



Similarity Report

Metadata

Name of the organization

Universitas Muhammadiyah Sidoarjo

Title

Artikel Rama Sudh fix tanpa daftar pustaka

Author(s)

Coordinator





perpustakaan umsidahanin

Organizational unit

Perpustakaan

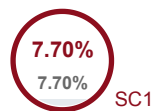
Alerts

In this section, you can find information regarding text modifications that may aim at temper with the analysis results. Invisible to the person evaluating the content of the document on a printout or in a file, they influence the phrases compared during text analysis (by causing intended misspellings) to conceal borrowings as well as to falsify values in the Similarity Report. It should be assessed whether the modifications are intentional or not.

Characters from another alphabet		0
Spreads		0
Micro spaces		7
Hidden characters		0
Paraphrases (SmartMarks)		44

Record of similarities

SCs indicate the percentage of the number of words found in other texts compared to the total number of words in the analysed document. Please note that high coefficient values do not automatically mean plagiarism. The report must be analyzed by an authorized person.

**25**

The phrase length for the SC 2

5131

Length in words

38011

Length in characters

Active lists of similarities

This list of sources below contains sources from various databases. The color of the text indicates in which source it was found. These sources and Similarity Coefficient values do not reflect direct plagiarism. It is necessary to open each source, analyze the content and correctness of the source crediting.

The 10 longest fragments

Color of the text

NO	TITLE OR SOURCE URL (DATABASE)	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
1	https://zenodo.org/records/10851075/files/8432-Article%20Text-25415-2-18-20240321%20Publish.pdf?download=1	25 0.49 %
2	https://zenodo.org/records/10851075/files/8432-Article%20Text-25415-2-18-20240321%20Publish.pdf?download=1	21 0.41 %
3	https://zenodo.org/records/10851075/files/8432-Article%20Text-25415-2-18-20240321%20Publish.pdf?download=1	20 0.39 %

4	https://zenodo.org/records/10851075/files/8432-Article%20Text-25415-2-18-20240321%20Publish.pdf?download=1	19 0.37 %
5	https://zenodo.org/records/10851075/files/8432-Article%20Text-25415-2-18-20240321%20Publish.pdf?download=1	19 0.37 %
6	https://zenodo.org/records/10851075/files/8432-Article%20Text-25415-2-18-20240321%20Publish.pdf?download=1	19 0.37 %
7	Analisis Sentimen Produk Permainan menggunakan Metode TF-IDF dan Algoritma K-Nearest Neighbor Rifki Kosasih, Anggi Alberto;	17 0.33 %
8	https://zenodo.org/records/10851075/files/8432-Article%20Text-25415-2-18-20240321%20Publish.pdf?download=1	15 0.29 %
9	https://zenodo.org/records/10851075/files/8432-Article%20Text-25415-2-18-20240321%20Publish.pdf?download=1	15 0.29 %
10	https://zenodo.org/records/10851075/files/8432-Article%20Text-25415-2-18-20240321%20Publish.pdf?download=1	14 0.27 %

from RefBooks database (1.13 %)

NO	TITLE	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
Source: Paperity		
1	Analisis Sentimen Produk Permainan menggunakan Metode TF-IDF dan Algoritma K-Nearest Neighbor Rifki Kosasih, Anggi Alberto;	27 (2) 0.53 %
2	ANALISIS PENGARUH BRAND AWARENESS, BRAND ASSOCIATION DAN BRAND AMBASSADOR TERHADAP KEPUTUSAN PEMBELIAN PADA PRODUK SKINCARE SCARLETT WHITENING Devi Apriliani, Hayuningtias Kristina Anindita;	12 (1) 0.23 %
3	PENERAPAN ALGORITMA NAIVE BAYES DAN METODE CRISP-DM DALAM PREDIKSI HASIL TES KEMAMPUAN BAHASA INGGRIS MAHASISWA Asriyanik Asriyanik,Sany Santiastry, Winda Apriandari;	8 (1) 0.16 %
4	OPTIMIZING THE COMPETENCE AND PROFESSIONALISM OF ACCOUNTING TEACHERS IN USING ACCURATE SOFTWARE: Optimalisasi Kompetensi dan Profesionalisme Guru Akuntansi Dalam Menggunakan Software Accurate Abidin Fityan Izza Noor, Anis Milasari,Rahayu Ruci Arizanda, Alshaf Pebrianggara;	6 (1) 0.12 %
5	ALUMNI JOB WAITING PERIOD PREDICTION USING NAÏVE BAYES CLASSIFIER AT COMPUTER SCIENCE STUDY PROGRAM UNIVERSITY OF NUSA CENDANA Rachmadiansyah Rachmadiansyah, Rumlaklak Nelci Dessy, Mauko Arfan Yeheskiel;	5 (1) 0.10 %

from the home database (0.00 %)

NO	TITLE	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
----	-------	---------------------------------------

from the Database Exchange Program (0.00 %)

NO	TITLE	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
----	-------	---------------------------------------

from the Internet (6.57 %)

NO	SOURCE URL	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
1	https://zenodo.org/records/10851075/files/8432-Article%20Text-25415-2-18-20240321%20Publish.pdf?download=1	275 (20) 5.36 %

2	https://timesindonesia.co.id/indonesia-positif/433000/universitas-muhammadiyah-sidoarjo-gelar-yudisium-fakultas-bisnis-hukum-dan-ilmu-sosial	20 (4) 0.39 %
3	https://nuramiril.blogspot.com/2023/12/tugas-penyusunan-apbdes-di-desa.html	13 (1) 0.25 %
4	https://jurnal.ar-raniry.ac.id/index.php/cyberspace/article/download/8805/5200	13 (1) 0.25 %
5	http://repository.teknokrat.ac.id/5906/1/skripsi19315031.pdf	11 (1) 0.21 %
6	https://archive.umsida.ac.id/index.php/archive/preprint/download/7096/50857/56614	5 (1) 0.10 %

List of accepted fragments (no accepted fragments)

NO	CONTENTS	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
----	----------	---------------------------------------

Sentiment Analysis of Consumer Satisfaction with Steak Hut Manyar Kertoarjo Restaurant Services Using the Naïve Bayes Method and Algorithm TF-idf

Analisis Sentimen Kepuasan Konsumen Terhadap Layanan Restoran Steak Hut Manyar Kertoarjo Menggunakan Metode Naive Bayes dan Algoritma TF-Idf

Rama Chikal Abimanyu
216120900007

Dosen Pembimbing
Istian Kriya Almanfaluti, S.Kom, M.Kom

Dosen Penguji 1
Bambang Bayu Prasajo, S.Si., M.Pd

Dosen Penguji 2
Alshaf Pebrianggara ,S.E., M.M.

