

Sentiment Analysis of Public Comments on Jetski Mahakam in TikTok Social Media Using a Lexicon-Based Approach

Siti Nur Atifah¹⁾, Wahyuni²⁾, Muhammad Ibnu Sa'ad³⁾

^{1,2,3}Sistem Informasi, STMIK Widya Cipta Dharma

Jl. M. Yamin, Gn. Kelua, Kec. Samarinda Ulu, Kota Samarinda, Kalimantan Timur 75123

E-mail: nuratifah1103@gmail.com

ABSTRACT

This study aims to conduct sentiment analysis on public comments on TikTok related to the Jetski Mahakam tourism activity. The analysis was carried out to understand public perception as reflected in user comments and to serve as an Evaluation tool for improving service quality and tourism promotion strategies. The research was conducted on the Jetski Mahakam tourism activity in Samarinda. Data collection techniques included interviews with the management, literature Reviews of relevant studies, and field studies by directly collecting comment data from the Jetski Mahakam TikTok account using Web scraping methods with the help of the Apify platform. This study employed a Lexicon-Based approach using the InSet (Indonesia Sentiment Lexicon) dictionary. The research process followed the CRISP-DM stages, namely business understanding, data understanding, data preparation, modelling, Evaluation, and deployment. A total of 505 comments were collected, and after data cleaning, 422 comments were analyzed. Each comment was classified into positive, negative, or neutral sentiment categories, either automatically using the Lexicon or manually. The research findings showed that the Manual Labelling produced 85 positive sentiments, 223 negative sentiments, and 114 neutral sentiments, Lexicon-Based Labelling using the InSet dictionary resulted in 60 positive sentiments, 290 negative sentiments, and 72 neutral sentiments. Meanwhile, GPT-Based Labelling generated 29 positive sentiments, 38 negative sentiments, and 355 neutral sentiments. In the model Evaluation phase, the comparison between manual and InSet Labelling yielded an accuracy of 71.56%, Manual and GPT Labelling showed 41.70% accuracy, and InSet and GPT Labelling resulted in 28.43% accuracy. These results indicate that the Lexicon-Based approach using the InSet dictionary is fairly effective for sentiment analysis. The system visualizes the results in the form of pie charts and word clouds, which help in understanding the distribution of sentiments and the dominant words in the comments.

Keywords : Sentiment Analysis, Lexicon, TikTok, InSet, CRISP-DM, Jetski Mahakam

Analisis Sentimen Komentar Masyarakat Terhadap Jetski Mahakam Pada Media Sosial TikTok Menggunakan Pendekatan Lexicon

ABSTRAK

Penelitian ini bertujuan untuk melakukan analisis sentimen terhadap komentar masyarakat pada media sosial TikTok terkait wisata Jetski Mahakam. Analisis ini dilakukan guna mengetahui persepsi publik yang terekam melalui komentar-komentar pengguna, serta sebagai bahan evaluasi untuk meningkatkan kualitas layanan dan strategi promosi wisata. Penelitian ini dilakukan pada wisata Jetski Mahakam di Samarinda. Teknik pengumpulan data yang digunakan meliputi wawancara dengan pihak pengelola, studi pustaka yang mengkaji literatur relevan, serta studi lapangan dengan melakukan pengumpulan data komentar secara langsung dari akun TikTok Jetski Mahakam menggunakan metode *Web scraping* dengan bantuan *platform Apify*. Dalam penelitian ini digunakan pendekatan Lexicon dengan memanfaatkan kamus InSet (Indonesia Sentiment Lexicon). Proses penelitian mengikuti tahapan CRISP-DM, yaitu *business understanding*, *data understanding*, *data preparation*, *modelling*, *Evaluation*, dan *deployment*. Komentar yang berhasil dikumpulkan sebanyak 505, dan setelah proses pembersihan data tersisa 422 komentar yang dianalisis. Setiap komentar diklasifikasikan ke dalam kategori sentimen positif, negatif, atau netral, baik secara otomatis menggunakan kamus Lexicon maupun secara manual. Hasil penelitian menunjukkan Hasil *Labelling* manual menghasilkan 85 sentimen positif, 223 sentimen negatif, dan 114 sentimen netral, hasil *Labelling* menggunakan kamus Lexicon InSet menghasilkan 60 sentimen positif, 290 sentimen negatif, dan 72 sentimen netral, dan *Labelling* GPT menghasilkan 29 sentimen positif, 38 sentimen negatif, dan 355 sentimen netral. Pada tahap model *Evaluation* dihasilkan perbandingan akurasi yaitu, perbandingan *Labelling* manual dan *Labelling* InSet menghasilkan akurasi 71.56%, perbandingan *Labelling* Manual dan *Labelling* GPT menghasilkan akurasi 41.70%, perbandingan *Labelling* InSet dan *Labelling* GPT menghasilkan akurasi 28.43%%. Dari hasil tersebut dapat disimpulkan bahwa pendekatan Lexicon menggunakan kamus InSet cukup efektif untuk analisis sentimen. *Visualisasi* di sistem dalam bentuk *pie chart* dan *word cloud* yang memudahkan pemahaman distribusi sentimen serta kata-kata dominan dalam komentar.

