

Metode C4.5 Dan Naive Bayes Untuk Klasifikasi Kesehatan Mental Karyawan Di Tempat Kerja

Meilany Nonsi Tentua¹, Vicky Fidianoro¹, Pradana Feri Ariyanto¹

^{1,2,3}Jurusan Informatika, Universitas PGRI Yogyakarta

e-mail: ¹meilani@upy.ac.id , ² vfidiantoro@gmail.com , ³
pradanaferi1888@gmail.com

Intisari

Kesehatan mental adalah suatu hal penting yang harus selalu kita perhatikan. Karena kesehatan mental sangat berhubungan dengan kondisi kesehatan tubuh kita. Banyak faktor yang dapat mempengaruhi kesehatan mental kita di tempat kerja. Contohnya yaitu mulai dari permasalahan gaji, lingkungan yang dirasa kurang nyaman, dan teman kerja. Dengan begitu, kesehatan mental kamu dapat terpengaruh. Pada penelitian ini digunakan metode Algoritma C4.5 dan Naïve Bayes untuk mengklasifikasikan kesehatan mental pada karyawan di lingkungan kerjanya. Data diperoleh dari Kaggle yang terdiri dari 1259 record dengan 28 field dan telah diubah menjadi 15 field. Pada penelitian ini dilakukan penghitungan menggunakan formula Microsoft Excel untuk menentukan algoritma C4.5 dan Naive Bayes..

Kata kunci— kesehatan mental, C4.5, Naive Bayes

Abstract

Mental health is an important thing that we must always pay attention to. Because mental health is closely related to the health condition of our bodies. Many factors can affect our mental health at work. Examples are starting from salary issues, an environment that feels uncomfortable, and work colleagues. That way, your mental health can be affected. In this study, the C4.5 and Naïve Bayes

Metode C4.5 Dan Naive Bayes Untuk Klasifikasi Kesehatan Mental Karyawan Di Tempat Kerja

(Meilany Nonsi Tentua, Vicky Fidianoro, Pradana Feri Ariyanto)

Algorithm methods were used to classify the health of employees in their work environment. Data obtained from Kaggle which consists of 1259 records with 28 fields and has been converted into 15 fields. In this study, calculations were carried out using the Microsoft Excel formula to determine the C4.5 and Naive Bayes algorithms.

Keywords— Mental health, C4.5, Naive Bayes

PENDAHULUAN

Kantor atau tempat kerja merupakan suatu tempat yang dirasa paling penting untuk mendiskusikan sebuah masalah kesehatan atau penyakit mental. Namun di satu sisi, banyak karyawan atau pegawai untuk menyembunyikan kesehatan mentalnya. Mereka takut untuk mendiskusikannya dengan tema kerjanya. alasannya seorang karyawan tidak ingin mendiskusinya dengan alasan mereka tidak mau kehilangan pekerjaannya dan merusak hubungan dengan teman kerja dan dengan atasannya. Data mining adalah suatu algoritma di dalam menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (database) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. Dalam penelitian kita menggunakan metode C4.5 dan metode Naive Bayes untuk mengklasifikasikan penyakit kesehatan mental pada tempat kerja..

METODE PENELITIAN

Dalam penelitian ini data kita menggunakan data yang sudah ada. Data yang kita peroleh bersumber dari alamat web <https://www.kaggle.com/osmi/mental-health-in-tech-survey>. Data ini merupakan hasil survey 1259 orang di tempat kerja.

Variabel yang terdapat pada data kesehatan mental tersebut sebagai berikut:

- a. Age (usia)
- b. Gender (Jenis Kelamin)
- c. self_employed
- d. family_history
- e. treatment
- f. work_interfere
- g. remote_work
- h. tech_company
- i. benefits
- j. care_options

Metode C4.5 Dan Naive Bayes Untuk Klasifikasi Kesehatan Mental Karyawan Di Tempat Kerja

(Meilany Nonsi Tentua, Vicky Fidianoro, Pradana Feri Ariyanto)

- k. wellness_program
- l. seek_help
- m. anonymity
- n. leave
- o. mental_health_consequence (Yes/No/Maybe)

Berdasarkan data yang diperoleh, tidak perlu lagi dilakukan data integration, hal ini disebabkan karena tempat penyimpanan yang digunakan hanya bersumber dari satu tempat penyimpanan saja, sehingga tidak diperlukan adanya proses penyatuan tempat penyimpanan. Sedangkan proses data cleaning dan data reduction perlu dilakukan. Hal ini disebabkan karena dalam data yang diperoleh masih terdapat kemungkinan adanya data yang bernilai kosong, tidak konsisten atau mungkin tupel yang kosong (missing values dan noisy) serta adanya kemungkinan terjadi duplikasi atau terdapat tupel yang sama. Data tersebut bisa di lihat pada tabel 1 berikut:

