

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Kajian Pustaka**

Penelitian dan pemanfaatan *Naive Bayes* sebagai sebuah metode klasifikasi yang digunakan untuk mengklasifikasikan data berdasarkan probabilitas, banyak diterapkan dalam berbagai aplikasi. Diantaranya seperti pengenalan teks, klasifikasi dokumen, klasifikasi spam, dan diagnosa medis. Dalam diagnosa medis, *Naive Bayes* digunakan untuk menentukan jenis penyakit berdasar pada gejala yang dialami oleh pasien. Salah satu penerapan *Naive Bayes* yaitu melalui pendekatan sistem pakar untuk diagnosa suatu penyakit yang tidak terbatas untuk manusia saja, hewan dan tumbuhan juga. Dalam penelitian ini, sejumlah penelitian terdahulu digunakan sebagai sumber referensi yang memiliki keterkaitan dengan pokok bahasan, berikut penjelasannya.

Penelitian pertama dan kedua yang dilakukan oleh Prasetya [9] tentang klasifikasi penderita penyakit diabetes dan oleh Tempola [10] tentang penentuan status gunung berapi. Hasil penelitian [9] menunjukkan bahwa model *KNN* lebih unggul dalam hal akurasi pada data *train* sebesar 95% dan *Naive Bayes* sebesar 93%, namun untuk akurasi pada data *test*, model *Naive Bayes* lebih unggul yaitu sebesar 100% dan *KNN* sebesar 80%. *Confusion Matrix* digunakan untuk melakukan pengujian akurasi dari hasil yang didapatkan Menurut penelitian [10], hasil tingkat akurasi model *Naive Bayes* lebih tinggi daripada model *KNN*, yaitu dengan rata-rata sebesar 79.71% untuk *Naive Bayes* dan 63.68% untuk *KNN*.

Penelitian ketiga, dilakukan oleh Putro [11] tentang klasifikasi pelanggan dengan menerapkan metode *Naive Bayes*. Metode tersebut bertugas untuk mengklasifikasikan pelanggan ke kelas potensial dan non-potensial dengan menggunakan atribut jumlah pembelian, interval waktu, dan lokasi. Setelah melakukan pengujian sebanyak 25 kali, hasil dari sistem klasifikasi tersebut berupa 23 reaksi benar dan 2 reaksi salah. Nilai akurasi yang didapat yaitu 92%, nilai presisi 100%, dan recall 91%.

Penelitian keempat, dilakukan oleh Janah [12] tentang klasifikasi emosi pengguna pada ulasan aplikasi Traveloka. Emosi tersebut diklasifikasikan sebagai emosi senang, sedih, terkejut, dan marah. Pengambilan data ulasan dilakukan melalui aplikasi *Google Play* sebanyak 2662 data dengan rentang dari bulan November 2019 sampai Januari 2020. Menggunakan metode *Naive Bayes* (*Multinomial NB*, *Bernoulli NB*, dan *Gaussian NB*) untuk mendeteksi emosi pengguna yang terkandung dalam ulasan tersebut. Melalui perhitungan *Confusion Matrix* didapatkan bahwa sentimen emosi senang merupakan kelas yang paling banyak diklasifikasikan benar, dengan jumlah data sebanyak 359. Sentimen emosi marah ada di posisi kedua dengan jumlah data sebanyak 63, diikuti oleh sentimen emosi sedih dengan 22 data, dan sentimen emosi terkejut dengan 16 data. *Multinomial NB* mendapatkan hasil akurasi tertinggi yakni sebesar 86%, diikuti dengan *Bernoulli NB* sebesar 69%, dan *Gaussian NB* sebesar 44%.

Penelitian kelima, yang dilakukan oleh Widiyawati [13] memiliki tujuan membangun perangkat lunak berbasis *web* untuk diagnosa penyakit kucing disertai dengan solusi penanganan penyakit yang muncul pada proses klasifikasi dan juga menilai seberapa baik performa *Naive Bayes* melalui perhitungan tingkat akurasi. Dalam sistem pakar ini, metode *Naive Bayes Classifier* digunakan untuk proses pengambilan keputusan dengan data yang dikumpulkan dari dokter hewan dan berbagai literatur pendukung lainnya. Pada *dataset* terdapat 20 data penyakit dan 73 data gejala yang digunakan. Dari 10 pasien, penelitian ini menghasilkan akurasi nilai sebesar 85%, hasil ini didapatkan dari membandingkan hasil uji coba antara hipotesis dokter hewan dengan perhitungan matematis [13].

