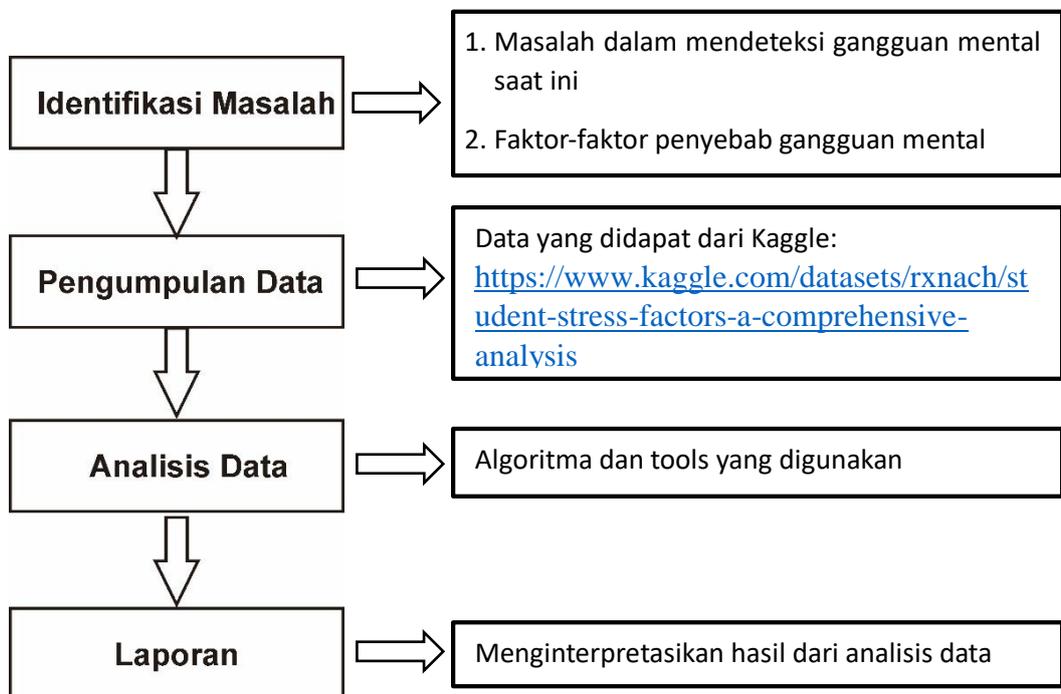


BAB III METODE PENELITIAN

3.1 Alur Penelitian

Pada bab ini akan membahas langkah-langkah dari proses penelitian yang akan dilaksanakan, dalam melakukan analisa dan mencari pola data untuk dijadikan sebuah dataset untuk memudahkan penelitian dan dapat berjalan dengan sistematis sehingga memenuhi tujuan yang diinginkan maka dibuat alur tahapan penelitian secara detail pada gambar 3.1 :



Gambar 3.1 Alur Penelitian

1. Identifikasi Masalah

Pada tahap identifikasi masalah, peneliti melakukan langkah-langkah untuk mengidentifikasi atau merumuskan masalah utama yang akan diselesaikan dalam penelitian. Dalam pengidentifikasian masalah, peneliti memastikan bahwa masalah tersebut diungkap dengan jelas dan terperinci. Untuk mencapai tujuan penelitian, peneliti harus menggali fakta-fakta empiris yang relevan dan mengumpulkan data yang diperlukan. Langkah awal dalam identifikasi masalah adalah melakukan observasi melalui media digital atau pengumpulan berbagai informasi yang didapat dari buku, jurnal atau artikel terdahulu. Tahap ini juga peneliti menyusun masalah secara jelas dan terperinci, sehingga tujuan penelitian dapat dicapai dengan efektif. Dalam hal ini ditemukan masalah bahwa kekurangan metode yang objektif untuk mendeteksi gangguan mental pada remaja, hal ini dapat mengakibatkan ketidakpastian dalam proses diagnosis. Kurangnya metode yang bersifat objektif dapat menghambat pemahaman yang tepat terkait gejala dan tingkat keparahan, sehingga mempersulit pemberian diagnosis yang akurat. Serta didapat juga beberapa faktor penyebab gangguan mental pada remaja yaitu

