ANALISIS SENTIMEN TERHADAP PENILAIAN CUSTOMER DI PHD KARAWITAN MENGGUNAKAN METODE K-NEAREST NEIGBOR

Taufik Hidayat Supratman

Teknik Informatika –Universitas Komputer Indonesia
Jl. Dipatiukur 112-114 Bandung
E-mail : akasia_a7x@yahoo.com

ABSTRAK

PHD (*Pizza hut delivery*) menyediakan wadah untuk berbagi pengalaman customer berupa penilaian terhadap outlet PHD yang dikunjunginya yang dapat disalurkan melalui website www.halopizzaindonesia.com. Opini yang masuk akan didistribusikan ke masing-masing outlet PHD termasuk outlet PHD Karawitan. Outlet Manager ditugaskan untuk menghitung persentase penilaian kepuasan *customer* dengan cara mempelajari opini yang disampaikan oleh customer dan mengelompokkan penilaian kedalam beberapa klasifikasi yaitu opini puas dan opini tidak puas. Jumlah penilaian yang banyak membuat outlet manager mengalami kesulitan dikarenakan harus mengelompokkan opini satu persatu dari sekian banyaknya opini yang ada lalu menghitung persentase kepuasan *customer* secara manual.

Untuk mempermudah pekerjaan *outlet manager* terutama khusus dalam pengelompokkan penilaian berupa opini *customer*, maka diperlukan suatu metode yang dapat menganalisis opini secara otomatis. Salah satu cara yang dapat digunakan untuk mengelompokan opini yaitu dengan menggunakan metode analisis sentimen. Metode analisis sentimen yang di gunakan dalam penelitian ini yaitu menggunakan algoritma K-nearest neigbor. Tahapannya terdiri dari preprocessing, pembobotan kata, dan klasifikasi k-nearest neigbor.

Berdasarkan hasil implementasi dan pengujian yang telah dilakukan algoritma k-nearest neigbor dapat diterapkan pada kasus penelitian *customer* di PHD Karawitan dan dapat membantu mempermudah *outlet manager* dengan pengelompokkan penilaian opini puas dan opini tidak puas secara otomatis serta dapat menghitung persentase opini tersebut.

Kata kunci : *K-nearest neigbor*, Opini, *Pizza Hut delivery*, *PHD*.

1. PENDAHULUAN

PHD (Pizza hut delivery) adalah perusahaan yang bergerak di bidang kuliner. Untuk mengetahui pelayanan yang diberikan dari tiap outlet, maka PHD menyediakan wadah untuk berbagi pengalaman customer berupa penilaian terhadap outlet PHD yang dikunjunginya. Penilaian customer dapat disalurkan melalui website www.halopizzaindonesia.com. Semua penilaian yang tersimpan di website tersebut akan didistribusikan ke masing-masing outlet PHD diseluruh Indonesia. Outlet Manager ditugaskan untuk menghitung persentase penilaian yang masuk seberapa besar kepuasan customer yang di dapat dengan cara mempelajari opini yang disampaikan oleh customer

Dalam penelitian ini outlet PHD yang akan dijadikan tempat penelitian yaitu outlet PHD Karawitan. Jumlah penilaian yang banyak membuat outlet manager mengalami kesulitan dikarenakan harus mengelompokkan opini satu persatu dari sekian banyaknya opini yang ada lalu menghitung persentase kepuasan customer secara manual. Salah satu cara yang dapat digunakan untuk mengelompokan opini yaitu dengan menggunakan metode analisis sentimen.

Analisis sentimen merupakan suatu metode dalam menganalisis pendapat, penilaian, dan sikap terhadap suatu entitas seperti produk, pelayanan, peristiwa, dan topik. [1] Berdasarkan penelitian sebelumnya pada pengklasifikasian text yang mencoba membandingkan tingkat akurasi dari 3 algoritma langsung yaitu metode naive bayes, metode k-nearest neighbor(KNN), metode support vector machine(SVM) didapat bahwa metode K-NN memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan metode naive bayes dan metode support vector machine(SVM). [2].

