



Analisis performa metode Naïve Bayesh Classifier pada Electronic Nose dalam identifikasi formalin pada tahu

Fadhila Tangguh Admojo^{a,1,*}, Sitti Rahma Jabir^{b,2}

^aUniversitas Bina Darma Jl. Jenderal Ahmad Yani No.3, 9/10 Ulu, Kota Palembang dan 3011, Indonesia

^bUniversitas Muslim Indonesia, Jl. Urip Sumoharjo KM.05, Makassar dan 90231, Indonesia

*fadhila.tangguh@binadarma.ac.id, rahmajabir@umi.ac.id;

INFORMASI ARTIKEL	ABSTRAK
Diterima : 05 – 02 – 2022 Direvisi : 19 – 02 – 2023 Diterbitkan : 31 – 03 – 2023	Penelitian ini bertujuan untuk menganalisis performa metode Naive Bayes Classifier (NBC) dalam identifikasi formalin pada tahu menggunakan Electronic Nose. Hasil dari penelitian ini menunjukkan bahwa performa NBC cukup moderat, dengan nilai akurasi sekitar 0,59 hingga 0,60, presisi sekitar 0,67 hingga 0,68, recall sekitar 0,59 hingga 0,60, dan F1-score sekitar 0,55. Ini menunjukkan bahwa model mampu mengklasifikasikan beberapa titik data dengan benar, tetapi tidak semua. Walaupun demikian masih ada ruang untuk perbaikan dan perlu dipertimbangkan untuk mencoba metode lain untuk meningkatkan hasil identifikasi formalin pada tahu. Hasil ini menunjukkan bahwa metode Naive Bayes Classifier pada Electronic Nose masih belum dapat memberikan hasil yang optimal dalam identifikasi formalin pada tahu, dan hasil yang diperoleh masih tidak lebih baik dari penelitian sebelumnya.
Kata Kunci: Naïve Bayesh Classifier, Crossvalidation Akurasi, Presisi, Recall, F-Measure, Penyakit Cardiovascular	 

I. Pendahuluan

Manusia hidup di bumi tidak lepas dari kebutuhan akan makanan. Makanan merupakan sebuah kebutuhan dasar bagi manusia. Sebagai kebutuhan dasar makanan haruslah makanan yang baik, bergizi, dan juga sehat. Makanan yang tidak baik dan tidak aman untuk dikonsumsi dapat menyebabkan gangguan kesehatan bahkan bisa menyebabkan keracunan dan kematian.

Salah satu makanan yang mengandung gizi adalah tahu. Tahu merupakan makanan olahan yang kaya akan protein nabati yang banyak dikonsumsi oleh masyarakat Indonesia. Tahu adalah makanan yang dibuat dari kacang kedelai yang difermentasikan dan diambil sarinya. Berbeda dengan tempe yang asli Indonesia, tahu berasal dari negara China. Tahu sangat populer di kalangan masyarakat Indonesia bukan hanya karena harganya yang murah, tetapi juga mudah diolah menjadi makanan lain.

Berdasarkan data yang diperoleh dari Biro Pusat Statistik dan Survei Sosial Ekonomi Nasional [1], tingkat konsumsi tahu di Indonesia mencapai 0,152 kg/kapita/minggu, dan tahu masuk kategori bahan makanan penting. Jumlah ini lebih dari empat kali lipat jika dibandingkan dengan tingkat konsumsi daging ayam dan daging sapi yang merupakan sumber protein hewani. Hal tersebut disebabkan harga tahu dan tempe jauh lebih terjangkau jika dibandingkan dengan harga daging [2].

Tahu merupakan bahan pangan dengan kandungan protein yang tinggi dan kadar air mencapai 85%, sehingga tahu tidak dapat bertahan lama. Kerusakan tahu ditandai dengan bau asam dan berlendir. Berdasarkan hal tersebut membuat beberapa oknum pedagang dan pembuat tahu yang mulai memasang formalin pada tahu buatannya, karena proses pembusukan tahu yang cepat. Tahu mulai mengalami kerusakan sehari setelah produksi. Perendaman tahu dalam air yang diberi formalin akan membuat tahu menjadi lebih keras dan kenyal, sehingga tidak mudah hancur dan tahan terhadap mikroorganisme, sehingga awet dan dapat bertahan hingga tujuh hari [3].

Ini sangat berpengaruh buruk terhadap seluruh konsumen tahu yakni masyarakat Indonesia yang menjadikan tahu sebagai makanan utama dalam keseharian. Kasus penyalahgunaan formalin sebagai bahan pengawet makanan banyak dilakukan di Indonesia. Hasil uji Badan Pengawas Obat dan Makanan (BPOM)

menyatakan dari 700 sampel produk makanan yang diambil dari Jawa, Sulawesi Selatan dan Lampung, 56% men[3]andung formalin. Bahaya dari penyalahgunaan formalin sebagai bahan pengawet makanan perlu adanya perhatian khusus, karena penggunaan formalin sebagai bahan pengawet makanan dapat menyebabkan beberapa penyakit, diantaranya efek kesehatan manusia langsung terlihat akut seperti (iritasi, alergi, mual, muntah, sakit perut dan pusing), dan efek kronik yaitu efek pada kesehatan manusia terlihat terkena dalam jangka waktu yang lama dan berulang, seperti gangguan pencernaan, hati, ginjal, pankreas, sistem saraf pusat.