1. Gangguan Kecemasan

Gangguan kecemasan adalah salah satu bentuk gangguan psikologis yang melibatkan reaksi tubuh terhadap situasi yang mengancam. Berbeda dari rasa cemas pada umumnya, gangguan kecemasan membuat penderitanya mengalami rasa takut dan reaksi cemas yang berlebihan. Kondisi ini juga dapat semakin buruk dari waktu ke waktu. Menurut *Anxiety and Depression Association of America*, gangguan kecemasan adalah gangguan mental yang paling umum di Amerika Serikat, memengaruhi 40 juta orang dewasa yang berusia 18 tahun ke atas. Gangguan kecemasan terdiri dari beberapa jenis yaitu: Gangguan fobia, gangguan panik, dan gangguan GAD (*Generalized anxiety disorder*). Gangguan fobia adalah ketakutan yang kuat atau keengganan terhadap akan suatu objek atau situasi tertentu. Gangguan panik adalah munculnya rasa ketakutan yang

intens yang datang dengan cepat dan hanya berlangsung dalam beberapa menit secara tiba-tiba. Sedangkan gangguan GAD adalah gangguan yang menunjukkan kecemasan atau kekhawatiran yang berlebihan hampir setiap hari setidaknya selama 6 bulan.

2. Gangguan Mood

Gangguan mood adalah kondisi perubahan suasana hati atau emosi yang tidak biasa. Di mana kondisi tersebut sampai mengganggu kegiatan sehari-hari seseorang. Gangguan mood terdiri dari 2 jenis yang paling umum terjadi yaitu, bipolar dan depresi. Bipolar adalah seseorang yang mengalami kondisi ini akan menunjukkan gejala perubahan mood atau suasana hati yang ekstrem. Depresi merupakan salah satu masalah kesehatan mental serius yang berbahaya dan banyak terjadi. Gangguan depresi terbagi menjadi depresi mayor dan depresi persisten atau disebut juga distimia. Depresi mayor adalah jenis depresi yang membuat penderitanya merasa sedih dan putus asa sepanjang waktu. Sedangkan depresi persisten atau distimia merupakan istilah yang digunakan untuk menggambarkan kondisi depresi yang bersifat kronis

3. Gangguan Somatoform

Gangguan somatoform merupakan kelainan psikologis pada seseorang yang ditandai dengan sekumpulan keluhan fisik yang tidak menentu, namun tidak tampak saat pemeriksaan fisik. Munculnya gangguan ini biasanya disebabkan oleh stres dan banyak pikiran. Gangguan somatoform terdiri dari gangguan nyeri, gangguan dismorfik tubuh, gangguan hipokondriasis, gangguan konversi dan gangguan somatisasi. Gangguan nyeri adalah suatu kondisi ketika seseorang merasakan rasa sakit terus menerus yang setelah diperiksa tidak bisa ditemukan penyebab fisiknya. Gangguan dismorfik tubuh merupakan kondisi ketika penderitanya lebih banyak menghabiskan waktu untuk khawatir terkait penampilan tubuhnya. Gangguan hipokondriasis adalah kondisi ketika seseorang takut bahwa gejala ringan yang dialaminya disebabkan oleh penyakit serius. Gangguan

konversi adalah kondisi ketika penderitanya memiliki gejala yang menunjukkan adanya penyakit serius pada sistem syaraf, namun tidak dapat ditelusuri penyebab medisnya. Gangguan somatisasi adalah keluhan fisik di berbagai bagian tubuh yang disebabkan oleh stres atau beban mental yang berat [1].

2. Pengumpulan Data

Setelah peneliti mengidentifikasi masalah, langkah selanjutnya adalah mencari data yang diperlukan untuk penelitian. Dalam hal ini, peneliti memilih untuk menggunakan dataset publik yang tersedia di platform Kaggle. Sebelum menggunakan dataset dari Kaggle, peneliti perlu mengamati sumber data tersebut dan memeriksa validitas dan keandalannya. Hal ini diperlukan untuk memastikan bahwa data yang digunakan sesuai dengan tujuan penelitian dan dapat diandalkan. Setelah memilih dataset yang sesuai, peneliti dapat mengunduhnya dari Kaggle dan menyimpannya dalam format yang sesuai dengan kebutuhan penelitian.

