan teknologi yang sangat maju dapat membantu segala aspek kehidupan, salah satunya di dunia pendidikan. Dunia pendidikan tentunya sangat membutuhkan peranan teknologi didalam mengelolah kegiatan-kegiatan dengan memanfaatkan teknologi informasi. Permasalahan yang muncul sistem untuk mengolah tumpukan data agar dapat mengetahui mampu atau tidaknya siswa tersebut mengikuti mata pelajaran pada SDN 105351 Bakaran Batu. Proses penerapan aplikasi data mining ini, hanya membahas tentang memprediksi tingkat kemampuan anak dalam mengikuti mata pelajaran. Data yang digunakan adalah data siswa kelas 5 SD dari 2 tahun terakhir (2018-2019) pada SDN 105351 Bakaran Batu. Salah satu model klasifikasi adalah dengan membuat pohon keputusan (*decision tree*). Dengan mengetahui tingkat kemampuan anak dalam mengikuti mata pelajaran siswa diharapkan dapat memberi solusi untuk kedepannya dalam pengolahan data siswa dengan system komputerisasi khususnya data mengenai tingkat kemampuan anak dalam mengikuti mata pelajaran. Cara kerja sistem yang dibangun menggunakan metode C4.5 serta aplikasi data mining yang akan dirancang dengan VB Net dan database Microsoft Access.

Kata kunci: kemampuan anak, data mining, metode C4.5, pohon keputusan

Abstract— The development of very advanced technology can help all aspects of life, one of which is in the world of education. The world of education certainly really needs the role of technology in managing activities by utilizing information technology. The problem that arises is the system for processing piles of data in order to know whether or not the student is able to follow the subjects at SDN 105351 Bakaran Batu. The process of implementing this data mining application only discusses predicting children's ability levels in following subjects. The data used were data from grade 5 elementary school students from the last 2 years (2018-2019) at SDN 105351 Bakaran Batu. One of the classification models is making a decision tree. By knowing the level of ability of children in following student subjects is expected to provide solutions for the future in processing student data with a computerized system, especially data regarding the level of children's ability in following subjects. The way the system works is built using the C4.5 method and data mining applications that will be designed with VB Net and Microsoft Access databases.

Keywords: children's ability, data mining, C4.5 method, decision tree

1. Pendahuluan

Data mining merupakan suatu proses penggalian data atau penyaringan data dengan memanfaatkan kumpulan data dengan ukuran yang cukup besar melalui serangkaian proses untuk mendapatkan informasi yang berharga dari data tersebut. Data mining bisa digunakan oleh sekolah untuk menggali data untuk mendapatkan informasi yang dapat menunjang dan meningkatkan proses pendidikan. *Decision Tree* merupakan metode klasifikasi yang paling sering digunakan. Dalam pengerjaannya tidak memerlukan waktu yang lama dan hasilnya pun mudah untuk dipahami dan banyak penelitian dalam kasus ini sering menggunakan *Decision Tree* untuk mendapatkan hasil yang maksimal [1]. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Setelah sebuah pohon keputusan dibangun maka dapat digunakan untuk mengklasifikasikan *record* yang belum ada kelasnya. Dimulai dari *Node root*, menggunakan tes terhadap atribut dari record yang belum ada kelasnya tersebut lalu mengikuti cabang yang sesuai dengan hasil dari tes tersebut, yang akan membawa kepada internal Node (Node yang memiliki satu cabang masuk dan dua atau lebih cabang yang keluar), dengan cara harus melakukan tes lagi terhadap atribut atau Node daun [2]. *Record* yang kelasnya tidak diketahui kemudian diberikan kelas yang sesuai dengan kelas yang ada pada Node daun. Pada pohon keputusan setiap simpul daun menandai label kelas. Proses dalam pohon keputusan yaitu mengubah bentuk data (tabel) menjadi model pohon (*tree*) kemudian mengubah model pohon menjadi aturan (*rule*) [3].