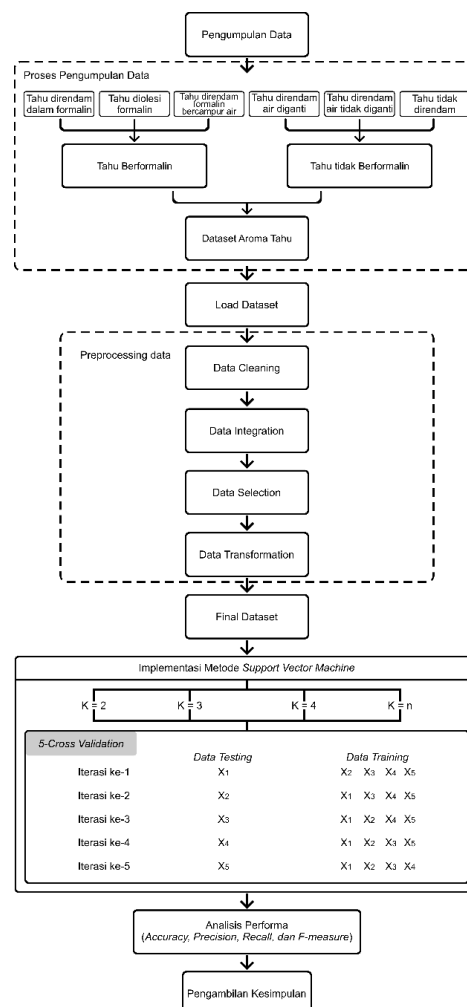
Data Mining adalah Serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basisdata dengan melakukan penggalian pola-pola dari data dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berharga yang diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basisdata[6]–[8].

Metode NBC (Naive Bayes Classifier) adalah sebuah metode klasifikasi yang menggunakan teori probabilitas untuk menentukan kelas suatu objek berdasarkan fitur-fitur yang dimilikinya[9]–[12]. Metode ini dikenal sebagai "naive" karena asumsi yang digunakan adalah bahwa setiap fitur adalah independen satu sama lain, meskipun dalam kenyataannya mungkin terdapat korelasi antar fitur. NBC digunakan dalam berbagai bidang seperti pengenalan wajah, analisis teks, dan perdagangan elektronik. Keuntungan dari metode ini adalah mudah dipahami dan diimplementasikan, serta membutuhkan data yang relatif sedikit untuk pelatihan.

Beberapa penelitian sebelumnya terkait data yang sama telah dilakukan, dimana metode yang digunakan adalah random forest classifier dan metode knn (Azis, 2020) yang memperoleh performa tertinggi yaitu 92% dilanjutkan oleh lasaiman (2021) dengan data yang sama menggunakan metode SVM dengan perolehan performa yaitu 94%.

Pada penelitian ini penulis tertarik untuk menganalisis performa metode NBC dengan menerapkan metode tersebut untuk mengklasifikasikan tahu berformalin dengan tahu yang tidak berformalin, data yang digunakan adalah data rekaman dari sensor gas, performa yang diukur dalam penelitian ini yaitu *Accuracy*, *Precision*, *Recall*, dan *F-Measure*[13]–[16].

II. Metode



Gambar 1. Tahapan Penelitian








Terdapat 4 bagian pada tahapan penelitian yang dilakukan yaitu:














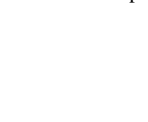
1. Tahap pengumpulan data dimana tahapan tersebut juga melakukan proses pembagian data berdasarkan label yang telah di tentukan sebelumnya
2. Tahapan preprocessing data dimana tahapn ini melakukan pembersihan data-data yang dikumpulkan.
3. Tahap penerapan algoritma dimana tahapan ini melakukan implementasi metode naïve bayesh classifier pada data yang telah dibersihkan
4. Tahap evaluasi, dimana dilakukan uji performa algoritma yang diterapkan yaitu akurasi, presisi, recall dan f-measure.

















A. Pengumpula Data












Pengumpulan data dilakukan di Laboratorium Terpadu Fakultas Ilmu Komputer Universitas Muslim Indonesia selama 15 hari, berikut rincian proses pengumpulan data pada penelitian ini ditujukan pada [Tabel 1](#) berikut:












Tabel 1. Rincian Pengumpulan data




Hari	Kategori	Perlakuan	Dokumentasi
Ke-1	Berformalin	Tahu direndam dalam formalin	
		Tahu diolesi formalin	
		Tahu direndam larutan formalin bercampur air	
Tidak Berformalin	Tidak Berformalin	Tahu direndam dengan air diganti tiap hari	
		Tahu direndam dengan air tidak diganti	
		Tahu tidak direndam	
Ke-2	Berformalin	Tahu direndam dalam formalin	

Hari	Kategori	Perlakuan	Dokumentasi
		Tahu diolesi formalin	
		Tahu direndam larutan formalin bercampur air	
	Tidak Berformalin	Tahu direndam dengan air diganti tiap hari	
		Tahu direndam dengan air tidak diganti	
		Tahu tidak direndam	
Ke-3	Berformalin	Tahu direndam dalam formalin	
		Tahu diolesi formalin	
		Tahu direndam larutan formalin bercampur air	
	Tidak Berformalin	Tahu direndam dengan air diganti tiap hari	
		Tahu direndam dengan air tidak diganti	
		Tahu tidak direndam	
Ke-4	Berformalin	Tahu direndam dalam formalin	
		Tahu diolesi formalin	
		Tahu direndam larutan formalin bercampur air	
	Tidak Berformalin	Tahu direndam dengan air diganti tiap hari	

Hari	Kategori	Perlakuan	Dokumentasi
		Tahu direndam dengan air tidak diganti	
		Tahu tidak direndam	
Ke-5	Berformalin	Tahu direndam dalam formalin	
		Tahu diolesi formalin	
		Tahu direndam larutan formalin bercampur air	
Tidak Berformalin		Tahu direndam dengan air diganti tiap hari	
		Tahu direndam dengan air tidak diganti	
		Tahu tidak direndam	
Ke-6	Berformalin	Tahu direndam dalam formalin	
		Tahu diolesi formalin	
		Tahu direndam larutan formalin bercampur air	
Tidak Berformalin		Tahu direndam dengan air diganti tiap hari	
		Tahu direndam dengan air tidak diganti	
		Tahu tidak direndam	Membusuk dan dibuang
Ke-7	Berformalin	Tahu direndam dalam formalin	
		Tahu diolesi formalin	
		Tahu direndam larutan formalin bercampur air	

