Penerapan Algoritma C4.5 dan *K-Nearest Neighbor* untuk Klasifikasi Peminatan Program Studi di Perguruan Tinggi Berdasarkan Nilai Rapor

Mulya Cahya Ramadanty
Universitas Buana Perjuangan
Karawang, Indonesia
if16.mulyaramadanty
@mhs.ubpkarawang.ac.id

Amril Mutoi Siregar
Universitas Buana Perjuangan
Karawang, Indonesia
amrilmutoi@ubpkarawang.ac.id

Dwi Sulistya Kusumaningrum
Universitas Buana Perjuangan
Karawang, Indonesia
dwi.sulistya@ubpkarawang.ac.id

***Abstract*—**
Setelah lulus Sekolah Menengah Kejuruan (SMK) siswa yang melanjutkan pendidikan ke Perguruan Tinggi sering merasa kebingungan dengan program studi yang akan mereka ambil. Berdasarkan hasil penelitian pada tahun 2017 yang dikutip oleh Murti, sebanyak 92% siswa SMA sederajat merasa bingung dan tidak tahu akan menjadi apa ke depannya. Pada penelitian lainnya ditemukan 87% mahasiswa Indonesia mengakui bahwa jurusan yang mereka ambil tidak sesuai dengan minatnya. Tujuan dari penelitian ini adalah Mengimplementasikan algoritma C4.5 dan K-*Nearest Neighbor* (K-NN) dalam klasifikasi peminatan program studi. Algoritma yang digunakan pada penelitian ini yaitu C4.5 dan K-NN. Data yang digunakan adalah nilai rapor Matematika dan mata pelajaran produktif siswa kelas XII jurusan Teknik Komputer Jaringan (TKJ), Teknik Elektronika Industri (TEI), dan Rekayasa Perangkat Lunak (RPL) Sekolah Menengah Kejuruan Negeri (SMKN) 1 Karawang. Hasil yang didapat dari pengujian menggunakan *tool* RapidMiner sebesar 98,04% untuk algoritma K-NN dan 100% untuk algoritma C4.5. Pada tahap implementasi algoritma K-NN ke program diperoleh hasil sebesar 98%.
Kata kunci — *Peminatan, Klasifikasi, Data, C4.5, K-Nearest Neighbor*

I. PENDAHULUAN

Siswa Sekolah Menengah Kejuruan (SMK) sejak awal masuk sudah diwajibkan untuk memilih jurusan sesuai dengan minat atau bakatnya. Pada SMK Teknik siswa akan diberikan pilihan berbagai macam jurusan seperti Teknik Komputer Jaringan (TKJ), Rekayasa Perangkat Lunak (RPL), Teknik Gambar Bangunan (TGB), Teknik Elektronika Industri (TEI) dan lain sebagainya. Namun, tidak semua siswa SMK setelah lulus dan melanjutkan pedidikan akan memilih jurusan yang sama dengan yang mereka pilih saat di SMK. Contohnya siswa SMK yang memilih program studi Teknik Informatika yang mana berbeda dengan jurusan yang dipilih saat di SMK. Hal tersebut tentunya dapat mempersulit siswa mempelajari materi yang ada dalam program studi tersebut. Berdasarkan hasil penetilian Indonesia *Career Center Network* (ICCN) tahun 2017 yang dikutip oleh Murti, terdapat 87% mahasiswa Indonesia mengakui bahwa jurusan yang mereka ambil tidak sesuai dengan minatnya. Ditemukan juga fakta bahwa 92% siswa SMA sederajat merasa bingung dan tidak tahu akan menjadi apa ke depannya [1].