Program Studi Bisnis Digital
Fakultas Bisnis Hukum dan Ilmu Sosial Universitas Muhammadiyah Sidoarjo

Daftar isi

DAFTAR ISI	ii
I. PENDAHULUAN	1
Rumusan Masalah	2
Pertanyaan	2
Kategori SDG's	2
II. LITERATURE REVIEW	2
Tripadvisor	2
Digital Marketing	3
Konsumen	3
Kepuasan Pelanggan	3
Kualitas Pelayanan	3
Analisis Sentimen	3
Brand Awareness	3
Metode Naive Bayes	3
Algoritma Tf-Idf	3
III. METODE	3
IV. HASIL DAN PEMBAHASAN	8
V. SIMPULAN	14
DAFTAR PUSTAKA	14

Page 2

Page 2

Page 2

Sentiment Analysis of Consumer Satisfaction with Steak Hut Manyar Kertoarjo Restaurant Services Using the TF-Idf Method

Rama Chikal Abimanyu1), Istian Kriya Almanfaluti2), Bayu Hari Prasoj3), Alshaf Pebrianggara4)

1)Program Studi Bisnis Digital, **Universitas Muhammadiyah Sidoarjo, Jawa Timur**, Indonesia

2)Program Studi Bisnis Digital, **Universitas Muhammadiyah Sidoarjo, Jawa Timur**, **Indonesia**

3)Program Studi Bisnis Digital, **Universitas Muhammadiyah Sidoarjo, Jawa Timur**, Indonesia

4)Program Studi Bisnis Digital, **Universitas Muhammadiyah Sidoarjo, Jawa Timur**, Indonesia

Email: istian.alman@umsida.ac.id

Abstract. On online shopping sites or often referred to as marketplaces, there is a column of comments and reviews of transactions that have been made by buyers for products that have been purchased. With this product assessment feature, buyers can consider decisions about the products they will buy. But at this time there is a problem with the review feature because many buyers give negative comments but give a five-star rating. This results in the feature of giving values from consumers being bad. For this reason, a sentiment analysis study was conducted on the review feature at the Steakhut Manyar restaurant using the naive Bayes method and the Tf-Idf algorithm. Based on the review of reviews at the Steakhut restaurant, 1000 review data have been collected which are divided into two, namely 700 training data and 300 test data. After that, the text preprocessing data stage is carried out, where the text preprocessing stage is collecting product and service review data on the web page (Cleaning data), changing uppercase letters to lowercase letters (Casefolding), separating sentences into single sentences (tokenizing), removing conjunctions that are not used for sentiment analysis (stopwords), changing words to basic words (stemming) and continuing to give weight to each word using the Tf-idf algorithm.

Kata Kunci - TF-Idf Sentimen Analysis Product Food

Abstrak. Pada situs belanja online atau yang sering disebut juga dengan marketplace terdapat sebuah kolom komentar dan ulasan dari transaksi yang telah dilakukan oleh pihak pembeli terhadap produk yang telah di beli. Dengan adanya fitur penilaian produk ini pembeli bisa mempertimbangkan keputusan terhadap produk yang akan dibeli. Tetapi pada saat ini muncul permasalahan dari fitur review ulasan hal ini dikarenakan banyak pembeli yang memberikan komentar negative tetapi memberikan rating bintang lima. Ini mengakibatkan fitur dari pemberian nilai dari konsumen menjadi buruk. Untuk itu dilakukan lah sebuah penelitian analisis sentimen terhadap fitur review ulasan pada restoran steakhut manyar dengan menggunakan metode naive bayes dan algoritma Tf-Idf. Berdasarkan review ulasan pada restoran steakhut telah dikumpulkan data sebanyak 1000 data ulasan yang dibedakan **menjadi dua yaitu 700 data latih dan 300 data uji**. Setelah itu dilakukan tahap text preprocessing data, yang dimana tahap text preprocessing adalah pengumpulan data ulasan produk dan layanan pada laman web (Cleaning data), perubahan huruf besar menjadi kecil (Casefolding), memisahkan kalimat menjadi kalimat tunggal (tokenizing), penghilangan kata sambung yang tidak digunakan untuk analisis sentimen (stopword), perubahan kata menjadi kata dasar (stemming) dan dilanjutkan pemberian bobot pada setiap kata menggunakan algoritma Tf-idf.
Kata Kunci - **TF-Idf** Analisis Sentimen Produk Makanan