1.1 Metode Pembangunan Analisis Sentimen

Sistem analisis sentimen yang akan dibangun akan melewati tahapan tahapan terdiri dari :

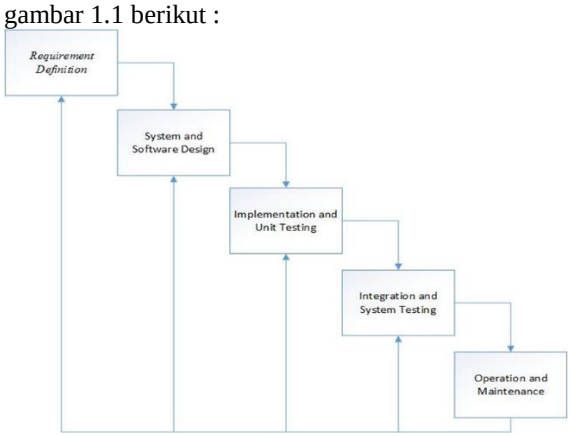
a. Analisis Sumber Data

Data yang digunakan adalah data yang diperoleh dari website www.halopizzaindonesia.com yang sudah didistribusikan ke outlet PHD Karawitan. File data berformat .xls dan tahapan pengolahan data untuk selanjutnya akan dilakukan proses import data.

b. Preprocessing
Pada tahap ini, data yang terkumpul akan diproses sehingga data yang didapat menjadi lebih terstruktur dan mudah untuk diolah. Langkah - langkah preprocessing terdiri dari case folding, tokenizing, stopword removal dan stemming.
c. Pembobotan Kata
Pada tahap ini akan dilakukan proses pengekstrakan keyword menggunakan nilai TF-IDF (Term Frequency – Inverse Document Frequency). Term (kata) di ambil dari hasil proses prepocessing terakhir yaitu stemming. Nilai dari hasil pembobotan akan digunakan sebagai tahapan pengklasifikasian menggunakan metode k-nearest neigbor.
d. Klasifikasi Sentimen
Langkah selanjutnya adalah proses pengklasifikasian yang akan diproses menggunakan metode k-nearest neigbor untuk menentukan mana yang termasuk opini puas dan mana yang termasuk opini tidak puas.
e. Visualisasi
Hasil dari tahap klasifikasi sentimen akan digambarkan dalam bentuk diagram lingkaran. Data diambil dari jumlah opini puas dan opini tidak puas lalu ditampilkan dalam bentuk persentase pada diagram

1.2 Metode Pembangunan Perangkat Lunak

Metode pengembangan perangkat lunak yang digunakan pada penelitian ini adalah metode waterfall menurut referensi Ian Sommerville [3]. Secara garis besar metode waterfall terlihat seperti gambar 1.1 berikut :



Gambar 1.1 Metode Waterfall

2 ISI PENELITIAN

2.1 Analisis Sentimen

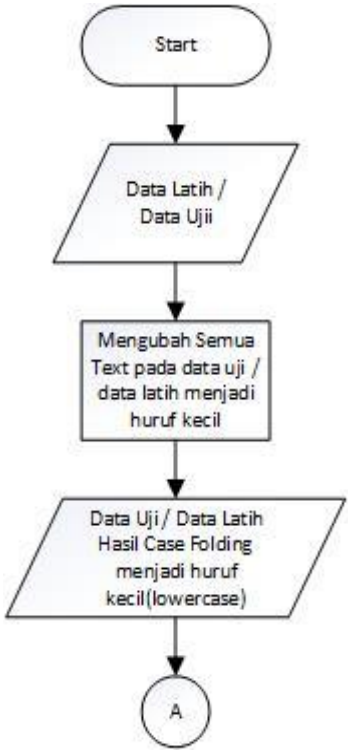
Analisis sentimen merupakan suatu cara dalam menganalisis pendapat, penilaian, dan sikap terhadap suatu entitas seperti produk, karakter orang, pelayanan, peristiwa dan topic. Salah satu contoh penggunaan analisis sentimen dalam dunia nyata adalah identifikasi kecenderungan pasar dan opini pasar terhadap suatu objek barang. [4]

