

Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam *Decision Tree* dan Algoritma C4.5

Nur Yanti Lumban Gaol[✉]

STMIK Triguna Dharma

ryanti2918@gmail.com

Abstract

Non-active students are students who do not attend the lecture process and do not pay tuition administration fees within two semesters or more. Reports on students who are not active will have an impact on the quantity of tertiary institutions. Students who are not registered in non-active students will potentially be expelled or dropped out. For this reason, this research was conducted to explore information on potentially non-active students by applying data mining science with the Decision Tree method and C4.5 algorithm. The tested data were sourced from Triguna Dharma Medan College of Information and Computer Management (STMIK). The results of the study get prediction rules for student data that are potentially non-active with a very good degree of accuracy. So this research can be used to avoid students dropping out unilaterally.

Keywords: Data Mining, Decision Tree, C4.5 Algorithm, Rapid Miner, Non-active students, *Drop out*.

Abstrak

Mahasiswa non aktif adalah mahasiswa yang tidak mengikuti proses perkuliahan dan tidak membayar biaya administrasi kuliah dalam waktu dua semester atau lebih. Laporan tentang mahasiswa yang tidak aktif akan berdampak terhadap kuantitas perguruan tinggi. Mahasiswa yang tidak terdata dalam mahasiswa non aktif akan berpotensi untuk dikeluarkan atau *drop out*. Untuk itu dilakukan penelitian ini untuk menggali informasi terhadap mahasiswa yang berpotensi non aktif dengan penerapan ilmu data mining dengan metode *Decision Tree* dan algoritma C4.5. Data yang diuji bersumber dari Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Triguna Dharma Medan. Hasil penelitian mendapatkan rule prediksi data mahasiswa yang berpotensi non aktif dengan tingkat keakurasian yang sangat baik. Sehingga penelitian ini dapat digunakan untuk menghindari mahasiswa *drop out* secara sepihak.

Kata kunci: Decision Tree, Algoritma C4.5, Rapid Miner, Mahasiswa non aktif, *Drop out*.

© 2020 JiDT

1. Pendahuluan

Mahasiswa non aktif merupakan mahasiswa tidak mengikuti perkuliahan paling lama 2 (dua) semester berturut-turut. Mahasiswa non aktif memiliki kemungkinan untuk dikeluarkan (*drop out*) pada sebuah perguruan tinggi. Mahasiswa yang berpotensi untuk di *drop out* dari kampus biasa dilatarbelakangi oleh terhentinya proses perkuliahan. Alasan utama terjadinya ketidakaktifan tersebut adalah tidak melakukan pembayaran biaya administrasi.

Banyaknya mahasiswa yang dikeluarkan akan berdampak terhadap mutu perguruan tinggi. Untuk itu pihak perguruan tinggi harus melakukan monitoring dan evaluasi dalam pelaksanaan sistem pembelajaran secara berkala dan mendata keaktifan mahasiswanya. Pendataan ini akan meningkatkan mutu sebuah perguruan tinggi [1].

Untuk menghindari terjadinya pemberhentian mahasiswa, maka perlu pendataan mahasiswa yang tidak aktif dan memprediksi akan terjadinya ketidakaktifan proses perkuliahan. Hal yang melatarbelakangi proses analisa terhadap

permasalahan tersebut akan dapat diketahui faktor penyebabnya.

Berdasarkan hal tersebut perlu dilakukan penelitian untuk memprediksi mahasiswa yang berpotensi non aktif. Metode yang digunakan untuk mengolah datanya adalah *Decision Tree* dengan algoritma C4.5.

Decision Tree merupakan sebuah *Knowledge Discovery In Database (KDD)*. Hasil dari metode ini akan mendapatkan sebuah pengetahuan dari database yang diolah. Database akan diekstraksi sehingga menghasilkan informasi atau pengetahuan yang berguna [2]. Tahap-tahapan yang dilakukan dalam KDD adalah data *mining*. Data *mining* merupakan suatu algoritma di dalam menggali informasi berharga yang terpendam (tersembunyi) pada koleksi data (*database*) yang sangat besar, sehingga ditemukan suatu pola menarik yang sebelumnya tidak diketahui [3]. Salah satu algoritma yang digunakan dalam data *mining* adalah algoritma C4.5.