Penelitian keenam yaitu yang dilakukan oleh Handoko [6] bertujuan membangun *website* sistem pakar yang berguna untuk diagnosa penyakit pada masa kehamilan. Hal ini didasari pada angka kematian ibu hamil yang cukup tinggi, karena kurang mendapatkan informasi mengenai kesehatan obstetri dan ginekologi. Juga rasa malu yang muncul pada ibu hamil untuk berkonsultasi langsung ke dokter obstetri dan ginekologi yang sebagian besar merupakan kaum pria. Penggunaan *Naive Bayes* pada penelitian ini berguna dalam memproses masukan gejala yang dialami ibu hamil sehingga muncul diagnosa penyakit pada masa kehamilan.

Dengan mewawancarai dokter spesialis kandungan dan penyakit kandungan, didapatkan *dataset* dengan 5 data penyakit dan 25 data gejala. Hasil pengujian pada penelitian ini sebesar 77%, yang merupakan perbandingan antara diagnosa hasil dari sistem dengan diagnosa hasil dari dokter [6].

Penelitian ketujuh yang dilakukan oleh Syarifudin [14] bertujuan untuk mengembangkan aplikasi berbasis android yang berguna untuk diagnosa penyakit pada tanaman jagung melalui sistem pakar. Selain itu, sistem yang dikembangkan diharapkan dapat memberikan solusi untuk menangani penyakit yang menjangkit pada tanaman jagung menggantikan seorang pakar. Penggunaan *Naive Bayes* sebagai metode utama dalam penelitian ini. Sistem akan mengolah data gejala yang diberikan untuk menghasilkan diagnosa penyakit serta solusi penanggulangannya. Seorang pakar yang berasal dari Universitas Brawijaya Malang diwawancarai oleh peneliti sehingga diperoleh hasil berupa *dataset*. *Dataset* yang didapat berjumlah 4 data penyakit dan 19 data gejala. Pengujian dalam penelitian ketujuh ini, menyatakan hasil bahwa kebutuhan fungsional melalui uji *black box* hasilnya valid, akurasi sistem yang didapat sebesar 96%, dan menghasilkan predikat sangat baik pada *usability testing* [14].

Penelitian kedelapan yaitu yang dilakukan oleh Andriyanto [15], bertujuan membangun suatu model sistem pakar yang mampu melakukan diagnosa penyakit diabetes melitus. Menggunakan *Naive Bayes* untuk proses perhitungan probabilitas penyakitnya. Selain itu, tujuan lainnya dapat memberikan wawasan terkait penyakit diabetes melitus dan memberikan solusi penanganan terkait penyakit ini apabila muncul saat proses klasifikasi. *Dataset* didapatkan dari Puskesmas Poncokusumo Malang dengan 15 data gejala. Penelitian ini memperoleh hasil yang terbilang tinggi, yaitu mencapai akurasi 100%. Akurasi tersebut didapat melalui pengujian dengan 140 data latih dan 10 data uji. Tingkat akurasi sistem akan semakin tinggi tergantung jumlah data latih dan data uji [15].

Penelitian kesembilan dan kesepuluh yang dilakukan oleh Simanungkalit [16] dan Wahyuni [17], bertujuan merancang *website* sistem pakar yang mampu untuk menentukan diagnosa suatu penyakit pada ayam kampung, metode yang digunakan yakni *Certainty Factor*, serta memberikan langkah pengobatan apabila ayam

kampung tersebut terjangkit penyakit yang muncul melalui prediksi sistem. Penggunaan *Certainty Factor* yaitu untuk menghitung derajat kepastian akan fakta atau aturan guna menggambarkan apa yang menjadi keyakinan tentang masalah yang akan diselesaikan oleh seorang pakar [17]. Data yang digunakan oleh Simanungkalit [16] didapatkan dari proses wawancara dengan pakar dengan diperoleh *dataset* penyakit sejumlah 10 dan data gejala sejumlah 37, sedangkan Wahyuni [17] melakukan wawancara dengan peternak ayam kampung sebagai data primer serta dengan dokter hewan sebagai data pendukung sehingga diperoleh *dataset* yang masing-masing 5 data penyakit dan 20 data gejala. Masing-masing penelitian tersebut melakukan pengujian dengan studi kasus yang menghasilkan persentase sebesar 97.50% berdasar masukan gejala yaitu penyakit Batuk Menahun oleh peneliti Simanungkalit [16], sedangkan penelitian oleh Wahyuni [17] menghasilkan persentase sebesar 99% yaitu penyakit Flu Burung. Untuk lebih singkatnya, dapat dilihat pada perbandingan penelitian beberapa waktu lalu yang digunakan pada penelitian ini di Tabel 2.1.

