

Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi menggunakan Algoritma K-Nearest Neighbor (K-NN)

Retno Sari

Teknik Informatika, STMIK Nusa Mandiri
Indonesia

E-mail:retno.rnr@nusamandiri.ac.id

Abstract

Reviews of tourist attractions make it easier for other visitors to know the state of the place and become input for other visitors. Sentiment analysis is needed to make the best decision. This study aims to determine the accuracy of sentiment analysis using the K-Nearest Neighbor algorithm, which reviews using Indonesian language texts. The advantages of k-NN have a simple principle, working based on the shortest distance from the test sample and training sample. In this research, preprocessing used is tokenize, stopword filter and bi-gram. The data used consisted of 50 positive reviews and 50 negative reviews. Experiments have been carried out by changing the value of k and obtained the greatest accuracy with a value of k = 7 with an accuracy of 77.01, precision 92.38, recall 61.56 and AUC value of 0.894.

Keywords: sentiment analysis; Review of attractions; k-NN

Abstrak

Review mengenai tempat wisata memudahkan pengunjung lain untuk mengetahui keadaan tempat tersebut dan menjadi masukan untuk pengunjung lain. Analisis sentimen diperlukan untuk mengambil keputusan yang terbaik. Penelitian ini bertujuan untuk mengetahui akurasi analisis sentimen dengan menggunakan algoritma K-Nearest Neighbor yang mana review menggunakan teks berbahasa Indonesia. Kelebihan k-NN memiliki prinsip sederhana, bekerja berdasarkan jarak terpendek dari sampel uji dan sampel latih. Pada penelitian ini *preprocessing* yang digunakan yaitu *tokenize*, *filter stopword* dan *bi-gram*. Data yang digunakan terdiri dari 50 review positif dan 50 review negative. Eksperimen telah dilakukan dengan mengubah nilai k dan didapat akurasi yang terbesar dengan nilai k=7 dengan akurasi 77.01, precision 92.38, recall 61.56 dan nilai AUC 0.894.

Kata kunci : analisis sentiment; review objek wisata; k-NN

1. Introduction

Banyaknya review mengenai tempat wisata yang telah dikunjungi, memudahkan pengunjung lain yang ingin berkunjung ke

tempat tersebut mengetahui keadaan di tempat tersebut dan menjadikan masukan untuk pengunjung lain untuk memutuskan ke tempat tersebut atau tidak.

Pengelolaan analisis sentimen menjadi suatu informasi untuk sesama pengunjung ataupun pelaku bisnis, untuk pelaku bisnis informasi yang didapat sangat berguna untuk menjalankan bisnisnya (Fitrianti et al., 2019).

Analisis sentimen merupakan penilaian publik, penilaian mengenai kesukaan atau ketidaksukaan (Romadloni et al., 2019). Tujuan dari teks mining adalah memproses *unstructured* data (tekstual) guna dicari pola makna serta ditindaklanjuti dengan pengambilan keputusan yang terbaik (Ipmawati et al., 2017).

Terdapat penelitian mengenai analisis sentimen sebelumnya menggunakan algoritma K-Nearest Neighbor (K-NN). Penelitian sebelumnya yaitu penerapan algoritma K-Nearest Neighbor pada sentimen review agen travel (Ernawati & Wati, 2018), Analisis sentimen kurikulum 2013 pada Twitter menggunakan *Ensemble Feature* dan metode K-Nearest Neighbor (Irfan et al., 2018).

Algoritma K-Nearest Neighbor (K-NN) merupakan metode *lazy learning* dimana tidak ada model yang dipelajari dari data testing, jadi hanya belajar dari contoh uji harus diklasifikasikan. Algoritma K-Nearest Neighbor (K-NN) bertujuan untuk mengklasifikasikan objek berdasarkan atribut dan data *training* (Salam et al., 2018).

Dari hasil penelitian sebelumnya yang berjudul penerapan algoritma K-Nearest Neighbor pada sentimen review agen travel didapatkan akurasi sebesar 87.00%, pada penelitian tersebut menggunakan bahasa Inggris dengan sentiment *fast, good, great, bad, cencel* dan *wait* (Ernawati & Wati, 2018).

Penelitian dengan judul analisis sentimen pada review objek wisata dunia fantasi menggunakan algoritma K-Nearest Neighbor bertujuan untuk mengetahui akurasi yang dihasilkan.

