

Analisa Sentimen Pengunjung Hotel Dengan K-Nearest Neighbor Studi Kasus Hotel Pop! Surabaya

¹Kevin Permana, ²Zuda Pradana Putra, ³Aryo Nugroho
^{1,2,3}Universitas Narotama

Alamat Surat

Email: kevinpermana17@fasikom.narotama.ac.id,
zudapradanaputra.17@fasikom.narotama.ac.id, aryo.nugroho@narotama.ac.id

Article History:

Diajukan: 9 Juli 2021; Direvisi: 30 Agustus 2021; Diterima: 03 September 2021

ABSTRAK

Keberadaan tripadvisor telah banyak digunakan masyarakat luas untuk mencari dan memesan hotel, salah satu hotel yang sering di pesan adalah hotel POP! Surabaya. Hotel yang dinilai merupakan hotel yang banyak orang tahu dan memiliki harga yang terjangkau sehingga tidak sedikit wisatawan dari luar maupun dalam kota yang memesan hotel tersebut. Oleh karena itu, penelitian ini mencoba untuk menganalisis hotel POP! Surabaya melalui ulasan berbahasa Indonesia yang memiliki banyak ulasan antara negatif dan positif. Pada penelitian ini terdiri dari beberapa tahap. Tahap pertama adalah proses pre-processing yang terdiri dari proses case folding, tokenisasi, stemming, normalisasi kata, dan stopword. Selanjutnya pada tahap kedua adalah dengan melakukan pembobotan kata menggunakan metode TF-IDF. Pada tahap terakhir yaitu melakukan klasifikasi dengan metode klasifikasi yang berisi tentang ulasan tertentu. Metode klasifikasi yang dilakukan dalam penelitian ini adalah K-Nearest Neighbor. Hasil yang diperoleh dari analisis sentimen terhadap hotel POP! Surabaya menggunakan algoritma K-Nearest Neighbor memperoleh hasil akurasi optimal pada nilai k=5 sampai dengan k=9 yaitu sebesar 93,3%.

Kata kunci: Klasifikasi; Hotel POP! Surabaya; TF-IDF; K-Nearest Neighbor

ABSTRACT

The existence of tripadvisor has been widely used by the public to find and book hotels, one of the hotels that is often ordered is the POP! Surabaya. The hotel that is assessed is a hotel that many people know and has an affordable price so that not a few tourists from outside and within the city book the hotel. Therefore, this study tries to analyze the hotel POP! Surabaya through reviews in Indonesian which have many reviews between negative and positive. This research consists of several stages. The first stage is the pre-processing process which consists of case folding, tokenization, stemming, word normalization, and stopword processes. Then in the second stage is to do word weighting using the TF-IDF method. The last stage is to classify with a classification method that contains certain reviews. The classification method used in this study is K-Nearest Neighbor. The results obtained from the sentiment analysis of the hotel POP! Surabaya using the K-Nearest Neighbor algorithm obtains the highest accuracy when using neighbor values of k=5 to k=9, which is 93.3%.

Keywords: Classification; Hotel POP! Surabaya; TF-IDF; K-Nearest Neighbor.

1. PENDAHULUAN

1.1 Latar Belakang

Saat ini perkembangan teknologi informasi sangat kompleks dan memberi banyak dampak bagi kehidupan manusia, termasuk pada website yang memuat mengenai objek wisata yang ada di Indonesia, dan salah satunya adalah Tripadvisor penyedia jasa layanan pemesanan hotel. Hotel adalah bangunan yang dikelola secara komersial untuk memberikan fasilitas penginapan kepada masyarakat umum dengan fasilitas di antaranya pelayanan barang bawaan, makanan dan minuman, fasilitas perabot dan hiasan yang ada di dalamnya, serta jasa pencucian pakaian, salah satu hotel termurah dan memiliki layanan yang bagus adalah Hotel POP! Surabaya, hotel tersebut memiliki beberapa keunggulan yaitu salah satu hotel termurah yang ada di Surabaya dan salah satu hotel yang ramah lingkungan.

