

**ANALISIS SENTIMEN ULASAN APLIKASI ONLINE TRAVEL AGENT (OTA) DI
GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA K-NEAREST
NEIGHBOUR**

SENTIMENT ANALYSIS OF ONLINE APP REVIEWS
TRAVEL AGENT (OTA) ON GOOGLE PLAY STORE
USING THE K-NEAREST NEIGHBOUR ALGORITHM

Hastri Cantya Danahiswari¹, Eka Dyar Wahyuni², Amalia Anjani Arifiyanti³

Sistem Informasi, Ilmu Komputer
Universitas Pembangunan Nasional “Veteran” Jawa Timur, Surabaya, Indonesia

ABSTRAK

Pada saat sekarang ini, Profil tiket.com dalam Google Play Store hingga akhir 2022 aplikasi Tiket.com telah digunakan lebih dari 10 juta pengguna dan mendapatkan rating 4.3 di Google PlayStore. Begitu juga dengan kompetitornya, Traveloka hingga akhir 2023 telah digunakan lebih dari 50 juta pengguna dan mendapatkan rating 4.8. Data tersebut menunjukkan aplikasi OTA yang merupakan aplikasi yang diminati. Akan tetapi, suatu aplikasi selalu memiliki kekurangan dan kelebihan masing masing yang dapat menimbulkan berbagai respon dari pengguna aplikasi seperti kepuasan dan kekecewaan terhadap aplikasi tersebut. Kolom ulasan komentar terhadap sebuah aplikasi menjadi salah satu tempat untuk melontarkan kepuasan dan kekecewaan pengguna atau opini terhadap aplikasi tersebut. Hal tersebut dapat dijadikan bahan analisis sentimen untuk dapat mempermudah, membantu serta memberikan informasi tentang emosi yang terkandung dalam ulasan tentang aplikasi OTA yang ada pada situs Google Play Store

Kata kunci: Analisis sentimen, OTA, K-Nearest Neighbour

ABSTRACT

Currently, the tiket.com profile in the Google Play Store until the end of 2022, the Tiket.com application has been used by more than 10 million users and has a rating of 4.3 on the Google PlayStore. Likewise with its competitor, Traveloka, by the end of 2023, has been used by more than 50 million users and received a rating of 4.8. This data shows that OTA applications are applications that are in demand. However, an application always has its own advantages and disadvantages which can give rise to various responses from application users such as satisfaction and disappointment with the application. The comments review column for an application is a place to express user satisfaction and disappointment or opinions about the application. This can be used as material for sentiment analysis to make things easier, helpful and provide information about the emotions contained in reviews of OTA applications on the Google Play Store site.

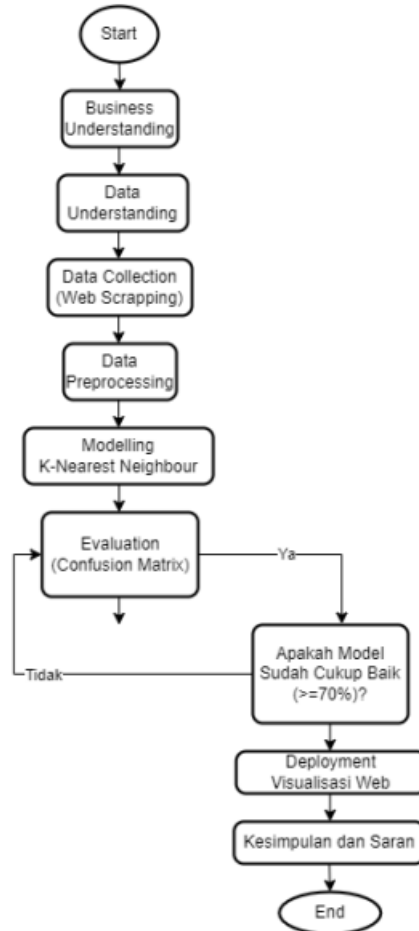
Keywords: Sentiment Analysis, OTA, K-Nearest Neighbour

