

# 1. BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Ekstraksi opini konsumen merupakan proses menghasilkan informasi baru dan penting dari opini yang berisi pernyataan mengenai klaim tentang layanan atau produk yang diberikan oleh *e-commerce* kepada konsumen (Vikas and Mungara, 2016). Informasi yang diperoleh akan memberikan manfaat bagi konsumen maupun penjual. Bagi penjual, informasi ini akan digunakan untuk mengelola kepuasan, kepercayaan serta loyalitas konsumen demi perkembangan jangka panjang *e-commerce* (Addepalli *et al.*, 2016). Bagi konsumen, informasi yang diperoleh dapat menjadi rekomendasi dalam proses pembelian (Sohail, Siddiqui & Ali, 2014), (Ramanathan & Meyyappan, 2019). Pemilihan opini berdasarkan pada kasus terdapat nilai *rating* yang tidak sesuai dengan opini (Fan, Chang & Cui, 2018). Menurut (Tama, Sibaroni & Adiwijaya, 2019), efek opini memberikan pengaruh terhadap pembelian sebesar 73% - 87%. Ekstraksi opini konsumen secara otomatis merupakan hal yang penting (Vikas & Mungara, 2016). Informasi yang akurat dapat diperoleh dengan membaca semua opini (Rajganesh, Nandhini & Sumitha, 2016). Selain itu, perlu memperhatikan bahasa pada opini yang tidak memperhatikan struktur kalimat, penggunaan bahasa *non-formal* serta penggunaan *emoticon* atau gambar (Yassine & Hajj, 2010), (AL-Rubaiee *et al.*, 2018). Dengan kondisi opini yang beragam, maka perlu menerapkan *text mining* untuk mengubah opini yang dipahami bahasa manusia ke opini yang dapat dipahami komputer untuk diproses lebih lanjut (Davydova, 2018).

Pada penelitian sebelumnya, penerapan *text mining* untuk mengolah opini konsumen dalam konteks bahasa Inggris sudah banyak diterapkan. Penelitian (Vikas & Mungara, 2016) yang bertujuan untuk mengekstrak opini dan menarik kesimpulan untuk peningkatan rekomendasi produk membandingkan metode *naive bayes* dan *k-*

*nearest neighbor* dengan melihat pada faktor *f-measure*. Hasil penelitian menghasilkan bahwa metode naive bayes menghasilkan nilai *f-measure* diatas 75%. Penelitian lainnya oleh (Rangu, Chatterjee & Valluru, 2017) dalam meningkatkan kualitas produk dan pengalaman konsumen dengan menerapkan klasifikasi pada opini dengan membandingkan tiga metode klasifikasi yaitu Support Vector Machine (SVM), Random Forest dan Bayes. Hasil dari penelitian menghasilkan akurasi terbaik diperoleh SVM dengan nilai akurasi 69%. Penelitian lainnya oleh (Xia & Jiang, 2018) dengan menerapkan metode Latent Aspect Rating Analysis (LARA), metode ini menerapkan algoritma segmentasi dan aspek *rating* untuk mendapatkan pemahaman lebih untuk opini konsumen. Dari beberapa penelitian di atas, topik ini sudah berkembang namun tidak demikian untuk konteks bahasa Indonesia. Penelitian opini konsumen dalam bahasa Indonesia masih sedikit dan terbatas. Penelitian (Indraloka & Santosa, 2017), menerapkan metode k-means untuk melakukan klasifikasi pada data tweet yang keluarannya berupa daftar kata dan jenis konten tweet yang sering muncul. Penelitian (Pamungkas & Putri, 2017), mengolah *dataset* opini berbahasa Indonesia namun diterjemahkan terlebih dahulu ke bahasa Inggris baru diproses klasifikasi.

Berdasarkan uraian di atas, maka diusulkan ekstraksi opini konsumen secara otomatis dengan menerapkan teknik *text mining* dengan pendekatan *lexicon*, dimana dalam mengolah opini memerlukan kamus kata. Dengan menggunakan pendekatan *lexicon*, terdapat dua informasi yang dapat diperoleh yaitu bobot opini dan klasifikasi. Bobot opini akan dihitung untuk mendapatkan klasifikasi yang akan mengelompokkan opini ke dalam kelas positif, negatif atau netral. Hasil klasifikasi opini akan dibandingkan dengan *rating* yang sudah diklasifikasikan untuk mendapatkan tingkat akurasi.

## 1.2. Masalah Penelitian

Berdasarkan uraian latar belakang di atas, maka masalah dalam penelitian ini dapat diuraikan menjadi dua bagian yaitu identifikasi masalah dan rumusan masalah.

