

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Pada era *Big Data* saat ini, dengan data yang besar kita dapat menggali informasi dengan salah satu *studi* yaitu *sentiment analysis* atau *opinion mining* adalah studi komputasional dari opini-opini orang, sentimen dan emosi melalui entitas dan atribut yang dimiliki yang diekspresikan dalam bentuk teks [1]. Analisis sentimen akan mengelompokkan polaritas dari teks yang ada dalam kalimat atau dokumen untuk mengetahui pendapat yang dikemukakan dalam kalimat atau dokumen tersebut apakah bersifat positif, negatif atau netral [2].

Analisis sentimen mulai populer pada tahun 2002, yang dilakukan analisis sentimen adalah proses menyajikan informasi dengan membangun sebuah sistem yang dapat mengklasifikasikan dokumen teks ke dalam dua kategori, yakni nilai positif dan negatif yang sesuai dengan keseluruhan sentimen yang dinyatakan di dalam setiap dokumen tersebut [3].

Seperti yang pernah dilakukan pada penelitian sebelumnya, yaitu tentang Klasifikasi Emosi dan Sentimen Pada Akun Media Sosial Teknik Informatika Dengan Algoritma *Synesketch*, dimana pada penelitian tersebut membahas mengenai klasifikasi emosi pada *tweet* dari media sosial twitter teknik informatika menggunakan Algoritma *Synesketch* yang mampu mengklasifikasikan emosi sesuai enam kategori emosi yaitu marah, jijik, takut, gembira, sedih dan terkejut, lalu dari data hasil klasifikasi emosi menjadi sentimen positif atau negatif melalui perhitungan valensi pada vektor emosi. Berdasarkan skenario pengujian yang

telah dirancang didapat tingkat akurasi sistem sebesar 89.63% dalam melakukan klasifikasi emosi dengan kategori emosi dominan gembira. Berdasarkan skenario pengujian yang telah dirancang didapat tingkat akurasi sistem sebesar 94.19% dengan sentimen dominan yaitu sentimen positif [4].

Dalam penelitian yang lain analisis sentimen terhadap data Twitter mengenai penggunaan transportasi umum darat dalam kota dapat dilakukan dengan metode Support Vector Machine, dengan akurasi mencapai 78,12% pada dataset Transjakarta. Hasil akurasi terhadap penggunaan metode Support Vector Machine dipengaruhi oleh beberapa hal, yaitu Komposisi jumlah data training dan testing Jumlah dataset yang digunakan Komposisi jumlah data positif dan negatif [5].

Penelitian yang dilakukan oleh Sucitra Cahaya dan Mochammad Wahyudi mengenai K-Nearest Neighbors Sebagai Analisis Sentiment Review Produk Appstore For Android Klasifikasi text dengan data berupa review aplikasi android, salah satu pengklasifikasian yang dapat digunakan adalah k-Nearest Neighbors (k-NN). Hal ini dikarenakan k-NN merupakan metode yang dapat sesuai dengan klasifikasi data dan mudah dipahami. k-NN juga sering digunakan pada beberapa peneliti dalam klasifikasi teks dan memiliki performa yang baik. Dari pengolahan data yang sudah dilakukan. Data review yang peneliti olah dapat diklasifikasi dengan baik ke dalam bentuk positif dan negatif. Akurasi k-NN pada data review produk pada Appstore For Android mencapai 74.50% dan nilai AUC sebesar 0.825 [6].

Naïve Bayes Classifier merupakan algoritma yang memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes,

yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Algoritma ini merupakan algoritma sederhana yang bekerja dengan baik untuk kategorisasi teks [7].