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan dan aturan – aturan (*rule*). Algoritma C4.5 memetakan nilai atribut

menjadi kelas yang dapat diterapkan untuk klasifikasi baru [4].

Beberapa penelitian yang menggunakan algoritma C4.5 diantaranya adalah Lorenza dkk (2014) menggunakan Algoritma C4.5 untuk memprediksi masa studi mahasiswa berdasarkan data nilai akademik. Hasil yang didapatkan melalui penelitian tersebut memberikan pengaruh besar terhadap nilai persentase kecocokan dan keakurasian dalam memprediksi masa studi mahasiswa berdasarkan data nilai akademik [5]. Harryanto & Hansun (2017) menggunakan algoritma C4.5 untuk memprediksi penerimaan calon pegawai baru pada PT.WISE, serta memberikan hasil pengujian terhadap keakurasian proses prediksi kepada calon pegawai [6]. Penelitian Melissa & Oetama (2013) menggunakan data *mining* untuk melakukan analisis data pembayaran kredit nasabah bank. Hasil penelitian dapat mengoptimalkan prediksi pembayaran kredit nasabah bank [7].

Penelitian Novriansyah dkk (2017) menggunakan data *mining* dan algoritma Naive Bayes untuk mengetahui minat beli pelanggan terhadap kartu internet XL. Hasil didapatkan bahwa metode klasifikasi dan algoritma Naive Bayes sangat cocok dipakai untuk memberikan prediksi yang dapat dipakai di masa depan [8].

2. Metodologi Penelitian

Penelitian ini menggunakan teknik klasifikasi yang terbagi menjadi pohon keputusan (*Decesion Tree*). Metode klasifikasi yang digunakan untuk proses permodelan dalam menggambarkan dan membedakan kelas data atau konsep label kelas yang tidak diketahui [9]. Pohon keputusan merupakan salah satu metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Tujuan klasifikasi ini untuk menemukan model dari

training set yang membedakan atribut ke dalam kategori atau kelas yang sesuai [10]. Pohon keputusan digunakan untuk mengubah fakta yang sangat besar untuk mempresentasikan aturan.

Algoritma C4.5 merupakan kelompok algoritma *decision tree* yang ditemukan oleh Quinlan [11]. Dalam membangun pohon keputusan memilih beberapa tahapan, yaitu 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 [12]. Algoritma mempunyai *input* berupa *training samples* dan *samples*. *Training samples* adalah berupa data contoh yang digunakan untuk membangun sebuah *tree* yang telah diuji kebenarannya. Sedangkan *samples* merupakan *field-field* data yang nantinya akan digunakan sebagai parameter dalam melakukan klasifikasi data.

3. Hasil dan Pembahasan

Data yang dianalisa adalah data mahasiswa yang aktif dan non aktif pada Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Triguna Dharma Medan. Data diklasifikasikan dengan menggunakan proses data *mining*, *Decision Tree* dan algoritma C4.5. Pengolahan data mahasiswa non aktif terdiri dari 4 kriteria yang disajikan pada tabel 1 dan data sampel disajikan pada tabel 2.

