

SENTIMENT ANALYSIS OF COFFEE SHOP REVIEWS USING RANDOM FOREST CLASSIFIER METHOD

Berliana Nur Isnayni¹, Nurirwan Saputra², Tri Hastono^{3*}

^{1,2,3}Department of Informatics, Universitas PGRI Yogyakarta, Indonesia
*bnisnayni@gmail.com*¹, *nurirwan@upy.ac.id*², *trihastono@upy.ac.id*³

**Corresponding author*

Manuscript Received March 17, 2024; Revised May 8, 2024; Accepted May 8, 2024; Published May 12, 2024

ABSTRACT

A Coffee shop is a place that serves drinks made from processed coffee grains, various drinks and various snacks to accompany coffee to consumers. Coffee shop reviews can help owners to find out how the community responds to the coffee shop and its services. The data used in this study was 2000 data taken on the old Google Maps Kopi Ampirano by scraping data using Instant Data Scraper. From the abundance of review data, it takes a long time to fully understand the polarity of positive, negative, and neutral reviews manually. Because of this, an accurate sentiment analysis model is needed to classify customer reviews into positive, negative, or neutral reviews. In this study, sentiment analysis used coffee shop reviews using the Random Forest Classifier method. The Preprocessing stage involves the process of case folding, tokenization, stopword removal and stemming. The results of this study are coffee shop reviews of the Random Forest Classifier method classification with an accuracy rate of 79% and a Precision value of 81%, Recall of 97% and while the F1 Score of 88%.

Keywords: *Sentiment analysis, coffee shop, random forest*

ABSTRAK

Kedai kopi merupakan sebuah tempat yang menyajikan minuman hasil yang dibuat berasal dari olahan bulir kopi, berbagai aneka minuman dan berbagai snack untuk menemani kopi kepada konsumen. Ulasan kedai kopi dapat membantu pemilik untuk mengetahui bagaimana respon masyarakat terhadap kedai kopi serta pelayanannya tersebut. Data yang digunakan pada penelitian ini sebanyak 2000 data yang diambil pada laman Google Maps Kopi Ampirano dengan cara scrapping data menggunakan Instan Data Sraper. Dari banyaknya data ulasan membutuhkan waktu yang lama untuk sepenuhnya memahami polaritas ulasan positif, negatif, dan netral secara manual. Karena itu, model analisis sentimen yang akurat diperlukan untuk mengklasifikasikan ulasan pelanggan menjadi ulasan positif, negatif, atau netral. Dalam penelitian ini analisis sentimen yang digunakan ulasan kedai kopi menggunakan metode Random Forest Classifier. Pada tahap Preprocessing melibatkan proses case folding, tokenisasi, stopword removal dan stemming. Hasil pada penelitian ini adalah ulasan kedai kopi dari klasifikasi metode Random Forest Classifier dengan tingkat akurasi sebesar 79% dan dari nilai Precision sebesar 81%, Recall sebesar 97% dan sedangkan F1 Score sebesar 88%.

Kata Kunci: *Analisis sentimen, kedai kopi, random forest*

PENDAHULUAN

Kedai kopi merupakan sebuah tempat yang menyajikan minuman hasil yang dibuat berasal dari olahan bulir kopi, berbagai aneka minuman dan berbagai snack untuk menemani kopi kepada pelanggan. Konsumen mungkin sesekali ingin tahu lebih banyak tentang kualitas kopi. Namun demikian, tak seluruhnya kedai kopi yang diminati konsumen mempunyai kualitas yang bagus. Konsumen juga kadang kala mau tahu banyak tentang kualitas kopi, apakah itu dari perspektif segi minuman, makanan atau dari fasilitas dan pelayanan yang ditawarkan sebelum mengunjungi kedai kopi tersebut. (Azhar et al., 2021). Meskipun di zaman modern ini, kedai kopi tak hanya dikenal sebagai tempat berkumpul dan peminum kopi, namun juga sebagai tempat yang nyaman untuk belajar atau bekerja.