1. PENDAHULUAN

Di era digital saat ini, berbagai macam aplikasi Online Travel Agent (OTA) di Google Play Store digunakan oleh masyarakat Indonesia untuk berbagai macam keperluan seperti berlibur atau travelling. Masing – masing aplikasi Online Travel Agent (OTA) memiliki produk dan fitur tersendiri yang membedakan antar aplikasi. Adapun, aplikasi OTA terpopuler di Indonesia, yaitu Traveloka dan Tiket.Com berdasarkan jumlah data pengunjung aplikasi tersebut. Agar konsumen mendapatkan pelayanan terbaik terhadap performa aplikasinya maka, penting untuk meningkatkan pelayanan kepada pengguna saat memesan tiket pada aplikasi tersebut dengan fokus pada kebutuhan dan emosi pengguna ketika menggunakan aplikasi tersebut sebagai salah satu kunci keberhasilan sebuah aplikasi (Maharani, 2022). Selama ini sebuah predikat untuk aplikasi terbaik diberikan kepada aplikasi dengan jumlah unduhan dan pemberian bintang paling tinggi di Play Store (Erfina et al., 2020). Berdasarkan profil tiket.com dalam Google Play Store bahwa hingga akhir 2022 aplikasi tiket.com telah digunakan oleh lebih dari 10 juta pengguna dan mendapatkan rating 4.3 di Google Play Store. Begitu juga dengan kompetitornya, Traveloka hingga akhir 2023 telah digunakan lebih dari 50 juta pengguna dan mendapatkan rating 4.8. Data tersebut menunjukkan Tiket.com dan Traveloka merupakan aplikasi OTA yang cukup diminati. Akan tetapi, suatu aplikasi selalu memiliki kekurangan dan kelebihan masing-masing yang dapat menimbulkan 11 berbagai respon dari pengguna aplikasi seperti kepuasan dan kekecewaan terhadap aplikasi tersebut (Giovani et al., 2020). Kolom ulasan komentar terhadap sebuah aplikasi di Google Play menjadi salah satu tempat untuk melontarkan kepuasan dan kekecewaan pengguna atau opini terhadap aplikasi tersebut. Hal tersebut dapat dijadikan bahan analisis sentimen terhadap aplikasi tiket.com untuk mengetahui apakah jumlah unduhan dan rating aplikasi tiket.com relevan untuk membuktikan bahwa kolom ulasan komentar merupakan tempat para pengguna aplikasi dalam melontarkan kepuasan dan kekecewaan.

2. METODOLOGI

Metode penelitian ini menggunakan tahapan:



1. Business Understanding
2. Data Understanding
3. Data Collection
4. Data Preprocessing
5. Modelling K-Nearest Neighbour
6. Evaluation (Confusion Matrix)
7. Deployment Visualisasi Web
8. Kesimpulan dan Saran

Web Scrapping

Pada tahap web scrapping bertujuan untuk mengambil data ulasan aplikasi Tiket.com di laman Google Play Store. Pada laman ini akan mengambil beberapa 34 data, seperti UserName, Content, Score, reviewCreatedVersion, at (timestamp Review user), repliedAt (timestamp review developer), replyContent, appVersion. Data yang akan discrape, yaitu dua studi kasus OTA (Tiket.com dan Traveloka) dengan masing – masing studi kasus berjumlah 1000 ulasan komentar. Dimana ulasan aplikasi

Tiket.com dengan rentang waktu 8 Februari 2023 – 10 Mei 2023 dan aplikasi Traveloka dengan rentang waktu 23 April 2023 – 10 Mei 2023. Komentar yang akan diambil merupakan komentar Bahasa Indonesia dan pengguna dari Indonesia. Adapun, cara web scraping akan menggunakan tools Google Colab dengan Bahasa pemrograman Python dengan memanfaatkan link gits dari Google Play Store. Data yang telah di scrape akan di generate menjadi CSV (Comma Separated Value) yang dapat dibuka pada Microsoft Excel. Pada tahap Preprocessing adalah proses perubahan bentuk data yang terstruktur sembarang menjadi data yang terstruktur sesuai kebutuhan untuk proses dalam text mining (Takdirillah, 2020). Tahapan Preprocessing yang digunakan antara lain:



1. Merge CSV

Merge CSV merupakan tahapan dalam penggabungan csv ulasan aplikasi tiket.com dan Traveloka.

2. Filtering Nama Aplikasi Filtering ini dilakukan untuk menghilangkan nama aplikasi yang dapat bersifat ambigu, yaitu kata “Tiket” pada nama aplikasi Tiket.com.

3. Transformation

a. Case Folding yaitu tahap mengubah semua huruf dalam dokumen menjadi huruf kecil dan karakter non-huruf akan dihilangkan.

4. Tokenizing Yang merupakan tahap pemotongan string input berdasarkan tiap kata yang menyusunnya (Afuan, 2018) dan secara garis besar memecah sekumpulan karakter dalam suatu teks ke dalam satuan kata.

5. Filtering

a. Stopword yaitu dengan mengambil kata-kata penting dari hasil tahap tokenizing yang dapat dilakukan dengan menghilangkan stoplist atau stopwords (Sugianto & Apandi, 2018).

6. Stemming Stemming adalah proses menyatukan atau mengubah varian bentuk kata dari umum menjadi kata dasar (Kannan and Gurusamy, 2015). Ada banyak macam stemming berdasarkan bahasa, mulai dari bahasa Inggris sampai bahasa Indonesia.

