# Laporan Tugas Besar

# NLP

# Abstrak

# Bab I : Pendahuluan

## Latar Belakang

Pada masa ini, aplikasi atau halaman web penyedia informasi sudah semakin berkembang dan banyak digunakan. Contoh informasi yang disediakan adalah informasi tempat wisata, hotel, harga tiket pesawat atau kereta dan lainnya. Penyedia informasi tersebut juga menambahkan kolom komentar atau *review* yang dapat diisi oleh setiap orang yang seharusnya sudah pernah memiliki pengalaman menggunakan barang atau jasa dari informasi yang disediakan. Komentar atau *review* tersebut juga dilengkapi dengan *rating* yang dapat mengelompokkan komentar – komentar dan penilaian dengan angka dari suatu produk atau jasa. Namun, terkadang ada saja komentar yang diberikan tidak sesuai dengan *rating* yang diberikan. Contohnya, komentar terhadap suatu hotel sudah sangat bagus dari segi pelayanan, kebersihan dan ruangan namun nilai yang diberikan hanya 3 dari 5, yang seharusnya bisa mencapai nilai 4 ataupun 5.

Oleh karena itu, sistem klasifikasi sentimen dari suatu komentar dapat berguna pada masalah ini. Dari komentar yang diberikan akan langsung ditentukan apakah sentimen atau nilai dari komentar tersebut. Pada kasus ini, kami mengambil contoh komentar pada halaman web atau aplikasi Zomato (penyedia informasi restoran – restoran). Dari komentar tersebut akan diklasifikasikan kelas sentimen positif (1) atau negatif (0). Sistem klasifikasi dibuat dengan bantuan penerapan ilmu *natural language processing* dan akan menghasilkan sebuah model yang paling cocok dengan kasus sentimen komentar.

## Rumusan Masalah

Aaa

## Tujuan

Penelitian ini memiliki tujuan untuk dapat membangun model *machine learning* yang cocok digunakan untuk klasifikasi sentimen pada komentar yang ada di halaman web Zomato berbahasa Indonesia.

# Bab II : Data

Pada penelitian ini, data diambil dengan melakukan *scrapping* pada halaman web Zomato dengan mengambil komentar – komentar pada setiap *review* restoran. Data yang diambil berjumlah 200 data dengan masing – masing kelas adalah 100 data. Dengan itu, digunakan 150 data sebagai data latih dan 50 data sebagai data validasi.

Data yang diambil berupa kalimat – kalimat dengan bahasa Indonesia yang semi-formal. Terdapat komentar yang menggunakan bahasa Indonesia yang formal (sesuai EYD) dan non-formal (menggunakan istilah sehari – hari). Pada kalimat yang menggunakan bahasa non-formal ataupun terdapat singkatan dapat mempersulit proses klasifikasi karena bisa saja kata yang sebenarnya memiliki arti sama dianggap berbeda, contohnya pada kalimat pertama terdapat kata “juga” dan kalimat kedua terdapat kata “jg”. Selain itu, pemilihan kata yang digunakan banyak terdapat pengulangan huruf dan angka, seperti “cekernyaaa pedes tapi enakkkkk. tempatnya kurang gede tapi nyaman. harganya ga terlalu mahal dan porsinya banyakkk. menu nya juga ga cuma ceker tapi ada yang lain juga seperti wings, dll” dan “Tempatnya unik dan kebetulan waktu pas dtg kesini lagi agak sepi, jadi enak2 aja lama2 disini. Porsi makanannya ngenyangin juga. Buat yang suka pasta, bisa mampir kesini.”.

Data dengan kelas negatif dan positif juga relatif berbeda. Pada data yang didapatkan, rata – rata jumlah kata untuk kelas negatif adalah 81,59 dan untuk kelas positif adalah 66,25. Hal tersebut karena, biasanya konsumen yang merasakan pengalaman yang buruk atau memberikan *review* yang buruk menggunakan kata yang lebih banyak daripada *review* positif untuk meluapkan pengalamannya.

# Bab III : Analisa dan Eksperimen

Penelitian ini menggunakan beberapa metode untuk mendapatkan fitur dan klasifikasi. Metode ekstraksi fitur yang digunakan adalah *CountVectorizer* dan *TF-IDF*. Dan metode klasifikasi yang digunakan adalah *Multinomial Naivè Bayes* (NB)dan *Support Vector Machine* (SVM). Selain itu, dilakukan *pre-processing* sebelum data diambil fiturnya untuk di klasifikasi menggunakan *classifier*.

*Pre-processing* yang digunakan pada data ini adalah menghapus tanda baca, mengubah semua huruf menjadi huruf kecil