

Deteksi Dini Tingkat Stres Pada Mahasiswa Menggunakan Metode Iterative Dichotomiser 3 dan K-Nearest Neighbour

¹ Alfiyan Aldi, ² S.R.Candra Nursari, ³ Febri Maspiyanti
Informatics Engineering
Pancasila University
Jakarta, Indonesia

¹ phillips.lamp42@gmail.com, ² sri.rezeki.candra.n@univpancasila, ³ febri.maspiyanti@univpancasila.ac.id

Abstract —Data mining merupakan cabang *artificial intelligence* yang digunakan untuk pencarian pola data atau model data dengan tugasnya melakukan klasifikasi, *clustering*, asosiasi dan deteksi anomali. Deteksi dini pada tingkat stres dapat dikategorikan sebagai klasifikasi dalam data mining. Penelitian deteksi dini tingkat stress pada mahasiswa ini menggunakan metode *Iterative Dichotomiser 3* (ID3) dan *K-Nearest Neighbour* (KNN). Teknik pengumpulan data yang dilakukan pada penelitian ini adalah wawancara pakar, studi literatur dan kuesioner dengan 300 mahasiswa sebagai responden. Penulis menggunakan teknik *cross-validation* 10 folds untuk menghitung nilai akurasi dari kedua metode tersebut. Nilai akurasi dengan metode K-NN rata-rata 91% sedangkan dengan metode ID3 hanya 61%. Akurasi tertinggi K-NN terdapat pada fold 10 dengan akurasi 100%. Sementara akurasi tertinggi pada ID3 ada pada fold 7 dengan akurasi 76.67%. Jadi, K-NN memiliki hasil akurasi klasifikasi yang lebih baik dari ID3.

Keywords—Data Mining, Klasifikasi, Tingkat Stres Mahasiswa, K-Nearest Neighbor, Iterative Dichotomiser 3.

I. PENDAHULUAN

Dalam upaya menghasilkan mahasiswa yang mampu bersaing di era yang penuh dengan perkembangan informasi, maka banyak kampus yang menetapkan standar akademik yang tinggi, baik dari segi nilai akademik, sertifikasi keahlian, penguasaan bahasa asing hingga kemampuan non akademik seperti *softskill* yang sudah harus dimiliki oleh setiap mahasiswa. Dengan peningkatan standar akademik yang tinggi, berbagai tanggung jawab, beban dan tuntutan hidup kerap kali menyebabkan mahasiswa mengalami stres.

Stres memiliki pengertian yang luas, bahkan setiap bidang mempunyai nama atau kategori kasus stres tersendiri. Dari beberapa kasus stres yang terjadi pada mahasiswa, stres akademik merupakan kasus stres yang umum dihadapi oleh setiap mahasiswa. Jumlah mahasiswa yang mengalami stres akademik meningkat setiap semesternya [1]. Stres akademik diartikan sebagai suatu keadaan hasil interaksi individu dengan lingkungan pendidikan yang menjadi stresor akademik individu yang dapat berdampak pada penyesuaian psikologi dan prestasi akademik individu [2] [3]. Stresor akademik meliputi manajemen waktu, masalah finansial, gangguan tidur dan aktivitas sosial [4].

Masalah yang terjadi pada mahasiswa adalah gejala-gejala stres yang dialami individu seringkali tidak disadari karena tahapan awal stres timbul secara lambat dan baru dapat disadari jika tahapan atau gejala tersebut sudah mengganggu kondisi kehidupan individu tersebut. Jika seseorang tidak sadar akan stres yang mereka alami hal ini dapat menyebabkan berbagai masalah atau gangguan. Masalah atau gangguan yang tidak diatasi pada tahap ringan akan meningkatkan stres ke tahap yang lebih tinggi dan jika hal tersebut dibiarkan terus menerus akan menimbulkan dampak yang lebih buruk. Salah satu masalah dampak yang paling sering dialami oleh orang stres adalah penurunan kekebalan tubuh. Jika kekebalan tubuh menurun maka seseorang dapat terjerang virus yang akan menyebabkan berbagai penyakit.

Dengan adanya data mining sistem dapat belajar dari data yang sudah ada kemudian menerapkan pola pembelajaran itu untuk melakukan proses terhadap data yang baru. Klasifikasi merupakan salah satu tugas yang paling umum dilakukan pada data mining. Klasifikasi dapat ditemukan dalam berbagai bidang, antara lain: perbankan, pendidikan, ilmu kedokteran, hukum [5]. K Nearest Neighbor (K-NN) dan Iterative Dichotomiser 3 (ID3) adalah dua contoh algoritma yang digunakan untuk melakukan proses klasifikasi dan prediksi.

Melihat permasalahan yang dihadapi oleh mahasiswa dan adanya hubungan antara data mining dengan bidang lain (ilmu kedokteran), peneliti tertarik untuk mengetahui deteksi dini tingkat stres pada mahasiswa menggunakan metode metode K-NN dan ID3 serta melakukan perbandingan akurasi terhadap dua metode tersebut.

