

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Seiring dengan berkembangnya zaman, teknologi baru semakin banyak bermunculan. Banyaknya teknologi baru tersebut mengakibatkan jumlah data yang dihasilkan semakin meningkat. Pada tahun 2003, jumlah data yang dihasilkan mencapai angka 5 miliar *gigabytes*. Jika data tersebut disimpan dalam bentuk *Compact Disc*, dapat dipastikan akan memenuhi lapangan sepak bola. Jumlah data tersebut, pada tahun 2011 meningkat setiap dua hari, sedangkan pada tahun 2013 jumlah tersebut terus meningkat setiap dua menit. Dapat dilihat dari rangkuman jumlah data tersebut, angka ini nantinya akan mengalami pertumbuhan yang sangat pesat. Munculnya media sosial bernama twitter pada tahun 2006 juga ikut menyumbang peningkatan jumlah data yang ada. Terhitung pada tahun 2019 total pengguna twitter mencapai 330 juta (*• Twitter by the Numbers (2020): Stats, Demographics & Fun Facts, n.d.*).

Penggunaan media sosial twitter ini dapat menjadikan penggunaanya berbagi tentang keseharian ataupun informasi dengan rekan atau kerabat secara realtime. Informasi yang dibagikan pada twitter disebut dengan kicauan (*tweet*) yang terdiri dari 140 karakter (Kwak & Moon, 2010) saat ini sudah maksimum 280 karakter (*Twitter 280 Karakter Resmi di Seluruh Dunia, n.d.*). Jumlah pengguna twitter yang sangat tinggi dapat menjadi sarana bagi masyarakat untuk melakukan jual beli (promosi), dapat juga untuk menuliskan perasaan apa yang dirasakan, produk apa saja yang sedang digunakan dan juga referensi tempat makan atau kuliner di Indonesia.

Dimana bisnis kuliner itu sendiri merupakan bisnis jasa yang mudah dilakukan. Untuk memulai bisnis kuliner itu sendiri, tidak perlu mempunyai modal besar dan memiliki kemampuan untuk memasak. Saat ini banyak sekali pebisnis yang memulai bisnis kuliner dari menjual berbagai masakan yang praktis, salah satunya jenis makanan *frozen*.

Makanan *frozen* ini dinilai sangat praktis dan cepat serta tidak mengubah cita rasa dari makanan tersebut.

Seiring berkembangnya teknologi, referensi kuliner dapat digunakan sebagai sarana untuk memilih tempat makan mana yang akan dikunjungi serta dapat memberikan penilaian terhadap tempat makan itu sendiri dari segi rasa hingga kisaran harga kuliner. Hal ini kemudian dimanfaatkan oleh calon pebisnis dalam mengetahui minat masyarakat saat ini. Tidak hanya bagi calon pebisnis, pebisnis lama pun dapat memanfaatkan untuk memperbarui menu yang sudah mereka jual sebelumnya, dikarenakan minat masyarakat yang sering berubah-ubah. Permasalahan tersebut dapat diketahui lewat analisis sentimen. Peran analisis sentimen disini sangatlah penting. Analisis sentimen merupakan penelitian yang digunakan untuk menghitung polaritas opini yang terbagi dalam positif dan negatif yang mana berasal dari dokumen teks (Perdana & Pinandito, 2017).

Terdapat proses pengambilan informasi yang berasal dari data tekstual, disebut dengan *text mining*. Dalam *text mining* dapat juga diketahui, permasalahan apa saja yang ada di dalam sebuah teks. Nantinya analisis sentimen juga membutuhkan *text mining*, yang membantu proses identifikasi emosional didalamnya (Zhang, et al, 2011). *Opinion mining* adalah istilah lain dari analisis sentimen, yang maknanya adalah opini yang ditampilkan di sebuah teks (Liu, B., 2012).