2.2.1 Preprocessing

Preprocessing merupakan proses menggali, mengolah, mengatur informasi dengan cara menganalisis hubungannya, aturan-aturan yang ada di data tekstual semi terstruktur atau tidak terstruktur. bagian – bagian tahapan dari preprocessing, dapat dijelaskan dengan menggunakan flowchart berjenis Flowchart Sistem, yaitu bagan yang menunjukkan alur kerja atau apa yang sedang dikerjakan di dalam sistem secara keseluruhan dan menjelaskan urutan dari prosedur-prosedur yang ada di dalam sistem [11].

1. Case Folding

Pada tahap ini, semua huruf akan diubah menjadi huruf kecil.
langkah-langkah pada tahap case folding adalah sebagai berikut :
1. Memeriksa ukuran setiap karakter dari awal sampai akhir karakter.
2. Jika ditemukan karakter yang menggunakan huruf kapital (uppercase), maka huruf tersebut akan diubah menjadi huruf kecil (lowercase). Berikut flowchart case folding terlihat pada gambar 2.1



Gambar 2.1*Flowchart Case folding*

2. Convert Negation

Convert Negation merupakan proses konversi kata-kata negasi yang terdapat pada suatu opini, karena kata negasi mempunyai pengaruh dalam merubah nilai sentimen pada suatu tweet. Jika terdapat kata negasi maka akan disatukan dengan kata setelahnya. Kata - kata negasi tersebut meliputi

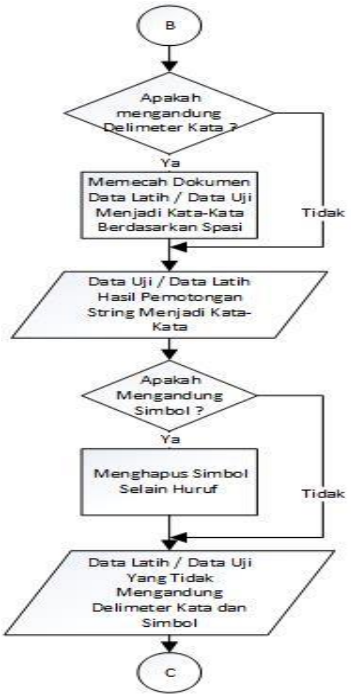
kata “bukan”, “tidak”, “tak”, “ga”,”gak”, “enggak”, “jangan”, dan ”nggak”.
Langkah – langkah pada tahap convert negation adalah sebagai berikut :
1. Kata yang digunakan adalah hasil dari case folding
2. Jika ditemukan opini yang mengandung kata – kata negasi maka akan disatukan kata negasi tersebut dengan kata setelah kata negasi tersebut.
Berikut *flowchart Convert Negation* terlihat pada gambar 2.2



Gambar 2.2 *flowchart Convert Negation*

3. Tokenizing

Tokenizing merupakan tahap pemotongan kalimat berdasarkan tiap kata yang menyusunnya. Proses ini melakukan penguraian deskripsi yang semula berupa kalimat-kalimat menjadi kata-kata dan menghilangkan simbol seperti titik(.), tanda seru(!), tanda tanya (?), koma(,), spasi, emoticon.
Langkah-langkah pada tahap tokenizing adalah sebagai berikut:
1. Kata yang digunakan adalah hasil dari convert negation.
2. Memotong setiap kata dalam kalimat berdasarkan pemisah kata yaitu spasi.
3. Menghilangkan simbol seperti titik(.), tanda seru(!), tanda tanya (?), koma(,), spasi, emoticon.
Berikut *flowchart Tokenizing* terlihat pada gambar 2.3



Gambar 2.3 *flowchart Tokenizing*

4. Stopword Removal

Stopword didefinisikan sebagai term yang tidak berhubungan dengan subyek utama dari database meskipun kata tersebut sering kali hadir di dalam dokumen dan kata yang dianggap tidak dapat memberikan pengaruh dalam menentukan suatu kategori sentimen.