Tabel 1. Tabel Data Kriteria Mahasiswa Berpotensi non aktif

No	Data Variabel
1	Jadwal Kuliah
2	Nilai Absensi
3	Nilai Gagal
4	Pembayaran Uang Kuliah

Tabel 2. Tabel Sample Data Mahasiswa

No	Nama	Alamat	Tanggal Lahir	Agama	Jadwal Kuliah	Absen	Nilai Gagal	Pembayaran Uang Kuliah	Status	Absen
1	Santi Gandhi	Kelurahan Kampung Lalang	28/06/1997 1998	Islam	sore	4	1	Menunggak	Non aktif	4
2	Andre Pradana	Jl. Resmi	05/10/1996 1994	Islam	Sore	4	0	Tidak Menunggak	Aktif	4
3	Reza Arielkhan	Jl. Budi Penerbangan No. 10 Kel.	15/03/1996 1996	Islam	Sore	2	0	Tidak Menunggak	Aktif	2
4	Dimas Pamungkas	Jl. TSutrisno	09/05/1993 1997	Islam	Pagi	12	11	Menunggak	Non aktif	12
5	Jaka	Jl. A. H. Nasutio Gg. Rapi Medan	17/09/1994 1995	Kristen	Sore	4	2	Tidak Menunggak	Aktif	4
6	Anisa	Jl. SM. Raja Simp. Limun No. 89	24/04/1996 1995	Kristen	Sore	3	0	Tidak Menunggak	Aktif	3
7	Eva Kamenia Pakpahan	Jl. Melayu Desa Tumpatan Kec. Beringin	13/07/1997 1998	Kristen	Sore	2	1	Tidak Menunggak	Aktif	2

Banyak data yang diuji berjumlah 57 orang mahasiswa, tetapi yang disajikan hanya 7 orang mahasiswa. Selanjutnya data tersebut dilakukan proses *cleaning* dan transformasi data. Data mahasiswa yang aktif dan non aktif akan dilakukan proses transformasi. Terlebih dahulu diklasifikasikan sesuai dengan peraturan dan ketentuan regulasi kampus. Klasifikasi pertama yaitu jadwal kuliah. Klasifikasi ini disajikan pada tabel 3. kedua yaitu nilai absensi, yang disajikan pada tabel 4, ketiga yaitu nilai gagal yang disajikan pada tabel 5, keempat yaitu status pembayaran uang kuliah yang disajikan pada tabel 6.

Tabel 3. Klasifikasi Jadwal Kuliah

No	Jadwal Kuliah	Klasifikasi
1	Kelas A	Pagi
2	Kelas B	Siang
3	Kelas C	Sore

Tabel 4. Klasifikasi Nilai Absensi

No	Nilai Absensi	Klasifikasi
1	0 – 4	Aman
2	5 – 8	SP
3	>8	Mengulang

Tabel 5. Klasifikasi Nilai

No	Nilai Gagal	Klasifikasi
1	≤ 5	Aman
2	6 – 10	SP
3	>10	Mengulang

Tabel 6. Klasifikasi Uang Kuliah

No	Pembayaran Uang Kuliah	Klasifikasi
1	Tepat waktu	Tidak Menunggak
2	Tidak Tepat Waktu	Menunggak

Selanjutnya dilakukan proses transformasi terhadap sampel data yang disesuaikan dengan Tabel klasifikasi yang telah ada. Hasil transformasi data sampel disajikan pada tabel 7.

Tabel 7. Sampel Data Cleaning dan Transformasi

No	Jadwal Kuliah	Absen	Nilai	Status Pembayaran Uang Kuliah	Status
1	Sore	Aman	Aman	Menunggak	Non aktif
2	Sore	Aman	Aman	Tidak Menunggak	Aktif
3	Sore	Aman	Aman	Tidak Menunggak	Aktif
4	Pagi	Mengulang	Aman	Menunggak	Non aktif
5	Sore	Aman	Aman	Tidak Menunggak	Aktif
6	Sore	Aman	Aman	Tidak Menunggak	Aktif
7	Sore	SP	Aman	Tidak Menunggak	Aktif

3.1. Mencari Entropy Total

Mencari nilai entropy total untuk dijadikan sebagai node 1. Jumlah data yang digunakan berjumlah 107,

dimana data mahasiswa aktif sebanyak 78 orang dan non aktif 29 orang mahasiswa. Pada tabel 8 dilakukan pengelompokan data mahasiswa.

