**1.2 Langkah-langkah Algoritma Naive Bayes**

Algoritma Naive Bayes merupakan suatu metode klasifikasi yang berdasarkan pada teorema Bayes dengan asumsi independensi kondisional antara fitur-fitur. Secara umum, langkah-langkah algoritma Naive Bayes melibatkan:

1. Persiapan Data

* Pengumpulan dataset yang akan digunakan pada pelatihan dan pengujian.
* Pengidetifikasian fitur-fitur relevan dalam data set yang dapat berupa nilai diskrit maupun kontinu.
* *Preprocessing* dada seperti pembersihan data, penghapusan noise dan pengisian data yang hilang.

1. Perhitungan Probabilitas

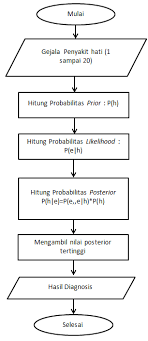
* Perhitungan probabilitas kelas-kelas pada dataset.
* Perhitungan probabilitas kondisional terhadap setiap kelas untuk setiap fitur.

1. Pelatihan Model

* Penggunaan dataset yang telah dipersiapkan dalam melatih model Naive Bayes.
* Perhitungan probabilitas posterior untuk setiap kelas dengan menggunakan teorema Bayes.
* Pembaruan probabilitas kondisional dan probabilitas kelas setiap ada pelatihan baru.

1. Klasifikasi

* Perhitungan probabilitas posterior pada setiap kelas menggunakan model yang telah dilatih Ketika terdapat data pengujian baru.
* Penentuan kelas dengan probabilitas posterior tertinggi sebagai prediksi.



**1.3 Model Klasifikasi Naive Bayes**

**1.3.1 Naïve Bayes Multinomial**

Naïve Bayes Multinomial merupakan model klasifikasi yang digunakan ketika data input memiliki atribut-atribut diskrit yang menggambarkan frekuensi atau jumlah kemunculan suatu fitur. Model ini mengasumsikan bahwa fitur-fitur bersifat multinomial secara independen terhadap kelasnya yang di mana perhitungan probabilitas menggunakan distribusi multinomial.

Berikut merupakan rumus perhitungan Naïve Bayes Multinomial :

Untuk menghindari adanya probabilitas nol, akan dilakukan *laplace smoothing* dengan penambahan +1 pada pembilang, dan |v| pada penyebut dengan rumus berikut :

Keterangan :

adalah jumlah kemunculan fitur w pada kategori c

adalahjumlah total kemunculan semua fitur pada kategori c

adalah banyaknya fitur unik

Naïve Bayes Multinomial memiliki beberapa kelebihan, yaitu sederhana dan mudah diimplementasikan. Selain itu model ini memiliki kinerja yang baik dalam menghadapi data berdimensi tinggi dan jarang terisi (sparse data). Model ini juga mampu mengatasi jumlah kelas yang besar, sehingga dapat diterapkan pada berbagai jenis masalah klasifikasi yang melibatkan banyak kelas.

Kelemahan Multinomial Naïve Bayes adalah fitur-fitur dalam data saling independen, dimana dalam praktiknya fitur-fitur sering didapati adanya keterkaitan, sehingga mengakibatkan prediksi yang tidak akurat. Selain itu, model ini membutuhkan jumlah data latih yang besar dalam mengestimasi parameter-parameter dengan akurat. Model ini juga dapat sensitive terhadap fitur-fitur yang tidak relevan.

**1.3.2 Naïve Bayes Gaussian**

Naïve Bayes Gaussian merupakan model klasifikasi yang digunakan ketika atribut input merupakan data kontinu yang mengikuti distribusi Gaussian yang independen satu sama lain. Algoritma ini bekerja diawali dengan memperkirakan distribusi probabilitas dari setiap fitur yang mempertimbangkan label kelas. Hal ini dilakukan dengan menghitung rata-rata dan standar deviasi dari setiap fitur untuk kelas tersebut.

Berikut merupakan rumus perhitungan Naïve Bayes Gaussian :

Keterangan :

adalah nilai phi 3.14…

adalah varian nilai fitur w pada kelas c

adalah nilai fitur w pada data uji

adalah rata-rata nilai fitur w pada kelas c

Naïve Bayes Gaussian memiliki beberapa kelebihan, yaitu sederhana, mudah dipahami, dan relative mudah untuk diimplementasikan. Algoritma ini efisien secara komputasional dan mampu menangani dataset besar dengan dimensi yang tinggi. Algoritma ini bekerja dengan baik terhadap dataset berukuran kecil hingga menengah, dan akan lebih efektif ketika jumlah fitur cenderung kecil jika dibandingkan dengan jumlah observasi. Naïve Bayes Gaussian juga mampu memberikan performa yang baik ketika asumsi independensi antar fitur tidak sepenuhnya terpenuhi. Selain itu, algoritma ini dapat mengatasi nilai data yang hilang dengan cara mengabaikannya selama proses pengembangan model.

