

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/322959527>

Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking

Article · August 2018

CITATIONS

13

READS

5,967

3 authors, including:



Muhammad Ali Fauzi
Brawijaya University

114 PUBLICATIONS 402 CITATIONS

[SEE PROFILE](#)



Sigit Adinugroho
Brawijaya University

38 PUBLICATIONS 125 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Image Processing [View project](#)



Image Processing and Computer Vision [View project](#)

Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking

Shima Fanissa¹, M. Ali Fauzi², Sigit Adinugroho³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹shimafanissa@gmail.com, ²moch.ali.fauzi@ub.ac.id, ³sigit.adinu@ub.ac.id

Abstrak

Pariwisata merupakan salah satu sektor usaha untuk memajukan suatu kota. Kota Malang pada saat ini memiliki sebuah *branding city* yang bernama “Beautiful Malang”. Masyarakat Indonesia memilih pariwisata Kota Malang sebagai destinasi dan mengulasnya pada *website* salah satunya adalah TripAdvisor. Dengan demikian penelitian ini mencoba menganalisis ulasan dari masyarakat tentang pariwisata Kota Malang melalui analisis sentimen dan diklasifikasikan menjadi dua kelas yaitu positif dan negatif. Pada penelitian ini metode yang digunakan adalah *Naive Bayes* dengan seleksi fitur *Query Expansion Ranking* untuk mengurangi jumlah fitur pada proses klasifikasi. Proses dari analisis sentimen terdiri dari *preprocessing*, seleksi fitur dengan metode *Query Expansion Ranking*, dan klasifikasi dengan *Naive Bayes*. Pengujian pada penelitian ini adalah uji akurasi dengan menggunakan variasi rasio seleksi fitur, hasilnya seleksi fitur 75% memiliki akurasi terbaik sebesar 86.6%.

Kata kunci: analisis sentimen, *query expansion ranking*, *naive bayes*

Abstract

Tourism is one of effort to promote a city. Malang currently has a branding city called "Beautiful Malang". Indonesian choose Malang tourism as a destination and review it on the website, one of them is TripAdvisor. Thus this research tried to analyze the reviews from the public about the tourism of Malang City through sentiment analysis and classified into two classes, that is positive and negative. In this research the method used is Naive Bayes with Query Expansion Ranking feature selection to reduce the number of features in the classification process. The process of sentiment analysis consists of preprocessing, feature selection with Query Expansion Ranking method, and classification with Naive Bayes. This research is testing the accuracy by using the variation of feature selection ratio, the result of 75% feature selection has the best accuracy of 86.6%.

Keywords: sentiment analysis, *query expansion ranking*, *naive bayes*

1. PENDAHULUAN

Word of Mouth yang diketahui sebagai penyampaian informasi dari mulut ke mulut (orang ke orang) adalah sebuah metode yang digunakan oleh konsumen untuk mengungkapkan perasaan dan pengalaman mereka mengenai sebuah produk, merek, atau layanan namun metode tersebut tidak bertujuan untuk memasarkan atau mempromosikan (Hung, 2017). Dengan berbagi dan menyebarkan dokumen *Word of Mouth*, masyarakat luas dapat menggunakan dokumen *Word of Mouth* tersebut sebagai referensi dan pengambilan keputusan. Terlebih lagi pada era sekarang

dimana pertumbuhan internet dan media sosial telah berkembang pesat sehingga menyediakan berbagai informasi opini orang lain.

Analisis Sentimen yaitu studi yang terdiri dari *Natural Language Processing*, komputasi linguistik, dan analisis teks untuk mengidentifikasi sentimen teks (Vinodhini dan Chandrasekaran, 2016) dapat membantu untuk mengidentifikasi opini mengenai suatu produk yang disampaikan oleh pengguna online untuk mengekskresikan emosi, perilaku, atau secara otomatis. Analisis sentimen biasanya terbagi menjadi dua kelas yaitu positif dan negatif.

Malang yaitu sebuah kota di yang terletak di Jawa Timur Indonesia sedang melakukan usaha untuk memajukan sektor pariwisatanya dengan membuat sebuah *branding city* bernama “Beautiful Malang” dengan tujuan Kota Malang dapat diposisikan menjadi kota yang indah, nyaman, ramah huni, tujuan wisata, mempunyai banyak *heritage* dan tempat kuliner (Aminudin, 2016). Maka dari itu penting bagi pemerintah kota dan masyarakat untuk mengetahui apa sentimen pariwisata di Kota Malang, sentimen tersebut bisa digunakan untuk pengambilan keputusan mengenai pariwisata Kota Malang.

