

BAB 1 PENDAHULUAN

1.1 Latar belakang

Dengan memanfaatkan perkembangan internet saat ini, kita bisa menggunakannya untuk melihat ulasan atau tanggapan seseorang terhadap suatu hotel. Dalam dunia bisnis, bagaimana pendapat atau perasaan seorang konsumen terhadap suatu produk bisa menjadi sumber untuk pengembangan bisnis. Bagi suatu perusahaan, ulasan akan produk mereka berupa hal yang positif akan memberikan gambaran yang bagus terkait produk tersebut bagi para konsumen lain. Ulasan yang positif juga berarti produk tersebut bagus untuk dipertahankan, namun jika ulasan tersebut berupa hal yang negatif, bisa menjadi bahan untuk evaluasi produk tersebut sehingga pengembangannya bisa lebih efektif. Positif dan negatif hasil dari pendapat seseorang dikenal sebagai sentimen. Untuk membantu suatu perusahaan dalam melihat bagaimana sentimen masyarakat terhadap produknya, perlu adanya sistem yang dapat merangkum bagaimana arah ulasan tersebut, yaitu dengan analisis sentimen pada tingkat aspek. Dengan itu kita bisa tahu aspek atau fitur apa saja yang banyak diulas terkait hotel tersebut dan bagaimana sentimennya, apakah positif atau negatif. Terdapat berbagai media di internet yang menyediakan sarana kepada para pengguna internet yang ingin memberikan ulasan. Salah satunya adalah TripAdvisor. Seiring dengan semakin banyaknya data atau ulasan yang masuk, semakin sulit pula jika seseorang ingin mendapatkan rangkuman ulasan dengan cara membacanya satu persatu. Untuk itu dibuatlah sistem yang dapat melakukan analisis terhadap sentimen suatu ulasan.

Berdasarkan data yang ada di situs TripAdvisor,¹ 5 peringkat teratas untuk hotel yang berada di Malang, antara lain, Hotel Tugu Malang (569 ulasan), Kertanegara Premium Guest House (777 ulasan), Hotel Santika Premiere Malang (418 ulasan), HARRIS Hotel & Conventions Malang (952 ulasan)², Atria Hotel Malang (780 ulasan). Peringkat tersebut berdasakan dari ulasan *traveler*, dengan lima macam komponen penilaian, yaitu “Luar biasa”, “Sangat bagus”, “Rata-rata”, “Buruk”, dan “Sangat buruk”. Dengan melihat kelima hotel tersebut di atas, dipilih HARRIS Hotel karena memiliki total ulasan terbanyak. HARRIS Hotel & Conventions Malang terletak di Jl. Jend. A. Yani Utara, Perum Riverside C-1, Malang, Indonesia. Dengan ramainya ulasan hotel ini, orang yang ingin mencari penginapan di Malang mungkin akan menemukan HARRIS Hotel sebagai rekomendasi.

Analisis sentimen sendiri dapat dibagi ke dalam tiga tingkatan, yaitu tingkat dokumen, tingkat kalimat, dan tingkat aspek (Schouten dan Frasinca, 2015). Pada tingkatan dokumen tujuannya adalah menemukan sentimen untuk keseluruhan

¹ <https://www.tripadvisor.co.id/>

² https://www.tripadvisor.co.id/Hotel_Review-g297710-d3870219-Reviews-HARRIS_Hotel_Conventions_Malang-Malang_East_Java_Java.html

dokumen. Pada tingkatan kalimat, tujuannya adalah menemukan sentimen untuk suatu kalimat yang dituju atau kalimat yang mengandung sentimen. Pada tingkatan aspek, aspek adalah objek yang dibahas pada suatu kalimat atau ulasan. Di tingkat aspek ini, terdapat tiga proses yang bisa dilakukan, yaitu deteksi atau identifikasi aspek, klasifikasi sentimen, dan perangkuman hasil klasifikasi berdasarkan aspek yang ada (Schouten dan Frasinca, 2015). Namun, tidak ketiga proses tersebut harus diimplementasikan, tergantung dengan keperluan penelitian. Pada penelitian ini akan dilakukan identifikasi aspek dan klasifikasi sentimen untuk dokumen. Berdasarkan penelitian oleh Farhadloo dan Rolland (2013) tentang analisis sentimen tingkat aspek, mereka mengusulkan penggunaan *Bag of Nouns* sebagai representasi untuk kalimat sehingga nantinya dapat diklusterisasi dengan K-Means. Kemudian mereka juga mengusulkan ekstraksi fitur dengan yang mereka sebut sebagai *score representation* untuk kemudian dapat dilakukan klasifikasi dengan SVM.