Analisis sentimen sering juga disebut sebagai opinion mining. Saat ini, pengguna sangat senang mengekspresikan perasaannya melalui platform online, seperti media sosial, e-commerce, dan website. Oleh karena itu, sentiment analysis dilakukan di platform-platform tersebut. Salah satu keunggulan dari analisis sentimen adalah menghemat waktu dan tenaga. Kini, analisis sentimen dapat dilakukan secara otomatis

K-Nearest Neighbor adalah salah satu metode untuk pengambilan keputusan yang dimana hasil dari data masukan yang baru diklasifikasi berdasarkan terdekat dalam data nilai. Algoritma *K-Nearest Neighbor* atau disingkat K-NN adalah sebuah metode untuk melakukan klasifikasi terhadap objek yang berdasarkan dari data pembelajaran yang jaraknya paling dekat dengan objek tersebut. K-NN termasuk juga dalam algoritma supervised learning yang nantinya hasil dari query instance yang akan diklasifikasi berdasarkan mayoritas dari kategori pada algoritma KNN. Dimana kelas yang paling banyak muncul nanti akan menjadi kelas hasil dari suatu klasifikasi yang nantinya juga akan diterapkan dalam penelitian ini. Berdasarkan penjelasan berikut diharapkan penelitian ini dapat bermanfaat bagi semua orang.

1.2 Rumusan Masalah

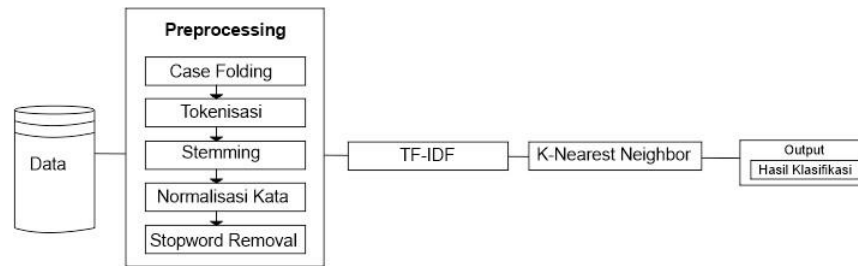
Berdasarkan latar belakang yang telah dipaparkan, maka permasalahan yang dapat diidentifikasi dari penelitian ini adalah bagaimana *K-Nearest Neighbor* dapat mengklasifikasikan komentar di Tripadvisor, dan bagaimana tingkat akurasi klasifikasi menggunakan metode *K-Nearest Neighbor*.

1.3 Tujuan

Tujuan yang ingin dicapai dari penelitian ini adalah untuk membuat sistem yang dapat mengklasifikasi review dengan menggunakan *K-Nearest Neighbor*, dan mengetahui hasil akurasi klasifikasi sentimen *K-Nearest Neighbor*.

2. METODE

Perencanaan alur penelitian merupakan serangkaian tahapan untuk menyelesaikan penelitian ini. Tahapan penelitian terdiri dari enam tahapan seperti yang digambarkan pada Gambar 1. Tahap pertama yaitu pengumpulan data, tahap kedua preprocessing yang terdiri dari case folding, tokenisasi, stemming, normalisasi kata dan stopword, tahap ketiga TF-IDF, tahap keempat *K-Nearest Neighbor* dan tahap terakhir adalah hasil klasifikasi.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Data yang digunakan berupa ulasan hotel POP! Surabaya yang diambil dari *Tripadvisor* dan diolah dengan cara crawling. Untuk contoh hasil yang digunakan ketika mengambil data bisa dilihat pada Gambar 2.