2. Materials and Methods

2.1. Materials

2.1.1. Text Mining

Teks merupakan sesuatu yang umum dalam melakukan pertukaran informasi. Syarat umum data dan teks mining adalah informasi yang diambil dan dapat menjadi data yang berguna. Teks mining merupakan proses menganalisa teks untuk menjadi informasi yang berguna untuk tujuan tertentu. Informasi yang diambil harus jelas dan eksplisit, karena teks mining merubah menjadi bentuk yang dapat digunakan oleh komputer atau orang yang tidak memiliki waktu untuk membaca full teks (Bramer, 2007).

Metode teks mining melihat format dokumen dan presentasi standar untuk

belajar yaitu format XML yang digunakan untuk dokumen (M.Weiss et al., 2010).

2.1.2. Analisis Sentimen

Analisis sentimen menentukan terhadap komentar mengenai permasalahan, apakah masalah tersebut memiliki kecenderungan positif atau negatif dan dapat dijadikan acuan untuk meningkatkan pelayanan atau kualitas produk (Nurjanah et al., 2017).

Sentimen analisis adalah riset komputasional dari opini, sentimen dan emosi yang di ekspresikan secara tekstual (Zulfa & Winarko, 2017).

Analisis sentimen merupakan penelitian dalam bentuk analisis, mengenai opini dan emosi banyak orang (Pravina et al., 2019).

2.1.3. K- Nearest Neighbor

Metode K-Nearest Neighbor (k-NN) pertama kali di perkenalkan awal tahun 1960. Metode ini bekerja intensif ketika diberikan data training yang besar dan tidak populer sampai dengan tahun 1960. Ketika tekin komputasi sedang berkembang (Han & Kamber, 2006).

Metode k-NN merupakan model supervised learning yang mana database populasi yang digunakan diidentifikasi terlebih dahulu ke dalam kelas tertentu (Nugraha et al., 2014).

k-NN memiliki prinsip sederhana, bekerja berdasarkan jarak terpendek dari

sampel uji ke sampel latih. Sedangkan kekurangan dari k-NN yaitu *computational cost* dan sensitif terhadap derau dan input yang tidak sesuai (Budianto et al., 2018).

2.1.4. Teks Preprocessing

Teks preprocessing merupakan tahap awal dari text mining, proses ini untuk mempersiapkan data (Rustiana & Rahayu, 2017).

2.1. Methods

Metode yang digunakan pada penelitian ini merupakan metode penelitian eksperimen, dengan tahapan sebagai berikut:

a. Pengumpulan Data

Penulis mengambil data dari situs review www.tripadvisor.com. Data yang diambil terdiri dari 50 review positif dan 50 review negative.

b. Pengolahan Awal Data

Preprocessing pada dataset yang dilakukan yaitu:

1) *Tokenize*

Tokenize merupakan tahap *preprocessing* dengan menghilangkan tanda baca maupun simbol apapun yang bukan huruf.

2) *Filter Stopword*

Pada tahap *filter stopwords* membuang kata-kata berupa kata sambung,

imbuan, akhiran. Seperti kata “dari”, “yang”, “di”, dan “ke”.

3) N-gram

Pada tahap ini dilakukan pemotongan karakter dengan bi-gram yaitu pemotongan string sebanyak 2 bagian.

c. Metode yang digunakan

Pada penelitian ini menggunakan metode K-Nearest Neighbor (k-NN) untuk mengetahui akurasi yang dihasilkan.

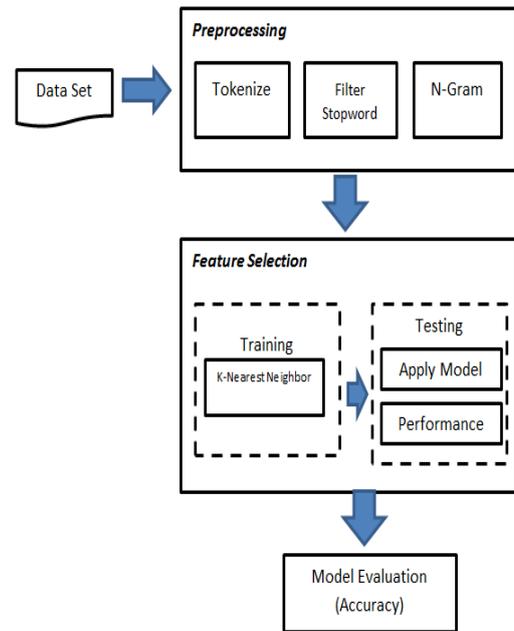
d. Eksperimen dan Pengujian Metode

Eksperimen pada penelitian ini menggunakan Rapid Miner 5 dalam pengolahan data.