Tabel 1 Dataset kesehatan mental

Age	Gender	Full employment	Family history	Treatment	Work stress	Stress work	Work company	Respects	Long service	Adverse opinion	Seek help	Anonymity	Leave	Mental health consequence
adult	Female	NA	No	Yes	Often	No	Yes	Yes	Not sure	No	Yes	Yes	Sometimes	NA
adult	Male	NA	No	No	Rarely	No	No	Difficult	NA	Difficult	Difficult	Difficult	Difficult	Maybe
adult	Male	NA	No	No	Rarely	No	Yes	NA	NA	No	No	Difficult	Difficult	NA
adult	Male	NA	Yes	Yes	Often	No	Yes	NA	Yes	No	No	No	Sometimes	Yes
adult	Male	NA	No	No	Never	Yes	Yes	NA	NA	Difficult	Difficult	Difficult	Difficult	NA
adult	Male	NA	Yes	No	Sometimes	No	Yes	NA	Not sure	No	Difficult	Difficult	Difficult	NA
adult	Female	NA	Yes	Yes	Sometimes	Yes	Yes	NA	NA	No	No	No	Sometimes	Maybe
adult	Male	NA	No	No	Never	Yes	Yes	NA	Yes	No	No	Yes	Difficult	NA
adult	Female	NA	Yes	Yes	Sometimes	No	Yes	Yes	Yes	No	No	No	Very difficult	Maybe
adult	Male	NA	No	No	Never	No	Yes	Difficult	NA	Difficult	Difficult	Difficult	Difficult	NA
adult	Male	NA	No	Yes	Sometimes	Yes	Yes	Difficult	NA	No	No	Difficult	Difficult	NA
adult	Male	NA	No	No	Never	Yes	Yes	Difficult	Not sure	No	No	Difficult	Difficult	NA
adult	Female	NA	Yes	Yes	Sometimes	No	No	Yes	Yes	No	No	Difficult	Sometimes	Yes
adult	Male	NA	Yes	No	Never	No	Yes	Difficult	Not sure	No	Difficult	Difficult	Difficult	NA
adult	Male	NA	No	No	Never	No	Yes	Difficult	Not sure	Difficult	Difficult	Difficult	Sometimes	NA
adult	Female	NA	Yes	Yes	Rarely	No	Yes	Yes	Not sure	No	No	Difficult	Sometimes	NA
adult	Male	NA	No	Yes	Sometimes	Yes	Yes	Difficult	NA	Difficult	Difficult	Difficult	Very easy	Maybe

Sumber : <https://www.kaggle.com/osmi/mental-health-in-tech-survey>.

1. Algoritma Naive Bayes

Harap mengirimkan naskah anda secara elektronik untuk direview sebagai attachments e-mail. Ketika anda mengirimkan dokumen naskah versi awal dalam format *Word.doc* satu kolom, termasuk gambar dan tabel.

Algoritma Naive Bayes adalah metode pembelajaran mesin yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris

Thomas Bayes, untuk memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Berikut rumus untuk menghitung probabilitas.

$$P(x|y) = \frac{P(y|x)p(x)}{p(y)} \quad (1)$$

Keterangan :

y = data dengan kelas yang belum diketahui

x = hipotesis data y merupakan suatu kelas spesifik

$P(x|y)$ = probabilitas hipotesis x berdasar kondisi y (*posteriori probability*)

$P(x)$ = probabilitas hipotesis x (*prior probability*)

$P(y|x)$ = probabilitas y berdasarkan kondisi pada hipotesis x

$P(y)$ = probabilitas dari y

2. Algoritma C4.5

Algoritma C4.5 adalah algoritma yang membangun pohon keputusan dan membentuk model pengetahuan untuk mengklasifikasikan data. Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

- a) Memilih atribut sebagai akar.
- b) Membuat cabang untuk tiap-tiap nilai.
- c) Membagi kasus dalam cabang.
- d) Mengulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih sebuah atribut menjadi akar, maka harus dilakukan perhitungan nilai dari atribut yang ada. Nilai gain yang paling tinggi dijadikan root pada pohon keputusan. Untuk menghitung nilai gain digunakan rumus:

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

Keterangan :

S : Himpunan Kasus

A : Atribut

N : Jumlah partisi dalam atribut

$|S_i|$: jumlah kasus pada partisi ke-i

$|S|$: Jumlah kasus

Sedangkan untuk perhitungan nilai entropy dapat dilakukan dengan rumus:

$$Entropy(S) = \sum_{i=1}^n -P_i \log_2 P_i \quad (3)$$

Keterangan:

S : Himpunan kasus,

n : Jumlah partisi dalam atribut,

P_i : Proporsi dari S_i terhadap S

Metode C4.5 Dan Naive Bayes Untuk Klasifikasi Kesehatan Mental Karyawan Di Tempat Kerja

(Meilany Nonsi Tentua, Vicky Fidianoro, Pradana Feri Ariyanto)

HASIL DAN PEMBAHASAN

1. Perhitungan manual algoritma C4.5.

Perhitungan secara manual dilakukan menggunakan aplikasi Microsoft excel. Perhitungan ini dilakukan untuk menentukan nilai gain dan entropy dari setiap atribut. Kemudian dari perhitungan tersebut akan didapatkan pohon keputusan sebagai aturan guna membantu penentuan kesehatan mental.