Penelitian tentang klasifikasi telah banyak ditemukan, beberapa diantaranya penelitian yang dilakukan oleh Novianti, Rismawan dan Bahri [2] tentang implementasi algoritma C4.5 untuk penjurusan siswa. Data yang digunakan yaitu nilai tes, nilai rata-rata rapor, dan nilai Ujian Nasional (UN). Hasil yang didapat dari penelitian tersebut yaitu nilai akurasi algoritma C4.5 sebesar 89,74%. Kemudian penelitian yang dilakukan oleh Siregar dan Fauzi [3] yang berjudul klasifikasi kabupaten kota provinsi jawa barat berdasarkan pendapatan dari sektor pertanian. Algoritma yang digunakan yaitu *decision tree*. Data yang digunakan yaitu *database* dari pemerintah bagian *statistic* di bidang pertanian, dan menggunakan 16 atribut. Penelitian tersebut menghasilkan nilai akurasi sebesar 90%. Penelitian lainnya oleh Sambani dan Nuraeni [4] menggunakan algoritma C4.5 tentang klasifikasi pola penjurusan di Sekolah Menengah Kejuruan (SMK). Atribut yang digunakan yaitu pilihan jurusan I, pilihan jurusan II, nilai rapor matematika, bahasa indonesia, bahasa inggris, Ilmu Pengetahuan Alam (IPA), tes kesehatan, dan tes olahraga. Dari penelitian tersebut diperoleh klasifikasi pola penjurusan SMK dengan akurasi 97,22%. Kemudian penelitian tentang seleksi karyawan pada perusahaan tekstil menggunakan algoritma C4.5 yang dilakukan oleh Amalia, Siregar dan Lestari [5]. Atribut yang digunakan yaitu usia, status, nilai tes, tinggi badan, pendidikan, sehat, dan jahit. Penelitian tersebut menghasilkan nilai akurasi dan nilai sensitivitas sebesar 100%. Selanjutnya penelitian dari Nurjanah, Siregar dan Kusumaningrum [6] tentang klasifikasi pencemaran udara di kota Jakarta. Algoritma yang digunakan yaitu K-*Nearest Neighbor* (K-NN). Data yang digunakan diperoleh dari Dinas Lingkungan Hidup kota Jakarta tahun 2018 yang terdiri dari 304 data dan 7 atribut. Hasil akurasi yang diperoleh dari penelitian tersebut yaitu 95,78% dengan menentukan K=7. Kemudian penelitian lainnya tentang Prediksi hasil produksi oleh Utari, Siregar dan Wahiddin [7] menggunakan algoritma K-*Nearest Neighbor* (K-NN). Data yang digunakan pada penelitian ini yaitu sebanyak 130 data dengan 9 atribut. Dari penelitian tersebut diperoleh hasil akurasi sebesar 100% dengan menentukan K=5.

Berdasarkan hasil yang diperoleh dari penelitian [2,3,4,5,6,7] menggunakan algoritma C4.5 dan K-*Nearest Neighbor* (K-NN) terbukti menghasilkan nilai akurasi yang tinggi, maka penelitian ini diberi judul "Penerapan Algoritma C4.5 dan K-*Nearest Neighbor* untuk Klasifikasi Peminatan Program Studi di Perguruan Tinggi Berdasarkan Nilai Rapor". Inti dari penelitian ini yaitu mengelompokan nilai siswa yang memenuhi kriteria program studi Teknik Informatika dan Sistem Informasi berdasarkan nilai rapor.

II. DATA DAN METODE

A. Algoritma C4.5

Algoritma C4.5 diperkenalkan oleh Quinlan sebagai versi perbaikan dari ID3. Dalam ID3, induksi *decision tree* hanya bisa dilakukan pada fitur bertipe kategorikal (nominal atau ordinal), sedangkan tipe numerik (interval atau rasio) tidak dapat digunakan. Perbaikan yang membedakan algoritma C4.5 dan ID3 adalah dapat menangani fitur dengan tipe numerik, melakukan pemotongan (*pruning*) *decision tree*, dan penurunan (*deriving*) *rule se*t. Algoritma C4.5 juga menggunakan kriteria gain dalam menentukan fitur yang menjadi pemecah node pada pohon yang diinduksi [8]. Adapun tahapan algoritma C4.5 sebagai berikut [9].

1. Memasukkan *data training*.
2. Mencari nilai entropy menggunakan rumus persamaan (1):

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

Keterangan:
S : himpunan kasus
A : atribut
n : jumlah partisi S
pi : proporsi dari Si terhadap S

3. Menghitung nilai *Gain* menggunakan rumus persamaan (2):

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

Keterangan:
S : himpunan kasus
A : atribut
n : jumlah partisi atribut A
|Si| : jumlah kasus pada partisi ke-i
|S| : jumlah kasus dalam S

4. Menentukan nilai *Gain* tertinggi dari setiap atribut.
5. Menentukan akar dari pohon keputusan.
6. Ulangi langkah kedua hingga semua tupel terpartisi

