

BAB III

ANALISIS DAN PERANCANGAN

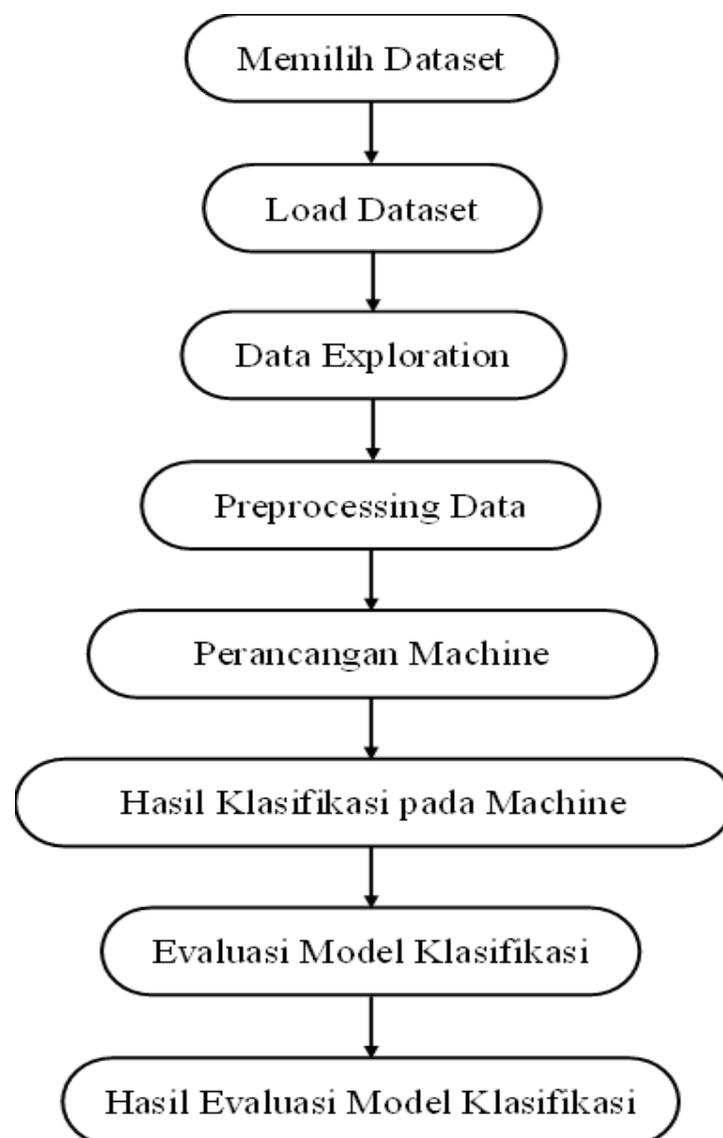
3.1. Kerangka Penelitian

Kerangka penelitian pada machine learning merangkum pendekatan dan langkah-langkah yang akan diambil dalam pengembangan model atau sistem berbasis machine learning. Pada awal penelitian, identifikasi permasalahan dan tujuan penelitian menjadi langkah awal dalam membentuk kerangka kerja. Penelitian ini kemudian melibatkan pengumpulan data yang relevan dengan permasalahan yang akan dipecahkan, yang bisa melibatkan dataset historis atau pengumpulan data baru. Setelah data terkumpul, tahap pra-pemrosesan data menjadi krusial untuk membersihkan, mentransformasi, dan mempersiapkan data agar sesuai dengan kebutuhan model machine learning yang akan diterapkan. Pemilihan model merupakan langkah berikutnya, di mana peneliti memilih algoritma yang paling sesuai dengan karakteristik data dan tujuan prediktifnya.

Pada tahap pelatihan model, data yang telah diproses digunakan untuk melatih model machine learning, yang kemudian diuji dan dievaluasi menggunakan data yang tidak terlibat dalam pelatihan. Evaluasi model melibatkan metrik kinerja seperti akurasi, presisi, recall, dan F1 score. Jika hasil evaluasi memenuhi kriteria keberhasilan, model tersebut dapat diimplementasikan untuk memecahkan permasalahan yang diidentifikasi pada awal penelitian.