user	judul	ulasan
Akbar	Transit ke surabaya	Pelayanannya baik. Terima kasih mas yus dan mbak dita sudah bantu saat keluar hotel. Kamarnya unik , lumayan bersih dan nyaman saat buat tidur, Yang bik
Yolanda	Hotel yang sangat keren	Pop Gubeng kami kira hotelnya biasa ternyata hotelnya sangat bagus...kebersihan kamar dan area lobi sangatt di jaga serta pelayanan karyawannya sangat
wijaya866	Cozy but lack of internet speed	Overall sudah enak dan nyaman, staf ramah, hanya disayangkan adalah akses koneksi inet yg super lemot aplagi tidak diberi pass wifi, sehingga smua org d
Dian M	Hotelnya bagus	Pop gubeng meskipun harganya murah tapi pelayanan yang diberikan sangat bagus .. mulai dari kamarnya yang bersih serta area lobi nya juga sangat dijaga
Reza R	Hotelnya bersih	Lokasinya sangat startegiss dan terutama hotel kamarnya sangat bersih serta karyawannya sangat ruamahh..Terbaiklha hotel pop gubeng di surabaya..tidal
hendrit p	proses hotel pop gubeng terjamin	hotel pop gubeng lokasi nya sangat strategis eekat dimana2 tapi sayang ppkm ini yang bikin kota susah..kami sekeluarga menginap di pop gubeng dengan l
Puri A	Kamar yang bagus	Overall hotel pop Gubeng sangattt bagus, harga yang ditawarkan juga sesuai dengan fasilitas yang diberikan, kamarnya pop Gubeng sangat unik dan bersih
Dovi Y	Check in check out smooth	Nice hotel di pusat kitaBagus dan nyaman secara keseluruhan n karena karyawan helpful dan hotel nya juga nyaman.. Cuma kehabisan kamar smoking, saya
Ika	Hotel aman di tengah pandemi	Budget hotel yang sangat oke karena fasilitas lengkap dan juga pelayanannya mantap semua service sangat ramah terutama resepsionis, harga sangat oke
Alexander K	Vacation	Kamar nyamanAkses mudah karena berada di pusat kotaPelayanan yang ramahRekomended untuk liburan dan tempat beristirahatHarga yang sangat terjan
anammm	Hotel pop bagus	Hotel pop gubeng pertama kali saya menginap di hotel pop..hotelnya sangat bagus,unik dan harga kamar yang sesuai..Kamar yang disediakan sangat bersih
Aditya A	hotelnya konsisten dengan servic	selama ppkm hotel sangat berdampakk dengan hotel.. sepi tidak seperti biasanya.. tapi hotel pop gubeng meskipun tidak ramai seperti biasanya hotelnya
rahma d	Pop gubeng prokes banget	Masa ppkm ini saya mencoba menginap di pop gubeng..hotel pop dengan harga yang murah tapi tidak murahan...kamarnya pop sangat unik, bersih serta an
Citytraveler	Pelayanan Ramah	Menginap di kamar 1262. Kamar lumayan oke karna ada bangku semi sofa yang biasanya tidak ada di kelas budget. Refill air mineral free, karna tidak disedi
Richard Dante	Lumayan	Hotel cukup bersih, pelayanan cukup baik, kamar juga bagus.Bbrpa masakan saya, untk perlengkapan mkan, tolong disediakan di lantai 1 karna sangat ribet
Farhan	Pelayanan hotel sangat baik	Customer service terbaik dan untuk kamar hotel juga lebih sangat nyaman dan view luar hotel bagus, hotel dekat dengan stasiun berada di pusat kota, ho
Audrey	Bisnis	Kamar oke sesuai pemesanan, kamar bersih, hotel berada di pusat kota mudah kemana manan, fasilitas ada kolam renang gabung dengan hotel harris, peli
Caca sheryly me	Pop hotel	Bersih dan nyaman. Pelayanannya okeViewnya bagus , kamar mandinya nyamanBersih dan nyaman. Pelayanannya sangat memuaskanCocok untuk staycati
Elsa O	Nyaman	Pelayanan nya bagus banget , cepat, karena ada kendala ac tapi resepsionis tanggap untuk mengganti kamar, sangat puas sih harga nya juga murah ya prom
Lentina T	Hotel pop terbaik	Hotel pop selalu menyenangkan dan untuk kamarnya serta area hotel selalu bersih dan yang utama karyawan nya sangat ramah,selalu sapa dan murah sen
Anisa N	kamar oke	pertama kalibmenginap di pop gubeng, kamarnya unik bersih dan sangat menjaga kualitas..tidak hanya kamar tapi karyawannya sangat ramah dan helpfull
yosafat w	Liburan & staycation	Sangat nyaman, pelayanannya sangat memuaskan. Mulai dari front office hingga semuanya memberikan kesan yang baik. Untuk kamarnya nyaman, cukup