I. PENDAHULUAN

Pada era digital yang sedang terjadi saat ini platform digital yang sering digunakan oleh masyarakat untuk membantu mencari restoran yang memiliki kualitas produk dan layanan yang terbaik adalah Tripadvisor. Produk berupa pelayanan dan makanan pada wisata kuliner memberikan dampak lebih terhadap kepuasan dan kesetiaan pelanggan. Tripadvisor adalah tempat untuk mengetahui tingkat popularitas sebuah restoran yang dilengkapi dengan review ulasan konsumen. Oleh karena itu kepuasan pelanggan terhadap pelayanan dan makanan dapat di analisis pada komentar yang terdapat pada platform digital seperti tripadvisor seperti text preprocessing. Review ulasan dan rating yang disediakan oleh Tripadvisor berguna untuk membantu pelanggan atau konsumen mengevaluasi dari segi pelayanan untuk menentukan pilihannya. Review ulasan ini pada umumnya menjadi parameter bagi para konsumen untuk menentukan pilihan. Untuk itu laman web tripadvisor memberikan fitur ulasan agar restoran bisa membuat brand awareness yang baik dan konsumen bisa mencari restoran yang lebih baik. Hampir 30 % pengguna internet selalu memberikan komentar atau ulasan terhadap produk atau layanan secara online. Komentar yang diberikan oleh konsumen terhadap fitur review ulasan tidak lain mengenai kualitas pelayanan, kualitas produk dan kualitas lingkungan. Kualitas pelayanan yang baik itu seperti ramah terhadap konsumen, lingkungan yang nyaman dan sesuai dengan kebutuhan pelanggan, hal tersebut sangat berdampak dalam kepuasan pelanggan. Pelayanan yang baik merupakan salah satu faktor penting sama dengan faktor kualitas makanan atau produk. Dimana pastinya setiap pelanggan atau konsumen ingin mendapatkan pelayanan yang terbaik. Analisis sentimen sendiri ialah gambaran polaritas dari teks atau kata[7]. Analisis sentimen yang diberikan oleh pengunjung restoran terhadap pelayanan dan makanan sangat penting untuk menghasilkan rekomendasi bagi pihak pemilik restoran untuk memberikan kualitas produk makanan dan pelayanan yang sesuai dengan keinginan pengunjung. Dari hasil pencarian tripadvisor umumnya muncul daftar restoran yang cocok dan biasanya ditampilkan berupa gambar yang disertakan nama restoran, alamat serta seluruh review keseluruhan makanan dan layanan dan sebuah hyperlink yang menuju ke halaman web yang profil dari restoran. Pada saat ini menunjukan bawah era Big data sangat memungkinkan pengembang bisnis untuk mengidentifikasi dan menganalisis preferensi konsumen terhadap pelayanan dan makanan pada restoran. Algoritma yang digunakan untuk melakukan proses analisis sentimen data ulasan sangat menentukan nilai akurasi, presisi, sensitivitas sebagai rujukan performa bagi algoritma yang digunakan. Tetapi terjadi kendala yaitu ketidaksesuaian antara pemberian review ulasan dengan rating yang diberikan terhadap produk mengakibatkan kita sebagai pembeli menjadi sulit untuk membedakan produk mana yang kualitasnya baik. Hal ini juga sangat mempengaruhi penilaian komentar dan rating terhadap restoran menjadi kurang tepat karena banyaknya ketidak sesuaian pemberian ulasan dengan penilaian rating restoran.

1.
- Fitur penilaiannya atau rating yang tersedia pada tripadvisor adalah bintang satu sampai dengan bintang lima, fitur penilaian atau rating bintang ini berguna bagi pembeli dalam menentukan restoran. Dengan terjadinya ketidaksesuaian antara review ulasan dan rating penilaian menjadi permasalahan dalam pemilihan restoran.

Untuk itu pembeli membutuhkan suatu sebuah metode untuk menganalisis sentimen agar penilaian menjadi lebih akurat dan efektif. Penelitian digunakan untuk menganalisis sentimen dari konsumen terhadap makanan, layanan dan kualitas lingkungan. Berdasarkan ulasan yang diberikan oleh konsumen terhadap kualitas makanan, kualitas layanan dan kualitas lingkungan akan diproses menggunakan metode naive bayes dan dibantu dengan algoritma Tf-Idf. Objek yang digunakan merupakan review ulasan yang diberikan oleh konsumen pada situs laman web Tripadvisor. Pada tahapan analisis sentimen terdapat 6 tahapan untuk mendapatkan data ulasan dari tripadvisor kategori aspek tersebut yang tersedia di Tripadvisor adalah

makanan, layanan, harga dan tempat. Dengan itu pengumpulan data opini dari data web untuk mengetahui perspektif konsumen terhadap restoran Steak Hut. Alasan pemilihan steak hut ialah karena restoran ini memiliki jumlah komentar yang sesuai dengan kebutuhan untuk melakukan metode naive bayes. Kesimpulannya adalah restoran dengan adanya informasi publik yang tersedia bisa menjadi bahan intropeksi guna restoran jauh lebih baik sebelumnya.

Rumusan Masalah :Ketidaksesuaian antara pemberian review ulasan dengan fitur penilaian rating yang mempersulit konsumen untuk menentukan pilihan.

Pertanyaan :

- 1.Apakah menggunakan metode naive bayes akurat dalam menentukan hasil perhitungan data review ulasan?
- 2.Apakah dengan tambahan fitur-fitur frekuensi data dan algoritma Tf-Idf dapat membantu meningkatkan kinerja dari metode naive bayes?

Kategori SDG's :Berdasarkan pendahuluan di atas rumusan masalah pada penelitian ini berfokus pada Suistanable development goals pada pilar Quality Educations yang meliputi point 4.

II.LITERATURE REVIEW

Tripadvisor

Tripadvisor adalah web yang menyajikan pengalaman perjalanan dan kunjungan terhadap restoran,hotel dan penginapan dalam bentuk ulasan online.web menyediakan banyak restoran dan hotel yang bisa menjadi rekomendasi untuk dikunjungi baik nasional maupun internasional.

Digital Marketing

Digital marketing adalah sebuah strategi yang dilakukan oleh pelaku bisnis atau umkm untuk memasarkan produknya dengan menggunakan teknologi seperti web,blog dan iklan media sosial. Digital marketing berguna untuk memperkenalkan atau memasarkan produk kepada konsumen dengan jangkauan yang lebih luas dan memanfaatkan teknologi.

Konsumen

Konsumen adalah seseorang atau kelompok orang yang melakukan sebuah transaksi pembelian terhadap barang atau jasa dari pihak penjual.konsumen juga merupakan pelaku yang melakukan pembelian pada platform marketplace.Konsumen juga menjadi target bagi para penjual yang untuk melakukan pembelian terhadap produknya.Konsumen juga berarti individu atau kelompok yang bertindak sebagai pengonsumsi produk dari penjual baik produk yang bisa dimakan atau jasa.