Rumus untuk mencari entropy menggunakan Microsoft Excel sebagai berikut:

	A	B	C	D	E	F	G
4							
5		children	1	0	1	0	0
6	Age	adult	1240	286	481	473	$=((-D6/C6)*LN(D6/C6))+((-E6/C6)*LN(E6/C6))+((-F6/C6)*LN(F6/C6))$
7		adolescents	18	6	8	4	1,0608569

Sedangkan untuk mencari Gain di Microsoft Excel menggunakan rumus sebagai berikut :

	A	B	C	D	E	F	G	H	I	J	K
4											
5		children	1	0	1	0					
6	Age	adult	1240	286	481	473					
7		adolescents	18	6	8	4	$=G3-((C6/C3*G6)+(C7/C3*G7)+(C5/C3*G5))$				
8											0,0851873

Langkah pertama yaitu melakukan perhitungan entropy dan gain dari masing- masing atribut. Berdasarkan hasil perhitungan node 1 pada Tabel 2 didapatkan nilai gain tertinggi tertinggi 0,0851873 pada Nilai Leave. Nilai Leave dapat dijadikan sebagai akar atau node untuk pohon keputusan. Nilai Leave memiliki 5 atribut yaitu Don't know, Somewhat difficult, Somewhat easy, Very difficult, Very easy. Dari Kelima atribut tersebut belum mempunyai keputusan, maka selanjutnya masih dilakukan perhitungan entropy dan gain untuk masing- masing atribut sampai menghasilkan keputusan. Perhitungan node 1 ditunjukkan pada Tabel 2.

Tabel 2 Perhitungan node 1

		mi	Yes	No	Maybe	Entropy	Gain
		1259	202	406	477	1,077941	
Age	children	1	0	1	0	0	
	adult	1240	286	403	471	1,0778038	
	adolescents	18	6	8	4	1,068669	0,00541
Gender	female	258	73	78	111	0,980451	
	Male	1031	129	438	366	1,063627	0,083011
	NA	18	2	11	1	0,980089	
self_employed	Yes	146	31	72	41	1,031888	
	No	1095	259	437	439	1,0758795	0,004879
family_history	Yes	492	134	105	293	1,0877522	
	No	797	168	375	284	1,0515649	0,008807
	NA	637	124	239	263	1,0829832	
treatment	Yes	622	138	280	231	1,0421938	
	No	1259	202	406	477	1,077941	0,000005
	NA	264	37	136	96	0,9663333	
work_interfere	Never	213	38	97	78	1,0318995	
	Often	144	50	47	13	1,0398045	
	Rarely	173	37	71	65	1,0631794	
remote_work	Sometimes	485	135	145	285	1,0832449	0,023289
	Yes	376	94	137	133	1,0772644	
	No	883	108	333	352	1,0660771	0,005972
tech_company	Yes	1031	125	430	376	1,0647872	
	No	228	67	80	301	1,0718796	0,007841
	NA	477	103	175	199	1,0633704	
benefits	Yes	376	112	139	133	1,0940771	
	No	408	77	176	155	1,0496844	0,007099
	Don't know	444	108	173	163	1,0369682	
care_options	Yes	521	127	180	194	1,0895662	
	Not sure	314	57	137	120	1,0392327	0,008219
	No	229	49	105	75	1,053038	
wellness_program	Yes	812	124	286	312	1,0498631	
	No	188	19	99	70	0,9370586	0,081138
	Don't know	250	47	120	83	1,0329821	
seek_help	Yes	646	180	221	245	1,0907332	
	No	883	45	189	143	1,0390002	
	Don't know	375	57	105	123	0,930205	0,008033
anonymity	Yes	45	42	8	15	0,8744117	
	No	819	199	397	319	1,0739832	0,049604
	Don't know	183	125	176	262	1,0328338	
leave	Somewhat difficult	125	54	35	37	1,0787717	
	Somewhat easy	266	34	127	105	0,9838427	
	Very difficult	96	61	11	26	0,8981138	
leave	Very easy	206	38	141	47	0,8096297	0,061187

Berdasarkan hasil perhitungan node 1.1 untuk kondisi Nilai Leave Don't Know didapatkan nilai gain tertinggi tertinggi 0,36445014 pada Nilai work_interfere. Ada lima nilai atribut pada nilai work_interfere yaitu NA, Never, Often, Rarely dan Sometimes. Dari kelima atribut tersebut juga belum menentukan keputusan sehingga masih dilakukan proses perhitungan.