Dalam upaya peningkatan kualitas pelayanan, segi hidangan maupun fasilitas di kedai kopi. Beberapa konsumen menilai positif terhadap kedai kopi dari pemantauan di ulasan google maps Kopi Ampirono tahu bahwa masyarakat menyampaikan beragam ulasan mulai dari positif hingga negatif pada kedai kopi. Komentar negatif bisa bermula dari kurangnya pada pelayanan kedai kopi, terdapat 14% ulasan negatif jika dipresentasikan terlihat pada Gambar 1.



Gambar 1. Rating Ulasan Kopi Ampirono
(Sumber : Google Maps, 2023)

Rating yang terdapat pada laman kopi ampirono diatas bahwa rating belum tentu juga menentukan ulasan positif, negatif maupun netral. Ulasan-ulasan berasal dari masyarakat yang dimaksud kemudian dapat dipelajari dan dianalisis menggunakan teknik yang dikenal sebagai analisis sentimen. Data yang diterima melalui ulasan google maps ini menyajikan sumber informasi yang sangat bermanfaat dan mempengaruhi kedai kopi. Perolehan informasi tentang opini publik dengan menganalisis data pada google maps dan secara otomatis mengklasifikasikan polaritas (positif, negatif, netral) dapat dilakukan dengan analisis sentimen (Elbagir & Yang, 2019).

Sentimen analisis merupakan teknik yang dapat digunakan untuk menyederhanakan tahap penentuan penerapan berbagai uji polaritas, seperti positif,

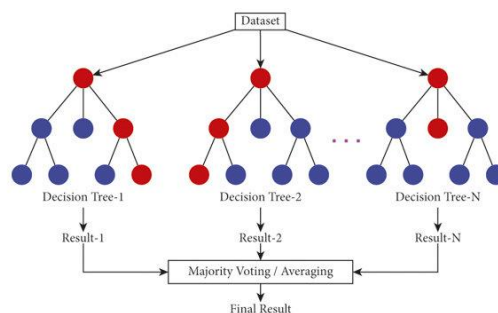
negatif, atau netral. Hanya ada satu teknik yang diperoleh untuk menangani sebuah masalah khusus ini berupa teks Bahasa Indonesia yang diambil dari ulasan google maps memanfaatkan pembelajaran mesin (Text Mining) dan teknik penambangan teks, serta teknik klasifikasi data. Dalam pengetahuan ini, biasanya terkait dengan pendapat masyarakat baik positif maupun negatif. Analisis sentimen sudah mulai terkenal di zaman sekarang, seperti opini terhadap tokoh publik terhadap pemerintah. (Saputra et al., 2022)

Beberapa penelitian tentang analisis sentimen kedai kopi sudah dilakukan dari beberapa peneliti dengan model pendekatan algoritma *merchine learning* seperti, *Random Forest* (Amaliah & Nusrang, 2022) *Naïve Bayes* (Bano et al., 2022) dan *K-Mean* (Habibi & Cahyo, 2019),. Penelitian dengan menggunakan metode *Random Forest* telah banyak dilakukan oleh para peneliti untuk menganalisis ulasan kedai kopi seperti peneliti pada penelitian tersebut menggunakan data komentar di website sebagai data penelitiannya, sedangkan dalam penelitian ini mengutamakan data yang didapatkan atau terkumpulnya dari ulasan google maps. Berdasarkan pada penjelasan diatas bahwa pada riset ini peneliti memilih judul penelitian “Analisis Sentimen Ulasan Kedai Kopi Menggunakan metode *Random Forest Classifier*”.

Tujuan dalam penelitian ini ialah untuk mengetahui tanggapan masyarakat mengenai kedai kopi agar dalam meningkatkan kualitas, baik dari segi hidangan, pelayanan, maupun fasilitas dengan membagi 3 kategorisasi sentimen yaitu positif, negatif dan netral menggunakan metode *Random Forest Classifier*.