Kepuasan Pelanggan

Kepuasan pelanggan merupakan suatu sesuatu yang berhubungan dengan minat tertentu dan biasanya terjadi penyebaran minat yang positif dari individu ke individu yang lain.Kepuasan pelanggan akan selalu berpengaruh untuk penilaian konsumen terhadap layanan dan kualitas makanan yang disajikan oleh pihak restaurant dan biasanya menjadi rekomendasi untuk tempat kunjungan kepada individu yang lain.

Kualitas Pelayanan

Kualitas pelayanan merupakan interaksi antara satu individu sebagai pekerja pada tempat kerja dengan indikasi kemampuan layanan yang piawai dan professional.Nilai yang diberikan oleh pelanggan terhadap restaurant yang bersangkutan dengan kualitas pelayanan.Kualitas pelayanan yang baik akan membantu menciptakan brand awareness yang baik.

Analisis Sentimen

Analisis sentimen merupakan Penelitian tentang produksi teks yang terkait dengan area yang lebih luas pengolahan kegiatan data tertentu.Analisis sentimen biasanya digunakan untuk mengukur sebuah review ulasan yang ada pada fitur marketplace dan e-commerce hal ini bertujuan untuk mengetahui seberapa akurat penilaian yang diberikan oleh platform terhadap outlet.

Brand Awareness Brand Awareness adalah Kemampuan yang dimiliki oleh calon pembeli untuk mengingat dan mengenali suatu merek pada produk tertentu.Brand awareness sangat berpengaruh terhadap penjualan hal ini dikarenakan brand adalah nama yang paling dikenal oleh konsumen apabila terjadi insiden positif ataupun insiden negative akan tetapi dengan adanya brand para pelaku bisnis bisa membuat konsumen mereka selalu mengingat produk dari brand pelaku bisnis

Metode Naive Bayes

Perhitungan naive bayes merupakan perhitungan yang dilakukan dengan cara mengklasifikasi statistik yang bisa digunakan buat probabilitas keanggotaan sesuatu kelas. Metode naive bayes dilakukan setelah pembobotan teks menggunakan algoritma Tf-Idf.

Algoritma Tf-Idf

TF-IDF merupakan metode pembandingan untuk proses pembobotan kata dengan menghitung frekuensi kemunculan suatu kata dalam sebuah dokumen serta frekuensi dari kebalikkan dokumen yang membanding.

III.METODE

Ditahap metode akan dibahas tahapan penelitian yang akan dilakukan untuk analisis sentimen produk makanan yang terdapat pada laman web tripadvisor.Analisis sentimen adalah metode yang berguna untuk menentukan ulasan negative dan positive. Metode yang digunakan ialah metode Naive Bayes dan menggunakan algoritma TF-IDF dari penilaian ulasan produk Steak Hut pada e-marketplace Tripadvisor. Naive berartikan bahwa pada saat munculnya sesuatu elemen tidak terdapat kaitannya dengan fitur yang lain, karena itu tiap elemen membagikan sumbangan secara perorangan dalam pengelompokan tanpa tergantung dengan fitur yang lainnya. Alasan kenapa menggunakan metode naive bayes adalah

karena pada metode naive bayes memiliki akurasi lebih tinggi walaupun data nya sedikit..

Obyek Penelitian

Objek Penelitian dari artikel ini adalah review ulasan pada restoran steak hut melalui web tripadvisor. Penelitian ini berfokus pada analisis sentimen yang diberikan oleh konsumen kepada restoran di web tripadvisor. Penelitian ini bertujuan untuk mengevaluasi kepuasan pelanggan dan rekomendasi perbaikan bagi pihak restoran. Data yang dianalisis mencakup 1000 data ulasan yang diklasifikasikan dengan metode naive bayes dan pemberian bobot dengan menggunakan algoritma tf-idf.

Model Penelitian

Model penelitian merupakan sebuah konsep yang mendasar didapatkan melalui kenyataan, pengamatan, dan tinjauan kepustakaan. Berikut adalah langkah-langkah yang hendak dilakukan dan praktekkan. pada gambar 2.1

Gambar 2.1 Model Penelitian

Keterangan:

1. Pengumpulan data (Scrapping Data)

Scrapping Data merupakan langkah pengambilan atau pengumpulan data dari suatu base yang digunakan untuk penelitian. Scraping juga berarti pengumpulan data dan informasi dari situs web secara otomatis.[20] Pada tahap pengumpulan data yang diperoleh dari ulasan produk dan layanan pada restoran di laman web tripadvisor sebanyak 1000 data yang dimana dari 1000 data tersebut dibagi menjadi dua bagian 700 data latih dan 300 data uji. Pengumpulan data menggunakan seleksi data, seleksi data menjadi alat lebih baik dan efisien dalam menentukan jumlah data latih[21]. Contoh penggunaan seleksi data adalah untuk menghilangkan data yang tidak digunakan untuk proses klasifikasi data[22]. Setelah mendapatkan data. Data akan dilaksanakan proses pelabelan data ulasan yang bertujuan untuk di lanjutkan ketahap berikutnya hingga akhir. Langkah berikutnya ialah melakukan proses case folding pada ulasan konsumen atau data mentah menjadi sebuah data yang bisa di olah untuk langkah berikutnya.

2. Pelabelan data ulasan

Pada langkah ini data yang telah diambil dari laman web akan dilakukan proses pembedahan yang dibedakan menjadi dua yaitu label positif dan negatif. Data ulasan yang sudah diberikan pelabelan akan dilakukan sebuah proses preprocessing data dimana data akan diolah menjadi data set yang lebih efektif. Dengan pelabelan data ulasan sangat membantu Preprocessing data.