Kata-kata tersebut dimasukkan kedalam daftar stopword yang biasanya berupa :
1. Kata ganti orang. Hanya dapat digunakan untuk mengganti nomina orang, nama orang, atau hal-hal lain yang dipersonifikasikan. Misalnya : ia, Saudara, Bapak, Ibu, Tuan, Nyonya, Mba, Mr, Mrs, karyawan, karyawati, pegawai dsb
2. Kata ganti penanya. Misalnya : apa, kapan, mengapa, siapa, bagaimana, berapa, di mana, ke mana, di dsb
3. Kata ganti petunjuk. Misalnya : ini, itu dsb
4. Kata ganti penghubung. Misalnya : yang, dan, atau dsb
5. Kata irrelevant. Misalnya : salah satu, karena, sangat, juga, agak, dengan, harus, dari, dgn, dg, yg, oke dsb
Langkah-langkah pada stopword removal adalah sebagai berikut:
1. Kata hasil tokenizing akan dibandingkan dengan daftar stopword.Dilakukan pengecekan apakah kata sama dengan daftar stopword atau tidak.
2. Jika kata sama dengan yang ada pada daftar stopword, maka akan dihilangkan.
Berikut *flowchart StopWord Removal* terlihat pada gambar 2.4



Gambar 2.4 *flowchart StopWord Removal*

5. Stemming

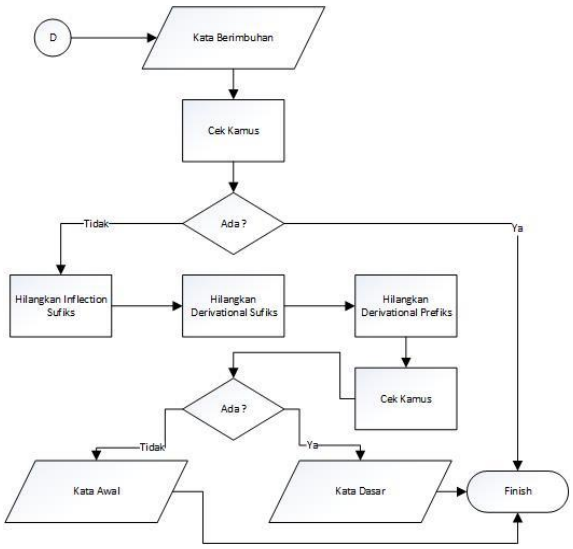
Stemming merupakan tahap untuk mentransformasi kata-kata yang terdapat dalam suatu dokumen ke kata-kata akarnya (root word) dengan menggunakan aturan-aturan tertentu. Dengan menggunakan stemming dapat mengurangi variasi kata yang sebenarnya memiliki kata dasar yang sama. Salah satu algoritma stemming yaitu Algoritma Nazief dan Adriani[5].

Langkah – langkah stemming menggunakan Algoritma Nazief dan Adriani adalah sebagai berikut :

1. Kata yang belum di-stemming dicari pada kamus. Jika kata itu langsung ditemukan, berarti kata tersebut adalah kata dasar. Kata tersebut dikembalikan dan algoritma dihentikan.
2. Hilangkan inflectional suffixes terlebih dahulu. Jika hal ini berhasil dan suffix adalah partikel (“lah” atau ”kah”), langkah ini dilakukan lagi untuk menghilangkan inflectional possessive pronoun suffixes (“ku”, “mu” atau”nya”).
3. Derivational suffix (“-i”, “-an” dan “kan”) kemudian dihilangkan. Lalu langkah ini dilanjutkan lagi untuk mengecek apakah masih ada derivational suffix yang tersisa, jika ada maka dihilangkan. Jika tidak ada lagi maka lakukan langkah selanjutnya.
4. Kemudian derivational prefix (“di-“,”ke-“,”se-“,”te-“,”be-“,”me-“ dan “per-“ “) dihilangkan. Lalu langkah ini dilanjutkan lagi untuk mengecek apakah masih ada derivational prefix yang tersisa, jika ada maka dihilangkan. Jika tidak ada lagi maka lakukan langkah selanjutnya.
5. Setelah tidak ada lagi imbuhan yang tersisa, maka algoritma ini dihentikan kemudian kata dasar tersebut dicari pada kamus, jika kata dasar tersebut ketemu berarti algoritma ini berhasil tapi jika kata dasar tersebut tidak ketemu pada kamus, maka dilakukan recoding.
6. Jika semua langkah telah dilakukan tetapi kata dasar tersebut tidak ditemukan pada kamus juga maka algoritma ini mengembalikan kata yang asli sebelum dilakukan stemming.

