

Sentiment Analysis of Mie Gacoan Juanda Bekasi Reviews on Google Reviews Using Naïve Bayes and C4.5 Methods

Nur Fadila¹, Wowon Priatna², Siti Setiawati³, Eka Nur A'ini⁴, Dhea Putri Aprilyana⁵

Universitas Bhayangkara Jakarta Raya¹²³⁴⁵

Email: nur.fadila19@mhs.ubharajaya.ac.id¹, wowon.priatna@dsn.ubharajaya.ac.id², siti.setiawati@dsn.ubharajaya.ac.id³,
eka.nur.a'ini19@mhs.ubharajaya.ac.id⁴, dhea.putri.aprilyana19@mhs.ubharajaya.ac.id⁵

ABSTRAKSI

Perkembangan bisnis kuliner yang semakin pesat mengakibatkan banyak pebisnis kuliner membuka restoran dengan menu-menu menarik. Salah satu restoran yang banyak diminati masyarakat adalah Mie Gacoan. Mie Gacoan sudah membuka banyak cabang, salah satunya berlokasi di Bekasi, khususnya di Juanda. Platform yang banyak digunakan para pebisnis sebagai sarana menerima feedback adalah Google Review. Google Review adalah salah satunya platform terbaik untuk mengumpulkan ulasan dari pelanggan mengenai produk atau layanan. Ulasan yang ada di Google Review menjadi referensi bagi pengunjung dan juga menjadi bahan evaluasi bagi pemilik usaha kuliner untuk melakukan perbaikan guna meningkatkan kenyamanan dan kepuasan pengunjung, baik dari segi rasa maupun fasilitas. Dalam penelitian ini, peneliti menggunakan metode Naïve Bayes dan C4.5 untuk menyelesaikan permasalahan tersebut dengan memfokuskan pada kinerja metode Naïve Bayes dalam mengklasifikasikan analisis sentimen. Data dikumpulkan dengan mengambil ulasan dari Google Ulasan menggunakan alat Pengikis Data Instan di platform Google Chrome. Sebanyak 1.291 ulasan diambil, yang mencakup ulasan positif dan negatif. Hasil pengujian menunjukkan bahwa sistem berhasil mengklasifikasikan sentimen menggunakan metode Naïve Bayes dan C4.5, dengan tingkat akurasi tertinggi sebesar 83% pada klasifikasi Naïve Bayes, sedangkan akurasi klasifikasi C4.5 sebesar 74%. Dengan hasil algoritma naïve Bayes 83% dan C4.5 74%, algoritma mampu mengklasifikasikan data dengan baik.

Kata Kunci: Analisis Sentimen, Mie Gacoan, Google Review, Naïve Bayes, C4.5.

ABSTRACT

The development of the culinary business is growing rapidly which has resulted in many culinary businessmen opening restaurants with attractive menus. One of the restaurants that is in great demand by the public is Mie Gacoan. Mie Gacoan has opened many branches, one of which is located in Bekasi, especially on Juanda. The platform that is widely used by business people as a means of receiving feedback is Google Review. Google Review is one of them. the best platform to collect reviews from customers regarding products or services. The reviews on Google Reviews serve as a reference for visitors and also serve as evaluation material for culinary business owners to make improvements to increase visitor comfort and satisfaction, both in terms of taste and facilities. In this study, researchers used the Naïve Bayes and C4.5 methods to solve these problems by focusing on the performance of the Naïve Bayes method in classifying sentiment analysis. Data is collected by fetching reviews from Google Reviews using the Instant Data Scraper tool on the Google Chrome platform. A total of 1,291 reviews were retrieved, which includes both positive and negative reviews. The test results show that the system succeeded in classifying sentiment using the Naïve Bayes and C4.5 methods, with the highest accuracy rate of 83% in the Naïve Bayes classification, while the accuracy of C4.5 classification is 74%. With the results of the naïve Bayes algorithm 83% and C4.5 74%, the algorithm is able to classify data properly.

Keywords: Sentiment Analysis, Mie Gacoan, Google Review, Naïve Bayes, C4.5.

Fadila dkk., Sentiment Analysis of Mie Gacaon Juanda Bekasi Reviews on Google Reviews Using the Naïve Bayes and C4.5 Methods



Tanggal Submit : 18/11/2023
Tanggal Diterima : 22/11/2023
Tanggal Terbit : 15/12/2023

Copyright: © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0) International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Publisher's Note: JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

I. PENDAHULUAN

Perkembangan teknologi yang pesat membawa perubahan dan kemudahan bagi khalayak dalam menjalankan aktivitas, salah satunya dalam aspek kuliner. Di Indonesia perkembangan dunia bisnis kuliner mengalami perkembangan yang sangat pesat, sehingga banyak pembisnis membuka restoran dengan berbagai menu menarik, salah satu restoran yang sedang banyak diminati pelanggan yaitu Mie Gacoan. Salah satu *platform* terbaik untuk menampung ulasan dari pelanggan mengenai suatu produk atau layanan yaitu *Google Review*. Ulasan yang terdapat pada *Google Review* juga menjadi bahan referensi pengunjung, ulasan pada situs *Google Review* juga menjadi bahan evaluasi dari pihak pengusaha kuliner agar dapat melakukan perbaikan demi kenyamanan dan kepuasan para pengunjung baik dari segi rasa dan fasilitasnya. Karena banyaknya ulasan atau *review* dari para konsumen, akan sangat membutuhkan waktu lama untuk membaca satu persatu ulasan atau *review* tersebut, maka dari itu diperlukan suatu teknik untuk mengumpulkan ulasan atau *review* dan menganalisis dengan jumlah data berskala besar. Salah satu metode yang dapat dilakukan untuk menganalisis hasil ulasan-ulasan tersebut yaitu analisis sentimen.

Analisis Sentimen merupakan sebuah metode yang digunakan untuk memahami, mengekstrak data opini, dan mengolah data tekstual secara otomatis untuk mendapatkan sebuah sentimen yang terkandung dalam sebuah opini [1]. Analisis sentimen ini juga digunakan untuk menganalisis klasifikasi jenis opini. dokumen, ulasan, atau pendapat berdasarkan emosi yang diungkapkan baik itu positif maupun negatif sehubungan dengan topik tertentu [2]. Klasifikasi sentimen bertujuan untuk mengatasi masalah ini dengan secara otomatis mengelompokkan *review* pengguna menjadi sentimen positif atau sentimen negatif [3].

Algoritma *Naive Bayes* dan C4.5 adalah dua metode yang populer digunakan dalam analisis sentimen karena kemampuan mereka untuk mengklasifikasikan ulasan ke dalam kategori sentimen yang relevan, seperti positif dan negatif[4]. Dalam penelitian ini peneliti menggunakan metode *Naive Bayes* dan C4.5 untuk menyelesaikan permasalahan tersebut dan peneliti berfokus pada seberapa baik performa metode *Naive Bayes* dalam pengklasifikasian analisis sentimen.

Maka tujuan penelitian ini adalah untuk mengetahui bagaimana tingkat akurasi metode *Naive Bayes* dan C4.5 dalam pengklasifikasian sentimen negatif maupun sentimen positif pada data ulasan atau *review* Mie Gacoan Juanda Bekasi.