METODE PENELITIAN

Metode *Random Forest* adalah algoritma pembelajaran *Ensemble* yang memerlukan struktur pohon dalam langkahnya. Di dalam penggunaannya, Pohon Keputusan dibuat dengan memisahkan atau memilah data sembarang. Untuk memilih sampel suatu data, pada *Random Forest* menggunakan sistem voting berdasarkan hasil *Decision Tree*. (Wiraguna et al., 2019) seperti terlihat pada Gambar 2 dibawah ini.



Gambar 2. Algoritma Random Forest

Meski begitu, Metode *Random Forest* bisa dibuat menggunakan *bagging* dengan penentuan atribut secara acak. Metode Classification and Regresion Tree (CART) sendiri bisa digunakan untuk membuat pohon keputusan, pohon keputusan pada akhirnya akan tumbuh dengan tingkatan maksimum dan tidak digunting kemudian menghasilkan

kumpulan pohon yang kemudian disebut *forest*. Akan halnya kelebihan dari metode *Random Forest* sebagai berikut:

- 1) Tinggi hasil akurasi
- 2) Dalam outliers dan noise korelasinya kuat
- 3) Dibandingkan dengan bagging dan boosting lebih cepat
- 4) Mudah dipahami dan simpel sifatnya

Adapun juga kekurangan dari Metode Random Forest sebagai berikut:

- 1) Pembelajaran bisa berjalan lambat
- 2) Tergantung pada parameter yang digunakan
- 3) Tidak bisa memperbaiki model yang dihasilkan secara berulang

Pengumpulan Data

Data dikumpulkan dari ulasan Google Maps pada laman Kopi Ampirono pada tanggal 11 Januari 2023 yang diambil melalui proses *scraping* menggunakan website *Instan Data Sraper*, seperti terlihat pada Gambar 3.



Gambar 3. Ulasan pada Fanpage Kopi Ampirono

Jumlah data mentah yang dikumpulkan sebanyak 2000 data dan kemudian disimpan dalam tabel dengan format *.xlsx*. Kemudian data dipilah dimana jika ulasan hanya rating bintang dan emoji saja di hapus disimpan dalam tabel dengan format *.csv*. Selanjutnya pada tahap ini sudah berjalan 100%, dalam arti data yang sudah terkumpul sejumlah 2000 ulasan.

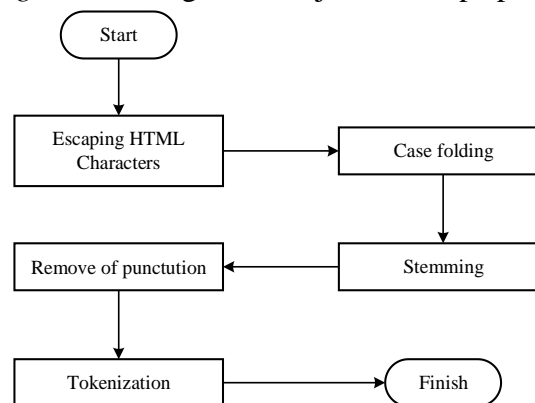
Pelabelan Data

Pelabelan data dilakukan secara manual dengan melibatkan dua orang anatator yang pertama penulis kemudian divalidasi oleh seorang yang ahli dalam bidangnya yang memberikan nilai “positif”, “negatif”, “netral”, terhadap setiap komentar ataupun ulasan. Pemilihan dengan tiga label tersebut dengan memanfaatkan pedoman label yang akan mempermudah proses pelabelan pada data. Data akan diklasifikasikan menjadi tiga *class rating*, yaitu *class 1* dengan ulasan positif (*rating 4 dan 5*), *class 2* dengan ulasan netral

(*rating* 3) dan *class* 3 dengan ulasan negatif (*rating* 1 dan 2) tetapi dalam *rating* tersebut terdapat ulasan yang belum mencerminkan ulasan positif, ulasan negatif dan ulasan netral. Dengan demikian, untuk untuk ulasan dengan sentimen positif (baik), label positif akan diberikan serta untuk ulasan dengan sentimen negatif (buruk) bahkan kalimat negatif, ataupun label negatif akan diberikan sebuah saran. Sebaliknya, istilah netral digunakan untuk subjek yang tidak menunjukkan perasaan negatif melainkan positif.