Hari	Kategori	Perlakuan	Dokumentasi
	Tidak Berformalin	Tahu direndam dengan air diganti tiap hari	
		Tahu direndam dengan air tidak diganti	Membusuk dan dibuang
		Tahu tidak direndam	Membusuk dan dibuang
Ke-8	Berformalin	Tahu direndam dalam formalin	
		Tahu diolesi formalin	
		Tahu direndam larutan formalin bercampur air	
	Tidak Berformalin	Tahu direndam dengan air diganti tiap hari	Membusuk dan dibuang
		Tahu direndam dengan air tidak diganti	Membusuk dan dibuang
		Tahu tidak direndam	Membusuk dan dibuang
Ke-9	Berformalin	Tahu direndam dalam formalin	
		Tahu diolesi formalin	
		Tahu direndam larutan formalin bercampur air	
	Tidak Berformalin	Tahu direndam dengan air diganti tiap hari	Membusuk dan dibuang
		Tahu direndam dengan air tidak diganti	Membusuk dan dibuang
		Tahu tidak direndam	Membusuk dan dibuang
Ke-10	Berformalin	Tahu direndam dalam formalin	
		Tahu diolesi formalin	
		Tahu direndam larutan formalin bercampur air	
	Tidak Berformalin	Tahu direndam dengan air diganti tiap hari	Membusuk dan dibuang
		Tahu direndam dengan air tidak diganti	Membusuk dan dibuang
		Tahu tidak direndam	Membusuk dan dibuang
Ke-11	Berformalin	Tahu direndam dalam formalin	

Hari	Kategori	Perlakuan	Dokumentasi
		Tahu diolesi formalin	
		Tahu direndam larutan formalin bercampur air	
	Tidak Berformalin	Tahu direndam dengan air diganti tiap hari	Membusuk dan dibuang
		Tahu direndam dengan air tidak diganti	Membusuk dan dibuang
		Tahu tidak direndam	Membusuk dan dibuang
Ke-12	Berformalin	Tahu direndam dalam formalin	
		Tahu diolesi formalin	
		Tahu direndam larutan formalin bercampur air	
	Tidak Berformalin	Tahu direndam dengan air diganti tiap hari	Membusuk dan dibuang
		Tahu direndam dengan air tidak diganti	Membusuk dan dibuang
		Tahu tidak direndam	Membusuk dan dibuang
Ke-13	Berformalin	Tahu direndam dalam formalin	
		Tahu diolesi formalin	
		Tahu direndam larutan formalin bercampur air	
	Tidak Berformalin	Tahu direndam dengan air diganti tiap hari	Membusuk dan dibuang
		Tahu direndam dengan air tidak diganti	Membusuk dan dibuang
		Tahu tidak direndam	Membusuk dan dibuang
Ke-14	Berformalin	Tahu direndam dalam formalin	
		Tahu diolesi formalin	
		Tahu direndam larutan formalin bercampur air	
	Tidak Berformalin	Tahu direndam dengan air diganti tiap hari	Membusuk dan dibuang
		Tahu direndam dengan air tidak diganti	Membusuk dan dibuang
		Tahu tidak direndam	Membusuk dan dibuang

Hari	Kategori	Perlakuan	Dokumentasi
Ke-15	Berformalin	Tahu direndam dalam formalin	
		Tahu diolesi formalin	
		Tahu direndam larutan formalin bercampur air	
Tidak Berformalin		Tahu direndam dengan air diganti tiap hari	Membusuk dan dibuang
		Tahu direndam dengan air tidak diganti	Membusuk dan dibuang
		Tahu tidak direndam	Membusuk dan dibuang

Data yang dikumpulkan dikelompokan berdasarkan kategori tahu mengandung formalin dan tahu tidak mengandung formalin, dan diperlakukan menjadi 6 variabel/sampel yakni: Tahu yang direndam keseluruhan bagian dengan formalin, Tahu yang hanya diolesi bagian luarnya dengan formalin dan tidak direndam, Tahu direndam larutan yang telah dicampurkan air dan formalin dengan perbandingan 50:50, Tahu yang hanya direndam dalam air dan diganti airnya tiap satu hari, Tahu yang hanya direndam dalam air tanpa diganti airnya sama sekali, Tahu yang dibiarkan tanpa air.

III. Hasil dan Pembahasan

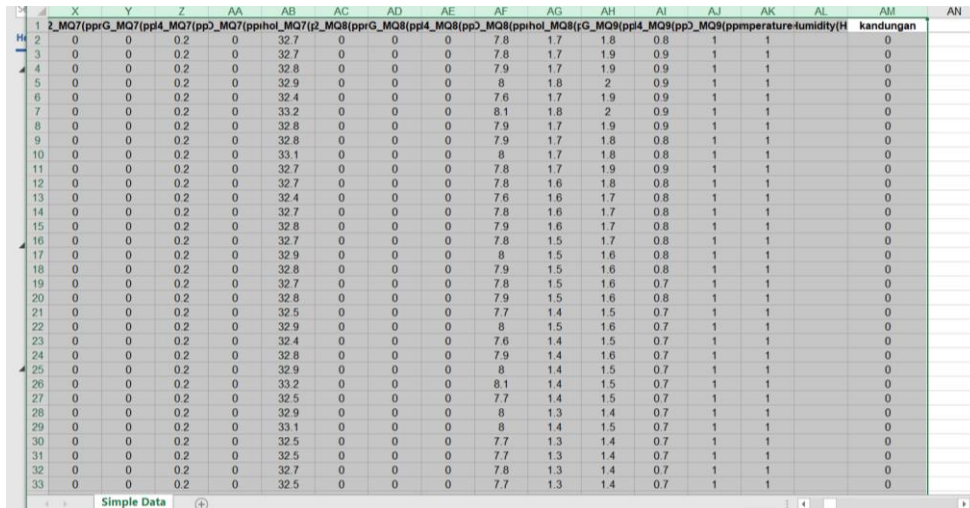
A. Hasil

1) Hasil Implementasi Preprocessing Data

Data yang di peroleh langsung dari penelitian selama 15 hari, diolah terlebih dahulu melalui tahap-tahap *data cleaning*, *data integration*, *data selection*, dan *data transformation*:

a) Data Cleaning

Dilakukan proses cleaning dari data awal yang diperoleh untuk mengatasi nilai yang hilang, noise dan data yang tidak konsisten. Maka dari data awal yang ditemukan 39 kolom perbaris yang ditunjukkan pada Gambar 2 selanjutnya dilakukan proses cleaning maka didapat data 9 kolom perbaris seperti yang ditunjukkan pada Gambar 3.