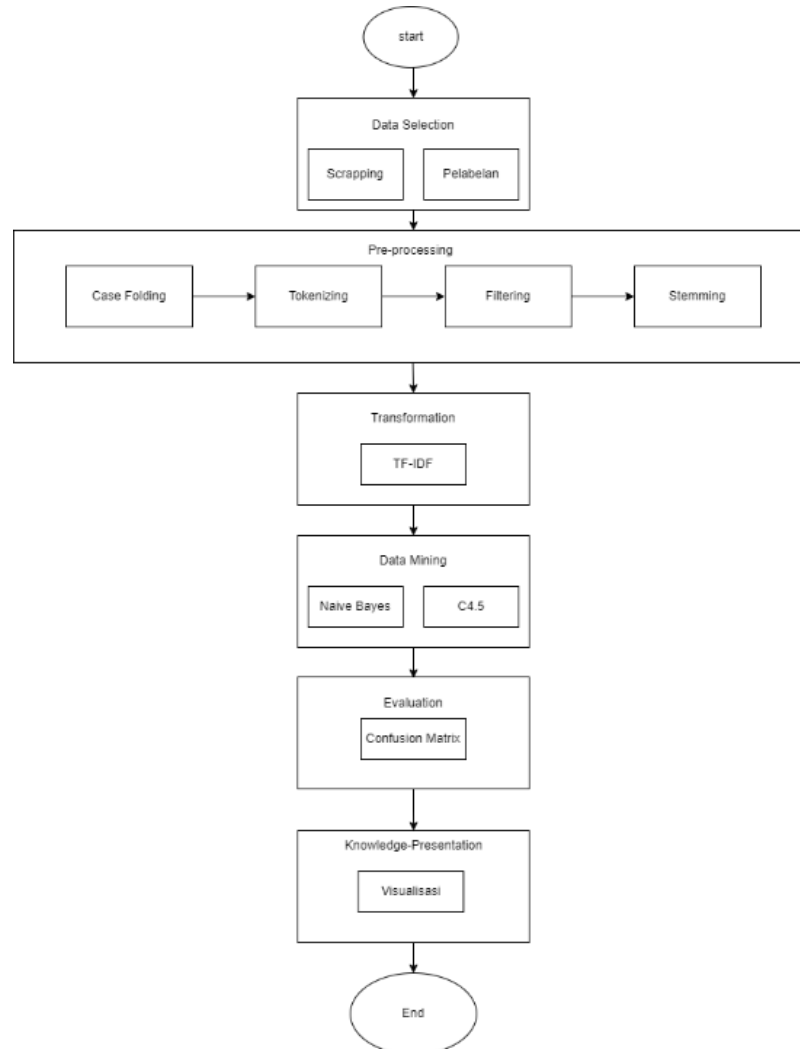
II. PENELITIAN YANG TERKAIT

Berikut penelitian terdahulu yang memiliki keterkaitan dengan penelitian ini:

1. Penelitian tentang analisis sentimen pada ulasan produk menggunakan algoritma *Naive Bayes* dilakukan oleh [5] dengan menggunakan *naive bayes* Dari hasil penelitian akhir yang diuji menghasilkan nilai akurasi terendah pada pengujian 5 kelas menggunakan dataset 80% latih dan 20% data uji sebesar 52.66%, sedangkan pada pengujian 3 kelas menggunakan dataset 90% data latih dan 10% data uji memiliki akurasi tertinggi sebesar 77.78%.
2. Penelitian lain yang pernah dilakukan oleh [6]. Dengan judul analisis sentimen ulasan media massa menggunakan metode C4.5 berbasis Forward Selection. Pengambilan data pada penelitian ini diambil pada situs review Yelp dengan masa waktu dari tahun 2010 sampai dengan 2018. Pada penelitian ini, proses analisis data menggunakan RapidMiner 6.5. Berdasarkan hasil analisis nilai akurasi dari C4.5 sebesar 80,00% dan hasil analisis forward selection sebesar 4,00%.
3. Selanjutnya penelitian yang dilakukan oleh [7] dengan judul Klasifikasi *Text Mining Review* Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma *Naive Bayes*. Nilai akurasi yang dihasilkan akan menjadi tolak ukur untuk mencari model pengujian terbaik untuk kasus klasifikasi sentimen. Evaluasi dilakukan menggunakan 10 *fold cross validation*. Pengukuran akurasi diukur dengan confusion matrix dan kurva ROC. Hasil penelitian menunjukkan penerapan generate *n-gram* pada tahap preprocessing mempengaruhi nilai akurasi dan nilai AUC yang dihasilkan. Nilai akurasi terbaik yang dihasilkan pada penelitian ini yaitu 90.50% dengan nilai AUC sebesar 0.715 pada penerapan generate *n-gram* = 2.

III. METODE PENELITIAN

Penelitian ini menggunakan metode KDD (*Knowledge Discovery In Databases*), dimana KDD didefinisikan sebagai ekstraksi informasi potensial, implisit dan tidak dikenal dari sekumpulan data [8]. Menurut [9] Seperti yang mungkin disadari, proses KDD adalah persimpangan teknik untuk mengolah dan menemukan pola dalam data. *Knowledge Discovery in Database (KDD)* akan diterapkan sebagai metodologi pada penelitian ini



Gambar. 1 Metodologi Penelitian

Berikut ini merupakan uraian alur penelitian yang menerapkan metode KDD:

1. *Data Selection*

Data selection yaitu proses seleksi atau pemilihan data yang dianggap relevan terhadap analisis [10]. *Data selection* merupakan tahapan pertama dalam KDD. Pada tahap ini dimulai dengan melakukan pengambilan data berupa ulasan atau *review* pada *platform google review* menggunakan *instant data scraper* untuk pengambilan datanya dan pemberian label manual pada setiap datanya dengan kategori sentimen positif dan negatif.

2. *Pre-Processing*

Data *pre-processing* adalah teknik *data mining* yang melibatkan transformasi data mentah menjadi format yang mudah dimengerti. Langkah data *pre-processing* diperlukan untuk menyelesaikan beberapa jenis masalah termasuk *noisy data*, data redundansi, nilai data yang hilang, dan lain-lain [5]. *Pre-Processing* merupakan tahapan ke dua dalam KDD. Di tahap ini dilakukan proses penyesuaian data dengan kriteria data yang dibutuhkan untuk penelitian, di tahap ini juga proses *cleaning* atau pembersihan data. Tahapan ini terdiri dari *tokenizing* yaitu pembersihan karakter spasi dan tanda baca, *case folding* yaitu perubahan huruf menjadi huruf kecil, *stemming* yaitu menyeragamkan kata, *filtering* yaitu penghapusan *term* yang tidak relevan atau tidak memiliki arti dan *labelling* yaitu proses pemberian label.

3. Transformation

Transformation merupakan tahapan ke tiga dalam KDD. Pada tahapan ini akan dilakukan pembobotan kata menggunakan TF-IDF. Untuk persamaan rumus TF-IDF dapat dilihat pada rumus sebagai berikut:

$$IDF = \log\left(\frac{N}{df}\right) \quad (1)$$

$$W = TF * IDF \quad (2)$$

Keterangan:

W = Pembobotan tf-idf

N = Jumlah Seluruh Dokumen

TF = Banyaknya kemunculan *term* (kata) dalam dokumen

IDF = *Inversed Document Frequency*

df = Banyak kata yang mengandung kata-i

4. Data Mining

Data Mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. Konsep *data mining* semakin dikenal sebagai *tools* penting dalam manajemen informasi karena jumlah informasi yang semakin besar jumlahnya [1].

Pada tahapan ini algoritma *Naïve Bayes* dan C4.5 diimplementasikan untuk mengklasifikasi data. *Naïve Bayes* dan C4.5 merupakan metode untuk melakukan klasifikasi. Ulasan atau *review* yang telah melalui *pre-processing* akan diklasifikasi sesuai kelasnya (sentimen positif dan negatif).

1) Algoritma Naïve Bayes

Naïve bayes yaitu salah satu pengklasifikasi statistik dimana pengklasifikasinya dapat melakukan prediksi probabilitas keanggotaan kelas dari suatu data *tuple* yang akan masuk ke dalam kelas tertentu sesuai dengan perhitungan probabilitas [11].

