

Analisis Sentimen Pengguna Google Terhadap Destinasi Wisata Di Kota Semarang Menggunakan Metode *K-Nearest Neighbor*

Slamet Widodo^{1*}, Budi Hartono²

Program Studi Teknik Informatika, Universitas Stikubank, Semarang, Indonesia

*e-mail *Corresponding Author*: dodobyst@gmail.com

Abstract

Sentiment is a form of application of text mining which aims to obtain opinions or user sentiments about the products and services available. Google is a search engine that is widely used to surf the internet, Google creates many products, one example is Google Maps. Google maps provides a review feature for places that have been visited in various places. Various reviews of places or tours can be written on google maps reviews, one of which is tourism in the city of Semarang. Several travelers have provided positive, negative and neutral reviews written on Google Reviews. To determine the grouping of tourists' reviews, it is necessary to apply sentiment analysis. This study uses the K-Nearest Neighbor algorithm to classify sentiment whether it is negative, positive, or neutral. The results showed that the K-Nearest Neighbor method produced the highest accuracy of 78% at K = 36 and a ratio of 90:10.

Keywords: *Sentiment analysis; Google maps review; Tourist attraction; K-Nearest Neighbor*

Abstrak

Sentiment merupakan salah satu bentuk aplikasi dari text mining yang bertujuan untuk memperoleh pendapat atau sentimen pengguna tentang produk dan layanan yang tersedia. Google adalah salah satu search engine yang banyak digunakan untuk berselancar di internet, google banyak menciptakan produk produk salah satu contohnya yaitu google maps. Pada google maps disediakan fitur review tempat yang telah dikunjungi di berbagai tempat. Berbagai review tempat atau wisata dapat di tulis pada google maps review, salah satunya yaitu wisata di kota Semarang. Beberapa wisatawan telah memberikan review positif, negatif dan netral yang ditulis di google review. Untuk menentukan pengelompokkan review para wisatawan, perlu dilakukan penerapan analisis sentiment. Penelitian ini menggunakan algoritme K-Nearest Neighbor untuk mengklasifikasikan sentiment apakah termasuk negatif, positif, atau netral. Hasil penelitian menunjukkan metode K-Nearest Neighbor menghasilkan akurasi tertinggi 78% pada nilai K=36 dan rasio 90:10.

Kata Kunci: *Analisis sentiment; Google maps review; Objek wisata; K-Nearest Neighbor*

1. Pendahuluan

Google adalah salah satu search engine yang saat ini banyak orang gunakan untuk berselancar di internet. Google banyak menciptakan produk produk yang sangat canggih salah satu contohnya yaitu google maps. Aplikasi pemetaan untuk wilayah diseluruh dunia ini tidak hanya menyediakan fitur maps tetapi juga memetakan beberapa tempat tak terkecuali tempat wisata di seluruh dunia, dimana di fitur google maps juga menyediakan fitur ulasan untuk para pengunjung yang telah mengunjungi tempat tersebut menyampaikan ulasan atau pendapatnya.

Kota Semarang secara geografis dan sosiologis memiliki daya tarik pariwisata dengan karakter dan keunikan tersendiri dibandingkan dengan kota kota lainnya di Indonesia. Kota Semarang memiliki kondisi geografis mulai dari daerah perbukitan sampai dengan daerah pantai sehingga memiliki potensi alam sebagai destinasi pariwisata bila dikelola dan dikembangkan dengan baik. Wisata Kota semarang terdiri dari beberapa objek diantaranya adalah Lawang Sewu, Saloka Theme Park, Klenteng Sampokong, Dusun Semilir, Masjid Agung Jawa Tengah, Candi Gedong Songo, Cimory on The Valley, Taman Bunga Celosia, Pagoda Avalokitesvara, Kota Lama Semarang dan beberapa lainnya. Banyaknya review di google maps mengenai

tempat wisata di Semarang 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. *Sentiment* merupakan salah satu bentuk aplikasi dari *text mining* yang bertujuan untuk memperoleh pendapat atau sentimen pengguna tentang produk dan layanan yang tersedia. Saat ini telah banyak orang yang memanfaatkan internet sebagai media penyampaian ulasan atau pendapatnya.

Metode *K-Nearest Neighbor* merupakan sebuah metode *machine learning* yang dapat melakukan klasifikasi terhadap objek berdasarkan jarak yang paling dekat dengan objek tersebut. Metode K-NN banyak digunakan dalam klasifikasi objek, seperti dalam [1] yang mengklasifikasi review para wisatawan yang berupa *text* menjadi suatu informasi.

Penelitian ini melakukan analisis sentimen dengan Bahasa Indonesia review pengguna *google maps* terhadap destinasi wisata di kota Semarang. Penelitian ini bertujuan untuk mendapatkan informasi apakah masyarakat cenderung bersifat positif, negatif, atau netral dalam mereview destinasi di kota Semarang, guna membantu para wisatawan yang akan berkunjung di kota Semarang dan juga untuk pengelola wisata di Semarang bisa memperoleh informasi tambahan apa saja yang perlu diperhatikan dan dipertahankan.