Pada penelitian ini akan dilakukan klusterisasi menggunakan algoritme K-Modes yang merupakan modifikasi dari K-Means. Penggunaan K-Modes di sini dikarenakan atribut yang digunakan untuk klusterisasi bertipe kategoris sehingga akan lebih cocok saat menggunakan metode K-Modes (Huang, 1998). Setiap kluster akan merepresentasikan aspek yang mewakili semua dokumen dalam klusternya. Jadi, klusterisasi untuk menemukan aspek disini akan menghasilkan label aspek untuk setiap ulasan berdasarkan klusternya. Untuk melakukan klasifikasi sentimen, pada penelitian ini akan menggunakan algoritme LVQ. Berdasarkan penelitian yang dilakukan oleh Jamadar dan Kakade (2015), LVQ dapat menyaingi algoritme SVM pada studi kasus deteksi tumor otak dengan memberikan akurasi yang lebih tinggi. Hasil dari SVM diperoleh dari klasifikasi yang tidak intuitif (Hammer, Strickert, dan Villmann, 2004) yang mana memerlukan data latih yang besar untuk hasil yang optimal. Sebagai alternatif dari SVM, terdapat algoritme LVQ yang merupakan klasifikasi yang intuitif karena LVQ merupakan salah satu algoritme jaringan saraf tiruan yang meniru cara kerja sistem saraf yang ada di otak kita sehingga dia dapat belajar. LVQ sendiri terdapat beberapa versi salah satunya yang merupakan modifikasi pertama adalah LVQ2. Berdasarkan penelitian dari Budianita dan Firdaus (2016) tentang penggunaan LVQ2 pada diagnosis penyakit jiwa menunjukkan hasil dari LVQ2 yang memiliki *window* dapat meningkatkan nilai akurasi dari LVQ dasar yang tanpa *window*.

Dengan itu dalam penelitian ini akan dilakukan analisis sentimen tingkat aspek menggunakan metode K-Modes untuk klusterisasi dengan fitur *Bag of Nouns* kemudian LVQ2 untuk klasifikasi dengan fitur *score representation*. Data yang akan diolah adalah ulasan dari HARRIS Hotel & Conventions Malang dengan Bahasa Indonesia. Ulasan tersebut akan diklasifikasikan ke dalam kelas positif dan negatif.

1.2 Rumusan masalah

Masalah yang akan dibahas pada penelitian ini adalah:

1. Berapa jumlah kluster ideal untuk mengelompokkan sekumpulan ulasan pada hotel HARRIS Hotels & Conventions dari situs TripAdvisor dengan klusterisasi menggunakan algoritme K-Modes?
2. Bagaimana evaluasi dari algoritme LVQ2 dengan fitur *score representation* dalam melakukan analisis sentimen ulasan pada hotel HARRIS Hotels & Conventions dari situs TripAdvisor dengan parameter yang optimal?

1.3 Tujuan

Tujuan yang ingin dicapai dari penelitian ini adalah:

1. Mendapatkan jumlah kluster ideal untuk mengelompokkan sekumpulan ulasan pada hotel HARRIS Hotels & Conventions dari situs TripAdvisor dengan klusterisasi menggunakan algoritme K-Modes.
2. Mendapatkan hasil evaluasi dari algoritme LVQ2 dengan fitur *score representation* dalam melakukan analisis sentimen ulasan pada hotel HARRIS Hotels & Conventions dari situs TripAdvisor dengan parameter yang optimal.

1.4 Manfaat

Manfaat yang didapatkan dari penelitian ini adalah:

1. Menjadi bahan pertimbangan bagi hotel yang menjadi objek ulasan untuk dapat meningkatkan pelayanan atau fasilitas di hotel.
2. Menjadi bahan pertimbangan bagi calon tamu hotel untuk dapat mempertimbangkan memilih hotel tersebut.
3. Mengetahui bagaimana penerapan metode K-Modes dan metode LVQ2.