3. Preprocessing Data

Text preprocessing data yang umum digunakan yaitu pra-pemrosesan, pelipatan kasus (cleaning), tokenisasi, normalisasi, penghapusan stopwords, dan stemming. Penyeragaman text atau Preprocessing data adalah tahap penyeragaman text review ulasan dalam bentuk dan formatnya yang kemudian akan dilanjutkan pada proses berikutnya. Hal ini dilakukan dengan tujuan untuk meningkatkan kualitas data, mencegah masalah yang akan terjadi selama pemrosesan data, dan membuat kumpulan data menjadi lebih efektif dan efisien. Dalam tahapan ini terjadi perubahan ulasan konsumen atau data mentah di olah menjadi data yang bisa di olah untuk tahap berikutnya.

1. Cleaning

Cleaning adalah tahap menghilangkan semua atribut yang ada pada text review ulasan yang tidak digunakan seperti hastag, simbol, tanda baca, angka dan url. Sehingga data lebih efisien dan data dapat diolah jadi lebih baik. Proses cleaning bisa kita lihat [pada tabel 2.1](#)

[Tabel 2.1 contoh data hasil cleaning 2. Case Folding](#)

[Case Folding](#) adalah [proses](#) merubah semua huruf kapital yang ada pada text review ulasan yang menjadi huruf kecil. Perihal ini bertujuan supaya dokumen pada bacaan review ulasan mempunyai wujud yang standar. Proses [case folding](#) bisa kita lihat [pada tabel 2.2](#)

[Tabel 2.2 contoh data hasil case folding 3. Normalisasi](#)

Normalisasi adalah proses perubahan kalimat yang tidak bisa dimengerti contoh kalimat yang disingkat, kalimat yang menggunakan bahasa asing dan istilah yang tidak standar menjadi standar. Proses normalisasi bisa kita lihat pada tabel 2.4

Tabel 2.4 contoh data hasil normalisasi

4. Tokenizing

Tokenizing ialah tahap pemecahan teks jadi unit terkecil atau token. Hal ini akan memudahkan pada langkah berikutnya yaitu langkah pemrosesan dan analisis data. proses tokenizing bisa kita lihat pada tabel 2.3

Tabel 2.3 contoh data hasil tokenizing

5. Stopword Removal

Stopword Removal adalah langkah penghapusan dan penghilangan kata yang tidak memiliki makna dan tidak berguna dalam tahap preprocessing data. Proses stopwords removal bisa kita lihat pada tabel 2.5

Tabel 2.5 contoh data hasil stopwords removal

6. Stemming

Stemming adalah Proses mengembalikan sebuah kata ke format kata dasarnya atau menghapus imbuhan (baik awalan maupun akhiran) dari kata yang telah di proses. Proses stemming bisa kita lihat pada tabel 2.6

Tabel 2.6 contoh data hasil stemming

4. TF-IDF

Klasifikasi tidak bisa memproses langsung dokumen aslinya oleh karena itu membutuhkan pembobotan terlebih dahulu yang bertujuan untuk mencari bobot kata yang muncul dalam sekumpulan dokumen. Metode pembobotan kata berdasarkan kombinasi. Tf-idf merupakan metode pembandingan

untuk proses pembobotan kata dengan menghitung frekuensi kemunculan suatu kata dalam sebuah dokumen serta frekuensi dari kebalikannya dokumen yang membanding. Pada algoritma tf-idf akan dilakukan proses pembobotan kata pada text review ulasan pada restoran steakhut pada laman web tripadvisor untuk mengetahui bobot dari teks review ulasan dari setiap kata tersebut. Tf-idf sendiri memiliki rumus untuk perhitungannya berikut ini adalah rumus perhitungan tf-idf 1

Rumus

$$TF - IDF(d,t) = TF(d,t) \times IDF(t) \quad (1)$$

Tabel rumus tf-idf

Keterangan:

Tf : jumlah kata dalam sebuah dokumen
IDF: nilai balik dari dokumen yang memuat kata itu
t : kata
d : dokumen

5. Classification

Classification dilakukan dengan cara mengklasifikasi data dengan memakai sebuah perhitungan algoritma Naive Bayes. Perhitungan naive bayes merupakan perhitungan yang dilakukan dengan cara mengklasifikasi statistik yang bisa digunakan buat probabilitas keanggotaan sesuatu kelas. Metode naive bayes juga bisa berarti sebuah tahap pengklasifikasian probabilistik sederhana yang disebut sebagai metode yang memiliki kecepatan pemrosesan yang tinggi dan mengasumsikan independensi antar fitur dalam kata. Tahap klasifikasi memerlukan data ulasan **yang akan digunakan untuk proses** klasifikasi metode naive bayes.

Rumus

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2)$$

Keterangan: B= Data kelas yang dimasukkan A=Data hipotesis P(A|B)=Probabilitas hipotesis A berdasarkan kondisi B. P(A)=Probabilitas hipotesis A P(B|A)=Probabilitas hipotesis B berdasarkan kondisi A P(B)=Probabilitas hipotesis B.

6. Evaluasi

Di tahap evaluasi dilakukan menggunakan sebuah Perhitungan meliputi akurasi, presisi, recall, dan F1-score yang dilakukan dengan menggunakan Confusion Matrix pada ulasan yang telah diklasifikasikan melalui algoritma Naive Bayes. Tujuan dari analisis ini adalah untuk mengevaluasi keakuratan dokumen setelah diklasifikasikan menjadi sentimen positif dan negatif. Dalam penelitian ini, digunakan proses evaluasi dengan matriks confusion. Matriks confusion adalah alat yang bermanfaat untuk menilai performa suatu metode klasifikasi. Perhitungan confusion matrix dapat kita lihat pada tabel 3.1.