Tabel 3. Perhitungan node 1.1 Atribut Leave nilai atribut Don't know

Leave [Don't know]		mi	Yes	No	Maybe	Entropy	Gain
Total		563	125	176	262	1,03811913	
Age	children	0	0	0	0	0	0,0000590
	adult	550	124	175	250	1,03360567	
	adolescents	4	1	1	2	1,039730771	
Gender	Female	117	32	22	63	1,003140	0,01077206
	Male	446	93	154	199	1,054132778	
self_employed	Yes	27	3	10	14	0,95250443	0,01087947
	No	537	122	174	240	1,03085148	
	NA	9	0	7	2	0,92950109	
family_history	Yes	260	52	48	104	1,03248703	0,00822205
	No	309	73	128	158	1,026811446	
treatment	Yes	254	64	62	128	1,03889375	0,00256833
	No	309	61	114	134	1,05001885	
work_interfere	NA	141	19	38	17	0,88948294	0,03645014
	Never	111	27	38	46	1,07589204	
	Often	45	13	5	27	0,90394856	
remote_work	Rarely	78	13	29	36	1,03343218	
	Sometimes						
tech_company	Yes	147	37	43	67	1,04912505	0,003948612
	No	416	88	133	195	1,048486132	
benefits	Yes	456	102	146	208	1,057664087	0,003807008
	No	107	23	54	30	1,03282318	
care_options	Yes	223	54	54	115	1,03848008	0,014303687
	No	108	30	29	49	1,057439005	
	Don't know	232	41	33	58	1,036758835	
wellness_program	Yes	148	35	36	77	1,03480	0,01008018
	No	212	64	76	112	1,07000	
	Not sure	163	26	64	73	1,01962	
seek_help	Yes	84	25	28	31	1,054718403	0,02404891
	No	372	90	97	189	1,041211341	
	Don't know	107	10	31	66	0,937828251	
anonymity	Yes	95	18	37	40	1,048657601	0,004755391
	No	244	61	66	117	1,023681097	
	Don't know	224	46	73	105	1,049639394	0,00345303

Metode C4.5 Dan Naive Bayes Untuk Klasifikasi Kesehatan Mental Karyawan Di Tempat Kerja

(Meilany Nonsi Tentua, Vicky Fidianoro, Pradana Feri Ariyanto)

Berdasarkan hasil perhitungan node 1.2 untuk kondisi Nilai Leave Somewhat difficult didapatkan nilai gain tertinggi tertinggi 0,053315186 pada Nilai anonymity. Ada 3 nilai atribut pada nilai anonymity yaitu Yes, No, Don't know. Dari kelima atribut tersebut juga belum menentukan keputusan sehingga masih dilakukan proses perhitungan.

Tabel 4. Perhitungan node 1.2 Atribut Leave nilai atribut Somewhat difficult

Leave (Somewhat difficult)						
	Yes	No	Maybe	Entropy	Gain	
Total	126	54	35	1,09271773		
children	0	0	0	0	0,00000000	
Age	171	51	34	1,06157596		
adolescents	5	0	1	0,90977379		
Gender	32	15	6	1,09610071	0,00746292	
Female	94	39	29	1,08129335		
Male	22	6	8	0,9804812	0,04829238	
Yes	100	46	21	1,05279786		
self_employed	4	2	1	1,08972077		
NA	60	29	11	1,02862587	0,00000000	
Family_history	66	25	24	1,08692596		
Yes	82	37	18	1,08775528	0,06216421	
No	44	17	17	1,07151912		
treatment	14	4	4	1,08093226	0,0132499	
NA	19	8	7	1,04379186		
Never	17	10	3	0,94801280		
Often	17	8	5	1,05130		
Rarely	63	29	16	1,06139543		
work_interfere	37	16	18	1,09525333	0,02617380	
Sometimes	89	40	18	1,07010486		
remote_work	98	40	27	1,05500	0,01457206	
Yes	28	14	6	0,99128		
No	29	17	7	1,08748752	0,02696489	
tech_company	59	27	12	1,04883907		
Yes	28	10	8	1,05847478		
benefits	40	22	10	1,04281233	0,007348421	
Don't know	21	10	17	1,02038145		
care_options	30	13	8	1,07603400		
Yes	17	8	3	1,0033996	0,02186123	
No	104	46	28	1,07215546		
wellness_program	5	2	1	1,04022346		
Don't know	23	1	1	1,06601348	0,01957528	
Yes	86	41	23	1,05602107		
seek_help	17	4	7	1,07339362		
Don't know	36	15	15	1,07790337	0,053315186	
anonymity	15	17	1	0,70791811		
Yes	26	10	10	1,06666667		

Berdasarkan hasil perhitungan node 1.3 untuk kondisi Nilai Leave Somewhat easy didapatkan nilai gain tertinggi tertinggi 0,053771400 pada Nilai work_interfere. Ada lima nilai atribut pada nilai work_interfere yaitu NA, Never, Often, Rarely dan Sometimes. kelima atribut tersebut belum menentukan keputusan sehingga masih dilakukan proses perhitungan.

