

## KOMPARASI KLASIFIKASI PENENTUAN KETERLAMBATAN SISWA SMA DATANG UPACARA MENGGUNAKAN ALGORITMA C4.5

Fajar Pramono<sup>1</sup>, Suwanda Aditya Saputra<sup>2</sup>, Burhanuddin<sup>3</sup>, Kusuma Ade<sup>4</sup>

<sup>1</sup>Program Studi Teknik Informatika, STMIK Bani Saleh Bekasi

Jl. Mayor Hasibuan No.68, Margahayu, Bekasi Timur, Telp. (021) 8800992

<sup>2,3,4</sup>Program Studi Ilmu Komputer, STMIK Nusa Mandiri Jakarta

Jl. Kramat Raya No.18, senen, Jakarta Pusat, Telp. (021) 31908575

Email: fajarpramono@gmail.com, wanda.adit@gmail.com, burhan.tugas@gmail.com, kusuma.ade@gmail.com

### ABSTRAKS

Proses upacara siswa akan berjalan dengan baik, jika diawali dengan proses kedatangan awal siswa di sekolah. dengan datang lebih awal di saat upacara mempengaruhi proses pembelajaran pada hari tersebut. Dari hasil penelitian yang dilakukan, dalam proses kedatangan siswa ke sekolah menggunakan algoritma C4.5 di dapatkan nilai Entropy dan Gain tertinggi yaitu 0.46179 terdapat pada atribut bangun tidur dan sampai di sekolah menggunakan cara manual, sedangkan menggunakan aplikasi rapid miner dihasilkan keputusan siswa sering telat datang upacara ke sekolah sebanyak 13 %, karena mereka bangun tidur lebih dari jam 6:17. Serta disebabkan juga jarak lokasi rumah ke sekolah lebih dekat 3.25 km atau sekitar 20%.

**Kata Kunci:** Algoritma C4.5, Siswa terlambat, KDD, Pohon Keputusan, Rapid miner

### 1. PENDAHULUAN

#### 1.1 Latar Belakang

Ketepatan kedatangan siswa pada saat upacara hari senin, mempengaruhi proses pelaksanaan upacara yang diadakan oleh SMA Global Prestasi di bekasi menjadi tepat waktu atau tidak.

Menurut Dafiaghor (2011), keterlambatan siswa saat datang sekolah juga dapat menyebabkan siswa tersebut tidak dapat maksimal mencapai prestasi, karena ketidakmampuan mengorganisasikan manajemen waktu terhadap sistem belajar yang berlaku di sekolah.

Menurut Ajay Kumar Pal (2013), dengan tepatnya siswa datang pada setiap upacara, maka kualitas pendidikan dapat diproduksi dengan cara yang efisien dan mereka dapat belajar tanpa ada masalah.

Banyak siswa yang selalu datang terlambat pada saat upacara hari senin. ada beberapa faktor yang menyebabkan mereka terlambat, diantaranya, karena bangun tidur, transportasi yang digunakan menuju sekolah, ataupun jarak tempuh dari rumah ke sekolah.

Menurut Dafiaghor (2011), ada beberapa macam penyebab siswa datang terlambat ke sekolah, yaitu :

a. Waktu tidur malam yang telat

Ini adalah yang menjadi faktor utama penyebab siswa datang terlambat ke sekolah.

b. Menonton film atau TV terlalu malam

Kebiasaan menonton film terlalu malam juga menyebabkan siswa terlambat bangun, dikarenakan terlalu asik dengan film yang ditonton sehingga lupa akan waktu tidur.

c. Kebiasaan bangun pagi telat

Banyak juga siswa telat bangun tidur di pagi hari, hingga enggan beranjak dari tempat tidurnya, sehingga ketika bangun tidur sudah telat waktunya.

d. Jarak dari rumah ke sekolah

Jarak juga menentukan ketepatan waktu seseorang, karena jika jarak rumah terlalu jauh, menyebabkan waktu yang dibutuhkan untuk sampai ke sekolah terlalu lama

e. Latar belakang budaya keluarga

Ada juga siswa yang telat, dikarenakan sering melihat beberapa anggota keluarga seperti ayah atau kakaknya, yang berangkat ke tempat kerja terlalu siang, sehingga mempengaruhi psikologi siswa menjadi enggan berangkat lebih awal.