3. Analisis Data

Pada tahap ini, peneliti akan melakukan analisis terhadap dataset yang telah ditentukan menggunakan perbandingan antara algoritma C4.5 dan Naïve Bayes dengan Particle Swarm Optimization (PSO) menggunakan tools RapidMiner. Analisis data dilakukan untuk melihat dan membandingkan hasil dari kedua algoritma tersebut guna memperoleh solusi terbaik bagi permasalahan yang telah diidentifikasi sebelumnya.

Algoritma C4.5 adalah salah satu algoritma pembelajaran mesin yang digunakan dalam pengklasifikasian data. Algoritma ini membentuk pohon keputusan yang berfokus pada pemisahan data berdasarkan atribut-atribut yang relevan. Sedangkan, algoritma Naïve Bayes adalah metode yang berdasarkan pada teorema Bayes dengan asumsi independensi antar atribut. Sedangkan Particle Swarm Optimization (PSO) ialah metode atau teknik untuk optimalisasi peningkatan akurasi penggunaan algoritma tersebut. Algoritma ini cocok untuk digunakan dalam pengklasifikasian data dengan

atribut yang tinggi. Dengan menggunakan tools RapidMiner, peneliti akan mengimplementasikan kedua algoritma tersebut pada dataset yang telah disiapkan. Setelah itu, peneliti akan melakukan analisis terhadap hasil yang diperoleh dari masing-masing algoritma. Dalam analisis data ini, peneliti akan memperhatikan beberapa faktor, seperti akurasi, presisi, recall, dan F1-score.

Dengan membandingkan hasil dari kedua algoritma tersebut, diharapkan peneliti dapat memilih algoritma yang paling sesuai untuk memecahkan permasalahan yang telah diidentifikasi. Pilihan algoritma yang tepat akan menghasilkan solusi yang lebih efektif dan efisien bagi permasalahan yang sedang dihadapi.

4. Laporan

Langkah selanjutnya, dari hasil analisis yang dilakukan, peneliti menginterpretasikan perbandingan antara algoritma C4.5 dan Naïve Bayes dengan mengambil makna dari temuan yang relevan. Pola, hubungan, atau tren yang ditemukan dalam konteks yang relevan, kemudian dijelaskan dengan detail. Hal ini bertujuan agar pemahaman terhadap hasil yang diperoleh dapat lebih mendalam.

Selain menjelaskan secara verbal, peneliti juga menggunakan visualisasi seperti grafik, tabel, atau diagram untuk mempresentasikan pola, hubungan, atau tren yang ditemukan. Grafik dapat membantu dalam melihat perbandingan antara hasil yang diperoleh menggunakan algoritma C4.5 dan Naïve Bayes. Tabel atau diagram juga dapat digunakan untuk menunjukkan perbedaan ataupun persamaan yang signifikan antara kedua algoritma tersebut.

Visualisasi ini juga dapat membantu dalam menggambarkan hubungan antara variabel-variabel yang terlibat dalam analisis, sehingga mempermudah pemahaman pembaca mengenai temuan yang ditemukan. Dengan bentuk visualisasi ini, informasi yang dihasilkan menjadi lebih mudah dipahami oleh pembaca. Dengan menjelaskan pola, hubungan, atau

tren yang dihasilkan dari perbandingan antara dua algoritma tersebut, hasil analisis menjadi lebih bermakna dan dapat dijadikan landasan untuk pengambilan keputusan atau langkah selanjutnya.

3.2 Pengujian Algoritma C4.5

Pengujian data yang digunakan dalam algoritma C4.5 adalah dataset yang didapatkan dari Kaggle. Data yang digunakan dalam pengujian ini dapat dilihat pada gambar berikut



