BAB 1

PENDAHULUAN

1.1 Latar Belakang

Perkembangan cepat yang dialami oleh internet dan teknologi dapat dirasakan dengan perkembangan-perkembangan dan inovasi baru yang muncul dikalangan masyarakat saat ini. Salah satu contohnya adalah *e-commerce*. Sejak berdirinya *e-commerce* pada tahun 1995 sampai sekarang, *e-commerce* membawa pengaruh yang sangat besar pada perubahan bisnis dan kebiasaan pelanggan dalam membeli maupun menjual barang yang mereka punya atau butuhkan (Laudon & Traver, 2014). Hal ini dapat dilihat dari perkembangan pesat yang dialami oleh beberapa *e-commerce* di Indonesia, karena layanan atau jasa yang *e-commerce* berikan kepada pelanggan untuk mempermudah mereka dalam berbelanja maupun menjual barang dagang mereka. Kepraktisan yang diberikan oleh *e-commerce* memang tidak dapat dipungkiri lagi, dari segi pengaksesan yang dapat dilakukan dimana-mana, metode-metode transaksi yang diberikan sangat jelas dan aman hingga pemilihan kurir yang disediakan untuk mengirim barang.

Untuk merasakan kepraktisan layanan dan jasa yang disediakan oleh *e-commerce*, pengguna dapat mengaksesnya melalui *website* maupun mengunduh aplikasi berbasis *mobile apps* yang sudah tersedia pada *play store* atau *app store* pada masing-masing *gadget* yang dimiliki.

Beberapa *e-commerce* yang dianggap penulis cukup populer dan sudah cukup diakui oleh masyarakat di Indonesia dalam segi layanan dan jasa yang diberikan yaitu Tokopedia, BukaLapak, dan, Shopee. Sampai saat ini tercatat pengunduhan yang dilakukan melalui *play store* untuk aplikasi Tokopedia adalah lebih dari 50 juta pengunduh dan lebih dari dua juta *reviews*, untuk aplikasi BukaLapak adalah lebih dari 10 juta pengunduh dan lebih dari satu juta *reviews*, dan untuk Shopee tercatat lebih dari 50 juta pengunduh dan lebih dari dua juta *reviews*.

Reviews yang diberikan beragam, mulai dari keluhan yang dialami selama pemakaian aplikasi, kritik, saran yang diajukan, sampai kepuasan yang diterima ketika menggunakan aplikasi tersebut. Data dari *reviews* diatas berjumlah banyak dan biasanya tidak berstruktur sehingga sulit untuk diolah oleh perusahaan *e-commerce*, hal ini menyebabkan banyak kritik, atau masukan yang diberikan oleh pengguna tidak tersampaikan atau tidak terlihat oleh pihak perusahaan. Oleh karena itu, *sentiment analysis* berbasis *machine learning* perlu dilakukan untuk mengolah data-data yang tidak berstruktur tadi menjadi data-data yang lebih terstruktur sehingga dapat dikumpulkan dan diolah lebih lanjut untuk memperoleh opini dan *sentiment* yang nantinya akan menghasilkan kesimpulan mengenai *experience* yang dirasakan pengguna selama menggunakan aplikasi *e-commerce* (Wahyu Handani, Intan Surya Saputra, Hasirun, Mega Arino, & Fiza Asyrofi Ramadhan, 2019).

Sentiment analysis adalah suatu proses analisis yang mengubah suatu data tidak berstruktur yang berasal dari berbagai sumber di internet, menjadi data terstruktur yang dapat digunakan untuk mengetahui opini dari pengguna mengenai kualitas suatu barang atau jasa yang mereka gunakan. Hasil dari pengolahan nantinya dapat dijadikan acuan atau referensi oleh pengguna-pengguna lain terhadap menilai suatu barang, jasa atau layanan yang diberikan oleh suatu perusahaan (Alamsyah, Rahmah, & Irawan, 2015). Bagi perusahaan pemanfaatan hasil *sentiment analysis* juga dapat dijadikan sebagai *feedback* yang membangun, umpan balik yang dapat dijadikan acuan terhadap keberlangsungan suatu produk dan juga bagaimana membangun layanan yang banyak diminati oleh masyarakat, secara tidak langsung melalui hasil dari analisis ini perusahaan dapat menerima rekapan opini dari hasil produk atau layanan yang mereka berikan sehingga untuk kedepannya mereka lebih memiliki gambaran harus bagaimana melayani pelanggan-pelanggan mereka.