Tabel 5.Perhitungan node 1.3 Atribut Leave nilai atribut Somewhat easy

Leave (Somewhat easy)						
		Yes	No	Maybe	Don't know	Gain
Total		269	36	127	129	0.98284272
Age	children	1	0	1	0	0.00961540
	adults	268	36	126	129	
Gender	Female	58	9	27	27	0.00962031
	Male	268	27	100	102	
self_employed	Yes	39	5	39	13	0.03089040
	No	229	31	88	116	
family_history	Yes	107	18	46	43	0.00805804
	No	162	18	81	86	
treatment	Yes	135	27	51	53	0.01823953
	No	134	9	76	76	
work_interface	NA	53	0	12	21	0.03771480
	Others	46	2	36	36	
remote_work	Yes	39	5	39	13	0.03089040
	No	229	31	88	116	
tech_company	Yes	219	21	115	63	0.02048636
	No	49	15	12	66	
benefits	Yes	109	15	53	47	0.00399131
	No	80	21	64	82	
care_options	Yes	111	16	56	41	0.00700748
	No	88	20	71	88	
wellness_program	Yes	60	14	33	27	0.00700748
	No	167	22	94	102	
work_help	Yes	49	15	36	26	0.03771480
	No	140	21	81	103	
anonymity	Yes	39	5	39	13	0.03089040
	No	229	31	88	116	

Berdasarkan hasil perhitungan node 1.4 untuk kondisi Nilai Leave Very difficult didapatkan nilai gain tertinggi tertinggi 0,064161935 pada Nilai self_employed. Ada tiga nilai atribut pada nilai Very difficult yaitu NA, Yes, No, Rarely dan Sometimes. Nilai atribut NA menghasilkan keputusan Maybe sedangkan Nilai atribut Yes dan No tersebut belum menentukan keputusan sehingga masih dilakukan proses perhitungan.

Tabel 5.Perhitungan node 1.4 Atribut Leave nilai atribut Very difficult

Metode C4.5 Dan Naive Bayes Untuk Klasifikasi Kesehatan Mental Karyawan Di Tempat Kerja

(Meilany Nonsi Tentua, Vicky Fidianoro, Pradana Feri Ariyanto)

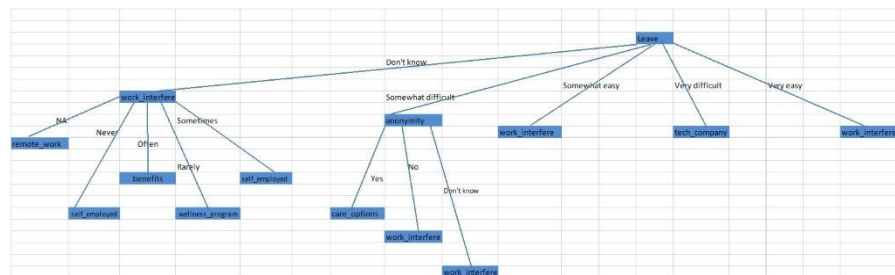
Leave (Very difficult)								
Total		jml	Yes	No	Maybe	Entropy	Gain	
		98	61	11	26	0,852013767		
Age	children	0	0	0	0	0	0	
	adult	98	61	11	26	0,852013767		
	adolescents	0	0	0	0	0		
Gender	Female	20	15	2	3	0,730588061	0,010714203	
	Male	78	46	9	23	0,920697386		
self_employed	Yes	25	15	7	3	0,917357388	0,064161595	
	No	72	46	4	22	0,809088124		
	NA	1	0	0	1	0		
family_history	Yes	43	29	4	10	0,82579064	0,00459353	
	No	55	32	7	16	0,936672467		
treatment	Yes	67	37	10	20	0,972085766	0,029351513	
	No	31	24	1	6	0,626766277		
work_interfere	NA	7	5	0	2	0,598270	0,038828543	
	Never	6	4	0	2	0,636514108		
	Often	24	15	5	4	0,919173829		
	Rarely	11	5	2	4	1,03619879		
	Sometimes	50	32	4	14	0,844112426		
remote_work	Yes	38	26	5	7	0,838136934	0,011207669	
	No	60	35	6	19	0,908809903		
tech_company	Yes	79	49	9	21	0,85991384	0,000071214	
	No	19	12	2	5	0,878525099		
benefits	Yes	21	11	1	9	0,846814381	0,025394025	
	No	61	39	9	13	0,897787519		
	Don't know	16	11	1	4	0,777462132		
care_options	Yes	48	27	8	13	0,976053399	0,01659077	
	No	36	25	2	9	0,760374183		
	Not sure	14	9	1	4	0,830471712		
wellness_program	Yes	10	3	3	4	1,088899975	0,038546871	
	No	80	54	6	20	0,806147375		
	Don't know	8	4	2	2	1,039720771		
seek_help	Yes	13	4	4	5	1,092880754	0,045222671	
	No	75	49	6	20	0,832629488		
	Don't know	10	8	1	1	0,63903186		
anonymity	Yes	26	11	5	10	1,048485442	0,063672598	
	No	21	18	2	1	0,501046914		
	Don't know	51	32	4	15	0,85203192		

Berdasarkan hasil perhitungan lanjutan node 1.5 untuk kondisi Nilai Leave Very easy didapatkan nilai gain tertinggi tertinggi 0,262567273 pada Nilai self_employed. pada Nilai work_interfere. Ada lima nilai atribut pada nilai work_interfere yaitu NA, Never, Often, Rarely dan Sometimes. kelima atribut tersebut belum menentukan keputusan sehingga masih dilakukan proses perhitungan.