Tabel 2.1 Penelitian Terdahulu

No.	Peneliti	Comparing	Contrasting	Criticize	Synthesize	Summarize
1.	Prasetya [9]	Penelitian ini melakukan klasifikasi penderita penyakit diabetes dengan <i>label</i> atau kelas <i>true</i> dan <i>false</i> .	Perbandingan algoritme <i>KNN</i> dan <i>Naive Bayes</i> .	Penambahan jumlah data atau variasi data <i>training</i> pada <i>dataset</i> agar dapat memperkuat hasil akurasi.	<i>Dataset</i> didapatkan dari seorang pakar bernama Ibu Saptarum Masalah Amd. Keb di Klinik Bidan Saptarum Masalah Jombang yang selanjutnya dilakukan tahap <i>preprocessing</i> , pembagian data menjadi 2 yaitu data <i>training</i> dan data <i>testing</i> . Tahap selanjutnya dengan mengimplementasikan algoritme <i>KNN</i> dan <i>Naive Bayes</i> dan tahap terakhir yaitu pengujian akurasi.	Hasil <i>KNN</i> lebih unggul dalam hal akurasi pada data <i>train</i> sebesar 95% dan <i>Naive Bayes</i> sebesar 93%, namun untuk akurasi pada data <i>test</i> , model <i>Naive Bayes</i> lebih unggul yaitu sebesar 100% dan <i>KNN</i> sebesar 80%.
2.	Tempola [10]	Penelitian ini melakukan klasifikasi untuk menentukan status gunung berapi dengan kelas (status) normal, waspada, dan siaga.	Perbandingan <i>KNN</i> dan <i>Naive Bayes</i> dengan K-Fold Cross Validation untuk validasi data.	Penulisan beberapa kata dan kalimat yang sama tidak konsisten.	<i>Dataset</i> bersifat publik yang didapatkan dari <i>website</i> Pusat Vulkanologi dan Mitigasi Bencana Geologi (PVMBG), kemudian dilakukan pembagian data menjadi data <i>training</i> dan data <i>testing</i> . Tahap selanjutnya yaitu mengimplementasikan algoritme <i>KNN</i> dan <i>Naive Bayes</i> , yang kemudian diakhiri dengan uji validasi data.	Hasil tingkat akurasi model <i>Naive Bayes</i> lebih tinggi daripada model <i>KNN</i> , yaitu dengan rata-rata sebesar 79.71% untuk <i>Naive Bayes</i> dan 63.68% untuk <i>KNN</i> .
3.	Putro [11]	Penelitian ini melakukan klasifikasi pelanggan ke kelas potensial dan non-potensial yang diimplementasikan dalam aplikasi menggunakan bahasa pemrograman <i>PHP</i> .	Penggunaan <i>Naive Bayes</i> sebagai algoritme klasifikasi.	Sistem yang dibuat pada penelitian ini terbilang cukup sederhana karena lebih fokus ke proses perhitungan dan	<i>Dataset</i> didapatkan dari proses tanya jawab kepada pemilik UD. Samodro yang kemudian data dibagi menjadi data <i>training</i> dan data <i>testing</i> . Setelah itu mengimplementasikan algoritme <i>Naive Bayes</i> dan diakhiri dengan pengujian akurasi sistem.	Sistem klasifikasi tersebut menghasilkan berupa 23 reaksi benar dan 2 reaksi salah. Nilai akurasi yang didapat yaitu 92%, nilai