Analisis Sentimen (Vader)

Dataset yang akan digunakan pada model Vader untuk analisis sentimen didapatkan dari pengumpulan dataset dengan melakukan web scraping oleh penulis. Dataset akan melewati tahap preprocessing terlebih dahulu karena, tahap ini penting sebelum diklasifikasikan data harus benar-benar bersih dan terstruktur. Vader berguna untuk labelling analisis sentiment dari suatu ulasan seperti, positif, negatif, dan netral.

Analisis Sentimen (Roberta)

Dataset yang akan digunakan pada model Roberta untuk analisis sentimen didapatkan dari pengumpulan dataset dengan melakukan web scraping oleh penulis. Dataset akan melewati tahap preprocessing terlebih dahulu karena, tahap ini penting sebelum diklasifikasikan data harus benar-benar bersih dan terstruktur. Vader berguna untuk labelling analisis sentiment dari suatu ulasan seperti, positif dan negatif.

Klasifikasi K-Nearest Neighbour

Klasifikasi data pada penelitian ini menggunakan metode K-Nearest Neighbour. Metode K-Nearest Neighbor (KNN), merupakan salah satu metode klasifikasi non parametrik. Prinsip operasinya adalah mengklasifikasikan objek pada tes set berdasarkan mayoritas kelas dari k tetangga terdekat pada training set. Nilai K adalah bilangan bulat positif yang ditentukan oleh pengguna. Perhatikan bahwa semakin kecil nilai k, semakin sensitif terhadap efek noise/runway, sedangkan semakin besar nilai k, maka semakin kecil efek noise/runway, tetapi batas antar kelas menjadi kabur.

Evaluasi Validasi (Confusion Matrix)

Tahapan yang terakhir dalam pembangunan model adalah evaluasi model klasifikasi dengan menggunakan confusion matrix. Evaluasi menggunakan confusion matrix untuk mengetahui perkiraan yang benar dan salah dari hasil klasifikasi. Evaluasi digunakan untuk mengetahui seberapa akurat model yang sudah dibuat. Hasil akurasi tersebut didapatkan dengan membandingkan data yang sudah di beri label benar atau salah pada model dengan hasil prediksi label. Tahap pengujian dilakukan dengan menggunakan data uji. Evaluasi model klasifikasi dianalisis dengan membandingkan nilai akurasi secara keseluruhan dan waktu yang diperlukan untuk melakukan training model.

Visualisasi (Web)

Setelah hasil dari tingkat akurasi kedua metode klasifikasi didapat maka langkah selanjutnya yaitu visualisasi. Proses visualisasi disini berarti proses pengkodean dan pembuatan tampilan implementasi antar muka agar para pengguna dapat memahami hasil dari klasifikasi ini. Hasil dari klasifikasi data dengan kedua 38 metode tersebut

akan divisualisasikan ke dalam bentuk grafik yang mudah dipahami dan dari klasifikasi juga ditampilkan kata yang menjadi pemicu sentimen positif, negative, dan netral dalam beropini yang akan diimplementasikan dalam suatu aplikasi berbasis web menggunakan Bahasa PHP. Fitur dari visualisasi yaitu website bisa memprediksi label dari ulasan yang dimasukkan ke dalam inputan dan menampilkan grafik batang dengan wordcloud dengan hasil tabel sentimen ulasan.

3. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Berikut ini merupakan contoh ulasan aplikasi Tiket.com yang diperoleh:

Username	Content	Score	Review Created Version	at	Reply Content	Replied At	appVersion
frenky Chandra	Aplikasi yg pusing, susah di pakai nya	1	4.60.0	2023-07-03 02:09:20	Hai Frenky Chandra, Terima kasih ya sudah luangkan waktumu berikan ulasan untuk kami. Jika kamu ada alami kendala saat pesan atau bingung dengan apps kami jangan sungkan hubungi customer care kami ya sobat. Kami siap bantu kamu 24/7 atau kamu bisa kirimkan email ke cs@ dengan subject : "Konfirmasi Review App".	2023-07-03 08:01:33	4.60.0

					Terima kasih ðŸ˜Š by : FA		
--	--	--	--	--	------------------------------	--	--

B. Merge CSV Ulasan Aplikasi Traveloka dan Tiket.com

Setelah Scrapping didapatkan data berupa file csv untuk aplikasi Traveloka dengan jumlah 7000 data dengan kurun waktu (November 2021 – Juli 2023) dan ulasan aplikasi Traveloka sejumlah 7000 dengan kurun waktu (Februari 2023 - Juli 2023).

C. Filtering Kata Traveloka dan Tiket.com

Filtering ini dilakukan untuk memfilter kata “Traveloka”, “Traveloka”, “TRAVELOKA”, “Tiket.com”, “tiket.Com”, “tiket.com”, “TIKET.COM”, “Tiketcom”, “tiketcom” agar pada ulasan fokus pada kalimat sentimen dan tidak menimbulkan ambigu terutama pada kata “Tiket” karena nama aplikasinya “Tiket.com”

D. Cleansing

Pada proses cleansing, terdapat beberapa langkah-langkah yang akan dijelaskan berdasarkan source code dan hasilnya. Langkah-langkah tersebut meliputi:

- Menghilangkan URL situs lain atau URL gambar dan video Untuk menghilangkan URL situs lain, gambar, dan video dari teks.
- Menghilangkan paragraf baru.
- Menghilangkan tanda baca dan angka.
- Menghilangkan spasi di depan dan di belakang untuk kalimat yang repetisi.

