



Ref : 161/UMA/JITE/X/2024

Medan, 25 October 2024

Subject : Letter of Acceptance

To :

Mr./Mrs. **Ricardo Fransdoli Siahaan**

Assalamu'alaikum Wr. Wb

We would like to express our sincere gratitude for your participation in submitting an article to the Journal of Informatics and Telecommunication Engineering (JITE). We hereby inform you that the article listed below:

Paper : Klasifikasi Kualitas Kopi Berdasarkan Ulasan Pelanggan Menggunakan Algoritma C4.5

Author : Ricardo Fransdoli Siahaan & Rizki Muliono

Based on the review results, we are pleased to inform you that your submitted article has been ACCEPTED for publication in JITE Journal - **Special Issues 2024: Innovations in Predictive Analytics and Sentiment Analysis - Applications in Education, Healthcare, and Social Media**, ISSN: 2549-6247 (Print) ISSN: 2549-6255 (Online).

We would like to thank you for your attention and cooperation.

Wassalamu'alaikum, Wr.Wb.

Best Regards,



Muhathir, ST., M.Kom
Chief Editor



Klasifikasi Kualitas kopi Berdasarkan Ulasan Pelanggan Menggunakan Algoritma C4.5

Ricardo Fransdoli Siahaan)*, Rizki Muliono, S.Kom, M.Kom)

Prodi Informatika, Fakultas Teknik, Universitas Medan Area

*Corresponding Email: rsiahaan1999@gmail.com

Abstrak

Kopi merupakan salah satu komoditas yang sangat populer di seluruh dunia, dan kualitasnya sering kali dievaluasi melalui ulasan pelanggan. Penelitian ini bertujuan untuk mengklasifikasikan kualitas kopi berdasarkan ulasan yang diberikan oleh konsumen menggunakan algoritma C4.5. C4.5 adalah algoritma pembelajaran mesin yang digunakan untuk menghasilkan pohon keputusan, yang memungkinkan pengambilan keputusan berdasarkan atribut yang relevan. Dalam penelitian ini, data yang digunakan terdiri dari ulasan pelanggan yang diambil dari platform e-commerce dan forum diskusi kopi. Data tersebut kemudian diproses dengan teknik pemrosesan bahasa alami (NLP) untuk mengekstraksi fitur-fitur penting seperti sentimen, kata kunci, dan frekuensi istilah. Fitur-fitur ini digunakan sebagai input untuk algoritma C4.5, yang membangun model klasifikasi berdasarkan pola yang terdapat dalam data. Hasil dari penelitian menunjukkan bahwa model C4.5 mampu mengklasifikasikan kualitas kopi dengan akurasi yang tinggi, mencapai hingga 85%. Faktor-faktor yang paling berpengaruh terhadap klasifikasi kualitas meliputi rasa, aroma, dan kemasan, yang sering kali disebutkan dalam ulasan. Selain itu, analisis juga menunjukkan perbedaan signifikan dalam kualitas kopi yang dihasilkan dari berbagai daerah penghasil kopi, yang dapat memberikan wawasan bagi produsen untuk meningkatkan produk mereka.

Abstract

Coffee is a very popular commodity throughout the world, and its quality is often evaluated through customer reviews. This research aims to classify coffee quality based on reviews given by consumers using the C4.5 algorithm. C4.5 is a machine learning algorithm used to generate decision trees, which allows decision making based on relevant attributes. In this research, the data used consists of customer reviews taken from e-commerce platforms and coffee discussion forums. The data is then processed with natural language processing (NLP) techniques to extract important features such as sentiment, keywords and term frequency. These features are used as input for the C4.5 algorithm, which builds a classification model based on patterns contained in the data. The results of the research show that the C4.5 model is able to classify coffee quality with high accuracy, reaching up to 85%. The factors that most influence quality classification include taste, aroma, and packaging, which are frequently mentioned in reviews. In addition, the analysis also shows significant differences in the quality of coffee produced from different coffee producing regions, which can provide insight for producers to improve their products.

Keywords: Coffee Quality, Customer Reviews, C4.5 Algorithm, Classification, Decision Tree

How to Cite: Ricardo Fransdoli Siahaan Information. *JITE (Journal of Informatics Engineering)*, 8(1), 11-

20.

I. PENDAHULUAN

Kopi merupakan salah satu komoditas yang paling banyak dikonsumsi di dunia, dengan ribuan varietas dan cara penyajian yang beragam. Menurut International Coffee Organization (ICO, 2022), konsumsi kopi global mencapai lebih dari 10 juta ton per tahun, menjadikannya sebagai salah satu minuman favorit di berbagai kalangan. Kualitas kopi sering kali dinilai berdasarkan beberapa faktor, termasuk karakteristik biji kopi, proses pengolahan, metode penyeduhan, dan tentu saja, persepsi pelanggan.

