

Implementasi Algoritma K - Nearest Neighbor Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia

Risma Putri Fitrianti¹, Ana Kurniawati², Dina Agusten³

^{1,2} Jurusan Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi.

³ Jurusan Teknik Informatika, Fakultas Teknologi Industri.

Universitas Gunadarma

Depok

rismaputrisma@gmail.com, ana@staff.gunadarma.ac.id, dina_agustin@staff.gunadarma.ac.id

Abstrak—*Review* atau sebuah ulasan dapat membantu pemilik restoran untuk dapat melihat bagaimana respon dari konsumen mengenai produknya. Pengklasifikasian dengan data yang berasal dari internet sangat sulit dilakukan karena memiliki informasi dalam berbagai bentuk. Implementasi algoritma K - Nearest Neighbor terhadap analisis sentimen *review* restoran dengan teks Bahasa Indonesia dapat membantu mengatasi hal tersebut dengan sistem otomatis yang dibuat. Hasil akurasi aplikasi pada analisis sentimen menggunakan algoritma K - Nearest Neighbor terhadap *review* restoran Taman Santap Rumah Kayu dengan confusion matrix sebesar 96,61%.

Kata kunci—*analisis sentimen; review restoran; algoritma K - Nearest Neighbor; knn; confusion matrix.*

I. PENDAHULUAN

Berdasarkan artikel yang ditulis Sansan bahwa saat ini perkembangan industri restoran menjadi salah satu sektor penting bagi peningkatan pendapatan nasional. Hasil data Kementerian Perindustrian industri bahwa restoran telah menyumbang 34,95% pada tahun 2017 untuk PDB industri nonmigas menjadi meningkat 4% dari tahun 2016, dapat dipastikan akan terus meningkat pada tahun 2018. *Investor* menilai bahwa industri restoran menjadi lahan subur dan usaha yang menjanjikan di bidang kuliner [1]. Tahun 2019 bisnis ini makin berkembang dengan masuknya banyak restoran asing dengan penyajian inovatif dan menarik bagi masyarakat lokal.

Tren Sentimen (emosi) yang seringkali berubah dapat dituangkan dalam sebuah ulasan (*review*). Adanya pengelolaan analisis sentimen menjadi jembatan komunikasi terhadap sesama pengunjung ataupun pelaku bisnis. Dampak *Review* ini menjadi begitu penting dalam menjalankan bisnis apa pun saat ini seiring berkembangnya teknologi dan internet.

Makalah ini membuat aplikasi untuk menentukan analisis sentimen *review* restoran Taman Santap Rumah Kayu – Summarecon Gading Serpong pada situs <https://www.google.com>, dengan pengklasifikasiannya

menggunakan stemming nazief adriani dan metode K - Nearest Neighbor. Taman Santap Rumah Kayu adalah salah satu restoran terbaik di Tangerang yang paling disukai berdasarkan hasil Google dengan rating 4.4. Pada makalah ini digunakan metode K - Nearest Neighbor yang ditujukan untuk mengolah dan mengklasifikasikan opini sehingga akan diketahui klasifikasi dari opini tersebut (positif atau negatif). Data *review* diambil dalam bahasa Indonesia yang memiliki kecenderungan positif atau negatif untuk membantu usaha dalam menganalisis riset pasar atas opini publik menggunakan metode K - Nearest Neighbor.

Tujuan dari implementasi algoritma K - Nearest Neighbor terhadap analisis sentimen *review* restoran dengan teks Bahasa Indonesia selain dalam hal klasifikasi juga meningkatkan tingkat akurasi yang sebelumnya sudah dilakukan beberapa peneliti lain. Hasil klasifikasi ditampilkan secara visualisasi.

Beberapa penelitian sebelumnya terhadap algoritma K - Nearest Neighbor telah dilakukan. Tahun 2013 Henny Leidiyana melakukan penelitian dengan judul Penerapan Algoritma K - Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor. Hasil penelitian memiliki nilai AUC antara 0.90-1.00, yaitu sebesar 0.98 dan dalam hal klasifikasi termasuk sangat baik tetapi dari tingkat akurasi sebesar 81,46%[2].