Algoritma C4.5 ini pernah dilakukan pada penelitian sebelumnya dalam bidang keuangan untuk menentukan resiko kredit kendaraan bermotor dan menghasilkan nilai akurasi yang tinggi. serta pada bidang pendidikan di india untuk menentukan hasil performa akademik siswa berdasarkan nilai ujian, sehingga bisa bermanfaat dalam hal penentuan tingkat akademik selanjutnya.

Oleh karena itu peneliti dalam hal ini ingin mencoba melakukan hal yang sama pada masalah komparasi klasifikasi keterlambatan siswa mengikuti upacara dengan menggunakan algoritma C4.5.

## 1.2 Tinjauan Pustaka

### 1.2.1 Data Mining

Menurut Han J (2006:5) “Data mining adalah proses menemukan pengetahuan menarik, seperti asosiasi, pola, perubahan, struktur dan anomali yang signifikan, dari sejumlah besar data yang tersimpan dalam database atau gudang data atau repositori informasi lainnya.”.

Menurut J Larose (2005) Data mining juga dapat diartikan sebagai 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.

Menurut Ajay Kumar Pal (2013), baru-baru ini, pendidikan tinggi telah meningkat cepat. Banyak institusi baru, perguruan tinggi dan universitas didirikan oleh swasta dan sektor pemerintah untuk meningkatkan pertumbuhan pendidikan dan kesejahteraan para siswa.

Dimana dengan pendidikan seorang guru menjelaskan materi dan siswa mengerti dan belajarpelajaran. Tidak ada skala mutlak dalam hal pengukuran pengetahuan, tapi nilai ujian adalah satu skala yang mana menunjukkan indikator kinerja siswa. penggunaan teknik data mining dapat dilakukan untuk meningkatkan efisiensi akademik kinerja di lembaga pendidikan. metode ini membantu untuk mengidentifikasi siswa yang membutuhkan perhatian atau nasehat khusus dari guru untuk meningkatkan kualitas pendidikan.

Menurut Kalpesh Adhatrao (2013), sebuah institusi pendidikan perlu memiliki perkiraan pengetahuan sebelumnya tentang siswa yang akan terdaftar, guna memprediksi tingkat akademik mereka di masa depan.

Ini membantu mereka untuk mengidentifikasi siswa dan juga menyediakannya mereka kesempatan untuk memperhatikan dan memperbaiki orang-orang yang mungkin akan mendapatkan nilai lebih rendah. sebagai solusi, telah di kembangkan sebuah sistem yang bisa memprediksi kinerja siswa dari sebelumnya.

Dengan menggunakan konsep teknik data mining di bawah Klasifikasi, dapat di analisisa kumpulan data yang berisi informasi tentang siswa, seperti jenis kelamin, tanda yang dicetak dalam ujian kelas X dan XII, tanda dan peringkat di ujian masuk dan menghasilkan tahun pertama dari angkatan sebelumnya siswa.

Dengan menerapkan algoritma klasifikasi ID3 (Iterative Dichotomiser 3) dan C4.5 sehingga dapat meramalkan kinerja umum dan individual siswa yang baru dan akan diterima di masa depan pada saat ujian ujian.

Menurut Hamidah Jantan (2014), dalam HRM antara tantangan bagi profesional HR adalah untuk mengelola bakat organisasi, terutama untuk memastikan orang yang tepat untuk pekerjaan yang tepat pada waktu yang tepat. Prediksi merupakan alternatif untuk menangani masalah ini. Karena itu alasan, klasifikasi dan prediksi dalam data mining yaitu biasanya digunakan di banyak daerah juga bisa diimplementasikan bakat manusia.

Ada banyak teknik klasifikasi dalam datateknik penambangan seperti Decision Tree, Neural Network, Rough Set Theory, teori Bayesian dan logika fuzzy. Pohon keputusan adalah salah satu teknik klasifikasi yang populer, yang bisa menghasilkan aturan interpretasi atau pernyataan logika.