B. Algoritma K-*Nearest Neighbor* (K-NN)

Algoritma K-*Nearest Neighbor* adalah algoritma yang sering digunakan untuk klasifikasi teks dan data. Tujuannya adalah mengklasifikasikan obyek berdasarkan atribut dan *training sample*. Algoritma K-*Nearest Neighbor* (K-NN) menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari *query instance* yang baru [10]. Adapun langkah-langkah algoritma K-NN sebagai berikut [11]:

1. Tentukan parameter K (dipilih secara manual)
2. Hitung jarak antara data yang akan dievaluasi dengan semua data pelatihan menggunakan rumus *Euclidean Distance*, dapat dilihat pada persamaan (3):

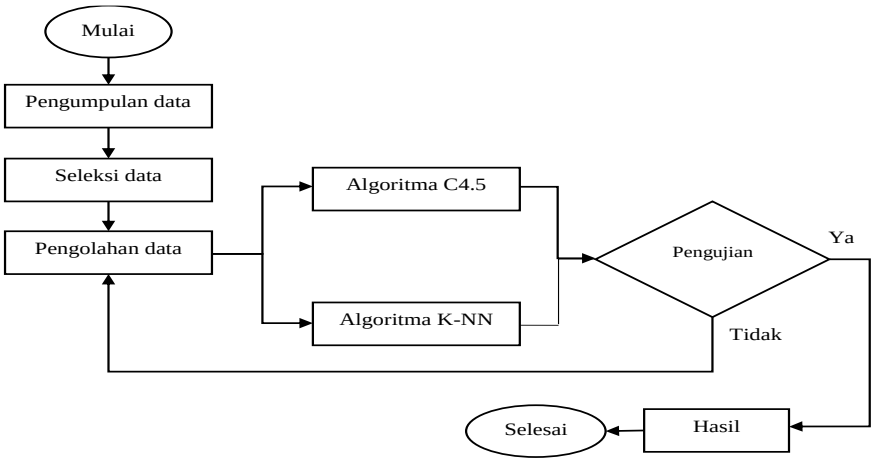
$$dij = \sqrt{[(x_i - x_j)^2 + (y_i - y_j)^2]} \quad (3)$$

Keterangan :
X_1 : Sampel data
X_2 : Data uji atau *testing*
I : Variabel data
d : Jarak
p : Dimensi data

3. Urutkan jarak yang terbentuk (urut naik)
4. Tentukan jarak terdekat sampai urutan K
5. Pasangkan kelas yang bersesuaian
6. Cari jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi.

C. Gambaran Umum Penelitian

Adapun gambaran umum penelitian yang digunakan terdapat pada Gambar 1



Gambar 1 Gambaran Umum Penelitian

D. Pengumpulan Data

Pada tahap pertama penelitian ini dimulai dari pengumpulan data yang diperoleh dari SMK Negeri 1 Karawang. Data yang digunakan yaitu data nilai rapor siswa kelas XII semester 5 dan 6 sebanyak 353 siswa, yang terdiri dari 119 siswa jurusan Teknik Elektronika Industri (TEI), 154 siswa jurusan Teknik Komputer Jaringan (TKJ) dan 80 siswa jurusan Rekayasa Perangkat Lunak (RPL) di tahun ajaran 2018/2019. Terdapat 10 mata pelajaran yang dijadikan atribut yaitu mata pelajaran Pendidikan Agama dan Budi Pekerti, Pendidikan Pancasila dan Kewarganegaraan, Bahasa Indonesia, Matematika, Bahasa Inggris, Administrasi Server, Rancang Bangun Jaringan, Jaringan Nirkabel, Troubleshooting Jaringan dan Sistem Operasi Jaringan.

E. Seleksi Data

Seleksi data dilakukan untuk memfokuskan data hanya pada atribut-atribut yang diperlukan saja. Atribut yang digunakan dalam penelitian ini yaitu Nama Siswa, Nilai rata-rata mata pelajaran Matematika (MTK) semester 5 dan 6, Nilai rata-rata mata pelajaran Administrasi Server (AS) semester 5 dan 6, Nilai rata-rata mata pelajaran Rancang Bangun Jaringan (RBJ) semester 5 dan 6 dan Nilai rata-rata mata pelajaran Jaringan Nirkabel (JN) semester 5 dan 6. Dataset setelah tahap seleksi tidak langsung digunakan, untuk mata pelajaran produktif yaitu Administrasi Server (AS), Rancang Bangun Jaringan (RBJ) dan Jaringan Nirkabel (JN) dibuat atribut baru yang dinamakan mata pelajaran produktif (MP Produktif). Hal tersebut dikarenakan ketiga mata pelajaran tersebut tidak dapat dijadikan tolak ukur seperti mata pelajaran Matematika. Nilai yang digunakan pada atribut mata pelajaran produktif yaitu rata-rata dari nilai mata pelajaran AS, RBJ dan JN. Atribut mata pelajaran Matematika dan MP Produktif terdapat pada Tabel 1.