No.	Peneliti	Comparing	Contrasting	Criticize	Synthesize	Summarize
				hasil perhitungan. Penulisan kata yang sama juga tidak konsisten.		presisi 100%, dan recall 91%.
4.	Janah [12]	Penelitian ini melakukan klasifikasi emosi pengguna pada ulasan aplikasi Traveloka dengan kelas emosi berupa senang, sedih, terkejut, dan marah.	Perbandingan model algoritme <i>Naive Bayes</i> untuk memperoleh nilai akurasi tertinggi dan sebaran <i>F1 score</i> yang rata.	Untuk kasus klasifikasi <i>text mining</i> , akurasi model <i>Gaussian NB</i> tergolong cukup rendah.	<i>Dataset</i> diambil dari data ulasan aplikasi Traveloka di <i>website Google Play</i> . Pelabelan dilakukan secara manual menghasilkan <i>label</i> emosi senang, sedih, terkejut, dan marah. Data dilakukan <i>preprocessing</i> , pembagian data <i>training</i> dan data <i>testing</i> , dan pembobotan TF-IDF. Tahap selanjutnya melatih data menggunakan algoritme <i>Naive Bayes</i> dengan 3 pendekatan yaitu <i>Multinomial NB</i> , <i>Bernoulli NB</i> , dan <i>Gaussian NB</i> , dan melakukan evaluasi model.	<i>Multinomial NB</i> mendapatkan hasil akurasi tertinggi yakni sebesar 86%, diikuti dengan <i>Bernoulli NB</i> sebesar 69%, dan <i>Gaussian NB</i> sebesar 44%.
5.	Widyawati [13]	Penelitian ini melakukan klasifikasi penyakit pada kucing dengan kelas penyakit sebanyak 20 kelas dan diimplementasikan ke dalam sistem pakar. Hasil diagnosa merupakan klasifikasi penyakit yang didapatkan melalui proses perhitungan dan diambil nilai tertinggi. Sistem berbasis <i>web</i> yang dibangun	Penggunaan <i>Naive Bayes Classifier</i> yang diimplementasikan dalam sistem pakar berbasis <i>web</i> dan menggunakan bahasa pemrograman <i>PHP</i> .	Sumber <i>dataset</i> tidak dijabarkan secara jelas, hanya berasal dari pakar, sehingga membuat asumsi bahwa data tersebut kurang kuat.	<i>Dataset</i> didapatkan dari pakar dengan total 20 kelas penyakit. Kemudian dilakukan proses perhitungan probabilitas menggunakan algoritme <i>Naive Bayes Classifier</i> dan dilakukan pengujian sistem.	Dari 10 pasien, penelitian ini menghasilkan akurasi nilai sebesar 85%, hasil ini didapatkan dari membandingkan hasil uji coba antara hipotesis dokter hewan dengan perhitungan matematis.

No.	Peneliti	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
		dengan bahasa pemrograman <i>PHP</i> .				
6.	Handoko [6]	Penelitian ini melakukan klasifikasi penyakit pada masa kehamilan dengan sebanyak 5 kelas penyakit dan diimplementasikan ke dalam sistem pakar berbasis <i>web</i> menggunakan bahasa pemrograman <i>PHP</i> .	Penggunaan <i>Naive Bayes</i> yang diimplementasikan dalam sistem pakar berbasis <i>web</i> dan menggunakan bahasa pemrograman <i>PHP</i> .	Penulisan kata yang sama kurang konsisten.	<i>Dataset</i> dikumpulkan dari proses wawancara dengan Dokter spesialis kebidanan dan penyakit kandungan Dr. Fonda Octarianingsih, Sp. OG. <i>Dataset</i> dibagi menjadi data <i>training</i> dan data <i>testing</i> , kemudian mengimplementasikan algoritme <i>Naive Bayes</i> . Tahap selanjutnya melakukan pengujian sistem dan validasi dengan pakar.	Pengujian dari penelitian ini menghasilkan sebesar 77%, yang merupakan perbandingan antara hasil diagnosa dari sistem dengan diagnosa dari dokter.
7.	Syarifudin [14]	Penelitian ini melakukan klasifikasi penyakit pada tanaman jagung dengan kelas penyakit sebanyak 4 kelas dan diimplementasikan ke dalam sistem pakar berbasis <i>android</i> .	Penggunaan <i>Naive Bayes</i> pada sistem pakar berbasis <i>android</i> .	Contoh perhitungan pada proses algoritme <i>Naive Bayes</i> tidak dicantumkan.	<i>Dataset</i> didapatkan melalui wawancara dengan seorang pakar yang berasal dari Universitas Brawijaya Malang bernama Luqman Qurata A, S.P., M.Si., Ph.D. Kemudian dilakukan proses perhitungan probabilitas menggunakan algoritme <i>Naive Bayes</i> . Tahap selanjutnya melakukan 3 jenis pengujian yaitu <i>black box</i> , akurasi, dan <i>usability testing</i> .	Pengujian dari penelitian tersebut, menunjukkan hasil bahwa semua kebutuhan fungsional melalui uji <i>black box</i> hasilnya valid, akurasi sistem yang didapat sebesar 96%, dan menghasilkan predikat sangat baik pada <i>usability testing</i> .
8.	Andriyanto [15]	Penelitian ini melakukan klasifikasi untuk menentukan penyakit Diabetes Mellitus dengan kelas penyakit sebanyak 2 kelas dan diimplementasikan	Penggunaan <i>Naive Bayes</i> pada model sistem pakar berbasis <i>web</i> .	<i>Dataset</i> yang dikumpulkan kurang jelas berasal dari pakar atau tidak dan contoh	<i>Dataset</i> dikumpulkan dari berbagai sumber, kemudian dilakukan proses perhitungan probabilitas menggunakan algoritme <i>Naive Bayes</i> . Tahap	Penelitian ini memperoleh hasil yang terbilang tinggi, yaitu mencapai akurasi 100%. Akurasi tersebut