50

- Menghilangkan tab pada kalimat.
- Mengubah huruf pada kalimat menjadi huruf kecil.

E. Case Folding

Mengubah semua karakter teks menjadi huruf kecil. Hal ini dilakukan untuk mengurangi kompleksitas dan perbedaan kapitalisasi teks.

F. Tokenization

Sebuah proses pemisahan teks menjadi kata, frasa, simbol menjadi elemen yang berbeda biasanya proses pemisahan setiap kata dipisahkan dengan tanda koma (,).

G. Stemming

Proses menyatukan atau mengubah varian bentuk kata dari umum menjadi kata dasar.

H. Labelling Menggunakan Library Roberta

Model ini awalnya merupakan model RoBERTa Base Indonesia yang telah dilatih sebelumnya, yang kemudian disempurnakan berdasarkan kumpulan data Analisis Sentimen Indolem yang terdiri dari tweet berbahasa Indonesia dan ulasan hotel (Koto et al., 2020).

No.	FINAL FIX	Sentiment	Confidence	Label
1.	sangat bantu diskon harga pantastis	1	0.999755	Positive
2.	sangat kecewa beli tiket dgn paylater reschedule...	0	0.999918	Negative
3.	travel yg tanggung jawab walaupun yg lbh murah...	1	0.987701	Positive
4.	trimakaasihh smg mjd bahagia smua umat amin	1	0.999656	Positive
5.	limit paylater kok turun kecewaaa	0	0.999933	Negative

I. Labelling Menggunakan Library Vader

Library Vader merupakan library yang dibuat untuk memberi label pada data berbahasa Inggris. Pada data penelitian ini menggunakan data berbahasa Indonesia. Maka dari itu, data perlu di translate terlebih dahulu sebelum di labelling. Disini menggunakan library translator

No.	Sebelum Translate	Setelah Translate
1.	sangat kecewa beli tiket dgn paylater reschedule tglnya dr juli juli ibu sakit keras ecommend bel paylater lg kena tagih dobel tiket yg baru ga masalah bayar yg sdh tiket khan sdh rubah sdngkan biaya reschedule rb orang tp say tetap harus bayar paylater yg sdh rubah jadwal khan aneh jadi asa beli tiket jt kalu sya byar dobel sejak malam sampe skrg no tanggap	very disappointed to buy a ticket with a rescheduled paylater date from July to July, the mother is seriously ill in Indonesia, the paylater bell is subject to being billed for double the new ticket, no problem paying for those who have already changed the ticket, while the rescheduling fee is thousands of people but I still have to pay the paylater, who has changed the schedule, it's strange so I hope to buy a ticket if I pay double since the evening until now no response
2.	sangat bantu diskon harga pantastis	very helpful discount reasonable prices
3.	lama lama jelek bgt aplikasi kmrn beli tiket pesawat pke aplikasi tulis blm pesan aktif skrg ecomme tiket dapat masalah koneksi ga ecommended lama lama aplikasi	it took a long time, the application was really bad, I bought plane tickets using the application, I haven't ordered yet, now I want cents for tickets, I got connection problems, it doesn't take long to recommend the application

No.	Translated Text	Sentimen
1.	very disappointed to buy a ticket with a rescheduled paylater date from July to July, the mother is seriously ill in Indonesia, the paylater bell is subject to being billed for double the new ticket, no problem paying for those who have already changed the ticket, while the rescheduling fee is thousands of people but I still have to pay the paylater, who has changed the schedule, it's strange so I hope to buy a ticket if I pay double since the evening until now no response	Negative
2.	very helpful discount reasonable prices	Positive
3.	it took a long time, the application was really bad, I bought plane tickets using the application, I haven't ordered yet, now I want cents for tickets, I got	Negative