Selain itu, kerangka penelitian pada machine learning mencakup langkah-langkah untuk validasi dan pengoptimalan model, yang melibatkan penyesuaian parameter dan teknik seperti validasi silang untuk meningkatkan generalisasi model

terhadap data yang belum pernah dilihat sebelumnya. Terakhir, penelitian ini juga mencakup dokumentasi hasil, interpretasi, dan komunikasi temuan kepada pemangku kepentingan dan masyarakat ilmiah. Keseluruhan, kerangka penelitian pada machine learning memberikan panduan sistematis untuk merancang, melatih, dan mengevaluasi model dengan tujuan mencapai solusi yang efektif dan efisien untuk permasalahan yang dihadapi.



Gambar 3. 1. Kerangka Penelitian

Adapun penjelasan dari tahapan ataupun kerangka penelitian pada machine learning yaitu sebagai berikut:

1. Memilih Dataset

Memilih dataset adalah tahap kritis dalam pengembangan model machine learning. Kesuksesan model secara langsung tergantung pada kualitas, relevansi, dan representativitas dataset yang digunakan. Saat memilih dataset, penting untuk mempertimbangkan tujuan analisis dan karakteristik masalah yang hendak dipecahkan. Pertama, identifikasi variabel-variabel yang memiliki dampak signifikan terhadap output yang diinginkan. Selanjutnya, pastikan dataset mencakup variasi yang cukup untuk memastikan generalisasi model. Keseragaman dan kebersihan dataset juga penting, termasuk penanganan nilai yang hilang dan deteksi outlier. Selain itu, dataset harus mencerminkan kondisi nyata dan konteks di mana model akan diterapkan. Pemahaman mendalam terhadap sumber data, proses pengumpulan, dan representasi dari setiap fitur akan membantu memastikan dataset yang dipilih dapat memberikan wawasan yang relevan dan dapat diandalkan dalam proses pembelajaran mesin.

2. Load Dataset

Proses memuat dataset merupakan langkah awal yang krusial dalam pengembangan model machine learning. Memilih dan memuat dataset yang tepat adalah fondasi penting untuk keberhasilan model. Saat memuat dataset, pertama-tama, identifikasi format file dan jenis dataset yang akan digunakan, apakah itu dalam bentuk CSV, Excel, atau format lainnya. Setelah itu, gunakan alat atau fungsi khusus, baik itu dalam bahasa pemrograman seperti Python dengan library seperti

Pandas, atau melalui platform seperti Orange, untuk membaca dan menyajikan dataset dalam format yang dapat dimengerti oleh algoritma machine learning. Pastikan untuk melakukan pemeriksaan awal terhadap dataset, termasuk melihat beberapa baris pertama data, mengecek nilai yang hilang, dan memahami struktur kolom. Proses ini membantu untuk memastikan bahwa dataset siap digunakan dan sesuai dengan tujuan analisis yang diinginkan. Pada dasarnya, langkah memuat dataset adalah fondasi pertama yang mendukung pembangunan model machine learning yang solid dan efektif.

3. Data Exprolation

Data exploration adalah tahap kritis dalam proses machine learning yang bertujuan untuk memahami karakteristik dan struktur dari dataset yang digunakan. Pada tahap ini, analisis awal dilakukan untuk mengidentifikasi pola, tren, serta sifat statistik dari data. Ini melibatkan visualisasi data menggunakan berbagai grafik dan plot, seperti histogram, scatter plot, dan box plot, untuk mendapatkan wawasan awal tentang distribusi dan hubungan antar variabel. Selain itu, statistik deskriptif, seperti mean, median, dan deviasi standar, juga sering digunakan untuk memberikan pemahaman yang lebih mendalam tentang properti dataset. Data exploration membantu dalam mengidentifikasi outliers, mengelompokkan variabel berdasarkan relevansi, dan membuat keputusan strategis tentang langkah-langkah preprocessing yang dibutuhkan sebelum membangun model. Keseluruhan, eksplorasi data memberikan landasan yang kuat bagi peneliti atau praktisi machine learning untuk memahami konteks masalah dan membuat keputusan yang tepat selama proses pengembangan model.

