

# 1. Pendahuluan

## 1.1. Latar Belakang

Setiap perusahaan yang bergerak di bidang jasa dan produk tentunya memerlukan penilaian berbentuk opini dari klien kepada perusahaan tersebut terhadap pelayanan (*service*) dan kualitas produknya. Opini yang diberikan klien terhadap perusahaan dapat digolongkan menjadi opini positif dan opini negatif yang sangat berpengaruh terhadap kemajuan kualitas perusahaan tersebut. Produk dapat dinilai berdasarkan letak lokasi yang digunakan sebagai pelayanan (*service*) dan kualitas produk tersebut. Maka dari itu dimungkinkan akan terdapat opini yang berbeda dari klien di kota A, kota B, dan bahkan kota C.

Ringkasan multi-opini bertujuan menyimpulkan seluruh opini yang dimaksud klien terhadap penilaian suatu produk. Ringkasan ini juga didasarkan pada lokasi perusahaan menjalankan bisnisnya. Output dari ringkasan ini akan menghasilkan ringkasan yang berguna untuk pengambilan keputusan yang akan dilakukan oleh calon klien dan perusahaan atas produk tersebut.

Beberapa penelitian telah dilakukan untuk melakukan klasifikasi opini terhadap review suatu produk oleh klien [1] [2] [3] [4]. Metode klasifikasi yang paling sering digunakan adalah Naïve Bayes dan Support Vector Machine (SVM). Naïve Bayes dan SVM masing-masing memiliki kelebihan yaitu kehandalan untuk menangani masalah dataset yang berdimensi besar dan tingkat akurasi prediksi hasil yang memuaskan [5]. Namun keduanya belum tangguh terhadap data latih (*training*) yang memiliki banyak *noise* [6]. Untuk itu, penulis mengusulkan metode *k-Nearest Neighbor* (k-NN) berbasis *Cosine Similarity* yang tangguh terhadap data yang mengalami banyak *noise* [6].

Penulis menerapkan pendekatan *Natural Language Processing* (NLP) dalam melakukan *pre-processing*. *Natural Language Processing* (NLP) atau pengolahan bahasa alami merupakan salah satu bidang ilmu *Artificial Intelligence* (Kecerdasan Buatan) yang mempelajari komunikasi antara manusia dengan komputer melalui bahasa alami. NLP tidak bertujuan untuk mentransformasikan bahasa yang diterima dalam bentuk suara menjadi data digital dan/atau sebaliknya pula; melainkan bertujuan untuk memahami arti dari ucapan yang diberikan dalam bahasa alami dan memberikan respon yang sesuai, misalnya dengan melakukan suatu aksi tertentu atau menampilkan data tertentu. Tahapan *pre-processing* tersebut antara lain mendeteksi kesesuaian opini dengan studi kasus yang dibutuhkan, mendeteksi opini apakah mengandung kalimat majemuk atau tidak, melakukan *tokenization*, *stopword removal* dan *stemming*, melakukan klasifikasi opini menggunakan k-NN, menyusun kembali data yang sudah terklasifikasi, dan melakukan ringkasan (*summarization*) berdasarkan opini positif dan opini negatif dengan dataset Go-Jek karena merupakan perusahaan yang sedang terkenal pada

saat ini. Data yang didapat diperoleh dengan teknik *web crawling* pada jejaring sosial Twitter.

## 1.2. Perumusan Masalah

Rumusan masalah yang menjadi acuan pelaksanaan tugas akhir ini adalah:

1. Bagaimana hasil penerapan, analisis, dan implementasi *multiple opinion summarization* dengan pendekatan *Natural Language Processing* (NLP)?
2. Bagaimana performansi k-Nearest Neighbor (k-NN) dalam mengklasifikasikan sentimen pada data teks?

## 1.3. Tujuan

Tujuan dari pelaksanaan tugas akhir ini adalah:

1. Mengetahui hasil penerapan, analisis, dan implementasi *multiple opinion summarization* dengan pendekatan *Natural Language Processing* (NLP).
2. Mengetahui performansi k-Nearest Neighbor (k-NN) dalam mengklasifikasikan sentimen pada data teks.

