



### **Hak cipta dan penggunaan kembali:**

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

### **Copyright and reuse:**

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

Sentimen analisis ataupun *opinion mining* merupakan ranah studi yang menganalisa opini, sentimen, evaluasi, sikap, dan emosi penggunanya dalam bentuk teks tertulis (Liu B. , 2012). Selain diimplementasikan dalam ranah bisnis, sentimen analisis juga kerap kali diperbincangkan dalam ranah penelitian (Vinodhini, 2012). Penelitian sebelumnya terkait klasifikasi sentimen *tweets* dengan menggunakan *Naïve Bayes* (Tunggawan, 2016). Namun demikian, *Naïve Bayes* merupakan metode yang mengasumsikan setiap kata yang muncul secara independen (Ginosar, 2012) dan lebih tepat diimplementasikan pada *datasets* dengan variasi data yang minim (Pang, 2002).

Penilaian *tweets* independen juga menunjukkan hasil yang tak optimal (Tang, 2015), terutama pada kasus berikut:

- Terjadinya misklasifikasi sentimen kalimat akibat adanya sentimen terhadap kata lainnya, disamping target kata.

Contoh: “*I bought a new camera. The picture quality is amazing but the battery life is too short.*”

Dalam contoh kalimat di atas, terdapat dua sentimen; (1) sentimen positif jika target kata merujuk pada: “*picture quality*”, dan (2) sentimen negatif jika target kata merujuk pada: “*battery life*”.

Untuk itu, diperlukan eksplorasi lebih lanjut atas pernyataan penelitian sebelumnya (Tang, 2015). Salah satu upaya eksplorasi yang dapat dilakukan yakni dengan merancang model yang mengasumsikan data secara dependen dan independen. Kedua model dilatih dan dites dengan menggunakan *datasets* yang serupa, dan hasil klasifikasi kedua model tersebut dibandingkan.

Berdasarkan pada penelitian serupa sebelumnya, metode *Long Short Term Memory* (LSTM) (Wang X. a., 2015) mampu mengklasifikasi sentimen yang kompleks. Perancangan model sentimen analisis kali ini menggunakan studi kasus pada *brand* Indomie. Pemilihan *brand* Indomie sendiri dimaksudkan agar *datasets* bersifat netral, tidak memihak beberapa penelitian sebelumnya yang menggunakan *datasets* berupa *tweets* mengenai pemilihan presiden Amerika (Tunggawan, 2016), dan *datasets* berupa *tweets* yang memuat konten mengenai selebriti, perusahaan, dan produk tertentu (Tang, 2015).

Perancangan model sentimen analisis diharapkan dapat mengeksplorasi argumen pada penelitian sebelumnya (Tang, 2015) dengan menggunakan *datasets* studi kasus *brand* Indomie yang berupa *tweets*, dan dapat dijadikan landasan bagi pengembangan sentimen analisis berikutnya yang lebih kompleks.

Penelitian ini terbagi atas dua bagian, mengingat besarnya skala penelitian. Terdapat 2 bagian pada penelitian ini, diantaranya:

1. Bagian pertama terdiri atas: pengumpulan data, pembersihan data, pemberian label pada data, pembentukan *corpus* dan *bag-of-words*, serta perancangan model *Naïve Bayes*.

Seluruh kegiatan yang terdapat pada bagian pertama dituntaskan pada mata kuliah: SKRIPSI I.

2. Bagian kedua terdiri atas: pembentukan *word2vec: skip-gram model*, perancangan model *Long Short Term Memory (LSTM)*.

Bagian kedua diselesaikan pada mata kuliah: SKRIPSI II.

## 1.2. Rumusan Masalah

Berikut rumusan masalah dalam penelitian yang diadakan:

1. Bagaimana menganalisa sentimen *tweets* dengan model *Naïve Bayes*?
2. Bagaimana menganalisa sentimen *tweets* dengan model *Long Short Term Memory*?
3. Bagaimana komparasi tingkat akurasi antar model yang dirancang terhadap *datasets* yang disediakan?

## 1.3. Batasan Masalah

Berikut batasan masalah yang akan menjadi ruang lingkup penelitian :

1. Klasifikasi polaritas sentimen pada *tweets* merupakan klasifikasi secara biner; sentimen positif dengan nilai: 1 dan sentimen negatif dengan nilai: 0. Klasifikasi biner tersebut bertujuan untuk mengetahui secara

jasas polaritas makna sentimen pada *tweets*. *Tweets* dengan makna bias tidak diikutsertakan dalam *datasets*.

2. *Datasets* terdiri atas kumpulan *tweets* yang memiliki kata kunci ‘indomie’. *Tweets* yang merujuk pada *brand* Indomie, namun mengandung kata kunci yang berbeda yakni kombinasi angka dan huruf tidak diikutsertakan dalam *datasets*. Contoh kata kunci kombinasi angka dan huruf yang dimaksud: “Ind0mi3”, dan sebagainya.
3. Hanya *URL* dan gambar yang dihilangkan dalam *tweets*. Seluruh *tweets* dalam *datasets* dijaga orisinalitasnya

#### 1.4. Tujuan dan Manfaat

Adapun tujuan dari penelitian berikut ialah:

- Mengetahui analisa sentimen *tweets* yang terdapat dalam *datasets* dengan model *Naïve Bayes* dan *Long Short Term Memory*.
- Mengetahui tingkat akurasi antar model dalam menganalisa sentimen terhadap *datasets* yang diberikan.

Manfaat utama dari penelitian ini ialah untuk pengembangan penelitian sebelumnya melalui eksplorasi perancangan model *Naïve Bayes* dan *Long Short Term Memory* dalam menganalisa sentimen *tweets*.