Kata Kunci: *Analisis Sentimen, Lexicon, TikTok, InSet, CRISP-DM, Jetski Mahakam*

1. PENDAHULUAN

Pariwisata berbasis olahraga air semakin berkembang di Indonesia, termasuk di Samarinda dengan hadirnya wisata Jetski Mahakam. Wisata ini menawarkan pengalaman unik bagi warga atau wisatawan yang ingin menikmati keindahan Sungai Mahakam dengan sensasi mengendarai jetski. Seiring dengan popularitasnya, promosi dan ulasan mengenai wisata ini banyak ditemukan di media sosial, terutama TikTok. Platform ini menjadi tempat utama bagi masyarakat untuk berbagi pengalaman, memberikan ulasan, serta menyampaikan opini mereka terkait wisata tersebut dalam bentuk komentar dan video. Analisis terhadap sentimen yang muncul dalam respons masyarakat dapat memberikan wawasan penting bagi pengelola wisata dalam meningkatkan pelayanan dan strategi pemasaran.

Perkembangan media sosial, khususnya TikTok, telah menjadi sarana utama bagi masyarakat untuk berbagi pengalaman, termasuk mengenai destinasi wisata. Salah satu destinasi wisata yang sedang berkembang di Samarinda adalah wisata Jetski di Sungai Mahakam. Meskipun memiliki potensi yang besar sebagai daya tarik wisata, respon masyarakat terhadap wisata ini belum banyak dianalisis secara sistematis. Oleh karena itu, penting untuk mengetahui bagaimana persepsi masyarakat mengenai wisata Jetski Mahakam melalui analisis sentimen di TikTok. Masyarakat sering kali berbagi pendapat mereka mengenai berbagai pengalaman wisata yang mereka alami, dan ini menciptakan peluang untuk memahami bagaimana persepsi publik terbentuk. Namun, tantangan utamanya adalah mengidentifikasi sentimen yang terkandung dalam komentar-komentar yang ditulis secara bebas, yang kadang tidak mudah dipahami karena adanya penggunaan bahasa gaul, ungkapan tidak baku, serta variasi dalam ekspresi emosi.

Untuk mengatasi permasalahan tersebut, analisis sentimen dengan menggunakan pendekatan Lexicon dapat diterapkan untuk mengklasifikasikan opini masyarakat menjadi kategori positif, negatif, atau netral. Metode ini mampu mengolah data teks dari berbagai platform secara otomatis, sehingga dapat memberikan gambaran yang lebih jelas mengenai sentimen publik terhadap wisata Jetski Mahakam Samarinda.

Diharapkan dengan penerapan analisis sentimen menggunakan Pendekatan Lexicon, pengelola wisata dapat memperoleh wawasan yang lebih mendalam mengenai kepuasan pelanggan. Hasil dari penelitian ini dapat menjadi bahan evaluasi untuk meningkatkan kualitas layanan, memperbaiki aspek yang kurang memuaskan, serta merancang strategi pemasaran yang lebih efektif guna menarik lebih banyak wisatawan ke Jetski Mahakam Samarinda.