Rumus umum *Naïve Bayes* sebagai berikut:

$$P(H | X) = \frac{P(H | X) P(H)}{P(X)} \quad (3)$$

Keterangan:

X = Data dengan kelas yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

P(H|X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

P(H) = Probabilitas hipotesis H (prior prob.)

P(X|H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

Untuk P(Vj) dan P(xi|Vj) dihitung pada saat pelatihan di mana persamaannya adalah sebagai berikut:

$$P(V_j) = \frac{|docs\ j|}{|contoh|} \quad (4)$$

$$P(xi|V_j) = \frac{nk+1}{n+|kosakata|} \quad (5)$$

Keterangan:

|dosc j| = jumlah dokumen setiap kategori j

|contoh| = jumlah dokumen dari semua kategori

nk = jumlah frekuensi kemunculan setiap kata

n = jumlah frekuensi kemunculan kata dari setiap kategori

|kosakata| = jumlah semua kata dari semua kategori

P(xi|Vj) = Probabilitas xi pada kategori Vj

P(Vj) = Probabilitas dari Vj

2) Algoritma C4.5

Algoritma C4.5 adalah algoritma menyediakan sekumpulan data berlabel dan menghasilkan pohon keputusan sebagai output [12]. Terdapat rumus dalam membentuk pohon keputusan C4.5.

$$entropy(T) = \sum_{i=0}^n - P_i * \text{Log}_2 P_i \quad (6)$$

Keterangan:

S = Himpunan Kelas

n = Jumlah Partisi S

P_i = Proporsi dari s_i terhadap S

$$Gain(S, A) = Entropy(S) - \sum_{i=0}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (7)$$

Keterangan:

S = Himpunan Kasus

A = Atribut

n = Jumlah Partisi Atribut A

$|S_i|$ = Jumlah Kasus pada Partisi ke-i

$|S|$ = Jumlah Kasus dalam S

5. Evaluation

Pada tahapan ini akan dilakukan pengujian menggunakan *Confusion Matrix*. *Confusion matrix* dilakukan untuk melihat kinerja dari *Classifier/Supervised Learning*. Dalam pengujian evaluasi ini, hasilnya akan berupa nilai *accuracy*, *precision*, *recall* dan *f-measure*.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (8)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (9)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (10)$$

$$F - Measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (11)$$

Keterangan:

TP (*True Positive*) = jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1

TN (*True Negative*) = jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0

FP (*False Positive*) = jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1

FN (*False Negative*) = jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0

6. Knowledge-Presentation

Knowledge Presentation merupakan tahap terakhir dalam KDD, dimana pada tahap ini data-data yang sudah di proses akan divisualisasikan agar mudah dipahami oleh pengguna dan hasil dari analisis sentimen.

IV. HASIL DAN PEMBAHASAN

1. Data Selection

Pengumpulan data dilakukan dengan mengambil data ulasan pada *Google Review* dan menggunakan bantuan platform *Google Chrome*, yaitu *Instant Data Scraper*. Platform *Instant Data Scraper* merupakan salah satu platform *Google Chrome* yang berfungsi sebagai *web scraper* untuk mengambil data atau konten yang terdapat pada situs online secara gratis. *Instant Data Scraper* ini mengekstraksi data dari halaman web HTML dan mengimpornya ke dalam *spreadsheet Microsoft Excel*.

	A
1	text
2	Review bagus / jeleknya mungkin tergantung dari jam berapa customer dateng. kalo mau udah pasti sep
4	First time makan disini itu pas weekday, ga terlalu antri dan makanan sampai tidak terlalu lama beda cer
5	Setiap lewAt sini, selalu penasarannya se-enak apaa sih produknyaa, sampai orang rela antri berjam2 d
6	Tolong kepada siapapun yang mau makan disini, kalian harus isi perut dulu sebelum ke tempat ini. Karn
7	Setelah sekian lama lihat postingan viralnya mi gacoan, akhirnya nyoba jugaFirst time ke Gacoan cabang
8	Mienya enak,, untuk tingkat kepedasan pas. Pesen yang iblis level 4, masih normal pedesnya. - pangsit g
9	Kesini jam 3 pagi karna kepo dan kita pikir bakal sepi, ternyata masih rame dong jam segitu ge ?? ...
10	Situasi tak terkendali pas weekend, karena antriannya panjang banget. Petugasnya juga kaya kewalahan
11	Pertama kali ketempat ini bareng sama bestie, dan pas lagi hujan, suasananya tidak terlalu ramai, pelaya
12	Pernah kesini sekitar jam 7 malem dan ngantri parah sampe keluar-keluar dan berujung nyerah. ...
13	Mie Gacoan yang ngehits di bekasi untuk anak muda, mie setan, mie iblis yang rasanya enak dan pedas, l
14	Dateng kesini nunggu sampe gak viral. Soalnya kalo pas viral itu gak antri . Jadinya sat set sat set. Hehe ...
15	jujur mie nya gk terlalu wah bgt tp karna pedesnya jd lumayan lah ya , kebetulan cobain mie iblis level 6
16	Udah sering kesini, tapi baru kali ini dapet bad experience ?? MAAF ULASAN AGAK PANJANG.Sy sama te
17	Pertama x ke mi gacoan, menghilangkan rasa penasaran karena selalu rame, gw bersama istri makan disi
18	Makannya enak, aku suka yg mie iblis karena ada rasa manisnya, level 3 sedeng sih, berasa pedesnya tap
19	Kesini waktu hari kerja jam 9 pagi dan jam 11 siang.Antri nya sama aja ya lumayan ramai. Waktu itu antri
20	Tempatnya Luas banget ternyata untuk sebuah restoran yang basic menunya adalah mie yaa... enak dan
21	datang hari rabu jam 16:30 anpengambilan nomor antriannya cari yang pakao jaket hijau pegang no.antri
22	Sudah dua kali makan disini. Pertama lewat ojek online dan satunya dine in. Kesan ketika pertama agak
23	Sangat worth it. Antrinya buat ane cukup cepet ya krn menunya tdk banyak. Jd proses pesan menu dan b
24	Kesini waktu awal mulai bisa Dine in. Dan antri bgtt nunggu masuk itu sekitar 40 menit, langsung pesen i

Gambar. 2 Hasil Scrapping Data

Untuk *mengidentifikasi* dan mengklasifikasi sentimen yang terkandung dalam teks atau data yang dianalisis peneliti melakukan dengan pelabelan manual dalam analisis sentimen. Pelabelan manual yang dilakukan menggunakan *Microsoft Excel Office*, dapat dilihat pada gambar 3 berikut di bawah ini.