4. Preprocessing Data

Preprocessing data adalah tahapan esensial dalam pengembangan model machine learning yang bertujuan untuk meningkatkan kualitas dan kesesuaian dataset sebelum proses pelatihan model. Tahap ini melibatkan sejumlah teknik dan operasi untuk membersihkan, merapikan, dan mempersiapkan data agar sesuai dengan kebutuhan algoritma machine learning. Langkah-langkah preprocessing meliputi penanganan nilai yang hilang, normalisasi data untuk mengatasi perbedaan skala, serta reduksi dimensi untuk mengurangi kompleksitas dan mempercepat proses analisis. Selain itu, encoding kategori, penanganan outlier, dan transformasi variabel dapat menjadi bagian dari preprocessing data. Tujuan akhir dari preprocessing adalah untuk memastikan bahwa dataset memiliki kualitas yang baik, tidak mengandung noise yang signifikan, dan dapat mendukung pembentukan model yang akurat dan robust. Dengan melakukan tahap preprocessing data secara cermat, pengguna dapat memaksimalkan performa model machine learning dan meningkatkan kemampuannya dalam menghasilkan prediksi atau klasifikasi yang relevan dan bermakna.

5. Perancangan Machine Learning

Perancangan machine learning merupakan langkah kunci dalam pengembangan model klasifikasi dengan metode K-Nearest Neighbors (KNN) dan Naive Bayes. Perancangan ini melibatkan serangkaian keputusan strategis untuk memastikan model yang dihasilkan dapat memahami serta menggeneralisasi pola-pola yang ada dalam data dengan efektif. Pertama-tama, pemilihan fitur-fitur yang paling relevan dan signifikan bagi tujuan klasifikasi menjadi fokus utama

perancangan. Selanjutnya, penentuan nilai parameter k pada KNN dan asumsi independensi fitur pada Naive Bayes menjadi langkah berikutnya yang kritis. Desain model KNN melibatkan pemilihan metrik jarak yang sesuai dan evaluasi dampaknya terhadap kinerja model. Sementara itu, perancangan model Naive Bayes mencakup pemahaman mendalam terhadap karakteristik data dan distribusi fitur-fitur yang saling bergantung. Pemilihan dataset pelatihan dan pengujian juga menjadi pertimbangan esensial untuk menjamin evaluasi yang akurat terhadap kinerja model. Dengan memperhatikan aspek-aspek ini dalam perancangan machine learning, diharapkan model klasifikasi yang dihasilkan mampu memberikan prediksi yang handal dan sesuai dengan konteks data yang dihadapi.

6. Hasil Klasifikasi pada Machine Learning

Hasil klasifikasi dalam machine learning mencerminkan efektivitas dan akurasi model dalam memprediksi atau mengelompokkan data baru berdasarkan pola yang telah dipelajari selama proses pelatihan. Metrik evaluasi, seperti akurasi, presisi, recall, dan F1-score, sering digunakan untuk menilai performa model. Akurasi mencerminkan sejauh mana model dapat mengenali kelas dengan benar, sementara presisi mengukur keakuratan prediksi positif, recall menilai kemampuan model untuk menemukan semua instance positif yang sebenarnya, dan F1-score memberikan keseimbangan antara presisi dan recall. Hasil klasifikasi yang baik ditandai oleh nilai-nilai metrik evaluasi yang tinggi, menunjukkan kemampuan model untuk mengatasi data yang belum pernah dilihat sebelumnya dengan baik. Analisis hasil klasifikasi juga melibatkan pemeriksaan matriks kebingungan (confusion matrix) untuk memahami sejauh mana model dapat mengidentifikasi

kelas-kelas dengan benar dan mengidentifikasi area di mana model mungkin perlu peningkatan. Kesimpulannya, hasil klasifikasi menjadi indikator penting dalam menilai kinerja model machine learning dan memberikan wawasan berharga untuk pengembangan selanjutnya atau pengoptimalan model.