Winda Estu Nurjanah, Rizal Setya Perdana, dan Mochammad Ali Fauzi pada tahun 2017, melakukan analisis sentimen terhadap tayangan televisi berdasarkan opini masyarakat pada media sosial twitter menggunakan metode K - Nearest Neighbor dan pembobotan jumlah retweet. Perhitungan akurasi menggunakan berbagai jenis pembobotan yaitu pembobotan tekstual, pembobotan nontekstual dan penggabungan kedua pembobotan. Pembobotan tekstual mencapai 82,50%, pembobotan non tekstual mencapai 60%, menggunakan penggabungan keduanya mencapai 83,33%[3].

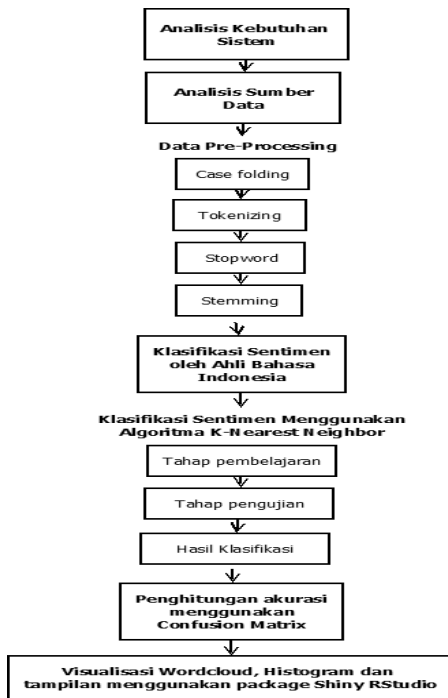
Text Mining Classification Sebagai Rekomendasi Dosen Pembimbing Tugas Akhir Program Studi Sistem Informasi pada tahun 2017 dilakukan oleh Zuliar Efendi dan Mustakim.

Dalam penelitiannya dilakukan pembobotan dengan menggunakan Term Frequency dan Inverse Document Frequency (TF-IDF) dengan hasil akurasi kecil yaitu 60% [4].

Hasil akurasi yang masih dibawah 90% menunjukkan bahwa hasil ke akuratan sebuah ulasan menjadi bahan pertimbangan dalam penelitian ini. Adanya sebuah analisis sentimen dapat membantu pengunjung dalam mendatangi sebuah tempat dan membantu pengusaha untuk meningkatkan hasil produk dan pelayanan agar pengunjung meningkat.

II. METODE PENELITIAN

Penelitian ini dilakukan dalam beberapa tahapan. Gambar 1 menggambarkan beberapa tahapan dalam membuat aplikasi analisis sentimen yang menggunakan algoritma K - Nearest Neighbor.



Gambar 1. Proses metode penelitian

Berikut ini adalah beberapa metode penunjang dalam penelitian:

A. Analisis Sentimen

Metode yang digunakan dalam memahami, mengekstrak data opini serta mengolah data tekstual secara otomatis sehingga didapatkan sebuah sentimen yang terkandung pada sebuah opini seringkali disebut sebagai analisis sentimen. Dalam analisis sentimen terdapat 3 opini yaitu opini negatif dan opini netral dimana dapat di ketahui respon masyarakat terhadap suatu pelayanan atau produk melalui respon balik masyarakat atau ahli[5].

B. Review

Fakta bahwa pengguna mempunyai pandangan berbeda terhadap berbagai jenis produk saat menulis ulasan (*review*) di web tidak dapat dipungkiri. Saat ini sebuah *review* tidak hanya terdiri dari teks saja tetapi dapat berupa *emoticon*, simbol, karakter, tanda baca dan sebagainya. Film, buku dan karya lainnya yang memiliki tujuan untuk mengetahui kualitas, kelebihan dan kekurangan yang ada pada karya tersebut serta untuk melakukan kritik terhadap suatu peristiwa atau karya seni bagi khalayak seringkali disebut sebagai *review* atau ulasan dalam bentuk sebuah teks yang digunakan untuk meninjau suatu karya baik [6].