Tabel 1 Atribut Nilai Matematika dan Mata Pelajaran Produktif

No.	Nama	Nilai Mata pelajaran	
		MTK	MP Produktif
1	Ade Karmila	79,5	84,33
2	Agil Ghalib Septiansyah	79	81,83
3	Ahmad Ruhyatul Aziz	78,5	87,33
4	Amalia Kartika Sandra	78,5	85,17
5	Amelia	79	84,50
6	Arsyi Rayhan Naufal	79,00	83,50
7	Atih Permata Sari	78,50	80,50
...	...	...	...
...	...	...	...
353	Wiwin Sumiyati	80	80,67

III. HASIL DAN PEMBAHASAN

A. Hasil Algoritma K-*Nearest Neighbor* (K-NN)

Proses perhitungan menggunakan algoritma K-NN diperlukan penentuan data *training* dan data *testing*. Pada perhitungan manual penelitian ini digunakan 30 data *training* dan 1 data *testing*. Data *testing* diambil dari dataset nomor urut 1. Data *training* dan data *testing* dapat dilihat pada tabel 2.

Tabel 2 Data *Training* dan Data *Testing*

No.	Nama	Nilai Mata pelajaran		Keterangan
		MTK	MP Produktif	
1	Ade Karmila	79,50	84,33	?
2	Agil Ghalib Septiansyah	79,00	81,83	SI
3	Ahmad Ruhyatul Aziz	78,50	87,33	IF
4	Amalia Kartika Sandra	78,50	85,17	IF
5	Amelia	79,00	84,50	IF
6	Arsyi Rayhan Naufal	79,00	83,50	IF
7	Atih Permata Sari	78,50	80,50	SI
...	...	...	...	...
...	...	...	...	...
31	Muhammad Andhika	82,00	83,50	IF

Menurut Kusumaningrum dan Lestari [12] Matematika Diskrit adalah salah satu matakuliah prasyarat yang memberikan landasan matematis untuk matakuliah-matakulia lain di rumpun ilmu komputer ataupun teknik informatika. Maka, untuk mengetahui keterangan IF dan SI pada tabel 2 yaitu dilihat dari nilai keempat mata pelajaran, terutama Matematika. Apabila nilai Matematika siswa > 80 dikategorikan IF, apabila nilai Matematika < 80 dikategorikan SI, dan apabila nilai Matematika < 80 tetapi rata-rata nilai ketiga mata pelajaran lainnya > 82 maka dapat dikategorikan IF. Data *testing* yang digunakan yaitu dari nilai siswa bernomor urut satu. Tanda tanya pada tabel 2 diartikan sebagai nilai yang akan dicari dan hasilnya akan dicocokkan dengan data asli.

Berdasarkan tahap-tahap algoritma K-NN, setelah menentukan data training dan data testing, menentukan nilai K (K=9) tahap selanjutnya yaitu menghitung jarak antara data *testing* dan seluruh data *training* menggunakan rumus *Euclidean Distance*.

Tabel 3 Perhitungan Jarak Menggunakan *Euclidean Distance*

Nilai Mata Pelajaran		Perhitungan *Euclidean Distance*
MTK	MP Produktif	
79	81,83	$\sqrt{(79-79{,}5)^2+(81{,}83-84{,}33)^2} = 2{,}5462$
78,5	87,33	$\sqrt{(78{,}5-79{,}5)^2+(87{,}33-84{,}33)^2} = 3{,}1654$
78,5	85,17	$\sqrt{(78{,}5-79{,}5)^2+(85{,}17-84{,}33)^2} = 1{,}3038$
79	84,50	$\sqrt{(79-79{,}5)^2+(84{,}50-84{,}33)^2} = 0{,}5281$
79	83,50	$\sqrt{(79-79{,}5)^2+(83{,}50-84{,}33)^2} = 0{,}9690$
...	...	…
...	...	…
82	83,50	$\sqrt{(82-79{,}5)^2+(83{,}50-84{,}33)^2} = 2{,}6342$

Setelah menghitung jarak, selanjutnya mengurutkan hasil perhitungan jarak yang terbentuk menggunakan metode *ranking*, diawali dengan data yang memiliki jarak *Euclidean* terkecil hingga jarak *Euclidean* terbesar.