7. *Evaluasi Model Klasifikasi*

Evaluasi model klasifikasi dalam machine learning adalah tahap kritis yang memberikan gambaran menyeluruh tentang sejauh mana model dapat memprediksi dengan akurat dan konsisten. Metode evaluasi umumnya melibatkan penggunaan metrik-metrik seperti akurasi, presisi, recall, dan F1-score, yang memberikan pandangan yang lebih mendalam tentang kinerja model. Akurasi mengukur persentase prediksi yang benar secara keseluruhan, presisi fokus pada keakuratan prediksi positif, recall mengukur kemampuan model untuk menemukan semua instance positif yang sebenarnya, dan F1-score mengukur keseimbangan antara presisi dan recall. Matriks kebingungan (confusion matrix) juga digunakan untuk memberikan gambaran lebih rinci tentang sejauh mana model dapat mengklasifikasikan instance dengan benar atau salah. Selama evaluasi, analisis terhadap kurva ROC (Receiver Operating Characteristic) dan area di bawah kurva (AUC-ROC) juga dapat memberikan informasi tambahan tentang sensitivitas dan spesifisitas model. Evaluasi model yang komprehensif membantu memahami kekuatan dan kelemahan model, memandu penyesuaian parameter, dan memungkinkan pengembangan model yang lebih baik dan lebih andal pada data yang belum pernah dilihat sebelumnya.

8. Hasil Evaluasi Model Klasifikasi

Hasil evaluasi model klasifikasi dalam machine learning memberikan pandangan mendalam tentang kinerja dan kehandalan model dalam memprediksi atau mengklasifikasikan data. Metrik evaluasi, seperti akurasi, presisi, recall, dan F1-score, memberikan informasi kuantitatif tentang seberapa baik model dapat mengenali kelas atau kategori data tertentu. Akurasi mencerminkan tingkat keseluruhan kebenaran prediksi, sementara presisi memberikan gambaran tentang seberapa banyak prediksi positif yang benar, recall menunjukkan kemampuan model untuk menemukan semua instance positif yang sebenarnya, dan F1-score mengukur keseimbangan antara presisi dan recall. Hasil evaluasi yang baik ditandai oleh nilai-nilai metrik yang tinggi. Analisis terhadap matriks kebingungan memberikan pemahaman lebih rinci tentang seberapa baik model dapat membedakan antara kelas-kelas yang berbeda. Selain itu, kurva ROC dan nilai AUC-ROC memberikan pandangan tentang sensitivitas dan spesifisitas model terhadap perubahan ambang batas keputusan. Dengan hasil evaluasi yang memuaskan, pemangku kepentingan dapat dengan lebih yakin menggunakan model tersebut untuk pengambilan keputusan atau memahami lebih baik area di mana model memerlukan peningkatan. Evaluasi model klasifikasi yang hati-hati dan terperinci adalah langkah penting dalam memastikan implementasi model yang efektif dan dapat diandalkan dalam berbagai konteks.

3.2. Pengumpulan Data

Pengumpulan data merupakan langkah kritis dalam proses penelitian dan analisis yang melibatkan pengumpulan informasi untuk menjawab pertanyaan

penelitian atau memecahkan masalah tertentu. Ada berbagai metode pengumpulan data yang dapat digunakan, tergantung pada tujuan penelitian, sumber daya yang tersedia, dan konteksnya. Salah satu metode pengumpulan data yang umum adalah survei. Survei dapat dilakukan melalui kuesioner atau wawancara, baik langsung atau melalui telepon atau daring. Survei berguna untuk mengumpulkan data dari responden yang mewakili kelompok tertentu dan dapat memberikan wawasan mendalam tentang pandangan, preferensi, atau perilaku mereka.

Selain itu, observasi juga merupakan metode efektif untuk pengumpulan data, di mana peneliti secara langsung mengamati dan mencatat perilaku, situasi, atau kejadian tertentu. Observasi dapat dilakukan secara partisipatif, di mana peneliti terlibat langsung, atau non-partisipatif, di mana peneliti hanya sebagai pengamat. Pengumpulan data juga dapat melibatkan penggunaan sumber data sekunder, yaitu data yang sudah ada dan dikumpulkan untuk tujuan lain. Sumber data sekunder dapat mencakup data statistik, literatur, atau catatan administratif yang dapat memberikan wawasan tambahan atau mendukung analisis penelitian. Namun, penting untuk mempertimbangkan etika pengumpulan data, termasuk privasi dan keamanan informasi. Peneliti harus memastikan bahwa partisipan memberikan persetujuan yang tepat dan bahwa data dikumpulkan dan disimpan dengan aman.