Gambar 2. Pengumpulan data

2.2 Preprocessing

Tahap pre-processing atau praproses data adalah proses mempersiapkan data dasar atau inti sebelum melakukan proses yang lain. Pada dasarnya, praproses data dapat dilakukan dengan cara menghilangkan data yang tidak sesuai atau mengubah data menjadi bentuk yang lebih mudah diproses oleh sistem.. Alat bantu untuk menjalankan preprocessing *Case Folding*, *Tokenisasi*, *Stemming*, *Normalisasi Kata* dan *Stopword*.(Mujilahwati, 2016)

2.3 TF-IDF

Term Frequency - Inverse Document Frequency atau TF - IDF adalah suatu metode algoritma yang berguna untuk menghitung bobot setiap kata yang umum digunakan.(Septian et al., n.d.) Metode ini juga terkenal efisien, mudah dan memiliki hasil yang akurat. Metode ini akan menghitung nilai Term Frequency (TF) dan Inverse Document Frequency (IDF) pada setiap token (kata) di setiap dokumen dalam korpus. Secara sederhana, metode TF-IDF digunakan untuk mengetahui berapa sering suatu kata muncul di dalam dokumen.(Andayani & Ryansyah, 2017)

2.4 K-Nearest Neighbor

Algoritma k-Nearest Neighbor adalah suatu algoritma *supervised learning* dimana hasil dari instance yang baru diklasifikasikan berdasarkan mayoritas dari kategori k-tetangga terdekat. Tujuan dari algoritma ini adalah untuk mengklasifikasikan obyek baru berdasarkan atribut dan sample-sample dari training data. Algoritma k-Nearest Neighbor menggunakan

Neighborhood Classification sebagai nilai prediksi dari nilai instance yang baru.(T et al., 2017)

2.5 Hasil dan Kesimpulan

Setelah semua langkah sudah dilakukan dapat ditarik hasil dan kesimpulan yang dapat membantu penelitian selanjutnya untuk mengembangkan penelitian yang sudah di buat.

3. HASIL DAN PEMBAHASAN

Bagian ini membahas hasil dari penelitian dan pada waktu yang sama juga memberikan pembahasan yang komprehensif. Hasil penelitian disajikan menggunakan gambar, grafik, tabel, dan lainnya yang membuat pembaca dapat memahami hasil penelitian dengan mudah. Pembahasan dapat dibuat dengan menggunakan beberapa sub-bab.

3.1 Hasil

Hasil dari penelitian ini adalah sebuah klasifikasi pada ulasan atau komentar di tripadvisor mengenai hotel POP! Surabaya yang menggunakan metode *k-Nearest Neighbor* untuk membantu proses pengolahan data.

a. Pelabelan Data

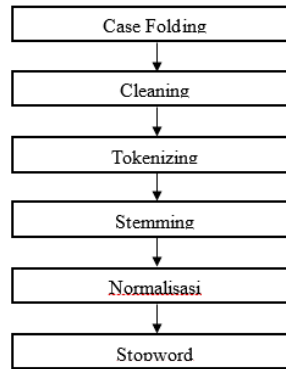
Proses pelabelan data dilakukan Ketika sudah melakukan pengumpulan data melalui crawling, proses pelabelan data ini dilakukan berdasarkan subjektifitas. Berikut adalah hasil dari pelabelan data.