Tabel 3.1. Confusion Matrix Fakta Prediksi

Negative Positive

Negatif True Negative False Positive Positif False Negative True Positive Untuk menilai kinerja klasifikasi Naive Bayes digunakan cara yaitu mencakup perhitungan akurasi (accuracy), presisi (precision), recall, dan F1-Score dengan rumus seperti berikut ini::

1. Perhitungan Akurasi(accuracy), adalah sebuah perhitungan jumlah estimasi yang akurat pada kumpulan data yang diambil dari tabel matriks confusion. Perhitungan akurasi dapat didefinisikan menggunakan rumus (3):

Rumus

$$Akurasi = \frac{TN + TP}{TP + TN + FN + FP} \times 100\% \quad (3)$$

2. Perhitungan Presisi(precision), adalah perhitungan rasio prediksi antara jumlah yang benar positif dan yang salah. Perhitungan ini bisa dilakukan menggunakan rumus (4):

Rumus

$$Presisi = \frac{TP}{TP + FP} \times 100\% \quad (4)$$

3. Perhitungan Recall, adalah Perhitungan rasio prediksi dilakukan dengan membagi jumlah positif yang benar dengan jumlah positif yang benar ditambahkan dengan jumlah negatif yang salah. Perhitungan ini bisa dilakukan menggunakan rumus (5):

Rumus

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (5)$$

4. Perhitungan F1-Score, Perhitungan yang digunakan untuk menyeimbangkan nilai ketepatan dan pengingatan. Maka dapat dihitung menggunakan persamaan (6):

Rumus

$$F1-Score = 2 \times (\text{recall} \times \text{presisi}) / (\text{recall} + \text{presisi}) \quad (6)$$

Keterangan:

TP=True Positive
TN=True Negative
FP=False Positive
FN=False Negative

IV.HASIL DAN PEMBAHASAN

Data Set

Setelah melakukan scrapping data pada laman web tripadvisor @steakhut manyar kertoarjo menggunakan tools web scrapper.Data yang terkumpul adalah data yang terdiri dan komentar dari pelanggan yang kemudian disusun menjadi dataset,dengan total 1000 data yang telah diambil pada laman web tripadvisor beberapa komen dikumpulkan dan ditampilkan pada gambar 3.1 . setelah proses pelabelan data,jumlah data yang akan di gunakan adalah 1000 data. Hal tersebut karena penelitian ini hanya menggunakan komentar dengan nilai positive dan negative.

Gambar 3.1 dataset komentar steakhut manyar

Pelabelan Data

Penelitian ini akan dilakukan pelabelan secara manual yang dilakukan oleh pelabel. Pelabel memberikan 2 label yaitu label positive untuk sentimen positive dan label negative untuk sentimen negative. Berikut ini adalah beberapa komentar pada laman steakhut yang telah di beri label dilihat pada gambar 3.2

Komentar	Label
Makanan di restoran ini sangat lezat! S aya mencoba menu spesial mereka dan rasanya luar biasa. Pelayanan juga cepat dan ramah.	Positive
Pengalaman saya di restoran ini mengecewakan. Makanan yang saya pesan datang terlambat dan tidak sesuai dengan harapan. .	Negative
The best place to eat dinner in sby @steakhut. Buka sampe jam 12 malem jd enak buat nongkrong. Harga juga terjangkau.	Positive
Tidak enak, karena tepungnya seperti bau tengik. Tempat luas, tetapi pelayanan lama. Soal rasa, tidak istimewa	Negative
Saya merasa harga yang ditawarkan tidak sebanding dengan kualitas makanan. Saya tidak akan kembali ke restoran ini.	Negative

Tabel 3.2 Proses Pelabelan data

Preprocessing Data

Pada proses preprocessing ini text akan di klasifikasi dan di bersihkan sebelum akan dilakukan analisis dokumen. Pada penelitian ini tahap preprocessing menggunakan fitur text preprocessing yang ada pada library NLTK (Natural Language Toolkit). Berikut ini adalah tahapan-tahapan yang dilakukan yaitu:

Cleaning Data

Dalam tahap cleaning data akan dilakukan proses penghilangan atribut,hashtag atau simbol pada data ulasan contoh tersebut bisa di lihat pada gambar tabel 3.3

Username	Tanggal	Ulasan Steakhut	Sentimen	Cleaning
Diana	20/7/17	Makanan di restoran ini sangat lezat! Saya mencoba menu spesial mereka dan rasanya luar biasa. Pelayanan juga cepat dan ramah.	Positive	Makanan di restoran ini sangat lezat Saya mencoba menu spesial mereka dan rasanya luar biasa Pelayanan juga cepat dan ramah
Budi	9/1/19	Pengalaman saya di restoran ini mengecewakan. Makanan yang saya pesan datang terlambat dan tidak sesuai dengan harapan.	Negative	Pengalaman saya di restoran ini mengecewakan Makanan yang saya pesan datang terlambat dan tidak sesuai dengan harapan
Lucas	1/15/20	The best place to eat dinner in sby @steakhut. Buka sampe jam 12 malem jd enak buat nongkrong. Harga juga terjangkau..	Positive	The best place to eat dinner in sby Buka sampe jam malem jd enak buat nongkrong Harga juga terjangkau
Indah	8/2/22	Tidak enak, karena tepungnya seperti bau tengik. Tempat luas, tetapi pelayanan lama. Soal rasa, tidak istimewa	Negative	Tidak enak karena tepungnya seperti bau tengik Tempat luas tetapi pelayanan lama Soal rasa tidak istimewa
Halimah	13/3/22	Saya merasa harga yang ditawarkan tidak sebanding dengan kualitas makanan. Saya tidak akan kembali ke restoran ini.	Negative	Saya merasa harga yang ditawarkan tidak sebanding dengan kualitas makanan Saya tidak akan kembali ke restoran ini

Tabel 3.3 Proses Cleaning Data

Case Folding

Pada tahap case folding data set yang telah dilakukan cleaning data akan di lakukan proses perubahan huruf kapital mejadi huruf kecil(non-capital). Proses case folding dapat kita lihat pada gambar tabel 3.4