Pengumpulan data dalam konteks machine learning sangat krusial karena kualitas dan relevansi data memainkan peran penting dalam keberhasilan model. Data yang tidak memadai atau bias dapat mengakibatkan model yang tidak akurat atau tidak dapat diandalkan. Ada beberapa langkah dan pertimbangan khusus yang harus diperhatikan dalam pengumpulan data untuk proyek machine learning.

Langkah pertama adalah mendefinisikan dengan jelas tujuan dan pertanyaan penelitian yang ingin dijawab oleh model machine learning. Hal ini akan membantu dalam menentukan jenis data yang diperlukan dan sumber daya yang dibutuhkan untuk pengumpulan data.

Selanjutnya, penting untuk menentukan variabel atau fitur apa yang akan digunakan oleh model. Variabel ini harus relevan dengan tujuan model dan mewakili informasi yang diperlukan untuk membuat prediksi atau keputusan yang diinginkan. Sumber data dapat berasal dari berbagai sumber, termasuk database internal perusahaan, data sensor, atau sumber eksternal seperti API dan data yang tersedia secara publik. Pemilihan sumber data harus mempertimbangkan keakuratan, kelengkapan, dan relevansi data tersebut terhadap konteks masalah yang dihadapi.

Tahap pengumpulan data juga melibatkan praproses data, termasuk pembersihan data, penghilangan nilai yang hilang, dan normalisasi. Proses ini bertujuan untuk memastikan bahwa data yang digunakan oleh model adalah berkualitas tinggi dan siap untuk diproses. Selama pengumpulan data, perlu diingat pentingnya etika data, termasuk privasi dan keamanan informasi. Pengumpulan data yang dilakukan harus mematuhi regulasi dan standar etika yang berlaku. Selama pengumpulan data, perlu diingat pentingnya etika data, termasuk privasi dan keamanan informasi. Pengumpulan data yang dilakukan harus mematuhi regulasi dan standar etika yang berlaku.

Selama pengumpulan data, perlu diingat pentingnya etika data, termasuk privasi dan keamanan informasi. Pengumpulan data yang dilakukan harus

mematuhi regulasi dan standar etika yang berlaku. Terakhir, dalam konteks machine learning, seringkali diperlukan jumlah data yang besar untuk melatih model dengan baik. Oleh karena itu, perlu dipertimbangkan bagaimana data tersebut dapat dihimpun dengan efisien dan memadai untuk mencakup berbagai skenario yang mungkin dihadapi oleh model. Dengan pengumpulan data yang teliti dan hati-hati, proyek machine learning dapat dimulai dengan landasan yang kuat, memungkinkan model untuk memberikan hasil yang akurat dan berguna dalam mengatasi tantangan yang dihadapi.

3.3. Metode yang Diusulkan

3.3.1. Metode Naive Bayes

Metode Naive Bayes adalah algoritma klasifikasi yang bersumber dari teorema probabilitas Bayes [14]. Meskipun sederhana, metode ini seringkali sangat efektif dan efisien, terutama dalam konteks klasifikasi teks dan analisis sentimen. Asumsi utama di balik "naive" dalam Naive Bayes adalah independensia kondisional antar fitur, artinya, setiap fitur dianggap tidak terkait satu sama lain setelah diberikan informasi tentang kelas yang diinginkan [15]. Langkah utama dalam metode Naive Bayes melibatkan penghitungan probabilitas posterior dari kelas yang diinginkan berdasarkan fitur-fitur yang diamati. Ini dilakukan dengan menerapkan teorema Bayes, yang menyatakan bahwa probabilitas suatu kelas setelah mengamati fitur adalah sebanding dengan probabilitas prior kelas tersebut dikalikan dengan probabilitas likelihood fitur-fitur tersebut bila diberikan kelas tersebut.

