

# BAB I

## PENDAHULUAN

### 1.1 Latar belakang

Susu merupakan sumber gizi yang bermanfaat bagi manusia yang mengandung protein hewani. Susu memiliki kandungan protein cukup tinggi, sehingga sangat membantu dalam pertumbuhan, kecerdasan, dan daya tahan tubuh manusia (Subagyo et al. 2022). Ada banyak protein yang terkandung dalam susu seperti kasein, protein whey, protein globul lemak susu, enzim dan protein minor lainnya (Purwantiningsih, Bria, and Kia 2022). Selain itu susu juga dapat mencegah *stunting* pada pertumbuhan anak usia dini (Gunawan, Gunawan, and Bima Aoron Hafiz 2020). Untuk mengetahui susu mana yang berkualitas dan tidak berkualitas, maka peneliti akan menggunakan *dataset* susu yang didapatkan dari *kaggle.com* untuk diuji kualitasnya sesuai atribut data yang didapatkan, tentu atribut akan di klasifikasikan terlebih dahulu sesuai atributnya, maka dari itu peneliti akan menggunakan *Algoritma Naive Bayes* untuk proses pengklasifikasian. *Algoritma Naive Bayes* adalah salah satu dari metode *machine learning* yang dapat digunakan untuk mengklasifikasikan suatu data. *Naive Bayes* adalah pengklasifikasi probabilistik sederhana berdasarkan penerapan teorema Bayes (dari statistik Bayes) (Zeniarta, Widia, and Sani 2020). Untuk mendapatkan hasil yang maksimal dalam menggunakan *Algoritma Naive Bayes* maka dibutuhkan fitur *Forward Selection* untuk hasil yang optimal (Saifudin et al. 2020).

Penelitian ini memusatkan perhatian pada beberapa *Algoritma* klasifikasi untuk menilai mana yang memiliki akurasi tertinggi dalam menentukan kualitas susu. Data mining menyediakan beragam *Algoritma* klasifikasi, seperti *K-Nearest Neighbors*, *Artificial Neural Network*, *Naive Bayes*, *Regresi Linear*, dan *C4.5*. Namun, fokus penelitian ini terbatas pada satu metode klasifikasi untuk memperoleh pemahaman yang lebih dalam tentang dampaknya terhadap penilaian kualitas susu. yaitu *Naive Bayes*. Setelah dilakukan pengolahan data memakai teknik klasifikasi data mining *Algoritma Naive Bayes*, hasil yang didapat untuk pembagian data menggunakan *K-Fold Cross Validation K=11* pada *Algoritma Naive Bayes* didapat lah hasil 75,78% dan untuk pengolahan menggunakan knn dengan

nilai  $K=25$  didapat hasil 74,48%. Dari hasil tersebut *Naive Bayes* lebih baik dibandingkan *K-Nearest Neighbor(KNN)*(Rinanda et al. 2022). Dengan ini peneliti memilih menggunakan *Algoritma Naive Bayes*. sehingga metode ini termasuk dalam kategori klasifikasi terbaik Dalam data mining.

Klasifikasi merupakan proses pengelompokan data dimana pengelompokannya harus berdasarkan karakteristik atribut data yang diolah(Dahlan et al. 2023). Klasifikasi ini bertujuan untuk mengelompokkan objek atau data kedalam kategori tertentu, memfasilitasi pemahaman, analisis, dan pengambilan keputusan dalam konteks yang berbeda(Ayu, Purnama, and Putra 2024) Dengan begitu, kitadapat menyimpulkan bahwa klasifikasi adalah proses mengelompokkan dan mendefinisikan golongan kelas dan menentukan hubungan antara kelas berdasarkan dengan sistem(Cholid Wahyudin 2020).

Riset ini dilaksanakan dengan tujuan untuk mengoptimalkan kinerja Algoritma klasifikasi *Naive Bayes* itu sendiri pada proses klasifikasi kualitas susu, dengan menambahkan metode seleksi fitur. Penelitian ini menerapkan metode seleksi fitur *Forward Selection* untuk meningkatkan akurasi Algoritma klasifikasi *Naive Bayes* dengan cara mengidentifikasi dan menghapus atribut yang tidak relevan. Metode ini bertujuan untuk mengoptimalkan performa Algoritma dengan mempertahankan atribut yang memberikan kontribusi signifikan dalam penilaian kualitas susu(Zeniarja, Widia, and Sani 2020).

## **1.2 Identifikasi masalah**

1. Performa fitur *Forward Selection* untuk *Naive Bayes*
2. Optimasi Algoritma *Naive Bayes*

## **1.3 Rumusan masalah**

1. Bagaimana mengklasifikasi kualitas susu menggunakan Algoritma *Naive Bayes* dengan seleksi fitur *Forward selection*?
2. Bagaimana mendapatkan akurasi yang optimal dalam klasifikasi kualitas susu?

## **1.4 Tujuan penelitian**

1. Menentukan atribut-atiribut yang paling berpengaruh dalam *dataset* susu dengan Algoritma *Naive Bayes* dengan seleksi fitur *Forward Selection*

2. Meningkatkan akurasi terhadap klasifikasi kualitas susu dengan metode *Naive Bayes* menggunakan seleksi fitur *Forward Selection*

## 1.5 Manfaat penelitian

### Manfaat Praktis:

1. Penelitian ini akan berkontribusi secara praktis bagi industri pengolahan susu dengan meningkatkan akurasi dalam menilai kualitas susu menggunakan Algoritma klasifikasi *Naive Bayes*.
2. Dengan menggunakan metode seleksi fitur *Forward Selection*, penelitian ini akan mengidentifikasi atribut-atribut yang paling berpengaruh dalam menentukan kualitas susu, memungkinkan fokus pada atribut yang relevan untuk pengambilan keputusan yang lebih efisien.
3. Penggunaan data yang lebih akurat dari hasil penelitian ini akan memperbaiki proses pengambilan keputusan terkait pengolahan dan distribusi susu di industri.

### Manfaat Teoritis:

1. Penelitian ini akan memberikan pemahaman mendalam tentang penerapan *Forward Selection* dalam meningkatkan performa Algoritma *Naive Bayes*, berkontribusi pada pengembangan praktik data mining.
2. Validitas hasil klasifikasi berdasarkan seleksi atribut akan diperkuat, mendukung keandalan penggunaan data dalam konteks evaluasi kualitas produk.
3. Kontribusi teoritisnya mencakup pengembangan metode klasifikasi untuk evaluasi kualitas produk, memberikan wawasan baru dalam penggunaan Algoritma klasifikasi dalam industri susu.