C. Klasifikasi

Klasifikasi merupakan proses ke dalam sebuah kelas dengan kesamaan sifat yang didalamnya terdapat data latih (data training) sebagai acuan pengelompokkan serta data uji (data testing) untuk pengujian. Tujuannya yaitu objek dari kelas yang belum mempunyai label atau kecenderungan data yang belum terklasifikasi dapat diprediksi [4].

D. K - Nearest Neighbor (KNN)

Penentu klasifikasi berdasarkan contoh dasar yang tidak membangun, representasi deklaratif eksplisit kategori, tetapi bergantung pada label kategori yang melekat pada dokumen pelatihan mirip dengan dokumen tes adalah algoritma K-Nearest Neighbor (KNN). Aplikasi menemukan k tetangga terdekat antara dokumen pelatihan pada tes dokumen. Rata kesamaan setiap dokumen tetangga terdekat dokumen uji digunakan sebagai berat kelas dokumen tetangga[7][8]. Persamaan untuk menghitung tingkat kemiripan tetangga antar 2 objek yaitu [9]:

$$\text{Similarity}(q,d) = \frac{\sum_{j=1}^n (q_{ij} \cdot d_{ij})}{\sqrt{\sum_{j=1}^n (q_{ij})^2 + \sum_{j=1}^n (d_{ij})^2}} \quad (1)$$

Keterangan :

q_{ij} = bobot istilah j pada dokumen i

d_{ij} = bobot istilah j pada dokumen i

E. Confusion matrix

Dalam melakukan klasifikasi, confusion matrix saat ini merupakan alat yang banyak digunakan dalam mengevaluasi performa. Tabel 1 adalah pengelompokkan confusion matrix sesuai karakteristik data *review* terdiri positif dan negatif pada kelas biner[10].

TABEL I. MODEL *CONFUSION MATRIX* [2]

Klasifikasi yang benar	Diklasifikasikan sebagai	
	+	-
+	True Positives	False Negatives
-	False Positives	True Negatives

TP dan TN merupakan hasil penjumlahan record positif dan negatif yang diklasifikasikan dengan tepat, FP dan FN adalah hasil penjumlahan record positif dan negatif yang tidak diklasifikasikan dengan tepat. Berdasarkan confusion matrix tersebut dapat ditentukan kriteria performa seperti *Accuracy*,

Precision, Recall, specificity[10][2]. Persamaan dalam melakukan perhitungan akurasi yaitu [11]:

$$\text{Akurasi} = \frac{TP + TN}{TP + FN + FP} \quad (2)$$

Keterangan:

TP = jumlah true positives, TN = jumlah *true negatives*

FN = jumlah *false negative*, FP = jumlah *false positive*

III. HASIL DAN PEMBAHASAN

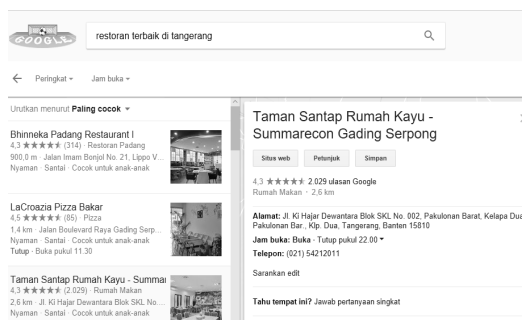
Tahapan yang terdapat dalam gambar 1 diimplementasikan sebagai berikut:

A. Analisis Kebutuhan Aplikasi

Pada analisis kebutuhan tahap - tahap yang akan berjalan pada aplikasi adalah analisis masalah, analisis aplikasi, pengumpulan data, dan analisis kebutuhan fungsional serta non-fungsional.

B. Analisis Sumber Data

Sumber Data pada analisis sentimen merupakan kumpulan *review* yang sudah diberi sentimen positif atau negatif. Data *review* diambil dari sebuah akun google dari restoran Taman Santap Rumah Kayu - Summarecon Gading Serpong yang merupakan salah satu restoran terbaik di Tangerang seperti gambar 2.