1.5 Batasan masalah

Untuk memfokuskan ruang lingkup penelitian ini, berikut beberapa batasan yang diberikan sehingga bisa lebih mudah dipahami:

1. Sumber data untuk dianalisis berasal dari ulasan yang ada di TripAdvisor untuk hotel HARRIS Hotel & Conventions Malang yang diambil pada tanggal 23 Agustus 2017.
2. Jumlah data sebanyak 277 ulasan Bahasa Indonesia.
3. Dalam mencocokkan kata tidak ada toleransi untuk kesalahan atau perbedaan huruf.
4. Metode yang digunakan adalah K-Modes dan LVQ2.
5. Pakar adalah mahasiswa S1 Fakultas Ilmu Komputer Universitas Brawijaya.

1.6 Sistematika pembahasan

Bagian ini berisi struktur skripsi ini mulai Bab Pendahuluan sampai Bab Penutup dan deskripsi singkat dari masing-masing bab. Diharapkan bagian ini dapat membantu pembaca dalam memahami sistematika pembahasan isi dalam skripsi ini.

BAB 1 PENDAHULUAN

Bab ini berisi tentang bagaimana kondisi terkini terkait topik analisis sentimen dan secara khusus tentang analisis sentimen pada tingkat aspek. Disampaikan juga tentang penelitian sebelumnya yang menjadi dasar untuk penelitian ini. Kemudian disampaikan tentang permasalahan dan hasil yang ingin didapat dalam penelitian ini.

BAB 2 LANDASAN KEPUSTAKAAN

Bab ini menjelaskan mengenai sumber penelitian yang mengawali penelitian tentang analisis sentimen pada tingkat aspek ini. Kemudian terdapat penjelasan terkait sumber pengambilan data dan hotel yang menjadi objek penelitian disini. Dijelaskan juga tentang algoritme yang akan digunakan disini.

BAB 3 METODOLOGI PENELITIAN

Bab ini menggambarkan bagaimana tahapan yang akan dilaksanakan untuk menyelesaikan penelitian dari awal hingga akhir. Mulai dari bagaimana pengumpulan data dilakukan, kemudian tahapan persiapan atau pre-processing- untuk selanjutnya di proses ke dalam klasterisasi dengan K-Modes guna mendapatkan aspek yang relevan hingga bagaimana proses klasifikasi pada LVQ2 dengan fitur *score representation* untuk dilakukan.

BAB 4 PERANCANGAN

Bab ini memberikan gambaran untuk sistem yang akan dibuat dengan memberikan rancangan sistem secara keseluruhan, kemudian rancangan penggunaan metode, hingga rancangan untuk bagaimana pengujian dilakukan. Rancangan penggunaan metode disini menunjukkan proses untuk metode yang digunakan dengan menggunakan sebagian kecil data sebagai manualisasi sehingga bisa dipahami alur metode nantinya.

BAB 5 IMPLEMENTASI

Bab ini menjelaskan proses implementasi metode yang digunakan ke dalam sistem perangkat lunak. Dijelaskan juga komponen-komponen yang terdapat untuk mendukung sistem tersebut. Kemudian, hasil implementasi tersebut seperti apa.

BAB 6 PENGUJIAN DAN ANALISIS

Bab ini menjelaskan hasil pengujian yang dilakukan terhadap penggunaan metode K-Modes dan LVQ2 dalam memberikan hasil klasifikasi yang baik terhadap ulasan yang memiliki dua kelas. Beberapa macam skenario pengujian dilakukan dengan ukuran berdasarkan *precision*, *recall*, dan *f1-score*. Kemudian disampaikan bagaimana analisisnya.

BAB 7 PENUTUP

Bab ini berisi kesimpulan dari hasil akhir pada penelitian analisis sentimen pada tingkat aspek dengan menggunakan K-Modes sebagai metode untuk mengambil aspek yang relevan dan LVQ2 sebagai metode klasifikasi dengan

memanfaatkan ekstraksi fitur *score representation*, kemudian saran untuk pengembangan selanjutnya seperti apa.