2. Tinjauan Pustaka

Pada penelitian yang dilakukan oleh Retno Sari pada tahun 2020 di Jurnal Sains dan Manajemen menjelaskan bahwa analisis sentimen diperlukan untuk mengambil keputusan yang terbaik. Penelitian dengan judul "Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi Menggunakan Algoritme *K-Nearest Neighbor* (K-NN)" ini bertujuan untuk mengetahui akurasi analisis sentimen menggunakan metode *K-Nearest Neighbor*. Dijelaskan juga dalam penelitian tersebut kelebihan metode KNN yaitu memiliki prinsip yang sederhana dan bekerja berdasarkan jarak terpendek dari data uji dan data latih. Dari 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. [1]

Pada penelitian yang dilakukan oleh Hidayatus Sibyan dan Nur Hasanah tahun 2021 di Jurnal Penelitian dan Pengabdian Kepada Masyarakat UNSIQ dengan judul "Analisis Sentimen Pada Wisata Dieng Dengan Algoritme *K-Nearest Neighbor* (K-NN)" menjelaskan tujuan dari penelitian tersebut yaitu untuk melakukan identifikasi dan klasifikasi ulasan pengunjung wisata sebagai bahan pertimbangan dalam melakukan perbaikan atas kekurangan pada tempat wisata yang dikelola, serta bisa bermanfaat bagi para wisatawan yang akan berkunjung sehingga para wisatawan tidak perlu untuk melihat review satu per satu. Pada penelitian tersebut menggunakan dataset positif dan negatif selanjutnya dilakukan preprocessing data dan menggunakan metode KNN untuk menguji tingkat akurasi nya. Hasil penelitian didapatkan akurasi sebesar 86% [2].

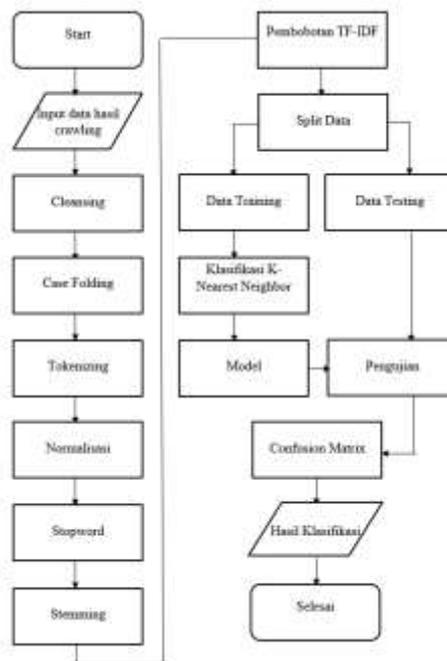
Pada penelitian yang dilakukan oleh Gergorius Kopong Pati dan Elfira Umar pada tahun 2022 di jurnal Media Informatika Budidarma dengan judul "Analisis Sentimen Komentar Pengunjung Terhadap Tempat Wisata Danau Weekuri Menggunakan Metode *Naive Bayes Classifier* dan *K-Nearest Neighbor*" menjelaskan bahwa kebiasaan para wisatawan yang setelah berkunjung yaitu berkomentar atau me-review pada salah satu aplikasi di situs *Trip Advisor*. Oleh karena itu penelitian tersebut bertujuan untuk menganalisis sentimen terhadap salah satu di kota Kupang yaitu Gua Kristal. Pada Penelitian ini menggunakan 2 metode yaitu *Naive Bayes* dan *K-Nearest Neighbor* untuk perbandingan terhadap akurasinya. Klasifikasi sentimen terdiri dari positif dan negatif. Hasil pengujian akurasi menggunakan metode *K-Nearest Neighbor* dimana diperoleh tingkat akurasinya 76.53% sedangkan tingkat akurasi dengan metode *Naive Bayes Classifier* sebesar 73.47%. oleh karena itu tingkat akurasi yang diperoleh dengan menggunakan metode K-NN lebih baik [3].

Pada penelitian yang dilakukan oleh Inosensius Karelo Hesay, Indriati, dan Sigit Adinugroho pada tahun 2021 di jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer dengan judul "Analisis Sentimen Ulasan Pengunjung Simpang Lima Gumul Kediri Menggunakan Metode Bm25 Dan *Neighbor-Weighted K-Nearest Neighbor*" menjelaskan pendatang yang mengunjungi Simpang Lima Gumul Kabupaten Kediri sering kali memberikan ulasan di *google review*. Adanya ulasan tersebut sangat membantu pengelola untuk mengetahui kekurangan dan kelebihan dari sarana dan prasarana yang ada. Namun pihak pengelola belum mempunyai cara atau sistem untuk mengelompokkan ulasan positif dan negatif secara otomatis [4].

