

1. PENDAHULUAN

1.1 Latar Belakang

Seiring dengan perkembangan dunia internet berpengaruh terhadap proses transaksi jual dan beli yang dimana proses sebelumnya bersifat konvensional berubah ke tingkat modern dengan adanya internet atau bisa disebut dengan *e-commerce*. Perkembangan *e-commerce* yang sangat cepat merupakan peluang besar bagi produsen untuk memasarkan produknya dan mempromosikannya kepada konsumen. Di dunia ini banyak sekali *e-commerce* salah satunya yang terkenal adalah Amazon. Pada website Amazon banyak sekali produk-produk yang dijual mulai dari barang-barang elektronik, produk kesehatan dan kecantikan, produk bayi, dan sebagainya. Pada website Amazon sendiri terdapat *review* dari para pengguna yang sudah membeli suatu produk untuk memberikan komentarnya terhadap produk tersebut. Dengan adanya kolom *review* akan menjadi masukan terhadap pihak Amazon serta para pembeli yang ingin membeli produk tersebut jadi lebih mengetahui sebelumnya berdasarkan *review* para pengguna. Akan tetapi dengan banyaknya *review* yang ada terkadang pembeli malas membaca satu per satu kolom ulasan tersebut dan hanya melihat komen-komen teratas dengan berbagai *rating* yang tinggi. Pada *review* juga terdapat banyak kata-kata yang menjadi masukan untuk pihak Amazon atau pihak yang memproduksi barang agar barangnya dapat ditingkatkan lagi kualitasnya.

Referensi dari Chris Nicolls dan Fei Song diterbitkan pada tahun 2009, dalam penelitiannya mengenai pemilihan *feature set* untuk meningkatkan performa dari *text classification*. Mereka menyebutkan bahwa pemilihan *feature set* pada *sentiment analysis* cukup berpengaruh terhadap performa serta menggunakan metode *maximum entropy* untuk metode pengklasifikasiannya. Hal ini dibuktikan dari penelitian mereka dengan menggunakan metode POS *Weighting* dengan kombinasi 1, 2, 3, 4, dan 5 untuk kategori *noun*, *verb*, *adv.*, dan *adj* menunjukkan akurasi sebesar 79.41%. Adapun *dataset* yang digunakan sebanyak 574 *reviews* untuk 13 produk yang berbeda dengan perbandingan *review* positif dan negatif

69.5% : 30.5% [1]. Kemudian Referensi dari Amit Gupte, Sourabh Joshi, Praktik Gadhul dan Akhay Kadam yang diterbitkan pada tahun 2014, pada penelitiannya membahas tentang perbandingan pada algoritma klasifikasi yang digunakan pada analisis sentimen. Adapun algoritma yang diuji yaitu *naïve bayes*, *max entropy*, *boosted trees* dan *random forest*. Beberapa aspek diuji seperti kesederhanaan, performansi, akurasi, memori yang dibutuhkan, hasil akurasi selama periode tertentu dan waktu yang dibutuhkan untuk mengklasifikasi data training. Metode *maximum entropy* mempunyai tingkat akurasi yang cukup tinggi. Lalu untuk performansinya masih lebih bagus dibanding dengan metode *naïve bayes* [2].

Oleh sebab itu pada penelitian tugas akhir ini akan membagi suatu ulasan produk berdasarkan opini positif dan opini negatif pada *data* ulasan produk Amazon mengklasifikasikannya ke dalam kelas positif atau negatif dengan menggunakan metode dalam pengkategorian opini yang disebut dengan *sentiment analysis*. *Sentiment analysis* merupakan proses memahami, mengekstrak dan mengolah *data* tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung di dalam suatu kalimat opini [3]. Pada permasalahan diatas peneliti ingin membagi suatu opini kedalam opini positif dan opini negatif pada website Amazon dengan kategori penjualan *cell phones & accessories*. Menggunakan metode *tf-idf* sebagai pembobotan fitur. Lalu *maximum entropy* untuk membangun model *classifier*. Kemudian akan dilakukan pemisahan label kelas positif dan negatif dengan syarat jika kata tersebut bernilai lebih dari sama dengan 0 maka hasilnya adalah positif dan jika bernilai kurang dari 0 maka hasilnya adalah negatif.