Tabel 1 Hasil pengolahan data

No	Komentar	Label
1	Pelayanannya baik. Terima kasih mas yus dan mbak dita sudah bantu saat keluar hotel. Kamarnya unik, lumayan bersih dan nyaman saat buat tidur, yang bikin keren di lobby ada kedai kopi ada kedai makanan. Sip pokoknya keren	Positif
2	Pop Gubeng kami kira hotelnya biasa ternyata hotelnya sangat bagus...kebersihan kamar dan area lobi sangatt di jaga serta pelayanan karyawannya sangat ramah dan juga dari sisi lokasi sangat strategis...keren lhaa	Positif
3	Overall sudah enak dan nyaman, staf ramah, hanya disayangkan adalah akses koneksi inet yg super lemot apalagi tidak diberi pass wifi, sehingga smua org disekitar hotel bs akses. Need to fix this :) karena beberapa orang pst perlu akses inet yg stabil untuk keperluan kerja.	Negatif
4	Pop gubeng meskipun harganya murah tapi pelayanan yang diberikan sangat bagus... mulai dari kamarnya yang bersih serta area lobinya juga sangat dijaga kebersihan apalagi situasi pndemi ini dituntut untuk jaga kebersihan...pelayanan dari staffnya sangat ramah juga tapi sayang wifi di kamar 1162 kurang kenceng saja tapi semuanya hotelnya sangat bagus	Positif
5	Lokasinya sangat startegiss dan terutama hitel kamarnya sangat bersihh serta karyawannya sangat ruamahh... Terbaiklha hotel pop gubeng di surabaya..tidak kapok ke surabaya pop gubeng lagi..cocok untuk berlibur	Positif

b. Preprocessing

Setelah dilakukannya proses pelabelan maka langka selanjutnya adalah melakukan preprocessing, berikut adalah tahapan dari preprocessing



Gambar 3. Tahapan preprocessing

c. Hasil K-Nearest Neighbor

Dari tahapan preprocessing yang sudah dilalui maka data yang memiliki format .xlsx dimasukkan kedalam system, berikut adalah source code dan hasil uji nya

```

from sklearn.neighbors import KNeighborsClassifier

knn = KNeighborsClassifier(n_neighbors=10).fit(X_train, y_train
)
pred_knn = knn.predict(X_test)

y_pred = knn.predict(X_test)
print(classification_report(y_test, y_pred))
score_knn = accuracy_score(pred_knn, y_test)
print("Accuracy with K-NN: ", score_knn)

```

Tabel 2 Source Code k-NN

Berikut ini adalah hasil uji:

precision	recall	f1-score	support		
	negatif	0.36	0.17	0.23	24
	positif	0.94	0.98	0.96	341
	accuracy			0.93	365
	macro avg	0.65	0.57	0.59	365
	weighted avg	0.91	0.93	0.91	365
Accuracy with K-NN:		0.9260273972602739			

precision	recall	f1-score	support		
	negatif	0.27	0.29	0.28	24
	positif	0.95	0.94	0.95	341
	accuracy			0.90	365
	macro avg	0.61	0.62	0.61	365
	weighted avg	0.91	0.90	0.90	365
Accuracy with K-NN:		0.9013698630136986			

precision	recall	f1-score	support		
negatif		0.50	0.17	0.25	24
positif		0.94	0.99	0.97	341
accuracy				0.93	365
macro avg		0.72	0.58	0.61	365
weighted avg		0.91	0.93	0.92	365
Accuracy with K-NN: 0.9342465753424658					