Aturan yang dihasilkan dari teknik yang dipilih dapat digunakan prediksi masa depan. Dalam hal ini, studi tentang bagaimana caranya potensi bakat manusia bisa diprediksi menggunakan pohon keputusan. Dengan menggunakan teknik ini, pola bakat dan kinerja dapat diidentifikasi melalui klasifikasi proses.

Dalam hal ini, pengetahuan yang tersembunyi dan berharga, ditemukan di database terkait dan dirangkum dalam struktur pohon keputusan C4.5 untuk menghasilkan aturan klasifikasi catatan kinerja bakat manusia.

Menurut D.Lavanya (2011), dalam data mining, klasifikasi adalah satu hal teknik yang signifikan dengan aplikasi dalam deteksi kecurangan, kecerdasan buatan, diagnosa medis dan banyak bidang lainnya.

Klasifikasi objek berdasarkan fitur mereka ke dalam kategori yang telah ditentukan adalah masalah yang dipelajari secara luas.

Pohon keputusan sangat berguna untuk mendiagnosis masalah pasien oleh dokter. Pengklasifikasian pohon keputusan digunakan secara ekstensif untuk diagnosis tumor payudara pada gambar ultrasonik, kanker ovarium dan diagnosa suara jantung.

Menurut Babak Dalvand (2014), Data analisis yang terselebung. dapat diterima untuk mengukur efisiensi dalam berbagai bidang. Dengan menggunakan algoritma C4.5 kita dapat mengatasi salah satu kekurangan dan memprediksi nilai pengambilan keputusan dari suatu units.

### 1.2.2 Pohon keputusan

Manusia selalu dihadapkan dengan berbagai macam masalah dari berbagai bidang kehidupan. Masalah ini juga memiliki variasi tingkat kesulitannya.

Menurut Fandy Ferdian (2017), Untuk menghadapi masalah ini manusia mulai mengembangkan sebuah sistem untuk membantu mereka menyelesaikan masalah-masalah ini, salah satu sistem tersebut adalah pohon keputusan.

Pohon keputusan adalah metode klasifikasi dan prediksi yang sudah terbukti powerfull dan sangat terkenal. Metode ini berfungsi untuk mengubah fakta menjadi pohon keputusan yang merepresentasikan aturan yang dapat mudah dimengerti dengan bahasa alami. Proses dari pohon keputusan ini dimulai dari node akar hingga node

daun yang dilakukan secara rekursif dimana setiap percabangan menyatakan kondisi dan setiap ujung pohon akan menyatakan keputusan.

Menurut Triisant (2017), Arsitektur pohon keputusan dibuat sedemikian rupa agar menyerupai pohon asli, dimana terdapat beberapa bagian yaitu:

- a. *Root Node* : Node ini terletak pada bagian paling atas dari pohon keputusan.
- b. *Internal Node*: Node ini merupakan percabangan dimana membutuhkan satu input dan mengeluarkan maksimal dua output.
- c. *Leaf Node* : Node ini merupakan node yang terletak pada ujung pohon. Node ini hanya memiliki satu input dan tidak memiliki output.

Menurut Triisant (2017), pohon keputusan ini memiliki beberapa kekurangan dan kelebihan, yakni:

1. Kelebihan Pohon Keputusan:
  - Daerah pengambilan keputusan yang kompleks dapat diubah menjadi sederhana.
  - Dapat menghilangkan perhitungan yang tidak penting karena proses pengujian hanya berdasarkan kriteria yang diperlukan saja.
  - Proses pemilihan fitur dari internal node yang berbeda lebih fleksibel. Fitur yang telah dipilih ini akan menjadi pembeda antara kriteria yang satu dengan kriteria lainnya.
  - Metode ini dapat menghindari munculnya permasalahan dengan cara menggunakan kriteria dengan jumlah yang sedikit pada node internal tanpa mengurangi kualitas keputusan yang dihasilkan.
2. Kekurangan Pohon Keputusan :
  - Dapat terjadi overlap apabila hasil keputusan dan kriteria yang digunakan jumlahnya sangat banyak. Hal ini juga dapat berakibat bertambahnya waktu yang digunakan untuk pengambilan keputusan dan jumlah memori yang dibutuhkan semakin tinggi.
  - Akumulasi jumlah error dari setiap tingkat pohon keputusan besar.
  - Mendesain pohon keputusan yang optimal sulit.
  - Kualitas keputusan yang didapatkan sangat tergantung dengan bagaimana pohon tersebut didesain.