II. K-NEAREAST NEIGHBOR (K-NN) AND ITERATIVE DICHOTOMISER 3 (ID3)

A. Metode K-Neareast Neighbor (K-NN)

K-Nearest Neighbor (K-NN) adalah suatu metode yang menggunakan algoritma *super-vised* dimana hasil dari *query instance* yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN. Tujuan dari algoritma ini adalah mengklasifikasikan objek baru berdasarkan atribut dan *training sample* [6]. Jarak *Euclidean* dapat dicari dengan menggunakan persamaan 1 berikut ini [7]:

$$D_{xy} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.1)$$

Keterangan :

D : jarak kedekatan

x : data *training*

y : data *testing*

n : jumlah atribut individu antara 1 s.d. n

f : fungsi *similarity* atribut i antara kasus X dan kasus Y

i : Atribut individu antara 1 sampai dengan n

Tahapan algoritma K-NN:

1. Menentukan parameter k .
2. Menghitung jarak *euclidean* objek terhadap data *training* yang diberikan.
3. Mengurutkan hasil jarak *euclidean* secara *ascending*.
4. Mengumpulkan objek klasifikasi *nearest neighbor* berdasarkan nilai k yang telah ditetapkan.
5. Mencari kelas mayoritas berdasarkan hasil prediksi

B. Metode Iterative Dichotomiser 3 (ID3)

Iterative Dichotomiser 3 (ID3) [8] adalah algoritma *decision tree learning* yang paling dasar. Algoritma ini melakukan pencarian secara menyeluruh (*greedy*) pada semua kemungkinan pohon keputusan. Salah satu algoritma induksi pohon keputusan yaitu ID3. ID3 dikembangkan oleh J. Ross Quinlan. Algoritma ID3 dapat diimplementasikan menggunakan fungsi rekursif [9].

Algoritma ID3 membangun *decision tree* secara *top-down*, mulai dengan pertanyaan : “atribut mana yang pertama kali harus dicek dan diletakkan pada root?” pertanyaan ini dijawab dengan mengevaluasi semua atribut yang ada dengan menggunakan suatu ukuran statistik (yang banyak digunakan adalah *information gain*) untuk mengukur efektivitas suatu atribut dalam mengklasifikasikan kumpulan sampel data

Tahapan ID3:

1. Input sampel *training*, label *training* dan atribut.
2. Hitung entropi dan informasi gain dari setiap atribut. Maka gain tertinggi akan menjadi simpul akar pada pohon keputusan.
3. Jika semua sampel berada dalam kelas yang sama, maka simpul ini menjadi daun dan dilabeli menjadi kelas. Jika tidak, *information gain* akan digunakan untuk memilih atribut terbaik dalam memisahkan data sampel menjadi kelas-kelas individu.
4. Cabang akan dibuat untuk setiap nilai pada atribut dan data sampel akan dipartisi lagi.
5. Algoritma ini menggunakan proses rekursif untuk membentuk pohon keputusan pada setiap data partisi. Jika sebuah atribut sudah digunakan disebuah simpul, maka atribut ini tidak akan digunakan lagi di simpul anak-anaknya.
6. Proses ini berhenti jika dicapai kondisi seperti berikut : Semua sampel pada simpul berada di dalam satu kelas Tidak ada atribut lainnya yang dapat digunakan untuk mempartisi sampel lebih lanjut. Dalam hal ini akan diterapkan suara terbanyak. Ini berarti mengubah sebuah simpul menjadi daun dan melabelinya dengan kelas pada suara terbanyak

C. Entropi dan gain informasi

Konsep Entropi yang digunakan untuk mengukur “seberapa informatifnya” sebuah node.

$$Entropy(S) = \sum_{j=1}^k -p_j \log_2 p_j \quad (2.2)$$

Dimana:

S = himpunan (dataset) kasus

k = banyaknya partisi S

p_j = probabilitas di dapat dari jumlah (Ya) / Total Kasus.

Setelah mendapat nilai entropi, pemilihan atribut dilakukan dengan nilai *information gain* terbesar.

$$Gain(A) = Entropy(S) = \sum_{i=1}^k \frac{|S_i|}{|S|} x Entropy(S_i) \quad (2.3)$$

Dimana:

S = ruang (data) sample yang digunakan untuk *training*.

A = atribut.

$|S_i|$ = jumlah sample untuk nilai V .

$|S|$ = jumlah seluruh sample data.

$Entropy(S_i)$ = entropi untuk sample yang memiliki nilai i

III. HASIL DAN PEMBAHASAN

Penelitian dilakukan di Fakultas Teknik Universitas Pancasila dengan menyebarkan kuesioner mengacu pada penelitian yang dilakukan oleh Wicaksana [2], kuesioner pada penelitian tersebut telah dilakukan analisis ulang karena terdapat beberapa item pertanyaan yang kurang menggambarkan stres akademik sehingga terpilih 14 item pertanyaan yang meliputi stres pengajar, stres hasil, stres ujian, stres belajar dalam kelompok, stres teman sebaya, stres manajemen waktu dan stres yang diakibatkan diri sendiri. dengan 3 label yaitu stres ringan, stres sedang dan stres berat dan memiliki 5 alternatif jawaban yaitu Sangat Sesuai (SS), Sesuai (S), Netral (N), Tidak Sesuai (TS) dan Sangat Tidak Sesuai (STS). Nilai tertinggi 5 diberikan untuk jawaban Sangat Sesuai (SS), nilai 4 diberikan untuk jawaban Sesuai (S), nilai 3 diberikan untuk jawaban Netral (N), nilai 2 diberikan untuk jawaban Tidak Sesuai (TS) dan nilai 1 diberikan untuk jawaban Sangat Tidak Sesuai (STS).