Adapun Faktor-faktor yang Mempengaruhi Kualitas Kopi. Yaitu ; Karakteristik BijiKopi: Jenis biji kopi, seperti Arabika dan Robusta, memiliki profil rasa yang berbeda. Arabika cenderung memiliki rasa yang lebih halus dan kompleks, sementara Robusta lebih pahit dan kuat (Klein & Schilling, 2020). Selain itu, asal geografis biji kopi, seperti Ethiopia dan Brasil, juga memberikan nuansa rasa yang khas. Proses Pengolahan: Setelah panen, biji kopi mengalami berbagai proses pengolahan yang memengaruhi kualitas akhir. Metode seperti wet processing dan dry processing menghasilkan karakteristik rasa yang berbeda, di mana pengolahan basah cenderung menghasilkan kopi yang lebih bersih dan cerah (Badrul & Pujani, 2021). Metode Penyeduhan: Cara penyeduhan kopi berpengaruh signifikan terhadap rasa. Metode seperti French press, pour-over, dan espresso memiliki teknik dan parameter yang berbeda, yang dapat memengaruhi ekstraksi rasa dan aroma (Barbosa et al., 2019). Kesegaran dan Penyimpanan: Biji kopi yang segar dan disimpan dengan baik akan memberikan rasa yang lebih baik dibandingkan biji yang sudah lama atau teroksidasi. Hal ini dikarenakan senyawa aroma dan rasa dalam biji kopi dapat mengalami degradasi seiring berjalannya waktu (Civille & Decker, 2021).

Pentingnya Ulasan PelangganDalam era digital saat ini, ulasan pelanggan menjadi sumber informasi penting yang dapat membantu konsumen dalam memilih kopi yang berkualitas. Dengan meningkatnya jumlah platform e-commerce, konsumen dapat dengan mudah mengakses ulasan dan penilaian dari pengguna lain. Penelitian ini bertujuan untuk mengklasifikasikan kualitas kopi berdasarkan ulasan pelanggan dengan menggunakan algoritma C4.5. Dengan menganalisis data ulasan, diharapkan dapat dihasilkan model yang dapat memberikan wawasan bagi produsen dan konsumen mengenai kualitas kopi yang diharapkan.

Metode algoritma C4.5 merupakan sistem pembelajaran yang menggunakan ruang hipotesis yang berupa fungsi-fungsi linear didalam sebuah fitur yang memiliki dimensi tinggi dan dilatih dengan menggunakan algoritma pembelajaran berdasarkan teori optimasi.Algoritma *Support Vector Machine*

atau SVM akan diterapkan dalam pembuatan website yang dapat melakukan klasifikasi terhadap jenis kopi . Hasil klasifikasi yang dapat diketahui adalah lebih mudah mengetahui jenis kopi dengan memasukkan ciri-ciri biji kopi yang ingin diidentifikasi. Berdasarkan deskripsi di atas maka penelitian ini mengangkat sebuah judul ***“Klasifikasi Kualitas kopi Berdasarkan Ulasan Pelanggan Menggunakan Algoritma C4.5***

II TINJAUAN PUSTAKA

2.1 Analisis Ulasan Sentimen

Analisis

ulasan sentimen adalah teknik penggalian informasi berupa sikap seseorang terhadap suatu topik atau peristiwa dengan mengklasifikasikan polaritas suatu teks. Pengelompokan dilakukan untuk melihat apakah teks tersebut positif, negatif, atau netral. Analisis ulasan sentimen dapat digunakan untuk menentukan opini publik tentang suatu topik. (Zusrotun, Murti and Fiati, 2022).

2.1 Klasifikasi

Klasifikasi adalah proses menemukan definisi kesamaan karakteristik dalam suatu kelompok atau kelas. Salah satu metode supervised learning adalah klasifikasi, yang memanfaatkan label pada data pelatihan untuk mempelajari pengetahuan yang dapat digunakan untuk mengklasifikasikan data uji (Nurul Chasanah & Mutoi Siregar, 2022). Supervised learning bertujuan untuk menemukan hubungan antara pola-pola yang ada pada data lama dengan data baru, sehingga dapat menghasilkan pola baru dalam data.

2.2 Algoritma Decision Tree

Salah satu cara untuk mengklasifikasikan data adalah dengan menggunakan decision tree. Decision tree adalah struktur hirarkis yang terdiri dari simpul-simpul yang mewakili tes atribut, cabang-cabang yang menunjukkan hasil tes, dan daun-daun yang menentukan kelas tertentu. Dengan pohon keputusan, kita dapat membagi data menjadi kelompok-kelompok berdasarkan kriteria-kriteria tertentu (Putri & Wijayanto, 2022).