Gambar 2. Akun google restoran taman santap rumah kayu - summarecon gading serpong

Review terdiri dari 205 data awal sejak pukul 10.00 WIB pada tanggal 1 Februari 2018 sampai 3 Maret 2018. Data tersebut dimasukan ke dalam microsoft excel secara manual. Proses pelabelan data dilakukan oleh ahli bahasa berdasarkan data excel tersebut. Contoh dari *review* yang termasuk ke dalam pelabelan kelas sentimen positif, "Sebuah resto dengan konsep alam terbuka yg KEREN di Serpong Tangerang, Makanannya enak cocok, terjangkau, ada playground n dikelilingi taman yg rindang. Cocok utk makan bersama teman, pasangan, keluarga dan rekan bisnis.. Pokoknya recommend deh...". Contoh *review* yang termasuk ke dalam pelabelan kelas sentimen negatif, "Proses masakannya Lama untuk takeaway #pemburupoint".

Hasil data pelabelan manual akan digunakan ke dalam sistem. Setelah proses preprocessing terdapat beberapa kalimat

yang hilang atau terhapus sehingga total data *review* menjadi 199 data *review*.

C. Data Pre-Processing

Metode klasifikasi baru dapat dilakukan jika tahap pre-processing sudah selesai diproses. Menyamakan bentuk kata, dihilangkannya gangguan dan meminimalkan jumlah kosa kata adalah tujuan dari tahap pre-processing. Case folding, tokenizing, penghapusan stopword, dan stemming adalah proses yang terdapat dalam tahapan pre-processing.

Mengubah huruf besar menjadi huruf kecil merupakan tahap pada case folding. Pengubahan hanya berlaku untuk huruf tidak berlaku untuk angka. Setelah tahap case folding, dilakukan tahap tokenizing. Tokenizing adalah proses pemotongan kalimat menjadi kata dengan menggunakan pemisah kata seperti tanda titik, tanda koma dan spasi. Angka dan spesial karakter pada kalimat akan dihilangkan.

Dari hasil tokenizing, akan dihilangkan kata-kata yang tidak mengandung makna. Kata-kata tersebut seperti kata sambung, kata depan dan kata ganti. Proses penghilangan kata-kata disebut dengan stopword removal. Cara menghilangkannya dengan membandingkan kata-kata hasil tokenizing dengan daftar kata stopword. Langkah terakhir dari pre-processing adalah mencari kata dasar atau stemming. Algoritma stemming yang digunakan pada penelitian ini mengacu pada algoritma Nazief Andriani.

D. Klasifikasi Sentimen

Klasifikasi dilakukan dengan 2 cara yaitu secara manual dan menggunakan algoritma K - Nearest Neighbor. Hasil klasifikasi manual dilakukan oleh ahli bahasa menggunakan 20 dokumen dari 144 dokumen yang tidak dapat dicantumkan, dapat dilihat pada tabel 2. Cara ini dilakukan untuk melihat perbandingan dari hasilnya.

TABEL II. HASIL KLASIFIKASI MANUAL OLEH AHLI BAHASA

No	Review	Sentimen
1	tempat bagus makan malam keluarga	positif
2	suasana sawah suara gemiricik air kodok ikan bakar	positif
3	makananya kurang enak kurang variatif	negatif
4	restauran gaya leseh utama citra sunda keluarga acara kantor	positif
5	enak makan suasana	positif
6	bagus tempat makan mahal harga makan	positif
7	tempat baik	positif
8	tempat nyaman makan makan enak teman kelompok	positif
9	tempat besar masak baik baik staf banjir butuh koordinasi staf lebih baik	Positif
10	tempat nyaman makan santai rindang pohon arena main anak	Positif
11	resto konsep alam buka keren serpong tangerang makan enak cocok jangkau keliling taman rindang cocok makan teman pasang keluarga rekan bisnis pokok	Positif

Similarity Dokumen 3 dan X

$$\text{Similarity}(q,d) = \frac{\sum_{i=1}^k (q_{ij} \cdot d_{ij})}{\sqrt{\sum_{i=1}^k (q_{ij})^2} \cdot \sqrt{\sum_{i=1}^k (d_{ij})^2}}$$

$$= \frac{(2 \times 0) + (1 \times 0) + (1 \times 0) + (0 \times 1) + (0 \times 1) + (0 \times 1) + (0 \times 1)}{\sqrt{2^2 + 1^2 + 1^2 + 0^2 + 0^2 + 0^2} \cdot \sqrt{0^2 + 0^2 + 0^2 + 1^2 + 1^2 + 1^2}} = 0$$