Tujuan dari algoritma penginduksi pohon keputusan adalah mengkonstruksi struktur data pohon yang dapat digunakan untuk memprediksi kelas dari sebuah kasus atau record baru yang belum memiliki kelas. C4.5 ini melakukan konstruksi pohon keputusan dengan metode *divide and conquer*. Aturan-aturan yang terbentuk dari pohon keputusan akan membentuk suatu kondisi dalam bentuk *if-then*. Aturan-aturan ini didapat dengan cara menelusuri pohon keputusan dari akar sampai daun [4]. Setiap *node* dan syarat percabangan akan membentuk suatu kondisi atau suatu *if*, sedangkan untuk nilai-nilai yang terdapat pada daun akan membentuk suatu hasil atau suatu *then*. Penelitian berjudul memprediksi tingkat peminat ekstrakurikuler pada siswa SMK Analisis Kesehatan Abdurrah menggunakan algoritma C4.5 (studi kasus: SMK analisis kesehatan Abdurrah) menjelaskan pengolahan data mining terdiri dari beberapa metode pengolahan, yaitu predictive modelling yang merupakan pengolahan data mining dengan melakukan prediksi atau peramalan [5]. Tujuan metode ini untuk membangun model prediksi suatu nilai yang mempunyai ciri-ciri tertentu. Contoh algoritmanya *Linear Regression*, *Neural Network*, *Support Vector Machine*, dan lain-lain. Informasi yang dihasilkan diperoleh dengan cara mengekstrak dan mengenali pola yang penting atau menarik dari data yang terdapat pada basis data [6]. Data mining terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering disebut *Knowledge Discovery Databases* (KDD). Metode C4.5 membangun pohon keputusan dimulai dari pemilihan atribut sebagai akar, membuat cabang untuk tiap-tiap nilai, membagi kasus dalam cabang dan mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama [7].

2. Tinjauan Pustaka

2.1. Data Mining

Data mining didefinisikan sebagai sebuah proses untuk menemukan hubungan, pola dan trend baru yang bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [8]. Data mining dapat diartikan sebagai serangkaian proses untuk menggali nilai tambah yang berupa informasi yang selamaini tidak diketahui secara manual dari suatu basis data dengan melakukan proses ekstraksi dan mengenali pola penting dari data yang ada. Data mining didefinisikan sebagai sebuah proses untuk menemukan hubungan, pola dan tren baru yang bermakna dengan menyaring data yang sangat besar, yang tersimpan dalam penyimpanan, menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [9]. Data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data.

2.2 Metode C4.5

Tahapan pohon keputusan metode C4.5 [10] [11]:

- Mempersiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
- Menghitung akar dari pohon. Akar akan diambil dari atribut yang akan terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung nilai entropy.

$$Entropy(S) = -\sum_{i=1}^n p_i * \log_2 p_i \quad (1)$$

$$Gain(S,A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

- Ulangi langkah ke 2 dan langkah ke 3 hingga semua *record* terpartisi.
- Proses partisi pohon keputusan akan berhenti saat:
 - Semua record dalam simpul N mendapat kelas yang sama.
 - Tidak ada atribut di dalam record yang dipartisi lagi.
 - Tidak ada record di dalam cabang yang kosong.

3. Metode Penelitian

Tahap dalam penelitian adalah sebagai berikut:

1. Analisa Masalah

Analisa masalah selama melakukan penelitian pada SDN 105351 Bakaran Batu diperoleh dari hasil studi observasi serta wawancara sehingga masalah-masalah dalam penelitian mendapat solusi. Penelitian membahas dan menyampaikan uraian hasil penelitian mengenai prediksi tingkat kemampuan anak dalam mengikuti mata pelajaran pada SDN 105351 Bakaran Batu.

2. Pegumpulan Data

Dalam memperoleh data-data, beberapa metode telah dilakukan guna mendapatkan data yang sesuai

dengan kebutuhan penelitian. Metode observasi digunakan untuk mengetahui dan mengamati secara langsung objek yang diteliti dan juga melakukan proses pengumpulan data. Metode wawancara, pengumpulan data dengan melakukan tanya jawab langsung. Proses wawancara dilakukan untuk memperoleh data yang diinginkan dalam penilaian kemampuan siswa dalam mengikuti matapelajaran pada SDN 105351 Bakaran Batu. Studi literatur yang dilakukan adalah metode pengumpulan data dengan cara membaca dan membandingkan literatur yang sebagian besar berasal dari artikel-artikel di internet, buku, jurnal dan E-book. Literatur tersebut berhubungan dengan penulisan penelitian yang mendukung topik prediksi kemampuan siswa dalam mengikuti mata pelajaran metode C4.5 data mining.

3. Analisis Data

Analisis data adalah upaya atau cara untuk mengolah data menjadi informasi sehingga dapat dipahami dan bermanfaat untuk solusi permasalahan

4. Menerapkan Metode C4.5

Penerapan metode C4.5 dalam penyelesaian masalah prediksi tingkat kemampuan anak dalam mengikuti mata pelajaran pada SDN 105351 Bakaran Batu.