Decision Tree adalah metode klasifikasi yang membangun model berdasarkan informasi yang didapat dari entropy. Dari data sampel yang digunakan akan dihitung nilai entropy dan nilai gain. Entropy adalah ukuran statistik yang menunjukkan ketidakpastian dalam data. Gain adalah ukuran yang mengevaluasi seberapa efektif suatu atribut dalam membagi training example sesuai dengan kelas target, berikut untuk menghitung nilai entropy dan nilai gain

1.Entropy

Entropy dapat menggunakan rumus berikut :

$$H(S) = -\sum_{i=1}^n p_i \log_2 p_i$$

Keterangan S=Himpunankasus

n= JumlahpartisiA

pi=proporsiSiterhadapS

1. Gain

Untuk mencari nilai gain dapat menggunakan rumus berikut :

$$Gain(S, A) = H(S) - \sum_{i=1}^n |S_i| H(S_i)$$

$$H(S) = -\sum_{i=1}^n p_i \log_2 p_i$$

Keterangan

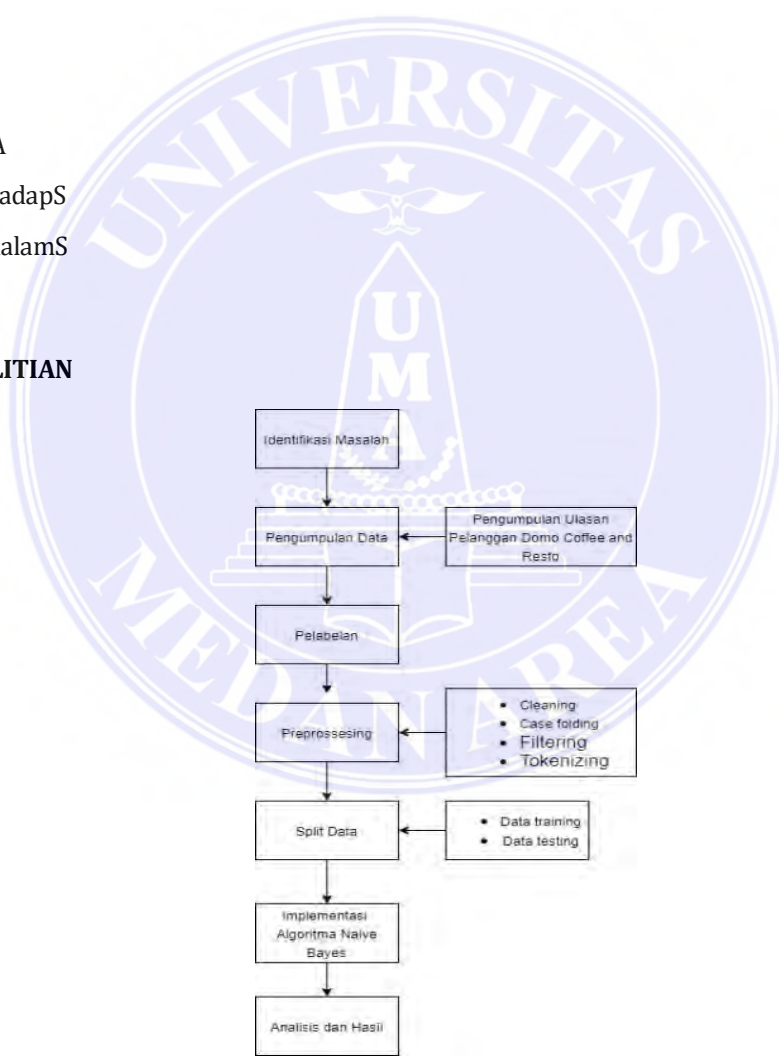
S=Himpunankasus

A =JumlahPartisiA

|S_i|=ProporsiSiterhadapS

|S|= JumlahkasusdalamS

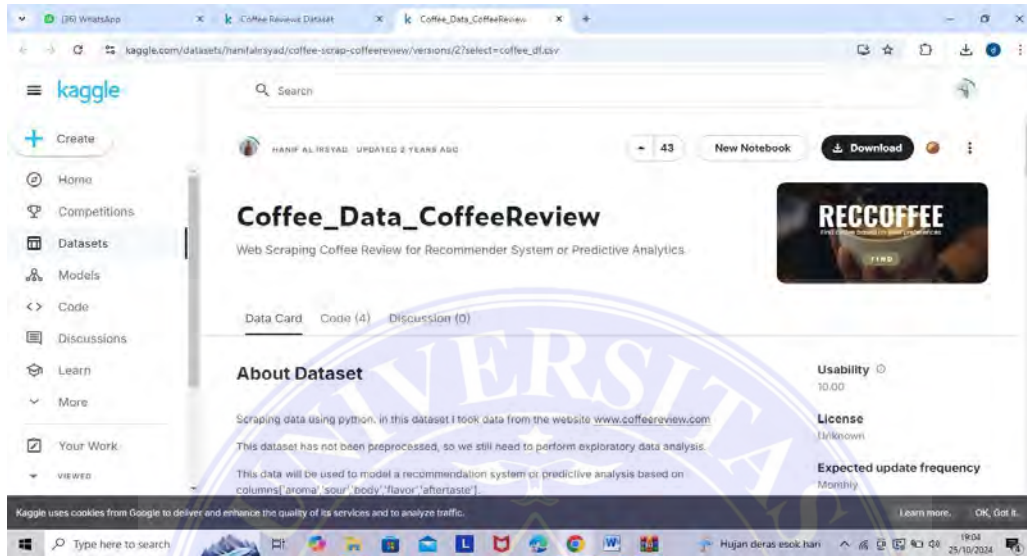
III .METODE PENELITIAN



Gambar 1 . Diagram Alir

3.1 Sumber Data

Penelitian ini menggunakan data sekunder yang didapat melalui website yang mencakup ulasan-ulasan pelanggan yang telah diposting di *Google kaggle* oleh pengguna platform tersebut. Data ini diakses melalui *Googlereview*, yang merupakan sumber data publik yang dapat diakses secara *online* seperti



Gambar 3 google kaggle