No.	Peneliti	<i>Comparing</i>	<i>Contrasting</i>	<i>Criticize</i>	<i>Synthesize</i>	<i>Summarize</i>
		ke dalam sistem pakar berbasis <i>web</i> .		perhitungan dari algoritme <i>Naive Bayes</i> tidak dicantumkan.	selanjutnya dilakukan pengujian akurasi dan evaluasi.	didapat melalui pengujian dengan 140 data <i>training</i> dan 10 data <i>testing</i> .
9.	Simanungkalit [16]	Penelitian ini melakukan penentuan diagnosa penyakit pada ayam kampung berdasar pada perhitungan derajat kepastian akan fakta dan diimplementasikan ke dalam sistem pakar.	Penggunaan <i>Certainty Factor</i> dalam sistem pakar dengan basis <i>web</i> .	<i>Dataset</i> yang diperoleh dari wawancara dengan pakar ini kurang jelas siapa yang diwawancarai.	<i>Dataset</i> diperoleh melalui proses wawancara dengan pakar. Kemudian dilakukan perhitungan menggunakan <i>Certainty Factor</i> melalui bobot keyakinan dari gejala yang diberikan oleh pakar. Hasil diagnosa merupakan nilai persentase kelas penyakit yang terbesar.	Hasil penelitian ini berdasar studi kasus yang dilakukan dengan menghasilkan persentase sebesar 97.50% berdasar masukan gejala yaitu penyakit Batuk Menahun.
10.	Wahyuni [17]	Penelitian ini melakukan diagnosa penyakit pada ayam kampung berdasarkan perhitungan derajat kepastian akan suatu fakta dan diimplementasikan ke dalam sistem pakar. Perancangan sistem pakar tersebut menggunakan metode <i>Waterfall</i> dengan basis <i>web</i> .	Penggunaan <i>Certainty Factor</i> beserta metode <i>Waterfall</i> pada sistem pakar dengan basis <i>web</i>	Penulisan beberapa kata yang sama tidak konsisten.	<i>Dataset</i> diperoleh melalui proses wawancara dengan pakar. Setelah itu dilakukan perhitungan menggunakan <i>Certainty Factor</i> melalui bobot keyakinan dari gejala yang diberikan oleh pakar. Hasil diagnosa merupakan nilai persentase kelas penyakit yang tertinggi.	Hasil penelitian ini dihasilkan melalui studi kasus dengan masukan gejala yang menghasilkan persentase sebesar 99% yaitu penyakit Flu Burung.

Berdasarkan Tabel 2.1, pada penelitian ini menggunakan penelitian terdahulu sebagai bahan referensi. Penelitian terdahulu yang digunakan berkaitan dengan klasifikasi dan sistem pakar.



## 2.2 Dasar Teori

### 2.2.1 Klasifikasi

Penerapan klasifikasi pertama kali dijumpai pada bidang botani, dimana spesies tertentu diklasifikasikan berdasarkan karakteristik fisik. Hal ini dilakukan oleh Carolus Linnaeus yang kemudian dikenal sebagai bapak dari klasifikasi [7]. Klasifikasi adalah proses penilaian dan evaluasi objek data untuk menempatkannya ke dalam suatu kelas yang sesuai dari sekumpulan kelas yang tersedia [11]. Adapun klasifikasi merupakan kegiatan mengelompokkan objek ke dalam data kelas tertentu menggunakan data latih untuk membangun suatu model. Menggunakan model tersebut untuk mengklasifikasikan label pada kelas data paling baru yang masih belum diketahui [9].