ulasan,sentimen
"Tempatnya enak banget, nongki sore2 beuuh. Harga juga terjangkau, siap2 antri aja, perjuangan bnget","positif"
"Lokasi sangat strategis berada di jl juanda Bekasi, lokasi jg termasuk daerah macet, saran agar bisa nyaman berkunjung ke restoran ini bisa pergi pada waktu pagi... Jika sudah siang pasti sangat ramai karena sdh masuk waktu makan siang dan jg ada serbuan dari kalangan anak muda selain dari para pekerja....","negatif"
"Pelayanannya sudah baik antrian juga tidak terlalu lama, mienya enak, sayang untuk hidangan di mangkoknya masih cemerot bekas pegangan pelayanan walaupun pelayanannya pake sarung tangan","positif"
"Nyobain mie goreng yang super pedas, harga terjangkau, rasa juga mantap, tempat sering rame, beberapa kali kesini sering lihat banyak yang ngantri, akhirnya mutusin untuk balik lagi, nunggu momen gak terlalu rame, tgl 10 kemarin baru kesampaian pesan walaupun take away... Yang intinya banyak bersabar karena antriannya","positif"
"Kalo menurut saya. Kurang profesional. Dan tempat masuk parkir seharusnya jangan pas lampu merah bikin bahaya dan bingung buat nyebrangnya. Karena baru kali ya. Ramenya luar biasa.,"negatif"
"tempatny termasuk luas, banyak meja & kursi buat dine-in, mushola jg ada. pelayanannya termasuk oke. walaupun antri tp masih terbilang yg biasa aja ko, ga sampai yg berjam2. untuk rasa mienya 7/10 tapi side dish nya enak2 semuuaa! worth it untuk harganya yg terjangkau.,"positif"
"Dateng jam 09 pagi belum antri. Tempat luas dan makanan dan gorengannya menurut saya enak. Harga murah meriah. Next pasti datang lagi","positif"
"Tempat nya enak luas.. makanannya sangat amat enak harga murah ... Tapi saran buat aku sih untuk ambil no antrian nya itu jangan di parkir lebih enak di depan pintu biar orang gak bingung .. dan no antrian Dine in sama take way di bedain biar gak nyampur begtu ribet","positif"
"Makannya enak harganya murah kalo bandingin harga dan rasa itu worth it banget tapi antrinya yg bikin males kalo beli mending lewat gofood aja
"Hari ini sebetulnya tdk d rencanakan untuk pergi kesini, karena pengalaman beberapa hari lalu sempat datang tp tdk kebagian no antrian karena sdh

Gambar. 3 Pelabelan Manual

2. Pre processing

Tahapan ini terdiri dari beberapa proses karena data ulasan tidak sepenuhnya menggunakan bahasa atau kata baku. Penerapan tahap *pre-processing* data pada penelitian ini dilakukan dengan melakukan 4 proses secara berurut, diantaranya sebagai berikut:

1) Case Folding

Pada proses *case folding* bertujuan untuk menyeragamkan karakter data. Proses *case folding* yaitu proses mengubah seluruh huruf menjadi huruf kecil dan menghilangkan tanda baca, emoji, dan lain-lain.

	ulasan	Sentimen
0	review bagus jeleknya mungkin tergantung dari...	1
1	tengah malem nyari cemilan pas muter muter lew...	1
2	pertama makan disini itu pas weekday ga terlal...	0
3	setiap lewat sini selalu penasarannya seenak apa...	1
4	tolong kepada siapapun yang mau makan disini k...	0
5	setelah sekian lama lihat postingan viralnya m...	1
6	mienya enak untuk tingkat kepedasan pas pesen ...	1
7	kesini jam pagi karna kepo dan kita pikir bak...	1
8	situasi tak terkendali pas weekend karena antr...	1
9	pertama kali ketempat ini bareng sama bestie d...	1

Gambar. 4 Hasil Case Folding

2) Tokenizing

Tahap *tokenizing* merupakan metode untuk melakukan pemisahan kata dalam suatu kalimat dengan tujuan untuk proses analisis teks lebih lanjut.

	ulasan	Sentimen
0	[review, bagus, , jeleknya, mungkin, tergantun...	1
1	[tengah, malem, nyari, cemilan, pas, muter, mu...	1
2	[pertama, makan, disini, itu, pas, weekday, ga...	0
3	[setiap, lewat, sini, selalu, penasarannya, seen...	1
4	[tolong, kepada, siapapun, yang, mau, makan, d...	0
5	[setelah, sekian, lama, lihat, postingan, vira...	1
6	[mienya, enak, untuk, tingkat, kepedasan, pas,...	1
7	[kesini, jam, , pagi, karna, kepo, dan, kita, ...	1
8	[situasi, tak, terkendali, pas, weekend, karen...	1
9	[pertama, kali, ketempat, ini, bareng, sama, b...	1

Gambar. 5 Hasil Tokenizing

3) Filtering

Tahap selanjutnya yaitu tahap *filtering*. Pada tahap ini, tujuan utama adalah untuk menghilangkan atau memfilter unsur-unsur yang tidak relevan atau tidak diperlukan agar fokus analisis lebih terarah.

index	
0	review, bagus, jeleknya, tergantung, jam, customer, dateng, kalo, udah, sepi, enak, makan, saran, dateng, jam, weekend, kondusif, keadaannya, antri, orang,
1	malem, nyari, cemilan, pas, muter, muter, daerah, proyek, bekasi, liat, yg, rame, rame, mie, gacoan, buka, rame, bgt
2	makan, pas, weekday, ga, antri, makanan, beda, cerita, take, away, kali, kesini, take, away, makanan, dibilang, beda, makan, ditempat, maybe, bcs, byk, jg, Shope
3	penasaraan, seenak, apaa, sih, produknnya, orang, relaa, ngantri, berjam, store, jarang, sepi, akhirnya, cobain, yg, brgkt, jam, siang, belain, gojek, krn, parkir
4	jadi, yg, berbeda, disarankan, makan, nyaman, tanpaa, antri, yg, berlebih, kerja, dipagi, hari naah, september, mie, gacoan, buka, jam, review, menu, mie, iblis
5	tolong, makan, isi, perut, karna, antrinya, aaaa, padet, pool, aku, kesini, senin, siang, ntah, aja, rame, yg, orang, kerja, sekolah, untuk, mie, nya, jujur, gak, enak, c
6	sekian, lihat, postingan, viralnya, mi, gacoan, nyoba, juga, first, time, gacoan, cabang, bekasi, pas, daftar, dipanggil, nomor, meja, jam, menit, weekend, rame, pe
7	, es, teh, manis, temenku, gak, manis, diatasnya, diaduk, wkwk, jd, manis, nya, dibawah, poin, temenku, es, lemon, tea, es, pocong, teh, poin, untuk, harganya
8	mienya, enak, tingkat, kepedasan, pas, pesen, iblis, level, normal, pedesnya, pangsit, gorengnyanya, jaura, banget, enak, murah, siamay, enak, udang, rar
9	kesini, jam, pagi, karna, kepo, pikir, sepi, rame, jam, segitu, ge, pelayanan, ramah, staff, masker, tau, senyum, ramah, staff, nya, gak, yg, ketus, wkwkwk, kebersil
10	tadi, pesen, mie, iblis, level, kena, charge, k, kalo, yg, level, harganya, k, sumpah, murah, banget, bener, menghibur, perut, dompet, minumannya, murah
11	situasi, terkendali, pas, weekend, antriannya, banget, petugasnya, kaya, kewalahan, nentuin, antriannya, parkir, motor, luas, banget, gak, cocok, acara, kelu
12	disclaimer, review, sejujurjujurnya, berdasarkan, pengalaman, alami, ditempat, diwaktu, hasil, review, berbeda, yg, alami, ditempat, sudut, pandang, subjektif,
13	kali, ditempat, bareng, bestie, pas, hujan, suasananya, ramai, pelayanan, pegawainya, oke, antrian, ditata, teratur, menunya, lumayan, harganya, murah, mur

Gambar. 6 Hasil Filtering

4) Stemming

Tahap terakhir dalam *text pre-processing* ialah tahap *stemming*. Tahap *stemming* dalam analisis sentimen adalah proses untuk mengubah kata-kata dalam teks menjadi bentuk dasar atau akar kata.