2. RUANG LINGKUP

Agar penelitian ini tidak melebar lebih luas dari pembahasan yang dimaksud, maka dibuatlah batasan

masalah ini untuk membatasi pada ruang lingkup penelitian sebagai berikut :

1. Data yang digunakan berupa data komentar dari akun TikTok Jetski Mahakam Samarinda dan hanya mengambil komentar dari konten dengan *viewers* tertinggi.
2. Analisis sentimen dilakukan dengan menggunakan pendekatan Lexicon menggunakan kamus InSet (Indonesia Sentiment Lexicon)
3. Tahapan *modelling* dilakukan dengan 3 cara yaitu *Labelling* menggunakan kamus InSet, *Labelling* Manual, dan *Labelling* menggunakan GPT (*Artificial Intelligence*).
4. Klasifikasi sentimen dibatasi dalam tiga kategori, yaitu positif, negatif, dan netral.
5. Data yang terkumpul terdiri dari 45 kolom tetapi hanya kolom yang berisi komentar yang digunakan untuk proses analisis.
6. Pengujian sistem hanya menggunakan metode *Black Box Testing*

3. BAHAN DAN METODE

3.1 Analisis Sentimen

Menurut Asriana, dkk (2024), Analisis Sentimen adalah proses yang digunakan untuk menganalisis data teks seperti *tweet*, komentar, ulasan dan lainnya untuk mengelompokkan polaritas dari suatu teks dalam kalimat. Analisis sentimen dapat memberikan informasi mengenai persepsi atau pandangan pengguna terhadap suatu topik tertentu. Sehingga dapat digunakan untuk memberikan *feedback* bagi pihak terkait untuk meningkatkan kualitas layanan yang diberikan.

3.2 Lexicon Based Features

Menurut Manullang dkk (2023), Lexicon-Based (berbasis Lexicon) merupakan salah satu metode dalam analisis sentimen yang menggunakan kamus Lexicon yang telah dikurasi secara manual. Kamus Lexicon ini berisi daftar kata-kata atau frasa yang memiliki hubungan dengan sentimen tertentu, seperti positif, negatif, atau netral. Setiap kata atau frase dalam kamus diberi label sentimen yang sesuai. Lexicon Based adalah fitur kata yang memiliki sentimen positif atau negatif berdasarkan kamus atau Lexicon.

3.3 CRISP-DM

Menurut Muttaqin dkk (2024), *The Cross Industry Standard Procces for Data Mining* atau CRISP-DM, adalah model proses yang digunakan secara luas dalam penggalian data dan tidak bergantung pada industri tertentu. Pertimbangan metode CRISP-DM karena merupakan metode yang secara spesifik menekankan pada konteks data dan tujuan pengolahan data sehingga fokus utama pembahasan terkait kualitas produk dan

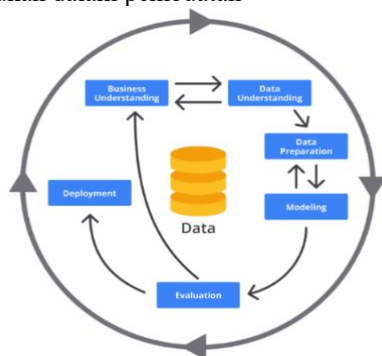
layanan yang berhubungan dengan manajemen destinasi wisata.

