

Klasifikasi Data Ulasan Positif dan Negatif Dengan Menggunakan Algoritma Naive Bayes

Muh. Adrian Juniarta Hidayat¹, Gilang Primajati², Ahmad Zuli Amrullah³

^{1,2,3} Fakultas Teknik dan Desain, Universitas Bumigora
m.adrian@universitasbumigora.ac.id¹, gilangprimajati@universitasbumigora.ac.id²,
zuli@universitasbumigora.ac.id³

Abstrak

Bagi para pelaku penyedia jasa, kualitas pelayanan merupakan suatu hal yang sangat penting karena dapat berdampak pada keberlangsungan bisnis mereka. Salah satu cara untuk menilai kualitas suatu layanan ialah dengan melihat ulasan dari para pelanggan yang pernah menggunakan layanan tersebut. Ulasan yang berbentuk teks dapat menjadi data yang berguna apabila diolah dengan teknik tertentu untuk melihat kualitas suatu layanan dengan mengelompokkan ulasan positif dan negatif dari suatu layanan. Pada tulisan ini mengusulkan untuk melihat kualitas layanan suatu hotel dengan mengolah data ulasan yang berbentuk teks dengan membandingkan rasio ulasan positif dan negatif dari pelanggan dengan menggunakan algoritma Naive Bayes. Hasil dari percobaan terlihat bahwa pengelompokan ulasan positif dan negatif dapat dilakukan dengan baik dengan memberikan bobot ranking tertentu pada setiap kata yang terdapat didalam ulasan, sehingga dapat diketahui bahwa kualitas layanan cenderung mendapat umpan balik positif atau negatif.

Kata kunci : Text Mining, Klasifikasi, Algoritma Naive Bayes.

Abstract

For service providers, service quality is very important because it can have an impact on the sustainability of their business. One way to assess the quality of a service is to look at reviews from customers who have used the service. Reviews in the form of text can be useful data if it is processed with certain techniques to see the quality of a service by grouping positive and negative reviews of a service. In this paper proposes to see the quality of a hotel service by processing text review data by comparing the ratio of positive and negative reviews from customers using the Naive Bayes algorithm. The results of the experiment show that the grouping of positive and negative reviews can be done well by giving a certain ranking weight to each word contained in the review, so it can be seen that the quality of service tends to get positive or negative feedback.

Keywords : Text mining, Classification, Naive Bayes Algorithm.

I. PENDAHULUAN

Menilai kepuasan pelanggan dengan menganalisis ulasan yang diberikan merupakan salah satu cara mengetahui kualitas layanan bagi para penyedia jasa seperti jasa penginapan/hotel. Setiap penyedia layanan yang berbasis teknologi saat ini pasti memiliki nilai ulasan, baik yang bersifat positif ataupun negatif. Ulasan dapat berupa teks, gambar, maupun *rating* dari kualitas layanan tersebut. Hasil ulasan pelanggan yang berupa teks tersebut dapat dimanfaatkan oleh para penyedia jasa untuk melihat bagaimana tanggapan pelanggan terhadap kualitas layanan mereka, sehingga dapat di perbaiki kualitas layanan apabila mayoritas pelanggan memberikan ulasan

yang negatif. Salah satu cara untuk mengolah data ulasan berupa teks tersebut adalah dengan teknik teks mining untuk mengoleksi datanya dan dengan menggunakan algoritma Naive Bayes untuk menganalisis dan mengklasifikasikan jenis ulasan tersebut, apakah itu ulasan positif atau negatif.

Text mining merupakan suatu proses ekstraksi pola berupa informasi yang dapat di mengerti dan dimanfaatkan dari sejumlah besar sumber data berupa teks.[2] Jenis masukan data untuk proses penambangan data ini ialah data dengan bentuk tidak terstruktur atau *raw* data. Tujuan dari *text mining* ialah untuk memperoleh informasi yang berguna dari sejumlah kumpulan dokumen. Hasil dari proses ekstraksi data dan pola

ini nantinya akan dilakukan klasifikasi berdasarkan kategori dengan menggunakan metode Naive Bayes.