Produk yang dijadikan bahan pengujian dalam penelitian ini adalah *review* pada akun resmi Persib Bandung yang terdapat pada twitter, dimana menurut web [www.jelasberita.com](http://www.jelasberita.com) PERSIB adalah satu-satunya klub di Indonesia yang bisa bersaing dengan kepopuleran klub-klub elit Eropa di internet. Terbukti, kata kunci Persib adalah yang paling banyak dicari di google selama beberapa tahun, mengalahkan popularitas Manchester United. Persib juga merupakan klub dengan Fans Page Facebook terbesar di Indonesia, mencapai 6 juta likers, mengalahkan jumlah FP Inter Milan. Akun twitter @Persib memiliki pengikut terbesar dibandingkan klub-klub yang lain di Indonesia, mencapai 1,2 juta followers [8]. Diharapkan dengan adanya penelitian ini manajemen bisa lebih memperhatikan opini masyarakat sehingga dapat mengetahui mana yang harus dipertahankan dan mana yang harus diperbaiki.

Namun, Analisis sentimen pada Twitter terdapat kelemahan dalam kata-kata yang terdapat pada kalimat yang diposting oleh pengguna situs tersebut. Twitter hanya memungkinkan pengguna menulis sebanyak 140 karakter, hal ini yang menyebabkan para pengguna sering menggunakan singkatan kata dan ejaan kata yang salah. Cara penulisan yang salah tersebut mengakibatkan terjadi kelemahan pada proses *Text Mining*, dimana dapat menyulitkan fitur yang diambil serta mengurangi ketepatan klasifikasi. Oleh karena itu disini penulis akan menggunakan metode *n-gram* karakter kata untuk mengambil fitur-fitur yang ada pada sebuah *Tweet* yang kemudian akan diklasifikasi dengan Algoritma *Naive*

*Bayes Classifier* dimana *algoritma* ini dirasa paling cocok untuk mengklasifikasi berdasarkan penelitian terdahulu.

Berdasarkan latar belakang diatas, maka penelitian ini mengambil judul **“Analisis Sentimen Tentang Opini Masyarakat Terhadap Klub Persib Bandung Menggunakan Metode *Naive Bayes Classifier*”**

## 1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dari penelitian ini adalah:

1. Bagaimana implementasi metode *Naive Bayes Classifier* untuk menentukan kalimat beropini positif, negative dan netral terhadap *tweet* tentang PERSIB?
2. Bagaimana menentukan fanatisme terhadap PERSIB di masyarakat dalam media sosial *Twitter*?
3. Bagaimana kinerja algoritma *Naive Bayes Classifier* untuk menentukan kalimat beropini positif, negative dan netral terhadap *tweet* tentang PERSIB?

## 1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah:

1. Mengimplementasikan metode *Naive Bayes Classifier* dalam menentukan kalimat beropini positif, negatif dan netral terhadap *tweet* tentang PERSIB.
2. Menentukan fanatisme terhadap PERSIB di masyarakat dalam media sosial *Twitter*.
3. Kinerja algoritma *Naive Bayes Classifier* untuk menentukan kalimat beropini positif, negative dan netral terhadap *tweet* tentang PERSIB