### 1.2.3 KDD (*Knowledge discovery in database*)

Menurut Han, J (2006), Istilah data mining sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar.

Proses KDD secara garis besar dapat dijelaskan sebagai berikut:

#### 1. *Data selection*

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

#### 2. *Pre-processing / Cleaning*

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi focus KDD. Proses cleaning mencakup antara lain, membuang duplikasi data, memeriksa data yang inkonsistensi, dan memperbaiki kesalahan pada data seperti kesalahan cetak (tipografi). Juga dilakukan proses enrichment, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eketernal.

#### 3. *Transformation*

Coding adalah proses transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

#### 4. *Data mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

#### 5. *Interpretation/Evaluation*

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation. Tahap mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

### 1.2.4 Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk melakukan proses klasifikasi data dengan menggunakan teknik pohon keputusan.

Menurut Kalpesh Adhatrao (2013), Algoritma ini merupakan perpanjangan dari ID3, dan menghasilkan pohon keputusan yang dapat digunakan untuk klasifikasi dan statistik klasifikasi.

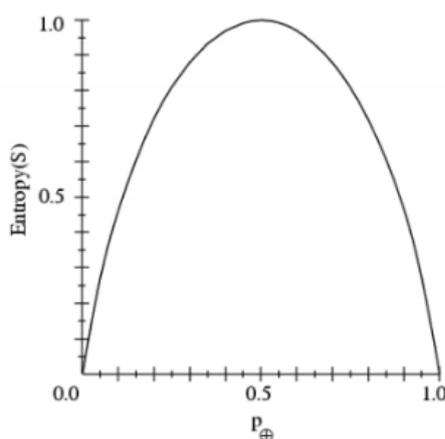
Ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai gain tertinggi berdasarkan nilai entropy atribut tersebut sebagai poros atribut klasifikasi.

Menurut Fandy Ferdian Harryanto (2017), terdapat empat langkah dalam proses pembuatan pohon keputusan pada algoritma C4.5, yaitu:

1. Memilih atribut sebagai akar
2. Membuat cabang untuk masing-masing nilai
3. Membagi setiap kasus dalam cabang
4. Mengulangi proses dalam setiap cabang sehingga semua kasus dalam cabang memiliki kelas yang sama.

### 1. Konsep Entropy

Menurut kamus IGI Global (*International Publisher of Progressive Academic*), *Entropy(S)* adalah jumlah data yang tidak relevan terhadap informasi dari suatu kumpulan data. Semakin kecil nilai entropy, maka akan semakin sering entropy digunakan untuk mengekstrak suatu kelas. Entropy digunakan untuk mengukur ketidakpastian S. sebagaimana tergambar dalam gambar 1 (Denis, 2013:50)



Gambar 1. Grafik Entropy

Adapun perhitungan Entropy dapat dilakukan dengan rumus :

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

Keterangan:

S= Himpunan kasus

n= Jumlah partisi S

$p_i$  = Proporsi  $S_i$  terhadap S

### 2. Konsep Gain

Gain adalah informasi yang didapatkan dari perubahan entropy pada suatu kumpulan data, baik melalui observasi atau bisa juga disimpulkan dengan cara melakukan partisipasi terhadap suatu set data [8].

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

Keterangan:

S= Himpunan kasus

A= Fitur

n= Jumlah partisi atribut A

$|S_i|$  = Proporsi  $S_i$  terhadap S

S= jumlah kasus dalam S

### 1.2.5 Rapid Miner

Perangkat lunak yang bersifat terbuka (*Open Source*). menurut Dennis (2013), Rapid Miner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi Rapid Miner menggunakan berbagai teknik deksriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik”.

Rapid Miner sebelumnya bernama YALE (*Yet Another Learning Environment*), dimana versi awalnya mulai dikembangkan pada tahun 2001 oleh Ralf Klinberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit dari University of Dortmund.