Username	Tanggal	Steakhut Ulasan	Cleaning	Case Folding
Diana	20/7/17	Makanan di restoran ini sangat lezat! Saya mencoba menu spesial mereka dan rasanya luar biasa. Pelayanan juga cepat dan ramah.	Makanan di restoran ini sangat lezat Saya mencoba menu spesial mereka dan rasanya luar biasa Pelayanan juga cepat dan ramah	makanan di restoran ini sangat lezat saya mencoba menu spesial mereka dan rasanya luar biasa pelayanan juga cepat dan ramah
Budi	9/1/19	Pengalaman saya di restoran ini mengecewakan. Makanan yang saya pesan datang terlambat dan tidak sesuai dengan harapan.	Pengalaman saya di restoran ini mengecewakan Makanan yang saya pesan datang terlambat dan tidak sesuai dengan harapan	pengalaman saya di restoran ini mengecewakan makanan yang saya pesan datang terlambat dan tidak sesuai dengan harapan
Lucas	1/15/20	The best place to eat dinner in sby @steakhut. Buka sampe jam 12 malem jd enak buat nongkrong. Harga juga terjangkau..	The best place to eat dinner in sby Buka sampe jam malem jd enak buat nongkrong Harga juga terjangkau	the best place to eat dinner in sby buka sampe jam malem jd enak buat nongkrong harga juga terjangkau
Indah	8/2/22	Tidak enak, karena tepungnya seperti bau tengik. Tempat luas, tetapi pelayanan lama. Soal rasa, tidak istimewa	Tidak enak karena tepungnya seperti bau tengik Tempat luas tetapi pelayanan lama Soal rasa tidak istimewa	tidak enak karena tepungnya seperti bau tengik tempat luas tetapi pelayanan lama soal rasa tidak istimewa
Halimah	13/3/22	Saya merasa harga yang ditawarkan tidak sebanding dengan kualitas makanan. Saya tidak akan kembali ke restoran ini.	Saya merasa harga yang ditawarkan tidak sebanding dengan kualitas makanan Saya tidak akan kembali ke restoran ini	saya merasa harga yang ditawarkan tidak sebanding dengan kualitas makanan saya tidak akan kembali ke restoran ini

Tabel 3.4 Hasil Proses Case Folding

Normalisasi

Dalam tahap normalisasi adalah perbuahan bahasa yang kurang dimengerti yang ada pada data set seperti singkatan, bahasa asing dan bahasa kurang baku. Hasil dari proses normalisasi dapat kita lihat melalui gambar tabel 3.5

Username Case Folding Normalisasi

Diana makanan di restoran ini sangat lezat saya mencoba menu spesial mereka dan rasanya luar biasa pelayanan juga cepat dan ramah
makanan di restoran ini sangat lezat saya mencoba menu spesial mereka dan rasanya luar biasa pelayanan juga cepat dan ramah
Budi makanan di restoran ini sangat lezat saya mencoba menu spesial mereka dan rasanya luar biasa pelayanan juga cepat dan ramah pengalaman
saya di restoran ini mengecewakan makanan yang saya pesan datang terlambat dan tidak sesuai dengan harapan
Lucas the best place to eat dinner in sby buka sampe jam malem jd enak buat nongkrong harga juga terjangkau the best place to eat dinner in sby
steakhut buka sampe jam 12 malem jd enak buat nongkrong harga juga terjangkau
Indah tidak enak karena tepungnya seperti bau tengik tempat luas tetapi pelayanan lama soal rasa tidak istimewa tidak enak karena tepungnya
seperti bau tengik tempat luas tetapi pelayanan lama soal rasa tidak istimewa
Halimah saya merasa harga yang ditawarkan tidak sebanding dengan kualitas makanan saya tidak akan kembali ke restoran ini saya merasa harga
yang ditawarkan tidak sebanding dengan kualitas makanan saya tidak akan kembali ke restoran ini

Tabel 3.5 Hasil Normalisasi

Tokenizing

Pada tahap **tokenizing tahap ini merupakan tahap pemecahan kalimat menjadi satuan kata yang menyusunnya hasil dari proses tokenizing dapat dilihat** melalui gambar tabel 3.6

Username Normalisasi Tokenisasi

Diana makanan di restoran ini sangat lezat saya mencoba menu spesial mereka dan rasanya luar biasa pelayanan juga cepat dan ramah
[makanan, di, restoran, ini, sangat, lezat, saya, mencoba, menu, spesial, mereka, dan, rasanya, luar, biasa, pelayanan, juga, cepat, dan, ramah]
Budi pengalaman saya di restoran ini mengecewakan makanan yang saya pesan datang terlambat dan tidak sesuai dengan harapan
[pengalaman, saya, di, restoran, ini, mengecewakan, makanan, yang, saya, pesan, datang, terlambat, dan, tidak, sesuai, dengan, harapan]
Lucas the best place to eat dinner in sby steakhut buka sampe jam 12 malem jd enak buat nongkrong harga juga terjangkau
[the, best, place, to, eat, dinner, in, sby, steakhut, buka, sampe, jam, 12, malem, jd, enak, buat, nonongkrong, harga, juga, terjangkau]
Indah tidak enak karena tepungnya seperti bau tengik tempat luas tetapi pelayanan lama soal rasa tidak istimewa
[tidak, enak, karena, tepungnya, seperti, bau, tengik, tempat, luas, tetapi, pelayanan, lama, soal, rasa, tidak, istimewa]
Halimah saya merasa harga yang ditawarkan tidak sebanding dengan kualitas makanan saya tidak akan kembali ke restoran ini
[saya, merasa, harga, yang, ditawarkan, tidak, sebanding, dengan, kualitas, makanan, saya, tidak, akan, kembali, ke, restoran, ini]

Tabel 3.6 Hasil Proses Tokenizing

Stopword Removal

Pada tahap stopwords removal dataset yang akan dilakukan penghapusan **kata kata yang terkandung pada daftar stopwords. Yang dalam hal ini peneliti akan menggunakan stopwords yang ada pada NLTK (Natural Language Toolkit) dengan dataset bahasa indonesia. Proses stopwords removal bisa kita lihat pada gambar tabel 3.7**