Tabel 8. Pengelompokan Data Mahasiswa

Node 1	Jumlah Kasus	Aktif	Non aktif
TOTAL	107	78	29
Jadwal			
Pagi	30	16	14
Sore	77	62	15
Absensi			
Aman	81	70	11
SP	13	8	5
Mengulang	13	0	13
Nilai gagal			
Aman	93	71	22
SP	8	7	1
Mengulang	6	0	6
Pembayaran			
Menunggak	15	0	15
Tidak Menunggak	92	78	14

Setelah data yang ditransformasikan dikelompokkan, selanjutnya mencari nilai entropy dari semua atribut untuk dijadikan entropy total.

Perhitungan Entropy dan Gain pada node 1 :

$$= (-78/107 * \log_2 78/107) + (-29/107 * \log_2 29/107) = 0,84293.$$

3.2. Mencari Entropy dan Gain pada semua atribut

Setelah menghitung nilai Entropy total, berikutnya menghitung nilai setiap atribut berdasarkan klasifikasi yang ditentukan. Klasifikasi pada atribut jadwal kuliah terdapat 2 klasifikasi yaitu pagi dan sore.

- Entropy jadwal kuliah (pagi)

$$= (-16/30 * \log_2 16/30) + (-14/30 * \log_2 14/30) = 0,99679$$

- Entropy jadwal kuliah (sore)

$$= (-62/77 * \log_2 62/77) + (-15/77 * \log_2 15/77) = 0,71141$$

Selanjutnya menghitung nilai *gain* untuk atribut jadwal kuliah, dan nilai Gain tertinggi akan menjadi node 1.

- Information Gain (total, jadwal kuliah)

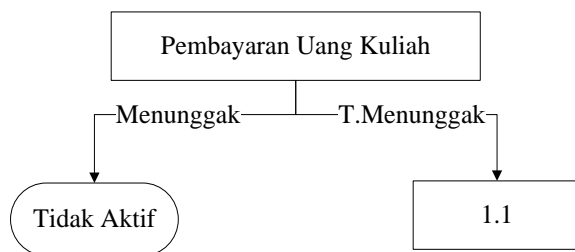
$$= 0,842936 - ((30/107 * 0,996792) + (77/107 * 0,711416)) = 0,05150$$

Setelah menghitung nilai Entropy dan nilai Gain untuk jadwal kuliah, selanjutnya lakukan hal yang sama untuk kembali menghitung nilai Entropy dan nilai Gain untuk atribut yang lainnya. Tabel 9 adalah tabel hasil perhitungan manual untuk nilai Entropy dan nilai Gain yang dipakai untuk menentukan node akar.

Tabel 8. Pengelompokan Data Mahasiswa

Perhitungan Node 1	Jumlah Kasus	Aktif	Non Aktif	Entropy	Gain
Total	107	78	29	0,842	
Jadwal Kuliah					0,051
Pagi	30	16	14	0,996	
Sore	77	62	15	0,711	
Absensi					0,292
Aman	81	70	11	0,573	
SP	13	8	5	0,961	
Mengulang	13	0	13	0	
Nilai Gagal					0,116
Aman	93	71	22	0,789	
SP	8	7	1	0,543	
Mengulang	6	0	6	0	
Pembayaran Uang Kuliah					0,313
Menunggak	15	0	15	0	
Tidak Menunggak	92	78	14	0,615	

Dari tabel 8 diketahui bahwa atribut dengan gain tertinggi adalah pembayaran uang kuliah sebesar 0,313. Hasil pohon keputusan pada gambar 1.



Gambar 1. Pohon Keputusan Node 1

3.2. Menentukan Node cabang

Setelah mendapatkan node 1, selanjutnya mencari atribut mana yang akan menjadi cabang level 1. Langkah - langkah untuk menghitung cabang level 1 adalah dengan menghitung kembali Entropy atribut pembayaran uang kuliah dengan klasifikasi tidak menunggak.