Komponen utama dari proses klasifikasi yaitu: Kelas, sebagai label hasil klasifikasi dan merupakan variabel tidak bebas, contohnya kelas loyalitas pelanggan, kelas penyakit, dan lain-lain; Prediktor, sebagai variabel bebas dalam model yang didasarkan pada atribut karakteristik data yang diklasifikasi, seperti status pendidikan, gejala, dan lain-lain; Data *Training*, sekumpulan data lengkap berisi kelas dan prediktor yang digunakan untuk melatih model dalam mengelompokkan data ke kelas yang sesuai; dan Data *Testing*, data baru untuk menguji akurasi model yang dibuat [7]. Dalam proses klasifikasi, terdapat dua tugas utama yang dilakukan, yakni membangun model yang berguna sebagai *prototype*, kemudian menyimpannya sebagai suatu memori lalu model tersebut digunakan untuk pengenalan/mengklasifikasikan objek data lain agar dapat diketahui di kelas apa objek data tersebut pada model yang sudah disimpan sebelumnya [11].

Berbagai metode atau model telah dikembangkan oleh para peneliti untuk mengatasi kasus klasifikasi. Metode-metode itu diantaranya, *Decision Tree*, Klasifikasi Bayes/*Naive Bayes*, Jaringan Syaraf Tiruan, Analisis Statistik, Algoritma Genetik, *Rough Set*, Klasifikasi *K-Nearest*

*Neighbour (KNN)*, *Memory Based Reasoning*, dan *Support Vector Machine* [7]. Dalam penelitian ini menggunakan *Naive Bayes* sebagai algoritme klasifikasi.

### 2.2.2 *Naive Bayes*

Klasifikasi dengan menggunakan *Naive Bayes* mengaplikasikan Teorema Bayes. Teorema Bayes adalah teori tentang perhitungan probabilitas yang bergantung pada kondisi (*posterior probability*), untuk memprediksi kemungkinan di masa depan yang didasari oleh pengalaman pada masa lampau [7][6]. *Naive Bayes* adalah suatu algoritme *Machine Learning* untuk mengklasifikasikan data berdasarkan probabilitas sederhana dengan asumsi bahwa setiap kelas tidak saling terkait (*independen*) dari kondisi lainnya [8]. Adapun *Naive Bayes* merupakan salah satu teknik penalaran probabilistik yang menghitung serangkaian probabilitas dengan melakukan penjumlahan frekuensi dan kombinasi nilai dari kumpulan data (*dataset*) yang disediakan [18][10]. Istilah *Naive* muncul karena itu adalah penyederhanaan dimana antara satu variabel tidak mempengaruhi variabel lain [7].

Dalam pengklasifikasian data, *Naive Bayes* memiliki kelebihan tertentu. Salah satunya yaitu selain mengklasifikasikan data, metode tersebut menunjukkan seberapa besar peluang suatu data masuk ke dalam kelas tertentu. Dalam proses pembelajaran model, *Naive Bayes* memanfaatkan bukti (*evidence*) yang ada yaitu data *training*. Untuk memastikan akurasi proses pembelajaran, data *training* yang tepat sangat penting karena *Naive Bayes* bergantung pada pengetahuan awal [7].

*Naive Bayes* dalam penelitian ini menggunakan salah satu pendekatan, yaitu *GaussianNB* yang memanfaatkan nilai probabilitas dan statistik. Fungsi *GaussianNB* merupakan alat yang dibutuhkan dalam pembuatan model *Naive Bayes*, yang digunakan untuk melakukan pengujian pada *dataset* baik dalam proses *training* ataupun *testing* [9]. Berikut merupakan garis besar persamaan dari *Naive Bayes* [19]:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2.1)$$

Keterangan:

$X$ : Data yang tidak diketahui kelasnya.

$H$ : Hipotesis data  $X$  adalah suatu kelas tertentu.

$P(H|X)$ : Probabilitas hipotesis  $H$  berdasarkan kondisi  $X$  (*posterior probability*).

$P(X|H)$ : Probabilitas  $X$  berdasarkan kondisi hipotesis  $H$ .

$P(H)$ : Probabilitas hipotesis  $H$  (prior probability).