### 1.2.1. Identifikasi Masalah

Adapun yang menjadi identifikasi masalah pada penelitian ini adalah sebagai berikut:

1. Opini konsumen yang banyak pada *e-commerce* menyebabkan konsumen kesulitan mendapatkan informasi.
2. Bahasa opini dengan bentuk yang beragam dan penggunaan kata tidak baku yang hanya dapat dipahami manusia.
3. Terdapat kasus dimana antara *rating* dengan opini tidak sesuai.
4. Ketersediaan *tools* yang hanya dapat mengolah opini konsumen dalam bahasa Inggris dan tidak dapat mengolah opini bahasa Indonesia.

### 1.2.2. Rumusan Masalah

Rumusan masalah pada penelitian ini adalah sulitnya mengekstrak opini konsumen pada *e-commerce* yang bentuknya beragam dan menggunakan kata yang tidak baku pada opini berbahasa Indonesia serta membandingkannya dengan *rating*.

## 1.3. Tujuan dan Manfaat Penelitian

### 1.3.1. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mampu menerapkan *text mining* dengan pendekatan *lexicon* untuk mengklasifikasikan opini konsumen yang berbahasa Indonesia.
2. Membandingkan *rating* yang diberikan konsumen dengan opini konsumen.

### 1.3.2. Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

1. Hasil klasifikasi dapat membantu konsumen dalam proses pembelian berdasarkan pengalaman konsumen sebelumnya.

2. Hasil klasifikasi dapat memudahkan pihak Shopee dan penjual untuk mengelola kepuasan, kepercayaan serta loyalitas konsumen demi perkembangan jangka panjang *e-commerce*.

#### 1.4. Ruang Lingkup

Ruang lingkup yang ditentukan untuk menghindari perluasan pembahasan dalam penelitian ini adalah sebagai berikut:

1. Data yang akan diklasifikasi hanya data opini konsumen Shopee berupa data sampel produk pada kategori “Pakaian Wanita” yang dikumpulkan secara *online* dari tanggal 1 Maret 2019 sampai 30 April 2019 yang berjumlah 125.144 opini yang memiliki komentar.
2. Opini yang diproses hanya dalam bentuk teks alphabet.
3. Perbaikan kesalahan eja berupa karakter yang mengalami perulangan.
4. Atribut yang diperlukan berupa *itemname*, *date*, *rating* dan opini.
5. Pendekatan *lexicon* yang digunakan berbasis kamus kata.
6. Opini akan diklasifikasikan ke dalam kelas positif, negatif atau netral.

#### 1.5. Metodologi Penelitian

Metodologi yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Studi Literatur

Pada tahap ini dilakukan pendalaman mengenai bagaimana cara kerja *text mining* dan proses klasifikasi pada beberapa penelitian sebelumnya.

2. Tahap Analisis

Pada tahap ini dilakukan proses untuk mengidentifikasi data yang dibutuhkan, masalah dan tantangan yang harus diselesaikan dan menjelaskan solusi yang diusulkan untuk menyelesaikan masalah dan tantangan yang ada. Proses dalam *text mining* akan digambarkan dalam bentuk *flowchart*.

### 3. Perancangan Model

Perancangan model dimulai dari mengumpulkan *dataset* opini, menerapkan *text preprocessing* kemudian ekstraksi opini sampai diperoleh pembobotan dan klasifikasi untuk setiap opini.

### 4. Pengujian

Pengujian tingkat akurasi teknik klasifikasi opini yang berbahasa Indonesia dengan *rating* serta pengujian dengan *confusion matrix* untuk memperoleh nilai akurasi, presisi dan *recall*.

### 5. Evaluasi

Pada tahap ini, dilakukan evaluasi terhadap hasil pengujian yang sudah dilakukan untuk mengambil kesimpulan dan saran.

## 1.6. Sistematika Penulisan

Sistematika penulisan laporan penelitian ini terdiri dari 5 bab, dimana secara garis besar masing-masing bab membahas hal – hal berikut ini. Bab 1 Pendahuluan, berisi penjelasan umum, masalah dan solusi yang akan dilakukan penelitian. Bab 2 berisi studi literatur dan tinjauan singkat terkait masalah dan metode yang berhubungan dengan penelitian yang akan dilakukan. Bab 3 Metodologi Penelitian, berisi identifikasi masalah, langkah-langkah dari metode yang diusulkan, data yang digunakan, alat-alat penelitian dan metode analisis. Bab 4 Hasil dan Pengujian, berisi hasil dari sistem yang dibangun dan analisis berdasarkan pengujian yang dilakukan. Bab 5 Kesimpulan dan Saran, berisi kesimpulan yang diperoleh dari hasil dan pengujian penelitian yang sudah dilakukan dan saran untuk hasil yang lebih baik dalam penelitian yang sejenis.