Rapid minner menyediakan GUI (*Graphical User Interface*) untuk merancang sebuah pipeline analisis. GUI ini akan menghasilkan file XML (*Extensible Markup Language*) yang mendefinisikan proses analisis keinginan pengguna untuk diterapkan ke data. File ini kemudian di baca oleh Rapid Miner untuk menjalankan analisis secara otomatis.

### 1.2.6 Populasi

Menurut Prasetyo (2011). Populasi adalah keseluruhan gejala atau satuan yang ingin diteliti. Populasi dalam penelitian ini merupakan data ketepatan datang siswa aktif kelas X.

### 1.2.7 Sample penelitian

Menurut J Hamidah Jantan (2014), Sample merupakan “bagian dari populasi yang ingin diteliti. Oleh karena itu, sample harus dilihat sebagai suatu pandangan pendugaan terhadap populasi dan bukan populasi itu sendiri.

**Tabel 1. Contoh data penelitian**

NO	NIS	NAMA LENGKAP	JENKEL	KE SEKOLAH DENGAN	JARAK KE SEKOLAH dalam KM	WAKTU BANGUN TIDUR	STATUS
1	1718.10.034	AAK	L	ONLINE	2.1	4:30	TIDAK TELAT
2	1718.10.012	AAP	L	MOBIL	2.2	6:20	TELAT
3	1718.10.020	ARA	L	SEPEDA	0.65	6:30	TELAT
4	1718.10.071	ISH	L	MOTOR	12.6	4:30	TIDAK TELAT
5	1718.10.016	LDD	L	DIANTAR SUPIR	5.5	5:00	TIDAK TELAT
6	1718.10.056	MCS	P	DIANTAR SUPIR	4	5:00	TIDAK TELAT
7	1718.10.032	MLS	P	MOTOR	6.4	5:30	TIDAK TELAT
8	1718.10.083	NWT	L	MOTOR	8.2	5:00	TIDAK TELAT
9	1718.10.059	NAH	P	MOTOR	7	5:15	TIDAK TELAT
10	1718.10.003	PAM	P	DIANTAR SUPIR	9.6	5:00	TIDAK TELAT
11	1718.10.021	RYA	L	ONLINE	2.5	5:00	TELAT
12	1718.10.027	SOB	L	DIANTAR SUPIR	4.3	6:15	TIDAK TELAT
13	1718.10.025	SOY	P	DIANTAR SUPIR	9	5:00	TIDAK TELAT
14	1718.10.037	VBR	L	DIANTAR SUPIR	6.2	5:00	TIDAK TELAT
15	1718.10.018	YAT	P	MOTOR	5	5:15	TIDAK TELAT

### 1.3 Metodologi Penelitian

Dalam penelitian ini menggunakan Model CRSIP-DM (*Cross-Industry Standard Proses for Data Mining*), yang terdiri dari 6 tahap yaitu:

#### 1. Business / Research Understanding Phase

Banyaknya anak yang sering terlambat pada saat upacara setiap hari senin di sekolah. Penulis menggunakan metode Algoritma C4.5 untuk mengklasifikasikan siswa yang datang tepat waktu dan telat datang ke sekolah

#### 2. Data Understanding Phase

Atribut-atribut yang menjadi parameter terlihat pada tabel 2 yaitu

**Tabel 2. Atribut dan nilai kategori**

No	Atribut	Nilai
1	KE SEKOLAH DENGAN	SEPEDA
		MOTOR
		DIANTAR SUPIR
		MOBIL
		ONLINE
2	JARAK KE SEKOLAH	0 - 3 KM
		4 - 7 KM
		8 -10 KM
		11 - 15 KM
3	WAKTU BANGUN TIDUR	03.00 - 04.30
		04.31 -05.30
		05.31 -06.30

## 2. PEMBAHASAN

Langkah untuk menentukan pohon keputusan dengan menggunakan algoritma C4.5 dengan menggunakan data training sebanyak 15 data yaitu,

- Menyiapkan data training sebanyak 15 data yang digunakan dalam penelitian ini. Data training biasanya diambil dari data primer yang sebelumnya sudah dikelompokkan
- Hitung nilai entropy dan gain

Setelah dilakukan perhitungan entropy dan gain didapat hasil entropy seperti hitungan berikut.