Preprocessing Data

Tahap *Preprocessing* data merangkum menjadi beberapa proses, sebagai berikut:



Gambar 4. Flowchart Preprocessing

1) *Escapping HTML characters*

Pada awal ekstrasi data (*crawling*) terdapat suatu tautan di ulasan, dalam hal ini proses *Escapping HTML characters* sangat penting karena proses ini untuk menghapus tautan URL dan karakter HTML apapun dalam suatu teks.

2) *Case folding*

Proses menyamaratakan kata di sebuah ulasan, misalnya dalam huruf yang digunakan dalam penulisan adalah huruf kecil (lowercase). Data teks yang berisi huruf besar akan dikonversi ke format huruf kecil. contoh: “Tempatnya Asyik Buat Nongkrong” akan diubah kedalam bentuk “tempatnya asyik buat nongkrong”.

3) *Stemming*

Proses pemilihan kata yang memiliki kata-kata seperti "ganti," "imbuhan," "kerja," "sambung," dan "dasar," sementara mengabaikan "awalan" atau "akhiran." Contoh “Tempat nongkrong terenak bareng temen-temen” diubah menjadi “tempat nongkrong enak bareng temen temen”.

4) *Remove of punctuation*

Proses khusus yang digunakan untuk menghapus karakter seperti (#). Selain itu dilakukan penghapusan terhadap angka numerik (0...9), karakter lainnya(+, \$, ~, dll) dan tanda baca (, . ? !, dll).

5) *Tokenization*

Proses tokenisasi kata dalam kalimat tertentu di mana setiap kata dalam kalimat kedalam bentuk token dan dipisahkan spasi. Stopwords removal ialah proses menghapus kata yang tidak berdampak negatif pada proses klasifikasi, seperti: dan, ke, atau, dari, yang, dll.

Klasifikasi/Pengujian

Model performa untuk melakukan sebuah klasifikasi bisa ditakar dengan menjumlah kelas yang diperkirakan secara pasti (true positive), jumlah perkiraan yang salah prediksi (false positive/false negative atau false netral), dan kelas tersebut dan benar (true negative & true netral).

Tabel 1. Confusion Matrik

		<i>Actual Values</i>		
		<i>Positive</i>	<i>Negative</i>	<i>Netral</i>
<i>Predicted Values</i>	<i>Positive</i>	<i>true positif (tp)</i>	<i>false negative (fn)</i>	<i>false netral (fnet)</i>
	<i>Negative</i>	<i>false positive (fp)</i>	<i>true negative (tn)</i>	<i>false netral (fnet)</i>
	<i>Netral</i>	<i>False positive (fp)</i>	<i>False negative (fn)</i>	<i>true netral (tnet)</i>

Metrik khusus ini dapat diturunkan menggunakan *confusion matrix*

$$\text{Accuracy} = \frac{TP + TN = TNET}{TP + FP + TN + FN + TNET + FNET}$$

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

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{F1 score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Accuracy metrik, precision, recall dan F1 digunakan untuk meningkatkan kinerja model saat membuat prediksi dalam kelas positif.