Proses stemming dalam penelitian ini merupakan proses terakhir dari tahap preprocessing, setelah opini hasil preprocessing sudah dilakukan maka dilakukan pembobotan kata agar opini bisa diklasifikasikan menggunakan metode KNN.

Berikut *flowchart Stemming* terlihat pada gambar 2.5



Gambar 2.5 *flowchart Stemming*

2.2.2 Pembobotan Kata(TF IDF)

Term weighting merupakan tahapan untuk memberikan suatu nilai/bobot pada term yang terdapat pada suatu dokumen setelah melewati preprocessing.

$$\text{Idf} = \log\left(\frac{N}{Dfi}\right) \quad (2.2)$$

IDF = inverse document frequency
N = Jumlah kalimat yang berisi term(t)
Dfi = Jumlah kemunculan term terhadap D

Pembobotan kata dilakukan setelah melalui tahap preprocessing, nilai dari hasil pembobotan kata maka akan digunakan untuk menghitung nilai kemiripan antar dokumen(Cosine Similarity) yang dimana merupakan tahap dalam pengklasifikasian opini menggunakan metode KNN.

Berikut merupakan langkah – langkah dalam pembobotan :

1. Buat Susunan per record kata – kata baik opini data latih ataupun data uji, kata – kata yang muncul lebih dari 1 hanya dituliskan 1x
2. Isi angka 1 apabila katanya muncul di setiap (tf) baik data latih atau data uji, apabila tidak muncul isi angka 0
3. Hitung jumlah Kata – kata yang muncul di semua data latih dan uji(df)
4. Hitung Idf menggunakan persamaan (2.2)

Kalikan masing – masing W dengan masing – masing Tf sesuai dengan pasanganya. Contoh : Tf(i) *W(i)

Contoh kasus :
k1 : layanan cepat ramah
k2 : layanan kurang muas kurang sopan

Tabel 2.1 PembobotaN Kata

Kata	Tfk1	Tf k2	Df	Idf	Wk1	Wk2
layanan	1	1	2	0	0	0
cepat	1	0	1	0.30	0.30	0
ramah	1	0	1	0.30	0.30	0
kurang	0	2	2	0.30	0	0.60
muas	0	1	1	0.30	0	0.30
sopan	0	1	1	0.30	0	0.30

Hasil nilai dari pembobotan akan digunakan untuk perhitungan cosine simmilirity pada metode KNN.

2.2.3 Klasifikasi K-Nearest Neigbor(KNN)

K-Nearest Neighbor (KNN) adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari query instance yang baru diklasifikan berdasarkan mayoritas dari kategori pada KNN. KNN memiliki beberapa kelebihan yaitu ketangguhan terhadap training data yang memiliki banyak noise dan efektif apabila training data-nya besar. Sedangkan, kelemahan KNN adalah KNN perlu menentukan nilai dari parameter k (jumlah dari tetangga terdekat), training berdasarkan jarak tidak jelas mengenai jenis jarak apa yang harus digunakan dan atribut mana yang harus digunakan untuk mendapatkan hasil terbaik, dan biaya komputasi cukup tinggi karena diperlukan perhitungan jarak dari tiap query instance pada keseluruhan training sample. [6].