Algoritma Naive Bayes ialah suatu teknik klasifikasi menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan asal Inggris yakni Thomas Bayes[1]. Algoritma Naive Bayes merupakan salah satu algoritma klasifikasi yang dapat digunakan untuk megelompokkan data-data yang berupa teks dengan kategori tertentu sesuai dengan kebutuhan [1]. Metode ini memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya.

Pada penelitian ini, akan dilakukan analisis terhadap dataset dari ulasan suatu hotel yang didapat dari portal teks analytics. [7]. Hasil akhir dari eksperimen ini ialah melihat seberapa perbandingan total ulasan positif dan negatif dari kumpulan data ulasa yang diproses. Sehingga kita dapat melihat bagaimana kualitas suatu layanan hotel tersebut berdasarkan dari analisis ulasan positif dan negatif nya.

II. METODOLOGI

Pada tulisan ini, algoritma Naive Bayes diusulkan sebagai metode untuk mengelompokkan data ulasan yang berbentuk teks menjadi kelompok ulasan positif dan negatif. Dengan menggunakan metode Naive Bayes, setiap kata yang muncul dalam ulasan teks akan di kategorikan dalam 2 kondisi yakni kata positif dan negatif. Beberapa kata seperti *good*, *nice*, *like it*, *confidence* dan lain sebagainya akan masuk dalam kategori positif. Sedangkan kata-kata seperti *bad*, *hate* dan lain sebagainya akan masuk kedalam kategori negatif. Sebelum melakukan proses klasifikasi, data yang berbentuk teks terlebih dahulu dilakukan preprocessing data untuk membersihkan data dan membuat data menjadi lebih terstruktur agar dapat hasil yang akurat.

2.1. Tokenize

Yakni proses mengumpulkan semua kata yang muncul dan menghilangkan tanda baca maupun simbol apapun yang bukan huruf.

2.2. Filter Stopwords

Yakni proses menghilangkan kata penghubung yang banyak muncul namun tidak memiliki makna seperti the, of, for, dll.

2.3. Stemming

Yakni proses untuk mencari root dari sebuah kata dengan menghilangkan imbuhan pada kata tersebut.

2.4. Transform cases

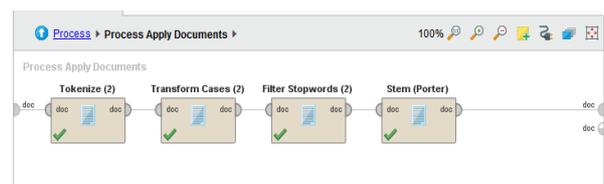
Yakni proses mengubah semua kata kedalam bentuk huruf kecil agar lebih mudah diproses.

Sedangkan untuk tahap transformasi dengan melakukan pembobotan TF-IDF pada masing-masing kata. Di mana prosesnya menghitung kehadiran atau ketidakhadiran sebuah kata di dalam dokumen. Berapa kali sebuah kata muncul di dalam suatu dokumen juga digunakan sebagai skema pembobotan dari data tekstual.

Tahapan terakhir ialah klasifikasi untuk memprediksi ulasan yang ada dan mengelompokkan berdasarkan polaritasnya masing-masing apakah bersifat positif atukah negatif dengan mempertimbangkan bobot dari tiap kata yang memiliki arti positif dan negatif.

III. HASIL DAN PEMBAHASAN

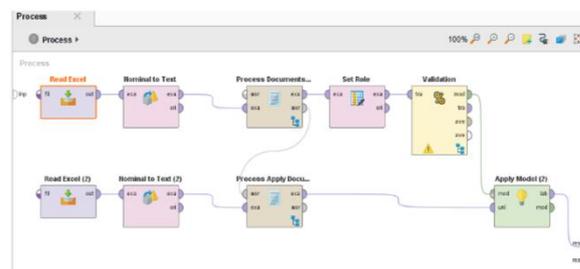
Pada bagian ini akan dipaparkan hasil percobaan pengolahan dan analisis data ulasan pelanggan. Percobaan dilakukan dengan menggunakan bantuan aplikasi Rapid Miner untuk mengolah dan mengklasifikasikan datanya. Adapun tahapan pertama yakni proses pre-processing data menggunakan aplikasi Rapid Miner dapat dilihat sebagai berikut.