Berikut merupakan tahapan dalam metode CRISP-DM

1. *Business Understanding*
Pemahaman bisnis merupakan kajian tentang topik penelitian yang sedang dilakukan saat ini. Menurut Adytia, dkk (2023), Fungsi dasar dalam *Business Understanding* yaitu menentukan objek bisnis, menentukan tujuan teknis, data *science*, dan membuat rencana proyek.
2. *Data Understanding*
Persiapan data adalah proses menyiapkan data agar menjadi akurat, bersih, dan siap digunakan dalam penelitian.
3. *Data Preparation*
Data Preparation adalah proses menyiapkan data untuk mendapatkan informasi yang akurat, menyeluruh, dan dapat diandalkan yang dapat digunakan.
4. *Modelling*
Pemilihan teknik pemodelan adalah langkah pertama yang akan diambil dan diikuti untuk pembuatan skenario pengujian dan memvalidasi kualitas sebuah model.
5. *Evaluation*
Penelitian ini akan menggunakan teknik *Cross Validation* dengan menggunakan *Naive Bayes Classifier* untuk melakukan pengujian model. Menurut Syah, dkk (2024) *Output* model akan digunakan sebagai evaluasi kinerja model yang dilakukan dengan menggunakan beragam metrik yang mencakup akurasi. Representasi matematis disajikan dalam persamaan

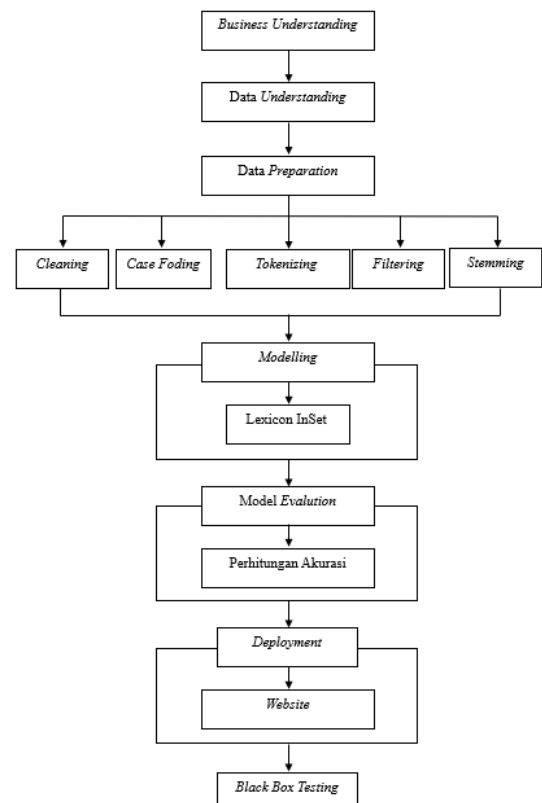
$$\text{Akurasi} = \frac{\text{Total Prediksi Benar}}{\text{Total Data}} \times 100\%$$

6. *Deployment*
Tahapan pengujian melalui eksperimen dan evaluasi dari model yang akan digunakan, dengan menghasilkan data yang dapat digunakan sebagai masukan dalam pembuatan



Gambar 2. 1 *Cross Standard Procces for Data Mining*
Sumber: Alwi, M. N, Dicoding Blog (2024)

4. HASIL DAN PEMBAHASAN



Gambar Alur Penelitian
Sumber : Dokumen Pribadi

4.1 *Busines Understanding*

Selama proses penelitian, dilakukan analisis terhadap permasalahan yang ada, pemahaman terhadap kebutuhan, serta perencanaan strategi guna mengatasi permasalahan tersebut. Berdasarkan data yang ada, ditemukan sejumlah keluhan dalam komentar di akun media sosial TikTok Jetski Mahakam. Keluhan tersebut meliputi berbagai aspek, seperti pengalaman pengguna, layanan, fasilitas, dan unsur lainnya yang dinilai oleh pengguna media sosial TikTok. Oleh karena itu, penelitian ini bertujuan untuk mengidentifikasi pola sentimen yang muncul dalam ulasan tersebut, naik yang bersifat positif, negatif, maupun netral.

Keluhan-keluhan yang ada di pada TikTok Jetski Mahakam dapat dijadikan bahan evaluasi bagi pihak institusi untuk ditindaklanjuti. Dengan menganalisis sentimen dalam ulasan tersebut, institusi dapat merumuskan langkah-langkah strategis untuk meningkatkan mutu layanan serta memperbaiki aspek-aspek yang menjadi sorotan utama. Tahapan ini mencakup penetapan tujuan spesifik, yaitu menyajikan hasil analisis sentimen yang dapat menjadi acuan bagi pengelola Jetski Mahakam dalam proses pengampilan Keputusan. Harapannya, penelitian ini dapat memberikan solusi yang aplikatif dalam menjawab permasalahan yang sedang dihadapi.