Similarity Dokumen 4 dan X

$$\text{Similarity}(q,d) = \frac{\sum_{i=1}^k (q_{ij} \cdot d_{ij})}{\sqrt{\sum_{i=1}^k (q_{ij})^2} \cdot \sqrt{\sum_{i=1}^k (d_{ij})^2}}$$

$$= \frac{(1 \times 0) - (1 \times 0) + (1 \times 0) + (0 \times 1) + (0 \times 1) + (0 \times 1) + (0 \times 1)}{\sqrt{1^2 + 2^2 + 1^2 + 0^2 + 0^2 + 0^2} \cdot \sqrt{0^2 + 0^2 + 0^2 + 1^2 + 1^2 + 1^2}} = 0$$

Jika nilai k=3 maka jarak terbesar adalah data ke 1,2,3 atau 1,2,4. Jumlah sentimen terbesar adalah POSITIF, sehingga data X berkategori POSITIF.

• Hasil Klasifikasi

Hasil dari klasifikasi sentimen oleh aplikasi menggunakan algoritma KNN. Pada penelitian ini data uji yang digunakan diambil secara acak sebanyak 59 dari dataset atau 30% dari dataset. Hasil data test dengan klasifikasi menurut ahli Bahasa dapat dilihat pada gambar 3 dan hasil data test dengan klasifikasi menggunakan KNN terdapat pada gambar 4.

```
> c1[test]
[1] positif positif positif negatif positif positif positif positif positif positif positif
[13] positif positif positif positif positif positif positif positif positif positif positif
[25] positif positif positif positif positif positif positif positif positif positif positif
[37] positif positif negatif positif positif positif positif positif positif positif positif
[49] positif positif positif positif positif positif positif positif positif positif positif
Levels: negatif positif
> table(c1[test])
negatif positif
2      57
```

Gambar 3. Hasil klasifikasi menurut ahli bahasa pada data test

Hasil data test dengan klasifikasi menurut ahli Bahasa terdapat 2 review dengan sentimen negatif dan 57 review dengan sentimen positif. Review dengan sentimen negatif terdapat pada urutan review ke-4 dan review ke-39.

```
> knn_pred <- knn(modeldata[train.], modeldata[test.], c1[train.], k=1)
> knn_pred
[1] positif positif positif negatif positif positif positif positif positif positif positif
[13] positif positif positif positif positif positif positif positif positif positif positif
[25] positif positif positif positif positif positif positif positif positif positif positif
[37] positif negatif positif positif positif positif positif positif positif positif positif
[49] positif positif positif positif positif positif positif positif positif positif positif
Levels: negatif positif
> table(knn_pred)
knn_pred
negatif positif
2      57
```

Gambar 4. Hasil klasifikasi menggunakan KNN pada data test

Hasil data test dengan klasifikasi menggunakan K - Nearest Neighbor terdapat 2 review dengan sentimen negatif dan 57 review dengan sentimen positif. Review dengan sentimen negatif terdapat pada urutan review ke-4 dan review ke-38.

E. Confusion matrix

Pengujian analisis sentimen dalam penelitian ini dilakukan menggunakan confusion matrix. Hasil perhitungan sentimen negatif dan positif dapat dilihat pada gambar 5.

Predicted	Actual negatif	positif	Row Total
negatif	1 0.500 0.500	1 0.500 0.018	2 0.034
positif	1 0.018 0.500	56 0.982 0.982	57 0.966
Column Total	2 0.034	57 0.966	59