Metode Naive Bayes memiliki beberapa variasi, tergantung pada distribusi probabilitas yang diasumsikan untuk fitur-fitur. Naive Bayes Gaussian cocok untuk data yang terdistribusi normal, sementara Naive Bayes Multinomial lebih sesuai untuk data kategorikal seperti data teks yang direpresentasikan sebagai model Bag-of-Words. Naive Bayes Bernoulli cocok untuk data biner, yang seringkali digunakan dalam klasifikasi dokumen. Kelebihan utama dari metode Naive Bayes adalah sederhananya dan kemampuannya mengatasi dimensi fitur yang tinggi. Meskipun asumsi independensi fitur seringkali tidak sepenuhnya terpenuhi dalam dunia nyata, metode ini tetap menghasilkan hasil klasifikasi yang cukup baik. Kekurangannya termasuk sensitivitas terhadap data yang tidak seimbang dan tidak dapat menangani hubungan non-linear antar fitur. Metode Naive Bayes diterapkan dalam berbagai konteks, termasuk klasifikasi teks, analisis sentimen, dan pengenalan pola. Terutama efektif dalam kasus-kasus di mana asumsi independensi fitur dapat diterima, metode Naive Bayes tetap menjadi pilihan yang populer dalam proyek-proyek machine learning yang memerlukan solusi sederhana dan cepat.

Metode Naive Bayes adalah salah satu pendekatan yang umum digunakan dalam machine learning untuk tugas klasifikasi. Metode ini berlandaskan pada teorema probabilitas Bayes, dengan asumsi naif (naive) bahwa semua fitur yang digunakan untuk klasifikasi adalah independen satu sama lain, meskipun ketergantungan yang sebenarnya mungkin ada di antara fitur-fitur tersebut. Meskipun sederhana, metode Naive Bayes sering memberikan kinerja yang baik dan dapat diterapkan dengan efisien dalam berbagai konteks. Dalam konteks klasifikasi, metode Naive Bayes digunakan untuk memprediksi kelas suatu instance

data berdasarkan distribusi probabilitas dari fitur-fitur yang diamati. Proses ini melibatkan perhitungan probabilitas prior (sebelum pengamatan fitur) dan probabilitas likelihood (probabilitas fitur terjadi bila kelas diketahui) untuk setiap kelas yang mungkin. Melalui aplikasi teorema Bayes, probabilitas posterior dari setiap kelas dapat dihitung, dan kelas dengan probabilitas tertinggi dianggap sebagai prediksi model.

Naive Bayes memiliki variasi berdasarkan distribusi probabilitas yang diasumsikan untuk fitur-fiturnya. Misalnya, Naive Bayes Gaussian cocok untuk data kontinu yang terdistribusi normal, sementara Naive Bayes Multinomial lebih sesuai untuk data kategorikal seperti kata-kata dalam teks. Naive Bayes Bernoulli cocok untuk data biner, yang seringkali muncul dalam konteks klasifikasi dokumen atau data kehadiran. Meskipun asumsi naif tentang independensi fitur, metode Naive Bayes dapat memberikan hasil yang mengesankan terutama dalam tugas klasifikasi teks, spam filtering, dan analisis sentimen. Kelebihan utamanya termasuk kecepatan komputasi tinggi, kemampuan menangani dimensi fitur yang tinggi, dan kemudahan interpretasi. Namun, metode Naive Bayes dapat kurang efektif dalam menangani data yang tidak seimbang dan tidak cocok untuk masalah di mana hubungan antar fitur sangat kompleks.

3.4. Eksperimen dan Pengujian Metode

Untuk tahapan pengujian yang merupakan tahapan pengolahan data dalam sistem manual, penulis membutuhkan 2 data set yang akan digunakan pada penelitian ini. Untuk data set yang pertama yaitu data testing dan data set yang kedua yaitu data training.

3.4.1. Data Testing

Data testing merupakan data sampel Penelitian yang digunakan untuk diolah nantinya. Untuk data testing yang digunakan penulis yaitu berjumlah sebanyak 30 data testing.

