

Analisis Pelabelan dalam Klasifikasi Sentimen Ulasan Produk dengan Menggunakan Algoritma Multinomial Naïve Bayes

Vicha Octavia Tama¹, Yuliant Sibaroni², Adiwijaya³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹octaviavicha@students.telkomuniversity.ac.id, ²yuliant@telkomuniversity.ac.id,

³adiwijaya@telkomuniversity.ac.id

Abstrak

Seiring dengan perkembangan teknologi, *e-commerce* turut mengalami perkembangan yang cukup pesat. Keberadaan *e-commerce* menjadi alternatif lain konsumen untuk mempermudah mereka berbelanja memenuhi kebutuhannya. Setelah membeli barang, konsumen bebas melakukan penilaian terhadap produk yang mereka beli. Ulasan produk dan *rating* yang diberikan oleh konsumen merupakan salah satu sarana yang dapat digunakan untuk meningkatkan penjualan dan dapat pula digunakan untuk menentukan keputusan dalam pembelian suatu produk dengan membaca ulasan produk tersebut. Namun dengan menggunakan *rating* dan ulasan saja tidak cukup untuk menyimpulkan pendapat seseorang. Maka dari itu pada Tugas Akhir ini dibangun sistem yang dapat mengklasifikasikan opini terhadap ulasan produk menjadi sentimen positif dan negatif dengan memanfaatkan *rating*. *Dataset* yang digunakan adalah *Grocery and Gourmet Food* dari Amazon sebanyak 50.000 yang kemudian akan dilabeli menggunakan Metode Pelabelan *Average* dan *Binary*. Klasifikasi opini ini menggunakan pendekatan *Supervised learning* Algoritma *Multinomial Naïve Bayes*. Hasil penelitian ini didapatkan bahwa pelabelan menggunakan Metode *Average* cocok untuk mengolah *Dataset Grocery and Gourmet Food* dan membuktikan bahwa jumlah terbaik penggunaan *feature selection* sebesar 20% berhasil menghasilkan akurasi sebesar 80.48%. **Kata kunci :** *sentimen analisis, review produk, multinomial naive bayes, pelabelan average, pelabelan binary, feature selection*

Abstract

Along with the development of technology, e-commerce also experienced a fairly rapid development. The existence of e-commerce becomes another consumer alternative to make it easier for them to fulfill their needs, participated in a fairly rapid development. After buying the goods, consumers care free to assess the products they buy. Product reviews and ratings provided by consumers are one means that can be used to increase sales and can also be used to determine the decision in purchasing a product by reading the product reviews. However, using ratings and reviews alone is not enough to summarize one's opinion. Therefore in this Final Project build a system that can classify opinions on products reviews into positive and negative sentiments by utilizing the rating. The dataset used is Grocery and Gourmet Food from Amazon as much as 50,000 which will then be labeled using Labeling Methods Average and Binary. The classification of this opinion uses the approach of Supervised learning Algorithm Multinomial Naïve Bayes. The results of this research shows that labeling using Method Average is suitable for processing Grocery and Gourmet Food Dataset and proves that the best amount of feature selection usage is 20% succeed to produce 80.48% accuracy.

Keywords: *analysis sentiment, product review, multinomial naive bayes, average labeling, binary labeling, feature selection*

1. Pendahuluan

Latar Belakang

Seiring berjalannya waktu, teknologi mengalami perkembangan yang pesat. Hal tersebut membuat masyarakat harus bisa beradaptasi dengan perkembangan tersebut. Salah satu perubahan yang dapat kita lihat adalah pergeseran perilaku masyarakat dalam membeli kebutuhan barang secara *offline shop* menjadi *online shop* atau *E-commerce* [1]. Alasan yang mendasari pergeseran perilaku tersebut adalah kemudahan yang ditawarkan oleh *online shop*. Namun selain memiliki kemudahan, terdapat pula kekurangan yang berbelanja secara *online*, yaitu terkait dengan keamanan pembayaran dan kualitas barang yang tidak sesuai dengan ekspektasi [1]. Berdasarkan uraian permasalahan tersebut, diperlukan suatu media yang dapat menimbulkan tingkat kepercayaan pada konsumen lain sebelum membeli suatu produk, salah satunya adalah dengan menggunakan *review* produk.

Ulasan terhadap produk adalah hal yang sangat penting dan sekitar 73%-87% dapat mempengaruhi keputusan dalam pembelian suatu produk hanya melalui membaca ulasan produk *online* untuk produk makanan, hotel, dan juga jasa [2]. Hal tersebut dapat membawa dampak yang baik bagi konsumen dan produsen. Belanja *online* dapat membantu konsumen untuk melihat produk yang akan mereka beli memiliki *review* atau opini produk yang baik. Sedangkan keuntungan *review* produk bagi produsen adalah dapat mengelompokkan *review* pembeli ke dalam kelas negatif dan kelas positif.