harga jangkau k porsi mienya ples pangsit side dishnya recommended dimsum udang keju udang rambutan jangkau pinggir :
mie gaco jam bgitu jam ramai unjung dimsum kotor lembab kaya gak bersih guna bagus tempat luas
bungkus kalo ngantri takut udah males duluan hahaha tp kadang coba makan wkwk
mo kampus lwt rame trs cuz enak murah mie iblis lvl turut udah pedes bgt udang rambutankeju enakk bgt wajib coba blr
dateng pas weekday jam antri tp cepet ko total antri sampe makan jam staffnya helpful ramah servicenya oke bgt deh r
mantab dah mie gaco enak banget toppingnya lengkap harga jangkau banget gak ngecewain lumayan antri aja gak gerak cep
makan ckup enak murah waiting list dlu makan ckup nya ramai
porsi yg kali coba antri engga tetep oke cocok suka hangout bareng teman ajak keluarga oke
antre klo udh laper dateng rush hour trus dine in ga rekomendasi nunggunya menu ga mie mie an dimsum emg enak sih kl
worth it harga layan improve yg signifikan masuk tolong sedia sampah tissue
harga murmer bgt porsi mie pangsit goreng rb minum jg murah cemilan spt lumpia dimsum rb aja tpmatnya luas enak tp :
mie nya enaaakantrian panjang tapi cepet banget dapet menu nyamura meriahternyata level iblis n setan level pedeeeee
kali nyoba antriny rame bgt sampe ga bagi dine in take away tp krn bangku yg kosong krn udah tutup ttp makan pantes:
antri menit antri bayar cash or cashless menu murah cuman meja nyaman oranh lantai kuga kadang licin
ruaameee poll antri sampe an luama antri nyaa mie enak harga jangkau
rame bgt outletnya antri mulu hihhi hanya pesan online by grab oke laah harga segitu mantaapp

Gambar. 7 Hasil Stemming

3. Transformation

Transformation ini melibatkan perubahan format, struktur, atau representasi data agar sesuai dengan kebutuhan analisis. Contohnya, melakukan *normalisasi* data numerik, melakukan *encoding* pada data kategorikal, atau mengubah representasi data teks menjadi representasi vektor menggunakan teknik seperti TF-IDF.

1) TF- IDF (Term Frequency-Inverse Document Frequency)

Term Frequency merupakan konsep pembobotan yang menentukan seberapa sering (frekuensi) term muncul dalam dokumen. *Document Frequency* merupakan jumlah dokumen dimana term muncul [13]. TF-IDF adalah salah satu teknik pembobotan data. Dalam metode TF-IDF, *Term Frequency* lebih berfokus pada istilah yang sering muncul dalam suatu dokumen sedangkan *Inverse Document Frequency* lebih berfokus pada pemberian bobot rendah untuk istilah yang muncul dalam banyak dokumen [5].

Tahap terakhir dalam pra-pemrosesan adalah pembobotan kata (indeksasi). Dalam tahap indeksasi ini, digunakan metode tf dan idf untuk melakukan pembobotan kata, di mana kata-kata diubah menjadi vektor.

Table 1 Contoh Dokumen

Dokumen	Ulasan	Sentimen
Dokumen 1 (D1)	ruaameee poll antri sampe an luama antri nyaa mie enak harga jangkau	Positif
Dokumen 2 (D2)	nyaman bersih luas parkir lega ramai layan makan enak ramah kantong	Positif
Dokumen 3 (D3)	gila gila gila antri banget nunggu jam mie yg fenomenal hahahaha	Negatif
Dokumen 4 (D4)	mie gaco enak murah sabar antri karna jam pagi antri bgt hahha	Positif
Dokumen 5 (D5)	udh bas kali ksni puas tp ya hrs sabar nunggu antri ordeer	Positif
Dokumen 6 (D6)	mie nya enak parkir luas sabar antri kalo makan antri rame bgt	Positif
Dokumen 7 (D7)	layan cepat antri ga	Positif
Dokumen 8 (D8)	ya sabar sabar antri sih kalo udah jam makan siang	Negatif
Dokumen 9 (D9)	makan enak nya nyaman nunggu antri aja	Positif
Dokumen 10 (D10)	harga murah banget enak banget suka mie iblis isi gak pelit pangsit enak	Positif

Table 2 Perhitungan TF-IDF

Token	tf										df	D/df	IDF (log D/df)
	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10			
ruaameee	1	0	0	0	0	0	0	0	0	0	1	10	1
poll	1	0	0	0	0	0	0	0	0	0	1	10	1
antri	2	0	1	2	1	2	1	1	1	0	8	1,25	0,096910013
sampe	1	0	0	0	0	0	0	0	0	0	1	10	1
an	1	0	0	0	0	0	0	0	0	0	1	10	1
luama	1	0	0	0	0	0	0	0	0	0	1	10	1
nyaa	1	0	0	0	0	0	0	0	0	0	1	10	1
mie	1	0	1	1	0	1	0	0	0	1	5	2	0,301029996
enak	1	1	0	1	0	1	0	0	1	2	6	1,666667	0,22184875
harga	1	0	0	0	0	0	0	0	0	1	2	5	0,698970004
jangkau	1	0	0	0	0	0	0	0	0	0	1	10	1
nyaman	0	1	0	0	0	0	0	0	1	0	2	5	0,698970004
bersih	0	1	0	0	0	0	0	0	0	0	1	10	1
luas	0	1	0	0	0	1	0	0	0	0	2	5	0,698970004
parkir	0	1	0	0	0	1	0	0	0	0	2	5	0,698970004
lega	0	1	0	0	0	0	0	0	0	0	1	10	1
ramai	0	1	0	0	0	0	0	0	0	0	1	10	1
layan	0	1	0	0	0	0	1	0	0	0	2	5	0,698970004
makan	0	1	0	0	0	1	0	1	1	0	4	2,5	0,397940009
ramah	0	1	0	0	0	0	0	0	0	0	1	10	1
kantong	0	1	0	0	0	0	0	0	0	0	1	10	1
gila	0	0	3	0	0	0	0	0	0	0	1	10	1
banget	0	0	1	0	0	0	0	0	0	2	2	5	0,698970004
nunggu	0	0	1	0	1	0	0	0	1	0	3	3,333333	0,522878745
jam	0	0	1	1	0	0	0	1	0	0	3	3,333333	0,522878745
yg	0	0	1	0	0	0	0	0	0	0	1	10	1
fenomenal	0	0	1	0	0	0	0	0	0	0	1	10	1
hahahaha	0	0	1	0	0	0	0	0	0	0	1	10	1
gaco	0	0	0	1	0	0	0	0	0	0	1	10	1
murah	0	0	0	1	0	0	0	0	0	1	2	5	0,698970004
sabar	0	0	0	1	1	1	0	2	0	0	4	2,5	0,397940009
karna	0	0	0	1	0	0	0	0	0	0	1	10	1

pagi	0	0	0	1	0	0	0	0	0	0	1	10	1
bgt	0	0	0	1	0	1	0	0	0	0	2	5	0,698970004
hahha	0	0	0	1	0	0	0	0	0	0	1	10	1
udh	0	0	0	0	1	0	0	0	0	0	1	10	1
bas	0	0	0	0	1	0	0	0	0	0	1	10	1
kali	0	0	0	0	1	0	0	0	0	0	1	10	1
ksni	0	0	0	0	1	0	0	0	0	0	1	10	1
puas	0	0	0	0	1	0	0	0	0	0	1	10	1
tp	0	0	0	0	1	0	0	0	0	0	1	10	1
ya	0	0	0	0	1	0	0	1	0	0	2	5	0,698970004
hrs	0	0	0	0	1	0	0	0	0	0	1	10	1
ordeer	0	0	0	0	1	0	0	0	0	0	1	10	1
nya	0	0	0	0	0	1	0	0	1	0	2	5	0,698970004
kalo	0	0	0	0	0	1	0	1	0	0	2	5	0,698970004
rame	0	0	0	0	0	1	0	0	0	0	1	10	1
cepat	0	0	0	0	0	0	1	0	0	0	1	10	1
ga	0	0	0	0	0	0	1	0	0	0	1	10	1
sih	0	0	0	0	0	0	0	1	0	0	1	10	1
udah	0	0	0	0	0	0	0	1	0	0	1	10	1
siang	0	0	0	0	0	0	0	1	0	0	1	10	1
aja	0	0	0	0	0	0	0	0	1	0	1	10	1
suka	0	0	0	0	0	0	0	0	0	1	1	10	1
iblis	0	0	0	0	0	0	0	0	0	1	1	10	1
isi	0	0	0	0	0	0	0	0	0	1	1	10	1
gak	0	0	0	0	0	0	0	0	0	1	1	10	1
pelit	0	0	0	0	0	0	0	0	0	1	1	10	1
pangsit	0	0	0	0	0	0	0	0	0	1	1	10	1