4.2 Data Understanding

Pengumpulan data pada penelitian ini dilakukan dengan menggunakan ulasan dari media sosial TikTok Jetski Mahakam. Ulasan yang digunakan yaitu komentar dari salah satu konten yang ada di akun Jetski Mahakam yang mempunyai viewers tertinggi. Total komentar yang berhasil dikumpulkan sebanyak 505 data yang masih berupa data mentah. Selanjutnya, data yang telah dikumpulkan akan diolah untuk analisis sentimen. Berikut tahapan yang dilakukan pada data *understanding*.

4.3 Data Preparation

Data preparation merupakan proses yang penting dilakukan sebelum dilakukan analisis sentiment. Pada tahap ini data komentar yang diperoleh akan melalui beberapa tahapan pembersihan dan penyesuaian agar dapat digunakan dalam pemodelan dan pengujian sistem. Adapun tahapan tersebut yaitu sebagai berikut:

4.3.1 Cleaning

Menghapus elemen yang tidak relevan atau tidak akan digunakan dalam analisis sentiment, seperti tautan (URL), *tag* HTML, emoji, simbol, dan angka dari dataset. Proses ini bertujuan agar dataset lebih bersih, tidak mengganggu pemrosesan data, dan siap digunakan untuk analisis sentimen.

4.3.2 Case Foding

Mengubah seluruh huruf dalam dataset menjadi huruf kecil yang bertujuan untuk menyamakan bentuk kata agar tidak terjadi perbedaan makna akibat perbedaan penulisan.

4.3.3 Normalisasi Kata

Menggantikan kata tidak baku (*salng*) menjadi bentuk kata baku. Kamus *slang* yang digunakan diambil dari GitHub berisi kata baku, kata tidak baku dan ditambahkan beberapa kata yang ada pada dataset komentar TikTok.

4.3.4 Tokenizing

Memecah teks menjadi kata-kata kecil yang disebut token, tahapan ini bertujuan agar setiap teks diubah menjadi kata-kata yang dapat digunakan pada tahapan *stopwords removal* dan proses analisis selanjutnya.

4.3.5 Stopwords

Penghapusan kata-kata pendek yang tidak memberikan pengaruh terhadap proses analisis, contohnya seperti *di*, *dan*, *ya*, *ke* dan kata-kata pendek lainnya yang tidak berpengaruh ke analisis sentimen. *Stopword* yang digunakan berasal dari file *combined_stop_words.txt*. yang berisi daftar *stopwords*, kata-kata yang termasuk dalam daftar *stopwords* tersebut akan dihapus dari kata yang akan dianalisis.

4.3.6 Stemming

Mengubah setiap kata atau kalimat seperti imbuhan (awalan dan akhiran) menjadi bentuk dasar (kata dasar) agar lebih mudah dianalisis.

4.4 Modelling

Modelling merupakan tahapan melakukan analisis dan klasifikasi sentimen berdasarkan data yang telah dipersiapkan sebelumnya melalui *data preparation*, tujuan dari tahapan ini yaitu untuk menghitung skor sentimen dan menentukan label sentimen (positif, negatif, netral). Pada penelitian ini *modelling* dilakukan dengan 3 cara yaitu menggunakan *Labelling InSet* dan *Labelling Manual*, dan *Labelling GPT*.

4.4.1 Labelling Manual

Pada *Labelling manual* setiap kata diberi skor secara manual dengan skor +1 = positif, -1 = negatif, dan 0 = netral. Setelah tahapan *Labelling manual* selanjutnya yaitu hasil *Labelling manual* dibandingkan dengan *Labelling InSet* untuk mengukur akurasi. Hasil *Labelling InSet* menghasilkan 223 sentimen negatif, 114 sentimen netral, dan 85 sentimen positif.

