

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang Masalah**

Analisis sentimen merupakan bidang studi yang menanalisis pendapat, sentiment, penilaian, evaluasi, sikap, dan emosi seseorang terkait suatu topik, layanan, produk, individu maupun organisasi tertentu. Analisis sentimen dilakukan untuk menentukan apakah opini atau komentar terhadap suatu topik memiliki kecenderungan positif atau negative dan dapat dijadikan sebagai acuan dalam meningkatkan suatu pelayanan, ataupun meningkatkan kualitas produk [1].

Pada saat melakukan proses analisis sentimen kita akan menemukan beberapa kalimat sarkasme, Sarkasme adalah jenis sentimen dimana seseorang mengungkapkan emosi atau tanggapan negatif menggunakan kata-kata yang positif. Penyampaian sarkasme dalam bentuk teks seringkali kompleks sehingga bisa saja menyebabkan pembaca salah memahami polaritas dari suatu teks[2]. algoritma untuk pendeteksiian sarkasme menggunakan random forest classifier.

Random forest merupakan salah satu metode yang digunakan untuk klasifikasi dan regresi. Metode ini merupakan sebuah ensemble (kumpulan) metode pembelajaran menggunakan pohon keputusan sebagai base classifier yang dibangun dan dikombinasikan[3].

Penelitian sebelumnya mengenai analisis sentimen telah banyak dilakukan, diantaranya adalah penelitian yang dilakukan oleh I. Olive, D. Putra [4] yang berjudul “Implementasi Text Mining Untuk Analisis Opini Masyarakat Terhadap Kinerja Layanan Transportasi Online Dengan Analisis Faktor” pada penelitian ini tidak berhenti sampai klasifikasi saja tetapi dilakukan analisis faktor, dan dapat menghasilkan akurasi rata-rata yaitu 71.59%. Selanjutnya, terdapat penelitian yang dilakukan oleh B. Gunawan [5] yang berjudul “Sistem

Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes” hasil penelitian ini menghasilkan akurasi tertinggi yaitu 77.78%.

Pada penelitian yang dilakukan oleh S. Fanissa [6] berjudul “Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking” memiliki akurasi tertinggi sebesar 86.6% yang didapat dari seleksi fitur 75%, tetapi terdapat saran dari penulis untuk penyempurnaan penelitiannya disarankan untuk memperhatikan kalimat sarkastik supaya hasil klasifikasinya lebih sempurna. Dan pada penelitian yang dilakukan oleh D. Alita [7] yang berjudul “Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier” melakukan analisis sentimen dengan menggunakan pendeteksian sarkasme dapat meningkatkan nilai rata-rata dari akurasi, presisi, recall dan f1score dengan kenaikan nilai rata-rata akurasi sebesar 16,61 %, kenaikan nilai presisi sebesar 5,45 %, kenaikan nilai recall sebesar 9,64% dan kenaikan nilai F1score sebesar 11,27%.

Berdasarkan penelitian sebelumnya yang sudah diuraikan dan permasalahan yang terjadi pada penelitian sebelumnya, maka pada penelitian ini akan dilakukan pengaruh pendeteksian sarkasme pada analisis sentimen masyarakat terhadap politikus Indonesia menggunakan random forest classifier.

## **1.2 Identifikasi Masalah**

Berdasarkan latar belakang yang telah dipaparkan maka dapat diidentifikasi masalah yaitu kalimat sarkasme yang menggambarkan situasi sebaliknya yang dapat mempengaruhi penentuan polaritas pada analisis sentimen.

## **1.3 Maksud dan Tujuan**

Maksud dari penelitian ini adalah untuk mengimplementasikan metode Naive Bayes dalam melakukan analisis sentimen dan metode random forest untuk pendeteksian sarkasme.

Adapun Tujuan pada penelitian ini sebagai berikut :

1. Mengukur kinerja algoritma klasifikasi Naïve Bayes pada kasus pengaruh pendeteksian sarkasme pada analisis sentimen masyarakat terhadap politikus Indonesia menggunakan random forest classifier.
2. Mengukur kinerja algoritma pendeteksian sarkasme yaitu Random Forest Classifier pada kasus pengaruh pendeteksian sarkasme pada analisis sentimen masyarakat terhadap politikus Indonesia menggunakan random forest classifier.