Naïve Bayes Gaussian memiliki beberapa kekurangan, seperti kurangnya keakuratan dalam beberapa kasus yang disebabkan oleh fitur-fitur pada data yang saling independen. Algoritma ini tidak dapat menangani fitur-fitur yang berkorelasi, sehingga diperlukan algoritma yang lebih canggih. Algoritma ini juga tidak cocok untuk kasus yang membutuhkan kalibrasi probabilitas. Selain itu, algoritma ini sensitif terhadap *outlier* pada data yang di mana dapat mempengaruhi rata-rata dan variasi dari nilai-nilai fitur, sehingga menghasilkan asil yang tidak akurat. Naïve Bayes Gaussian juga tidak cocok dalam mengatasi data teks atau gambar, yang di mana ruang fitur dapat sangat besar dan fitur-fiturnya tidak selalu saling independen.

**1.3.3 Naïve Bayes Bernoulli**

Naïve Bayes Bernoulli merupakan model klasifikasi yang digunakan ketika atribut inputnya merupakan data biner atau data yang dapat diubah menjadi biner, sehingga fitur hanya memiliki dua nilai, yaitu 0 atau 1. Model ini cocok untuk kasus yang mengidentifikasi keberadaan atau ketiadaan fitur tertentu dalam dataset.

Berikut merupakan rumus perhitungan Naïve Bayes Gaussian :

Untuk menghindari adanya probabilitas nol, akan dilakukan *laplace smoothing* dengan penambahan +1 pada pembilang, dan +2 pada penyebut dengan rumus berikut :

Keterangan :

Jika fitur *w* ada pada data berkategori c, maka = 1

Jika fitur *w* tidak ada pada data berkategori c, maka = 0

= banyaknya data berkategori c

= banyaknya data berkategori c yang mengandung fitur w

Naïve Bayes Bernoulli memiliki beberapa kelebihan, yaitu memiliki efisiensi komputasi yang tinggi, sehingga menjadi pilihan yang baik untuk dataset besar. Model ini memiliki kinerja yang baik pada data biner, yaitu data yang hanya memiliki dua nilai kemungkinan, yaitu 0 dan 1. Selain itu, model ini membutuhkan jumlah data latih yang lebih sedikit jika dibandingkan dengan algoritma klasifikasi lainnya. Model ini juga toleransi terhadap fitur-fitur yang tidak relevan dan dapat menghasilkan hasil yang akurat.

Naïve Bayes Bernoulli memiliki beberapa kekurangan, yaitu asumsi fitur-fitur pada data saling independen, di mana dapat mengakibatkan hasil yang tidak akurat. Selain itu, model ini tidak dapat menangani data kontinu dan tidak dapat menangkap hubungan kompleks antara fitur-fitur dan variable target. Bernoulli Naïve Bayes juga membutuhkan seleksi fitur, di mana hanya fitur-fitur yang relevan yang digunakan menjadi input.

**1.4 Kelebihan dan Kekurangan Naive Bayes**

**1.4.1 Kelebihan Naive Bayes**

1. Sederhana dan cepat

Naïve Bayes merupakan suatu metode klasifikasi yang relative sederhana dan cepat jika dilihat dalam hal waktu komputasi, sehingga memungkinkan penggunaan Naïve Bayes pada dataset dengan ukuran yang besar.

1. Efektif dengan data berdimensi tinggi

Naïve bayes memiliki kecenderungan bekerja dengan baik pada dataset yang memiliki banyak fitur atau berdimensi tinggi, sehingga metode ini dapat digunakan pada masalah klasifikasi yang kompleks.

1. Efisien dengan jumlah data yang besar

Naïve bayes dapat mengatasi jumlah data yang besar dengan efisien di mana komputasi yang diperlukan hanya bergantung pada jumlah fitur, bukan jumlah data, sehingga metode ini efisien dalam memproses data set besar.

1. *Robust* terhadap atribut yang tidak relevan

Naïve bayes dapat mengatasi atribut yang tidak relevan atau tidak memiliki dampak besar terhadap hasil prediksi dalam dataset.

1. Tidak sensitif terhadap perbedaan skala

Naïve Bayes tidak terlalu berpengaruh pada perbedaan skala antara fitur dalam dataset, sehingga tidak memerlukan normalisasi data sebelum menggunakan Naïve Bayes.

**1.5.1 Kekurangan Naive Bayes**

1. Asumsi independensi fitur

Naïve Bayes mengasumsikan bahwa seluruh fitur pada dataset adalah independent secara kondisional terhadap kelas target. Hal ini jarang terpenuhi di dunia nyata, sehingga dapat mengakibatkan performa yang tidak terlalu akurat.

1. Tidak mampu mengatasi ketergantungan

Naïve Bayes tidak dapat menangkap ketergantungan kontekstual antar fitur, sehingga menyebabkan performa yang buruk dalam berbagai kasus di mana konteks sangat penting.

1. Sensitif terhadap fitur yang hilang

Naïve Bayes tidak dapat memberikan prediksi yang akurat, sehingga metode ini membutuhkan informasi mengenai seluruh fitur dalam dataset.

1. Tidak dapat menghasilkan probabilitas yang optimal

Probabilitas yang dihasilkan Naïve Bayes tidak selalu optimal meskipun memberikan prediksi kelas yang cukup baik, terutama pada kasus dataset yang tidak memenuhi asumsi independensi. Dalam hal ini, probabilitas yang diestimasikan oleh Naïve Bayes dapat menjadi bias atau tidak akurat.