Gambar 1 Pre-processing data.

Word	Attribute Name	Total Occurrences	Document Occurrences
abit	abit	1	1
abl	abl	1	1
absolut	absolut	1	1
accept	accept	1	1
ad	ad	1	1
ada	ada	1	1
add	add	1	1
adequ	adequ	5	1
align	align	1	1
allow	allow	1	1

Gambar 2 Hasil *pre-processing* data



Gambar 3 Proses klasifikasi

Pada gambar 1, merupakan proses pembersihan data meliputi proses tokenize, filter stopwords, stemming dan transform cases. Tujuan

dari proses ini ialah untuk membersihkan data agar menjadi lebih terstruktur dan dapat lebih mudah untuk dilakukan proses klasifikasinya. Hasil preprocessing data didapatkan seperti gambar dibawah ini.

Pada gambar 2, merupakan hasil dari data yang sudah dilakukan pre-processing dengan membuang semua simbol, kata penghubung dan pencarian root kata serta mengubah semua huruf menjadi huruf kecil agar lebih memudahkan untuk dilakukan proses selanjutnya.

Pada gambar 3, merupakan pemodelan proses klasifikasi pada aplikasi Rapid Miner dengan mengaplikasikan metode Naive Bayes serta dilakukan proses validasi data untuk ditampilkan hasil klasifikasi data. Adapun hasil klasifikasi data terlihat pada gambar dibawah ini.

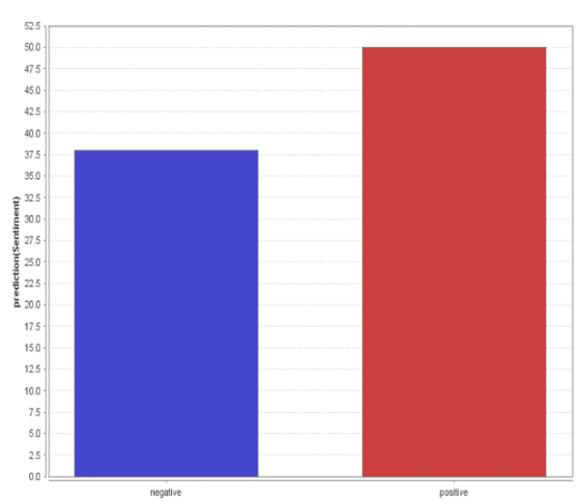
Pada gambar 4, merupakan hasil klasifikasi data berdasarkan kategori positif ataupun negatif. Dari gambar tersebut terlihat bahwa setiap kata yang diproses akan diberikan bobot nilai sesuai dengan bobot tiap kata yang menyusunnya apakah lebih banyak kata yang memiliki bobot negatif atau positif.

Row No.	prediction(S...	confidence_...	confidence_...	Sentiment	a	aaaaaah	aalfiyah	aamir	aardman	abandon
1	negative	1	0	?	0	0	0	0	0	0
2	negative	1	0	?	0	0	0	0	0	0
3	positive	0	1	?	0	0	0	0	0	0
4	negative	1	0	?	0	0	0	0	0	0
5	positive	0	1	?	0	0	0	0	0	0
6	positive	0	1	?	0	0	0	0	0	0
7	positive	0	1	?	0	0	0	0	0	0
8	negative	1	0	?	0	0	0	0	0	0
9	positive	0	1	?	0	0	0	0	0	0
10	positive	0	1	?	0	0	0	0	0	0
11	positive	0	1	?	0	0	0	0	0	0
12	positive	0	1	?	0	0	0	0	0	0
13	negative	1	0	?	0	0	0	0	0	0
14	positive	0	1	?	0	0	0	0	0	0
15	negative	1	0	?	0	0	0	0	0	0
16	positive	0	1	?	0	0	0	0	0	0