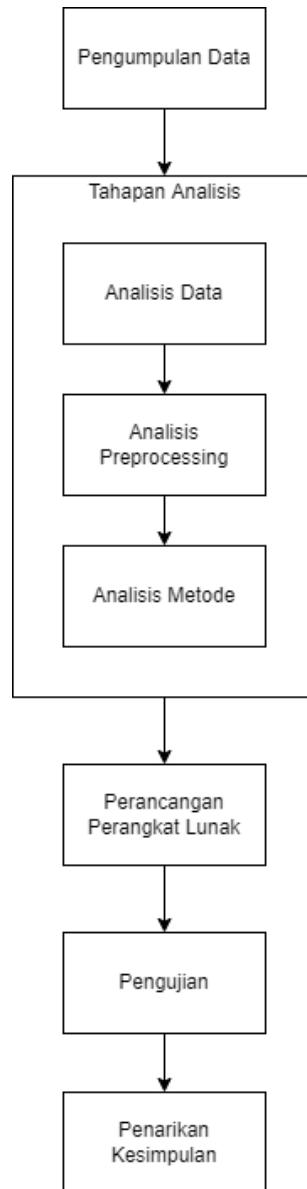
#### **1.4 Batasan Masalah**

Agar penelitian yang dilakukan lebih terarah, maka diperlukan suatu pembatasan masalah atau ruang lingkup kajian sebagai berikut :

1. Metode klasifikasi data menggunakan algoritma Naïve Bayes.
2. Kelas untuk pelabelan pada analisis sentimen dibagi menjadi dua kelas yaitu positif dan negatif.
3. Data yang diambil berupa kalimat berbahasa Indonesia.
4. Sumber Data diambil dari twitter menggunakan twitter api dengan tweet yang mengandung kata atau kalimat para politikus di Indonesia dengan pengambilan 7 tokoh politik.
5. Format file untuk data latih dan data uji adalah .csv.
6. Sumber data diambil pada bulan Januari-Mei 2022.
7. Pembobotan kata menggunakan TF untuk klasifikasi sentimen dan TF-IDF untuk deteksi sarkas.

## 1.5 Metode Penelitian

Berikut alur penelitian yang akan dilakukan seperti pada Gambar 1.1.



Gambar 1.1 Metode Penelitian

### 1.5.1 Metode Pengumpulan Data

Metode pengumpulan data yang digunakan pada penelitian ini adalah sebagai berikut :

1. Studi Literatur, yaitu Teknik pengumpulan data dengan cara mengumpulkan literatur, jurnal, paper, dan buku yang berkaitan dengan analisis sentimen.

2. Pengumpulan *Dataset*, *dataset* yang digunakan pada penelitian ini berupa opini masyarakat terhadap politikus di Indonesia yang diambil dari twitter menggunakan api Tweepy dan Snsrape dengan Bahasa pemrograman python dengan jumlah 7 tokoh politik di indonesia selama bulan Januari-Mei 2022.

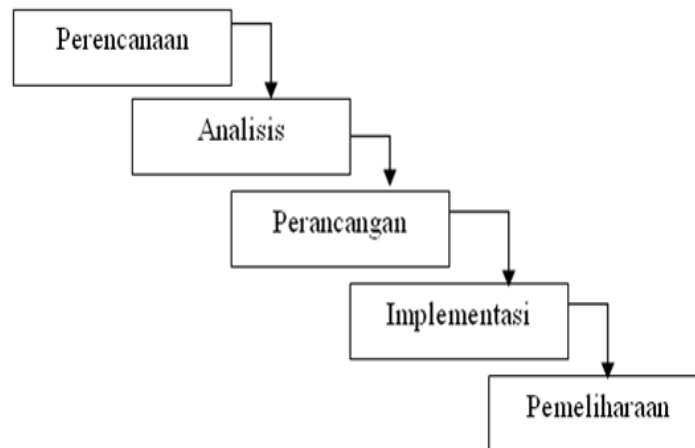
### 1.5.2 Metode Tahapan Analisis

Berdasarkan pengkajian dan evaluasi setelah membaca beberapa sumber seperti jurnal, literatur, Tahapan analisis pada penelitian ini adalah:

1. Analisis Data, merupakan analisis yang dilakukan terhadap pola kalimat dari data yang didapat, data yang dianalisis adalah *tweet* dari *twitter* yang mengandung *keyword* politikus di Indonesia.
2. Analisis *Preprocessing*, *preprocessing* berfungsi untuk mengubah data teks yang tidak terstruktur atau sembarang menjadi data yang terstruktur. Proses yang dilakukan dalam tahapan *preprocessing* adalah sebagai berikut ; *Cleaning*, *Spelling Normalization*, *Case Folding*, *Tokenizing*, *Filtering*, *Convert Negation*, *Stemming*. Selanjutnya dilakukan Pembobotan teks dilakukan menggunakan TF dan TF-IDF
3. Analisis Metode, Pada tahap ini akan dilakukan klasifikasi sentimen menggunakan metode Naïve Bayes untuk menghasilkan kelas positif dan negative, lalu kelas dengan label positif di klasifikasikan kembali kedalam kelas sarkasme atau tidak sarkasme. Jika masuk kedalam kelas sarkasme ubah hasil akhir klasifikasi kedalam kelas negatif, jika masuk kedalam kelas tidak sarkasme hasil akhir klasifikasi tidak berubah.