5. Perancangan Sistem

Pada perancangan sistem tahapan yang menggambarkan rancangan bangun sistem yang dibuat dengan metode UML (*Unified Modelling Language*). Perancangan sistem ini terdiri dari perancangan *interface*, *Use Case Diagram*, *Activity Diagram*, dan *Class Diagram*.

6. Pembuatan Aplikasi

Pada tahap pembuatan aplikasi melakukan perancangan sistem mengenai sistem aktivitas yang ada dibangun sistem dengan bahasa pemrograman *Visual Basic 2010* menggunakan *database access 2010*.

7. Pengujian Sistem

Pada tahap pengujian terhadap sistem prediksi tingkat kemampuan anak dalam mengikuti mata pelajaran pada SDN 105351 Bakaran Batu menggunakan metode *Black Box*. Proses pengujian membuktikan bahwa aplikasi telah sesuai dengan rancangan awal dari sistem yang telah dirancang dan telah memenuhi kebutuhan yang diharapkan.

8. Implementasi Sistem

Pada tahap Implementasi program diterapkan sesuai dengan tahap dimana sistem siap untuk dioperasikan di SDN 105351 Bakaran Batu untuk membantu dalam prediksi tingkat kemampuan anak dalam mengikuti mata pelajaran.

4. Hasil dan Pembahasan

Data nilai data siswa dalam bentuk angka untuk mempermudah didalam pemahan dan analisa data, maka nilai-nilai dalam bentuk angka dikompersi menjadi bentuk nilai dalam bentuk keterangan atau nilai dalam bentuk yudisium dengan range nilai Nilai 0-69 nilainya "Cukup" Nilai 70-79 nilainya "Baik" Nilai 80-100 nilainya "Sangat Baik". Pada data mata pelajaran yang digunakan untuk memprediksi tingkat pemahaman siswa dalam mengikuti mata pelajaran, yaitu Agama, PPKn, Bahasa Indonesia, Matematika, IPA, IPS dan Muatan Lokal. Setelah seluruh nilai Entropy dan Gain diperoleh, selanjutnya hasil dari perhitungan tersebut dimasukkan ke dalam Nilai Akhir Node 1.

Tabel 1. Nilai Akhir Node 1

NODE			Jml Kasus (S)	Baik (B)	Cukup (C)	Entropy	Gain
1	TOTAL		30	16	14	0,99679	
	Agama						0,06734
		Sangat Baik	25	15	10	0,97095	
		Baik	5	1	4	0,72193	
	PPKn						0,17213
		Sangat Baik	5	5	0	0	
		Baik	25	11	14	0,98959	
	B. Indonesia						0,38121
		Sangat Baik	13	12	1	0,39124	
		Baik	17	4	13	0,78713	
	Matematika						0.13382

		Sangat Baik	4	4	0	0	
		Baik	26	12	14	0,99573	
	IPA						0.53686
		Sangat Baik	15	14	1	0,35336	
		Baik	15	2	13	0,56651	
	IPS						0.39261
		Sangat Baik	9	9	0	0	
		Baik	21	6	15	0,86312	
	B. Inggris						0.37687
		Sangat Baik	22	16	6	0,84535	
		Baik	8	0	8	0	

Dari perhitungan Tabel 1 dapat diketahui bahwa atribut dengan Gain tertinggi adalah Mata Pelajaran IPA dengan nilai 0,53686, sehingga atribut Mata pelajaran IPA dapat menjadi Node Akar. Pada atribut IPA yang dijadikan akar terdapat 3 nilai yaitu “Sangat Baik”, dan “Baik”. Dari nilai atribut tersebut, nilai “Sangat Baik” mempunyai Hasil = “Baik” sehingga tidak perlu dilakukan perhitungan kembali. Sedangkan nilai “Baik” harus dilakukan perhitungan kembali.

Gambar 1. Pohon Keputusan Hasil Perhitungan Node 1

Dari gambar 1 bahwa yang dijadikan Node akar adalah IPA, IPA ini dibagi ke dalam 2 tipe yaitu “Sangat Baik” dan “Baik”. Untuk nilai “Sangat Baik” tidak perlu dilakukan perhitungan kembali dikarenakan hasil dari nilai tersebut adalah “Baik” sedangkan untuk nilai “Baik” harus dilakukan perhitungan kembali dikarenakan hasilnya belum bernilai “Baik”.