Dari beberapa sumber referensi yang dirujuk diatas dapat disimpulkan bahwa penggunaan algoritme K-NN dapat diterapkan untuk klasifikasi sentiment dengan nilai akurasi yang cukup akurat. Dari beberapa referensi yang dirujuk menunjukkan bahwa jumlah data, proses teks preprocessing dataset dan parameter klasifikasi sangat mempengaruhi dalam hasil akurasi. Pada penelitian ini perbedaan dengan penelitian sebelumnya yaitu menambah jumlah data serta mengubah obyek data, menambah proses dalam teks preprocessing dan mengubah parameter klasifikasi yang digunakan dengan tujuan mengetahui model terbaik dalam penerapannya.

3. Metodologi

Penelitian dimulai dengan pengumpulan dataset, *preprocessing* data, pembobotan tf-idf, klasifikasi dengan metode K-NN, uji validasi dan hasil klasifikasi. Proses pada penelitian ini bersifat ekperimental untuk menguji dan mengevaluasi algoritme K-NN dalam menentukan klasifikasi sentiment. Diagram alur perancangan sistem dapat dilihat dalam gambar berikut:



Gambar 1. Diagram Alur Proses Klasifikasi menggunakan Metode K-NN

3.1 Pengumpulan Dataset

Penelitian ini menggunakan data berupa data teks hasil dari ulasan masyarakat pada google maps. Ulasan yang digunakan sebagai dataset adalah ulasan yang ditujukan untuk destinasi wisata kota Semarang oleh beberapa wisatawan. Pengambilan dataset di *google* ini dikenal dengan istilah *scraper* atau *crawling*. *Crawling* data bertujuan untuk mengumpulkan data berupa teks dari ulasan yang di tulis oleh beberapa wisatawan di *google maps*. Pengumpulan data dengan mengambil kategori ulasan paling relevan, terbaik, dan terburuk.

Pada penelitian ini untuk mengumpulkan data menggunakan *web crawler* yaitu *apify*. *Web crawler* adalah sebuah aplikasi yang secara otomatis melintasi web dengan mengunduh dokumen dan mengikuti link dari halaman ke halaman, sehingga *web crawler* dapat dijadikan sebuah alat untuk mengambil konten-konten yang ditampilkan oleh halaman website, kemudian konten-konten tersebut dikelompokkan dalam satu atau lebih atribut

Proses pengambilan data di *google maps* melalui *web crawler apify* dilakukan sebanyak 30 kali dengan mengambil 3 kategori yaitu paling relevan, terbaik, dan terburuk untuk 10 tempat wisata di kota Semarang. Penulis mengambil ulasan dari tempat wisata di kota Semarang yaitu Lawang Sewu, Saloka Theme Park, Klenteng Sampokong, Dusun Semilir, Masjid Agung Jawa Tengah, Candi Gedong Songo, *Cimory on The Valley*, Taman Bunga Celosia, Pagoda Avalokitesvara, Kota Lama Semarang. Data tersebut kemudian digabungkan menjadi satu dalam format .xlsx. Data hasil proses pengambilan data didapatkan dataset sejumlah 1500 data.

3.2 Preprocessing teks

Preprocessing merupakan tahapan awal dari text mining yang mana proses yang penting dalam analisis sentiment karena pada tahapan ini sangat menentukan hasil analisis yang akan dihasilkan. *Preprocessing* merupakan proses membersihkan, menyaring, dan mengolah data dari noise ataupun merubah format data agar data yang dihasilkan bisa digunakan pada tahap selanjutnya dengan maksimal. Adapun tahapan *preprocessing* dataset disajikan pada Tabel 1.

Tabel 1. Penjelasan Proses *Preprocessing*

Proses	Penjelasan
Cleansing	Proses dimana karakter dan tanda baca yang tidak diperlukan dihilangkan dari teks. Contoh URL, tag (#), karakter atau symbol seperti koma (,), titik (.), emoticon dan angka dan tanda baca lainnya.
Case Folding	Proses penyamaan case dalam sebuah dokumen. Contoh perbedaan besar dan kecil huruf alphabet.
Tokenizing	Proses memecah string input berdasarkan tiap kata yang menyusunnya pada suatu kalimat.
Normalisasi	Proses mengubah kata yang tidak baku menjadi kata yang baku.
Stopword	Proses menghilangkan kata-kata yang sering muncul dan yang tidak memiliki makna serta tidak memiliki pengaruh apaun dalam analisis sentiment.
Stemming	Proses pengembalian bentuk kata yang mendapatkan imbuhan menjadi bentuk kata dasar.