## 1.4. Batasan Masalah

Adapun batasan masalah pada pelaksanaan penelitian tugas akhir ini adalah:

1. Implementasi dilakukan menggunakan bahasa pemrograman Python.
2. Data berupa file .xlsx berbahasa Indonesia yang berisi kalimat opini.
3. Klasifikasi dilakukan berdasarkan aspek opini positif, negatif, dan non-opini.
4. Ringkasan multi-opini dilakukan secara *offline*.
5. Review dilakukan pada perusahaan Go-Jek dengan teknik mengambil dataset secara *web crawling*.
6. Lokasi data yang menjadi dataset opini terhadap pelayanan Go-Jek hanya tiga lokasi, yaitu Jakarta, Bandung, dan Bali.
7. Hasil evaluasi sistem menggunakan tools ROUGE 2.0.
8. Dataset opini diperoleh mulai tanggal 26 Januari 2016 sampai dengan 1 Agustus 2016.

## 1.5. Metodologi Penyelesaian Masalah

Metodologi yang digunakan pada pelaksanaan tugas akhir ini adalah:

1. Studi Literatur

Mempelajari literatur yang relevan dengan topik *opinion summarization* guna memahami dan menganalisis langkah kerja sistem dalam meringkas opini, *preprocessing* data, klasifikasi menggunakan metode *k-Nearest Neighbor* (k-NN) berbasis *Cosine Similarity*, ekstraksi *multiple opinion summarization* menggunakan metode *Natural Language Processing* (NLP) serta kemungkinan bentuk implementasinya.

2. Pengumpulan Data  
Pengumpulan dataset yang akan digunakan dalam tugas akhir ini dengan melakukan teknik *Web Crawling* yang akan mengambil data sesuai dengan keyword yang penulis masukan. Penulis mengambil data tersebut sesuai dengan berbagai macam lokasi para klien dalam melakukan opini terhadap jasa Go-Jek.
3. Analisis Awal dan Perancangan Sistem  
Analisis dilakukan guna mengetahui kebutuhan sistem. Setelah itu, akan dilakukan perancangan sistem sesuai dengan kebutuhan dan masalah yang ada. Sistem yang akan dirancang sesuai dengan pendekatan *Natural Language Processing (NLP)* dan teknik *summarization multiple opinion*.
4. Implementasi  
Melakukan implementasi terhadap rancangan sistem yang telah dibuat sebelumnya. Implementasi dilakukan dengan menggunakan Bahasa pemrograman Python 2.7.
5. Pengujian dan Analisis Hasil  
Sistem yang sudah dirancang akan diimplementasikan kemudian dilakukan pengujian (*testing*). Sistem diuji dengan data uji yang telah dikumpulkan. Setelah itu, sistem akan dianalisis hasil performansinya berupa akurasi dari sistem.
6. Penarikan Kesimpulan  
Pada tahap ini, dilakukan analisis terhadap hasil pengujian sehingga dapat ditarik kesimpulan berdasarkan pengujian dari implementasi sistem yang telah dibuat.
7. Penyusunan Laporan Tugas Akhir  
Penyusunan laporan tugas akhir dilakukan dengan menuliskan setiap hal penting dalam pelaksanaan tugas akhir ini dan menjabarkan semua dengan detail penelitian yang dilakukan dan kesimpulan dari hasil analisa akhir.

## **1.6. Sistematika Penulisan**

Buku laporan tugas akhir ini ditulis dengan sistematika sebagai berikut:

### **1. Pendahuluan**

Menjelaskan latar belakang, perumusan masalah, batasan masalah, tujuan, metodologi penelitian, dan sistematika penulisan.

## **2. Landasan Teori**

Menjelaskan dasar teori yang diambil dari beberapa kutipan buku, konsep dasar sistem, dan definisi lainnya yang berkaitan dengan permasalahan dan pembahasan dari analisis sentimen dan ringkasan multi-opini dalam mendukung penyelesaian tugas akhir ini.

## **3. Perancangan Sistem**

Menjelaskan perancangan dari sistem yang dibangun, dimulai dari preprocessing menggunakan pendekatan Natural Language Preprocessing (NLP), klasifikasi dengan k-NN hingga ringkasan multi-opini dan rencana pengujian sistem yang dilakukan.

## **4. Analisis Hasil Pengujian**

Membahas hasil pengujian sistem yang telah dilakukan dengan menganalisis skenario pengujian perubahan nilai parameter yang mempengaruhi performansi sistem yang dilakukan dalam mencapai tujuan akhir yang diinginkan pada tugas akhir ini.

## **5. Kesimpulan dan Saran**

Membahas kesimpulan akhir yang didapat dari pengerjaan tugas akhir, serta saran yang diusulkan dalam mendukung pengembangan yang lebih lanjut.