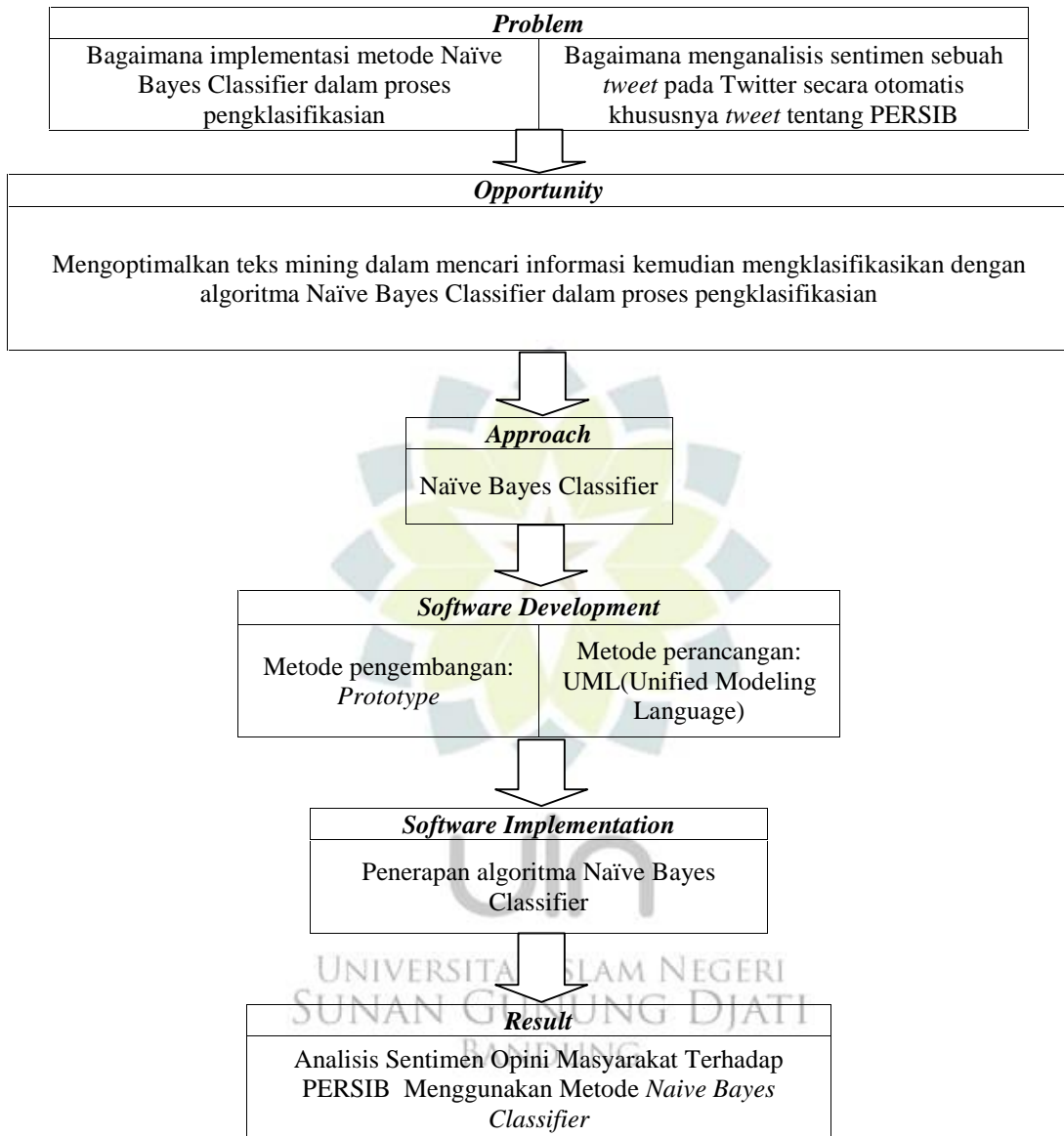
#### 1.4 Batasan Masalah

Agar penelitian ini tidak menyimpang dan lebih teratur, maka pembuatan aplikasi harus dibatasi, diantaranya:

1. Aplikasi ini dibangun dengan menggunakan bahasa pemrograman PHP menggunakan Framework LARAVEL
2. Aplikasi ini dirancang menggunakan model perancangan UML (*Unified Modeling Language*).
3. Aspek proses pengembangan sistem menggunakan metode pengembangan perangkat lunak *Prototype*.
4. Algoritma yang digunakan pada proses klasifikasi yaitu *Naive Bayes Classifier*.
5. Aspek input yang digunakan yaitu data yang di ambil dari media sosial twitter dengan kata kunci #persib, #persibday, #persibofficial, #persibbandung.
6. Aspek output dari sistem dapat mengklasifikasikan sentimen positif, negatif dan netral.
7. Aspek output dari sistem dapat menentukan fanatisme dari akun.
8. Aspek output disajikan dalam bentuk tabel dan diagram.
9. Data yang di ambil dengan jangka waktu tweet selama 10 hari terakhir