3.3 Pembobotan TF-IDF

Pembobotan kata dilakukan dengan mengubah kata menjadi angka dengan proses pembobotan kata yang bertujuan untuk pemberian nilai pada jumlah kemunculan suatu kata dalam sebuah dokumen. Metode pembobotan kata yang digunakan dalam penelitian ini adalah TF-IDF. Pada proses ini ada 3 tahapan yaitu menghitung nilai TF (*Term Frequency*), menghitung nilai IDF (*Invers Document Frequency*), dan menghitung nilai TF-IDF (*Term Frequency-Invers Document Frequency*). Frekuensi kemunculan kata (term) dalam dokumen digunakan untuk menentukan pembobotan nilai kata(term) pada dokumen. Bobot nilai suatu kata(term) semakin besar sesuai dengan banyak frekuensi kemunculannya dalam suatu dokumen. *Invers Document Frequency* merupakan kebalikan dari *term frequency* metode ini untuk menghitung penyebaran term pada sebuah dokumen. Untuk rumus menghitung IDF sebagai berikut:

$$\log \frac{n}{df_t} \dots\dots\dots (1)$$

Pada penelitian ini penulis menggunakan library scikit learn untuk mengimplementasikan IDF pada program *python*, sehingga rumus IDF pada scikit learn sedikit berbeda, untuk detail rumus dapat dilihat sebagai berikut:

$$\ln \frac{n+1}{df_t+1} + 1 \dots\dots\dots (2)$$

Pada library scikit learn juga menerapkan normalisasi L2 untuk IDF, untuk rumus normalisasi dijelaskan sebagai berikut:

$$v_{norm} = \frac{v}{\|v\|_2} = \frac{v}{\sqrt{v_1^2+v_2^2+\dots+v_n^2}} \dots\dots\dots (3)$$

Dimana v adalah vektor term pada suatu dokumen. Dan tahapan terakhir pada pembobotan kata adalah TF-IDF. TF-IDF merupakan perkalian nilai TF_t dengan nilai IDF_t yang sudah dihitung pada proses TF dan IDF.

3.4 Klasifikasi *K-Nearest Neighbor*

Pada penelitian ini menggunakan algoritme *k-nearest neighbor* untuk menentukan klasifikasi pada kelas label negatif, netral, dan positif. Algoritme *K-NN* memiliki prinsip sederhana yaitu bekerja berdasarkan jarak terpendek dari sampel uji ke sampel latih. Untuk *metric* yang

digunakan pada penelitian ini yaitu *metric minkowski* yang mana *metric* tersebut secara default adalah generalisasi dari *Euclidean distance* dan *Manhattan distance*. Rumus *metric minkowski* adalah sebagai berikut:

$$d(x,y) = (\sum_{i=1}^n |x_i - y_i|^m)^{\frac{1}{m}} \dots\dots(4)$$

Dimana x dan y masing masing adalah dua titik yang akan dicari jaraknya. Pada rumus tersebut terdapat parameter m dimana m=1 ruang jarak *minkowski* sama dengan *Manhattan distance*, ketika m=2 menjadi *Euclidean distance* dan untuk m=∞ (tak terhingga) dapat dikatakan sebagai *Chebyshev distance*. Pada penelitian ini akan menggunakan m=2 sebagai parameternya. Sehingga rumus untuk menghitung nya menjadi:

$$d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots\dots\dots (5)$$

3.5 Uji Validasi

Pengujian algoritme *K-Nearest Neighbor* ini diperlukan untuk mengetahui akurasi dari model yang dihasilkan pada proses klasifikasi. Pengujian menggunakan *confusion matrix* dengan nilai k tetangga 1 sampai dengan 40 dengan rasio 90:10, 80:20, 70:30, 60:40, 50:50. *Confusion matrix* berbentuk tabel matriks yang menghasilkan output berupa dua kelas atau lebih. Dalam penelitian ini *confusion matrix* menghasilkan 9 *output* yang dapat dilihat di tabel berikut:

Tabel 2. Output Tabel Confusion Matrix

Realita	Prediksi		
	Negatif	Netral	Positif
Negatif	Tneg	Fnet	Fpos
Netral	Fneg	Tnet	Fpos
Positif	Fneg	Fnet	Tpos

Dijelaskan Tneg adalah singkatan dari true negatif, data sebenarnya bernilai negatif dan diprediksi negatif. Tnet adalah singkatan dari true netral, data sebenarnya bernilai netral dan diprediksi netral. Tpos adalah singkatan dari true positif, data sebenarnya bernilai positif dan diprediksi positif. Fneg adalah singkatan dari false negatif, data sebenarnya bernilai netral atau positif dan diprediksi negatif. Fnet adalah singkatan dari false netral, data sebenarnya bernilai negatif atau positif dan diprediksi netral. Fpos adalah singkatan dari false positif, data sebenarnya bernilai negatif atau netral dan diprediksi positif. Dari tabel *confusion matrix*, rumus menghitung nilai akurasi, *precision*, *recall* dan *f-measure* adalah sebagai berikut:

Precision Negatif = $\frac{Tneg}{Tneg+Fneg} \times 100\%$

Recall Negatif = $\frac{Tneg}{Tneg+Fnet+Fpos} \times 100\%$

F1-Score Negatif = $\frac{(2 \times Precision\ Negatif \times Recall\ Negatif)}{Precision\ Negatif + Recall\ Negatif} \times 100\%$

Precision Netral = $\frac{Tnet}{Tnet+Fnet} \times 100\%$

Recall Netral = $\frac{Tnet}{Tnet+Fneg+Fpos} \times 100\%$

F1-Score Netral = $\frac{(2 \times Precision\ Netral \times Recall\ Netral)}{Precision\ Netral + Recall\ Netral} \times 100\%$