Ulasan atau dokumen *Word of Mouth* mengenai pariwisata di Kota Malang jumlahnya sangat banyak di internet, karena semua orang bebas untuk menuliskan pendapat mereka. Maka dari itu dipilihlah Trip Advisor sebagai sumber data dalam penelitian ini karena TripAdvisor adalah *website* yang menyediakan berbagai informasi mengenai pariwisata yang meliputi restoran, objek wisata, tempat penginapan, dll.

Untuk membuat sebuah analisis sentimen banyak hal yang harus dipersiapkan terlebih dahulu, salah satunya dengan memilih *classifier* yang akan digunakan. *Classifier* metode yang dapat melakukan klasifikasi data menjadi beberapa kelas. Dalam penelitian ini, *classifier* yang dipilih adalah *Naive Bayes*, sebuah metode sederhana yang dikembangkan berdasarkan aturan Bayes dengan melihat kondisi-kondisi yang ada dan peluang-peluang setiap kondisinya.

Penelitian sebelumnya mengenai analisis sentimen dengan menggunakan *Naive Bayes* sebagai *classifier*-nya telah banyak dilakukan, diantaranya adalah penelitian menggunakan data twitter berjudul “Analisis Sentimen Tentang Opini Film pada Dokumen Twitter Berbahasa Indonesia Menggunakan *Naive Bayes* dengan Perbaikan Kata Tidak Baku” oleh (Antinasari et al., 2017). Penelitian ini menganalisa opini twitter dan membagginya

menjadi dua kelas yaitu positif dan negatif yang menghasilkan akurasi hingga 91.67%. Selanjutnya, terdapat penelitian analisis sentimen pada *review* aplikasi *mobile* yang dilakukan oleh (Firmansyah et al., 2016) berjudul “Sentiment Analysis Pada Review Aplikasi Mobile Menggunakan Metode *Naive Bayes* Dan *Query Expansion*” hasil penelitian ini menghasilkan akurasi 95% dan 98% jika tanpa *Query Expansion*.

Penelitian mengenai analisis sentimen dengan *Naive Bayes* juga diterapkan pada *Big Data* berjudul “*Scalable Sentiment Classification for Big Data Analysis Using Naive Bayes Classifier*” oleh (Liu et al., 2013). Hasil penelitian ini dapat berjalan dengan baik pada *Big Data* dan menghasilkan akurasi sebesar 82%.

Pada dasarnya, analisis sentimen merupakan klasifikasi teks yang membagi dokumen menjadi dua kelas yaitu positif dan negatif yang dilakukan dengan menghitung nilai setiap fiturnya. Dengan begitu, tingginya dimensi fitur pada proses klasifikasi menjadi masalah pada analisis sentimen. Oleh karena itu, pada beberapa penelitian digunakan seleksi fitur untuk mengurangi jumlah fitur dalam proses klasifikasi.

Seleksi fitur ini berguna untuk mereduksi fitur-fitur sehingga proses klasifikasi bisa lebih efektif dan efisien. Pada penelitian ini, metode seleksi fitur yang dipilih adalah *Query Expansion Ranking* (QER) yang pernah diterapkan dalam penelitian (Parlar dan Özel, 2016) berjudul “*A New Feature Selection Method for Sentiment Analysis of Turkish Reviews*”. Penelitian ini menerapkan seleksi fitur QER untuk analisis sentimen empat kategori. Setelah itu, metode QER dibandingkan dengan metode seleksi fitur lain yaitu *Chi-Square* dan *Document Frequency Difference* dan hasilnya QER menghasilkan akurasi terbaik dengan akurasi tertinggi 91%.

2. ANALISIS SENTIMEN

Analisis sentimen adalah bidang interdisipliner, sebuah bidang dimana

pendekatan pemecahan masalahnya dengan menggunakan tinjauan dari berbagai sudut pandang ilmu serumpun secara relevan dan terpadu. Analisis sentimen terdiri dari pemrosesan bahasa alami, analisis teks dan komputasi linguistik untuk mengidentifikasi sentimen dari suatu dokumen (Vinodhini dan Chandrasekaran, 2016).

3. MULTINOMIAL NAIVE BAYES

Multinomial Naive Bayes merupakan sebuah metode yang bekerja dengan cara menghitung frekuensi setiap term pada dokumen (McCallum dan Nigam, 1998). Sebagai contoh, frekuensi kata “jaringan” pada berita teknologi. Sehingga peran tokenisasi dalam Multinomial Naive Bayes ini sangat penting.