Saat ini *text mining* dan *Natural Language Processing* banyak dimanfaatkan dengan menggabungkan keduanya untuk mencapai berbagai tujuan. Salah satu hasil dari pemanfaatan keduanya adalah analisis sentimen pada sosial media. Analisis sentimen disini nantinya akan dilakukan pada banyak data atau sekumpulan data yang mana bentuknya adalah teks guna mengidentifikasikan berbagai macam hal, yaitu emosi dan opini pada suatu topik. Hasil dari analisis sentimen tersebut akan dapat memberikan keuntungan untuk organisasi atau yang sedang menggeluti bidang tersebut. Pada hakikatnya *text mining* dan *Natural Language Processing* adalah teknologi *Artificial Intelligence* (AI) yang

pemanfaatannya adalah mengubah teks dokumen menjadi data kuantitatif. Pada proses mengubah teks dokumen tersebut tidak membutuhkan waktu yang lama. Data kuantitatif yang dihasilkan akan dimanfaatkan sesuai kebutuhan pengguna.

Bayangkan jika perusahaan atau organisasi pada bisnis kuliner memiliki banyak cabang, dan harus menemukan informasi dari sejumlah data yang besar. Jika hal tersebut dilakukan oleh karyawan atau secara manual maka hal itu sangat membuang waktu dan tenaga. Informasi yang dikumpulkan oleh karyawan tersebut juga belum tentu 100% akurat. Maka dari itu proses tersebut dapat dilakukan dengan bantuan *keyword search*, dimana karyawan tersebut dapat menemukan informasi sesuai dengan *keyword* yang sudah ditentukan. Informasi yang berhasil dikumpulkan tersebut, kemudian dilakukan analisis sentimen yang bertujuan untuk meningkatkan efisiensi analisis dokumen teks dengan memberikan kemampuan otomatisasi pada proses tersebut.

Hal penting yang sangat perlu dilakukan dalam analisis sentimen adalah klasifikasi. Metode klasifikasi yang digunakan dalam penelitian ini adalah metode Naive Bayes yang sering disebut *Naive Bayes Classifier* (NBC). NBC merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yaitu memprediksi peluang berdasarkan pengalaman di masa sebelumnya (Teorema Bayes) dengan ciri utamanya adalah asumsi yang kuat (naif) akan ketergantungan dari masing-masing kejadian.

Naive Bayes memiliki kelebihan antara lain, sederhana, cepat, dan berakurasi tinggi (Artikel, 2015). Kelebihan Naive Bayes juga dapat digunakan untuk data kuantitatif maupun kualitatif, tidak memerlukan jumlah data yang banyak, jika terdapat nilai yang hilang, maka bisa diabaikan dalam perhitungan serta dapat digunakan untuk klasifikasi masalah biner ataupun *multiclass* (Algoritma Naive Bayes | BINUS UNIVERSITY BANDUNG - Kampus Teknologi Kreatif, n.d.).

Penggunaan *Natural Language Processing* dan Naive Bayes *Classifier* (NBC) sangat berguna dalam mengolah sumber data pada penelitian ini. Penggunaan *Natural Language Processing* berguna untuk menganalisis data berbasis bahasa berjumlah banyak dan konsisten. Jika penelitian ini dilakukan tanpa menggunakan *Natural Language Processing* dapat dipastikan analisis data tidak berjalan maksimal, pasalnya analisis dilakukan dengan waktu yang lama dan hasil yang di dapat tidak konsisten. Tidak hanya *Natural Language Processing* yang dibutuhkan penggunaan Naive Bayes *Classifier* (NBC) juga sangat penting, karena Naive Bayes *Classifier* (NBC) ini memiliki tingkat akurasi yang tinggi dan lebih baik daripada *Decision Trees* dan *Neural Networks* dimana hasil akurasi tertinggi (95,18%), *recall* (95,36%), *precision* (99,43%) dan F-Score (97,36%) dengan menggunakan data dari organisasi *Apricot Training Management* (Xhemali et al., 2009).

Pada penelitian terdahulu (Statistika & Diponegoro, 2020) yang berjudul Analisis Sentimen Pemindahan Ibu Kota Negara dengan Klasifikasi Naive Bayes untuk Model Bernoulli dan Multinomial menyimpulkan bahwa klasifikasi menggunakan Bernoulli Naive Bayes menghasilkan performa dengan tingkat sensitifitas 90,19%, sementara Multinomial Naive Bayes menghasilkan performa dengan tingkat sensitifitas 93,45%. Sehingga, baik model Bernoulli maupun Multinomial memiliki performa yang baik dalam melakukan klasifikasi teks untuk kelas sentimen dengan pembagian datasetnya menggunakan data latih 80% dan data uji 20%. Sedangkan pada penelitian (Sari & Hayuningtyas, 2019) yang berjudul Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website menyimpulkan bahwa penerapan algoritma Naive Bayes untuk analisis sentimen mendapatkan hasil yang terbaik dalam mengklasifikasikan suatu ulasan informasi mengenai opini.