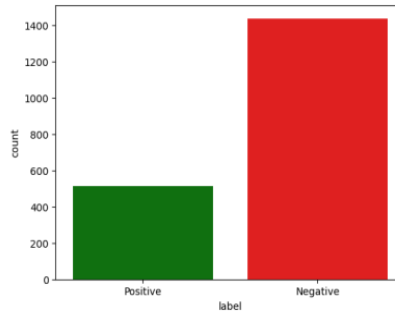
J. Perbandingan Labelling

Perbandingan labelling dilakukan untuk validasi bahwa hasil labelling vader maupun roberta tidak jauh berbeda

aplikasi tipu kedok inestasi emas	Negative	Positive	Negative
selalu beri mudah utk cari kamar hotel pergi makas	Positive	Positive	Positive
limit gak pakai terus dpt notif terus biar pake limit kan aneh refundnya lamaaaaaaaaaaaaaaa banget	Negative	Negative	Negative
bagus pokok lancar jaya sukses trus	Positive	Positive	Positive
mudah sesuai expetasi kamar layan	Negative	Positive	Positive
kalau tambah tiket kapal laut pak	Positive	Negative	Positive
harga nya lebih murah dr aplikasi yg metode beli gak ribet	Positive	Positive	Positive
no hp blokir kok gk ganti verifikasi lewat email aja sungguh aneh	Negative	Negative	Negative
puas dgn aplikasi mudah mesan hotel jalan keluar kota selalu dapat hotel baik	Positive	Positive	Positive
cepat proses nya terima kasih	Negative	Positive	Positive

K. EDA (Exploratory Data Analysis)

Pada grafik tersebut dapat terlihat bahwa data ulasan yang telah melewati preprocessing data mengalami imbalanced dataset (tidak seimbang). Maka dari itu, perlu dilakukan penyeimbangan data seperti, undersampling dan oversampling.



L. Pemodelan Dengan Algoritma K-Nearest Neighbour

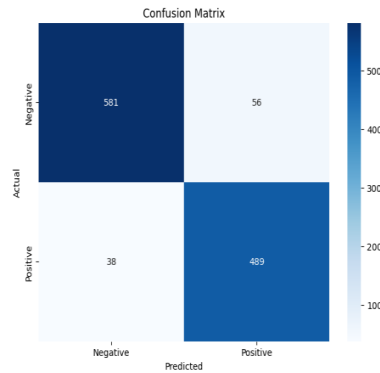
Pada metode ini, dilakukan pengujian pemodelan KNN nilai k=1 hingga k=25 dengan menggunakan skenario data mentah, data labelling Vader, dan data labelling Roberta. Berikut merupakan hasil pengujian nilai akurasi dengan algoritma KNN

No.	Nilai K	Roberta Under Sampling	Vader Under Sampling	Data Mentah (Imbalanced)	Roberta Over Sampling	Vader Over Sampling
1.	1	0.64	0.78	0.77	0.97	0.77
2.	2	0.60	0.65	0.74	0.87	0.74
3.	3	0.50	0.63	0.75	0.94	0.75
4.	4	0.49	0.74	0.75	0.91	0.73
5.	5	0.49	0.71	0.75	0.82	0.72
6.	6	0.49	0.75	0.74	0.92	0.71
7.	7	0.48	0.70	0.74	0.72	0.71
8.	8	0.69	0.75	0.74	0.72	0.71
9.	9	0.65	0.72	0.74	0.49	0.71
10.	10	0.74	0.76	0.74	0.80	0.71
11.	11	0.72	0.74	0.74	0.79	0.71
12.	12	0.75	0.78	0.74	0.80	0.71
13.	13	0.74	0.76	0.74	0.76	0.71
14.	14	0.77	0.76	0.75	0.79	0.71
15.	15	0.75	0.76	0.74	0.76	0.73
16.	16	0.77	0.77	0.75	0.81	0.72
17.	17	0.77	0.76	0.74	0.79	0.74
18.	18	0.79	0.75	0.65	0.80	0.73
19.	19	0.77	0.75	0.77	0.78	0.74
20.	20	0.78	0.75	0.73	0.80	0.74
21.	21	0.78	0.75	0.74	0.78	0.76
22.	22	0.79	0.77	0.75	0.80	0.75
23.	23	0.79	0.75	0.74	0.80	0.77
24.	24	0.79	0.76	0.74	0.82	0.76
25.	25	0.80	0.75	0.78	0.80	0.78

J. Confusion Matrix

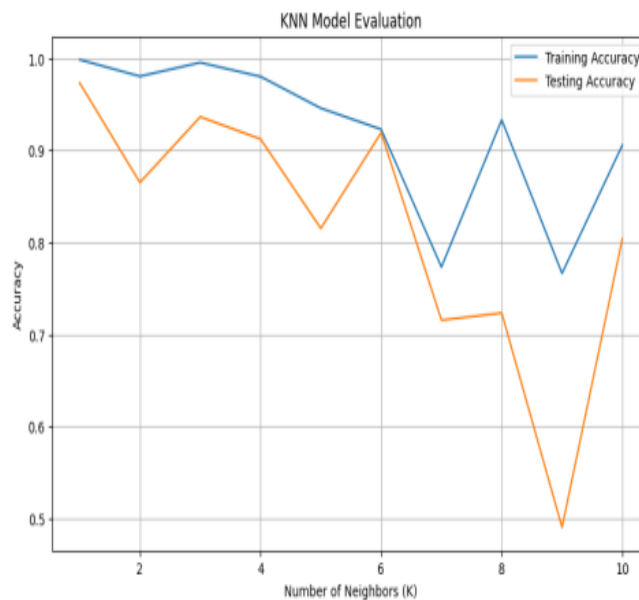
Pada penelitian ini dilakukan perhitungan confusion matrix dengan hasil sebagai berikut: Dengan jumlah data 4653 dengan data testing yang digunakan 25%, sehingga hasil perhitungan confusion matrix didapatkan nilai:

- TP (True Positive) dimana prediksi benar positif dengan jumlah 581
- FP (False Positive) dimana prediksi salah positif dengan jumlah 56
- FN (False Negative) dimana prediksi salah negative dengan jumlah 38
- TN (True Negative) dimana prediksi benar negative dengan jumlah 489



K. Evaluasi

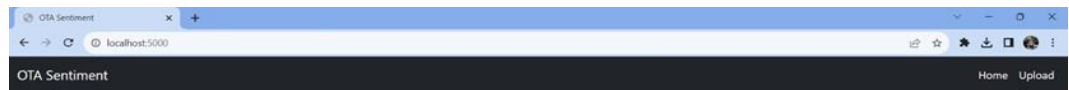
Pada grafik tersebut dapat terlihat bahwa nilai $k=6$ akurasi training dan akurasi testingnya menemukan titik Tengah dimana nilai akurasinya = 92%. Adapun, pada grafik tersebut dapat terlihat bahwa nilai $k=1$ data trainnya memiliki akurasi = 100% namun, testing akurasinya memiliki akurasi = 97%. Maka dari itu, nilai k yang digunakan pada penelitian ini, yaitu $k=6$. Untuk informasi lebih rinci mengenai report hasil klasifikasi dari setiap skenario dapat dilihat dari tabel berikut



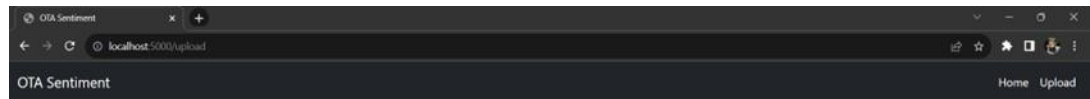
Percobaan nilai K		Akurasi	Waktu Proses	Presisi	Recall	F1-Score	Support
k=1	Negative	97%	1.97 s	96%	100%	98%	637
	Positive			99%	95%	97%	527
k=5	Negative	82%	1.42 s	93%	72%	81%	637
	Positive			73%	93%	82%	527
k=6	Negative	92%	1.19 s	94%	91%	93%	637
	Positive			90%	93%	91%	527
k=7	Negative	72%	1.94 s	93%	52%	67%	637
	Positive			62%	95%	75%	527

L. Visualisasi

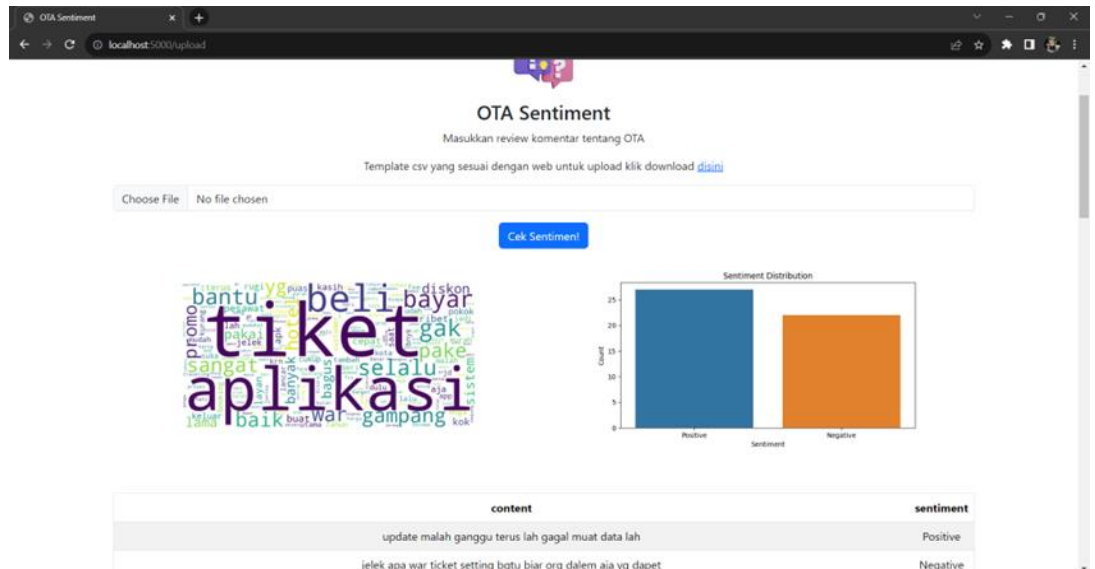
Pada tahapan deployment dan visualisasi dalam bentuk web yang dapat digunakan oleh user. Terdapat dua halaman, yaitu halaman index untuk user mengecek sentimen per kalimat dan halaman upload untuk user mengecek sentimen per csv.



sangat memuaskan, transaksi mudah, banyak promo. toplah pokoknya
Positive



Cek Sentimen!