Tabel 4 Urutan Jarak *Euclidean* Terkecil Hingga Terbesar

MTK	MP Produktif	Jarak (Euclidean Distance)	Rank
79	81,83	2,5462	8
78,5	87,33	3,1654	10
78,5	85,17	1,3038	4
79	84,50	0,5281	2
79	83,50	0,9690	3
...	...	...	...
...	...	...	...

MTK	MP Produktif	Jarak (Euclidean Distance)	Rank
79	88,33	4,0344	17

Setelah diperoleh jarak Euclidean terkecil hingga terbesar, tahap selanjutnya menentukan tetangga terdekat berdasarkan jarak minimum ke-K. Karena K yang telah ditentukan adalah 9, maka apabila hasil perhitungan jarak < 9 dikategorikan Ya, dan > 9 dikategorikan Tidak.

Tabel 5 Kategori Tetangga Terdekat ke-K

MTK	MP Produktif	Jarak (*Euclidean Distance*)	Rank	Termasuk 9-NN
79	81,83	2,5462	8	Ya
78,5	87,33	3,1654	10	Ya
78,5	85,17	1,3038	4	Ya
79	84,50	0,5281	2	Ya
79	83,50	0,9690	3	Ya
...	...	...	...	...
...	...	...	...	...
79	88,33	4,0344	17	Ya

Selanjutnya menentukan kelas mayoritas sebagai klasifikasi objek atau data baru. Setelah menentukan jarak *Euclidean* pada tabel 5 termasuk kategori Ya atau Tidak, selanjutnya menentukan kelas mayoritas.

Tabel 6 Kelas Mayoritas Data Baru

MTK	MP Produktif	Jarak (*Euclidean Distance*)	Rank	Termasuk 9-NN	Kategori Ya untuk K-NN
79	81,83	2,5462	8	Ya	SI
78,5	87,33	3,1654	10	Ya	IF
78,5	85,17	1,3038	4	Ya	IF
79	84,50	0,5281	2	Ya	IF
79	83,50	0,9690	3	Ya	IF
...	...	...	...	...	...
...	...	...	...	...	...
79	88,33	4,0344	17	Ya	IF

Pada tabel 6 data disesuaikan dengan dataset awal. Hasilnya terdapat 29 data yang sesuai dengan dataset, yaitu 26 untuk kategori IF dan 3 untuk kategori SI. Karena dilihat dari kelas mayoritas, maka dapat ditarik kesimpulan bahwa data *testing* termasuk kedalam kategori IF.

Tabel 7 Hasil Klasifikasi K-NN

No.	Nama	Nilai Mata pelajaran		Keterangan
		MTK	MP Produktif	
1	Ade Karmila	79,5	84,33	IF

Hasil akurasi yang diperoleh dari tabel tetangga terdekat yaitu:

$$Accuracy = \frac{\text{Total data benar}}{\text{Total data keseluruhan}} \times 100$$
$$= \frac{26}{30} \times 100\% = 86{,}67\%$$

B. Hasil Algoritma C4.5

Pada perhitungan manual algoritma C4.5 digunakan keseluruhan data, yaitu 353 siswa dari ketiga jurusan. Data yang digunakan untuk algoritma C4.5 tidak sama seperti yang digunakan pada algoritma K-NN. Pada algoritma C4.5 data yang bertipe numerik harus di kategorikan terlebih dahulu untuk memudakan proses klasifikasi.