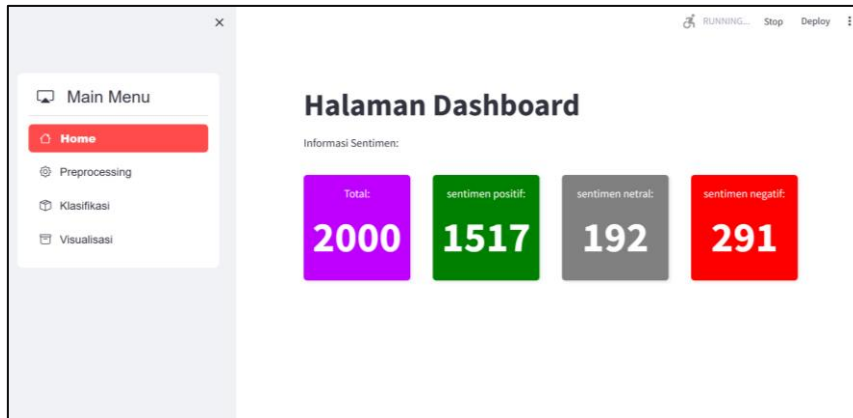
HASIL DAN PEMBAHASAN

Implementasi

Penelitian yang diselesaikan peneliti adalah membuat sistem analisis sentimen kedai kopi dengan baik sebagaimana kegunaanya. Tampilan sistem dapat dilihat sebagai berikut:

Halaman Dashbord

Jumlah data sentimen positif ada 1517, data sentimen negatif ada 291, dan data sentimen netral ada 192.



Gambar 5. Halaman Dashbord

Halaman Preprocessing

Bagian ini terdiri dari berbagai tahapan karena data ulasan tidak memanfaatkan tata bahasa yang tepat. Langkah preprocessing dilakukan menggunakan library untuk bahasa pemrograman Python. Dalam penelitian ini, langkah preprocessing data dilakukan dengan melakukan empat tahapan langsung, yaitu sebagai berikut:

1) Text Cleaning

Tujuan langkah pembersihan teks (text cleaning) ini adalah untuk mengekstrak dataset ulasan dari beberapa hal yang tidak penting, seperti teks yang akan ditulis, normalisasi unicode, dan sebagainya. Ada empat langkah yang terlibat dalam proses pembersihan ini untuk mencapai hasil maksimal ialah membersihkan tanda baca, membersihkan angka, membersihkan kelebihan spasi, dan membuat huruf lebih kecil secara keseluruhan.

No	sentimen	Tweet
1	1	makanannya enak dan fasilitas lengkap
2	1	tempatny bagus murah meriah dan nyaman benerbener bisa foto sambil ngopi buki
3	1	lokasi tidak begitu jauh dari kota tempatnya nyaman terutama yg diatas harga camli
4	2	kasirnya yg cewe ngrgas bgt ajg pdhl tnya baik klo ganiat nglayani gsh kerja dongzzzz
5	1	tempatny bagus makanannya enak murah jooss
6	1	tempat istimewa buat orang istimewa
7	1	selalu ramai tapi tetep nyaman terutama duduk di bagian atas untuk cemilan sih blaz
8	1	viewnya mantep ditambah makanan ala desa harga terjangkau
9	1	pemandangannya bagus parkirannya luas toilet bersih makanannya relatif murah dan er
10	0	pelayanan kurang baik menu kurang lengkap tempat duduk minta di tambahkan su

Gambar 6. Hasil Text Cleaning

2) Tokenisasi (Unigram)

Tokenisasi adalah proses untuk memecahkan sebuah kalimat kedalam kepingan-kepingan kata-kata.

No	sentimen	Tweet_token
1	1	makanannya enak dan fasilitas lengkap
2	1	tempatya bagus murah meriah dan nyaman benerbener bisa
3	1	lokasi tidak begitu jauh dari kota tempatnya nyaman terata
4	2	kasirnya yg cewe ngegas bgt ag pdhl tnya baik klo gar
5	1	tempatya bagus makanannya enak murah jooss
6	1	tempat istimewa buat orang istimewa
7	1	selalu ramai tapi tetep nyaman terutama duduk di bagian
8	1	viewnya mantep ditambah makanan ala desa harga terjangkau
9	1	pemandangannya bagus parkirannya luas toilet bersih makanannya
10	0	pelayanan kurang baik menu kurang lengkap tempat duduk

Gambar 7. Hasil Tokenisasi (Unigram)

3) Remove Stopword

Proses pembuatan stopwords dari data dilakukan dengan mendefinisikan kata-kata yang muncul lebih sering daripada tidak. Dalam hal ini, setiap daftar stopwords disimpan dalam file dengan nama stopwords.