4. KESIMPULAN DAN SARAN

Kesimpulan

- Tahapan membangun model analisis sentimen adalah studi literatur, analisis kebutuhan, perancangan model yang meliputi pengumpulan data, penyaringan data, text preprocessing, EDA, pelabelan data, klasifikasi model, pembagian data hold out, evaluasi model klasifikasi, dan visualisasi.
- Dalam pengembangan model klasifikasi menggunakan algoritma KNN dengan nilai $k=6$.
- Proses visualisasi ulasan aplikasi OTA disajikan dengan hasil sentimen, grafik untuk setiap kelas sentimen, wordcloud pada dataset yang diupload, dan tabel kelas sentimen dalam bentuk web. Dengan adanya aplikasi OTA ini diharapkan dapat mempermudah, membantu serta memberikan informasi tentang emosi yang terkandung dalam ulasan tentang aplikasi OTA yang ada pada situs Google Play Store

Saran

- Memperluas kemampuan sistem untuk mengatasi ulasan dengan kata-kata singkatan dan kesalahan pengejaan (typo), sehingga dapat meningkatkan akurasi dalam mengolah data yang tidak standar.
- Menambahkan kata tidak baku seperti “gak” saat preprocessing, sehingga dapat meningkatkan akurasi dalam pemberian label sentiment.
- Menambah kelas sentimen netral ketika library sudah menyediakannya di kemudian hari.

5. DAFTAR RUJUKAN

- Afshoh, F. (2017). Analisa Sentimen Menggunakan Naïve Bayes. *Informatika, Program Studi Komunikasi, Fakultas Informatika, D A N Surakarta, Universitas Muhammadiyah*, 12, 17.
- Afuan, L. (2018). Sentimen Analisis Di Twitter (Studi Kasus: Resepsi Pernikahan Putri Presiden Ri Ke-7). *Semnasteknomedia Online*, 6(1), 1-6.
- Aisha Alfani W. P. R., Fahrur Rozi, F. S. (2021). Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 6(1), 155–160.
- Alasmari, S. F., Dahab, M. 2017. Sentiment Detection, Recognition, and Aspect Identification. *International Journal of Computer Applications*, Vol. 177, No. 2, hal. 31-38.
- Amrullah, A. Z., Anas, A. S., & Hidayat, M. A. J. (2020). Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square. *Jurnal BITE*, 2(1), 40–44. <https://doi.org/10.30812/bite.v2i1.804>
- Andriana, H., Hilab, S. S., & ... (2023). Penerapan Metode K-Nearest Neighbor pada Sentimen Analisis Pengguna Twitter terhadap KTT G20 di Indonesia. *JURIKOM (Jurnal ..., 10(1)*, 60–67. <https://doi.org/10.30865/jurikom.v10i1.5427>
- Aqlan, A. A. Q., Manjula, B., Naik, R. L. 2019. *A Study of Sentiment Analysis: Concepts, Techniques, and Challenges. Proceedings of International Conference on Computational Intelligence and Data Engineering*, hal. 147-162.
- Cahyo, J. (2022). *Analisis Sentimen Pengguna Twitter Terhadap Penerapan Pembelajaran Tatap Muka Pada Masa Pandemi Covid-19 Menggunakan* http://repository.upnjatim.ac.id/id/eprint/6923%0Ahttp://repository.upnjatim.ac.id/6923/1/18082010034_Cover.pdf
- Cici, H., & Hwan, S. Y. (2023). *Mitigating Class Imbalance in Sentiment Analysis through GPT-3-Generated Synthetic Sentences*. <https://doi.org/10.3390/app13179766>
- Erfina, A., Basryah, E. S., Saepulrohman, A., & Lestari, D. (2020). Analisis Sentimen Aplikasi Pembelajaran Online di Play Store pada Masa Pandemi Covid-19 Menggunakan Algoritma Support Vector Machine. *Seminar Nasional Informatika (SEMNASIF)*, 1(1), 145-152.
- Ernawati, S., & Wati, R. (2018). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel. *Jurnal Khatulistiwa Informatika*, 6(1), 64–69.
- Emarketer. (2018). *Global Digital Travel Sales 2018 Ease of Mobile Payment Options Boost Online Bookings*.
- Fajri, K., Jey, H. L., & Timothy, B. (2020). *In Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing (AACL/IJCNLP 2020)*.
- Feldman, R., & Sanger, J. (2007). *Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. New York: Cambridge University Press,
- Firdaus, M. F. El, Nurfaizah, N., & Sarmini, S. (2022). Analisis Sentimen Tokopedia Pada Ulasan di Google Playstore Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor. *JURIKOM (Jurnal Riset Komputer)*, 9(5), 1329. <https://doi.org/10.30865/jurikom.v9i5.4774>
- Giovani, A. P., Ardiansyah, A., Haryanti, T., Kurniawati, L., & Gata, W. (2020). Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi. *Jurnal Teknoinfo*, 14(2), 115. <https://doi.org/10.33365/jti.v14i2.679>