- Entropy pembayaran uang kuliah (tidak menunggak)

$$= (-78/92 * \log_2 78/92) + (-14/92 * \log_2 14/92) = 0,61525$$

Atribut yang masih dapat digunakan sebagai cabang level 1 yaitu jadwal kuliah, nilai gagal dan nilai absensi. Atribut jadwal kuliah terdapat 2 klasifikasi yaitu pagi dan sore.

• Entropy jadwal kuliah (pagi)

$$= (-16/24 * \log_2 16/24) + (-8/24 * \log_2 8/24) = 0,91829$$

• Entropy jadwal kuliah (sore)

$$= (-62/68 * \log_2 62/68) + (-6/68 * \log_2 6/68) = 0,43055.$$

Selanjutnya menghitung nilai gain untuk atribut jadwal kuliah. Nilai Gain akan digunakan sebagai perbandingan dengan atribut lainnya. Nilai Gain tertinggi akan menjadi cabang level 1.

• Information Gain (total, jadwal kuliah)

$$= 0,615254 - ((24/92 * 0,91829) + (68/92 * 0,43055)) = 0,05746$$

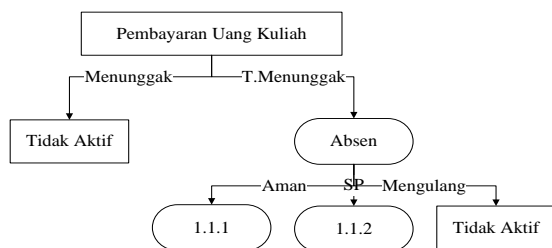
Selanjutnya lakukan hal yang sama untuk kembali menghitung nilai entropy dan nilai gain untuk atribut yang lainnya. tabel 9 adalah tabel untuk mendapatkan pohon keputusan Node 1.1, dapat menggunakan nilai Gain yang tertinggi.

Tabel 9. Hasil Perhitungan Node 1.1

Perhitungan Node 1.1	Jumlah Kasus	Aktif	Non Aktif	Entropy	Gain
Total	92	78	14	0,615	
Jadwal Kuliah					0,051
Pagi	24	16	8	0,918	
Sore	68	62	6	0,430	
Absensi					0,369
Aman	73	70	3	0,247	
SP	9	8	1	0,503	
Mengulang	10	0	10	0	
Nilai Gagal					0,140
Aman	81	71	10	0,539	
SP	7	7	0	0	
Mengulang	4	4	0	0	

Berdasarkan tabel diatas diketahui bahwa atribut dengan Gain tertinggi adalah atribut nilai absensi sebesar 0,369, maka nilai absensi dijadikan sebagai Node 1.1. Dari hasil tersebut didapatlah pohon keputusan seperti pada gambar 2.

Setelah mendapatkan node 1.1, selanjutnya dilanjutkan dengan perhitungan nilai Entropy pada nilai absensi (aman), Entropy jadwal kuliah (pagi), Entropy jadwal kuliah (sore) untuk mencari cabang Level 1.1.1 seperti proses pada tahapan sebelumnya. Hingga didapatlah hasil perhitungan untuk dijadikan sebagai Node 1.1.1 seperti pada tabel 10.