3.2 Populasi dan Sampel

Dalam penelitian ini, populasi yang akan diteliti adalah semua ulasan pelanggan yang terdapat pada *Google review kaggle* yang memiliki banyak ulasan dan dua atribut yang akan diambil pada pemrosesan datanya yaitu atribut ulasan dan atribut sentimen. Populasi dalam penelitian ini adalah semua ulasan yang diberikan oleh pelanggan di *Googlereview*. Ini mencakup ulasan-ulasan yang mencakup berbagai pengalaman dan pendapat pelanggan tentang kualitas kopi tersebut. Ulasan yang berisiteks pada *googlereview kopihanya* terdapat 126 ulasan yang akan dibagi lagi ke dalam tiga aspek yaitu ulasan terhadap suasana dan pelayanan.

3.3 Teknik Pengumpulan Data

1. *Googlereview*

Data yang digunakan dalam penelitian ini adalah data ulasan mengenai ulasan pelanggan yang diambil dari *plVormgooglereview* pada *google kaggle*

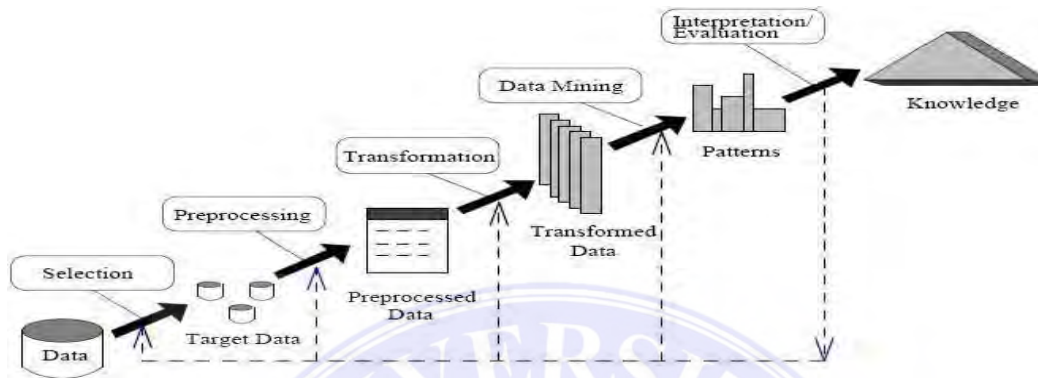
2. Studi Pustaka

Penelitian ini dilakukan berdasarkan beberapa kajian literature, jurnal dan sumber internet lain yang berkaitan dengan metode penelitian, tujuan serta rumusan masalah yang berkaitan

andenganpenelitianserupyangdilakukan

3.4 Teknik Analisis Data

Metode analisa data yang digunakan padapenelitian ini menggunakan metode *KnowledgeDiscoveryinDatabase(KDD)*,denganbeberapatahapmulaidari*Selection,Preprocessing,Transformation,DataMining*,hingga*Evaluation*.