	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN
1	0	0	0.2	0	32.7	0	0	0	7.8	1.7	1.8	0.8	1	1	0		
2	0	0	0.2	0	32.7	0	0	0	7.8	1.7	1.9	0.9	1	1	0		
3	0	0	0.2	0	32.8	0	0	0	7.9	1.7	1.9	0.9	1	1	0		
4	0	0	0.2	0	32.9	0	0	0	8	1.8	2	0.9	1	1	0		
5	0	0	0.2	0	32.4	0	0	0	7.6	1.7	1.9	0.9	1	1	0		
6	0	0	0.2	0	33.2	0	0	0	8.1	1.8	2	0.9	1	1	0		
7	0	0	0.2	0	32.8	0	0	0	7.9	1.7	1.9	0.9	1	1	0		
8	0	0	0.2	0	32.8	0	0	0	7.9	1.7	1.8	0.8	1	1	0		
9	0	0	0.2	0	33.1	0	0	0	8	1.7	1.8	0.8	1	1	0		
10	0	0	0.2	0	32.7	0	0	0	7.8	1.7	1.9	0.9	1	1	0		
11	0	0	0.2	0	32.7	0	0	0	7.8	1.6	1.8	0.8	1	1	0		
12	0	0	0.2	0	32.4	0	0	0	7.6	1.6	1.7	0.8	1	1	0		
13	0	0	0.2	0	32.7	0	0	0	7.8	1.6	1.7	0.8	1	1	0		
14	0	0	0.2	0	32.8	0	0	0	7.9	1.6	1.7	0.8	1	1	0		
15	0	0	0.2	0	32.7	0	0	0	7.8	1.5	1.7	0.8	1	1	0		
16	0	0	0.2	0	32.9	0	0	0	8	1.5	1.8	0.8	1	1	0		
17	0	0	0.2	0	32.8	0	0	0	7.9	1.5	1.6	0.8	1	1	0		
18	0	0	0.2	0	32.7	0	0	0	7.8	1.5	1.6	0.7	1	1	0		
19	0	0	0.2	0	32.8	0	0	0	7.9	1.5	1.6	0.8	1	1	0		
20	0	0	0.2	0	32.5	0	0	0	7.7	1.4	1.5	0.7	1	1	0		
21	0	0	0.2	0	32.9	0	0	0	8	1.5	1.6	0.7	1	1	0		
22	0	0	0.2	0	32.4	0	0	0	7.6	1.4	1.5	0.7	1	1	0		
23	0	0	0.2	0	32.8	0	0	0	7.9	1.4	1.6	0.7	1	1	0		
24	0	0	0.2	0	32.9	0	0	0	8	1.4	1.5	0.7	1	1	0		
25	0	0	0.2	0	33.2	0	0	0	8.1	1.4	1.5	0.7	1	1	0		
26	0	0	0.2	0	32.5	0	0	0	7.7	1.4	1.5	0.7	1	1	0		
27	0	0	0.2	0	32.9	0	0	0	8	1.3	1.4	0.7	1	1	0		
28	0	0	0.2	0	33.1	0	0	0	8	1.4	1.5	0.7	1	1	0		
29	0	0	0.2	0	32.5	0	0	0	7.7	1.3	1.4	0.7	1	1	0		
30	0	0	0.2	0	32.5	0	0	0	7.7	1.3	1.4	0.7	1	1	0		
31	0	0	0.2	0	32.7	0	0	0	7.8	1.3	1.4	0.7	1	1	0		
32	0	0	0.2	0	32.5	0	0	0	7.7	1.3	1.4	0.7	1	1	0		
33	0	0	0.2	0	32.5	0	0	0	7.7	1.3	1.4	0.7	1	1	0		