Tabel 8 Perubahan Tipe Data

Mata Pelajaran	Nilai	Kategori
Matematika	> 80	Baik
	< 80	Cukup
Mata Pelajaran Produktif	> 82	Baik
	< 82	Cukup

Tabel 9 Dataset Setelah Perubahan Tipe Data

No.	Nama	MTK	MP Produktif	Keterangan
1	Abdul Rizal Mustopa	Cukup	Cukup	SI
2	Ade Karmila	Cukup	Baik	IF
3	Agil Ghalib Septiansyah	Cukup	Cukup	SI
4	Ahmad Ruhyatul Aziz	Cukup	Baik	IF
5	Amalia Kartika Sandra	Cukup	Baik	IF
...	...	...	...	...
...	...	...	...	...
353	Wiwin Sumiyati	Cukup	Cukup	SI

Setelah data dikategorikan menjadi Baik dan Cukup, proses perhitungan dapat dimulai. Algoritma C4.5 merupakan salah satu algoritma yang menampilkan hasil akhir berupa pohon keputusan, sebelum membuat pohon keputusan (*decision tree*), hal pertama yang harus dilakukan yaitu menentukan entropy dari setiap atribut.

1. Menghitung Entropy Node 1
   a. Entropy S (Total)

$$= \left(-\frac{175}{353} \times log_2\left(\frac{175}{353}\right)\right) + \left(-\frac{178}{353} \times log_2\left(\frac{178}{353}\right)\right) = 0{,}99995$$

   b. Entropy Nilai Matematika
   Entropy (Baik)

$$= \left(-\frac{121}{121} \times log_2\left(\frac{121}{121}\right)\right) + \left(-\frac{0}{121} \times log_2\left(\frac{0}{121}\right)\right) = 0$$

   Entropy (Cukup)

$$= \left(-\frac{54}{232} \times log_2\left(\frac{54}{232}\right)\right) + \left(-\frac{178}{232} \times log_2\left(\frac{178}{232}\right)\right) = 0{,}78279$$

   c. Entropy Nilai Mata Pelajaran Produktif
   Entropy (Baik)

$$= \left(-\frac{114}{114} \times log_2\left(\frac{114}{114}\right)\right) + \left(-\frac{0}{114} \times log_2\left(\frac{0}{114}\right)\right) = 0$$

   Entropy (Cukup)

$$= \left(-\frac{61}{239} \times log_2\left(\frac{61}{239}\right)\right) + \left(-\frac{178}{239} \times log_2\left(\frac{178}{239}\right)\right) = 0{,}81946$$

2. Menghitung *Information Gain*

Setelah mendapatkan entropy dari setiap atribut, selanjutnya menentukan *information gain*, hasil dari *information gain* tertinggi yang akan dijadikan node (akar) dari pohon keputusan.

   a. *Gain* Nilai Matematika

$$= 0{,}99995 - \left(\frac{232}{353} \times 0{,}78279\right) = 0{,}48548$$

   b. *Gain* Nilai Mata Pelajaran Produktif

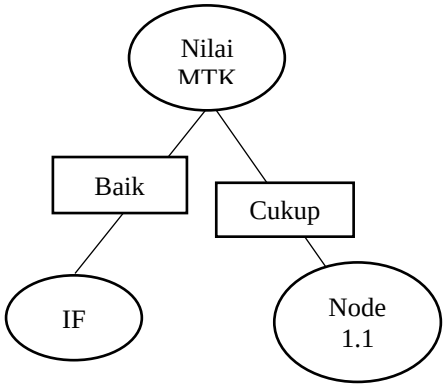
$$= 0{,}99995 - \left(\frac{239}{353} \times 0{,}81946\right) = 0{,}44513$$

Setelah mendapatkan nilai entropy dan *information gain*, dibuat tabel node 1.

Tabel 10 Node 1

Node			Jumlah	IF	SI	Entropy	*Gain*
1	Total		353	175	178		
	Nilai Matematika						
		Baik	121	121	0	0	0,48548
		Cukup	232	54	178	0,78279	
	Nilai MP Produktif						
		Baik	114	114	0	0	0,44513
		Cukup	239	61	178	0,81946	

Berdasarkan tabel 10 dapat dilihat bahwa atribut yang memiliki *information gain* tertinggi adalah Nilai Matematika, maka akar dari pohon keputusan adalah Nilai Matematika. Pohon keputusan node 1 dapat dilihat pada gambar 2.



Gambar 2 Node 1

Berdasarkan pohon keputusan node 1 jika nilai mata pelajaran matematika Baik maka dikategorikan IF, jika cukup harus dilakukan perhitungan selanjutnya untuk node 1.1.

3. Menghitung Entropy Node 1.1

Untuk node 1.1 entropy total yang digunakan tidak sama seperti entropy total node 1. Entropy total pada node 1.1 diambil dari atribut nilai matematika cukup, dataset yang digunakan hanya pada data yang mempunyai kategori cukup saja pada mata pelajaran matematika.