Row No.	stress_level	prediction(s...	confidence(1)	confidence(2)	confidence(0)	anxiety_level	self_esteem	mental_heal...	depression	headache	blood_press...	sleep_quality	breathing_D
1	1	1	1	0	0	14	20	0	11	2	1	2	4
2	2	2	0.008	0.992	0	15	8	1	15	5	3	1	4
3	1	1	1	0	0	12	18	1	14	2	1	2	2
4	1	1	0.381	0.286	0.333	16	28	0	7	2	3	5	1
5	2	2	0.008	0.992	0	20	13	1	21	3	3	1	4
6	0	0	0	0	1	4	26	0	6	1	2	4	1
7	2	2	0.008	0.992	0	17	3	1	22	4	3	1	5
8	1	2	0.008	0.992	0	6	8	0	27	4	3	1	2
9	2	2	0.008	0.992	0	17	12	1	25	4	3	1	3
10	2	2	0.008	0.992	0	17	15	1	22	3	3	1	5
11	0	0	0	0	1	5	28	0	8	1	2	4	2
12	2	2	0.008	0.992	0	9	23	1	24	4	3	1	0
13	1	1	1	0	0	11	21	0	14	3	1	2	4
14	0	0	0	0	1	6	28	0	1	1	2	4	2
15	0	0	0	0	1	7	25	0	3	1	2	4	2
16	1	1	1	0	0	11	23	0	12	3	1	2	2
17	2	2	0.008	0.992	0	21	1	1	25	4	3	1	4
18	0	0	0	0	1	3	27	0	0	1	2	4	1
19	2	2	0.008	0.992	0	18	1	1	21	4	3	1	3
20	0	0	0	0	1	7	27	0	5	1	2	4	1
21	2	2	0.008	0.992	0	20	5	1	26	3	3	1	4
22	2	2	0.008	0.992	0	18	6	1	27	5	3	1	5
23	0	1	0.381	0.286	0.333	7	28	0	20	2	3	3	1
24	1	1	1	0	0	13	23	1	14	2	1	3	2
25	0	0	0	0	1	0	27	0	3	1	2	5	1

Gambar 3.2 Pengujian Algoritma C4.5

Berikut penjelasan singkat mengenai masing-masing atribut:

1. *Anxiety Level*: Tingkat kecemasan seseorang, dapat berkaitan dengan ketegangan, kekhawatiran, dan gejala fisik seperti detak jantung meningkat, pernapasan cepat, dan keringat dingin.
2. *Self-esteem*: Tingkat harga diri seseorang, sejauh mana seseorang memiliki persepsi positif atau negatif tentang dirinya sendiri.

3. *Mental Health History*: Riwayat kesehatan mental seseorang, termasuk riwayat gangguan mental sebelumnya atau masalah psikologis yang telah dialami.
4. *Depression*: Gangguan suasana hati yang serius, ditandai dengan perasaan sedih, kehilangan minat atau kegembiraan, gangguan tidur, dan perubahan berat badan.
5. *Headache*: Sakit kepala, gejala yang dapat disebabkan oleh berbagai faktor seperti stres, ketegangan otot, atau masalah kesehatan lainnya.
6. *Blood Pressure*: Tekanan darah seseorang, yang bisa menjadi indikator kesehatan fisik dan faktor risiko untuk berbagai penyakit jantung dan vaskular.
7. *Sleep Quality*: Kualitas tidur seseorang, mencakup durasi, kedalaman, dan kepuasan tidur.
8. *Breathing Problem*: Masalah pernapasan yang bisa disebabkan oleh berbagai faktor termasuk asma, alergi, atau kondisi medis lainnya.
9. *Noise Level*: Tingkat kebisingan lingkungan sekitar seseorang, yang dapat mempengaruhi kesehatan fisik dan mental.
10. *Living Conditions*: Kondisi tempat tinggal seseorang, mencakup aspek seperti kebersihan, keamanan, dan kenyamanan lingkungan.
11. *Safety*: Tingkat keamanan seseorang dalam lingkungan mereka, termasuk dari ancaman fisik dan psikologis.
12. *Basic Needs*: Kepuasan kebutuhan dasar seseorang, seperti makanan, tempat tinggal, dan keamanan.
13. *Academic Performance*: Kinerja akademik seseorang di sekolah atau perguruan tinggi.
14. *Study Load*: Beban belajar atau tugas akademik yang harus dihadapi seseorang.
15. *Teacher-Student Relationship*: Hubungan antara guru dan siswa, yang dapat mempengaruhi motivasi belajar dan kesejahteraan emosional siswa.
16. *Future Career Concerns*: Kecemasan atau perhatian terkait pilihan karier di masa depan seseorang.