Gambar 2. Dataset Sebelum Proses *Cleaning*

	A	B	C	D	E	F	G	H	I
1	H2_MQ2(ppm)	LPG_MQ2(ppm)	CO_MQ2(ppm)	Alcohol_MQ2(ppm)	Propane_MQ2(ppm)	CH4_MQ4(ppm)	Smoke_MQ4(ppm)	Temperature(C)	label
2	1.6	0.8	3.5	1.2	1	2.4	0.4	1.00	1
3	1.6	0.8	3.5	1.2	1	2.4	0.4	1.00	1
4	1.6	0.8	3.7	1.3	1.1	2.3	0.4	1.00	1
5	1.6	0.8	3.6	1.3	1	2.4	0.4	1.00	1
6	1.6	0.8	3.5	1.2	1	2.3	0.4	1.00	1
7	1.6	0.8	3.6	1.3	1	2.3	0.4	1.00	1
8	1.5	0.7	3.4	1.2	1	2.3	0.4	1.00	1
9	1.6	0.8	3.6	1.3	1	2.3	0.4	1.00	1
10	1.7	0.8	3.8	1.3	1.1	2.5	0.5	1.00	1
11	1.6	0.8	3.6	1.3	1	2.4	0.5	1.00	1
12	1.6	0.8	3.6	1.3	1	2.3	0.4	1.00	1
13	1.7	0.8	3.8	1.3	1.1	2.3	0.4	1.00	1
14	1.6	0.8	3.6	1.3	1	2.2	0.3	1.00	1
15	1.6	0.8	3.6	1.3	1	2.3	0.4	1.00	1
16	1.6	0.8	3.5	1.2	1	2.2	0.3	1.00	1
17	1.6	0.8	3.6	1.3	1	2.2	0.4	1.00	1
18	1.6	0.8	3.6	1.3	1	2.2	0.4	1.00	1
19	1.6	0.8	3.5	1.2	1	2.2	0.3	1.00	1
20	1.5	0.7	3.4	1.2	1	2.2	0.3	1.00	1
21	1.6	0.8	3.5	1.2	1	2.1	0.3	1.00	1
22	1.6	0.8	3.5	1.2	1	2.2	0.3	1.00	1
23	1.5	0.7	3.4	1.2	1	2.2	0.3	1.00	1
24	1.6	0.8	3.5	1.2	1	2.2	0.3	1.00	1
25	1.6	0.8	3.5	1.2	1	2.2	0.3	1.00	1
26	1.5	0.7	3.4	1.2	1	2.2	0.3	1.00	1
27	1.5	0.7	3.4	1.2	1	2.2	0.3	1.00	1
28	1.5	0.7	3.4	1.2	1	2.2	0.4	1.00	1
29	1.5	0.7	3.4	1.2	1	2.2	0.4	1.00	1
30	1.6	0.8	3.5	1.2	1	2.2	0.3	1.00	1

Gambar 3. Dataset Setelah Proses *Celaning*

b) *Data Integration*

Dilakukan proses *Integration* untuk menggabungkan data yang diambil sehingga terampung total 91.500 baris data yang ditujukan pada [Gambar 4](#).

1	CO_MQ2(ppm)	Alcohol_MQ2(ppm)	Propane_MQ2(ppm)	CH4_MQ4(ppm)	Smoke_MQ4(ppm)	Temperature(C)	label
91481	3.6	1.3	1.1	1.7	0.2	1.00	0
91482	3.5	1.2	1	1.7	0.2	1.00	0
91483	3.6	1.3	1.1	1.8	0.2	1.00	1
91484	3.8	1.4	1.1	1.8	0.2	1.00	1
91485	3.6	1.3	1.1	1.8	0.2	1.00	1
91486	3.6	1.3	1.1	1.7	0.2	1.00	0
91487	3.6	1.3	1.1	1.8	0.2	1.00	1
91488	3.5	1.2	1	1.8	0.2	1.00	1
91489	4	1.4	1.1	1.8	0.2	1.00	1
91490	3.6	1.3	1.1	1.8	0.2	1.00	1
91491	3.8	1.4	1.1	1.8	0.2	1.00	1
91492	3.6	1.3	1.1	1.8	0.2	1.00	1
91493	3.8	1.4	1.1	1.8	0.2	1.00	1
91494	3.8	1.4	1.1	1.8	0.2	1.00	1
91495	3.6	1.3	1.1	1.8	0.2	1.00	1
91496	4	1.4	1.1	1.8	0.2	1.00	1
91497	3.8	1.4	1.1	1.7	0.2	1.00	1
91498	3.6	1.3	1.1	1.8	0.2	1.00	1
91499	3.8	1.4	1.1	1.8	0.2	1.00	1
91500	3.6	1.3	1.1	1.7	0.2	1.00	1
91501	3.8	1.4	1.1	1.7	0.2	1.00	1
91502							

Gambar 4. Dataset Setelah Proses *Integration*

c) Data Selection

Data selection yang digunakan yaitu proses sampling, dari jumlah data 91.500 menjadi 45.000 data yang ditunjukkan pada Gambar 5.

	H2_MQ2(ppm)	LPG_MQ2(ppm)	CO_MQ2(ppm)	Alcohol_MQ2(ppm)	Propane_MQ2(ppm)	CH4_MQ4(ppm)	Smoke_MQ4(ppm)	Temperature(C)	label
44973	1.5	0.7	3.3	1.2	1	1	0	1.00	1
44974	1.5	0.7	3.2	1.1	1	1	0	1.00	1
44975	1.5	0.7	3.3	1.2	1	1	0	1.00	1
44976	1.5	0.7	3.2	1.1	1	1	0	1.00	1
44977	1.5	0.7	3.2	1.1	1	1	0	1.00	1
44978	1.5	0.7	3.3	1.2	1	1	0	1.00	1
44979	1.5	0.7	3.3	1.2	1	1	0	1.00	1
44980	1.5	0.7	3.2	1.1	1	1	0	1.00	1
44981	1.5	0.7	3.2	1.1	1	1	0	1.00	1
44982	1.5	0.7	3.3	1.2	1	1	0	1.00	1
44983	1.5	0.7	3.3	1.2	1	1	0	1.00	1
44984	1.5	0.7	3.2	1.1	1	1	0	1.00	1
44985	1.4	0.7	3	1.1	0.9	1	0	1.00	1
44986	1.6	0.8	3.5	1.3	1	1	0	1.00	1
44987	1.5	0.7	3.3	1.2	1	1.1	0	1.00	1
44988	1.5	0.7	3.2	1.1	1	1	0	1.00	1
44989	1.5	0.7	3.3	1.2	1	1	0	1.00	1
44990	1.5	0.7	3.3	1.2	1	1	0	1.00	1
44991	1.5	0.7	3.3	1.2	1	1	0	1.00	1
44992	1.5	0.7	3.3	1.2	1	1	0	1.00	1
44993	1.5	0.7	3.3	1.2	1	1	0	1.00	1
44994	1.5	0.7	3.3	1.2	1	1	0	1.00	1
44995	1.5	0.7	3.3	1.2	1	1	0	1.00	1
44996	1.6	0.8	3.5	1.3	1	1	0	1.00	1
44997	1.5	0.7	3.3	1.2	1	1.1	0	1.00	1
44998	1.5	0.7	3.2	1.1	1	1.1	0	1.00	1
44999	1.5	0.7	3.3	1.2	1	1.1	0	1.00	1
45000	1.5	0.7	3.3	1.2	1	1.1	0	1.00	1
45001	1.5	0.7	3.3	1.2	1	1.1	0	1.00	1
45002									