Gambar 5. Hasil perhitungan confusion matrix 1

Berdasarkan hasil perhitungan confusion matrix didapat jumlah sentimen true negatif sebesar 1 review, false negatif sebesar 1 review, false positif sebesar 1 review dan true positif sebesar 56 review. Hasil perhitungan confusion matrix dapat dilihat pada gambar 6.

```
> conf.mat <- table("Predictions"= knn_pred, "Actual" = c1[test])
> conf.mat
Predictions \ Actual
negatif      1
positif     56
Actual
negatif      1
positif     56
> confu.mat <- confusionmatrix(knn_pred, c1[test])
> confu.mat$byclass
Sensitivity Specificity Pos Pred Value Neg Pred Value Precision
0.50000000 0.98245614 0.50000000 0.98245614 0.50000000
Recall    Prevalance Detection Rate Detection Prevalance
0.50000000 0.50000000 0.03389831 0.01694915 0.03389831
Balanced Accuracy
0.7432907
> confu.mat$overall
Accuracy Kappa AccuracyLower AccuracyUpper AccuracyNull AccuracyPValue McNemarPValue
0.9661017 0.4824561 0.8828515 0.9958681 0.9661017 0.6767032 1.0000000
> confu.mat$overall["Accuracy"]
Accuracy
0.9661017
```

Gambar 6. Hasil perhitungan confusion matrix

F. Visualisasi

Hasil visualisasi dengan menggunakan package wordcloud di R Studio digunakan untuk menunjukkan kelompok kata-kata yang terkandung dalam sentimen negatif atau kata-kata dalam sentimen positif. Hasil visualisasi wordcloud positif dapat dilihat pada gambar 7.



Gambar 7. Hasil visualisasi wordcloud positif

Hasil wordcloud positif terdapat kata tempat, enak, bagus, keluarga, nyaman, anak, harga dan baik. Gambar 8 hasil visualisasi wordcloud negatif.



Gambar 8. Hasil visualisasi wordcloud negatif

G. Hasil Pengujian

Pengujian akurasi dilakukan dengan metode confusion matrix yang membandingkan hasil pengujian sebanyak 59 data uji. Hasil pengujian terdapat pada tabel 4.

TABEL IV. CONFUSION MATRIX

Jumlah Data Uji : 59	Sentimen Hasil	Sentimen Hasil
	Analisis POSITIF	Analisis NEGATIF
Sentimen Asli POSITIF	True Positives = 56	False Negatives = 1
Sentimen Asli NEGATIF	False Positives = 1	True Negatives = 1

Setelah aplikasi selesai pada tahap klasifikasi, maka hitung nilai akurasi menggunakan (2):

- Menghitung Akurasi

$$\text{Akurasi} = \frac{TP+TN}{TP+FN+FP}$$

$$\text{Akurasi} = \frac{56+1}{56+1+1} = 0,9661$$

- Menghitung tingkat kesalahan pada aplikasi

$$\text{Error Rate} = 1 - \text{Akurasi}$$

$$\text{Error Rate} = 1 - 0,9661 = 0,0339$$

Data pengujian akurasi yang digunakan pada tabel 4 sebanyak 59 *review*. Berdasarkan pengujian akurasi, didapatkan hasil akurasi klasifikasi *review* dari aplikasi analisis sentimen dengan menggunakan KNN sebesar 96,61%. Hasil confusion matrix oleh aplikasi dapat dilihat pada gambar 14.

```
> conf.mat <- table("Predictions"= knn_pred, "Actual" = c1[test] )
> conf.mat
      Actual
Predictions negatif positif
negatif      1      1
positif      1     56
> confu.mat <- confusionMatrix(knn_pred, c1[test])
> confu.mat$byClass
      sensitivity      specificity      Pos Pred Value      Neg Pred Value      Precision
0.50000000      0.98245614      0.50000000      0.98245614      0.50000000
recall      1.00000000      prevalence      0.03389831      detection Rate      0.03389831
balanced accuracy      0.74122807
> confu.mat$overall
      Accuracy      Kappa      AccuracyLower      AccuracyUpper      AccuracyNull      AccuracyPValue      McNemarPValue
0.9661017      0.4824561      0.8828515      0.9958681      0.9661017      0.6767032      1.0000000
> confu.mat$overall[,"Accuracy"]
      Accuracy
0.9661017
```

Gambar 14. Hasil confusion matrix

IV. KESIMPULAN

Dalam makalah ini pembuatan aplikasi analisis sentimen terhadap *review* pelanggan restoran dengan teks Bahasa Indonesia menggunakan algoritma K - Nearest Neighbor telah berhasil dilakukan. Hasil dari penelitian ini adalah:

- Melakukan tahapan penelitian yang terdiri dari analisis kebutuhan, analisis sumber data, data pre-processing, klasifikasi manual oleh ahli Bahasa, klasifikasi sentimen menggunakan algoritma k-nearest neighbor, penghitungan akurasi dengan confusion matrix dan visualisasi wordcloud, histogram dan tampilan menggunakan package shiny R Studio.
- Menggunakan data *review* pelanggan restoran sebanyak 205 data dan setelah melalui tahap pre-processing menjadi 199 data dikarenakan beberapa data mengalami reduksi karena terdapat noise atau kata-kata pada Bahasa Indonesia yang tidak sesuai.