### 1.5.3 Metode Pembangunan Perangkat Lunak

Metode pembangunan perangkat lunak yang digunakan dalam penelitian ini adalah model *waterfall*. Tahapan-tahapan pada metode *Waterfall* sebagai berikut [8]:



Gambar 1.2 Metode Pembangunan Perangkat Lunak

#### 1. Perencanaan

Pada tahap ini dibuat rencana tentang kebutuhan dari perangkat lunak yang akan dibuat agar perangkat lunak yang dibuat sesuai dengan kebutuhan pengguna.

#### 2. Analisis

Pada tahap dilakukan analisa kebutuhan apa saja yang diperlukan dalam pembangunan aplikasi secara lengkap berdasarkan masalah yang telah disampaikan.

#### 3. Perancangan

Pada tahap ini setelah selesai menganalisa kebutuhan secara lengkap, pengembang melakukan pembuatan desain sistem secara keseluruhan dan menentukan alur dari perangkat lunak yang akan dibangun. Desain yang dibangun berupa alur data dan juga tampilan antar muka.

#### 4. Implementasi

Pada tahapan ini semua data informasi yang telah dikumpulkan dan didesain oleh pengembang akan diterjemahkan kedalam bahasa pemrograman dan di integrasikan sehingga nantinya akan terbentuk aplikasi.

#### 5. Pemeliharaan

Proses ini merupakan tahapan perbaikan apabila terdapat kesalahan

yang ditemukan pada aplikasi setelah digunakan oleh user.

#### **1.5.4 Metode Pengujian Akurasi**

Dalam tahapan ini akan dilakukan pengujian akurasi metode menggunakan confusion matrix dan untuk validasi menggunakan k-fold cross validation pada hasil sentimen dengan klasifikasi menggunakan metode naïve bayes dan pendeteksian sarkasme menggunakan random forest classifier pada analisis sentimen masyarakat terhadap politikus Indonesia.

#### **1.5.5 Metode Penarikan Kesimpulan**

Pada tahap ini akan dilakukan penarikan kesimpulan berdasarkan hasil dari tahap pengujian dalam penerapan metode naïve bayes dan random forest classifier pada analisis sentimen dan pendeteksian sarkasme.

### **1.6 Sistematika Penulisan**

Sistematika penulisan disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan penelitian ini adalah sebagai berikut :

#### **BAB 1 PENDAHULUAN**

Pada Bab 1 akan dijelaskan mengenai latar belakang masalah, identifikasi masalah, maksud dan tujuan, manfaat, batasan masalah, metodologi penelitian dan sistematika penulisan.

#### **BAB 2 TINJAUAN PUSTAKA**

Pada Bab 2 akan dijelaskan tentang pembahasan berbagai konsep dasar mengenai analisis sentimen, *Naïve Bayes*, Sarkasme, *Random Forest Classifier*, twitter, metode pengujian k-fold cross validation dan teori-teori pendukung lainnya yang berkaitan dengan topik.

#### **BAB 3 ANALISIS DAN PERANCANGAN**

Pada Bab 3 ini berisi tentang analisis dan perancangan aplikasi yang dibangun meliputi analisis masalah, analisis data, analisis sistem, analisis prerprocessing, analisis algoritma Naïve Bayes, analisis algoritma Random Forest, analisis non fungsional dan fungsional.

#### **BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM**

Pada Bab 4 ini berisi tentang gambaran dari hasil implementasi dan pengujian pada sistem yang mulai pada proses pengklasifikasian dengan metode Naive Bayes, proses deteksi sarkasme dengan metode Random Forest dan melakukan tahapan pengujian dengan metode pengujian yang akan menampilkan hasil keluaran seperti nilai recall, precision dan f1-score pada sistem aplikasi yang telah deprogram untuk menguji akurasi dari analisis sentimen dan pendeteksian sarkasme.

#### **BAB 5 KESIMPULAN DAN SARAN**

Pada Bab 5 akan dijelaskan mengenai kesimpulan yang sudah diperoleh dari hasil penulisan tugas akhir dan saran agar dapat dilakukan pengembangan untuk penelitian selanjutnya.