Gambar 2.6 Alir Diagram K-Nearest Neigbor(KNN)

Langkah – langkah metode KNN :
1. Hitung jarak antara data sampel (data uji) dengan data latih yang telah dibangun. Salah satu persamaan dalam menghitung jarak kedekatan Cosine Similarity.

$$Cos(i,k) = \frac{\sum_k (d_i d_k)}{\sqrt{\sum_k d^2{}_{ik}}\sqrt{\sum_k d^2{}_{jk}}}$$

Keterangan :

$\sum_k (d_i d_k)$: Vector dot produk dari i dan k

$\sqrt{\sum_k d^2{}_{ik}}$: Panjang vector i

$\sqrt{\sum_k d^2{}_{jk}}$: Panjang vector k

i : data uji ke-i
k : data latih ke-j

2. Menentukan parameter nilai k = jumlah tetanggaan terdekat bebas.
3. Mengurutkan jarak terkecil dari data uji
4. Pasangkan kategori sesuai dengan kesesuaian

Cari jumlah terbanyak dari tetanggaan terdekat Kemudian tetapkan kategori. [7]

2.2.4 Pengujian *Confusion Matrix*

Confusion Matrix [8] berisi informasi tentang klasifikasi aktual dan yang telah diprediksi yang dilakukan oleh sistem klasifikasi. Kinerja sistem tersebut umumnya dievaluasi dengan menggunakan data dalam matriks. Tabel berikut menunjukkan confusion matrix untuk klasifikasi dua kelas.

		Diklasifikasikan sebagai (*Predicted Class*)	
		Puas	Tidak Puas
Klasifikasi yang benar (*Actual Class*)	Puas	True Puas	False Tidak Puas
	Tidak Puas	False Puas	True Tidak Puas

Tabel 2.2 *Confusion Matrix*

2.2.5 Pengujian *K-Fold Cross-Validation*

Dalam pendekatan K-fold cross-validation, setiap record digunakan beberapa kali dalam jumlah yang sama untuk training dan tepat sekali untuk testing. Untuk mengilustrasikan metode ini, anggaplah seperti mempartisi data ke dalam dua subset yang berukuran sama. Pertama, pilih satu dari kedua subset tersebut untuk training dan satu lagi untuk testing. Kemudian dilakukan pertukaran fungsi dari subset sedemikian sehingga subset yang

sebelumnya sebagai training set menjadi test set demikian sebaliknya. Pendekatan ini dinamakan two-fold cross validation. Total error diperoleh dengan menjumlahkan error –error untuk kedua proses tersebut.Setiap record digunakan tepat satu kali untuk training dan satu kali untuk testing. Metode k-fold cross-validation mengeneralisasi pendekatan ini dengan mensegmentasi data ke dalam k partisi berukuran sama. Selama proses, salah satu dari partisi dipilih untuk testing, sedangkan sisanya digunakan untuk training. Prosedur ini diulangi k kali sedemikian sehingga setiap partisi digunakan untuk testing tepat satu kali. Total error ditentukan dengan menjumlahkan error untuk semua k proses tersebut [9].

Metode k-fold cross-validation menetapkan k = N, ukuran dari data set. Metode ini dinamakan pendekatan leave-one-out, setiap test set hanya mengandung satu record. Pendekatan ini memiliki keuntungan dalam pengunaan sebanyak mungkin data untuk training. Test set bersifat mutually exclusive dan secara efektif mencakup keseluruhan data set. Kekurangan dari pendekatan ini adalah banyaknya komputasi untuk mengulangi prosedur sebanyak N kali.