Gambar 5. Dataset Setelah Proses Selection

d) Data Transformation

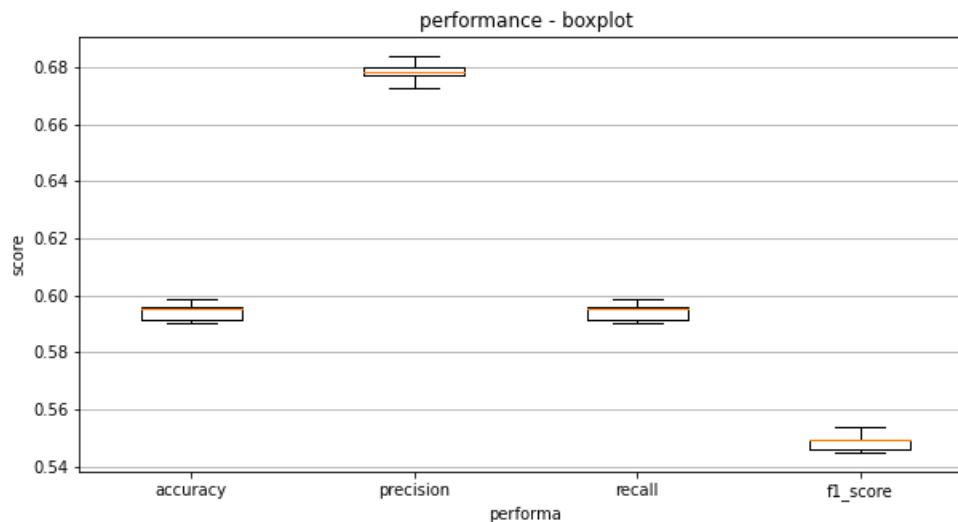
Data transformation yang digunakan yaitu proses *scaling* untuk membuat numerical data pada dataset dengan tetap mempresentasikan nilai yang sama.

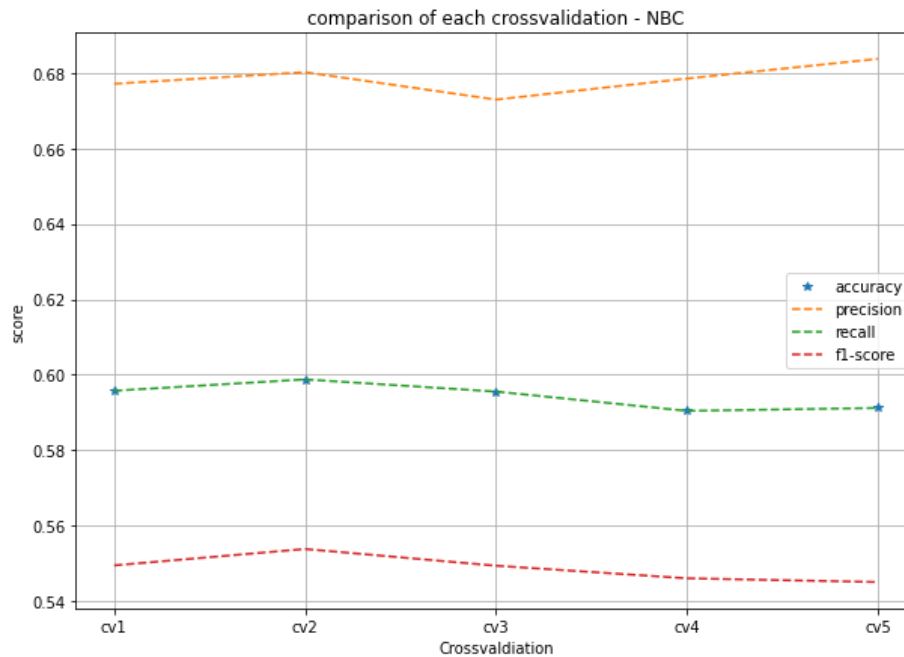
2) Implementasi Metode Naïve Bayes Classifier

Implementasi metode NBC dilakukan pada dataset yang telah melewati proses preprocessing data dengan jumlah total 45.000 data. dataset akan dibagi terdiri dari 80% data training dan 20% data testing dengan konsep Crossvalidation.

3) Hasil Pengujian Performa

Pengujian performa pada penelitian ini dikumpulkan dari setiap bagian *crossvalidation*. Gambar 6 menunjukkan hasil performa yang diperoleh yang divisualisasikan dalam boxplot dan linchart





Gambar 6. Visualisasi Performa Boxplot dan Linechart

Tabel 2 menunjukkan hasil performa akurasi terbaik disetiap *crossvalidation* dengan nilai K-fold = 5.

Tabel 2. Hasil uji performa metode NBC dan CV K-Fold =5.

No.	Simulasi Crossvalidation	Akurasi	Presisi	Recall	F-measure
1.	Cv I	59%	67%	59%	54%
2.	Cv II	59%	68%	59%	55%
3.	Cv III	59%	67%	59%	54%
4.	Cv IV	59%	67%	59%	54%
5.	Cv V	59%	68%	59%	54%

Berdasarkan pada Tabel 4.1 maka dapat disimpulkan bahwa nilai $k = 2$ hingga $k=900$ serta penerapan metode *crossvalidation* menjadi 10 bagian maka di peroleh performa terbaik pada *crossvalidation* VII, $K=13$ dimana akurasi tertinggi mencapai 82%, presisi 82%, recall 82% dan *f-measure* 82%. Gambar 4.1 meunjukkan grafik perbandingan performa setiap *crossvalidation* pada rasio 50:50.