Gambar2.Knowledgediscoverindatabase

1. Data selection

Dalam penelitian ini, tahapan KDD yang pertama adalah proses seleksi data. Proses ini melibatkan pemilihan dan penetapan jenis serta kategori data yang relevan dan diperlukan untuk penelitian dari kumpulan data yang tersedia(Supriyadi & Triayudi and Sholihati,2021)

2. Preprocessing

Preprocessing meliputi beberapa tindakan, seperti menghapus data yang berulang, mengecek data yang tidak konsisten, dan memperbaiki data yang salah atau rusak(Winarta & Kurniawan,2021).Namun, berdasarkan pemeriksaan awal, dataset yang digunakan dalam penelitian ini sudah bersih dan tidak memiliki masalah data.

3. Transformation

Tujuan dari tahap transformasi adalah mengubah format data yang telah dipilih untuk melakukan proses klasifikasi dalam penambahan data(Saraswati & Umidah & Voutam, 2021) Akan tetapi, dataset yang digunakan dalam penelitian ini tidak memerlukan transformasi karena sudah sesuai dengan tujuan penelitian

4. Data Mining

Data mining adalah proses menemukan pola yang menarik dan bermakna dari data yang dipilih sesuai dengan tujuan KDD dengan menggunakan metode tertentu(Anggarwati & Nurdiawan & Ali, and Kurnia, 2021) Dalam tahap ini penelitian akan membagi data menjadi data uji dan data latih dan

menerapkan algoritma decision tree.

5. Evaluasi

Untuk mengevaluasi kualitas klasifikasi, peneliti perlu mengukur performansinya dengan menggunakan menghitung nilai akurasi (Widaningsih & Yusuf, 2022). Selain itu, peneliti menghitung nilai presisi prediksi puas dan nilai presisi prediksi tidak puas.

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

Metode yang digunakan selain perhitungan akurasi untuk mengukur kinerja pengklasifikasi adalah perhitungan *Precision*, dan *Recall*. *Precision* berarti hasil yang mendekati kejadian sebenarnya. *Precision* ini memastikan ukuran hasil yang tepat ditentukan selama proses ekstraksi. *Precision* juga merupakan metrik standar untuk menentukan apakah dokumen atau kumpulan data relevan dengan tujuan yang dimaksudkan.

Precision

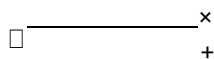
(Presisi) adalah perbandingan antara jumlah data relevan yang ditemukan dan jumlah data yang ditemukan. *Precision* dihitung dengan membagi data benar dengan nilai positif dengan data benar dengan nilai positif dan data salah dengan nilai positif. Nilai data positif palsu untuk diambil dari jumlah nilai yang berbeda dari kolom positif benar yang sesuai dengan setiap kelas. Untuk menghitung nilai *Precision* dapat menggunakan persamaan 2.

$$Presisi = \frac{TP}{TP + FP}$$

Perhitungan lain yang memberikan hasil yang baik adalah *callback*. *Recall* adalah ukuran pentingnya dokumen yang ada untuk hasil tertentu. *Recall* adalah perbandingan dari jumlah materi relevan yang ditemukan terhadap jumlah materi yang relevan. Perhitungan *Recall* dilakukan dengan cara membagi data benar bernilai positif dengan hasil penjumlahan dari data benar yang bernilai positif dan data salah yang bernilai negatif. Nilai dari data salah yang bernilai negatif diambil dari jumlah nilai selain true positive baris yang sesuai tiap kelasnya. Perhitungan *Recall* dapat menggunakan Persamaan 3.

$$Recall = \frac{TP}{TP + FN}$$

F1-score merupakan nilai rata-rata dari penjumlahan presisi dan *Recall*. *F1-score* digunakan untuk menyeimbangkan nilai presisi dan juga *Recall*. Persamaan 4 dapat digunakan untuk menghitung *F1-score*.



Keterangan :

1. TP adalah *True Positive*, yaitu jumlah data yang nilai aktualnya adalah positif
2. TN adalah *True Negative*, yaitu jumlah data yang nilai aktualnya adalah negatif dan terklasifikasi sebagai negatif.
3. FN adalah *False Negative*, yaitu jumlah data yang nilai aktualnya adalah negatif namun terklasifikasi salah.
4. FP adalah *False Positive*, yaitu jumlah data yang nilai aktualnya adalah positif namun terklasifikasi salah.