Nama	Batuk	Sesak Nafas	Hidung tersumbat	Gatal-Gatal
Adi Nugroho	Ya	Ya	Ya	Ya
Arif Maulana	Tidak	Ya	Ya	Ya
Benny Hardi	Ya	Tidak	Ya	Ya
Bunga Citra	Ya	Ya	Ya	Ya
Cindy Oktavia	Ya	Ya	Ya	Tidak
Dede Supriadi	Tidak	Ya	Ya	Ya
Elly Kristiani	Tidak	Ya	Ya	Ya
Fenny Marlina	Ya	Tidak	Ya	Ya
Guntur Prasetyo	Ya	Ya	Tidak	Ya
Hera Sari	Tidak	Tidak	Tidak	Tidak
Intan Dewi	Tidak	Tidak	Tidak	Tidak
Joni Iskandar	Ya	Tidak	Tidak	Tidak
Kusuma Wardani	Tidak	Ya	Tidak	Tidak
Lusi Ananda	Tidak	Tidak	Ya	Tidak
Miko Permana	Tidak	Tidak	Tidak	Ya
Nana Suryana	Ya	Ya	Ya	Ya
Oka Yudistira	Ya	Ya	Ya	Ya
Putu Aris	Ya	Ya	Ya	Ya
Rani Fitri	Ya	Ya	Ya	Ya
Sandi Maulana	Ya	Ya	Ya	Ya
Tita Andriani	Ya	Ya	Ya	Ya
Umar Dani	Ya	Ya	Ya	Ya
Vera Anggraeni	Ya	Ya	Ya	Ya
Vina Permata	Ya	Ya	Ya	Ya
Wahyu Ramadhan	Ya	Ya	Ya	Ya
Wawan Hermawan	Ya	Ya	Ya	Ya
Yoga Pratama	Ya	Ya	Ya	Ya
Yulia Rachma	Ya	Ya	Ya	Ya
Zaki Fadli	Ya	Ya	Tidak	Ya
Zulkifli Hasan	Ya	Ya	Ya	Ya

3.4.2. Data Training

Data training merupakan data latihan yang digunakan untuk membantu proses perhitungan data. Dengan adanya data training, maka data testing dapat diolah dan dihitung. Untuk data training yang akan digunakan pada Penelitian ini yaitu sebanyak 20 data training.

Nama	Batuk	Sesak Nafas	Hidung tersumbat	Gatal-Gatal	Keterangan
Andi Setiawan	Ya	Ya	Ya	Ya	Alergi
Budi Santoso	Ya	Ya	Ya	Ya	Alergi
Chandra Wijaya	Ya	Ya	Ya	Ya	Alergi
Dian Purnama	Ya	Ya	Tidak	Ya	Alergi
Eka Putra	Tidak	Ya	Ya	Ya	Alergi
Fajar Hidayat	Ya	Tidak	Ya	Ya	Alergi
Gita Lestari	Ya	Ya	Ya	Tidak	Alergi
Hadi Saputra	Tidak	Ya	Ya	Ya	Alergi
Indra Gunawan	Tidak	Ya	Ya	Ya	Alergi
Joko Susanto	Ya	Tidak	Ya	Ya	Alergi
Kiki Amelia	Ya	Ya	Tidak	Ya	Alergi
Laila Rahma	Tidak	Tidak	Tidak	Tidak	Tidak Alergi
Made Surya	Tidak	Tidak	Tidak	Tidak	Tidak Alergi
Nia Saraswati	Ya	Tidak	Tidak	Tidak	Tidak Alergi
Oki Pratama	Tidak	Ya	Tidak	Tidak	Tidak Alergi
Putri Melati	Tidak	Tidak	Ya	Tidak	Tidak Alergi
Rizki Fadillah	Tidak	Tidak	Tidak	Ya	Tidak Alergi
Sari Utami	Ya	Tidak	Tidak	Tidak	Tidak Alergi
Tono Wibowo	Tidak	Tidak	Ya	Tidak	Tidak Alergi
Udin Jafar	Tidak	Tidak	Tidak	Ya	Tidak Alergi

Untuk data training diatas, tidak dapat langsung digunakan untuk pengolahan data. Jadi data training harus di pisah dari setiap atributnya. Untuk data training yang akan dipisah dapat dilihat pada table dibawah ini.