17. *Social Support*: Dukungan sosial yang diterima seseorang dari keluarga, teman, atau masyarakat dalam mengatasi masalah dan tantangan.
18. *Peer Pressure*: Tekanan dari teman sebaya yang bisa mempengaruhi perilaku dan keputusan seseorang.
19. *Extracurricular Activities*: Aktivitas di luar kurikulum akademik yang diikuti seseorang, seperti olahraga, klub, atau organisasi.
20. *Bullying*: Perlakuan yang merendahkan dan merugikan seseorang secara fisik, verbal, atau sosial oleh orang lain.
21. *Stress Level*: Tingkat stres yang dialami seseorang, dapat berasal dari berbagai sumber seperti sekolah, pekerjaan, atau masalah pribadi.

Semua atribut ini dapat saling terkait dan mempengaruhi kesejahteraan dan kinerja seseorang secara keseluruhan. Penting untuk memahami dan mengelola faktor-faktor ini dengan baik untuk mempromosikan kesejahteraan fisik, mental, dan sosial yang optimal.

Perhitungan Manual Algoritma C4.5

Tabel 3.1 Perhitungan Manual Algoritma C4.5

Attribute		Jumlah	0	1	2	Entropy	Info Gain	Split Info	Gain Ratio
		15	3	5	7	1,5058231			
anxiety_level									
	<11	5	3	1	1	1,3710	0,141530204	0,918295834	0,154122668
	>=11	10	1	4	5	1,3610			
self_esteem									
	<18	7	1	1	5	1,1488	0,220026002	0,996791632	0,220734198
	>=18	8	3	4	1	1,4056			
mental_health_history									
	0	6	2	3	1	1,4591	0,187527266	0,970950594	0,193137805
	1	9	1	2	6	1,2244			
depression									
	<13	6	2	3	1	1,4591	0,187527266	0,970950594	0,193137805
	>=13	9	1	2	6	1,2244			
headache									
	1	3	1	1	1	1,5850	1,1888306	2,181937064	0,544851004
	2	3	1	1	0	0,0000			
	3	3	0	1	2	0,0000			
	4	5	0	1	4	0,0000			
	5	1	0	0	1	0,0000			

Tabel 3.1 Perhitungan Manual Algoritma C4.5 (Lanjutan)

blood pressure	1	3	1	1	1	1,5850	0,137201433	1,370950594	0,100077591
	2	3	1	1	1	1,5850			
	3	9	1	2	6	1,2244			
sleep quality									
	1	8	1	1	6	1,0613	0,6228156	1,672905595	0,372295724
	2	3	1	1	1	0,0000			
	3	3	1	1	1	1,5850			
	4	1	0	1	0	0,0000			
breathing problem									
	0	1	0	0	1	0,0000	1,1058231	2,440223929	0,4531646
	1	2	1	1	0	0,0000			
	2	4	1	2	1	1,5000			
	3	2	0	0	2	0,0000			
	4	4	1	0	3	0,0000			
	5	2	0	0	2	0,0000			
noise level									
	0	1	1	0	0	0,0000	0,191856037	2,40971477	0,079617737
	1	3	2	0	1	0,0000			
	2	4	1	2	1	1,5000			
	3	5	1	1	3	1,3710			
	4	2	0	0	2	0,0000			
	5	1	0	0	1	0,0000			

Tabel 3.1 Perhitungan Manual Algoritma C4.5 (Lanjutan)

living conditions									
	1	2	0	0	2	0,0000	0,462390803	1,421620806	0,325256075
	2	7	1	2	4	1,3788			
	3	4	2	1	1	1,5000			
	4	1	1	0	0	0,0000			
	5	1	0	1	0	0,0000			
safety									
	1	3	0	0	2	0,0000	1,1888306	1,74185331	0,682509023
	2	4	0	1	3	0,0000			
	3	3	1	1	1	1,5850			
	4	4	0	2	2	0,0000			
	5	1	1	0	0	0,0000			
basic needs									
	1	4	1	0	3	0,0000	1,1888306	1,684869778	0,705591978
	2	5	0	2	3	0,0000			
	3	3	1	1	1	1,5850			
	4	2	0	1	1	0,0000			
	5	1	1	0	0	0,0000			
academic performance									
	1	5	1	0	4	0,0000	1,5058231	1,811995818	0,831030119
	2	4	0	2	2	0,0000			
	3	2	0	2	0	0,0000			
	4	2	1	1	0	0,0000			
	5	2	2	0	0	0,0000			