Tabel 6. Perhitungan node 1.5 Atribut Leave nilai atribut Very easy

Leave (Very easy)						
		Yes	No	Missing	Entropy	Gain
Total		205	18	341	0.762625006	
Age	children	0	0	0	0	(0.0)0.00000
	adult	201	17	337	0.642914795	(0.0)0.00000
	adolescents	5	1	4	0.60905829	
Gender	Female	31	2	17	0.613671001	(0.0)0.00000
	Male	175	16	324	0.642914795	
Self-employed	Yes	14	2	21	0.60905829	(0.0)0.00000
	No	191	16	320	0.642914795	
Family_history	Yes	16	5	15	0.60905829	(0.0)0.00000
	No	189	13	326	0.642914795	
Treatment	Yes	101	9	60	0.62411161	(0.0)0.00000
	No	104	9	281	0.642914795	
Risk	NA	49	4	31	0.62886646	(0.0)0.00000
	Normal	17	2	26	0.60905829	
	Often	21	2	13	0.60905829	
	Rarely	28	1	19	0.60905829	
Work_interfere	Sometimes	71	7	48	0.642914795	
	Never	79	5	58	0.62886646	
Remote_work	Yes	127	13	81	0.60905829	(0.0)0.00000
	No	78	5	260	0.642914795	
Tech_company	Yes	129	13	129	0.60905829	(0.0)0.00000
	No	76	5	212	0.642914795	
Benefit	Yes	86	8	60	0.60905829	(0.0)0.00000
	Don't know	69	5	42	0.60905829	
New_options	Yes	69	8	41	0.60905829	(0.0)0.00000
	No	67	5	45	0.60905829	
Wellness_program	Yes	42	4	38	0.60905829	(0.0)0.00000
	No	113	12	21	0.642914795	
Work_help	Yes	11	3	33	0.60905829	(0.0)0.00000
	Don't know	46	11	42	0.60905829	
Anonymity	Yes	107	8	82	0.60905829	(0.0)0.00000
	Don't know	98	10	259	0.642914795	

Setelah dilakukan perhitungan entropy dan gain maka akan terbentuk pohon keputusan seperti pada gambar 1.



Gambar 1. Pohon keputusan

2. Perhitungan Naive Bayes

Pada perhitungan Naive Bayes data dibagi menjadi 2 bagian yaitu data training dan data testing. Untuk data kesehatan mental yang berjumlah 1259 kita bagi menjadi dua yaitu 1200 untuk data training dan 59 untuk data testing. Selanjutnya menghitung Probabilitas dari masing masing atribut dengan rumus seperti dibawah ini.

$$P(x|y) = \frac{P(y|x)p(x)}{p(y)}$$

Keterangan :

y = data dengan kelas yang belum diketahui

x = hipotesis data y merupakan suatu kelas spesifik

Metode C4.5 Dan Naive Bayes Untuk Klasifikasi Kesehatan Mental Karyawan Di Tempat Kerja

(Meilany Nonsi Tentua, Vicky Fidianoro, Pradana Feri Ariyanto)

$P(x|y)$ = probabilitas hipotesis x berdasar kondisi y (*posteriori probanility*)

$P(x)$ = probabilitas hipotesis x (*prior probability*)

$P(y|x)$ = probablitas y berdasarkan kondisi pada hipotesis x

$P(y)$ = probabilitas dari y

Untuk perhitungan Naïve Bayes kita lakukan menggunakan Microsoft Excel. Dari Data Training yang berjumlah 1200 terdapat 14 atribut yaitu Age, Gender, self_employed, Family_History, treatment, work_interfere, remote_work, tech_company, benefit, care options, wellnes_program, seek_help, anonymity, dan keputusan kesehatan mental atau mental_health_consequence.

Dalam metode Naive Bayes kita lakukan perhitungan probabilitas masing masing klasifikasi. Klasifikasi tersebut mempunyai tiga nilai yaitu Yes, No, Maybe.

$$P(\text{Yes}) = 277/1200 = 0,2308$$

$$P(\text{No}) = 471/1200 = 0,3925$$

$$P(\text{Maybe}) = 452/1200 = 0,3767$$

Setelah menghitung probabilitas pada klasifikasi ,selanjutnya menghitung probabilitas tiap-tiap atribut.