B. Pembahasan

Beberapa teknik pada penelitian ini menggunakan bahasa pemrograman *python* dan *library scikitlearn* mulai dari tahap *load* dataset *cardiovascular*, visualisasi data, *split* data implementasi simulasi *crossvalidation*, implementasi metode NBC hingga menghitung nilai performa.

1) Pembahasan Source Code

Pembahasan *source code* program pada tahap ini yang terdiri dari *source code* menghubungkan *google driver* ke *google colab*, deklarasi *library*, *load* dataset, visualisasi data, menganalisis menggunakan metode NBC dan *crossvalidation*. Source Code *google driver* dan *google colab*

Source code	1	from google.colab import drive drive.mount('/content/drive')
Keterangan	1	Mengubungkan <i>google drive</i> dengan <i>google colab</i> sehingga data yang tersimpan di <i>google driver</i> dapat dibaca oleh <i>google colab</i>

Source Code Deklarasi Library

Source code	1	<code>import pandas as pd</code>
	2	<code>import numpy as np</code>
	3	<code>from sklearn.model_selection import train_test_split</code>
	4	<code>from sklearn.preprocessing import StandardScaler</code>
	5	<code>from sklearn.neighbors import KNeighborsClassifier</code>
	6	<code>from sklearn.metrics import confusion_matrix</code>
	7	<code>from sklearn.metrics import f1_score</code>
	8	<code>from sklearn.metrics import accuracy_score</code>
	9	<code>from sklearn.metrics import precision_score</code>
	10	<code>from sklearn.metrics import recall_score</code>
	11	<code>from sklearn.metrics import classification_report</code>
	12	<code>from matplotlib import pyplot as plt</code>
Keterangan	1-2	Deklarasi library pengolahan struktur dan analisis data
	3	Deklarasi library untuk membagi dataset menjadi data training dan testing
	4	
	5	Deklarasi library untuk menyamakan skala data pada atribut
	6-	Deklarasi library metode k- nearest
	11	Deklarasi library perhitungan performa
	12	Deklarasi library membuat visualisasi (grafik)

Source Code Load DataSet

Source code	1	<code>datatrain = pd.read_csv('/content/drive/MyDrive/dataset/cardio/dataset5050train_1.csv')</code>
	2	<code>datatest = pd.read_csv('/content/drive/MyDrive/dataset/cardio/dataset5050test_1.csv')</code>

	3	<code>print (len(d atatrain))</code>
	4	<code>print (len(datatest))</code>
Keterangan	1-2	Proses load (memasukkan) dataset cardiovascular sekaligus memasukkan data training dan testing
	2-4	Proses untuk mencetak informasi jumlah data yang ada dalam variabel

Source Code Source code split target and attribute

Source code	1-4	<pre>x_train = datatrain.iloc[:, 1:12] y_train = datatrain.iloc[:, 12] x_test = datatest.iloc[:, 1:12] y_test = datatest.iloc[:, 12]</pre>
Keterangan	1-4	Proses Memisahkan attribute dengan target

Source Code method and fitting

Source code	1	<code>gnb = GaussianNB()</code>
	2	<code>kfold=KFold(n_splits=5, shuffle=True, random_state=0)</code>
	3	<code>gnb.fit(X_train,y_train)</code>
	4	<code>y_pred=svc.predict(X_test)</code>
	5	
Keterangan	1	Menentukan model atau metode yang digunakan yaitu Naïve Bayesh Classifier
	2	Mencocokkan pasangan x-train(attribute) dan y-train(hasil)

Source Code testing data

Source code	1	<code>y_pred = classifier.predict(x_test)</code>
--------------------	---	--

Keterangan	1	Melakukan testing dari data hitung x_train dan y_test 3 , memprediksi hasil dari attribute yang ada dalam x_test dan outputnya disimpan dalam y_pred
-------------------	---	--

Source code print result

Source code	1	<code>cm = confusion_matrix(y_test, y_pred)</code>
	2	<code>print("k = ", x)</code> <code>print (cm)</code> <code>print("Akurasi :", accuracy_score(y_test, y_pred))</code> <code>print("presisi :", precision_score(y_test, y_pred,average=None))</code> <code>print("recall :", recall_score(y_test, y_pred,average=None))</code> <code>print("F-Score :", f1_score(y_test, y_pred,average=None))</code> <code>print("-----")</code>
Keterangan	1	Proses untuk menunjukkan confusion matrix yang mana jadi TP, TN, FP, FN
	2	Proses mencetak semua performa hasil perbandingan antara y_test dan y_pred

Source Code Implementasi Performa

Source Code	1	<code>cm = confusion_matrix(y_test, y_pred)</code>
	2	<code>print (cm)</code>
	3	<code>print("Akurasi :",</code> <code>accuracy_score(y_test, y_pred))</code>
	4	<code>print("Presisi :",</code> <code>precision_score(y_test, y_pred,</code>
	5	<code>average=None))</code>

	6	<pre>print("Recall :", recall_score(y_test, y_pred, average=None)) print("F-measure :", f1_score(y_test, y_pred, average=None))</pre>
Keterangan	1	Membuat matrix dari hasil hitung y_test dan y_pred
	2-6	Menampilkan hasil akurasi,presisi,recall dan f-measure

2) Pembahasan Kesimpulan Pengujian

Berdasarkan hasil performa yang diperoleh dari menggunakan Naive Bayes Classifier (NBC), dapat disimpulkan bahwa performa model secara keseluruhan cukup moderat. Nilai akurasi, presisi, recall, dan F1-score berada pada rentang 0,59 hingga 0,68, yang menunjukkan bahwa model mampu mengklasifikasikan beberapa titik data dengan benar, tetapi tidak semua.