Precision Positif = $\frac{Tpos}{Tpos+Fpos} \times 100\%$

Recall Positif = $\frac{Tpos}{Tpos+Fneg+Fnet} \times 100\%$

$$F1\text{-Score Positif} = \frac{(2 \times \text{Precision Positif} \times \text{Recall Positif})}{\text{Precision Positif} + \text{Recall Positif}} \times 100\%$$

$$\text{Akurasi} = \frac{(\text{Tneg} + \text{Tnet} + \text{Tpos})}{\text{Tneg} + \text{Tnet} + \text{Tpos} + \text{Fneg} + \text{Fnet} + \text{Fpos}} \times 100\%$$

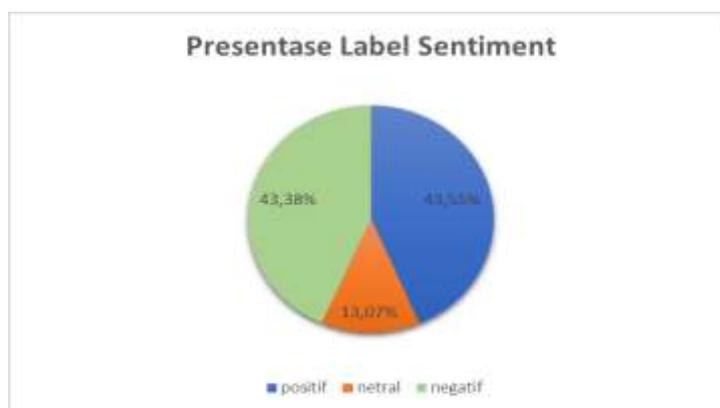
3.6 Hasil Klasifikasi

Setelah melakukan proses *preprocessing*, pembobotan kata dengan TF-IDF, klasifikasi menggunakan algoritme K-NN dan pengujian menggunakan *confusion matrix* sehingga didapatkan akurasi terbaik dalam model klasifikasi yang akan di bahas pada bab selanjutnya.

4. Hasil dan Pembahasan

Pada penelitian ini untuk implementasi nya menggunakan *python* dan *tools* yang digunakan adalah *google colab*. Untuk memulai program *python* pada *google colab* diperlukan library. Library merupakan kumpulan gabungan dari *package* dan *module* berupa kode yang dapat digunakan berulang kali dalam program yang berbeda. Saat melakukan analisis sentimen dengan *Python*, beberapa library berikut harus diinstal yaitu *csv*, *pandas*, *numpy*, *sastrawi*, *scikit learn* atau *sklearn*. Untuk install library dapat menggunakan perintah “pip” contoh “pip install <nama paket>”. Setelah terinstall lalu import library dengan perintah “import” contoh “import <nama library>”.

Setelah *library* terinstal pada *google colab* tahapan selanjutnya yaitu membaca data yang akan dilakukan *preprocessing* pada tahap selanjutnya. Untuk membaca data library yang digunakan adalah *pandas*. Sebelum membaca data dalam program *python*, terlebih dahulu membaca atau mengolah hasil *crawling* di dalam komputer yaitu menggunakan *microsoft excel*. Proses pelabelan dibagi menjadi tiga kategori sentiment yaitu negatif, netral dan positif. Pada tahap ini menggunakan metode manual untuk pelabelan sentiment dengan didasarkan pada rating yang diberikan oleh para wisatawan itu sendiri yaitu, kategori positif dengan rentang rating 5 dan 4, netral dengan rating 3 dan negative dengan rating 2 dan 1 sehingga didapat 500 review positif, 500 review negatif dan 500 review netral. Kemudian di sortir kembali menjadi 500 review positif, 498 review negatif dan 150 review netral. Sehingga presentase pembagian data dapat dilihat pada grafik Gambar 2.



Gambar 2. Grafik Presentase Label Sentiment

Ketika dataset sudah diberi label tahapan selanjutnya adalah *preprocessing*. Pada tahap ini akan dilakukan 6 tahapan yaitu *cleansing*, *casefolding*, *tokenizing*, normalisasi, *stopword* dan *stemming*. Hasil dari masing masing tahapan *preprocessing* dapat dilihat pada Tabel 3.