Tabel 1. Pertanyaan Kuesioner

No	Pertanyaan
1	Saya merasa kalau saya tidak mengerti sebagian besar materi yang diberikan dosen
2	Beberapa dosen menyediakan materi yang terlalu banyak. hal ini menyebabkan saya tidak dapat menyelesaikan pembelajaran dan menyerap pengetahuan
3	Saya merasa kalau ada banyak tekanan karena beberapa mata kuliah menggunakan bahasa asing
4	Saya merasa kalau saya tidak dapat menyesuaikan dengan beberapa metode pembelajaran dosen
5	Saya merasa ketika saya masuk ke universitas, saya tidak bisa mengikuti kecepatan mengajar dosen
6	Saya khawatir kalau hasil akademik saya tidak akan memenuhi harapan orang tua saya
7	Saya tidak mendapatkan tidur yang cukup di malam hari karena saya khawatir akan ujian
8	Saya kuatir kalau saya harus mengulang mata kuliah wajib di mana saya gagal
9	Ketika kerja berkelompok diperlukan untuk menyelesaikan latihan atau tugas, saya khawatir kalau saya tidak dapat menemukan anggota kelompok yang tepat

10	Ketika saya memberikan pidato atau presentasi, saya khawatir kalau teman sekelas saya akan menertawakan ketidakmampuan saya menampilkan yang terbaik
11	Terkadang, kata-kata yang digunakan teman sekelas saya dengan mudah membuat saya tidak percaya diri dan menyebabkan kerugian bagi saya
12	Saya merasa gugup ketika harus membuat pidato atau memberikan presentasi
13	Ketika saya ingin belajar sendiri, saya sering terganggu dengan obrolan teman sekelas
14	Saya sangat khawatir kalau hasil akademik saya tidak sebaik teman sekelas saya

Kuesioner disebarakan pada 300 mahasiswa 6 prodi di Fakultas Teknik Universitas Pancasila yang masing-masing prodinya 50 data mahasiswa. Kemudian dari setiap kuesioner ditetapkan labelnya. Penentuan label dilakukan berdasarkan ketentuan dari pakar, ketentuan tersebut adalah jika jumlah > dari 0 & < 24 maka stres ringan, jika jumlah > 24 & < 48 maka stres sedang, dan jika lain dari itu maka termasuk dalam stres berat. Setelah proses pengumpulan data. Dilakukan pembagian data *training* dan data *testing* menggunakan teknik *corss-validation 10 folds*. Setelah itu data dihitung dengan metode K-NN dan ID3.

A. Metode K-Neareast Neighbor (K-NN)

Pada perhitungan ini peneliti menggunakan 10 data dimana 9 data sebagai data training dan 1 data sebagai data testing. Berikut adalah dataset yang digunakan :

Tabel 2. Dataset K-NN

mhs	a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12	a13	a14	label
mhs1	1	2	1	1	2	2	1	3	1	1	1	1	1	1	Ringan
mhs2	5	4	3	5	4	1	2	1	2	2	1	2	3	1	Sedang
mhs3	5	4	5	5	5	5	4	5	3	2	2	5	3	4	Berat
mhs4	2	1	2	2	2	2	1	2	2	2	1	2	1	2	Ringan
mhs5	4	3	5	3	1	2	2	5	1	5	1	2	2	3	Sedang
mhs6	2	2	5	2	2	5	5	5	5	2	4	2	4	4	Berat
mhs7	3	2	1	2	2	3	1	1	1	1	1	2	2	2	Ringan
mhs8	3	4	1	4	2	5	2	2	2	1	2	1	2	2	Sedang
mhs9	4	5	1	4	4	5	2	5	5	3	5	5	5	5	Berat
mhs10	3	4	4	2	1	2	1	5	1	2	1	2	3	2	Sedang

Ada 14 atribut (a1 s.d.14) yang digunakan, nilai atribut antara lain : 1=STS, 2=TS, 3=N, 4=S, 5=SS. Proses perhitungan K-NN adalah :

1. Peneliti menentukan K sama dengan 1. Karena dari ujicoba yang dilakukan menggunakan sistem k=10% dari data training adalah yang paling baik akurasiya.
2. Menghitung jarak euclidean objek terhadap data training yang diberikan, maka akan diperoleh hasil seperti tabel dibawah ini.

$$\text{Jarak mhs1, mhs10} = \sqrt{(1-3)^2 + (2-4)^2 + (1-4)^2 + (1-2)^2 + (2-1)^2 + (2-2)^2 + (1-1)^2 + (3-5)^2 + (1-1)^2 + (1-2)^2 + (1-1)^2 + (1-2)^2 + (1-3)^2 + (1-2)^2}$$