- Menggunakan data training sebesar 70% dari dataset yaitu sebanyak 140 data dan data test sebesar 30% dari dataset sebanyak 59 data. Pembagian data training dan data test dilakukan secara acak dari dataset.
- Melakukan pengujian menggunakan confusion matrix dan didapatkan hasil akurasi dari analisis sentimen terhadap *review* restoran dengan teks Bahasa Indonesia menggunakan K - Nearest Neighbor dengan persentase 96,61%, nilai k = 1 dan nilai error rate sebesar 3,39%.

Berdasarkan hasil analisis sentimen penelitian ini dapat disimpulkan bahwa restoran Taman Santap Rumah Kayu Tangerang memiliki kecenderungan sentimen positif. Adanya analisis sentimen dapat membantu masyarakat dan pengusaha dalam keakuratan hasil respon pengunjung karena jika hanya melihat rating terkadang tidak sesuai dengan komentar yang diberikan contoh memberi rating 4 tetapi kecenderungan kalimat tidak positif.

REFERENSI

- Sansan, "Perkembangan Industri Restoran", 24 September 2018, <http://keysoft.co.id/perkembangan-industri-restoran/>, [diakses: Mei 2019].
- H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor", Jurnal Penelitian Ilmu Komputer, System Embedded & Logic, 2013, vol. 1, No. 1, pages. 65-76.
- W. E. Nurjanah, R. S. Perdana, dan M. A. Fauzi, 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, vol. 1, No. 12, hlm. 1750-1757.
- Z. Efendi dan Mustakim, "Text Mining Classification Sebagai Rekomendasi Dosen Pembimbing Tugas Akhir Program Studi Informasi", Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 9, 18-19 Mei 2017, hal. 235-242, ISSN (Online) : 2579-5406.
- N. T. Romadloni, I. Santoso, dan S. Budilaksono, "Perbandingan Metode Naive Bayes, KNN dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line.", IKRA-ITH INFORMATIKA: Jurnal Komputer dan Informatika 3.2 (2019): 1-9.
- R. I. Pristiyanti, M. A. Fauzi dan L. Muflikhah, "Sentiment Analysis Peringkasan *Review* Film Menggunakan Metode Information Gain dan K-Nearest Neighbor", Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 03 Maret 2018, vol. 2, no. 3, hlm. 1179-1186.
- S. Sahara dan M. Wahyudi. "K-Nearest Neighbors Sebagai Analisis Sentimen *Review* Produk Appstore For Android." Konferensi Nasional Ilmu Pengetahuan dan Teknologi 1.1 (2015): 77-82.
- T. Songbo, and J. Zhang. "An empirical study of sentiment analysis for chinese documents." Expert Systems with applications 34.4 (2008): 2622-2629.
- O. Nurdiana, Jumadi dan D. Nusanika, "Perbandingan Metode Cosine Similarity Dengan Metode Jaccard Similarity Pada Aplikasi Pencarian Terjemah Al-Qur'an Dalam Bahasa Indonesia", 2016, JOIN. vol 1, no. 1.
- R. Siringoringo dan J. Jamaludin, "Text Mining dan Klusterisasi Sentimen Pada Ulasan Produk Toko Online", Jurnal Teknologi dan Ilmu Komputer Prima (JUTIKOMP), 03 April 2019, vol.2(1), pp.41-8.
- P. Mayadewi dan E. Rosely, "Prediksi Nilai Proyek Akhir Mahasiswa Menggunakan Algoritma Klasifikasi Data Mining" Seminar Nasional Informasi Indonesia, 2015.