Gambar 2. Pohon Keputusan Node 1.1

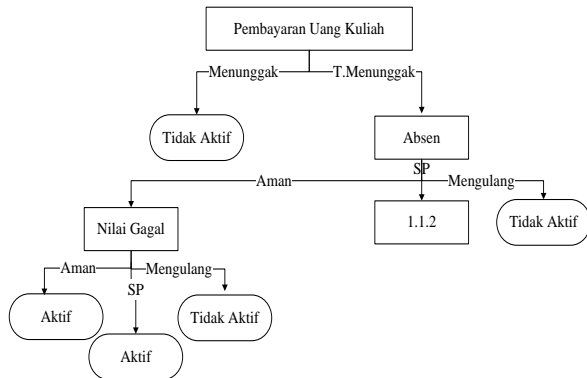
T

Tabel 10. Hasil Perhitungan Node 1.1.1

Perhitungan Node 1.1	Jumlah Kasus	Aktif	Non Aktif	Entropy	Gain
Total	73	70	3	0,247	
Jadwal Kuliah					0,010
Pagi	12	12	0	0,918	
Sore	61	58	3	0,282	
Nilai Gagal					0,247
Aman	63	63	0	0	
SP	7	7	0	0	
Mengulang	3	0	3	0	

Berdasarkan tabel diatas diketahui bahwa atribut dengan gain tertinggi adalah atribut nilai gagal sebesar 0,247, maka nilai gagal dijadikan sebagai Node 1.1.1.

Dari hasil tersebut didapatkan pohon keputusan seperti pada Gambar 3.



Gambar 3. Pohon Keputusan Node 1.1.1

Selanjutnya menghitung nilai Entropy pada Node 1.1.2. Klasifikasi nilai aman dan mengulang tidak perlu lagi

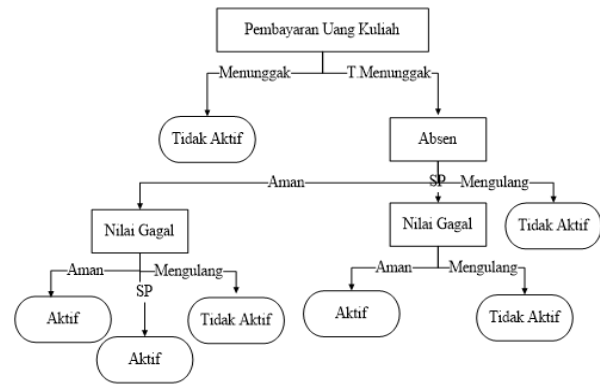
dihitung karena sudah bernilai 0. Selanjutnya yang dihitung adalah nilai Entropy pada atribut yang lain,

dengan menggunakan nilai dari atribut nilai gagal dengan klasifikasi aman sebagai nilai Entropy total untuk mendapatkan Node 1.1.2 pohon keputusan Node 1.1.2 terlihat pada Tabel 11.

Tabel 11. Hasil Perhitungan Node 1.1.2

Perhitungan Node 1.1.2	Jumlah Kasus	Aktif	Non Aktif	Entropy	Gain
Nilai Absensi (SP)	9	8	1	0,503	
Jadwal Kuliah					0,010
Pagi	5	4	1	0,721	
Sore	4	4	0	0	
Nilai Gagal					0,503
Aman	8	8	0	0	
SP	0	0	0	0	
Mengulang	1	0	1	0	

Berdasarkan tabel diatas diketahui bahwa atribut dengan Gain tertinggi adalah atribut nilai gagal sebesar 0,503. Dengan demikian atribut nilai absen dijadikan Node 1.1.2. Dari hasil tersebut didapatkan pohon keputusan seperti pada Gambar 4.



Gambar 4. Pohon Keputusan Node 1.1.2

Setelah proses perhitungan selesai, maka didapatkan sebuah pohon keputusan dengan aturan (*rule*) yaitu :

- IF “Pembayaran Uang Kuliah “Menunggak” THEN “Non aktif”
- IF “Pembayaran Uang Kuliah “Tidak Menunggak” AND Absen “Aman” AND Nilai “Aman” THEN “Aktif”
- IF “Pembayaran Uang Kuliah “Tidak Menunggak” AND Absen “Aman” AND Nilai “SP” THEN “Aktif”
- IF “Pembayaran Uang Kuliah Tidak Menunggak” AND Absen “ Aman” AND Nilai “Mengulang ” THEN “Non aktif”
- IF “Pembayaran Uang Kuliah Tidak Menunggak” AND Absen “SP” AND Nilai “Aman” THEN “Aktif”
- IF “Pembayaran Uang Kuliah Tidak Menunggak” AND Absen “SP” AND Nilai “Mengulang” THEN “ Non aktif”
- IF “Pembayaran Uang Kuliah Tidak Menunggak” AND Absen “Mengulang” THEN “ Non aktif”