Tabel 3.1 Perhitungan Manual Algoritma C4.5 (Lanjutan)

study load									
	1	1	1	0	0	0,0000	0,364528735	1,436824992	0,253704339
	2	5	2	2	1	1,5219			
	3	6	2	2	2	1,5850			
	4	2	0	0	2	0,0000			
	5	1	0	0	1	0,0000			
teacher student relationship									
	1	6	1	2	3	1,4591	0,922163933	1,641201643	0,561883384
	2	3	0	1	2	0,0000			
	3	3	0	2	1	0,0000			
	4	2	2	0	0	0,0000			
	5	1	1	0	0	0,0000			
future career concerns									
	1	3	1	1	1	1,5850	1,1888306	1,82486081	0,651463714
	2	2	0	2	0	0,0000			
	3	3	0	2	1	0,0000			
	4	4	1	0	3	0,0000			
	5	3	0	1	2	0,0000			
social support									
	0	1	0	0	1	0,0000	0,413132465	1,252638516	0,329809805
	1	8	1	2	5	1,2988			
	2	2	0	2	0	0,0000			
	3	4	2	1	1	1,5000			
peer pressure									

Tabel 3.1 Perhitungan Manual Algoritma C4.5 (Lanjutan)

	1	3	2	0	1	0,0000	1,1888306	2,181937064	0,544851004
	2	1	1	0	0	0,0000			
	3	3	1	1	1	1,5850			
	4	5	1	0	4	0,0000			
	5	3	0	2	1	0,0000			
extracurricular activities									
	0	2	0	1	1	0,0000	1,1058231	1,93171977	0,572455238
	1	1	1	0	0	0,0000			
	2	4	1	2	1	1,5000			
	3	2	0	2	0	0,0000			
	4	4	1	0	3	0,0000			
	5	2	0	0	2	0,0000			
bullying									
	1	4	1	1	2	1,5000	0,356490798	1,889246429	0,188694705
	2	3	1	1	1	1,5850			
	4	2	0	1	1	0,0000			
	5	6	1	1	5	1,0808			

Gain Rtio MAX

0,831030119

Berikut adalah penjelasan tabel perhitungan manual diatas:

1. *Attribute*: Merupakan atribut atau faktor yang menjadi pertimbangan dalam analisis data.
2. Jumlah: Menunjukkan jumlah total data yang ada dalam masing-masing kategori atribut.
3. 0, 1, 2: Merupakan nilai atau kategori dari atribut yang sedang dianalisis.
4. *Entropy*: Merupakan ukuran dari ketidakpastian atau keacakan di dalam data. Semakin rendah nilai *entropy*, semakin pasti atau teratur data tersebut.
5. *Info Gain (Information Gain)*: Merupakan perbedaan antara *entropy* sebelum dan sesudah pemisahan data menggunakan atribut tertentu. *Info Gain* yang tinggi menunjukkan bahwa atribut tersebut memberikan informasi yang berguna dalam klasifikasi data.
6. *Split Info*: Merupakan ukuran seberapa baik suatu atribut membagi data. Semakin tinggi nilai *Split Info*, semakin baik atribut tersebut membagi data.
7. *Gain Ratio*: Merupakan rasio antara *Info Gain* dan *Split Info*. *Gain Ratio* yang tinggi menunjukkan bahwa atribut tersebut lebih baik dalam membagi data dibandingkan dengan atribut lainnya.

Dalam tabel tersebut, terdapat beberapa atribut yang dianalisis, yaitu:

- *Anxiety Level*: Tingkat kecemasan.
- *Self Esteem*: Penghargaan diri.
- *Mental Health Histor*: Riwayat kesehatan mental.
- *Depression*: Depresi.
- *Headache*: Sakit kepala.

Setiap atribut memiliki nilai-nilai tertentu yang dibagi berdasarkan kategori atau rentang tertentu, dan kemudian dihitung nilai *entropy*, *info gain*, *split info*, dan *gain ratio*-nya untuk mengevaluasi seberapa baik atribut tersebut dalam mempengaruhi klasifikasi data.