**Tabel 3. Nilai Entropy dan gain untuk menentukan node awal**

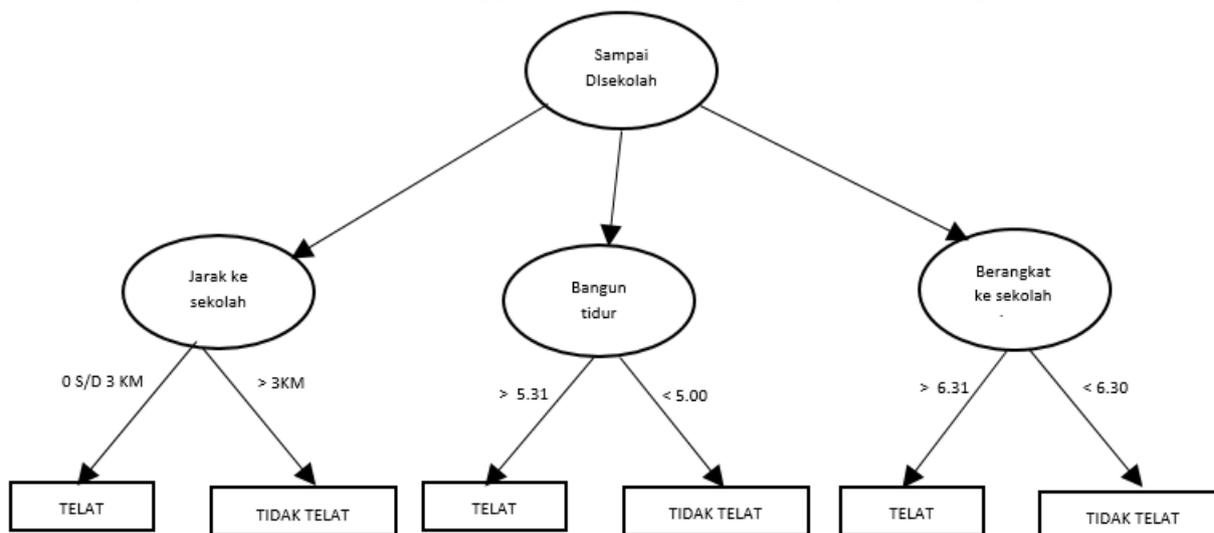
		JUMLAH	SUM(TELAT)	SUM(TIDAK)	ENTROPHY	GAIN
TOTAL		15	3	12	0.721928095	
KE SEKOLAH DENGAN						0.588594762
	SEPEDA	1	1	0	0	
	MOTOR	5	5	0	0	
	DIANTAR SUPIR	6	6	0	0	
	MOBIL	1	1	0	0	
	ONLINE	2	1	1	1	
JARAK KE SEKOLAH						0.598783262
	0 - 3 KM	4	3	1	0.461793122	
	4 - 7 KM	7	0	7	0	
	8 -10 KM	3	3	0	0	
	11 - 15 KM	1	0	1	0	
WAKTU BANGUN TIDUR						0.363208644
	03.00 - 04.30	2	0	2	0	
	04.31 -05.30	10	1	9	0.373374551	
	05.31 -06.30	3	2	1	0.549015419	

Dari tabel 3, dapat dilihat nilai gain tertinggi terdapat pada atribut jarak ke sekolah yakni 0,0598783 sehingga dapat disimpulkan sebagai akar dari pohon keputusan atau node awal. Kemudian dilakukan perhitungan nilai entropy dan gain untuk menentukan node 1.1. nilai yang dihitung berdasarkan sekolah dengan nilai atribut 0-3 km.