Tabel 3. Hasil *Preprocessing Text*

Proses	Dataset Asli	Dataset Hasil Preprocessing
Cleansing	Salah satu tmpt wisata akhir pekan yg menarik, banyak jajanan nya, spot foto, dll	Salah satu tmpt wisata akhir pekan yg menarik banyak jajanan nya spot foto dll

Proses	Dataset Asli	Dataset Hasil Preprocessing
Case Folding	Salah satu tmpt wisata akhir pekan yg menarik banyak jajanan nya spot foto dll	salah satu tmpt wisata akhir pekan yg menarik banyak jajanan nya spot foto dll
Tokenizing	salah satu tmpt wisata akhir pekan yg menarik banyak jajanan nya spot foto dll	'salah', 'satu', 'tmpt', 'wisata', 'akhir', 'pekan', 'yg', 'menarik', 'banyak', 'jajanan', 'nya', 'spot', 'foto', 'dll'
Normalisasi	'salah', 'satu', 'tmpt', 'wisata', 'akhir', 'pekan', 'yg', 'menarik', 'banyak', 'jajanan', 'nya', 'spot', 'foto', 'dll'	'salah', 'satu', 'tempat', 'wisata', 'akhir', 'pekan', 'yang', 'menarik', 'banyak', 'jajanan', 'nya', 'spot', 'foto', 'dll'
Stopword	'salah', 'satu', 'tempat', 'wisata', 'akhir', 'pekan', 'yang', 'menarik', 'banyak', 'jajanan', 'nya', 'spot', 'foto', 'dll'	'salah', 'tempat', 'wisata', 'pekan', 'menarik', 'jajanan', 'nya', 'spot', 'foto', 'dll'
Stemming	'salah', 'tempat', 'wisata', 'pekan', 'menarik', 'jajanan', 'nya', 'spot', 'foto', 'dll'	salah tempat wisata pekan tarik jajan nya spot foto dll

Data yang sudah melewati tahapan preprocessing akan dilanjutkan dengan tahapan pembobotan kata dengan TF-IDF. Disiapkan data training dan data uji pada Tabel 4 dan Tabel 5.

Tabel 4. Contoh Data Training

No Dokumen	Ulasan	Label
D1	Wahana banyak bagus	positif
D2	Tempatnya wisata nyaman bersih keren	positif
D3	Wisata indah tapi kotor	netral
D4	wisata buruk	negatif

Tabel 5. Contoh Data Uji

No Dokumen	Ulasan	Label
D5	wisata semarang terburuk kecewa	?

Data training dan data uji akan dilakukan pembobotan kata TF-IDF dan menerapkan rumus normalisasi L2 sehingga hasil selengkapnya dijelaskan pada Tabel 6.

Tabel 6. Hasil TF-IDF

ID Dokumen	D1	D2	D3	D4	D5
bagus	0,57735	0	0	0	0
banyak	0,57735	0	0	0	0
bersih	0	0,48127	0	0	0
buruk	0	0	0	0,87125	0
indah	0	0	0,54904	0	0
kecewa	0	0	0	0	0,54904
keren	0	0,48127	0	0	0
kotor	0	0	0,54904	0	0
nyaman	0	0,48127	0	0	0
semarang	0	0	0	0	0,54904
tapi	0	0	0,54904	0	0

ID Dokumen	D1	D2	D3	D4	D5
tempatnya	0	0,48127	0	0	0
terburuk	0	0	0	0	0,54904
wahana	0,57735	0	0	0	0
wisata	0	0,27114	0,30932	0,49085	0,30932

Setelah setiap term diberi nilai pembobotan tahapan selanjutnya yaitu proses klasifikasi menggunakan algoritme K-NN. Pada penelitian ini menggunakan jarak *euclidean* sebagai parameter sehingga term data uji akan dicari nilai terdekat dengan data training kemudian dicari nilai similaritas antara data uji dengan data training dan diurutkan dari nilai yang paling besar ke kecil. Untuk hasil selengkapnya disajikan pada Tabel 7.

Tabel 7. Hasil Nilai Similaritas

No Dokumen	Ulasan	Label	Nilai Similarity
D4	wisata buruk	negatif	0,43432214
D3	Wisata indah tapi kotor	netral	0,4264649
D2	Tempatnya wisata nyaman bersih keren	positif	0,42487898
D1	Wahana banyak bagus	positif	0,41421356

Maka dapat disimpulkan bahwa data uji "wisata semarang terburuk kecewa" dapat diklasifikasikan sebagai nilai sentiment negatif

Untuk menguji performa algoritme *K-Nearest Neighbor* ini diperlukan pengujian, untuk mengetahui akurasi dari model yang dihasilkan pada proses klasifikasi. Sebelum pengujian, dataset sebanyak 1148 akan di split terlebih dahulu menjadi dua bagian yaitu data training untuk membuat model dan data uji untuk menguji model yang telah dibuat. Dataset pada hasil *crawling* didapatkan sebanyak 1500 dataset akan dikurangi pada sentiment netral dikarenakan nilai akurasi yang dihasilkan sangat buruk sehingga tidak bisa dijadikan dalam sebuah model untuk klasifikasi sentimen. Disini peneliti menggunakan 5 rasio split antara data training dan data uji yaitu 90:10, 80:20, 70:30, 60:40, dan 50:50. Pengujian akan dilakukan menggunakan nilai K = 1 sampai 40 untuk masing masing rasio split. Pada pengujian ini menggunakan report akurasi *confusion matrix*, data seluruh akurasi pada pengujian ini dapat dilihat pada Tabel 8.