Dalam *Multinomial Naive Bayes*, dokumen urutan kejadian munculnya kata dalam dokumen tidak dipedulikan, jadi dokumen dianggap seperti “*bag of words*”, sehingga setiap kata diolah menggunakan distribusi multinomial. Secara umum persamaan *Multinomial Naive Bayes* adalah sebagai berikut:

$$P(c|d) = P(c) \prod_{i=1}^n P(w_i|c) \quad (1)$$

$P(c|d)$ adalah peluang kelas c berdasarkan dokumen d , n adalah jumlah seluruh kata pada dokumen

$$P(c) = \frac{N_c}{N} \quad (2)$$

$P(c)$ adalah peluang kelas c , c adalah kelas N_c adalah jumlah dokumen kelas c , N adalah jumlah seluruh dokumen.

$$P(w_i|c) = \frac{\text{count}(w_i, c) + 1}{\text{count}(c) + |V|} \quad (3)$$

$P(w_i|c)$ adalah peluang kata ke- i pada kelas c , $\text{count}(w_i, c)$ adalah jumlah kata ke- i pada kelas c , $\text{count}(c)$ adalah jumlah seluruh kata pada kelas c , $|V|$ adalah jumlah kata unik pada semua kelas.

4. QUERY EXPANSION RANKING

Query Expansion Ranking adalah sebuah metode seleksi fitur yang berguna untuk mengurangi kompleksitas komputasi tanpa mengurangi kualitas dari analisis sentimen (Fauzi et al., 2017). *Query Expansion Ranking* terinspirasi dari metode *Query Expansion* yang berguna untuk meningkatkan kualitas *query* yang dimasukkan oleh pengguna kemudian

digabung dengan cara *probabilistic weighting model* untuk memberi skor pada setiap fitur (Parlar dan Özel, 2016). Persamaan dari QER sebagai berikut:

$$\text{Score}_f = \frac{p_f + q_f}{p_f - q_f} \quad (4)$$

Score_f adalah skor atau nilai QER, p_f adalah peluang fitur f dalam dokumen kelas positif, q_f adalah peluang fitur f dalam dokumen kelas negatif.

Nilai-nilai diatas dihitung berdasarkan persamaan 5 dan 6 berikut.

$$p_f = \frac{df_+^f + 0.5}{n^+ + 1.0} \quad (5)$$

$$q_f = \frac{df_-^f + 0.5}{n^- + 0.5} \quad (6)$$

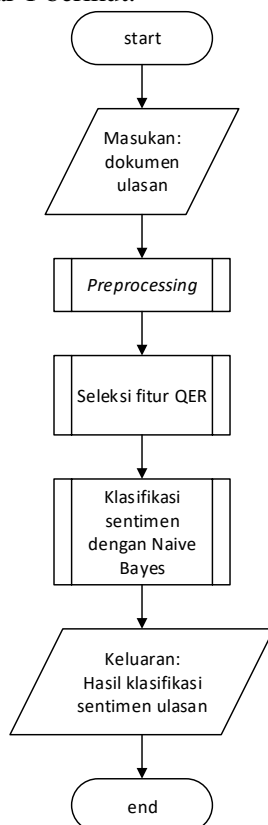
df_+^f adalah jumlah dokumen positif yang mengandung fitur f , df_-^f adalah jumlah dokumen negative yang mengandung fitur f , n^+ adalah jumlah dokumen positif, n^- adalah jumlah dokumen negatif.

5. PERANCANGAN SISTEM

Sistem yang akan dibuat pada penelitian Analisis Sentimen Pariwisata di Kota Malang menggunakan Metode *Naive Bayes* dan seleksi fitur *Query Expansion Ranking* adalah sistem yang mampu memuat dokumen dan melakukan serangkaian proses sehingga dokumen tersebut bisa diklasifikasikan menjadi dua kelas, positif atau negatif. Sistem ini memiliki beberapa tahap dari awal hingga selesai, yaitu:

1. *Preprocessing* terdiri dari tokenisasi, *filtering*, dan *stemming*. Tokenisasi untuk memisahkan dokumen menurut tokennya. *Filtering* yaitu membuang kata-kata yang tidak berguna dalam proses klasifikasi. *Stemming* yaitu merubah setiap kata ke dalam kata dasarnya, pada penelitian ini proses *stemming* menggunakan *library* Sastrawi.
2. Seleksi fitur menggunakan metode *Query Expansion Ranking* yang bertujuan untuk mengurangi jumlah fitur sehingga dapat meningkatkan akurasi.
3. Klasifikasi menggunakan *Naive Bayes*.