Tabel 11 Dataset Matematika Cukup

No.	Nama	MTK	MP Produktif	Keterangan
1	Abdul Rizal Mustopa	Cukup	Cukup	SI
2	Ade Karmila	Cukup	Baik	IF
3	Agil Ghalib Septiansyah	Cukup	Cukup	SI
4	Ahmad Ruhyatul Aziz	Cukup	Baik	IF
5	Amalia Kartika Sandra	Cukup	Baik	IF
...	...	...	...	...
...	...	...	...	...
353	Wiwin Sumiyati	Cukup	Cukup	SI

a. Entropy S (Total)

$$= \left(-\frac{54}{252} \times log_2\left(\frac{54}{252}\right)\right) + \left(-\frac{178}{252} \times log_2\left(\frac{178}{252}\right)\right) = 0{,}78279$$

b. Entropy Nilai Mata Pelajaran Produktif

Entropy (Baik)

$$= \left(-\frac{54}{54} \times log_2\left(\frac{54}{54}\right)\right) + \left(-\frac{0}{54} \times log_2\left(\frac{0}{54}\right)\right) = 0$$

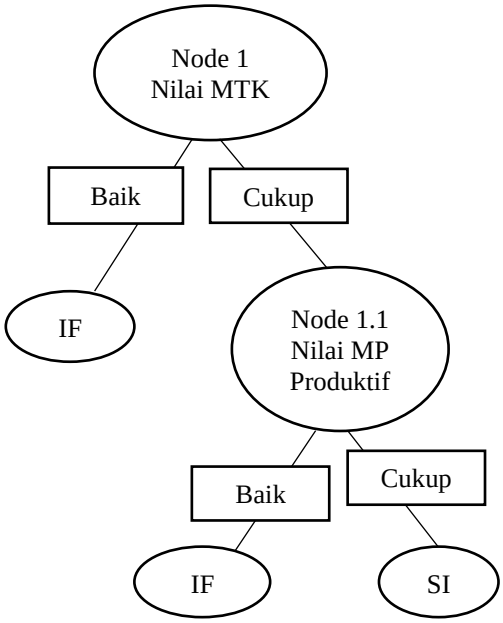
Entropy (Cukup)

$$= \left(-\frac{0}{178} \times log_2\left(\frac{0}{178}\right)\right) + \left(-\frac{178}{178} \times log_2\left(\frac{178}{178}\right)\right) = 0$$

Tabel 12 Node 1.1

Node			Jumlah	IF	SI	Entropy	*Gain*
1.1	Total (MTK)	Cukup	232	54	178	0,78279	
	Nilai MP Produktif						
		Baik	54	54	0	0	0
		Cukup	178	0	178	0	

Berdasarkan tabel 12 nilai mata pelajaran produktif menjadi atribut terakhir yang perlu dihitung dan menjadi akar dari node 1.1.



Gambar 3 Node 1.1

Dari gambar 4.3 dapat disimpulkan bahwa pada node 1 apabila nilai mata pelajaran Matematika baik dikategorikan IF, apabila cukup dilihat dari nilai mata pelajaran produktif. Jika nilai mata pelajaran produktif baik dikategorikan IF, jika cukup dikategorikan SI.

C. Pengujian Menggunakan *Tool* RapidMiner

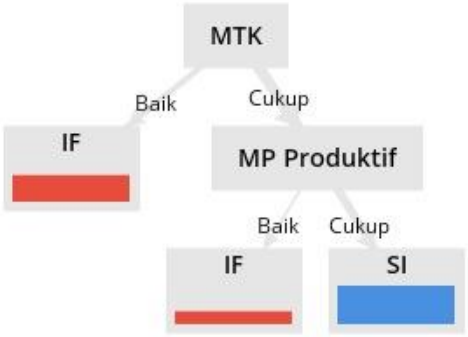
Pengujian menggunakan tool RapidMiner menghasilkan nilai akurasi yang baik, yaitu 98% untuk algoritma K-NN dan 100% untuk algoritma C4.5.