e. Evaluasi dan Validasi Hasil

Validasi dilakukan menggunakan *fold cross validation*, untuk melakukan pengukuran akurasi menggunakan *confusion matrix* dan kurva ROC untuk mengukur nilai AUC.

Berikut adalah gambar model dari penelitian ini:



Gambar 1. Model Penelitian

3. Results and Discussion

Pada penelitian ini menggunakan data dari www.tripadvisor.com yang terdiri dari 50 data review positif dan 50 data review negative dengan teks berbahasa Indonesia. Data review tersebut dikelompokkan ke dalam folder positif untuk review positif dan dikelompokkan ke dalam folder negative untuk review negative.

Proses pengolahan awal terdiri dari proses *tokenize*, *filter stopwords* dan *n-gram*. Pada proses *tokenize* semua review dihilangkan tanda baca maupun simbol apapun yang bukan huruf. Sehingga membentuk kumpulan kata yang memiliki arti. Pada tabel 1 dapat dilihat hasil dari proses *tokenize*.

Pada penelitian ini pengujian dilakukan dengan model *k fold cross validation* dimana proses ini membagi data secara acak kedalam k bagian. Proses pengujian dimulai dari pembentukan model dengan data pada bagian pertama. Setelah dilakukan training dan testing maka dapat di ukur akurasi.

Pengujian metode pada algoritma K Nearest Neighbor. Berikut adalah hasil dari eksperimen pengujian metode K –Nearest Neighbor dengan *fold cross validation*.

Tabel 4. Nilai Akurasi pada K Nearest Neighbor

Nilai K	Accuracy	Precision	Recall	AUC
1	71.88	71.13	77.88	0.927
2	64.77	74.44	52.05	0.868
3	70.77	73.87	67.86	0.877
4	75.92	91.67	58.07	0.875
5	73.84	78.45	65.48	0.881
6	74.94	89.52	57.84	0.894
7	77.01	92.38	61.56	0.894
8	73.56	86.25	58.57	0.894
9	76.09	89.81	59.63	0.910
10	76.00	91.00	60.00	0.912

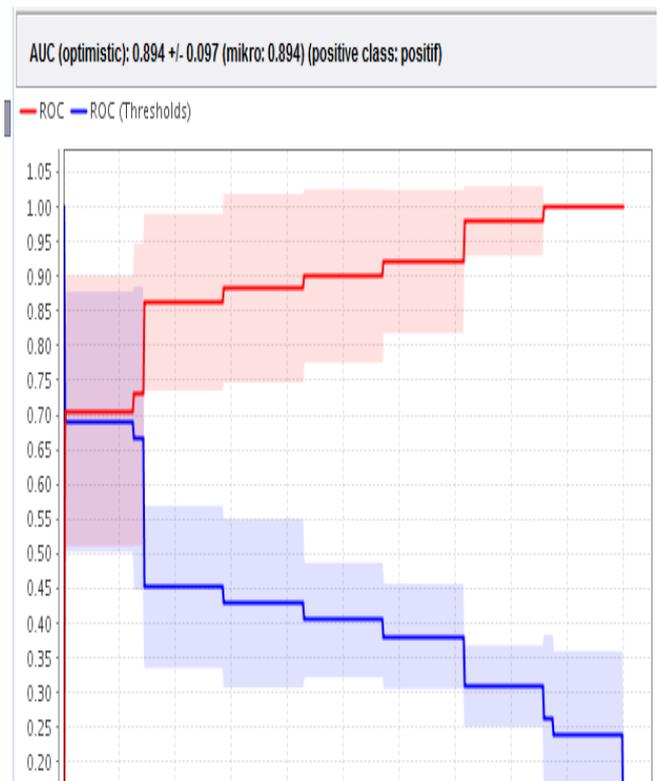
Berdasarkan tabel akurasi dengan K-Nearest Neighbor dengan mengubah nilai K. Pada tabel diatas dapat dilihat akurasi yang nilai K nya 7 memiliki akurasi yang paling besar yaitu 77.01%.