4. Kesimpulan

Berdasarkan penelitian diatas disimpulkan bahwa rule – rule yang dihasilkan menghasilkan kriteria-kriteria yang tepat dalam menganalisa mahasiswa yang berpotensi non aktif. Kriterianya adalah jadwal kuliah, nilai absensi, nilai gagal, pembayaran uang kuliah yang dapat mempengaruhi keputusan pada status mahasiswa aktif dan non aktif.

Daftar Rujukan

[1] Departemen Pendidikan Nasional, (2007) *Buku I Naskah Akademik Akreditasi Institusi Perguruan Tinggi*. Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT) .

[2] Artha K.J., Indrawan G & Dantes D.R. (2016), *Data Mining Rekomendasi Calon Mahasiswa Berprestasi di STMIK Denpasar Menggunakan Metode Technique For Others Preference By Similarity to Ideal Solution*. *Jurnal Sains dan*

- Teknologi*, 15(2), 2303-3142. <http://dx.doi.org/10.23887/jst-undiksha.v5i2.8549>.
- [3] Elmande Y & Widodo P. (2012), **Pemilihan Criteria Splitting dalam Algoritma Iterative Dichotomiser 3 (ID3) untuk Penentuan Kualitas Beras Perum Bulog Drive Lampung**. *Jurnal Telematika MKOM*, 06(1), 2085-7250.
- [4] Novandya A, (2017), **Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi**. *Konferensi Nasional Ilmu Sosial dan Teknologi*. Vol 1, no 1.
- [5] Lorena, S., Zarman, W., & Hamidah, I. (2014). **Analisis Dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Memprediksi Masa Studi Mahasiswa Berdasarkan Data Nilai Akademik**. *Prosiding Seminar Nasional Aplikasi Sains Dan Teknologi (SNAST)*, (November), 263–272.
- [6] Harryanto, F. F., & Hansun, S. (2017). **Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE**. *Jurnal Teknik Informatika Dan Sistem Informasi*, 3(2), 95–103.
- [7] Melissa, I., & Oetama, R. S. (2013). **Analisis Data Pembayaran Kredit Nasabah Bank Menggunakan Metode Data Mining**. *Jurnal ULTIMA InfoSys*, 4(1), 18–27. <https://doi.org/10.31937/si.v4i1.238>.
- [8] Novriansyah D., Erwansyah K., & Ramadhan M. (2016), **Penerapan Data Mining dengan Algoritma Naive Bayes Clasifier untuk Mengetahui Minat Beli Pelanggan terhadap Kartu Internet XL (Studi Kasus di CV. Sumber Utama Telekomunikasi)**. *Jurnal SAINTIKOM*, 15(2).
- [9] Utama T.D., Sari W S, & Afrizal D., (2014), **Implementasi Iterative Dichotomiser 3 Pada Penyeleksian Program Mahasiswa Wirausaha UNS**. *Jurnal ITS MART*, 3.
- [10] Rhamadianoor & Boy A.R., (2015), **Sistem Pendukung Keputusan Penerima Beasiswa Dengan Metode Decision Tree**. *Jurnal Ilmiah Pendidikan*, 12 (2).
- [11] Lakshmi, B. N., Indumathi, T. S., & Ravi, N. (2016). **An Hybrid Approach for Prediction Based Health Monitoring in Pregnant Women**. *Procedia Technology*, 24, 1635–1642. <https://doi.org/10.1016/j.protcy.2016.05.171>.
- [12] Kusrini & Lutfhi E.T, (2019). **Algoritma dan Data Mining**. Yogyakarta: Andi Publisher.