Tabel 3. Jarak Euclidean

mhs	label	jarak
mhs1	Ringan	5.477226
mhs2	Sedang	6.557439
mhs3	Berat	8.124038
mhs4	Ringan	5.385165
mhs5	Sedang	4
mhs6	Berat	7.874008
mhs7	Ringan	5.744563
mhs8	Sedang	6.164414
mhs9	Berat	9.433981

3. Mengurutkan hasil jarak euclidean secara ascending.

Tabel 4. Urutan Ascending Data Mhs

mhs	label	jarak
mhs5	Sedang	4
mhs4	Ringan	5.385165
mhs1	Ringan	5.477226
mhs7	Ringan	5.744563
mhs8	Sedang	6.164414
mhs2	Sedang	6.557439
mhs6	Berat	7.874008
mhs3	Berat	8.124038
mhs9	Berat	9.433981

4. Mengumpulkan objek K-NN berdasarkan nilai k yang sudah ditetapkan dan mencari kelas mayoritas berdasarkan hasil prediksi.

Tabel 5. Penentuan Prediksi K-NN

mhs	label	jarak
mhs5	Sedang	4
mhs4	Ringan	5.385165
mhs1	Ringan	5.477226
mhs7	Ringan	5.744563
mhs8	Sedang	6.164414
mhs2	Sedang	6.557439
mhs6	Berat	7.874008
mhs3	Berat	8.124038
mhs9	Berat	9.433981

Dari tabel diatas dapat dilihat hasil urut dari nilai terkecil, karena tadi nilai k tadi sudah ditetapkan sama dengan 1 maka hasil mhs5 dengan label sedang yang terpilih. Maka prediksi mhs 10 memiliki label sedang adalah benar.

B. Metode Iterative Dichotomiser 3 (ID3)

Pada perhitungan ini peneliti menggunakan 300 data dimana 270 data sebagai data training dan 30 data sebagai data testing. Terdapat 14 pertanyaan atau atribut pada penelitian ini dimana A1=tidak mengerti materi, A2=materi banyak tidak terserap, A3=tekanan bahasa asing, A4=tidak sesuai dengan metode belajar dosen, A5=tidak dapat mengikuti kecepatan mengajar dosen, A6=nilai < harapan orang tua, A7=gangguan tidur akan ujian, A8=khawatir mengulang matakuliah, A9=takut gagal menemukan kelompok, A10=khawatir presentasi, A11=kata-kata teman mem-

buat saya down, **A12**= presentasi/pidato, **A13**=gangguan orobolan saat belajar, **A14**=khawatir nilai < teman

Nilai tertinggi 5 diberikan untuk jawaban Sangat Sesuai (SS), nilai 4 diberikan untuk jawaban Sesuai (S), nilai 3 diberikan untuk jawaban Netral (N), nilai 2 diberikan untuk jawaban Tidak Sesuai (TS) dan nilai 1 diberikan untuk jawaban Sangat Tidak Sesuai (STS).

Tabel 6. Dataset ID3

mhs	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	label
Mhs1	2	1	4	2	3	1	1	1	4	4	2	5	2	4	Sedang
Mhs2	3	1	5	1	2	1	2	3	1	2	3	2	1	2	Sedang
Mhs3	1	2	1	1	2	2	1	3	1	1	1	1	1	1	Ringan
Mhs4	5	4	3	5	4	1	2	1	2	2	1	2	3	1	Sedang
Mhs5	3	3	4	3	2	1	2	1	4	3	5	4	2	4	Sedang
Mhs6	5	4	5	5	5	5	4	5	3	2	2	5	3	4	Berat
Mhs7	4	4	4	4	5	3	5	4	3	3	5	4	5	4	Berat
Mhs8	2	1	1	3	3	2	4	2	3	3	3	4	3	3	Sedang
Mhs9	2	1	2	4	1	3	2	4	2	4	1	3	1	5	Sedang
.....	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
Mhs300	3	4	5	4	1	3	1	3	3	5	1	2	3	3	Sedang

Kemudian menghitung entropi.

$$Entropi (S) = (- (25/270) \times \log_2 (25/270) + (- (154/270) \times \log_2 (154/270)) + (- (91/270) \times \log_2 (91/270)) = 1.308701$$

Tabel 7. Analisis Total Kasus Node Root

Total kasus	Sum(Ringan)	Sum(Sedang)	Sum(Berat)	Entropi total
270	25	154	91	1.308701

Setelah mendapatkan entropi dari keseluruhan kasus, maka melakukan analisis pada setiap atribut dan nilai-nilai di hitung entropinya.