1.2 Perumusan Masalah

Rumusan masalah yang akan dibahas dalam penelitian ini berupa :

- a. Bagaimana cara membangun *maximum entropy* untuk mengklasifikasikan data ulasan produk toko *online*?
- b. Bagaimanakah performa *maximum entropy* dalam mengklasifikasikan sentimen pada data ulasan produk toko *online*?

Adapun batasan masalah yang dipakai pada Tugas Akhir ini yaitu;

- a. *Data* yang diambil merupakan opini berbahasa Inggris.
- b. *Data* teks yang dipakai berasal dari <http://www.amazon.com> dengan mengambil kategori *cell phones & accessories* [4].
- c. Penelitian ini merupakan sentimen level kalimat.
- d. Ekstraksi fitur menggunakan *tf-idf*.
- e. Metode klasifikasi yang digunakan yaitu *maximum entropy*.

1.3 Tujuan

Tujuan pembuatan tugas akhir ini adalah :

- a. Dapat membangun *classifier* yang mampu mengklasifikasikan sentimen pada ulasan produk secara otomatis.
- b. Untuk menghasilkan performa klasifikasi yang baik dari *maximum entropy*

1.4 Metodologi Penyelesaian Masalah

Tahapan yang akan dilakukan di dalam penyelesaian tugas akhir ini yaitu:

- a. Tahap Studi Literatur
Pada tahap ini, penyusun akan mencari referensi-referensi yang terkait dengan tugas akhir ini dengan mengumpulkan berbagai referensi seperti jurnal, *paper* dan buku tugas akhir senior. Hal tersebut dilakukan untuk mendapatkan pemahaman tentang metode yang tepat dalam menyelesaikan masalah yang diangkat.
- b. Tahap Pencarian dan Pengumpulan *Data*
Pada tahap ini akan dilakukan pencarian *data* dan pengumpulan *data* yang terdapat pada situs Amazon dengan mengambil kategori *Cell Phones & Accessories*.

c. Tahap Pengolahan *Data*

Pada tahap ini akan dilakukan *preprocessing data*. *Preprocessing* dilakukan agar *data* tersebut bisa diolah oleh sistem dengan melakukan *data cleaning, case folding, tokenization, remove punctuation, stopword removal, lemmatization, dan stemming*.

d. Tahap Perancangan Sistem

Pada tahap ini akan dibuat rancangan suatu sistem dengan memperhatikan diagram alir yang telah dibuat

e. Tahap Implementasi

Pada tahap ini, implementasi sistem dilakukan setelah perancangan sistem selesai dilakukan dengan menuliskan kode program berdasarkan diagram alir yang telah dibuat. Implementasi menggunakan bahasa pemrograman Python.

f. Tahap Analisis Hasil Pengujian

Pada tahap ini, dilakukan analisis dari pengujian yang sudah dilakukan dan menarik suatu kesimpulan berdasarkan metode yang diimplementasikan kedalam sistem tersebut.

g. Tahap Pembuatan Laporan

Pada tahap ini, penyusun akan melakukan proses penyusunan laporan akhir serta pengumpulan dokumentasi dengan berdasarkan hasil dari penelitian yang dilakukan.

1.5 Sistematika Penulisan

Sistematika penulisan Tugas Akhir ini secara garis besar akan diuraikan sebagai berikut :

BAB 1 : PENDAHULUAN

Pada bab ini berisi latar belakang pemilihan tema, perumusan masalah dan Batasan masalah, tujuan, metodologi penyelesaian masalah dan sistematika penulisan.

BAB 2 : TINJAUAN PUSTAKA

Pada bab ini menjelaskan mengenai teori-teori yang menjadi pedoman dalam penulisan tugas akhir yang meliputi penjelasan teori. Adapun teori yang dijelaskan yaitu amazon, *text preprocessing*, *tf-idf*, *sentiment analysis*, klasifikasi dan *maximum entropy*.

BAB 3 : PERANCANGAN SISTEM

Bab ini menjelaskan mengenai penelitian yang dilakukan sehingga dapat dibuat perancangan sistem dari segi gambaran umum, rancangan, dan fungsionalitas dari sistem itu sendiri.

BAB 4 : PENGUJIAN DAN ANALISIS

Pada bab ini akan membahas mengenai Analisa hasil penelitian dengan melakukan skenario pengujian. Sehingga akan diketahui apakah sesuai dengan tujuan yang diharapkan atau belum.

BAB 5 : KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran dari seluruh kegiatan penelitian tugas akhir ini yang dibutuhkan terhadap hasil akhir dari penelitian yang telah dibuat.