Gambar 4 Hasil klasifikasi data

Name	Type	Missing	Statistics	Filter (13,756 / 13,756 attributes)	Search for Attributes
prediction(Sentiment)	Polynomial	0	Least negative (38)	Most positive (50)	Value positive (50), negative (38)
confidence_negative	Real	0	Min 0	Max 1	Average 0.432
confidence_positive	Real	0	Min 0	Max 1	Average 0.568
Sentiment	Polynomial	0	Least ? (88)	Most ? (88)	Value ? (88)
a	Real	0	Min 0	Max 0	Average 0
aaaaah	Real	0	Min 0	Max 0	Average 0
aallyah	Real	0	Min 0	Max 0	Average 0
amir	Real	0	Min 0	Max 0	Average 0

Gambar 5. Detail data



Gambar 5 Perbandingan ulasan positif dan negatif.

Pada gambar 5, merupakan detail dari setiap data yang telah diolah didalam sistem seperti jumlah kemunculan kata, jenis kata, kategori dan nilainya. Adapun jumlah perbandingan dari ulasan positif dan negatif yang didapatkan seperti gambar dibawah ini.

Pada gambar 6, merupakan perbandingan jumlah perkiraan ulasan dengan kategori positif dan negatif dari total seluruh data yang sudah di proses. Dari gambar tersebut terlihat bawah ulasan dengan kategori positif lebih banyak dibandingkan ulasan dengan kategori negatif.

Dari semua bentuk representasi dari hasil analisis text ulasan tersebut dapat diketahui bahwa apa saja jenis komentar dari para pengguna layanan hotel, bagian kata apa saja yang paling banyak muncul sebagai komentar, apakah kata-

kata tersebut tergolong positif atau negatif dan diagram secara keseluruhan dapat dilihat secara detail. Dengan hasil analisis dari teknik teks mining ini pengelola layanan jasa dapat melihat apa saja jenis komentar yang banyak muncul sehingga dapat diketahui bagian layanan maupun fasilitas yang mana saja harus diperbaiki dan ditingkatkan. Dengan demikian, pengelola dapat berbenah pada hal-hal tertentu yang lebih spesifik yang disukai dan tidak disukai oleh pelanggan.

IV. KESIMPULAN

Dari hasil percobaan dalam mengklasifikasikan data untuk melihat perbandingan jumlah ulasan positif dan negatif menggunakan metode Naive Bayes didapatkan hasil yang baik. Algoritma Naive Bayes dapat bekerja dengan baik dan memberikan hasil yang akurat dalam proses klasifikasi data. Dari hasil tersebut dapat dilihat bahwa dengan menggunakan teknik teks mining kita dapat menganalisis data untuk mendapatkan informasi yang lebih akurat dan pasti untuk melihat tingkat kepuasan pelanggan dalam menggunakan layanan yang kita sediakan.

REFERENSI

- [1]. Shilpa Dang, Peerzada Hamid A. Text Mining: Technique and its Application, International Journal of Engineering & Technology Innovation, Vol. 1 Issue 4, 2014.
- [2]. Chen, J., Huang, H., Tian, S., & Qu, Y. (2009). Feature selection for text classification with Naïve Bayes. Expert Systems with Applications, 36, 5432–5435.
- [3]. Muthia, D. A. (2013). Analisis Sentimen Pada Review Buku Menggunakan Algoritma. Sistem Informasi, Sistem Informasi, 1–9.
- [4]. Zhang, & Gao, F. (2011). An Improvement to NB for Text Classification. Procedia Engineering, 15, 2160–2164.
- [5]. Santoso, Budi, Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu. 2007.

- [6]. Berry, M.W. &Kogan, J. 2010. Text Mining Application and theory. WILEY: United Kingdom.
- [7]. Feldman, R & Sanger, J. 2007. The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press: New York.
- [8]. http://text-analytics101.rxnlp.com/2011/07/user-review-datasets_20.html.