Pemerosesan opini-opini tadi dilakukan pada tahap *sentiment analysis polarity classification* dimana pada proses ini penentuan teks dilakukan untuk memisahkan mana teks yang objektif (ekspresi yang berupa fakta yang dinyatakan pada sesuatu) dan mana teks yang subjektif (ekspresi yang berupa opini yang dinyatakan pada sesuatu). Selain itu, penentuan juga dilakukan pada opini dilihat mana opini yang negatif dan mana opini yang positif (Abbasi, Chen, & Salem,

2008). Selain pada teks, *sentiment analysis polarity classification* juga dapat dilakukan pada dokumen, kalimat, dan beberapa kalimat (Abbasi et al., 2008). Seperti pada dokumen pengklasifikasian dapat dilakukan pada review pada suatu film seperti contohnya pada saat pembelian tiket di TIX ID dimana kita dapat melihat review yang diberikan oleh para penonton yang telah menonton film tersebut, lalu pada posting forum *web* dimana biasanya kita dapat memberikan nilai atau tanggapan terhadap suatu komentar yang diberikan pada forum *web* (Abbasi et al., 2008). Pada kalimat pengklasifikasian hampir mirip pada tahap yang dilakukan pada teks, dimana penentuan dilakukan pada kalimat yang bersifat subjektif maupun objektif selain itu juga ditentukan kalimat mana yang memiliki makna positif dan kalimat mana yang memiliki makna negatif (Abbasi et al., 2008). Pada beberapa kalimat, pengklasifikasian dapat dilihat beberapa makna yaitu negatif maupun positif yang terdapat pada suatu kalimat pernyataan mengenai suatu hal (Abbasi et al., 2008).

Pada penelitian ini akan dilakukan klasifikasi *sentiment analysis* terhadap *review* yang didapatkan dari masing-masing *e-commerce* pada *play store*. Klasifikasi *sentiment* yang dimaksud tergolong menjadi dua yaitu negatif dan positif.

Dalam tahap pemrosesan klasifikasi dari *sentiment analysis* ada beragam teknik algoritma yang dapat digunakan, teknik algoritma yang paling umum adalah *linear regression*, *linear regression* merupakan teknik statistik yang sering digunakan untuk memprediksi adanya pengaruh yang diberikan suatu variabel terhadap variabel lainnya (Fikri, 2013). dan beberapa teknik lain yang sering digunakan saat ini adalah algoritma *machine learning*, seperti *support vector machine (SVM)*, *naive bayes*, dan *deep learning* (Abbasi et al., 2008). *Naive bayes* merupakan algoritma probalistik yang menggunakan teori *bayes* untuk melakukan prediksi pada suatu teks (Abbasi et al., 2008), *support vector machine* merupakan teknik yang digunakan untuk mencari *hyperlane* yang berfungsi sebagai pemisah dari dua set data yang berasal dari dua buah kelas yang berbeda (Octaviani, Yuciana Wilandari, & Ispriyanti, 2014), *deep learning* merupakan teknik yang saat ini cukup populer pada algoritma *machine learning* dimana *deep learning* ini erat hubungannya dengan *artificial intelligence* yang mempelajari tingkatan representasi pada suatu level rendah hingga mencapai level representasi yang tinggi yaitu

abstrak, guna memahami dan memprediksi suatu teks, gambar, dan suara (Lecun, Bengio, & Hinton, 2015). Hasil prediksi menjajikan yang dihasilkan oleh *deep learning* sudah terbukti mengalahkan beberapa teknik *mechine learning* lainnya. Beberapa contohnya seperti, memprediksi gambar, proses pengklasifikasian, dan penerjemahan bahasa (Lecun et al., 2015).

Sebelum teknik-teknik pengklasifikasian diterapkan, ada yang dikenal dengan istilah perepresentasi kata, ada beberapa teknik yang sering dipakai untuk merepresentasikan kata. Beberapa teknik yang sudah sering digunakan adalah

1. *Bag-of-Words (BoW)*

Bag-of-Words merupakan salah satu teknik representasi kata yang cukup sering digunakan. Pada teknik ini tiap dokumen direpresentasikan sebagai tas/kantung yang berisi informasi-informasi penting mengenai penjelasan terhadap sesuatu, yang kata-katanya tidak berurutan. Dalam pengaplikasiannya, *Bag-of-Words* terkenal dalam merepresentasikan setiap gambar sebagai sebuah histogram kata yang ada pada gambar tersebut (Li, Mei, Kweon, & Hua, 2011).