4.4.2 Labelling Lexicon Inset

Penelitian ini menggunakan pendekatan berbasis Lexicon dengan menggunakan kamus sentimen dari GitHub milik InSet (Indonesia Sentimen Lexicon) yang menyediakan daftar kata-kata positif dan negatif dalam Bahasa Indonesia. Tahapan ini juga bertujuan untuk menghitung jumlah kata positif dan negatif pada setiap kata, Jika skor > 0 = positif, < 0 = negatif, 0 = netral. Hasil *Labelling InSet* menghasilkan 290 sentimen negatif, 72 sentimen netral, dan 60 sentimen positif.

4.4.3 Labelling GPT

Labelling GPT bertujuan untuk memberikan penilaian terhadap komentar terkait jetski Mahakam, pada *Labelling GPT* setiap kata diberi skor oleh sistem. Setelah tahapan *Labelling GPT* selanjutnya hasil dari *Labelling GPT* dibandingkan dengan *Labelling InSet* untuk mengukur akurasi. Hasil *Labelling InSet* menghasilkan 38 sentimen negatif, 355 sentimen netral, dan 29 sentimen positif.

4.5 Model Evaluation

Tahapan model *Evaluation* bertujuan untuk mengukur performa model yang telah dibangun menggunakan pendekatan lexicon. Evaluasi dilakukan dengan cara membandingkan hasil prediksi sentimen lexicon InSet dengan hasil *Labelling Manual* dan GPT.

4.5.1 Perbandingan Labelling Manual dan labelling InSet

Prediksi yang benar sebanyak 302 dan jumlah *Prediksi* yang salah sebanyak 120, Akurasi yang dihasilkan dari perbandingan antara *Labelling manual* dan *Labelling InSet* sebesar 71.56%.

4.5.2 Perbandingan Labelling Manual dan labelling GPT

Prediksi yang benar sebanyak 176 dan jumlah *Prediksi* yang salah sebanyak 246, Akurasi yang dihasilkan dari perbandingan antara *Labelling* GPT dan *Labelling* Manual sebesar 41.70%.

4.5.3 Perbandingan Labelling InSet dan labelling GPT

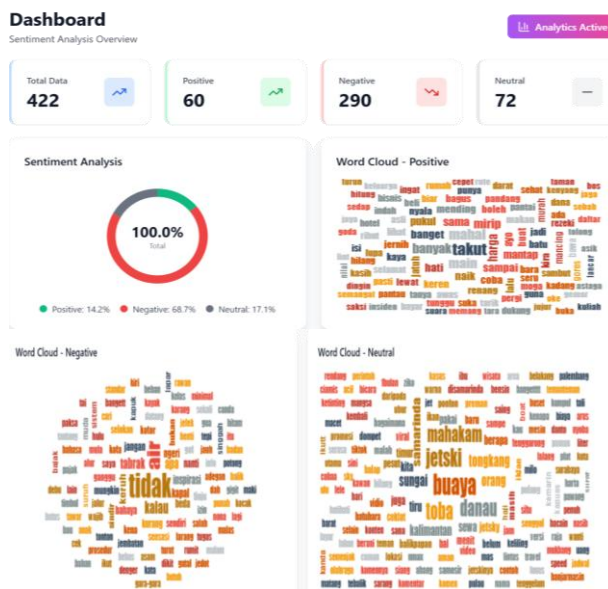
Prediksi yang benar sebanyak 120 dan jumlah *Prediksi* yang salah sebanyak 302, Akurasi yang dihasilkan dari perbandingan antara *Labelling* GPT dan *Labelling* InSet sebesar 28.43%.

4.6 Deployment

Deployment merupakan tahapan akhir yang bertujuan mengimplementasikan model atau insight yang diperoleh ke sistem agar dapat digunakan untuk pengambilan keputusan.