3.3 Pengujian Algoritma Naïve Bayes

Pengujian data yang digunakan dalam algoritma Naïve Bayes adalah yang didapatkan dari Kaggle. Data yang digunakan dalam pengujian ini dapat dilihat pada gambar berikut.



Row No.	stress_level	prediction(s...	confidence(1)	confidence(2)	confidence(0)	anxiety_level	self_esteem	mental_heat...	depression	headache	blood_press...	sleep_quality	breathing_p
1	1	1	1	0	0	14	20	0	11	2	1	2	4
2	2	2	0.008	0.992	0	15	8	1	15	5	3	1	4
3	1	1	1	0	0	12	18	1	14	2	1	2	2
4	1	1	0.381	0.286	0.333	16	28	0	7	2	3	5	1
5	2	2	0.008	0.992	0	20	13	1	21	3	3	1	4
6	0	0	0	0	1	4	26	0	6	1	2	4	1
7	2	2	0.008	0.992	0	17	3	1	22	4	3	1	5
8	1	2	0.008	0.992	0	6	8	0	27	4	3	1	2
9	2	2	0.008	0.992	0	17	12	1	25	4	3	1	3
10	2	2	0.008	0.992	0	17	15	1	22	3	3	1	5
11	0	0	0	0	1	5	28	0	8	1	2	4	2
12	2	2	0.008	0.992	0	9	23	1	24	4	3	1	0
13	1	1	1	0	0	11	21	0	14	3	1	2	4
14	0	0	0	0	1	6	28	0	1	1	2	4	2
15	0	0	0	0	1	7	25	0	3	1	2	4	2
16	1	1	1	0	0	11	23	0	12	3	1	2	2
17	2	2	0.008	0.992	0	21	1	1	25	4	3	1	4
18	0	0	0	0	1	3	27	0	0	1	2	4	1
19	2	2	0.008	0.992	0	18	1	1	21	4	3	1	3
20	0	0	0	0	1	7	27	0	5	1	2	4	1
21	2	2	0.008	0.992	0	20	5	1	26	3	3	1	4
22	2	2	0.008	0.992	0	18	6	1	27	5	3	1	5
23	0	1	0.381	0.286	0.333	7	28	0	20	2	3	3	1
24	1	1	1	0	0	13	23	1	14	2	1	3	2
25	0	0	0	0	1	0	27	0	3	1	2	5	1

Gambar 3.3 Pengujian Algoritma Naïve Bayes

Diatas adalah tampilan datas yang akan dilakukan pengujian manual menggunakan algoritma Naïve Bayes. Terdapat beberapa atribut yang sudah dijelaskan pada halaman sebelumnya yang mana atribut ini dapat saling terkait dan mempengaruhi kesejahteraan dan kinerja seseorang secara keseluruhan.