Tabel 8. Analisis Node Root

Node	Atribut	Nilai Atribut	Total Kasus	Ringan	Sedang	Berat	Entropi	gain
	A01	1	26	11	14	1	1.18672	0.255359
		2	45	11	30	4	1.19717	
		3	80	3	57	20	1.02607	
		4	72	0	34	38	0.99777	
		5	47	0	19	28	0.97338	
	A02	1	21	7	13	1	1.165784	0.254219
		2	57	15	36	6	1.267444	
		3	71	3	49	19	1.071067	
		4	78	0	45	33	0.982859	
		5	43	0	11	32	0.820364	
	A03	1	29	13	14	2	1.292164	0.234308
		2	68	12	43	13	1.316071	
		3	72	0	51	21	0.870864	
		4	53	0	24	29	0.99357	
		5	48	0	22	26	0.994985	
	A04	1	25	10	14	1	1.182966	0.174106
		2	62	13	36	13	1.400507	
		3	76	2	47	27	1.097301	
		4	77	0	42	35	0.99403	

A05	5	30	0	15	15	1	0.179653
	1	34	7	25	2	1.036054	
	2	77	16	46	15	1.374745	
	3	81	2	50	29	1.09202	
	4	50	0	24	26	0.998846	
A06	5	28	0	9	19	0.905928	0.156012
	1	36	10	22	4	1.299737	
	2	52	10	34	8	1.273651	
	3	71	4	40	27	1.230605	
	4	61	0	34	27	0.99048	
A07	5	50	1	24	25	1.008269	0.183829
	1	42	14	23	5	1.36959	
	2	63	9	42	12	1.246705	
	3	63	2	40	21	1.102426	
	4	68	0	38	30	0.989993	
A08	5	34	0	11	23	0.908178	0.158769
	1	31	9	20	2	1.181039	
	2	43	9	25	9	1.399403	
	3	66	7	40	19	1.298345	
	4	73	0	43	30	0.977001	
A09	5	57	0	26	31	0.994442	0.232326
	1	41	12	26	3	1.211555	
	2	63	10	47	6	1.059903	
	3	61	3	36	22	1.193365	
	4	65	0	32	33	0.999829	
A10	5	40	0	13	27	0.909736	0.283512
	1	43	17	25	1	1.110375	
	2	69	7	52	10	1.046298	
	3	74	1	45	28	1.050814	
	4	47	0	21	26	0.991821	
A11	5	37	0	11	26	0.877962	0.223043
	1	48	16	29	3	1.217539	
	2	76	8	52	16	1.189732	
	3	65	1	42	22	1.028755	
	4	51	0	21	30	0.977418	
A12	5	30	0	10	20	0.918296	0.304737
	1	37	17	19	1	1.150062	
	2	63	7	50	6	0.939914	
	3	67	1	43	23	1.030697	
	4	63	0	27	36	0.985228	
A13	5	40	0	15	25	0.954434	0.208487
	1	37	10	24	3	1.209095	
	2	68	14	45	9	1.249732	
	3	56	1	35	20	1.058007	
	4	68	0	33	35	0.999376	
A14	5	41	0	17	24	0.97887	0.164434
	1	27	10	15	2	1.279975	
	2	61	10	38	13	1.32834	
	3	66	5	44	17	1.176038	
	4	60	0	31	29	0.999198	
	5	56	0	26	30	0.996317	

$$Entropi (S,A01, 1) = (- (11/26) \times \log_2 (11/26) + (- (14/26) \times \log_2 (14/26) + (- (1/26) \times \log_2 (1/26)) = 1.186719845$$

$$Entropi (S,A01, 2) = (- (11/45) \times \log_2 (11/45) + (- (30/45) \times \log_2 (30/45) + (- (4/45) \times \log_2 (4/45)) = 1.197176081$$

$$Entropi (S, A01, 3) = (- (3/80) \times \log_2 (3/80) + (- (57/80) \times \log_2 (57/80) + (- (20/80) \times \log_2 (20/80))) = 1.026075842$$

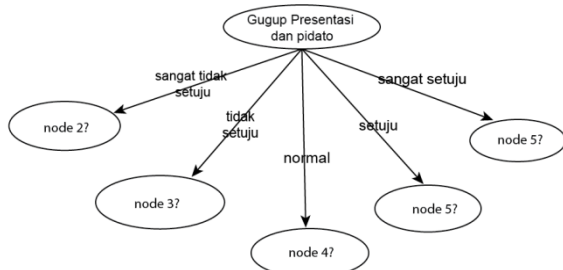
$$Entropi (S, A01, 4) = (- (0/72) \times \log_2 (0/72) + (- (34/72) \times \log_2 (34/72) + (- (38/72) \times \log_2 (38/72))) = 0.997772472$$

$$Entropi (S, A01, 5) = (- (0/47) \times \log_2 (0/47) + (- (19/47) \times \log_2 (19/47) + (- (28/47) \times \log_2 (28/47))) = 0.973385435$$

Setelah mendapatkan nilai entropi dari setiap nilai atribut yang ada pada atribut A01, hitung pula nilai entropi pada setiap atribut seperti Tabel 8, selanjutnya hitung nilai informasi gain dari setiap variabel.

$$Gain (A01) = 1.308701 - (((26/270) \times 1.186719845) + ((45/270) \times 1.197176081) + ((80/270) \times 1.026075842) + ((72/270) \times 0.997772472) + ((47/270) \times 0.973385435)) = 0.258521037$$

Hitung pula Gain (A02), Gain (A03), Gain (A04) hingga Gain (A14) Karena nilai gain terbesar adalah Gain (A12), maka atribut “A12” menjadi node akar (root node).