$P(X)$ : Probabilitas  $X$

$P(X|H)$  adalah probabilitas  $X$  dengan kondisi  $H$ , yang sering disebut sebagai *likelihood*. Hasilnya disesuaikan dengan perbandingan  $P(H)$  dan  $P(X)$  [7].

Dalam kasus perhitungan sistem pakar, persamaan (2.1) dapat dituliskan sebagai berikut [20]:

$$P = \frac{\text{Jumlah kemungkinan penyakit yang muncul}}{\text{Jumlah semua penyakit}} \quad (2.2)$$

$$G = \frac{\text{Jumlah kemungkinan}}{\text{Jumlah kemungkinan penyakit akibat gejala}} \quad (2.3)$$

$$P(p'|g') = \frac{[P(g'|p') \cdot P(p')]}{[P(g'|p') \cdot P(p')] + [P(g'|p'') \cdot P(p'')]} \quad (2.4)$$

Keterangan:

$P$ : Nilai probabilitas kelas penyakit

$G$ : Nilai probabilitas gejala

Dalam penelitian ini, maka hasil klasifikasi dengan nilai probabilitas tertinggi dari tiap-tiap kelas penyakit dijadikan sebagai hasil diagnosa penyakit pada ayam kampung.

### 2.2.3 Confusion Matrix

*Confusion Matrix* adalah tabel yang menunjukkan hasil dari perbandingan nilai prediksi dan nilai aktual yang berbeda-beda. Saat digunakan untuk mengukur performa, ada empat istilah yang digunakan, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. *TN* menunjukkan jumlah data negatif yang terklasifikasi dengan benar, sementara *FP* adalah data negatif yang terklasifikasi sebagai positif. *TP* adalah jumlah data positif yang terklasifikasi dengan benar, sementara *FN* adalah data yang sebenarnya positif namun terklasifikasi sebagai negatif [21]. Penggambarannya sebagai berikut:

		Predicted	
		Positive	Negative
Actual	Positive	TP (True Positive)	FP (False Positive)
	Negative	FN (False Negative)	TN (True Negative)

Gambar 2.1 Confusion Matrix pada 2 Kelas

Komponen di atas dibutuhkan untuk memperoleh nilai akurasi, presisi, dan *recall* yang dapat menandakan kinerja model itu baik atau tidak. Perhitungannya sebagai berikut:

a. Akurasi

Menunjukkan seberapa tepat model dalam mengklasifikasikan data dengan akurat.

$$\text{Akurasi} = \frac{TP}{\text{Total Dataset}} \quad (2.5)$$

b. Presisi

Menggambarkan rasio hasil prediksi dari model dibandingkan dengan seluruh hasil prediksi positif.

$$Presisi = \frac{TP}{FP + TP} \quad (2.6)$$

c. Recall

Menggambarkan rasio hasil prediksi dari model dibandingkan dengan seluruh data benar positif.

$$Recall = \frac{TP}{FN + TP} \quad (2.6)$$

#### 2.2.4 Penyakit Ayam Kampung

Ayam Kampung (*Gallus Domesticus*) yang dikenal sebagai ayam buras atau ayam bukan ras, adalah jenis ayam yang umum ditanakkan ataupun dipelihara oleh masyarakat di Indonesia [17]. Ada yang ditanakkan untuk kebutuhan pasar, dan ada juga yang memeliharanya untuk kebutuhan pribadi. Jenis ayam ini lebih sering hidup di lingkungan sekitar kandang, baru saat menjelang malam, ayam tersebut masuk ke dalam kandang. Ayam kampung biasanya ditanakkan untuk diambil daging, telur, dan bulunya. Namun, perlu diperhatikan juga tentang pemeliharanya, terkhusus apabila terserang penyakit. Memelihara ayam memiliki risiko penyakit yang harus dihadapi, dapat dikatakan penyakit tersebut berlaku juga untuk jenis ayam kampung [3]. Jika ayam terserang penyakit dan mati, akan berdampak berupa kerugian. Berikut penyakit yang teridentifikasi pada penelitian ini:

a. Flu Burung (*Avian Influenza*)

Flu Burung (*Avian Influenza*) adalah salah satu penyakit yang menyerang unggas, disebabkan oleh infeksi virus *influenza* bertipe A dari famili *Orthomyxoviridae*. Virus ini dapat menyebabkan gangguan pernapasan mulai dari ringan (*lowly pathogenic*) sampai akut (*highly pathogenic*). Dimana tingkat kematian sangat tinggi

atau hampir 90% di beberapa peternakan ayam dan dapat mengakibatkan kerugian material bagi peternak. Penyakit flu burung mempunyai sifat *zoonosis* atau dapat menular pada manusia dan penularannya bisa terjadi apabila virus bermutasi [5].