Tabel 13 Hasil Pengujian Algoritma K-NN di RapidMiner

Accuracy : 98.04%+/-2.95% (*micro average* : 98.02%)			
	True SI	*True* IF	*Class Precision*
Predi SI	177	6	96.72%
Pred IF	1	169	99.41%
Class recall	99.44%	96.57%	

Tabel 14 Hasil Pengujian Algoritma C4.5 di RapidMiner

Accuracy : 100.00%			
	True SI	*True* IF	*Class Precision*
Pred SI	53	0	100.00%
Pred IF	3	53	100.00%
Class recall	100.00%	100.00%	



Gambar 4 Hasil Pohon Keputusan di RapidMiner

IV. KESIMPULAN DAN SARAN

Berdasarkan hasil perhitungan dan pengujian menggunakan algoritma C4.5 dan K-*Nearest Neighbor*, dapat disimpulkan bahwa algoritma C4.5 dan K-*Nearest Neighbor* mampu melakukan klasifikasi peminatan program studi dengan sangat baik. Kedua algoritma tersebut menghasilkan nilai akurasi yang tinggi dengan selisih hasil yang tipis, yaitu 100% dari algoritma C4.5 dan 98% dari algoritna K-*Nearest Neighbor*.

PENGAKUAN

Naskah ilmiah ini adalah sebagian dari penelitian Tugas Akhir milik Mulya Cahya Ramadanty dengan judul Penerapan Algoritma C4.5 dan K-*Nearest Neighbor* untuk Klasifikasi Peminatan Program Studi di Perguruan Tinggi Berdasarkan Nilai Rapor, yang dibimbing oleh Amril Mutoi Siregar dan Dwi Sulistya Kusumaningrum.

DAFTAR PUSTAKA

[1] Murti, A. T. A. 2019. Fenomena Salah Jurusan di Kalangan Mahasiswa. *https://mahasiswaindonesia.id/*. 26 Januari 2020 (20:30).

[2] Novianti. B., T. Rismawan, dan S. Bahri. 2016. Implementasi Data Mining Dengan Algoritma C4.5 Untuk. *Jurnal Coding, Sistem Komputer Untan* 4(3): 75-84.

[3] Siregar, A. M. dan A. Fauzi. 2020. Klasifikasi Kabupaten Kota Provinsi Jawa Barat Berdasarkan Pendapatan Dari Sektor Pertanian Dengan Algoritma *Decision Tree*. *Faktor Exacta* 13(1): 1-8.

[4] Sambani, E. B. dan F. Nuraeni. 2017. Penerapan Algoritma C4.5 Untuk Klasifikasi Pola Penjurusan di Sekolah Menengah Kejuruan (SMK) Kota Tasikmalaya. *Computer Science Research and Its Development Journal* 9(3): 149-157.

[5] Amalia, W., A. M. Siregar, dan S. A. P. Lestari. 2020. Seleksi Karyawan Menggunakan Algoritme C4.5 Pada Perusahaan Tekstil. *Scientific Student Journal for Information, Technology and Science*. 1(2): 102-107.

[6] Nurjanah, S., A. M. Siregar, dan D. S. Kusumaningrum. 2020. Penerapan Algoritma K – *Nearest Neighbor* (K-NN) untuk Klasifikasi Pencemaran Udara Di Kota Jakarta. *Scientific Student Journal for Information, Technology and Science*. 1(2): 71-76.

[7] Utari, F. D., A.M. Siregar, dan D. Wahiddin. 2020. Implementasi Algoritme K – *Nearest Neighbor* (K-NN) untuk Prediksi Hasil Produksi. *Scientific Student Journal for Information, Technology and Science*. 1(1): 21-25.

[8] Prasetyo, E. 2014. *Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab.* 1st ed. Penerbit Andi. Jakarta.

[9] Haryati, S., A. Sudarsono, dan E. Suryana. 2015. Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu). *Jurnal Media Infotama* 11(2): 130-138.

[10] Ernawati, S. dan R. Wati. 2018. Penerapan Algoritma K-*Nearest Neighbors* Pada Analisis Sentimen *Review* Agen Travel. *JURNAL KHATULISTIWA INFORMATIKA* 6(1): 64-69.

[11] Lestari, M. 2015. Penerapan Algoritma Klasifikasi *Nearest Neighbor* (K-NN) Untuk Mendeteksi Penyakit Jantung. *Faktor Exacta* 7(4): 367-371.

[12] Kusumaningrum, D. S. dan S. A. P. Lestari. 2019. Analisis Kesulitan Belajar Matematika Diskrit Mahasiswa Teknik Informatika. *PRISMA*. 8(2): 96-110.