Tabel 1. Perhitungan Probabilitas pada nilai Age

C1 Age	Jumlah Kejadian Dipilih			Probabilitas		
	Yes	No	Maybe	Yes	No	Maybe
adult	271	462	448	0,9783	0,9809	0,9912
children	0	1	0	0	0,0021	0
adolescents	6	8	4	0,0217	0,017	0,0088
JML	277	471	452	1	1	1

Tabel 2. Perhitungan Probabilitas pada nilai Gender

C2						
Gender	Jumlah Kejadian Dipilih			Probabilitas		
	Yes	No	Maybe	Yes	No	Maybe
Female	70	70	108	0,2527	0,1486	0,2443
Male	207	401	334	0,7473	0,8514	0,7557
JML	277	471	442	1	1	1

Tabel 3. Perhitungan Probabilitas pada nilai self_employed

C3						
self_employed	Jumlah Kejadian Dipilih			Probabilitas		
	Yes	No	Maybe	Yes	No	Maybe
NA	2	11	5	0,0072	0,0234	0,0111
Yes	31	69	41	0,1119	0,1465	0,0907
No	244	391	406	0,8809	0,8301	0,8982
JML	277	471	452	1	1	1

Tabel 4. Perhitungan Probabilitas pada nilai Family_history

C4						
family_history	Jumlah Kejadian Dipilih			Probabilitas		
	Yes	No	Maybe	Yes	No	Maybe
Yes	129	155	183	0,4657	0,3291	0,4049
No	148	316	269	0,5343	0,6709	0,5951
JML	277	471	452	1	1	1

Tabel 5. Perhitungan Probabilitas pada nilai treatment

C5						
treatment	Jumlah Kejadian Dipilih			Probabilitas		
	Yes	No	Maybe	Yes	No	Maybe
Yes	165	194	239	0,5957	0,4119	0,5288
No	112	277	213	0,4043	0,5881	0,4712
JML	277	471	452	1	1	1

Tabel 6. Perhitungan Probabilitas pada nilai Work_interfere

C6						
work_interfere	Jumlah Kejadian Dipilih			Probabilitas		
	Yes	No	Maybe	Yes	No	Maybe
NA	30	134	90	0,1083	0,2845	0,1991
Never	36	96	75	0,13	0,2038	0,1659
Often	47	37	46	0,1637	0,0786	0,1018
Rarely	36	66	63	0,13	0,1401	0,1394
Sometimes	128	138	178	0,4621	0,293	0,3938
JML	277	471	452	1	1	1

Tabel 7. Perhitungan Probabilitas pada nilai Remote_work

Metode C4.5 Dan Naive Bayes Untuk Klasifikasi Kesehatan Mental Karyawan Di Tempat Kerja

(Meilany Nonsi Tentua, Vicky Fidianoro, Pradana Feri Ariyanto)

C7						
remote_work	Jumlah Kejadian Dipilih			Probabilitas		
	Yes	No	Maybe	Yes	No	Maybe
Yes	89	151	119	0,3213	0,3206	0,2633
No	188	320	333	0,6787	0,6794	0,7367
JML	277	471	452	1	1	1

Tabel 8. Perhitungan Probabilitas pada nilai tech_company

C8						
tech_company	Jumlah Kejadian Dipilih			Probabilitas		
	Yes	No	Maybe	Yes	No	Maybe
Yes	212	414	358	0,7653	0,879	0,792
No	65	57	94	0,2347	0,121	0,208
JML	277	471	452	1	1	1

Tabel 9. Perhitungan Probabilitas pada nilai Benefit

C9						
benefits	Jumlah Kejadian Dipilih			Probabilitas		
	Yes	No	Maybe	Yes	No	Maybe
Yes	98	168	192	0,3538	0,3567	0,4248
No	106	131	114	0,3827	0,2781	0,2522
Don't know	73	172	146	0,2635	0,3652	0,323
JML	277	471	452	1	1	1

Tabel 10. Perhitungan Probabilitas pada nilai care_options

C10						
care_options	Jumlah Kejadian Dipilih			Probabilitas		
	Yes	No	Maybe	Yes	No	Maybe
Yes	100	162	155	0,361	0,3439	0,3429
No	121	176	183	0,4368	0,3737	0,4049
Not sure	56	133	114	0,2022	0,2824	0,2522
JML	277	471	452	1	1	1

Tabel 11. Perhitungan Probabilitas pada nilai wellness_program dan seek_help

C11						
wellness_program	Jumlah Kejadian Dipilih			Probabilitas		
	Yes	No	Maybe	Yes	No	Maybe
Yes	47	101	70	0,1697	0,2144	0,1549
No	211	275	313	0,7617	0,5839	0,6325
Don't know	19	95	69	0,0686	0,2017	0,1527
JML	277	471	452	1	1	1
C12						
seek_help	Jumlah Kejadian Dipilih			Probabilitas		
	Yes	No	Maybe	Yes	No	Maybe
Yes	46	112	78	0,1661	0,2378	0,1726
No	168	211	230	0,6065	0,448	0,5088
Don't know	63	148	144	0,2274	0,3142	0,3186
JML	277	471	452	1	1	1