K-fold cross-validation adalah salah satu teknik untuk mengevaluasi keakuratan model, dengan ciri-ciri [10]:

1. Mempartisi data secara random ke dalam k buah himpunan/fold yaitu D1, D2, ..Dk. Setiap kelompok mempunyai jumlah yang hampir sama.
2. Pada perulangan i, gunakan Di sebagai data uji dan himpunan lainnya sebagai data pelatihan Contoh :

a. Pada perulangan ke-1 : D1 sebagai data uji dan D2 s.d .Dk sebagai data pelatihan
b. Pada perulangan ke-2 : D2 sebagai data uji dan D1, D3 s.d. Dk sebagai data pelatihanan dan seterusnya
c. Melakukan training dan pengujian sebanyak k kali
d. Menghitung keakuratan dengan rumus.

Sebagai gambaran jika kita melakukan 3-Fold Cross-Validation maka desain data eksperimenya adalah sebagai berikut :

Dataset : Tabel 1.3 Data Set

K1	K2	K3

Data Subset : Tabel **Error! No text of specified style in document.**.4 Data Subset

Eksperimen ke	Training	Testing
1.	K2, K3	K1
2.	K1, K3	K2
3.	K1, K2	K3

2.2.6 Pengujian Blackbox

Pengujian black box memiliki beberapa jenis. Pengujian blackbox yang akan digunakan dalam penelitian ini yaitu pengujian fungsional. Pada pengujian ini, perangkat lunak di uji untuk persyaratan fungsional. Pengujian dilakukan dalam bentuk tertulis untuk memeriksa apakah aplikasi berjalan seperti yang diharapkan. Pengujian fungsional meliputi seberapa baik sistem melaksanakan fungsinya. [12] .

contoh kasus pengujian blackbox :

Pengujian black box yang dilakukan pada aplikasi yang dibuat dibagi menjadi 2 bagian, yaitu skenario pengujian, kasus dan hasil pengujian.

1. Skenario Pengujian

Skenario pengujian menjelaskan pengujian terhadap sistem yang ada pada aplikasi analisis sentimen ini. Skenario pengujian yang akan dilakukan pada aplikasi ini selengkapnya dapat dilihat pada Tabel 2.5.

No.	Komponen Yang Di uji	Jenis Pengujian
1.	Proses Pengolahan Data Latih	*Black Box*
2.	Proses Pengolahan Data Uji	*Black Box*
3.	Proses Klasifikasi	*Black Box*
4.	Proses Visualisasi	*Black Box*

Tabel 2.5 Skenario Pengujian *blackbox*

2. Kasus dan Hasil Pengujian

Kasus dan hasil pengujian berisi pemaparan dari rencana pengujian yang telah disusun pada skenario pengujian. Pengujian ini dilakukan secara black box dengan memperhatikan setiap kasus yang akan di uji. Berikut ini pemaparan dari setiap butir pengujian yang terdapat pada skenario pengujian:

a. Pengujian Pengolahan Data Latih

Pengujian Pengolahan Data Latih memaparkan pengujian yang dilakukan terhadap aktivitas pengguna saat akan melakukan pengambilan data latih. Pengujian pengolahan data latih yang akan dilakukan pada aplikasi ini selengkapnya dapat dilihat pada Tabel 2.6.

Tabel 2.6 Pengujian Pengolahan Data Latih

Kasus Yang Akan Di Uji	Hasil Uji	
	Hasil yang diharapkan	Kesimpulan
Pengolahan Data Latih	sistem berhasil melakukan penyimpanan data latih yang telah ter-*processing* otomatis pada database	[√] Berhasil [ ] Tidak Berhasil

b. Pengujian Pengolahan Data Uji

Pengujian Pengolahan Data uji memaparkan pengujian yang dilakukan terhadap aktivitas pengguna saat akan melakukan pengambilan data uji. Pengujian pengolahan data uji yang akan dilakukan pada aplikasi ini selengkapnya dapat dilihat pada Tabel 2.7.

Tabel 2.7 Pengujian Ambil Data Uji

Kasus Yang Akan Di Uji	Hasil Uji	
	Hasil yang diharapkan	Kesimpulan
Pengolahan Data Uji	sistem berhasil melakukan penyimpanan data uji pada database	[√] Berhasil [ ] Tidak Berhasil

c. Pengujian Klasifikasi

Pengujian klasifikasi memaparkan pengujian yang dilakukan terhadap aktivitas pengguna saat akan melakukan klasifikasi opini. Pengujian klasifikasi yang akan dilakukan pada aplikasi ini selengkapnya dapat dilihat pada Tabel 2.8.