Tabel 8. Hasil Pengujian

Data Training/Testing	Rasio				
K	0,1	0,2	0,3	0,4	0,5
1	60%	57%	56%	54%	55%
2	54%	52%	51%	52%	51%
3	65%	58%	56%	60%	57%
4	64%	61%	58%	62%	60%
5	67%	64%	63%	63%	60%
6	64%	70%	65%	67%	64%
7	71%	71%	64%	67%	64%
8	69%	70%	68%	69%	66%
9	71%	71%	66%	70%	66%
10	69%	72%	68%	69%	67%
11	71%	72%	70%	68%	65%
12	75%	73%	70%	69%	67%
13	76%	72%	70%	69%	66%
14	76%	71%	68%	68%	67%
15	75%	71%	67%	69%	66%
16	71%	72%	69%	70%	68%
17	73%	72%	69%	71%	67%

	Data		Rasio		
	Training	Testing			
18	73%	73%	70%	71%	68%
19	72%	72%	70%	71%	68%
20	72%	73%	71%	71%	69%
21	74%	73%	71%	69%	68%
22	74%	75%	70%	70%	67%
23	73%	75%	70%	69%	68%
24	77%	74%	70%	68%	68%
25	76%	74%	71%	69%	68%
26	76%	74%	71%	69%	68%
27	76%	74%	70%	70%	69%
28	75%	74%	69%	69%	69%
29	74%	75%	69%	70%	69%
30	76%	74%	70%	70%	69%
31	77%	75%	70%	71%	69%
32	77%	75%	70%	70%	70%
33	77%	75%	70%	70%	70%
34	77%	76%	70%	70%	69%
35	76%	74%	70%	69%	70%
36	78%	74%	69%	70%	70%
37	77%	73%	70%	70%	70%
38	77%	73%	69%	69%	70%
39	77%	73%	68%	70%	70%
40	77%	73%	69%	70%	69%

Hasil pengujian penelitian pada Tabel 8 di atas menjelaskan algoritme *K-Nearest Neighbor* dalam pengklasifikasian teks dengan kelas sentiment Negatif, Netral, dan Positif. Dari Tabel 8 tersebut, nilai akurasi tertinggi yang didapatkan dalam masing masing rasio dan nilai K = 1 sampai dengan 40 adalah rasio 90:10 dengan nilai K=36 yaitu sebesar 78% sedangkan akurasi terendah ada pada rasio 50:50 dengan nilai K=2. Dari tabel 8 dapat disimpulkan bahwa data training yang semakin besar dan data uji semakin sedikit akan membuat model dengan tingkat akurasi yang tinggi sedangkan penggunaan nilai K yang semakin tinggi akan membuat akurasi yang tinggi, hal ini dikarenakan ketika nilai K yang tinggi, data yang masuk pada k tetangga terdekat semakin banyak dan bisa mempresentasikan kelas pada data uji dengan baik. Sehingga penelitian analisis sentiment pengguna *google* terhadap destinasi wisata di kota semarang menggunakan metode K-NN menghasilkan akurasi tertinggi 78% pada nilai K=36 dan rasio 90:10

5. Kesimpulan

Hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa platform *google* merupakan salah satu media yang dapat digunakan untuk memperoleh data review suatu tempat di lokasi yang berbeda dalam bentuk teks. Sentiment yang didapatkan dalam proses *crawling dataset* cukup beragam sehingga memberi label dalam 3 kelas yaitu positif, negatif dan netral. Pengklasifikasian sentiment menggunakan algoritme K-NN dengan kelas label yaitu positif, netral, dan negatif memiliki performa sistem klasifikasi yang cukup baik. Hasil penelitian menunjukkan metode K-NN menghasilkan akurasi tertinggi 78% pada nilai K=36 dengan rasio 90:10.

Daftar Referensi

- [1] R. Sari, "Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi menggunakan Algoritme K-Nearest Neighbor (K-NN)," *Jurnal Sains dan Manajemen*, vol. 8, no. 1, pp. 10–17, 2020, Accessed: Jul. 01, 2023. [Online]. Available: ejournal.bsi.ac.id/ejurnal/index.php/khatulistiwa/issue/archive/index.php/evolusi/article/view/7371
- [2] H. Sibyan and N. Hasanah, "Analisis Sentimen Pada Wisata Dieng Dengan Algoritme K-Nearest Neighbor (K-NN)," *Jurnal Penelitian dan Pengabdian Kepada Masyarakat UNSIQ*, vol. 9, no. 1, pp. 38–47, 2021, Accessed: Jul. 01, 2023. [Online]. Available: <https://ojs.unsiq.ac.id/index.php/ppkm/article/view/2218>
- [3] P. Gergorius Kopong and E. Umar, "Analisis Sentimen Komentar Pengunjung Terhadap Tempat Wisata Danau Weekuri Menggunakan Metode Naive Bayes Classifier Dan K-