4.6.1 Halaman Dashboard

Halaman *Dashboard* menyajikan informasi hasil analisis sentimen dari data komentar yang telah diproses. Pada halaman ini pengguna (admin) dapat langsung melihat statistik dan *visualisasi* seperti total data yang menampilkan jumlah keseluruhan komentar yang dianalisis, selanjutnya jumlah sentimen (positif, negatif, netral), selanjutnya *pie chart* yang menampilkan persentase distribusi dari masing-masing sentimen, selanjutnya *word cloud* (positif, negatif, netral).

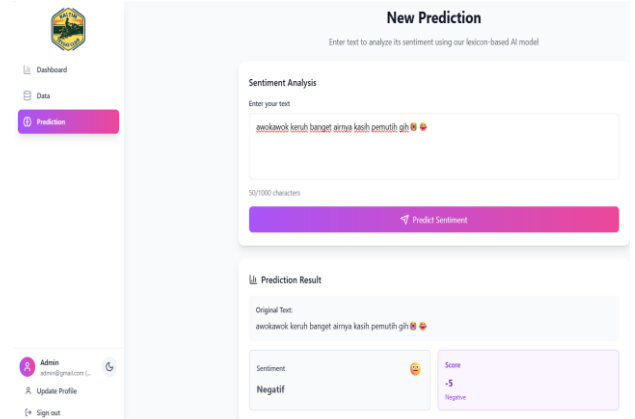


Sumber : Dokumen pribadi (2025)

4.6.2 Halaman Prediction

Tampilan halaman *Prediction* berfungsi untuk menampilkan hasil analisis sentimen dari komentar-komentar yang terdapat dalam dataset yang telah diimpor kedalam sistem. Pada halaman ini, pengguna tidak dapat

menginput komentar baru secara langsung kedalam dataset.



Sumber : Dokumen pribadi (2025)

4.7 Evaluation

Tahapan akhir dari penelitian ini adalah melakukan pengujian terhadap *Website* yang telah dibuat. Pengujian dilakukan melalui 1 cara, yaitu *Black Box Testing* yang difokuskan pada pengujian tampilan dan fungsionalitas antarmuka pengguna

No	Menu	Fungsi	Hasil yang diharapkan	Hasil Pengujian
1	Login	Memvalidasi keberhasilan Login	Pengguna berhasil Login	Lulus
2	Dashboard	Menampilkan total data komentar dan visualisasi data	Jumlah total komentar per kategori sentimen dan hasil visualisasi	Lulus
3	Menu Data	Menampilkan jumlah data, Import data dan hapus data	Data komentar berhasil ditampilkan	Lulus
4	Menu Prediksi	MemPrediksi komentar baru	Komentar berhasil diPrediksi dan ditampilkan hasilnya	Lulus
5	Menu Update Profil	Mengubah data pengguna seperti nama, email dan password	Data pengguna berhasil diperbarui	Lulus

Sumber : Dokumen Pribadi (2025)

5. KESIMPULAN

Berdasarkan hasil penelitian dan analisis yang telah dilakukan terhadap komentar di media sosial TikTok Jetski Mahakam, dapat diambil beberapa kesimpulan sebagai berikut:

1. Penelitian ini berhasil menerapkan metode pendekatan Lexicon menggunakan kamus InSet (Indonesia Sentiment Lexicon) untuk melakukan analisis sentimen komentar terhadap wisata Jetski Mahakam pada media sosial TikTok.
2. Proses analisis dilakukan melalui tahapan CRISP-DM yang terdiri dari *business understanding*, data

understanding, data preparation, modelling, Evaluation, dan deployment.