Tabel 2.8 Pengujian Klasifikasi

Kasus Yang Akan Di Uji	Hasil Uji	
	Hasil yang diharapkan	Kesimpulan
Mengklasifikasikan penilaian *customer* (Data uji)	sistem berhasil mengklasifikasikan penilaian customer(data uji) dalam sentimen puas atau tidak puas	[√] Berhasil [ ] Tidak Berhasil

d. Pengujian Visualisasi

Pengujian visualisasi memaparkan pengujian yang dilakukan terhadap aktivitas pengguna saat akan melakukan visualisasi ke dalam diagram lingkaran. Pengujian Visualisasi yang akan dilakukan pada aplikasi ini selengkapnya dapat dilihat pada Tabel 2.9.

Tabel 2.9 Pengujian Visualisasi

Kasus Yang Akan Di Uji	Hasil Uji	
	Hasil yang diharapkan	Kesimpulan
Menampilkan Diagram persentase sentimen puas dan sentimen tidak puas	sistem berhasil Menampilkan Diagram persentase sentimen puas dan sentimen tidak puas sesuai dengan data yang sudah diklasifikasikan sebelumnya	[√] Berhasil [ ] Tidak Berhasil

\

3 PENUTUP

3.1 KESIMPULAN

Berdasarkan hasil implementasi dan pengujian yang telah dilakukan pada sistem analisis sentimen dengan menggunakan metode *K-nearest neigbor* ini maka dapat diambil kesimpulan bahwa, sistem analisis sentimen ini dapat mengklasifikasikan opini secara otomatis dan dapat memberikan informasi mengenai persentase sentimen puas dan sentimen tidak puas sehingga dapat membantu mempermudah pekerjaan *Outlet Manager* khususnya dalam pengklasifikasian penilaian customer(opini).

3.2 SARAN

Saran untuk pengembangan lebih lanjut, yaitu sebagai berikut:

1. Penambahan jumlah data latih untuk mendapatkan hasil yang lebih baik saat klasifikasi penilaian customer.
2. Pengoptimalan waktu dalam pemprosesan sistem.

DAFTAR PUSTAKA

[1] B. Liu, "Sentiment Analysis and Opinion Mining, Morgan & Claypool," 2012.

[2] K. Khamar, "Short Text Classification Using KNN Based on Distance Function," International journal of advanced research in computer and communication engineering, 2013.

[3] I. Sommerville, Software Engineering, jakarta: Erlangga, 2003.

[4] B. P. a. L. Lee, ""Opinion Mining and Sentiment Analysis"," Foundations and Trends in Information Retrieval, vol II, no 1-2, pp. 1-135.

[5] Agusta, "Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia," Konferensi Nasional Sistem dan Informatika, 2009.

[6] Evan, Buku TA : K-Nearest Neigbor(KNN), Jakarta: Informatika, 2010.

[7] D. E. Garcia, "The Classic Vector Space Model (Description, Advantages and Limitations of the Classic Vector Space Model)," 2005.

[8] K. a. Provost, "On Applied Research in Machine Learning," In Editorial for the Special Issue on

Applications of Machine Learning and the Knowledge Discovery Process, 1998.

[9] J.Han, "Data Mining : Concepts and techniques, Morgan Kaufman,," 2006.

[10] Nugroho, "SVM : Paradigma Baru dalam Softcomputing dan Aplikasinya," Konferensi Nasional Sistem & Informatika, 2008.

[11] Sudarsono, "[Online] Analisis Dan Perancangan Sistem," Available : sdarsono.staff.gunadarma.ac.id/Downloads/files/16512/Flowchart.pdf [Accessed 15 Agustus 2015].

[12]J. Simarmata, rekayasa perangkat lunak, Yogayakarta: Andi.