Table 3 Perhitungan Bobot Kata

W = IDF									
D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
0,1938	0	0,0969	0,194	0,0969	0,1938	0,097	0,097	0,097	0
1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
0,301	0	0,301	0,301	0	0,301	0	0	0	0,301
0,2218	0,2218	0	0,222	0	0,2218	0	0	0,222	0,444
0,699	0	0	0	0	0	0	0	0	0,699
1	0	0	0	0	0	0	0	0	0
0	0,699	0	0	0	0	0	0	0,699	0
0	1	0	0	0	0	0	0	0	0
0	0,699	0	0	0	0,699	0	0	0	0
0	0,699	0	0	0	0,699	0	0	0	0

4. Data Mining

Data mining merupakan salah satu proses penting yang dilakukan untuk mengekstraksi pola, pengetahuan, atau informasi yang berharga dari data yang telah dipersiapkan melalui tahap *preprocessing* dan *transformation*. Dalam penelitian ini menggunakan teknik atau algoritma klasifikasi yaitu *Naïve Bayes* dan *C4.5*.

1) Algoritma Naïve Bayes

Table 4 Data Training Naive Bayes

Dokumen	Ulasan	Sentimen
Dokumen 1 (D1)	ruaameee poll antri sampe an luama antri nyaa mie enak harga jangkau	Positif
Dokumen 2 (D2)	nyaman bersih luas parkir lega ramai layan makan enak ramah kantong	Positif
Dokumen 3 (D3)	gila gila gila antri banget nunggu jam mie yg fenomenal hahahaha	Negatif
Dokumen 4 (D4)	mie gaco enak murah sabar antri karna jam pagi antri bgt hahha	Positif
Dokumen 5 (D5)	udh bas kali ksni puas tp ya hrs sabar nunggu antri ordeer	Positif
Dokumen 6 (D6)	mie nya enak parkir luas sabar antri kalo makan antri rame bgt	Positif
Dokumen 7 (D7)	makan enak nya nyaman nunggu antri aja	Positif
Dokumen 8 (D8)	harga murah banget enak banget suka mie iblis isi gak pelit pangsit enak	Positif

Table 5 Tabel Prediksi Naive Bayes

Prediksi	Aktual		Total
	Positif	Negatif	
ruaameee	1	0	1
poll	1	0	1
antri	5	1	6
sampe	1	0	1
an	1	0	1
luama	1	0	1
nyaa	1	0	1
mie	4	1	5
enak	6	0	6
harga	2	0	2
jangkau	1	0	1
nyaman	2	0	2
bersih	1	0	1
luas	2	0	2
parkir	2	0	2
lega	1	0	1
ramai	1	0	1
layan	1	0	1
makan	3	0	3
ramah	1	0	1
kantong	1	0	1
gila	0	3	3
banget	1	1	2
nunggu	2	1	3
jam	1	1	2
yg	0	1	1
fenomenal	0	1	1
hahahaha	0	1	1
gaco	1	0	1
murah	2	0	2
sabar	3	0	3
karna	1	0	1
pagi	1	0	1
bgt	2	0	2
hahha	1	0	1
udh	1	0	1
bas	1	0	1

kali	1	0	1
ksni	1	0	1
puas	1	0	1
tp	1	0	1
ya	1	0	1
hrs	1	0	1
ordeer	1	0	1
nya	2	0	2
kalo	1	0	1
rame	1	0	1
aja	1	0	1
suka	1	0	1
iblis	1	0	1
isi	1	0	1
gak	1	0	1
pelit	1	0	1
pangsit	1	0	1
Jumlah Term	74	11	85

Selanjutnya, untuk membandingkan probabilitas posterior dari setiap kelas dapat menghitung Prior Probability.

Table 6 Prior Probability

	$P_{(Positif)}$	$P_{(Negatif)}$
Prior Probability	$\frac{74}{85} = 0,87058824$	$\frac{11}{85} = 0,12941176$

Model probabilitas dihasilkan dengan menggunakan persamaan (3).

Kata 'antri':

$$P(a_{antri}|V_{Positif}) = \frac{5 + 1}{74 + 85} = \frac{6}{159} = 0,03773585$$

$$P(a_{antri}|V_{Negatif}) = \frac{1 + 1}{11 + 85} = \frac{2}{96} = 0,02083333$$

Kata 'mie':

$$P(a_{mie}|V_{Positif}) = \frac{4 + 1}{74 + 85} = \frac{5}{159} = 0,03144654$$

$$P(a_{mie}|V_{Negatif}) = \frac{1 + 1}{11 + 85} = \frac{2}{96} = 0,02083333$$

Kata 'enak':

$$P(a_{enak}|V_{Positif}) = \frac{6 + 1}{74 + 85} = \frac{7}{159} = 0,04402516$$

$$P(a_{enak}|V_{Negatif}) = \frac{0 + 1}{11 + 85} = \frac{1}{96} = 0,01041667$$

Kata 'makan':

$$P(a_{makan}|V_{Positif}) = \frac{3 + 1}{74 + 85} = \frac{4}{159} = 0,02515723$$

$$P(a_{makan}|V_{Negatif}) = \frac{0 + 1}{11 + 85} = \frac{1}{96} = 0,01041667$$

Kata 'gila':

$$P(a_{gila}|V_{Positif}) = \frac{0 + 1}{74 + 85} = \frac{1}{159} = 0,00628931$$

$$P(a_{gila}|V_{Negatif}) = \frac{3 + 1}{11 + 85} = \frac{4}{96} = 0,04166667$$

Kata ‘murah’:

$$P(a_{murah}|V_{Positif}) = \frac{2+1}{74+85} = \frac{3}{159} = 0,01886792$$

$$P(a_{murah}|V_{Negatif}) = \frac{0+1}{11+85} = \frac{1}{96} = 0,1041667$$

Table 7 Data Uji Naive Bayes

Dokumen	Ulasan	Sentimen
Dokumen 1 (D1)	enak murah	?
Dokumen 2 (D2)	antri gila	?

$$P(Uji | positif)$$

$$= P(positif) \times P(enak | positif) \times P(murah | positif)$$

$$= 0,87058824 \times 0,0440251 \times 0,01886792$$

$$= 0,00072316$$

$$P(Uji | negatif)$$

$$= P(negatif) \times P(enak | negatif) \times P(murah | negatif)$$

$$= 0,12941176 \times 0,01041667 \times 0,1041667$$

$$= 0,00014042$$

$$P(Uji | positif)$$

$$= P(positif) \times P(antri | positif) \times P(gila | positif)$$

$$= 0,87058824 \times 0,03773585 \times 0,00628931$$

$$= 0,00020662$$

$$P(Uji | negatif)$$

$$= P(negatif) \times P(antri | negatif) \times P(gila | negatif)$$

$$= 0,12941176 \times 0,02083333 \times 0,04166667$$

$$= 0,00011234$$

Dari perhitungan di atas dapat disimpulkan bahwa data uji dokumen 1 (D1) prediksi sentimen positif dan pada dokumen 2 (D2) prediksi sentimen positif.