Untuk memudahkan menerjemahkan *review* konsumen, *e-commerce* memanfaatkan *rating*. *Rating* yang digunakan berupa bintang dari skala 1 hingga 5 [14]. Namun dengan menggunakan *rating* saja belum cukup menjadi acuan untuk menentukan sentimen, karena terkadang terdapat ketidaksesuaian antara *rating* dan *review* yang diberikan. Maka dari itu perlu adanya pengklasifikasian teks dari *review* produk untuk menentukan suatu kalimat termasuk ke dalam sentimen positif dan negatif. Untuk mengklasifikasikan opini tersebut digunakan sebuah teknik analisis sentimen. Analisis sentimen sangat penting digunakan dalam pemasaran produk, mengukur prediksi pasar dan dapat pula digunakan dalam bidang politik.

Dalam melakukan klasifikasi sentimen terdapat salah satu hal yang perlu diperhatikan, yaitu pelabelan. Pelabelan merupakan kegiatan memberikan label pada suatu *dataset*. Metode yang sering digunakan untuk melakukan pelabelan adalah secara manual dengan menggunakan bantuan tenaga ahli bahasa [3]. Pelabelan secara manual menghasilkan data yang akurat karena manusia dapat membedakan dengan tepat suatu kalimat termasuk ke dalam suatu sentimen [4]. Namun hal tersebut sangat sulit jika dilakukan dengan menggunakan jumlah data yang besar, karena akan membutuhkan banyak ahli bahasa dan waktu yang cukup lama [3]. Metode lain yang dapat digunakan untuk pelabelan yaitu pelabelan secara otomatis. Keuntungan melakukan pelabelan secara otomatis adalah waktu yang dibutuhkan relatif sebentar. Namun pada penelitian sebelumnya yang berkaitan dengan pelabelan otomatis dan klasifikasi teks menggunakan *Support Vector Machine (SVM)* menghasilkan akurasi yang rendah, sebesar 67,83% [4]. Maka dari itu, dalam penelitian ini diusulkan metode pelabelan otomatis dengan menggunakan metode pelabelan lain, yaitu secara *Average* [5] dan secara *Binary* [6]. Pelabelan otomatis yang akan digunakan adalah melabeli teks secara manual berdasarkan acuan tertentu. Setelah melakukan pelabelan kemudian akan dilakukan klasifikasi sentimen produk termasuk ke dalam sentimen positif dan sentimen negatif dengan menggunakan pendekatan *Supervised learning*.

Pada penelitian sebelumnya yang berkaitan dengan prediksi *rating* membahas tentang penggunaan pelabelan secara *Average* pada beberapa metode seperti *Naïve Bayes* dan *SVM* [5]. Dari penelitiannya didapatkan kebenaran prediksi *rating* terbaik dengan akurasi sebesar 78,00% menggunakan metode *Naïve Bayes Classifier* [5]. Sehingga algoritma klasifikasi yang akan digunakan pada penelitian ini adalah *Multinomial Naïve Bayes Classifier*. Algoritma *Multinomial Naïve Bayes* merupakan perkembangan dari *Naïve Bayes Classifier (NBC)* yang cocok digunakan untuk menangani data teks, mudah diimplementasikan, dan memiliki *error rate* yang minimum [7].

Topik dan Batasannya

Berdasarkan latar belakang yang sudah dijelaskan, penelitian ini akan membahas topik mengenai Analisis Sentimen. Analisis sentimen atau *opinion mining* merupakan proses memahami, mengekstrak, dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu opini [9]. Sehingga analisis sentimen berfungsi untuk membantu menentukan persepsi masyarakat umum mengenai penilaian pendapat suatu objek.

Terdapat tiga rumusan masalah pada penelitian ini. Rumusan masalah yang pertama, metode pelabelan manakah yang paling tepat digunakan untuk melabeli data uji. Kedua, apakah pengaruh penggunaan *feature selection* dan tanpa *feature selection* pada akurasi sistem dengan pelabelan terbaik. Ketiga yaitu bagaimana hasil klasifikasi sistem dengan mengimplementasikan metode pelabelan dan penggunaan fitur yang terbaik dengan menggunakan Algoritma *Multinomial Naïve Bayes*.

Pada penelitian ini terdapat beberapa batasan masalah. Batasan masalah yang pertama, yaitu data uji yang digunakan adalah Data *Grocery and Gourmet Food* milik Amazon. Kedua, pada *dataset* ini akan dilabeli dengan angka 1 sebagai positif dan angka 0 sebagai negatif.

Tujuan

Terdapat tiga tujuan pada penelitian ini, antara lain :

1. Mengetahui metode pelabelan yang paling tepat digunakan untuk melabeli data uji.
2. Mengetahui pengaruh penggunaan *feature selection* dan tanpa *feature selection* pada akurasi sistem dengan menerapkan pelabelan terbaik.
3. Mengetahui hasil klasifikasi sistem dengan mengimplementasikan metode pelabelan dan penggunaan fitur yang terbaik dengan menggunakan Algoritma *Multinomial Naïve Bayes*.

Organisasi Tulisan

Selanjutnya pada bagian 2 berisi tentang studi terkait penelitian ini, pada bagian 3 membahas tentang perancangan sistem, pada bagian 4 membahas evaluasi, dan bagian 5 membahas kesimpulan dan saran.