Perhitungan dilakukan untuk mencari node cabang dari nilai atribut baik. Perhitungan dilakukan dengan mencari nilai dari atribut selain yang menjadi node akar (IPA), yaitu dengan mencari jumlah kasus untuk hasil Baik dan Cukup, dan nilai Entropy dari semua kasus saat IPA = Baik. Dan kemudian lakukan perhitungan untuk mencari nilai Gain, dan atribut dengan nilai Gain terbesar, maka akan menjadi node cabang dari nilai atribut Baik. Jumlah Kasus IPA = Baik = 15. Jumlah Kasus IPA = Baik, yang hasilnya Baik = 2. Jumlah Kasus IPA = Baik, yang hasilnya Cukup = 13.

Setelah seluruh nilai Entropy dan Gain diperoleh, selanjutnya hasil dari perhitungan tersebut dimasukkan ke dalam Tabel 2.

Tabel 2. Nilai Akhir Node 1.2

NODE			Jml Kasus (S)	Baik (B)	Cukup (C)	Entropy	Gain
2	TOTAL		15	2	13	0,56651	
	Agama						0,06488
		Sangat Baik	11	2	9	0,68404	
		Baik	4	0	4	0	
	PPKn						0
		Sangat Baik	0	0	0	0	
		Baik	15	2	13	0.56651	
	B. Indonesia						0,0941
		Sangat Baik	2	1	1	1	
		Baik	13	1	12	0,39124	
	Matematika						0

		Sangat Baik	0	0	0	0	
		Baik	15	2	13	0.56651	
	IPS						0.03777
		Sangat Baik	1	1	0	0	
		Baik	14	1	13	0.37123	
	B. Inggris						0.16372
		Sangat Baik	7	2	5	0,86312	
		Baik	8	0	8	0	

Pada perhitungan pada Tabel 2 dapat diketahui bahwa atribut dengan Gain tertinggi adalah Mata Pelajaran B. Inggris dengan nilai 0,16372, sehingga atribut Mata pelajaran B. Inggris dapat menjadi Node Akar berikutnya.

Gambar 2. Pohon Keputusan Akhir

Ada 2 atribut dari B. Inggris yaitu “Sangat Baik” dan “Baik”. Nilai “Sangat Baik” mempunyai Hasil “Baik” sedangkan nilai “Baik” mempunyai Hasil = “Cukup”. Apabila dilihat dari atribut selanjutnya yaitu Agama diperoleh bahwa Jika Agama bernilai “Sangat Baik” dan PPkn “Baik” maka Hasil = “Cukup” sedangkan Bahasa Indonesia yang nilai “Baik” dan Matematika Bernilai “Baik” maka Hasil = “Cukup”. Berdasarkan dari pohon keputusan yang terbentuk seperti pada gambar 2, maka dapat dihasilkan rule

1. If IPA = “Sangat Baik” then Hasil = “Baik”
2. If IPA = “Baik” and B. Inggris = “Sangat Baik” then Hasil = “Baik”
3. If IPA = “Baik” and B. Inggris = “Baik” and Bahasa Indonesia = “Sangat Baik” then Hasil = “Cukup”
4. If IPA = “Baik” and B. Inggris = “Baik” and Bahasa Indonesia = “Sangat Baik” and Agama = “Sangat Baik” then Hasil = “Cukup”
5. If IPA = “Baik” and B. Inggris = “Baik” and Bahasa Indonesia = “Sangat Baik” and Agama = “Sangat Baik” and IPS = “Sangat Baik” then Hasil = “Cukup”.

Tampilan aplikasi Form Prediksi merupakan form yang berfungsi untuk mengelolah data prediksi, berupa proses data prediksi. Form prediksi akan tampil melalui menu proses pada form utama, kemudian pilih pada sub menu form prediksi, Gambar 3.

Data Awal

No Indeks Siswa	Nama Siswa	Nilai Agama	Nilai PPKn	Nilai B. Indonesia	Nilai Matematika	Nilai IPA	Nilai IPS	Nilai B. Inggris	Nilai Hasil
3070	Ali Syahri	69	76	78	76	67	78	69	79
3083	Aisyah Anggia Kasih	77	76	75	75	74	75	76	75
3084	Azi Fransah Anggara	75	69	68	73	73	74	76	73
3085	Azi Pratama	77	74	75	74	75	76	77	75
3110	Ashila Dirana	76	73	73	69	76	67	76	73