Diagram alir dari sistem ini ditunjukkan dengan Gambar 1 berikut.



Gambar 1. Diagram alir sistem

6. PENGUJIAN DAN ANALISIS

Tabel 1. Hasil akurasi sistem

Fitur	akurasi
25%	83.3%
50%	83.3%
75%	86.6%
100%	83.3%

Pengujian pada penelitian ini menggunakan pengujian seleksi fitur dengan mengganti rasio seleksi fitur antara 0-100 menggunakan jumlah data latih dan data uji pada setiap pengujian yang sama, yaitu 200 data latih (100 data positif dan 100 data negatif) dan data uji sejumlah 30 dokumen.

Semua percobaan variasi rasio seleksi fitur dilihat dan dicatat perubahan akurasi kemudian dianalisis. Untuk hasil dari pengujian ini dapat dilihat pada Tabel 1.

7. KESIMPULAN

Metode *Multinomial Naive Bayes* dapat diterapkan pada proses analisis sentimen pariwisata Malang menggunakan data latih dari

ulasan-ulasan pada website TripAdvisor. Data latih juga harus dilabeli kelas sentimennya oleh pakar untuk proses klasifikasi. Tahap dalam analisis sentimen penelitian ini adalah preprocessing (tokenisasi, *filtering*, *stemming*), seleksi fitur *Query Expansion Ranking*, dan *Naive Bayes*.

Pemberian seleksi fitur dalam proses analisis sentimen bertujuan untuk mengurangi dimensi fitur, dan metode *Query Expansion Ranking* bekerja dengan baik bersama metode *Multinomial Naive Bayes*. Berdasarkan pengujian, algoritma *Query Expansion Ranking* menghasilkan akurasi tertinggi sebesar 86.6 pada seleksi fitur 75%.

Dalam penyempurnaan penelitian ini maka penelitian selanjutnya disarankan untuk memperhatikan singkatan, gabungan dua kata atau lebih (bigram, trigram, n-gram), kata-kata ambigu, dan kalimat sarkastik supaya hasil klasifikasinya lebih sempurna.

8. DAFTAR PUSTAKA

- Aminudin, M. 2016. *Beautiful Malang Branding City Kota Menjual Keindahan*. [News] detiknews. Tersedia di: <<http://news.detik.com/berita-jawa-timur/3207407/beautiful-malang-branding-city-kota-menjual-keindahan>> [Diakses 17 Mei 2017].
- Antinasari, P., Perdana, R. & Fauzi, M. 2017. Analisis Sentimen Tentang Opini Film pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naive Bayes dengan Perbaikan Kata Tidak Baku. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIHK)*, 1: 1733–1741.
- Fauzi, M.A., Arifin, A.Z., Gosaria, S.C. & Prabowo, I.S. 2017. Indonesian News Classification Using Naïve Bayes and Two-Phase Feature Selection Model. *Indonesian Journal of Electrical Engineering and Computer Science*, 8(3): 610–615.
- Firmansyah, R.F.N., Fauzi, M. & Afirianto, T. 2016. SENTIMENT ANALYSIS PADA REVIEW APLIKASI MOBILE MENGGUNAKAN METODE NAÏVE BAYES DAN QUERY EXPANSION. *DORO PTIHK*, 8.
- Hung, C. 2017. Word of mouth quality

- classification based on contextual sentiment lexicons. *Information Processing & Management*, 53(4): 751–763.
- Liu, B., Blasch, E., Chen, Y., Shen, D. & Chen, G. 2013. Scalable sentiment classification for Big Data analysis using Naive Bayes Classifier. *2013 IEEE International Conference on Big Data*. 2013 IEEE International Conference on Big Data. hal.99–104.
- McCallum, A. & Nigam, K. 1998. A Comparison of Event Models for Naive Bayes Text Classification. *Learning for Text Categorization: Papers from the 1998 AAAI Workshop*. hal.41–48.
Tersedia di:
<<http://www.kamalnigam.com/papers/multinomial-aaaiws98.pdf>>.
- Parlar, T. & Özel, S.A. 2016. A new feature selection method for sentiment analysis of Turkish reviews. *2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*. 2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA). hal.1–6.
- Vinodhini, G. & Chandrasekaran, R.M. 2016. A comparative performance evaluation of neural network based approach for sentiment classification of online reviews. *Journal of King Saud University - Computer and Information Sciences*, 28(1): 2–12.