Sumber: (Aldi & Maspiyanti, 2018)

Gambar 4. Implementasi Tes Stres

Kemudian dihitung nilai entropi atribut Gugup Presentasi dan Pidato atau “A12” dan nilai atribut sangat tidak setuju “1” serta hitung entropi setiap atribut dan gainnya. Setelah itu tentukan pilih atribut yang memiliki gain tertinggi untuk dibuatkan node 2. Kemudian menghitung total kasus pada A12 dengan nilai atribut 1 ada berapa banyak, untuk menghitung entropi seluruh kasus.

$$Entropi (S) = (- (17/37) \times \log_2 (17/37) + (- (19/37) \times \log_2 (19/37)) + (- (1/37) \times \log_2 (1/37))) = 1.150061715$$

Tabel 9. Analisis Total Kasus Node 2

Total kasus	Sum(Ringan)	Sum(Sedang)	Sum(Berat)	Entropi total
37	17	19	1	1.150061715

Setelah mendapatkan entropi dari keseluruhan kasus, lakukan analisis pada setiap atribut dan nilai-nilainya dan hitung entropinya. Karena mencari kasus dan menghitung atribut A12 maka A12 tidak termasuk dalam tabel dibawah.

Tabel 10. Analisis Node 2

Node	Atribut	Nilai Atribut	Total Kasus	Ringan	Sedang	Berat	Entropi	gain
	A01	1	10	6	4	0	0.970951	0.510832
		2	12	10	2	0	0.650022	

	A02	3	7	1	6	0	0.591673	0.33417
		4	6	0	6	0	0	
		5	2	0	1	1	1	
		1	6	4	2	0	0.918296	
		2	17	11	6	0	0.936667	
	A03	3	6	2	4	0	0.918296	0.503558
		4	4	0	4	0	0	
		5	4	0	3	1	0.811278	
		1	11	9	2	0	0.684038	
		2	13	8	5	0	0.961237	
	A04	3	6	0	5	1	0.650022	0.36305
		4	6	0	6	0	0	
		5	1	0	1	0	0	
		1	10	7	3	0	0.881291	
		2	12	8	3	1	1.188722	
	A05	3	7	2	5	0	0.863121	0.242558
		4	3	0	3	0	0	
		5	5	0	5	0	0	
		1	9	5	4	0	0.991076	
		2	17	10	7	0	0.977418	
	A06	3	7	2	5	0	0.863121	0.365956
		4	2	0	2	0	0	
		5	2	0	1	1	1	
		1	9	6	3	0	0.918296	
		2	11	7	4	0	0.94566	
	A07	3	5	3	2	0	0.970951	0.447661
		4	4	0	4	0	0	
		5	6	0	6	0	0	
		1	11	9	2	0	0.684038	
		2	10	6	4	0	0.970951	
	A08	3	6	2	3	1	1.459148	0.471517
		4	4	0	4	0	0	
		5	6	0	6	0	0	
		1	5	5	0	0	0	
		2	7	5	2	0	0.863121	
	A09	3	16	7	9	0	0.988699	0.205263
		4	4	0	3	1	0.811278	
		5	5	0	5	0	0	
		1	17	9	8	0	0.997503	
		2	12	6	6	0	1	
	A10	3	4	2	2	0	1	0.253374
		4	2	0	2	0	0	
		5	2	0	1	1	1	
		1	20	12	8	0	0.970951	
		2	12	5	7	0	0.979869	
	A11	3	3	0	3	0	0	0.217269
		4	0	0	0	0	0	
		5	2	0	1	1	1	
		1	20	10	10	0	1	
		2	12	7	5	0	0.979869	
	A12	3	3	0	2	1	0.918296	
		4	1	0	1	0	0	
	A12	1						
		2						

		3						0.442398
		4						
		5						
	A13	1	12	8	4	0	0.918296	
		2	14	9	5	0	0.940286	
		3	5	0	5	0	0	
		4	2	0	1	1	1	
		5	4	0	4	0	0	
	A14	1	12	9	3	0	0.811278	
		2	8	4	3	1	1.405639	
		3	8	4	4	0	1	
		4	3	0	3	0	0	
		5	6	0	6	0	0	

$$Entropi (S, A01, 1) = (-6/10) \times \log_2 (6/10) + (-4/10) \times \log_2 (4/10) + (-0/10) \times \log_2 (0/10) = 0.970950594$$

$$Entropi (S, A01, 2) = (-10/12) \times \log_2 (10/12) + (-2/12) \times \log_2 (2/12) + (-0/12) \times \log_2 (0/12) = 0.650022422$$

$$Entropi (S, A01, 3) = (-1/7) \times \log_2 (1/7) + (-6/7) \times \log_2 (6/7) + (-0/7) \times \log_2 (0/7) = 0.591672779$$