Data Akhir

No Indeks Siswa	Nama Siswa	Nilai Agama	Nilai PPKn	Nilai B. Indonesia	Nilai Matematika	Nilai IPA	Nilai IPS	Nilai B. Inggris	Nilai Hasil
3070	Ali Syahri	Cukup	Baik	Baik	Baik	Cukup	Baik	Cukup	Baik
3083	Aisyah Anggia Kasih	Baik	Baik	Baik	Baik	Baik	Baik	Baik	Baik
3084	Azi Fransah Anggara	Baik	Cukup	Cukup	Baik	Baik	Baik	Baik	Baik
3085	Azi Pratama	Baik	Baik	Baik	Baik	Baik	Baik	Baik	Baik
3110	Ashila Dirana	Baik	Baik	Baik	Cukup	Baik	Cukup	Baik	Baik

Gambar 3. Tampilan Form Prediksi

5. Kesimpulan

Kesimpulan dari penelitian :

1. Algoritma c4.5 telah berhasil diterapkan untuk mendapatkan klasifikasi data dengan membentuk pohon keputusan dari seluruh data maupun sebagian data. Dari hasil klasifikasi berupa pohon keputusan tersebut dapat digunakan untuk memprediksi hasil dari tingkat kemampuan anak terhadap mata pelajaran.
2. Hasil dari data mining ini dengan menggunakan metode C4.5 merupakan sebuah susunan urutan kegiatan yang saling mendukung penilaian siswa sehingga lebih mudah dipahami dengan melihat tahapan gambar pohon keputusan.

Daftar Pustaka

- [1] A. E. Wicaksono, "Implementasi Data Mining Dalam Pengelompokan Peserta Didik di Sekolah untuk Memprediksi Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means (Studi Kasus SMA N 6 Bekasi)," *Jur. Tek. Inform. Univ. Gunadarma*, vol. 21, no. 3, pp. 206–216, 2016.
- [2] H. Susanto and S. Sudiyatno, "Data mining untuk memprediksi prestasi siswa berdasarkan sosial ekonomi, motivasi, kedisiplinan dan prestasi masa lalu," *J. Pendidik. Vokasi*, vol. 4, no. 2, pp. 222–231, 2014.
- [3] O. OSCARIO, J. JASMIR, and Y. NOVIANTO, "PENERAPAN ALGORITMA C4.5 UNTUK MEMPREDIKSI KECOCOKAN GAYA BELAJAR BAGI SISWA SISWI SEKOLAH DASAR (STUDI KASUS : SD SARIPUTRA Algoritma C4.5 Untuk Memprediksi Kecocokan Gaya Belajar Bagi Siswa Siswi Sekolah Dasar (Studi Kasus : SD Sariputra Jambi)," *J. Process.*, vol. 14, no. 2, p. 141, 2019.
- [4] S. M. Hutabarat and A. Sindar, "Data Mining Penjualan Suku Cadang Sepeda Motor Menggunakan Algoritma K-Means," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 2, no. 2, pp. 126–132, 2019.
- [5] N. Yahya and A. Jananto, "Komparasi Kinerja Algoritma C.45 Dan Naive Bayes Untuk Prediksi Kegiatan Penerimaan mahasiswa Baru (Studi Kasus : Universitas Stikubank Semarang)," *Pros. SENDI*, no. 2014, pp. 978–979, 2019.
- [6] M. H. Luluk Elvitaria, "Smk Analisis Kesehatan Abdurrah Menggunakan Algoritma," (*Jurnal Teknol. dan Sist. Inf. Univrab*), vol. 2, no. 2, pp. 220–233, 2017.
- [7] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," *Sistemasi*, vol. 7, no. 3, p. 238, 2018.
- [8] S. Butsianto and N. T. Mayangwulan, "Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 3, no. 3, pp. 187–

- 201, 2020.
- [9] B. Hermanto, A. S. Azhari SN, and F. P. Putra, "Analisis Kinerja Decision Tree C4.5 dalam Prediksi Potensi Pelunasan Kredit Calon Debitur," *INOVTEK Polbeng - Seri Inform.*, vol. 2, no. 2, p. 189, 2017.
- [10] H. Dhika and F. Destiwati, "Penerapan Algoritma C45 Untuk Penilaian Karyawan Pada Restoran Cepat Saji," no. September, pp. 55–59, 2018.
- [11] J. S. Parapat and A. S. RMS, "Data Mining Klasifikasi Data Nasabah Kredit KSU Taman Mandiri Menggunakan Algoritma C4.5," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 4, no. 2, p. 150, 2019.