No	sentimen	Tweet_no_stopwords
1	1	makananakfasilitaslengkap
2	1	tempatbagusmurahriahnyamanbenerbenerfotongopipurapurangopiselfiedoang
3	1	lokasikotatempatnyamanygatahargacamilmurah5rbananakmendoanjuaramaafay
4	2	kasirygcwengegasbgtajgpdhltnyaklogianatnglayanigshkerjadonggggg
5	1	tempatbagusmakanenakmurahjooss
6	1	istimewaorangistimewa
7	1	ramaltetepnyamandudukcemilansihpisangmendoansmgeblekenakitu
8	1	viewnyamantepditambahmakanaladesahargajangkau
9	1	pandangbagusparkirluastolletbersihmakanrelatifmurahenakterkadangyggeblekker
10	0	layanmenulengkapdudukdambahtunggu

Gambar 8. Hasil Remove Stopword

4) Stemming

Stemming adalah proses memecah infleksi kata, tetapi bentuk kata berbeda dari akar kata.



Hasil Setelah Stemming

No	sentimen	Tweet
1	1	makan enak fasilitas lengkap
2	1	tempat bagus murah/rah nyaman benerbener foto ngopi pi
3	1	lokasi kota tempat nyaman yg atas harga camil murah di
4	2	kasir yg cewe nggapi bgt agi pdhl brye klo gantel ngli
5	1	tempat bagus makan enak murah joss
6	1	istimewa orang istimewa
7	1	ramai tetep nyaman duduk cemilan sih piang mendoan sm
8	1	viewnya mantep ditambah makan ala desa harga terjangkau
9	1	pandang bagus parkir luas toilet bersih makan relatif murah
10	9	layan menu lengkap duduk tambah tunggu

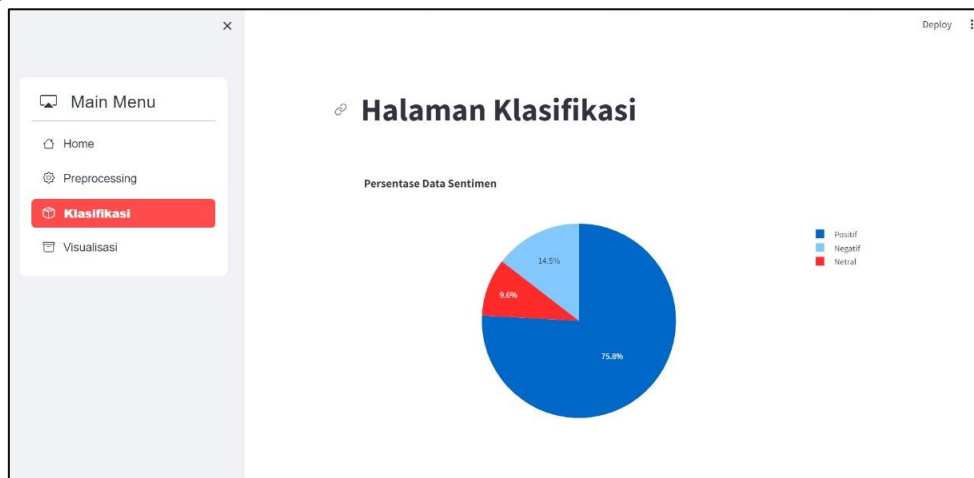
Hasil Setelah Unlisting:

No	sentimen	Tweet
1	1	makan_enak,fasilitas_lengkap
2	1	tempat_bagus,murah/rah_nyaman,benerbener,foto,ngopi,pi
3	1	lokasi_kota,tempat_nyaman,yg_atas,harga_camil,murah,di
4	2	kasir_yg_cewe,nggapi_bgt,agi_pdh,brye_klo_gantel,ngli
5	1	tempat_bagus,makan_enak,murah,joss
6	1	istimewa_orang,istimewa
7	1	ramai,tetep_nyaman,duduk,cemilan,sih,piang,mendoan,sm
8	1	viewnya_mantep,ditambah,makan,ala_desa,harga_terjangkau
9	1	pandang_bagus,parir,luas,toilet_bersih,makan,relatif_murah
10	9	layan_menu,lengkap,duduk,tambah,tunggu