**Tabel 4. Nilai entropy dan gain untuk menentukan node 1.1**

		JUMLAH	SUM(TELAT)	SUM(TIDAK)	ENTROPHY	GAIN
KE SEKOLAH DENGAN						-0.038206878
	SEPEDA	1	1	0	0	
	MOBIL	1	1	0	0	
	ONLINE	2	1	1	1	
WAKTU BANGUN TIDUR						0.461793122
	03.00 - 04.30	1	0	1	0	
	04.31 -05.30	1	1	0	0	
	05.31 -06.30	2	2	0	0	
SAMPAI DI SEKOLAH						0.461793122
	05.30 -06.30	1	0	1	0	
	06.31 - 07.00	2	2	0	0	
	07.00 -07.30	1	1	0	0	

Setelah dilakukan perhitungan menggunakan entropy dan gain ternyata hanya cukup dilakukan satu kali sudah mendapatkan nilai node terakhir, sehingga akan menghasilkan pohon keputusan sebagai berikut:



**Gambar 2. Keputusan perhitungan manual**

Selain itu penulis juga menggunakan aplikasi Rapid miner untuk mencari perbandingan hasil pohon keputusan yang dihasilkan.



Gambar 3. Pohon keputusan menggunakan Rapidminer dan algoritma C4.5

### 3. KESIMPULAN

Dari hasil percobaan pencarian hasil pohon keputusan dari data kedatangan siswa menggunakan algoritma C4.5 di hasilkan nilai Entrophy dan Gain tertinggi yaitu 0.46179 terdapat pada atribut bangun tidur dan sampai di sekolah menggunakan cara manual.

Sedangkan menggunakan aplikasi rapid miner dihasilkan keputusan siswa sering telat datang upacara ke sekolah sebanyak 13 %, karena mereka bangun tidur lebih dari jam 6:17. Serta disebabkan juga jarak lokasi rumah ke sekolah lebih dekat 3.25 km atau sekitar 20%.

Saran untuk penelitian selanjutnya, guna memperbaiki kondisi keterlambatan siswa maka diperlukan kedisiplinan dari para siswa agar bisa bangun tidur kurang dari jam 6.17, serta bisa juga dihasilkan hasil keputusan yang lebih baik dengan menggunakan .

### PUSTAKA

- Ajay Kumar Pal, Saurabh Pal, 2013, "Analysis and Mining of Educational Data for Predicting the Performance of Students", Sai Nath University, Ranchi, Jharkhand, India.
- Babak Dalvand dan Gholamreza, 2014, "Using C\$.5 Algorithm for Predicting Efficiency Score of DMUS in DEA", Departemen of Mathematics, Islamic Azad UNiversity, Techrn, Iran.
- D.Lavanya and Dr. K.Usha Rani, 2011, "Performance Evaluation of Decision Tree Classifiers on Medical Datasets", Andhra Pradesh, India.
- Dafiaghor, Kose Famous, 2011, "Lateness: A major problem confronting school administrators in Delta State, Nigeria", Delta State University, Abraka, Delta State, Nigeria.
- Dennis, Aprilia, Donny Aji Baskoro, Lia Ambarwati dan I Wayan Simri Wicaksana. 2013. *Belajar Data Mining dengan Rapid Miner*, Jakarta:Gramedia Pustaka Utama.
- Fandy Ferdian Harryanto, 2017 "Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE", Universitas Multimedia Nusantara, Tangerang.
- Hamidah Jantan dan Abdul Razak Hamdan, 2014, "Human Talent Prediction in HRM using C4.5 Classification Algorithm", Faculty of Information Science and Technology Universiti Kebangsaan Malaysia (UKM)
- Han, J. and Kamber, M., (2006) "Data Mining: Concepts and Techniques, Elsevier". San Francisco: Elsevier Publishing Co.
- Kalpesh Adhatrao, Aditya Gaykar, Amiraj Dhawan, Rohit Jha and Vipul Honrao, 2013, "Predicting Students' Performance Using ID3 and C4.5 Classification Algorithms", Department of Computer Engineering, India.
- Kusrini dan Emha Taufiq Luthfi. 2009. *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- Larose, 2005, "Discovering Knowledge in Data: An Introduction to Data Mining", John Willey & Sons, Inc.
- Prasetyo, Bambang, Lina Miftahul Jannah, 2011. *Metode Penelitian Kuantitatif: Teori dan aplikasi*, Jakarta: PT Raja Grafindo Persada, 2011