2) Algoritma C4.5

Table 8 Table 8 Data Training C4.5

Dokumen	Data Training	Kelas
Dokumen 1 (D1)	enak tapi antri	positif
Dokumen 2 (D2)	mie gacoan biasa aja	negatif
Dokumen 3 (D3)	pelayanan nya buruk mie nya biasa aja	negatif
Dokumen 4 (D4)	mie nya si oke pelayanan nya yang lama	positif
Dokumen 5 (D5)	gacoan biasa aja	negatif

Table 9 Nilai Data Training C4.5

D	enak	antri	mie	gacoan	pelayanan	buruk	lama	Kelas
D1	1	1	0	0	0	0	0	positif
D2	0	0	1	1	0	0	0	negatif
D3	0	0	1	0	1	1	0	negatif
D4	0	0	1	0	1	0	1	positif
D5	0	0	0	1	0	0	0	negatif

Perhitungan entropy untuk seluruh training data :

$$\begin{aligned} \text{entropy (total)} &= \left(-\frac{2}{5}\right) \times \log_2\left(\frac{2}{5}\right) + \left(-\frac{2}{5}\right) \times \log_2\left(\frac{2}{5}\right) + \left(-\frac{3}{5}\right) \times \log_2\left(\frac{3}{5}\right) + \left(-\frac{3}{5}\right) \times \log_2\left(\frac{3}{5}\right) \\ &+ \left(-\frac{1}{5}\right) \times \log_2\left(\frac{1}{5}\right) = 2,406 \end{aligned}$$

Perhitungan entropy dan gain pada setiap atribut untuk menentukan simpul akar:

a. enak

Perhitungan entropy atribut “enak” dengan nilai 0 dan 1:

$$\text{Entropy (0)} = \left(-\frac{3}{4}\right) \times \log_2\left(\frac{3}{4}\right) + \left(-\frac{1}{4}\right) \times \log_2\left(\frac{1}{4}\right) = 0,881$$

$$\text{Entropy (1)} = 0$$

$$\text{Gain (total, enak)} = (2,406) - \left(\frac{4}{5}\right) \times (0,811) + \left(\frac{1}{5}\right) \times (0) = 1,757$$

b. antri

Perhitungan entropy atribut “antri” dengan nilai 0 dan 1:

$$\text{Entropy (0)} = \left(-\frac{3}{4}\right) \times \log_2\left(\frac{3}{4}\right) + \left(-\frac{1}{4}\right) \times \log_2\left(\frac{1}{4}\right) = 0,881$$

$$\text{Entropy (1)} = 0$$

$$\text{Gain (total, antri)} = (2,406) - \left(\frac{4}{5}\right) \times (0,811) + \left(\frac{1}{5}\right) \times (0) = 1,757$$

c. mie

Perhitungan entropy atribut “mie” dengan nilai 0 dan 1:

$$\text{Entropy (0)} = \left(-\frac{1}{2}\right) \times \log_2\left(\frac{1}{2}\right) + \left(-\frac{1}{2}\right) \times \log_2\left(\frac{1}{2}\right) = 1$$

$$\text{Entropy (1)} = \left(-\frac{2}{3}\right) \times \log_2\left(\frac{2}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2\left(\frac{1}{3}\right) = 0,918$$

$$\text{Gain (total, mie)} = (2,406) - \left(\frac{2}{5}\right) \times (1) + \left(\frac{1}{5}\right) \times (0,918) = 2,557$$

d. gacoan

Perhitungan entropy atribut “gacoan” dengan nilai 0 dan 1:

$$\text{Entropy (0)} = \left(-\frac{1}{3}\right) \times \log_2\left(\frac{1}{3}\right) + \left(-\frac{2}{3}\right) \times \log_2\left(\frac{2}{3}\right) = 0,918$$

$$\text{Entropy (1)} = 0$$

$$\text{Gain (total, gacoan)} = (2,406) - \left(\frac{3}{5}\right) \times (0,918) + \left(\frac{2}{5}\right) \times (0) = 1,855$$

5. pelayanan

Perhitungan entropy atribut “pelayanan” dengan nilai 0 dan 1:

$$\text{Entropy (0)} = \left(-\frac{2}{3}\right) \times \log_2\left(\frac{2}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2\left(\frac{1}{3}\right) = 0,918$$

$$\text{Entropy (1)} = \left(-\frac{1}{2}\right) \times \log_2\left(\frac{1}{2}\right) + \left(-\frac{1}{2}\right) \times \log_2\left(\frac{1}{2}\right) = 1$$

$$\text{Gain (total, pelayanan)} = (2,406) - \left(\frac{3}{5}\right) \times (0,918) + \left(\frac{2}{5}\right) \times (1) = 2,553$$

6. buruk

Perhitungan entropy atribut “buruk” dengan nilai 0 dan 1:

$$\text{Entropy (0)} = \left(-\frac{2}{4}\right) \times \log_2\left(\frac{2}{4}\right) + \left(-\frac{2}{4}\right) \times \log_2\left(\frac{2}{4}\right) = 1$$

$$\text{Entropy (1)} = 0$$

$$\text{Gain (total, buruk)} = (2,406) - \left(\frac{4}{5}\right) \times (1) + \left(\frac{1}{5}\right) \times (0) = 1,606$$

7. lama

Perhitungan entropy atribut “lama” dengan nilai 0 dan 1:

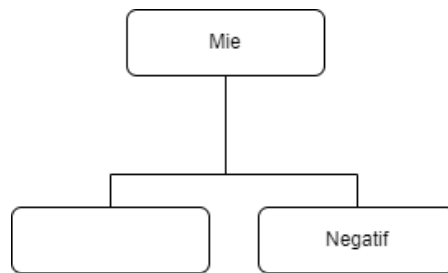
$$\text{Entropy (0)} = \left(-\frac{3}{4}\right) \times \log_2\left(\frac{3}{4}\right) + \left(-\frac{1}{4}\right) \times \log_2\left(\frac{1}{4}\right) = 0,811$$

$$\text{Entropy (1)} = 0$$

$$\text{Gain (total, lama)} = (2,406) - \left(\frac{4}{5}\right) \times (0,811) + \left(\frac{1}{5}\right) \times (0) = 1,757$$

Table 10 Perhitungan Gain untuk Simpul Akar

Atribut	Neg(0)	Pos(0)	Entropy(0)	Neg(1)	Pos(1)	Entropy(1)	Gain
enak	3	1	0.811	0	1	0	1.757
antri	3	1	0.811	0	1	0	1.755
mie	1	1	1	2	1	0.918	2.557
gacoan	1	2	0.918	2	0	0	1.855
pelayanan	2	1	0.918	1	1	1	2.255
buruk	2	2	1	1	0	0	1.606
lama	3	1	0.811	0	1	0	1.757
Gain Tertinggi	: 2.557						
Node	: mie						



Gambar. 8 Hasil Simpul Akar

Table 11 Nilai Data Training

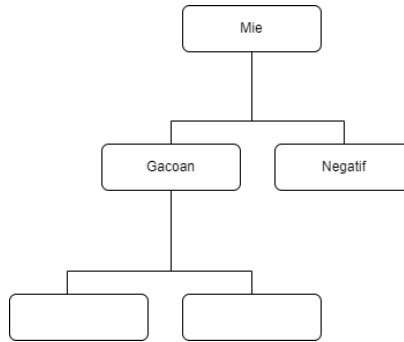
D	enak	antri	gacoan	pelayanan	buruk	lama	Kelas
D1	1	1	0	0	0	0	positif
D4	0	0	0	1	0	1	positif
D5	0	0	1	0	0	0	negatif

Perhitungan entropy total sebagai berikut:

$$\text{entropy (total)} = \left(-\frac{2}{3}\right) \times \log_2 \left(\frac{2}{3}\right) + \left(-\frac{2}{3}\right) \times \log_2 \left(\frac{2}{3}\right) + \left(-\frac{1}{3}\right) \times \log_2 \left(\frac{1}{3}\right) = 1,308$$