IV HASIL DAN PEMBAHASAN

4.1 Data Selection

Dataset ulasan dari pelanggan diperoleh dari review pada Situs *Google review*. Mendapati 126 data ulasan dari pelanggan yang sudah sempat mengunjunginya, namun ulasan yang memiliki teks dalam penelitian ini hanya mendapati 126 data dimana didalamnya terdapat ulasan yang memiliki sentiment Positif dan Negatif kemudian diinput dan diberikan label secara manual. Setiap ulasan memiliki tiga pandangan dan penilaian oleh sebab itu ulasan akan dibagi menjadi tiga aspek yaitu ulasan terhadap kopi. Namun perlu diingat bahwa tidak semua ulasan jugamemiliki tiga penilaian atau komentar sekaligus, beberapa ulasan hanya tertuju pada satu aspek saja. Data ulasan yang sudah diinput, dilakukan pembagian aspek dan dilakukan pelabelan kemudian disimpan dalam *Microsoft Excel*. Pada Tabel 1 merupakan hasil dari penginputan data secara

Table 1. Hasil Input

No	Text
1	A floral-driven straight shot, amplified with notes of stone fruit and chocolate in cappuccino-scaled milk. ☑
2	Lavender-like flowers and hints of zesty pink peppercorn animate this crisply sweet wet-processed Ethiopia cup.:D
3	A deeply and generously lush cup saved from sentimentality by a crisp, incense-like, slightly savory edge. Very wide range of aromatics.
4	A lively and crisply sweet, fruit-forward natural-processed Ethiopia cup.
5	A deeply sweet natural-processed Ethiopia cup, with stone fruit and brandy notes as parallel throughlines.:V
6	A delicate, richly sweet wet-processed Ethiopia: spice, herb, fruit and flowers in intricate balance.

4.2 Preprocessing

Padatahapinidigunakanfiturreplaceuntukmenghilangkankarakter,tandabaca,angka dan emoji yang tidak diperlukan

Gambar 5 Data Cleansing



Hasil dari penggunaan fitur replace untuk menghilangkan karakter, tanda baca, angka yang tidak diperlukan dapat dilihat pada Tabel 2.

Table 2. Hasil Replace

<i>Input</i>	<i>Output</i>
A floral-driven STRAIGHT shot, amplified with notes of stone fruit and chocolate in cappuccino-scaled milk. ☺	A floral-driven straight shot amplified with notes of stone fruit and chocolate in cappuccino-scaled milk

2. Labeling Manual

Setelah data terkumpul tahap selanjutnya yaitu dengan melakukan pelabelan manual. Pelabelan dilakukan dengan mengklasifikasikan komentar ke dalam dua kelas, yaitu positif dan negatif. Cara kerja dari pelabelan ini yaitu dengan memahami makna dari kalimat, bukan dilakukan per kata. (Apriliani et al., 2023) Proses pelabelan data berjumlah 126 dibagi menjadi tiga aspek dan dilakukan secara manual dengan kategori sentimen positif dan negatif. (Irvandi, Irawan and Nurdiawan, 2023)

Table 3, Labeling Manual

No	Ulasan	Rating(1-5)	...	Classification quality	Sentiment	Fitur tambahan
1	Evaluated as	5	...	Tall	Positif	long
	espresso. Sweet-toned, deeply rich, chocolaty					review 40 say

2	Crisply sweet, cocoa-toned. Lemon blossom, roasted cacao nib, date, rice candy, white peppercorn in aroma and cup	2	...	Low	Negatif	long review 15 say
3	Delicate, sweetly spiced. Pink peppercorn, date, myrrh, lavender, roasted cacao nib in aroma and cup	3	...	Low	Negatif	long review 10 say
4	Deeply sweet, subtly pungent. Honey, pear, tangerine zest, dark chocolate, pistachio in aroma and cup	5	...	Tall	Positif	long review 8
5	Delicate, richly and sweetly tart. Dried hibiscus, fine musk, almond, sandalwood, raspberry in aroma and cup	2	...	Tall	Positif	long review 6 say
...
122	High-toned, floral. Dried apricot, magnolia, almond butter, maple syrup, cherry brandy in aroma and cup	1	...	Low	Positif	long review 7 say
123	Very delicate, sweetly savory. Lemon verbena,	5	...	Tall	Negatif	long review 12 say

	allspice, dried persimmon, dogwood, baker's chocolate in aroma and cup					
124	Almost no coffee taste	1	...	Tall	Positif	long review 5 say
125	The perfect coffee for the morning.	5	...	Low	Negatif	long review 9 say
126	High-toned, crisply sweet. Lemon blossom, apricot, frankincense, spearmint, roasted cacao nib in aroma and cup	5	...	Tall	Positif	long review 8 say

