

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Peningkatan internet yang masif beberapa tahun terakhir ini telah menambah pertumbuhan *e-commerce*. Internet bukan hanya untuk menjadi sumber informasi lagi, tetapi menjadi sesuatu yang penting untuk berbelanja, belajar, berkomunikasi dan untuk mendapatkan layanan dari penyedia jasa seperti tukang kayu, dokter, kurir, logistik dan banyak pelayanan yang bisa didapat menggunakan internet.[1]

Persaingan *e-commerce* mengakibatkan banyaknya penyedia layanan *e-commerce* yang menawarkan jasa logistik. Asosiasi Perusahaan Jasa Pengiriman Ekspres Indonesia(Asperindo) memiliki lebih dari 1000 perusahaan yang menjadi anggota asosiasi[2].

Per Juli 2014, Twitter melaporkan 500 juta tweet per hari dan 255 juta pengguna aktif bulanan [3]. Tidak jarang bahwa banyak penyedia jasa yang melakukan pelayanan menggunakan media sosial sebagai alternatif terbaik untuk meningkatkan kepuasan pelanggan. Salah satu unsur penting dalam kepuasan pelanggan adalah kualitas. Kepuasan pelanggan akan semakin meningkat seiring dengan semakin tingginya kualitas yang diberikan oleh perusahaan. Kualitas layanan yang diciptakan tidak bisa diukur dari sudut pandang perusahaan, namun harus dari sudut pandang pelanggan.[4]

*Random Forest* adalah metode pembelajaran untuk klasifikasi dan regresi. Metode ini sangat mudah dipelajari bagi orang awam yang tidak memiliki latar belakang statistik yang kuat dengan sedikit riset dan pemrograman yang diperlukan [5]. Dari berbagai literatur *Random Forest* lebih unggul dengan akurasi yang tinggi dibandingkan dengan pohon tunggal. Dan dengan data *noisy* dan sentiment yang bernilai positif atau negative *Random Forest* mencapai akurasi sebesar 82,64% [5].

Berdasarkan penjelasan tersebut, peneliti menentukan untuk melakukan analisis sentimen pengguna ter menggunakan metode *Random Forest* dengan hyperparameter bernilai 10, 50, 100, 250, 500, 700, dan 1000 dengan nilai sentiment positif, negatif dan netral.

### 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah peneliti jabarkan, maka rumusan masalah yang didapat adalah "bagaimana algoritma random forest dapat mengklasifikasi sentimen tweet terhadap pelayanan jasa kurir JNE?".

### 1.3 Batasan Masalah

Batasan masalah penelitian ini adalah :

1. Pengambilan data tweet yang hanya mengandung #(hashtag)jnecare
2. Proses klasifikasi hanya data teks berbahasa Indonesia.
3. Data yang digunakan dalam penelitian ini berjumlah 1200 tweet.

4. Analisis sentimen dengan algoritma *Random Forest* menggunakan package yang sudah tersedia dari *python*.
5. Klasifikasi opini dibagi menjadi 3 kategori yaitu opini positif ,opini negative dan opini netral.
6. Pengujian menggunakan confusion matrix dan validasi menggunakan k-fold.

#### **1.4 Maksud dan Tujuan Penelitian**

Berdasarkan rumusan masalah diatas maka tujuan dari penelitian ini adalah mengetahui ketepatan akurasi klasifikasi menggunakan algoritma Random Forest dengan metode information gain dan gini impurity.

#### **1.5 Manfaat Penelitian**

Adapun manfaat penelitian tersebut adalah:

1. Hasil dari penelitian dapat digunakan oleh pihak JNE untuk mengetahui tingkat kepuasan pelanggan melalui opini yang di ambil dari tweet pelanggan sebagai rujukan dalam mengambil keputusan untuk langkah skedepanya.
2. Bagi peneliti dapat meningkatkan kemampuan data mining dan analisis terhadap sentiment masyarakat.

#### **1.6 Metodologi Penelitian**

1. Pengumpulan Data
  - a. Studi Literatur

Studi literatur dilakukan untuk mempelajari penelitian dengan tema sejenis yang telah dilakukan sebelumnya. Penelitian tersebut dijadikan sebagai referensi untuk penelitian sekarang dengan contoh menambahkan elemen baru pada penelitian yang sedang dikerjakan. Sumber yang digunakan bisa berupa jurnal penelitian, website, dan buk-buku yang berkaitan dengan penelitian.

b. Pengumpulan Data Tweet

Data yang dikumpulkan dari twitter berupa *tweet* masyarakat yang berkaitan dengan *mention @jnecare*. Data yang sudah dikumpulkan akan dilabeli secara manual untuk data yang akan dilatih. Data dibagi menjadi 80% untuk *data training* dan 20% untuk *data test*. Pada *data test* sentimen masyarakat tidak dilebali seperti *data training*.

2. Eksperimen

a. Alat dan Bahan

Pada penelitian ini menggunakan bahasa pemrograman Python 3.6 dengan IDE JetBrains Pycharm Community Edition 2019, dengan *library* yang digunakan adalah Scikit-learn untuk proses *machine learning* dan *library* Sastrawi versi 2.0.1 untuk *preprocessing data*. Sedangkan untuk data tweet diambil menggunakan API twitter pihak ke tiga yaitu TAGS v.1.6.1.

b. Tahap penelitian

Pada proses ini akan dibuat rancangan model untuk selanjutnya akan dilakukan tahap pengujian menggunakan *confusion matrix*. Tahapan dalam rancangan model tersebut antara lain yaitu *preprocessing*, seleksi fitur menggunakan *gain ratio*, pembobotan kata menggunakan TF-IDF, dan melakukan klasifikasi data latih menggunakan algoritma *random forest*.

c. Pengujian

Model yang telah dibuat kemudian diuji dan dievaluasi. Hasil uji klasifikasi adalah akurasi yang akan dilakukan menggunakan metode *confusion matrix*. Dan proses validasi model yang telah dibangun akan menggunakan metode K-fold dengan  $K=10$  untuk memeriksa apakah model melakukan klasifikasi secara *overfitting*.

d. Hasil dan Pembahasan

Pada tahap ini memaparkan dan membahas hasil dari eksperimen yang telah dilakukan. Pembahasan yang dilakukan mengenai akurasi model.

e. Penulisan Laporan

Hasil penelitian kemudian dituliskan ke dalam sebuah laporan. Dalam penulisan laporan penelitian akan ditarik pula suatu

kesimpulan berdasarkan hasil pembahasan dan pengujian, serta saran untuk penelitian berikutnya

### 1.7 Sistematika Penulisan

Sistematika yang digunakan dalam penelitian ini adalah sebagai berikut:

#### BAB I PENDAHULUAN

Memuat latar belakang masalah, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika penulisan.

#### BAB II LANDASAN TEORI

Menguraikan kajian pustaka dan dasar teori yang berkaitan dengan penelitian, seperti referensi-referensi dengan tema yang sama serta definisi yang berhubungan dengan ilmu dan permasalahan yang sedang diteliti.

#### BAB III ANALISIS PENELITIAN

Pada bab ini akan diuraikan tentang analisis penelitian .

#### BAB IV HASIL DAN PEMBAHASAN

Membahas tentang implementasi dari hasil analisis pada BAB III, selain itu juga berisi script program yang dibuat.

## BAB V PENUTUP

Bab ini menyajikan kesimpulan dalam penelitian serta saran-saran yang didapat dari hasil analisis, perancangan dan implementasi.

## DAFTAR PUSTAKA.