Tabel 12. Perhitungan Probabilitas pada kolom anonymity dan seek_help

C13						
anonymity	Jumlah Kejadian Dipilih			Probabilitas		
	Yes	No	Maybe	Yes	No	Maybe
Yes	53	189	118	0,1913	0,4013	0,2611
No	40	6	14	0,1444	0,0127	0,031
Don't know	184	276	320	0,6643	0,586	0,708
JML	277	471	452	1	1	1
C14						
leave	Jumlah Kejadian Dipilih			Probabilitas		
	Yes	No	Maybe	Yes	No	Maybe
Don't know	117	169	250	0,4224	0,3588	0,5531
Somewhat difficult	52	34	35	0,1877	0,0722	0,0774
Somewhat easy	31	120	95	0,1119	0,2548	0,2102
Very difficult	53	11	26	0,213	0,0234	0,0575
Very easy	18	137	46	0,065	0,2909	0,1018
JML	277	471	452	1	1	1

Berdasarkan data tersebut apabila:

Age(adult), Gender(Male), self_employed (No), Family_history(No),
treatmen(Yes), work_interfere(Often), remote_work(Yes), tech_company(Yes),
benefits(No), care_options(Yes), wellnes(No), seek_help(No), anonyity(Don't
know), leave(Dont know) maka dapat dihitung :

Yes =

$$0,9783*0,7473*0,8809*0,5343*0,5957*0,1697*0,3213*0,7653*0,3827*0,361*0,7617*0,6065*0,6643*0,4224 = 0,000153138$$

No =

$$0,9809*0,8514*0,8301*0,6709*0,4119*0,0786*0,3206*0,879*0,2781*0,3439*0,5839*0,448*0,586*0,3588 = 0,00002231$$

Metode C4.5 Dan Naive Bayes Untuk Klasifikasi Kesehatan Mental Karyawan Di Tempat Kerja

(Meilany Nonsi Tentua, Vicky Fidianoro, Pradana Feri Ariyanto)

$$\begin{aligned} \text{Maybe} &= \\ &0,9912 * 0,7557 * 0,8982 * 0,5951 * 0,5288 * 0,1018 * 0,2633 * 0,792 * 0,2522 * 0,3429 * 0, \\ &6925 * 0,5088 * 0,708 * 0,5531 = 0,00005361 \end{aligned}$$

Dari hasil perhitungan tersebut nilai P(Yes) lebih besar dari nilai No dan Maybe. Jadi dapat disimpulkan bahwa Age(adult), Gender(Male), self_employed (No), Family_history(No), treatmen(Yes), work_interfere(Often), remote_work(Yes), tech_company(Yes), benefits(No), care_options(Yes), wellnes(No), seek_help(No), anonyity(Don't know), leave(Dont know) masuk kedalam klasifikasi Yes.

Dalam penelitian ini, dapat disimpulkan bahwa :

1. untuk menentukan kesehatan mental di tempat kerja dapat diprediksi dengan memanfaatkan teknik data mining dengan menggunakan metode algoritma C4.5 dan Naive Bayes.
2. Berdasarkan metode decision tree atribut berat badan merupakan atribut yang paling berpengaruh dalam penelitian ini. Dibuktikan dengan atribut Leave menduduki root node.
3. Pada perhitungan Naive Bayes data dibagi menjadi 2 bagian yaitu data training dan data testing.
4. Untuk data kesehatan mental yang berjumlah 1259 kita bagi menjadi dua yaitu 1200 untuk data training dan 59 untuk data testing. Selanjutnya menghitung Probabilitas dari masing masing atribut.

DAFTAR PUSTAKA

- [1] E. Rahmawati, "Vol. XII No. 2, September 2015 Jurnal Techno Nusa Mandiri," *Techno Nusa Mandiri*, vol. XII, no. 2, pp. 21–26, 2015.
- [2] R. Kurniasari and A. Fatmawati, "Penerapan Algoritma C4.5 Untuk Penjurusan Siswa Sekolah Menengah Atas," *Komputa J. Ilm. Komput. dan Inform.*, vol. 8, no. 1, pp. 19–27, 2019, doi: 10.34010/komputa.v8i1.3045.
- [3] B. Sugara, D. Widyatmoko, B. S. Prakoso, and D. M. Saputro, "Penerapan Algoritma C4.5 untuk Deteksi Dini Autisme Pada Anak," *Semin. Nas. Teknol. Inf. dan Komun.*, vol. 2018, no. Sentika, pp. 87–96, 2018.
- [4] cnnindonesia.com.(2016,22 Agustus).Pentingnya Bicara Masalah Mental di Lingkungan Kerja.Diakses pada 27 Desember 2020 dari <https://www.cnnindonesia.com/gaya-hidup/20160822153230-255-153027/pentingnya-bicara-masalah-mental-di-lingkungan-kerja>
- [5] halodoc.com.(2019 Desember 16).Ini Hubungan Kesehatan Mental dengan Lingkungan Kerja.Diakses pada 27 Desember 2020 dari <https://www.halodoc.com/artikel/ini-hubungan-kesehatan-mental-dengan-lingkungan-kerja>