2. *N-Gram*

Teknik ini berfungsi dengan cara memisahkan teks menjadi *string* dengan panjang N yang ditentukan berdasarkan teks tersebut. Terdapat beberapa pembagian dalam *N-Gram*. Seperti unigram $n=1$, bigram $n=2$, dan trigram $n=3$ (Lisangan, 2015). Contoh kalimatnya “tupai pandai dalam melompat”

Dibagi menjadi bigram menjadi,

- tupai pandai
- pandai dalam
- dalam melompat

3. *Word Embedding*

teknik lain yang terkenal pada saat ini adalah *word embedding*. Pada *word embedding* kata nantinya akan di representasi dengan *vector* yang berbentuk angka. Nantinya kata yang memiliki makna mirip

akan terletak berdekatan(Huang, Qiu, & Huang, 2014). Pada penelitian ini teknik representasi yang akan dipakai adalah *word embedding* karena dari beberapa penelitian, *word embedding* memberikan akurasi yang lebih tinggi dibandingkan kedua teknik lainnya berdasarkan jumlah data tertentu

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan diatas, maka rumusan masalah yang dapat diambil adalah sebagai berikut :

1. Bagaimana proses representasi *text mining* dengan pendekatan *word-embedding*?
2. Bagaimana merancang algoritma berbasis *text mining* dengan representasi *word embedding* dan *machine learning* dalam mengekstraksi opini pelanggan?
3. Bagaimana mengukur *performance* dari sistem ekstraksi opini yang dikembangkan?
4. Bagaimana menerapkan sistem baru dalam proses ekstraksi opini pada beberapa *e-commerce*?

1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengetahui cara dan tahapan-tahapan untuk membangun sebuah sistem *sentiment analysis* yang dapat digunakan untuk melakukan *review* beberapa perusahaan *e-commerce* yang berada di Indonesia. Tingkat tahapan-tahapannya berawal dari

1. Mencari tahu cara untuk melakukan representasi *text review* dengan pendekatan *word embedding*
2. Mencari tahu cara merancang suatu algoritma berbasis *text mining* dengan representasi *word embedding* dan *machine learning* dalam mengekstraksi opini pelanggan.
3. Mencari tahu cara untuk mengukur *performance* dari sistem ekstraksi opini yang dikembangkan.

4. Mencari tahu cara untuk penerapan sistem baru dalam proses ekstraksi opini pada beberapa *e-commerce*.

1.4 Manfaat Penelitian

Ada beberapa manfaat yang dapat diambil dari penelitian ini.

1. Pada penelitian ini pelatihan model *word2vec* dari *corpus wikipedia* menghasilkan beragam dimensi *vector*. Dengan ini, dapat dibandingkan hasil evaluasi mana dimensi *vector* yang memiliki tingkat keakuratan tertinggi dan paling baik digunakan untuk proses *sentiment analysis*
2. Penelitian ini menggunakan dua model algoritma yang terdapat pada *deep learning*. Dengan ini, dapat dilihat hasil perbandingan dari evaluasi penggunaan model, model mana yang memiliki tingkat keakuratan tertinggi dan paling baik digunakan untuk proses *sentiment analysis*
3. Penelitian ini membandingkan juga antara penggunaan *stopword removal* pada proses pembersihan data dan tanpa penggunaan *stopword removal* pada proses pembersihan data. Dari hasil evaluasi dapat dilihat mana yang lebih memiliki tingkat keakuratan yang tinggi dan paling baik untuk proses *sentiment analysis*
4. Manfaat dari penelitian ini bagi perusahaan *e-commerce* adalah mengetahui opini-opini yang diberikan pengguna terhadap layanan atau jasa yang mereka berikan pada aplikasi yang mereka buat agar kedepannya perusahaan *e-commerce* dapat lebih matang dalam menyajikan layanan atau jasa kepada pengguna sesuai dengan ekspektasi yang pengguna inginkan.
5. Manfaat bagi penulis adalah penulis jadi mempelajari hal-hal baru mengenai pembuatan dan penggunaan *sentiment analysis*.
6. Manfaat bagi dosen pembimbing sekiranya dapat dijadikan acuan untuk penelitian selanjutnya dan dapat menghasilkan hasil *sentiment analysis* yang baik.

1.5 Ruang Lingkup Penelitian

Ruang lingkup penelitian ini meliputi

1. Data *review* yang berasal dari beberapa aplikasi *e-commerce* yang berada di *play store* seperti Bukalapak, Tokopedia, dan Shopee dengan kurun waktu tiga bulan. Jumlah data yang diperoleh sebanyak 73.596
2. Penggunaan metode pada penelitian ini adalah *CRISP-DM*. Penggunaan metode ini ditujukan untuk mendapatkan hasil klasifikasi dari *review* yang diberikan berdasarkan sentimen pengguna (positif dan negatif)
3. Ada dua model algoritma *deep learning* yang digunakan pada penelitian ini yaitu model *artificial neural network* sederhana *flatten* dan model *conv1d*
4. Penggunaan teknik *word representation* yang digunakan pada penelitian ini adalah *word2vec*
5. Pada penelitian ini tahap akhir pada metode *CRISP-DM* yang berupa tahap *deployment* yang dimaksud adalah rekomendasi perancangan sistem, tidak sampai tahap pembagungan sistem.