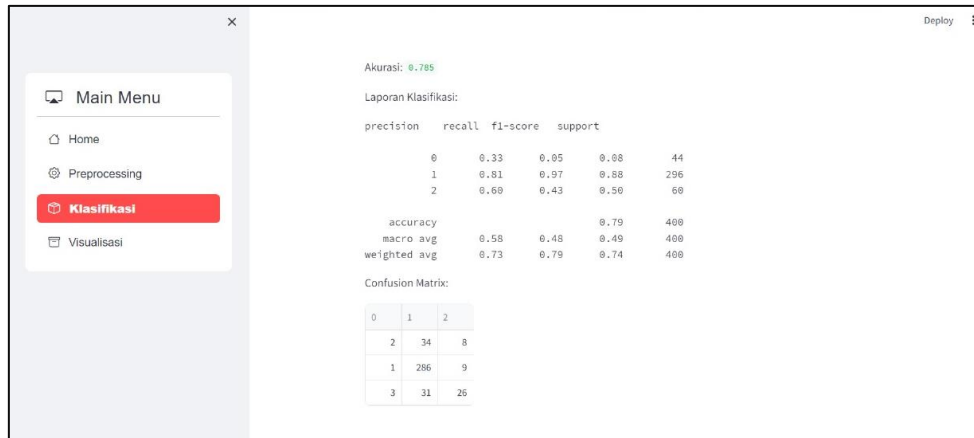
Gambar 9. Hasil Stemming

Halaman Klasifikasi

Halaman Klasifikasi merupakan halaman dimana terdapat hasil presentase data sentimen serta confusion matrik dengan hasil tingkat akurasi dengan model *random forest classifier* adalah sebesar 0,79.



Gambar 10. Halaman Klasifikasi



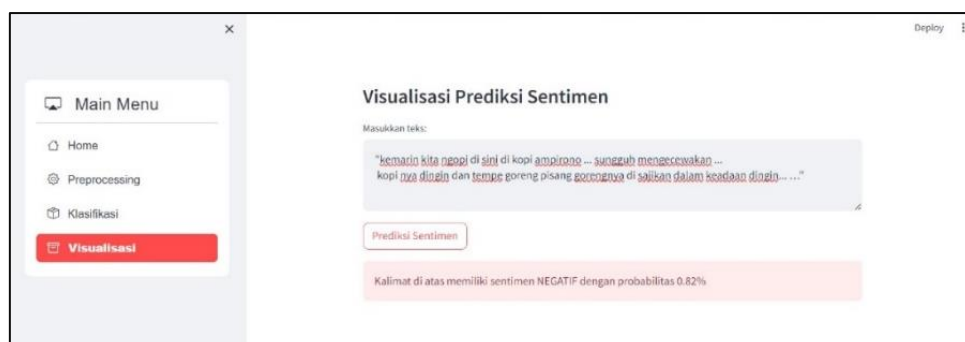
Gambar 11. Confusion Matrix

Halaman Visualisasi

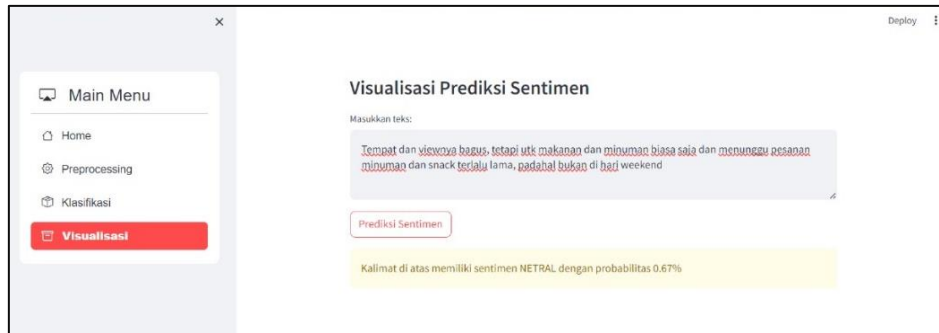
Halaman Visualisasi merupakan hasil prediksi sentimen dengan memberikan sebuah ulasan untuk menentukan bahwa ulasan tersebut termasuk ulasan positif, ulasan negatif maupun ulasan netral.