Akurasi model, yang diukur dengan proporsi prediksi yang benar, sekitar 0,59 hingga 0,60, yang berarti model mampu mengklasifikasikan sekitar 60% dari titik data dengan benar. Presisi, yang merupakan proporsi true positive dari semua prediksi positif, sekitar 0,67 hingga 0,68, yang berarti model mampu mengidentifikasi dengan benar sekitar 67% hingga 68% dari kasus positif. Recall, yang merupakan proporsi true positive dari semua kasus positif yang sebenarnya, sekitar 0,59 hingga 0,60, yang berarti model mampu mengidentifikasi dengan benar sekitar 59% hingga 60% dari kasus positif yang sebenarnya.

F1-score, yang merupakan rata-rata harmonis dari presisi dan recall, sekitar 0,55, yang relatif rendah. Ini berarti bahwa model memiliki keseimbangan yang relatif rendah antara presisi dan recall, yang merupakan tanda bahwa model mungkin kurang konsisten dalam mengidentifikasi kasus positif. Secara keseluruhan, hasil menunjukkan bahwa performa model cukup moderat dan masih ada ruang untuk perbaikan. Hal penting untuk dipertimbangkan adalah menggunakan algoritma atau teknik lain untuk meningkatkan performa model.

IV. Kesimpulan

Dapat disimpulkan bahwa performa model Naive Bayes Classifier (NBC) dalam klasifikasi aroma tahu cukup moderat, dengan nilai akurasi, presisi, recall, dan F1-score berada pada rentang 0,59 hingga 0,68. Ini menunjukkan bahwa model mampu mengklasifikasikan beberapa titik data dengan benar, tetapi tidak semua..

Daftar Pustaka

- [1] BPS - Statistics Indonesia, "Indo_05_25." p. 1, 2020.
- [2] L. Vinet and A. Zhedanov, "A 'missing' family of classical orthogonal polynomials," *J. Phys. A Math. Theor.*, vol. 44, no. 8, pp. 1–8, 2011, doi: 10.1088/1751-8113/44/8/085201.
- [3] N. M. Saptarini, Y. Wardati, and U. Supriatna, "Deteksi Formalin Dalam Tahu di Pasar Tradisional Parwakarta," *J. Penelit. Sains dan Teknol.*, vol. 12, pp. 37–44, 2011.
- [4] W. Astuti, D. Lenono, and F. Faizah, "Identifikasi Tahu Berformalin dengan Electronic Nose Menggunakan Jaringan Syaraf Tiruan Backpropagation," *IJEIS (Indonesian J. Electron. Instrum. Syst.)*, vol. 6, no. 2, p. 211, 2016, doi: 10.22146/ijeis.15330.
- [5] D. D. Novita, A. B. Sesunan, M. Telaumbanua, S. Triyono, and T. W. Saputra, "Identifikasi Jenis

-
- Kopi Menggunakan Sensor E-Nose Dengan Metode Pembelajaran Jaringan Syaraf Tiruan Backpropagation,” *J. Ilm. Rekayasa Pertan. dan Biosist.*, vol. 9, no. 2, pp. 205–217, 2021, doi: 10.29303/jrpb.v9i2.241.
- [6] Rizky Ade Putranto, Triastiti Wuryandari, and Sudarno, “Perbandingan Analisis Klasifikasi Antara Decision Tree Dan Support Vector Machine Multiclass Untuk Penentuan Jurusan Pada Siswa Sma,” *J. Gaussian*, vol. 4, no. 4, pp. 1007–1016, 2015.
- [7] A. Fitria and H. Azis, “Analisis Kinerja Sistem Klasifikasi Skripsi menggunakan Metode Naïve Bayes Classifier,” *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 2, pp. 102–106, 2018.
- [8] R. Rahmadewi and R. Kurnia, “Klasifikasi Penyakit Paru Berdasarkan Citra Rontgen dengan Metoda Segmentasi Sobel,” *J. Nas. Tek. Elektro*, vol. 5, no. 1, p. 7, 2016, doi: 10.25077/jnte.v5n1.174.2016.
- [9] H. Azis, F. T. Admojo, and E. Susanti, “Analisis Perbandingan Performa Metode Klasifikasi pada Dataset Multiclass Citra Busur Panah,” *Techno.Com*, vol. 19, no. 3, 2020.
- [10] M. Yusa, E. Utami, and E. T. Luthfi, “Analisis Komparatif Evaluasi Performa Algoritma Klasifikasi pada Readmisi Pasien Diabetes,” *J. Buana Inform.*, vol. 7, no. 4, pp. 293–302, 2016, doi: 10.24002/jbi.v7i4.770.
- [11] A. Fitria and H. Azis, “Analisis Kinerja Sistem Klasifikasi Skripsi menggunakan Metode Naïve Bayes Classifier,” *Pros. Semin. Nas. Ilmu Komput. dan Teknol. Inf.*, vol. 3, pp. 102–106, 2018.
- [12] Nirmalasari, “Analisis Performa Metode Naïve Bayes Classifier Dan Cross Validation Pada Data Penyakit Cardiovascular,” *Indones. J. data Sci.*, 2020.
- [13] S. Sahar, “Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Clasiffier Pada Dataset Penyakit Jantung,” *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 79–86, 2020, doi: 10.33096/ijodas.v1i3.20.
- [14] P. A. Flach and M. Kull, “Precision-Recall-Gain curves: PR analysis done right,” *Adv. Neural Inf. Process. Syst.*, vol. 2015-Janua, pp. 838–846, 2015.
- [15] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, “The balanced accuracy and its posterior distribution,” *Proc. - Int. Conf. Pattern Recognit.*, pp. 3121–3124, 2010, doi: 10.1109/ICPR.2010.764.
- [16] N.-D. Nguyen, T. Do, T. D. Ngo, and D.-D. Le, “An Evaluation of Deep Learning Methods for Small Object Detection,” *J. Electr. Comput. Eng.*, vol. 2020, p. 3189691, 2020, doi: 10.1155/2020/3189691.