- Nearest Neighbor,” *Jurnal Media Informatika Budidarma*, vol. 6, no. 4, pp. 2309–2315, Oct. 2022, doi: 10.30865/mib.v6i4.4635.
- [4] I. Karelo Hesay and S. Adinugroho, “Analisis Sentimen Ulasan Pengunjung Simpang Lima Gumul Kediri menggunakan Metode BM25 dan Neighbor-Weighted K-Nearest Neighbor,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 7, pp. 3160–3169, 2021, Accessed: Jul. 01, 2023. [Online]. Available: jptiik.multi.web.id/index.php/jptiik/article/view/9506
- [5] A. Mbay Ndapamuri, D. Manongga, and A. Iriani, “Analisis Sentimen Ulasan Aplikasi Tripadvisor Dengan Metode Support Vector Machine, K-Nearest Neighbor, Dan Naive Bayes,” *Jurnal Inovtek Polbeng - Seri Informatika*, vol. 8, no. 1, pp. 127–140, 2023, Accessed: Jul. 01, 2023. [Online]. Available: <http://ejournal.polbeng.ac.id/index.php/ISI/article/view/3260>
- [6] D. Era, S. Andryana, and A. Rubhasy, “Perbandingan Algoritme Naïve Bayes Dan K-Nearest Neighbor pada Analisis Sentimen Pembukaan Pariwisata Di Masa Pandemi Covid 19,” 2023. Accessed: Jul. 01, 2023. [Online]. Available: <http://tunasbangsa.ac.id/ejurnal/index.php/jsakti/article/view/590>
- [7] M. Furqan and S. Mayang Sari, “Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia,” *Techno.COM*, vol. 21, no. 1, pp. 52–61, 2022, [Online]. Available: <https://doi.org/10.33633/tc.v21i1.5446>
- [8] R. Prihatini Saputri, W. Setya Winahju, and Fithriasari, “Klasifikasi Sentimen Wisatawan Candi Borobudur pada Situs tripAdvisor Menggunakan Support Vector Machine dan KNN,” *JURNAL SAINS DAN SENI ITS*, vol. 8, no. 2, pp. 2337–3520, 2019, Accessed: Jul. 01, 2023. [Online]. Available: http://ejournal.its.ac.id/index.php/sains_seni/article/view/44391
- [9] F. Nur Rozi and D. Harini Sulistyawati, “Klasifikasi Berita Hoax Pilpres Menggunakan Metode Modified K-Nearest Neighbor Dan Pembobotan Menggunakan TF-IDF,” *KONVERGENSI*, vol. 15, no. 1, pp. 1–10, 2019, Accessed: Jul. 01, 2023. [Online]. Available: jurnal.untag-sby.ac.id/index.php/KONVERGENSI/article/view/2828
- [10] W. A. Istiqhfarani, I. Cholissodin, and F. Abdurrachman Bachtiar, “Klasifikasi Penyakit Dental caries menggunakan Algoritme Modified K-Nearest Neighbor,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 5, pp. 1499–1506, 2020, Accessed: Jul. 01, 2023. [Online]. Available: <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/7265>
- [11] Hozairi, Anwari, and S. Alim, “Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes,” *Jurnal Ilmiah NERO*, vol. 6, no. 2, pp. 133–144, 2021, Accessed: Jul. 01, 2023. [Online]. Available: <https://nero.trunojoyo.ac.id/index.php/nero/article/view/237>
- [12] B. Sifa Amalia, Y. Umaidah, R. Mayasari, S. Karawang JI HSRonggo Waluyo, K. Telukjambe Timur, and K. Karawang, “Analisis Sentimen Review Pelanggan Restoran Menggunakan Algoritme Support Vector Machine Dan K-Nearest Neighbor,” *Jurnal Sains, Teknologi dan Industri*, vol. 19, no. 1, pp. 28–34, 2021, Accessed: Jul. 01, 2023. [Online]. Available: ejournal.uin-suska.ac.id/index.php/sitekin/article/view/14861
- [13] K. Permana, Z. Pradana Putra, and A. Nugroho, “Analisa Sentimen Pengunjung Hotel Dengan K-Nearest Neighbor Studi Kasus Hotel Pop! Surabaya,” *Jurnal Ilmu Komputer dan Bisnis*, vol. 12, no. 2, pp. 159–166, Nov. 2021, doi: 10.47927/jikb.v12i2.170.
- [14] Y. Mardiyani and Murwatiningsih, “Pengaruh Fasilitas Dan Promosi Terhadap Kepuasan Pengunjung Melalui Keputusan Berkunjung Sebagai Variabel Intervening Pada Objek Wisata Kota Semarang,” *Management Analysis Journal*, vol. 4, no. 1, pp. 65–75, 2015, Accessed: Jul. 01, 2023. [Online]. Available: journal.unnes.ac.id/sju/index.php/maj/article/view/7220
- [15] P. Vidya Sakta, Indriati, and Marji, “Analisis Sentimen Pariwisata di Kabupaten Malang dengan Menggunakan Metode BM25F, Neighbor Weighted K-Nearest Neighbor dan Seleksi Fitur Chi-Square,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 4, no. 10, pp. 3659–3666, 2020, Accessed: Jul. 01, 2023. [Online]. Available: jptiik.ub.ac.id/index.php/j-ptiik/article/view/8083