Confusion matrix dengan menggunakan algoritma *K-Nearest Neighbor* dapat dilihat dari tabel 5.

Tabel 5. *Confusion Matrix K-Nearest Neighbor*

Accuracy : 77.01% +/-7.00%(mikro:77.00%)		
	True negative	True positif
Pred negative	46	19
Pred positif	4	31

Grafik ROC dari eksperimen menggunakan K-Nearest Neighbor dapat dilihat pada gambar 2.



Gambar 2. Grafik ROC K-Nearest Neighbor

4. Conclusions

Pada penelitian ini review taman rekreasi dunia fantasi menggunakan K-Nearest Neighbor dengan teks berbahasa

Indonesia menghasilkan akurasi sebesar 77.01% dengan nilai K 7 menggunakan 10 fold cross validation.

5. References

- Bramer, M. (2007). *Principles of Data Mining*. Springer.
- Budianto, A., Maryono, D., & Ariyuana, R. (2018). Perbandingan K-Nearest Neighbor (KNN) Dan Support Vector Machine (SVM) Dalam Pengenalan Karakter Plat Kendaraan Bermotor. *Jurnal Ilmiah Pendidikan Teknik Kejuruan (JIPTEK)*, 11(1), 42–52. <https://doi.org/10.20961/jiptek.v11i1.21088>
- Ernawati, S., & Wati, R. (2018). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel. *Jurnal Khatulistiwa Informatika*, VI(1), 64–69.
- Fitrianti, R. P., Kurniawati, A., & Agusten, D. (2019). Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi) 2019*, 27–32.
- Han, J., & Kamber, M. (2006). *Data Mining Concepts dan Technique*. Elsevier.
- Ipmawati, J., Kusriani, & Taufiq Luthfi, E. (2017). Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen. *Indonesian Journal on Networking and Security*, 6(1), 28–36.
- Irfan, M. R., Fauzi, M. A., & Tibyani. (2018). Analisis Sentimen Kurikulum 2013 pada Twitter menggunakan Ensemble Feature dan Metode K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(9), 3006–3014. <http://j-ptiik.ub.ac.id>
- M.Weiss, S., Indurkha, N., & Zhang, T. (2010). *Texts in Computer Science - Fundamentals of Predictive Text Mining*. Springer.
- Nugraha, K. A., Hapsari, W., & Haryono, N. A. (2014). Analisis Tekstur Pada Citra Motif Batik Untuk Klasifikasi K-NN. *Informatika*, 10(2), 135–140.
- Nurjanah, W. E., Perdana, R. S., & Fauzi, M. A. (2017). Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya*, 1(12), 1750–1757. <https://doi.org/10.1074/jbc.M20949820>

- 0
- Pravina, A. M., Cholissodin, I., & Adikara, P. P. (2019). Analisis Sentimen Tentang Opini Maskapai Penerbangan pada Dokumen Twitter Menggunakan Algoritme Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIHK) Universitas Brawijaya*, 3(3), 2789–2797.
- Romadloni, N. T., Santoso, I., & Budilaksono, S. (2019). PERBANDINGAN METODE NAIVE BAYES , KNN DAN DECISION TREE TERHADAP ANALISIS SENTIMEN TRANSPORTASI KRL. *Jurnal IKRA-ITH Informatika*, 3(2), 1–9.
- Rustiana, D., & Rahayu, N. (2017). Analisis Sentimen Pasar Otomotif Mobil: Tweet Twitter Menggunakan Naïve Bayes. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 8(1), 113–120. <https://doi.org/10.24176/simet.v8i1.841>
- Salam, A., Zeniarja, J., Septiyan, R., Khasanah, U., Studi, P., Informatika, T., Komputer, F. I., Nuswantoro, U. D., & Analysis, S. (2018). ANALISIS SENTIMEN DATA KOMENTAR SOSIAL MEDIA FACEBOOK DENGAN K-NEAREST NEIGHBOR (STUDI KASUS PADA AKUN JASA. *SINTAK 2018*, 480–486.
- Zulfa, I., & Winarko, E. (2017). Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 11(2), 187. <https://doi.org/10.22146/ijccs.24716>