4.3 Transformation

Pada tahap ini dilakukan *Term Weighting* atau mengimplementasikan metode *TF-IDF*. Pada table 4 menghitung *V-idf* secara manual

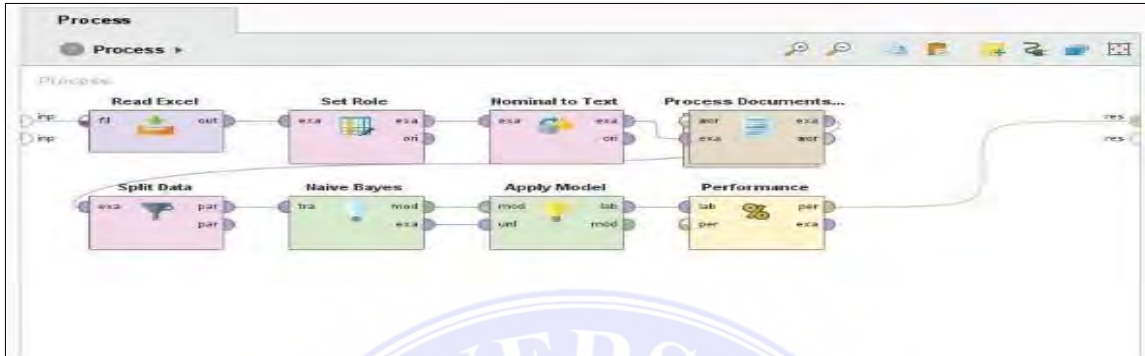
Table 4 Menghitung TF-IDF manual

Jens Kopi	TF			df	N/df	Idf
	d1	d2	d3			
Arabika	1	1	1	3	1	0
Gayo	1	0	1	2	1,5	0,176
Robusta	1	1	0	2	1,5	0,176
Liberika	1	0	0	1	3,0	0,477
esspreso	1	0	0	1	3,0	0,477
Bali kintami	1	0	0	1	3,0	0,477
Sidikalang	0	1	0	1	3,0	0,477
Papua wamena	0	1	1	2	1,5	0,176
Toraja	0	1	0	1	3,0	0,477
Florres bejawa	0	1	0	1	3,0	0,477

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Setiap kopi akan diberikan bobot guna mengubah data ke dalam bentuk numerik. Sehingga diketahui nilai kepentingan dari suatu jenis pada sebuah dokumen berdasarkan intensitas kemunculan jenis kopi

4.4 Evaluation

Model Klasifikasi algoritma c4.5. Analisis pengklasifikasian menggunakan RapidMiner yang operasinya sudah disesuaikan dapat dilihat pada Gambar 8.



Gambar 8. Tahapan Evaluasi

Keterangan

- a. *Read Excel*: melakukan pembacaan data berextension file xlsx yang merupakan dataset hasil pemrosesan data cleansing yang sudah diberikan keterangan sentimen positif dan negatif.
- b. *Set Role*: mengonversikan kolom sentimen menjadi sebagai label pada pemrosesan algoritma Naive Bayes
- c. *Nominal to Text*: mengubah jenis data nominal menjadi text yang akan diteruskan pada operator berikutnya.
- d. *Process Documents from Data*: operator ini mengonversi data text yang digunakan dalam dengan pembobotan TF-IDF, memiliki subproses yang diisi sebagai manata haptext preprocessing.
- e. *Split data*: Split data dilakukan untuk membagi dataset menjadi partisi data latihan dan data uji sesuai porsi yang ditentukan
- f. *Implementasi Naive Bayes*: Menerapkan algoritma yang dipakai
- g. *Apply Model*: Operator apply model digunakan untuk menerapkan model yang telah dilatih sebelumnya menggunakan data training pada data testing.
- h. *Performance*: digunakan untuk mengevaluasi kinerja model yang memberikan daftar nilai kriteria kinerja secara otomatis sesuai dengan tugas yang diberikan.

Table 5 ulasan

Aspek	Sentimen		TN	TP	FN	FP
	P	N				
Riview	374	58	46	241	0	58
Riview	469	40	32	339	0	36
Riview	253	99	74	202	5	0

