BAB 1
PENDAHULUAN

1.1 Latar Belakang

Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif. Salah satu contoh penggunaan analisis sentimen dalam dunia nyata adalah identifikasi opini publik terhadap kandidat capres. Akibatnya timbul kebutuhan untuk mengatur dan mengelompokkan opini berdasarkan sentimennya agar seseorang dapat mencari dan menilai kualitas kandidat capres dari opini publik, akan tetapi jika pengelompokkan berita dilakukan secara manual memakan waktu yang lama dan mahal maka dari itu diperlukan klasifikasi dokumen.

Klasifikasi dokumen secara otomatis sekiranya diperlukan untuk mengurangi biaya dan mempercepat pengaturan informasi. Salah satu metode yang dapat digunakan dalam pengklasifikasian yaitu *Naive Bayes*. *Naive Bayes* merupakan metode klasifikasi dengan tingkat keakuratan tinggi dan penghitungan yang sederhana. Metode ini mengasumsikan kemunculan suatu kata tidak mempengaruhi kemunculan kata yang lain. Walaupun asumsi ini melanggar aturan disetiap bahasa, asumsi ini tidak berpengaruh besar terhadap tingkat keakuratan metode ini [1].

Pada metode *Naive Bayes* yang menjadi masalah adalah keberadaan kata berperan penting pada proses klasifikasi. Salah satu faktor yang mempengaruhi keberadaan kata yaitu *stemming*. Dengan *stemming* bentuk kata baku yang beragam akan menjadi seragam (kata dasar). Namun faktanya pada *posting twitter* ditemukan banyak kata-kata yang tidak baku seperti kata “tidak” yang mempunyai variasi “tdk”, “gak”, “ngga”, “kagak”, “ga”, “nda”, “gag” yang tidak dapat ditangani oleh praproses *Stopwords*. Keberadaan kata yang tidak baku akan mempengaruhi akurasi hasil perhitungan *Naive Bayes*. Oleh karena itu diperlukan modifikasi *stemming* untuk menangani kata yang tidak baku.

Berdasarkan masalah di atas maka dibutuhkan optimasi *preprocessing text* yaitu *stemming* dalam analisis sentimen pada *posting twitter*.

1.2 Rumusan Masalah

Bagaimana penerapan *preprocessing* teks yaitu *stemming* sebelum melakukan klasifikasi pada metode *Naïve Bayes* dilakukan untuk menghilangkan variasi kata yang tidak baku dan menggantinya dengan kata yang baku.

1.3 Maksud dan tujuan

Berdasarkan masalah yang diteliti maka maksud dari penelitian tugas akhir ini adalah mengoptimasi analisis sentimen pada *posting twitter* menggunakan metode *Naive Bayes.*

Sedangkan tujuan yang akan dicapai dalam penelitian ini adalah untuk mengetahui pengaruh, performansi yang dihasilkan dengan melihat persentase akurasi dalam implementasi metode *Naive bayes* pada sentimen analisis dengan melakukan *stemming*.

1.4 Batasan masalah

Penelitian yang dilakukan untuk implementasikan *Naive Bayes* maka memiliki batasan masalah sebagai berikut :

1. Studi kasus yang digunakan kandidat Capres 2014 yaitu Jokowi dan Prabowo.
2. Klasifikasi sentimen yang digunakan 3 kategori positif, negatif dan non opini.
3. Informasi yang didapatkan adalah jumlah sentimen dari masing-masing kelas, akurasi dan kecepatan proses.
4. *Tweets* yang digunakan hanya berbahasa indonesia.
5. Metode *Stemming* menggunakan Algoritma Nazief dan Adriani
6. Evaluasi kinerja menggunakan *Accuracy and Error rate*
7. Variasi kata tidak baku yang ada pada kamus sama dengan variasi kata tidak baku pada data uji dan data pelatihan

1.5 Metodologi penelitian

Metodologi penelitian yang digunakan pada penelitian ini yaitu metodologi penelitian untuk mengumpulkan dasar dan data penelitian yaitu berupa ekstraksi informasi dari *tweets*. Metodologi yang digunakan untuk mengumpulkan dasar dan data penelitian, yaitu :

1. Studi masalah dan solusi

   Melakukan eksplorasi terhadap masalah yang akan dilakukan observasi dan menentukan solusi yang mungkin. Menentukan hipotesa yang mungkin dihasilkan dari solusi yang diambil terhadap masalah yang ada.Sehingga masalah tersebut dapat diselesaikan dan mendapatkan detail hasil eksplorasi dengan melakukan eksperimen dari penerapan solusi tersebut terhadap masalah yang diambil dan mendapatkan hasil secara kuantitatif dari efisiensi solusi hingga efektifitas yang dihasilkan dari penerapan solusi tersebut.

2. Pengumpulan data

   Pengumpulan data yang akan digunakan dalam penelitian ini adalah dengan mengambil *tweets* masyarakat mengenai salah satu capres 2014 secara random dari *twitter*.

3. *Learning Modeling*

   Membangun *learning modeling* yaitu model probabilistik yang dihasilkan dari data pelatihan sebagai hipotesis awal.

4. *Eksperiment*

   Merupakan tahap eksplorasi terhadap objek penelitian yang berupa teks, dengan melakukan *processing* terhadap teks tersebut dan menerapakan metode *learning* yang akan diuji coba. Eksplorasi dilakukan dengan membangun perangkat lunak yang mengimplementasikan metoda dibuat. Mengambil hasil dari uji coba tersebut dan melakukan analisis terhadap hasil uji tersebut. Dengan melakukan implementasi program yang dibangun untuk menerapkan teori yang ada pada penelitian ini.

1.6 Sistematika penulisan

Sistematika penulisan tugas akhir ini disusun untuk memberikan gambaran umum tentang permasalahan dan pemecahannya. Sistematika penulisan tugas akhir ini adalah sebagai berikut :

BAB 1 PENDAHULUAN

Bab ini berisi tentang latar belakang masalah, identifikasi masalah, maksud dan tujuan, batasan masalah dan metode penelitian yang dipergunakan dalam penelitian mengimplementasikan analisis sentimen pada *posting twitter* menggunakan metode *Naive Bayes*.

BAB II TINJAUAN PUSTAKA

Menguraikan tinjauan terhadap berbagai literatur pendukung yang digunakan untuk mengoptimasi analisis sentimen pada posting *twitter* menggunakan Metode *Naive Bayes* dan *Stemming*.

BAB III ANALISIS METODE

Pada bagian ini dijelaskan secara rinci mengenai tahapan pengerjaan dengan menggunakan metode yang digunakan penulis dalam mengerjakan penelitian sesuai dengan pendekata analisis orientasi berbasis objek, yaitu dengan analisis kebutuhan fungsionalitas atau *Use case, Class diagram, Sequence Diagram*. Selain itu terdapat juga perancangan untuk sistem yang akan dibangun sesuai dengan hasil analisis dan antarmuka untuk sistem yang akan dibangun.

BAB IV PENGUJIAN

Pada bagian dilakukan pengujian terhadap prototipe dari program dengan menggunakan data uji yang belum diketahui kategorinya agar diketahui apakah sudah optimal akurasi yang diteliti.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan yang diperoleh dari tujuan dan permasalahan yang telah terselesaikan.