Table 12 Perhitungan Gain untuk Simpul Akar

Atribut	Neg(0)	Pos(0)	Entropy(0)	Neg(1)	Pos(1)	Entropy(1)	Gain
enak	1	1	1	0	1	0	0.641
antri	1	1	1	0	1	0	0.641
gacoan	0	2	0	1	0	0	1.308
pelayanan	1	1	1	0	1	0	0.641
buruk	1	2	0.918	0	0	0	0.389
lama	1	1	1	0	1	0	0.641
Gain Tertinggi	: 1.308						
Node	: gacoan						



Gambar. 9 Hasil Simpul Akar Gacoan

Table 13 Nilai Data Training

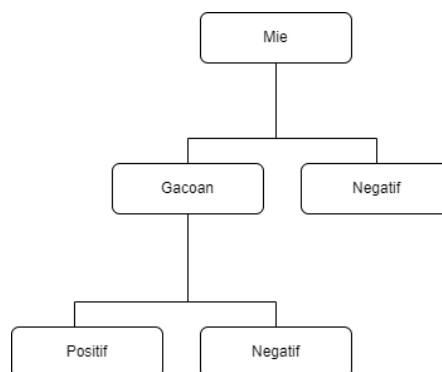
D	enak	antri	pelayanan	buruk	lama	Kelas
D1	1	1	0	0	0	positif
D2	0	0	1	0	1	positif

Perhitungan entropy total sebagai berikut:

$$entropy (total) = \left(-\frac{2}{2}\right) \times \log_2 \left(\frac{2}{2}\right) + \left(-\frac{2}{2}\right) \times \log_2 \left(\frac{2}{2}\right) = 0$$

Table 14 Perhitungan Gain untuk Simpul Akar

Atribut	Neg(0)	Pos(0)	Entropy(0)	Neg(1)	Pos(1)	Entropy(1)	Gain
enak	0	1	0	0	1	0	0
antri	0	1	0	0	1	0	0
pelayanan	0	1	0	0	1	0	0
buruk	0	2	0	0	0	0	0
lama	0	1	0	0	1	0	0



Gambar. 10 Hasil Simpul Akar Gacoan

Karena nilai entropi untuk cabang simpul "Gacoan" telah mencapai hasil yang sama dengan 0, kedua cabang atribut "Gacoan" akan menjadi simpul daun dan tidak ada perhitungan gain yang dilakukan untuk menentukan simpul selanjutnya.

5. Evaluation

Evaluation merupakan proses penting untuk mengukur kualitas dan efektivitas dari hasil analisis dan ekstraksi pengetahuan yang diperoleh melalui proses KDD. Evaluasi dilakukan untuk memastikan bahwa pengetahuan yang dihasilkan adalah valid, berguna, dan dapat diandalkan untuk pengambilan keputusan

Confusion Matrix

Berikut ini contoh perhitungan manual *confusion matrix Naïve Bayes*

$$\text{Accuracy} = \frac{105+110}{105+31+13+110} \times 100\%$$

$$\text{Accuracy} = \frac{215}{259} \times 100\%$$

$$\text{Accuracy} = 83 \%$$

$$\text{Precision} = \frac{105}{105+31} \times 100\%$$

$$\text{Precision} = \frac{105}{136} \times 100\%$$

$$\text{Precision} = 77 \%$$

$$\text{Recall} = \frac{105}{105+13} \times 100\%$$

$$\text{Recall} = \frac{105}{118} \times 100\%$$

$$\text{Recall} = 88 \%$$

$$F - \text{Measure} = 2 \times \frac{77 \times 88}{77+88} \times 100\%$$

$$F - \text{Measure} = 2 \times \frac{6.776}{165} \times 100\%$$

$$F - \text{Measure} = 82 \%$$

Berikut ini contoh perhitungan manual *confusion matrix C4.5*

$$\text{Accuracy} = \frac{83+108}{83+33+35+108} \times 100\%$$

$$\text{Accuracy} = \frac{191}{259} \times 100\%$$

$$\text{Accuracy} = 73 \%$$

$$\text{Precision} = \frac{83}{83+33} \times 100\%$$

$$\text{Precision} = \frac{83}{116} \times 100\%$$

$$\text{Precision} = 71 \%$$

$$\text{Recall} = \frac{83}{83+35} \times 100\%$$

$$\text{Recall} = \frac{83}{118} \times 100\%$$

$$\text{Recall} = 70 \%$$

$$F - \text{Measure} = 2 \times \frac{71 \times 70}{71+70} \times 100\%$$

$$F - \text{Measure} = 2 \times \frac{4.970}{141} \times 100\%$$

$$F - \text{Measure} = 70 \%$$

6. Knowledge Presentation

Wordcloud adalah salah satu metode untuk memvisualisasikan data teks agar lebih menarik dan mudah dipahami. Dengan menggunakan *wordcloud*, dapat diketahui frekuensi kata-kata yang sering muncul dalam kategori positif dan negatif. Tujuan dari pembuatan *wordcloud* ini adalah untuk mengetahui kata-kata yang sering muncul yang berpengaruh dalam model klasifikasi uji coba. Pembuatan visualisasi ini dilakukan menggunakan bahasa *Python* di *Google Colab*. Pada gambar 9 berikut hasil *wordcloud* penelitian.

- [6] A. Rakhman and M. R. Tsani, “ANALISIS SENTIMEN REVIEW MEDIA MASSA,” vol. 8, no. 2, pp. 78–82, 2019.
- [7] E. Indrayuni, “Klasifikasi Text Mining Review Produk Kosmetik Untuk Teks Bahasa Indonesia Menggunakan Algoritma Naïve Bayes,” vol. VII, no. 1, pp. 29–36, 2019.
- [8] M. Marsono, “Penerapan Data Mining Pengaturan Pola Tata Letak Barang Pada Berkah Swalayan Untuk Strategi Penjualan Menggunakan Algoritma Apriori,” *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 3, no. 2, pp. 170–175, 2019, doi: 10.30743/infotekjar.v3i2.908.
- [9] S. Ramos *et al.*, “Data mining techniques for electricity customer characterization,” *Procedia Comput. Sci.*, vol. 186, pp. 475–488, 2021, doi: 10.1016/j.procs.2021.04.168.
- [10] A. Kulsumarwati, I. Purnamasari, and B. A. Darmawan, “Penerapan SVM dan Information Gain Pada Analisis Sentimen Pelaksanaan Pilkada Saat Pandemi,” *J. Teknol. Inform. dan Komput.*, vol. 7, no. 2, pp. 101–109, 2021, doi: 10.37012/jtik.v7i2.641.
- [11] H. Herlawati, R. T. Handayanto, P. D. Atika, F. N. Khasanah, A. Y. P. Yusuf, and D. Y. Septia, “Analisis Sentimen Pada Situs Google Review dengan Naïve Bayes dan Support Vector Machine,” *J. Komtika (Komputasi dan Inform.*, vol. 5, no. 2, pp. 153–163, 2021, doi: 10.31603/komtika.v5i2.6280.
- [12] F. Albasithu and A. Wibowo, “Perbandingan Algoritma Naïve Bayes Dan C4 . 5 pada Analisis Sentimen Presiden 3 Periode di Twitter Comparison of Naive Bayes and C4 . 5 Algorithm on Presidential Sentiment Analysis 3 Period on Twitter,” no. September, pp. 510–516, 2022.
- [13] M. Hamka, N. Alfatari, and D. Ratna Sari, “Analisis Sentimen Produk Kecantikan Jenis Serum Menggunakan Algoritma Naïve Bayes Classifier,” *J. Sist. Komput. dan Inform.*, vol. 4, no. 1, p. 64, 2022, doi: 10.30865/json.v4i1.4740.