b. Tetelo atau *Newcastle Disease (ND)*

Tetelo atau *Newcastle Disease (ND)* adalah salah satu penyakit menular sangat merugikan bagi peternak, karena infeksi yang ditimbulkan dapat menyebabkan kematian ayam hingga dapat mencapai 50% pada ayam kampung. Penyakit ini timbul disebabkan oleh *strain virulen* dari *Avian Paramyxovirus tipe 1 (APMV-1)* dari genus *Avulavirus*, subfamili *Paramyxovirinae*, famili *Paramyxoviridae*. *Paramyxovirus* yang diisolasi dari spesies unggas dianalisis secara serologis dan filogenetik menjadi sepuluh subtipe, antara lain *APMV-1* hingga *APMV-10*. Di Indonesia, penyakit ini masih mewabah karena kasusnya tercatat sepanjang tahun, terjadi karena beberapa faktor seperti kualitas vaksin *ND* yang buruk, tindakan terhadap vaksin yang tidak memenuhi standar, dan kesalahan vaksinator [22].

c. Berak Kapur (*Pullorum*)

Penyakit *Pullorum* merupakan penyakit bakteri septikemia (*Septicaemic Bacterial Diseases*) yang umum ditemukan pada ayam atau kalkun dan penyebabnya yaitu bakteri *Salmonella Pullorum*. Ditemukan oleh Rettger pertama kali pada tahun 1899 dan dikenal di Australia pada tahun 1920 sebagai *Bacillary White Diarrhea* sesuai dengan gejala klinis penyakitnya yaitu diare berwarna putih. Penyebarannya sangat luas, hampir di seluruh dunia. *Pullorum* masih menjadi masalah utama pada industri peternakan dan termasuk ke dalam *Food Borne Diseases* yang memiliki sifat *zoonosis* [23][24].

d. Gumboro

Penyakit Gumboro merupakan penyakit klasik yang menyerang pada ayam dan menyebabkan kegagalan pada sistem imun ayam (*immunosuppressive*). Penyakit Gumboro menular dengan cepat dan dapat ditularkan dari ayam, manusia, atau peralatan yang sudah terinfeksi oleh virus Gumboro. Gumboro sulit diatasi jika terlambat untuk mengetahui kondisi ayam, sehingga dapat menyebabkan kerugian apabila ayam tersebut mati [25].

e. *Snot (Infeksius Coryza)*

*Snot (Infectious Coryza)* merupakan penyakit pernapasan yang menyerang ayam, disebabkan oleh *Avibacterium Paragallinarum (Av. Paragallinarum)*, yang dahulu dikenal dengan nama *Haemophilus Paragallinarum*. Penyakit ini memiliki sifat menular, yang mempengaruhi saluran pernapasan bagian atas dan berkisar dari akut sampai kronis. Secara umum, *Snot* merupakan penyakit dengan kasus ditemukannya tinggi, tetapi tingkat kematiannya rendah. Penyakit ini berdampak merugikan bagi industri peternakan [26].

### 2.2.5 Sistem Pakar

Sistem pakar merupakan salah satu cabang dari *Artificial Intelligence (AI)* atau Kecerdasan Buatan, pengembangannya dimulai pada tahun 1960, sehingga sistem ini bisa dikatakan sudah cukup tua [27]. Sistem pakar adalah program komputer yang berisi pengetahuan dari pakar manusia dengan tujuan untuk menuntaskan permasalahan yang membutuhkan kecerdasan pakar [6]. Dengan sistem pakar yang mengimplementasikan pengetahuan pakar/ahli pada suatu bidang tertentu ini, diharapkan mampu untuk membantu dalam mengambil suatu keputusan [6].

Di dalam sistem pakar, terdapat istilah yang bernama mesin inferensi. Mesin inferensi memegang peran sebagai inti dari sistem pakar. Dengan memandu proses berpikir mengenai situasi tertentu yang

didasarkan pada basis pengetahuan, atau dalam penelitian ini merupakan *dataset* [28]. Pada penelitian ini, *GaussianNB* memiliki peran sebagai mesin inferensi.