- He, H., Zhang, W., & Zhang, S. (2018). *A novel ensemble method for credit scoring: adaption of different imbalance ratios*. *Expert Systems With Applications*. doi:10.1016/j.eswa.2018.01.012
- Hartanto. (2017). Text Mining dan Sentimen Analisis Twitter pada Gerakan LGBT. *INTUISI: Jurnal Psikologi Ilmiah*, 9(1), 18–25.
- Hutto, C.J. and Gilbert, E. (2014). *VADER: A Parsimonious Rule-based Model for*. *Eighth International AAAI Conference on Weblogs and Social Media*, 18. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8109>
- Ilmawan, L. B., & Winarko, E. (2015). Aplikasi Mobile untuk Analisis Sentimen pada Google Play. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 9(1), 53. <https://doi.org/10.22146/ijccs.6640>
- Kotsiantis, S. B., D. Kanellopoulos dan P. E. Pintelas. 2006. *Data Preprocessing for Supervised Learning*. *International Journal of Computer Science* Volume 1 Number 2. Patras: University of Patras.
- Lidya, K.,S, Opim, S., & Syahril, E. (2015). Sentiment Analysis Pada Text Bahasa Indonesia Menggunakan Support Vector Machine (SVM) dan K-Nearest Neighbor (K-NN). Seminar Nasional Teknologi Informasi dan Komunikasi 2015, Yogyakarta.
- M. Bramer, *Principles of Data Mining*, Springer, 2007.
- Maharani, L. (2022). Evaluasi User Experience pada Aplikasi Pemesanan Tiket Pesawat Traveloka dan Tiket.com Menggunakan User Experience Questionnaire (UEQ). *Universitas Islam Negeri Syarif Hidayatullah Jakarta*. <https://repository.uinjkt.ac.id/dspace/handle/123456789/65499>
- Manning, C, Raghavan, P, dan Schutze, H. 2009. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press. Dapat diunduh dari: <https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf>.
- Marisa, F., Maukar, A. L., Farhan, A., Widodo, E. A., Sa'adah, I., & Dasilva, R. T. L. (2021). Pengukuran Tingkat Kematangan Kopi Arabika Menggunakan Algoritma K-Nearest Neighbour. *JIMP: Jurnal Informatika Merdeka Pasuruan*, 6(3), 1-5. <http://dx.doi.org/10.37438/jimp.v6i3.346>
- Obviously.ai. (2022). Cara Mengetahui Apakah Model Machine Learning Anda Memiliki Performa Baik. <https://www.obviously.ai/post/machine-learning-model-performance>.
- Prasetio Adhi, & Wita, N. (2022). Analisis Minat Pengguna OTA Tiket, cin di Indonesia Menggunakan Model Pendekatan Modifikasi UTAUT2. *Jurnal Manajemen Dan Keuangan*, 11(1), 36–54.
- Rahayu, S., MZ, Y., Bororing, J. E., & Hadiyat, R. (2022). Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 98–106. <https://doi.org/10.29408/edumatic.v6i1.5433>
- Ratnawati, F. (2018). Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter. *INOVTEK Polbeng - Seri Informatika*, 3(1), 50. <https://doi.org/10.35314/isi.v3i1.335>
- Sepri, D. (2020). Penerapan Algoritma Naïve Byase Untuk Analisis Kepuasan Penggunaan Aplikasi Bank. *JOSYC (Journal of Computer System and Informatics)* 2(1), 135-139.
- Sugianto, C. A., & Apandi, T. H. (2018). Pengaruh Tokenisasi Kata N-Grams Spam SMS Menggunakan Support Vector Machine [Preprint]. *INA-Rxiv*. <https://doi.org/10.31227/osf.io/vjc7k>
- Takdirillah, R. (2020). Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan. *Edumatic: Jurnal Pendidikan Informatika*, 4(1), 37–46. <https://doi.org/10.29408/edumatic.v4i1.2081>

- Tiket.com. (2023). Tiket.com – Tentang Kami. <https://www.tiket.com/info/about-us>.
- Traveloka. (2021). Traveloka - Tentang Kami. <https://www.traveloka.com/id-id/about-us>.
- Wahyudi, R., & Kusumawardana, G. (2021). Analisis Sentimen pada Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. *Jurnal Informatika*, 8(2), 200–207. <https://doi.org/10.31294/ji.v8i2.9681>