Dari pemaparan latar belakang dan penelitian di atas, metode Naive Bayes memiliki tingkat akurasi yang tinggi dan paling baik oleh karena itu pada skripsi ini peneliti ingin melakukan penelitian mengenai analisis sentimen pada twitter mengenai *tweet* yang berisi *hashtag* Kuliner

di Indonesia. Dimana penelitian ini akan membandingkan dua metode yaitu Multinomial Naive Bayes dan Bernoulli Naive Bayes. Sebelum melakukan pemodelan, pada penelitian ini juga melakukan tahapan pembobotan terlebih dahulu. Hal ini dilakukan sebagai variasi skenario didalam pembuatan model pada penelitian ini.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya dapat dirumuskan masalah yaitu bagaimana merancang dan membangun sebuah model untuk analisis sentimen terhadap *tweet* dengan *hashtag* Kuliner secara otomatis, serta untuk mengetahui perbandingan sentimen terhadap kuliner di Indonesia.

## 1.3 Batasan Masalah

Berdasarkan permasalahan yang telah didapatkan diatas maka penulis harus membatasi masalah agar tidak keluar dari topik yang dibicarakan. Adapun batasan masalah tersebut ialah:

1. *Tweet* yang diambil hanya *tweet* berbahasa Indonesia kemudian di terjemahkan ke dalam bahasa Inggris untuk dianalisis. *Tweet* tersebut diterjemahkan secara manual dengan *tools* pada laman web [onlinedoctranslator.com](http://onlinedoctranslator.com).
2. Metode klasifikasi yang digunakan adalah metode Multinomial Naive Bayes dan Bernoulli Naive Bayes.
3. Sentimen twitter hanya *tweet* yang memiliki *hashtag* Kuliner.
4. Media yang digunakan untuk mengambil data *tweet* pada twitter ialah dengan menggunakan library “tweepy” dengan teknik *crawling*.
5. Bahasa pemrograman yang digunakan adalah bahasa pemrograman Python versi 3.7.3.
6. *Tweet* hanya dikelompokkan kedalam 3 kelas, yaitu; netral, positif dan negatif.
7. Jenis pembagian data yang digunakan menggunakan metode *holdout*.

## 1.4 Tujuan

Tujuan dari penelitian ini adalah:

1. Untuk mengetahui tingkat sentimen netral, positif dan negatif terhadap *tweet* di twitter yang memiliki *hashtag* Kuliner secara otomatis.
2. Melakukan pengukuran akurasi serta melakukan perbandingan metode Multinomial Naive Bayes dan Bernoulli Naive Bayes pada analisis sentimen yang memiliki *hashtag* Kuliner.
3. Mengetahui referensi kuliner yang sedang ramai diperbincangkan.

## 1.5 Sistematika Penulisan

Adapun sistematika penulisan laporan yang akan dibuat adalah sebagai berikut :

### BAB I PENDAHULUAN

Bab pendahuluan mendeskripsikan mengenai latar belakang, rumusan masalah, batasan, tujuan dan sistematika penulisan.

### BAB II TINJAUAN PUSTAKA

Bab tinjauan pustaka memuat dasar teori dan penelitian terdahulu. Uraian mengenai landasan teori dan landasan empiris yang mendukung pendekatan pemecahan masalah, antara lain pengertian dan definisi dari Twitter, analisis sentimen, *machine learning*, *big data*, *text mining*, *crawling*, *text preprocessing*, *Natural Language Processing*, Naive Bayes, dan bahasa pemrograman Python.

### BAB III METODOLOGI PENELITIAN

Metodologi Penelitian berisi tentang penjelasan langkah-langkah yang akan digunakan dalam melaksanakan penelitian.

### BAB IV HASIL DAN PEMBAHASAN

Bagian ini berisi hasil yang dilakukan berdasarkan metodologi penelitian dan pembahasan mengenai hasil perancangan dan hasil sistem yang dibuat.

### BAB V PENUTUP

Pada bab ini berisikan beberapa kesimpulan dan saran dari hasil penelitian.

DAFTAR PUSTAKA  
LAMPIRAN