Atribut	Partisi	Alergi	Tidak Alergi	P (Alergi)	P (Tidak Alergi)
Batuk	Ya	8	2	8/11	2/9
	Tidak	3	7	3/11	7/9
	Total	11	9	100%	100%

Atribut	Partisi	Alergi	Tidak Alergi	P (Alergi)	P (Tidak Alergi)
Sesak Nafas	Ya	9	1	9/11	1/9
	Tidak	2	8	2/11	8/9
	Total	11	9	100%	100%

Atribut	Partisi	Alergi	Tidak Alergi	P (Alergi)	P (Tidak Alergi)
Hidung Tersumbat	Ya	9	2	9/11	1/9
	Tidak	2	7	2/11	7/9
	Total	11	9	100%	100%

Atribut	Partisi	Alergi	Tidak Alergi	P (Alergi)	P (Tidak Alergi)
Gatal-Gatal	Ya	10	2	10/11	2/9
	Tidak	1	7	1/11	7/9
	Total	11	9	100%	100%

Kategori		P (Alergi & Tidak Alergi)
Alergi	11	11/20
Tidak Alergi	9	9/20

3.4.3. Perhitungan

Untuk perhitungan pada metode Naive Bayes menggunakan rumus sebagai berikut.

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$P(A|B)$ = Probabilitas A bersyarat yang diberikan oleh B

$P(B|A)$ = Probabilitas B bersyarat yang diberikan oleh A

$P(A)$ = Probabilitas kejadian A

$P(B)$ = Probabilitas kejadian B

Untuk perhitungan yang akan dilakukan, pertama penulis akan menghitung data Adi Nugroho. Adapun perhitungannya sebagai berikut.

$$P(\text{Kategori}) = P(\text{Batuk}|\text{Sulit}) \times P(\text{Sesak Nafaslah}|\text{Lengkap}) \times P(\text{Hidung Tersumbat}|\text{Kurang Bagus}) \times P(\text{Gatal-Gatal}|\text{Setuju}) \times P(\text{Kategori}|\text{Alergi})$$

$$P(\text{Alergi}) = P(\text{Ya}|\text{Alergi}) \times P(\text{Ya}|\text{Alergi}) \times P(\text{Ya}|\text{Alergi}) \times P(\text{Ya}|\text{Alergi}) \times P(\text{Kategori}|\text{Alergi})$$

$$= \left(\frac{8}{11}\right) \times \left(\frac{9}{11}\right) \times \left(\frac{9}{11}\right) \times \left(\frac{10}{11}\right) \times \left(\frac{11}{20}\right)$$

$$= 0,243425 \text{ (Nilai Alergi)}$$

$$\begin{aligned}
P(\text{Tidak Alergi}) &= P(\text{Sulit} | \text{Tidak Alergi}) \times P(\text{Lengkap} | \text{Tidak Alergi}) \times P(\text{Kurang} \\
&\quad \text{Bagus} | \text{Tidak Alergi}) \times P(\text{Setuju} | \text{Tidak Alergi}) \times \\
&\quad P(\text{Mahal} | \text{Tidak Alergi}) \times P(\text{Kategori} | \text{Tidak Alergi}) \\
&= \binom{2}{9} \times \binom{1}{9} \times \binom{1}{9} \times \binom{2}{9} \times \binom{9}{20} \\
&= 0,022222 \text{ (Nilai Tidak Alergi)}
\end{aligned}$$

Dari hasil perhitungan dan pengolahan data pada metode naive bayes dengan menghitung data dari Adi Nugroho, bahwasanya dari hasil yang diperoleh nilai Alergi yaitu sebesar 0,243425 dan nilai dari Tidak Alergi yaitu sebesar 0,022222. Jadi dari hasil pengolahan data bahwasanya Adi Nugroho terkena penyakit alergi.