Username Tokenisasi Stopword Removal

Diana [makanan, di, restoran, ini, sangat, lezat, saya, mencoba, menu, spesial, mereka, dan, rasanya, luar, biasa, pelayanan, juga, cepat, dan, ramah]
[makanan, restoran, sangat, lezat, mencoba, menu, spesial, rasanya, luar, biasa, pelayanan, cepat, dan, ramah]
Budi [pengalaman, saya, di, restoran, ini, mengecewakan, makanan, yang, saya, pesan, datang, terlambat, dan, tidak, sesuai, dengan, harapan]
[pengalaman, di, restoran, mengecewakan, makanan, saya, pesan, datang, terlambat, tidak, sesuai, harapan]
Lucas [the, best, place, to, eat, dinner, in, sby, steakhut, buka, sampe, jam, 12, malem, jd, enak, buat, nonongkrong, harga, juga, terjangkau]
[the, best, place, to, eat, dinner, in, sby, steakhut, buka, sampe, jam, 12, malem, jd, enak, buat, nongkrong, harga, terjangkau]
Indah [tidak, enak, karena, tepungnya, seperti, bau, tengik, tempat, luas, tetapi, pelayanan, lama, soal, rasa, tidak, istimewa]
[enak, tepungnya, bau, tengik, tempat, luas, pelayanan, lama, soal, rasa, istimewa]
Halimah [saya, merasa, harga, yang, ditawarkan, tidak, sebanding, dengan, kualitas, makanan, saya, tidak, akan, kembali, ke, restoran, ini]
[merasa, harga, ditawarkan, sebanding, kualitas, makanan, tidak, kembali, restoran]

Tabel 3.7 Hasil Stopword Removal

Stemming

Pada tahap stemming pengembalian suatu kata menjadi ke bentuk akar katanya atau penghilangan kata imbuhan pada awal atau akhiran kata. Proses stemming bisa kita lihat melalui gambar tabel 3.8

Username Stopword Removal Stemming

Diana [makanan, restoran, sangat, lezat, mencoba, menu, spesial, rasanya, luar, biasa, pelayanan, cepat, dan, ramah] makan restoran sangat lezat coba
menu spesial rasa luar biasa layanan cepat dan ramah
Budi [pengalaman, di, restoran, mengecewakan, makanan, saya, pesan, datang, terlambat, tidak, sesuai, harapan] alam di restoran kecewa makan saya
pesan datang lambat tidak sesuai harap

Lucas [the,best,place,to,eat,dinner,in,sby,steakhut,buka,sampe,jam,12,malem,jd,enak,buat,nongkrong,harga,terjangkau] the best place to eat dinner in sby steakhut buka sampe jam 12 malem jd enak buat nongkrong harga terjangkau
 Indah [enak,tepungnya,bau,tengik,tempat,luas,pelayanan,lama,soal,rasa,istimewa] enak tepung bau tengik tempat luas layan lama soal rasa istimewa
 Halimah [merasa,harga,ditawarkan,sebanding,kualitas,makanan,tidak,kembali,restoran] rasa harga tawar banding kualitas makan tidak kembali restoran

Tabel 3.8 Hasil Prosesl Stemming

Pembobotan Tf-idf

Dataset yang telah dilakukan proses preprocessing akan dilanjutkan dengan pemberian bobot pada dataset agar bisa dilanjutkan pada metode naive bayes hasil dari perhitungan pembobotan tf-idf bisa kita lihat pada gambar tabel 3.9

	Document	Term	TF	DF	IDF	TF-IDF
0	0	makan	1	551	0.620713	0.620713
1	0	restoran	1	80	2.550421	2.550421
2	0	lezat	1	40	3.243568	3.243568
3	0	coba	1	57	2.889397	2.889397
4	0	menu	1	147	1.942015	1.942015

Gambar 3.9 Tabel Hasil Pembobotan Tf-Idf

Evaluasi

Evaluasi performa model menggunakan multinomial Naive Bayes. setelah dataset melalui tahapan preprocessing dan pembobotan tf-idf. Pengklasifikasian metode naive bayes akan di dapatkan **nilai akurasi,presisi,recall dan f1-score dengan memanfaatkan confusion matrix seperti pada gambar 4.0**

Accuracy : 73.66% Precision : 81.15% Recall : 73.66% F1-Score : 67.37%

Classification Report :

	Precision	Recall	F1-Score	Support
Negative	0.72	1.00	0.83	136
Postitive	1.00	0.22	0.36	69

Accuracy	0.74	205		
Macro avg	0.86	0.61	0.60	205
Weighted avg	0.81	0.74	0.67	205

Tabel 4.0 hasil pengujian evaluasi performa

Testing

Pada proses testing ialah memasukkan ulasan baru secara random yang kemudian akan langsung dilabeli dengan label sentimen negative atau sentimen positive. Hal ini bisa di lihat pada gambar di bawah ini.

Gambar 4.1 Hasil dari testing memasukkan ulasan baru

V.SIMPULAN

Dari hasil yang telah di uraikan, maka dapat disimpulkan bahwa metode naive bayes memiliki akurasi yang akurat dalam menentukan keakurasian dari penilaian rating terhadap ulasan yang di berikan oleh konsumen terhadap suatu perusahaan.dan dapat membantu konsumen untuk menentukan penilaian terhadap perusahaan. Dengan menggunakan metode naive bayes memiliki keakuratan 73,66% hingga 81,15%. Metode naive bayes juga mampu memberikan nilai dan hasil yang akurat dalam proses pelatihan dan proses pengujian. Namun pada proses penilaian confusion matrix true negative dan false negative mendapatkan hasil yang kurang optimal karena tidak mendapatkan nilai yang sempurna.

Penelitian ini berfokus pada analisis sentimen menggunakan metode naive bayes, metode yang lebih canggih untuk menentukan keakurasian ulasan yang diberikan oleh konsumen untuk menilai kinerja suatu perusahaan. Berdasarkan penelitian ini diharapkan pada penelitian mendatang dapat mendapatkan hasil yang jauh lebih sempurna untuk menentukan sebuah akurasian dari analisis sentimen pada sebuah perusahaan. Serta dapat lebih membantu konsumen untuk menentukan kinerja dari sebuah perusahaan.