precision	recall	f1-score	support		
negatif		0.54	0.29	0.38	24
positif		0.95	0.98	0.97	341
accuracy				0.94	365
macro avg		0.75	0.64	0.67	365
weighted avg		0.92	0.94	0.93	365
Accuracy with K-NN: 0.936986301369863					

precision	recall	f1-score	support		
negatif		1.00	0.08	0.15	24
positif		0.94	1.00	0.97	341
accuracy				0.94	365
macro avg		0.97	0.54	0.56	365
weighted avg		0.94	0.94	0.92	365
Accuracy with K-NN: 0.9397260273972603					

precision	recall	f1-score	support		
negatif		0.75	0.12	0.21	24
positif		0.94	1.00	0.97	341
accuracy				0.94	365
macro avg		0.85	0.56	0.59	365
weighted avg		0.93	0.94	0.92	365
Accuracy with K-NN: 0.9397260273972603					

precision	recall	f1-score	support		
negatif		1.00	0.08	0.15	24
positif		0.94	1.00	0.97	341
accuracy				0.94	365
macro avg		0.97	0.54	0.56	365
weighted avg		0.94	0.94	0.92	365
Accuracy with K-NN: 0.9397260273972603					

precision	recall	f1-score	support		
negatif		1.00	0.08	0.15	24

positif	0.94	1.00	0.97	341
accuracy			0.94	365
macro avg	0.97	0.54	0.56	365
weighted avg	0.94	0.94	0.92	365
Accuracy with K-NN: 0.9397260273972603				

precision	recall	f1-score	support	
negatif	0.00	0.00	0.00	24
positif	0.93	1.00	0.97	341
accuracy			0.93	365
macro avg	0.47	0.50	0.48	365
weighted avg	0.87	0.93	0.90	365
Accuracy with K-NN: 0.9342465753424658				

precision	recall	f1-score	support	
negatif	0.00	0.00	0.00	24
positif	0.93	1.00	0.97	341
accuracy			0.93	365
macro avg	0.47	0.50	0.48	365
weighted avg	0.87	0.93	0.90	365
Accuracy with K-NN: 0.9342465753424658				

Ketetangaan	Akurasi
k=1	0.9260273972602739
k=2	0.9013698630136986
k=3	0.9342465753424658
k=4	0.936986301369863
k=5	0.9397260273972603
k=6	0.9397260273972603
k=7	0.9397260273972603
k=8	0.9397260273972603
k=9	0.9397260273972603
k=10	0.9342465753424658

Tabel 2 Data Ketetangaan

4. KESIMPULAN

Pengujian klasifikasi sentiment analisa terhadap ulasan hotel POP! Surabaya menggunakan algoritma K-Nearest Neighbor. Berdasarkan hasil percobaan, didapatkan akurasi tertinggi pada dataset ketetangaan 5 (k=5) sampai dengan ketetangaan 9 (k=9) dengan pembagian dataset antara jumlah label positif dan label negatif, dan masing masing menghasikan akurasi terbaik sebesar 93%.

5. DAFTAR PUSTAKA

- Andayani, S., & Ryansyah, A. (2017). Implementasi Algoritma TF-IDF Pada Pengukuran Kesamaan Dokumen. *JuSiTik : Jurnal Sistem Dan Teknologi Informasi Komunikasi*, 1(1), 53. <https://doi.org/10.32524/jusitik.v1i1.218>
- Mujilahwati, S. (2016). Pre-Processing Text Mining Pada Data Twitter. *Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016*(Sentika), 2089–9815.
- Septian, J. A., Fahrudin, T. M., & Nugroho, A. (n.d.). *Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor*. 43–49.
- T, A. J., Yanosma, D., & Anggriani, K. (2017). Implementasi Metode K-Nearest Neighbor (Knn) Dan Simple Additive Weighting (Saw) Dalam Pengambilan Keputusan Seleksi Penerimaan Anggota Paskibraka. *Pseudocode*, 3(2), 98–112. <https://doi.org/10.33369/pseudocode.3.2.98-112>