Berdasarkan penelitian pada pengujian nilai

0.8 dan 0.2 diperoleh ulasan Domo Coffee and restop pada kelas *sentiment* positif dan diklasifikasikan dengan benar (*True Positive / TP*) sebanyak 241 untuk aspek makanan, 339 untuk aspek suasana dan 202 untuk aspek pelayanan. Sedangkan hasil klasifikasi salah (*False Positive / FP*) pada kelas *sentimen* positif adalah sebanyak 58 ulasan untuk aspek makanan, 36 untuk aspek suasana dan 0 untuk aspek pelayanan. Untuk hasil klasifikasi benar pada kelas *sentiment* negatif (*True Negative / TN*) terdiri dari 46 ulasan untuk aspek makanan, 32 untuk aspek suasana dan 74 untuk aspek pelayanan, sedangkan hasil klasifikasi salah pada kelas *sentimen* negatif (*False Negative*) adalah 0 ulasan untuk aspek makanan, 0 untuk aspek suasana dan 5 untuk aspek pelayanan. Dari hasil Confusion Matrix maka dilakukan perhitungan nilai akurasi dengan persamaan 1 sehingga diperoleh akurasi 83,19% untuk makanan, 91,15% untuk suasana dan 98,22% untuk pelayanan

V. KESIMPULAN

5.1 Kesimpulan

Berdasarkan uji yang telah dilakukan, data ulasan pelanggan pada kualitas kopi dapat dilakukan klasifikasi menggunakan algoritma *decision tree* melalui *software* RapidMiner. Dari pemodelan klasifikasi menggunakan algoritma *decision tree* tingkat akurasi mencapai 86.96%, dengan nilai presisi prediksi ulasan enak sebesar 90%, dan nilai presisi prediksi tidak enak sebesar 66.67%. Menunjukkan bahwa ulasan pelanggan Kopi merasa puas menikmati Kopi karena terdapat rasa yang enak.

Dengan tingkat ulasan prediksi positif sebesar 90%, peneliti menyarankan kepada pelanggan Kopi yang belum merasakan Kopi untuk segera mencoba merasakan kopi tersebut.

5.2 Saran

Untuk penelitian mendatang, peneliti dapat meningkatkan jumlah responden kuesioner dan dapat menambahkan variabel-variabel lain yang berpotensi berpengaruh terhadap kualitas kopi

DAFTAR PUSTAKA

- Anggarwati, Nurdiawan, Ali, & Kurnia. (2021). Penerapan Algoritma K-Means Dalam Prediksi Penjualan Karoseri. *Jurnal Data Science & Informatika (Jdsi)*, vol. 1, no. 2, pp. 58–62.
- Badrul, & Pujani. (2021). Analisis Sentimen Ulasan Produk Menggunakan Metode C4.5. *Jurnal Teknologi Informasi*.
- Barbosa, Lemos, & Silva. (2019). Impact of Different Brewing Methods on the Quality of Coffee. *Coffee Science*, 14(321-328).
- Civille, & Decker. (2022). Coffee Quality: A Guide to Grading, Scoring, and Testing. *Food Quality and Preference*, 85.

- Febriani , & Sulistiani. (2021). Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5 . *Jurnal Teknologi Dan Sistem Informasi (JTSI)*, vol. 2, no. 4(pp. 89–95).
- ICO. (2022). Coffee Market Report. Retrieved from. *International Coffee Organization* .
- Klein, & Schilling. (2020). he Effect of Coffee Bean Origin on Quality and Flavor. *Journal of Food Science*, 85.
- Nurul Chasanah, & A. Mutoi Siregar. (2022). Klasifikasi Kelayakan Siswa dalam Menentukan Kelas Unggulan Menggunakan . *Algoritma K-Nearest Neighbor*, vol. III, no. 1, p 51.
- Putri, & Wijayanto. (2022). Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing. *Komputika : Jurnal Sistem Komputer*, vol. 11, no. 1, pp. 59–66. <https://doi.org/10.34010/komputika.v11i1.4350>
- Saputra, Hardinata, & Saragih, A. (2020). Klasifikasi Algoritma C4.5 Dalam Penerapan Tingkat Kepuasan Siswa Terhadap Media Pembelajaran Online . *Ilmiah Informatika Dan Komputer, Kajian*, vol. 1, no. 3,(pp. 113–118). available: <https://djournals.com/klik>
- Saraswati, Umaidah, & Voutama. (2021). Penerapan Algoritma Artificial Neural Network untuk Klasifikasi Opini Publik Terhadap Covid- 19. *Generation Journal*, vol. 5, no., pp. 2580– 4952.
- Supriyadi, Triayudi, & Sholihati. (2021). Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas. *Jipi (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, vol. 6, no. 2, pp. 229–240.
- Widaningsih, & Yusuf. (2022). Penerapan Data Mining Untuk Memprediksi Siswa Berprestasi Dengan Menggunakan Algoritma K Nearest Neighbor. *Jurnal Teknik Informatika Dan Sistem Informasi*, vol. 9, no. 3. <http://jurnal.mdp.ac.id/>
- Winarta, & Kurniawan. (2021). Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python. *Jurnal Teknik Informatika Kaputama (JTIK)*, vol. 5, no. 1.