Gambar 12. Hasil Analisis Positif



Gambar 13. Hasil Analisis Negatif



Gambar 4.1 Hasil Analisis Netral

Pembahasan

Berdasarkan implementasi diatas berikut ini adalah kelebihan dan kekurangan sistem, yaitu:

Kelebihan Sistem:

- 1) Memiliki akurasi sekitar 79%, presisi 81%, recall 97%, dan Skor F1 88%, sehingga cocok untuk digunakan dalam sistem apa pun.
- 2) Metode *Random Forest Classifier* yang digunakan dalam penelitian ini didasarkan pada algoritma yang sangat baik, terlihat dari hasil yang menunjukkan akurasi 79%, presisi 81%, recall 97%, dan skor F1 88%.

Kekurangan Sistem:

- 1) Sistem masih memiliki nilai ketidakakuratan sebesar 21%.
- 2) Dalam proses stemming lama hasilnya karena jika semakin banyak data maka semakin lama juga runningnya dan tergantung koneksi internet.

KESIMPULAN

Penelitian ini menggunakan metode *Random Forest Classifier* untuk mengklasifikasikan komentar pada dataset ulasan kedai kopi di Kopi Ampirano. Penelitian ini mengumpulkan data melalui scraping dan preprocessing sebanyak 2000 ulasan. Klasifikasi yang paling umum adalah positif. Hasil penelitian menemukan bahwa metode *Random Forest Classifier* memiliki nilai akurasi, presisi, recall, dan skor F1 yang baik. Akurasi keseluruhan classifier adalah 79%, dengan nilai persentase tertinggi diperoleh untuk label negatif, yaitu presisi 81%, recall 97%, dan skor f1 88%.

DAFTAR PUSTAKA

- Amaliah, S., & Nusrang, M. (2022). *Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi Di Kedai Kopi Konijiwa Bantaeng*. 4(2), 121–127. <https://doi.org/10.35580/variainsium31>
- Azhar, N., Adikara, P. P., Adinugroho, S., Studi, P., Informatika, T., Komputer, F. I., Brawijaya, U., & Korespondensi, P. (2021). *SENTIMENT ANALYSIS FOR COFFEE SHOP REVIEWS USING NAÏVE BAYES*. 8(3), 609–618. <https://doi.org/10.25126/jtiik.202184436>

- Bano, N. J., Sukwadi, R., & Park, A. (2022). Analisis Perbaikan Kualitas Layanan Bluemoon Container Café: Model Integrasi Analisis Sentimen dan Quality Function Deployment. *Jurnal INTECH Teknik Industri Universitas Serang Raya*, 8(1), 75–82. <https://doi.org/10.30656/intech.v8i1.4569>
- Elbagir, S., & Yang, J. (2019). Language Toolkit and VADER Sentiment. *Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS)*, 0958, 12–16.
- Habibi, M., & Cahyo, P. W. (2019). Clustering User Characteristics Based on the influence of Hashtags on the Instagram Platform. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(4), 399. <https://doi.org/10.22146/ijccs.50574>
- Saputra, N., Nurbagja, K., & Turiyan, T. (2022). Sentiment Analysis of Presidential Candidates Anies Baswedan and Ganjar Pranowo Using Naïve Bayes Method. *JURNAL SISFOTEK GLOBAL*, 12(2), 114. <https://doi.org/10.38101/sisfotek.v12i2.552>
- Wiraguna, A., Faraby, S. Al, & Adiwijaya. (2019). Klasifikasi Topik Multi Label pada Hadis Bukhari dalam Terjemahan Bahasa Indonesia Menggunakan Random Forest. *E-Proceeding of Engineering*, 6(1), 2144–2153. <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/8507%0Ahttps://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/download/8507/8377>