Perhitungan Manual Algoritma Naïve Bayes

Tabel 3.2 Perhitungan Manual Algoritma Naïve Bayes

MEAN																					
Stres Level	anxiety_level	self_esteem	mental_health_histor	depression	headache	blood_pressure	sleep_quality	breathing_problem	noise_level	living_conditions	safety	basic_needs	academic_performance	study_load	teacher_student_relationship	future_career_concerns	social_support	peer_pressure	extracurricular_activities	bullying	
0	14	14,66666667	0,333333333	7,333333333	0,333333333	2,666666667	4,333333333	3,333333333	3	4	3,333333333	2,666666667	3,333333333	1,666666667	2,333333333	2	1,333333333	1,333333333	3,666666667	1,333333333	
1	12,2	19,2	0,4	14,2	2,6	1,8	2,4	2,6	2	3	3	2,4	2,8	2,6	2	3	1,8	3,8	2	3	
2	13	10,71428571	0,714285714	21,42857143	3,285714286	2,857142857	1,714285714	3,428571429	3,285714286	2,428571429	2	2,285714286	1,428571429	3,285714286	1,857142857	3,571428571	0,857142857	2,714285714	3,285714286	3,285714286	
STANDAR DEVIASI																					
Stres Level	anxiety_level	self_esteem	mental_health_histor	depression	headache	blood_pressure	sleep_quality	breathing_problem	noise_level	living_conditions	safety	basic_needs	academic_performance	study_load	teacher_student_relationship	future_career_concerns	social_support	peer_pressure	extracurricular_activities	bullying	
0	8,524474568	11,46976702	0,471404521	4,18993503	0,471404521	0,471404521	0,471404521	1,247219129	1,632993162	0,816496581	0,471404521	0,942809042	0,942809042	1,247219129	2,054804668	1,414213562	1,247219129	0,471404521	1,247219129	0,471404521	
1	3,370459909	6,523802572	0,489897949	6,794115101	0,8	0,979795897	1,356465997	1,2	1,095445115	1,095445115	0,632455532	0,489897949	0,748331477	0,489897949	0,894427191	1,095445115	0,748331477	0,979795897	1,095445115	1,264911064	
2	4,810702354	6,712765785	0,451753951	4,370588155	1,484614978	0,349927106	1,160576915	1,590789818	1,577908717	1,178030179	1,069044968	1,030157507	0,728431359	1,030157507	1,124858268	0,903507903	0,349927106	1,829464068	1,665986256	1,749635531	
PROBABILITAS STRES LEVEL																					
Stres Level	NILAI																				
0	0,133333333																				
1	0,333333333																				
2	0,466666667																				
Jumlah	0,933333333																				
Distribusi Gaussian																					
Data Tes																					
	anxiety_level	self_esteem	mental_health_histor	depression	headache	blood_pressure	sleep_quality	breathing_problem	noise_level	living_conditions	safety	basic_needs	academic_performance	study_load	teacher_student_relationship	future_career_concerns	social_support	peer_pressure	extracurricular_activities	bullying	stress_level
	2	28	0	3	1	2	4	2	1	3	4	4	4	2	5	1	3	1	2	1	?
0	0,050743012	0,059951584	0,452636619	0,114196146	0,213810399	0,213810399	0,452636619	0,201781117	0,147505133	0,20860376	0,213810399	0,151186783	0,320062423	0,344776994	0,119925969	0,261329874	0,146313994	0,452636619	0,146313994	0,452636619	1,32076E-15
1	0,002230711	0,06290132	0,408509857	0,039341885	0,060379042	0,394824931	0,170882592	0,321471699	0,251344179	0,381263181	0,143759674	0,002752518	0,127524771	0,269306296	0,001521743	0,072011313	0,127524771	0,006793338	0,381263181	0,10165344	4,71671E-24
2	0,013322458	0,005593588	0,17009874	2,63101E-05	0,100113783	0,0335852	0,053261713	0,211395157	0,111256659	0,326850009	0,067066686	0,098454947	0,000920105	0,180434581	0,007591399	0,007314111	4,853E-09	0,190192053	0,229537755	0,128513404	3,55076E-37
																					Nilai Max
																					1,32076E-15

Penjelasan perhitungan manual algoritma Naïve Bayes dari tabel yang diberikan adalah sebagai berikut:

1. *Mean* (Rata-rata): Rata-rata dihitung untuk setiap fitur (kolom) dalam data training untuk setiap kelas *stress level*.
2. *Standar Deviasi*: *Standar deviasi* dihitung untuk setiap fitur (kolom) dalam data training untuk setiap kelas *stress level*. Standar deviasi mengukur seberapa jauh titik data tersebar dari rata-ratanya.
3. *Probabilitas Stres Level*: *Probabilitas* setiap kelas *stress level* dihitung dengan membagi jumlah sampel pada kelas *stress level* tersebut dengan jumlah total sampel.
4. *Distribusi Gaussian*: *Distribusi Gaussian* digunakan untuk menghitung *probabilitas* fitur tertentu yang diberikan *stress level* tertentu pada data tes. *Distribusi Gaussian* digunakan karena fitur-fitur tersebut dianggap terdistribusi normal.

Dalam contoh ini, data tes memiliki nilai untuk setiap fitur. Kemudian, probabilitas fitur yang diberikan stres level tertentu dihitung menggunakan *distribusi Gaussian* dengan memasukkan nilai fitur ke dalam rumus *distribusi Gaussian* untuk setiap fitur dan *stress level*. Proses ini membantu dalam mengklasifikasikan data tes ke dalam kelas *stress level* yang sesuai berdasarkan *probabilitas* fitur yang diamati pada data *training*.