## 1.5 Kerangka Pemikiran

Adapun kerangka pemikiran yang digambarkan seperti dibawah ini:



**Gambar 1.1 Kerangka Pemikiran**

## 1.6 Metodologi Penelitian

### 1.6.1 Metode Pengumpulan Data

Metode pengumpulan data yang akan digunakan dalam penelitian ini adalah sebagai berikut:

#### 1. Observasi

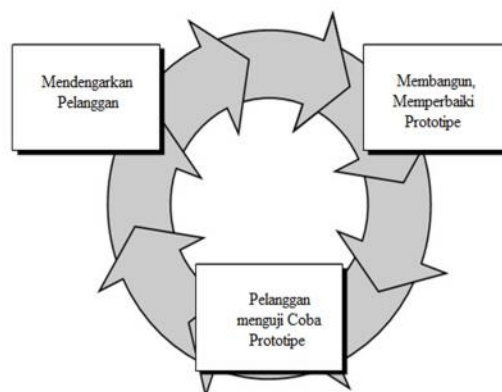
Teknik pengumpulan data dengan mengadakan penelitian dan peninjauan langsung terhadap permasalahan yang diambil.

#### 2. Studi Literatur

Pengumpulan data dengan cara mengumpulkan literatur, jurnal, dan bacaan-bacaan yang ada kaitannya dengan judul penelitian.

### 1.6.2 Metode Pengembangan Perangkat Lunak

Metode pengembangan perangkat lunak yang digunakan untuk penelitian ini, diantaranya adalah model *prototype*. *Prototype* merupakan suatu metode yang diawali dengan mendengarkan pelanggan untuk mengetahui program yang akan di buat. Kemudian membangun dan memperbaiki program yang sesuai dengan keinginan pelanggan. Langkah selanjutnya ialah uji coba dan evaluasi program oleh pengguna.



**Gambar 1.2 Paradigma *Prototyping***

Dengan *model prototype* ini perancang dan pengguna bertemu untuk mendefinisikan secara *obyektif* keseluruhan perangkat lunak, mengidentifikasi kebutuhan yang diketahui, dan area lebih besar dimana definisi lebih jauh merupakan keharusan kemudian dilakukan perancangan kilat berupa maket atau prototipe sistem untuk kemudian dievaluasi pengguna untuk menyaring kebutuhan pengembangan perangkat lunak.

## **1.7 Sistematika Penulisan**

Untuk memberikan gambaran dan sistematika yang jelas, peneliti akan menyusun penelitian ini menjadi 5 (lima) bab dengan urutan sebagai berikut :

### **BAB I : PENDAHULUAN**

Dalam bab ini berisi uraian tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan penelitian, kerangka pemikiran, metodologi penelitian, dan sistematika penulisan.

### **BAB II : STUDI PUSTAKA**

Dalam bab ini akan diuraikan secara singkat mengenai teori, *state of the art*, landasan, paradigma, dan cara pandang serta metode-metode yang telah ada atau akan digunakan dalam penyelesaian laporan pembuatan sistem aplikasi, perangkat keras, dan perangkat lunak yang dibangun.

### **BAB III : ANALISIS DAN PERANCANGAN SISTEM**

Bab ini mengungkapkan permasalahan lebih khusus guna mencari alternatif pemecahan masalah serta rancangan suatu pemecahan masalah yang mungkin dilakukan.



#### BAB IV : IMPLEMENTASI SISTEM

Bab ini memuat implementasi dari perancangan yang telah dibuat dan pembahasannya. Bab ini juga mencakup gambar tampilan dari program serta modul program yang mendukung.

#### BAB V : PENUTUP

Dalam bab ini berisi kesimpulan yang merupakan rangkuman keseluruhan isi yang sudah dibahas serta saran seputar perluasan, pengembangan, pendalaman, dan pengkajian ulang.