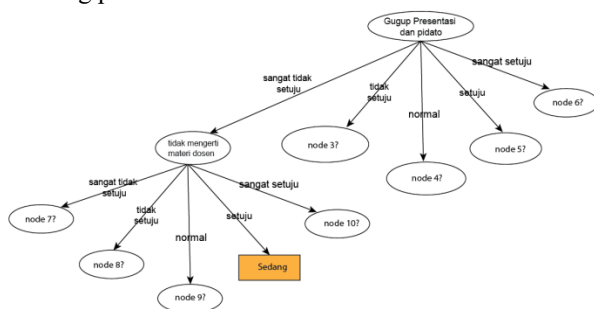
$$Entropi (S, A01, 4) = (-0/6) \times \log_2 (0/6) + (-6/6) \times \log_2 (6/6) + (-0/6) \times \log_2 (0/6) = 0$$

$$Entropi (S, A01, 5) = (-0/2) \times \log_2 (0/2) + (-1/2) \times \log_2 (1/2) + (-1/2) \times \log_2 (1/2) = 1$$

Setelah mendapatkan nilai entropi dari setiap nilai atribut yang ada pada atribut A01, hitung pula nilai entropi pada setiap atribut seperti Tabel 10, selanjutnya hitung nilai informasi gain dari setiap variabel.

$$Gain (A01) = 1.150061715 - (((10/37) \times 0.970951) + ((12/37) \times 0.650022422) + ((7/37) \times 0.591672779) + ((6/37) \times 0) + ((2/37) \times 1)) = 0.510832406$$

Hitung Gain (A02), Gain (A03), Gain (A04) hingga Gain (A14) Karena nilai gain terbesar adalah Gain (A01), maka atribut "A01" atau tidak mengerti materi dosen menjadi node cabang pada node 2.



Gambar 5. Tree node ke 2 Bercabang

Untuk menganalisis node-node selanjutnya, lakukan lagi langkah-langkah yang sama seperti sebelumnya hingga semua node berbentuk node leaf atau bisa dikatakan, lakukan langkah yang sama seperti sebelumnya hingga mendapatkan label pada node leaf berdasarkan analisis data.

C. Evaluasi Hasil

Penentuan atau pemilihan jumlah k pada metode K-NN untuk menentukan seberapa banyak tetangga yang akan dilakukan pengurutan berdasarkan jarak *euclidean* terdekat yang akan mempengaruhi hasil klasifikasi dari setiap data. Berdasarkan pengujian yang dilakukan seperti pada tabel 2, jika jumlah k terlalu besar atau jumlah k lebih dari 50% dari total data training maka akurasi yang diperoleh lebih buruk jika dibandingkan dengan k kurang dari 50%.

Hal ini terjadi karena jumlah label kelas stres ringan hanya 25 dari 300 data. Melihat label stres ringan yang sedikit maka dilakukan pengujian dengan jumlah k 5, 3 dan 1, akurasi yang diperoleh adalah 91%, 86.3% dan 86%. Jadi jumlah k yang tepat pada penelitian ini adalah 5.

Tabel 11. Tabel akurasi berdasarkan jumlah K pada K-NN

Jumlah K (%)	10	20	30	40	50	60	70	80	90	100
Akurasi (%)	92	84.67	85.33	84.67	83	73.67	57	57.33	57.33	57.33

Teknik evaluasi yang digunakan pada penelitian ini adalah Akurasi [10] menggunakan *cross-validation 10 folds* [11], dengan rumus sebagai berikut:

$$Akurasi = \frac{\text{jumlah prediksi benar}}{\text{jumlah total data yang diprediksi}} \times 100 \dots\dots\dots(3)$$

Dengan menggunakan rumus diatas, akurasi yang diperoleh menggunakan metode K-NN adalah:

Tabel 12. Tabel akurasi K-NN

Ranking	Folds ke-	Akurasi
1	10	100 %
2	6	96.67 %
3	9	93.33 %
4	4	93.33 %
5	2	93.33 %
6	7	90 %
7	1	90 %
8	5	86.67 %
9	3	86.67 %
10	8	80 %
Rata-rata		91 %

Sumber: (Aldi & Maspiyanti, 2018)

Berdasarkan evaluasi hasil pengukuran akurasi dengan *cross validation 10 folds* pada Tabel 3. Akurasi tertinggi terdapat pada *fold* ke 10 dengan akurasi 100%, akurasi terendah terdapat pada *fold* ke 8 dengan akurasi 80% dan rata-rata akurasi dari seluruh *fold* adalah 91%. Sedangkan hasil akurasi ID3 dapat dilihat pada Tabel 4 dibawah ini.

Tabel 13. Tabel akurasi ID3

Ranking	Folds ke-	Akurasi
1	7	76.67 %
2	1	70 %

3	9	70 %
4	6	66.67 %
5	2	66.67 %
6	8	60 %
7	5	56.67 %
8	4	53.33 %
9	3	50 %
10	10	46.67 %
Rata-rata		61.33 %

Berdasarkan evaluasi hasil pengukuran akurasi dengan cross validation 10 *folds* pada tabel 13. Akurasi tertinggi terdapat pada *fold* ke 7 dengan akurasi 76.67%, akurasi terendah terdapat pada *fold* ke 10 dengan akurasi 46.67% dan rata-rata akurasi dari seluruh *fold* adalah 61.33%.

Seperti yang terlihat pada tabel 12 dan 13 perbandingan akurasi rata-rata antara K-NN dan ID3 cukup jauh berbeda, K-NN memiliki akurasi rata-rata yang lebih baik yaitu 91% sedangkan akurasi rata-rata ID3 hanya 61.33%. Akurasi Tertinggi pada metode K-NN ada pada *fold* ke 10 dengan akurasi 100% sementara pada algoritma ID3 akurasi tertinggi ada pada *fold* ke 7 dengan akurasi 76.67 %. Hal ini membuktikan bahwa K-NN adalah metode yang lebih baik atau memiliki akurasi yang lebih tinggi daripada ID3 dalam melakukan klasifikasi tingkat stres mahasiswa dengan 300 data yang dilakukan pada Fakultas Teknik Universitas Pancasila.

IV. KESIMPULAN

metode K-NN dan ID3 dapat diimplementasikan pada klasifikasi data *mining* dengan kasus tingkat stres pada mahasiswa.

Berdasarkan evaluasi hasil pengukuran akurasi dengan teknik *cross-validation* 10 *folds*. Perbandingan hasilnya cukup jauh berbeda, K-NN memiliki akurasi rata-rata 91% sementara ID3 hanya 61%. Akurasi tertinggi pada K-NN terdapat pada *fold* 10 dengan akurasi 100%. Sementara akurasi tertinggi pada ID3 ada pada *fold* 7 dengan akurasi 76.67%. Jadi, K-NN memiliki hasil akurasi klasifikasi yang lebih baik dari ID3.

Penentuan atau pemilihan jumlah *k* pada metode K-NN untuk menentukan seberapa banyak tetangga yang akan dilakukan pengurutan berdasarkan jarak *euclidean* terdekat yang akan mempengaruhi terhadap hasil penentuan kelas dari setiap data. Berdasarkan pengujian yang dilakukan oleh sistem, jika jumlah *k* terlalu besar atau jumlah *k* lebih dari

50% dari total data *training* maka akurasi yang diperoleh lebih buruk jika dibandingkan dengan *k* kurang dari 50%. Hal ini terjadi karena jumlah label kelas stres ringan hanya 25 dari 300 data. Melihat label stres ringan yang sedikit maka dilakukan pengujian dengan jumlah *k* 5, 3 dan 1, akurasi yang diperoleh adalah 91%, 86.3% dan 86%. Jadi jumlah *k* yang tepat pada penelitian ini adalah 5 atau *k* sama dengan atau kurang dari 10% dari total data *training*.

DAFTAR PUSTAKA

- [1] Suwartika, I. Nurdin, A & Ruhmadi, E . (2014). Analisis Faktor Yang Berhubungan dengan Tingkat Stress Akademik Mahasiswa Reguler Program Studi DIII Keperawatan Cirebon Poltekkes Kemenkes. Tasikmalaya: The Soedirman Journal Of Nursing Volume 9, No.3.
- [2] Wicaksana, B. Y. (2017). Hubungan Antara Stres Akademik dan Kecenderungan Impulsive Buying Pada Mahasiswa. Yogyakarta: Universitas Sanata Dharma.
- [3] Anelia, N. (2012). Hubungan Tingkat Stres Dengan Mekanisme Koping Pada Mahasiswa Reguler Program Profesi Ners FIK UI Tahun Akademik 2011/2012. Depok: Universitas Indonesia.
- [4] Womble, L. P. (2001). Impact of Stress Factors on College Students Academic Performance. University of North Carolina at Charlotte.
- [5] Larose, D. T dan Larose, C. D. (2014). Discovering Knowledge in Data An Intoduction to Data Mining (2nd Edition). New Jersey: John Wiley & Sons, Inc.
- [6] Kartika, J. Santoso, E. & Sutrisno. (2017). Penentuan Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor dan Weighted Product (Studi Kasus : SMP Negeri 3 Mejayan). Malang: Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN: 2548-964X Vol. 1, No. 5, Mei 2017, hlm. 352-360.
- [7] Johar, T. A. Yanosma, D. & Anggriani, K. (2016). Implementasi Metode K-Nearest Neighbor (K-NN) dan Simple Additive Weighting (Saw) dalam Pengambilan Keputusan Seleksi Penerimaan Anggota Paskibraka (Studi Kasus: Dinas Pemuda dan Olahraga Provinsi Bengkulu). Bengkulu: Jurnal Pseudocode, Volume III Nomor 2, ISSN 2355-5920.
- [8] Utama, T.D. (2014). Implementasi Algoritma Iterative Dichotomiser 3 pada Penyeleksian Program Mahasiswa Wirausaha UNS. Surakarta: Jurnal ITSMART Vol 3. No 2 ISSN : 2301–7201.
- [9] Tyasti. A. & Ispriyanti. D. (2015). Algoritma Iterative Dichotomiser 3 (ID3) Untuk Mengidentifikasi Data Rekam Medis (Studi Kasus Penyakit Diabetes Mellitus Di Balai Kesehatan Kementerian Perindustrian, Jakarta). Semarang: ISSN: 2339-2541.
- [10] Eko Prasetyo. 2014. Data Mining – Mengolah Data Menjadi Informasi Menggunakan Matlab. ANDI, Yogyakarta.
- [11] Wibowo, A. 2017. 10 Cross Validation, (Online), (<https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/>), Diakses 27 juni 2019).