3. Sebanyak 505 data komentar berhasil dikumpulkan melalui teknik *scrapping* menggunakan *platform Apify*. Setelah dilakukan pembersihan data, diperoleh 422 komentar yang siap dianalisis.
4. Komentar berhasil diklasifikasikan kedalam 3 kategori sentimen, yaitu positif, negatif, dan netral.
5. Hasil *Labelling* Manual menghasilkan 85 sentimen positif, 223 sentimen negatif, dan 114 sentimen netral, *Labelling* menggunakan kamus *Lexicon InSet* menghasilkan 60 sentimen positif, 290 sentimen negatif, dan 72 sentimen netral, hasil, dan *Labelling* GPT menghasilkan 29 sentimen positif, 38 sentimen negatif, dan 355 sentimen netral.
6. Pada tahap model *Evaluation* dihasilkan perbandingan akurasi yaitu, perbandingan *Labelling* Manual dan *Labelling* InSet menghasilkan akurasi 71.56%, perbandingan *Labelling* Manual dan *Labelling* GPT menghasilkan akurasi 41.70%, perbandingan *Labelling* InSet dan *Labelling* GPT menghasilkan akurasi 28.43%. Dari hasil tersebut dapat disimpulkan bahwa pendekatan *Lexicon* menggunakan kamus *InSet* cukup efektif untuk analisis sentimen.
7. Hasil evaluasi klasifikasi sentimen menunjukkan bahwa akurasi tertinggi diperoleh dari perbandingan Manual dan InSet yaitu sebesar 71.56%, karena keduanya menggunakan pendekatan berbasis Bahasa Indonesia. Sebaliknya, akurasi Manual terhadap GPT hanya 41.70% dan InSet terhadap GPT sebesar 28.43%. Rendahnya akurasi GPT disebabkan oleh pendekatan yang berbeda serta kecenderungan model untuk memberikan label netral secara berlebihan. Akibatnya, jumlah *Prediksi* yang benar dengan label manual maupun InSet menjadi sedikit, sehingga menghasilkan akurasi yang rendah.
8. *Visualisasi* di sistem dalam bentuk *pie chart* dan *word cloud* yang memudahkan pemahaman distribusi sentimen serta kata-kata dominan dalam komentar.
9. Tidak ditemukan *error* pada saat sistem saat pengujian dilakukan. Namun, terkadang terdapat kendala berupa gangguan atau *error* pada xampp yang bersifat eksternal dan tidak berasal dari sistem yang dikembangkan.

6. SARAN

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan lebih lanjut:

1. Mengembangkan model klasifikasi dengan menggabungkan pendekatan *Lexicon-Based* dan *machine learning* seperti *naive bayes* atau *svm* agar dapat menangani bahasa tidak baku dan menghasilkan akurasi yang tinggi.

2. Banyak komentar TikTok yang menggunakan bahasa tidak formal atau gaul, diharapkan dapat memperluas kamus normalisasi dan menambahkan pemrosesan khusus menangani istilah kekinian agar hasil analisis sentimen lebih akurat.
3. Menerapkan metode evaluasi seperti *k-fold cross-validation*, untuk mengukur kinerja model dengan akurasi, presisi, *recall*, dan *f1-score*.

7. REFERENSI

- Adytia, P., Wahyuni, W., Sussolaikah, K., & Satria, Y. (2023). Klasifikasi Penggunaan Data Trafik Internet Menggunakan Algoritma Support Vector Machine. *J-Icon: Jurnal Komputer dan Informatika*, 11(1), 96-102.
- Agusia, P., Manurung, M. U. A., Calista, V., & Mawardi, V. C. (2024, October). Pemanfaatan *Word cloud* Pada Analisis Sentimen Dalam Menggali Persepsi Publik. In *Seminar Nasional Penelitian (SEMNAS CORISINDO 2024)* (pp. 25-30).
- Anggoro, D., Alessandro, A., Aditya, P., & Wijaya, A. (2025, April). Analisis Sentimen Film Squid Game Melalui Platform X Menggunakan Metode Lexicon Based. In *MDP Student Conference* (Vol. 4, No. 1, pp. 378-383).
- Albanna, I., & Laksono, R. T. H. (2022, November). Implementasi Pandas Data frame sebagai Agregasi dan Tabulasi Penyajian Data Luaran Survei Kepuasan Pengguna Proses Pembelajaran dalam Pendidikan Tinggi. In *Prosiding Seminar Nasional Sains dan Teknologi Terapan*.
- Alita, D. (2021). Multiclass SVM Algorithm for Sarcasm Text in Twitter. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(1), 118-128.
- Alwi, M. N. (2024, June 18). *CRISP-DM: Tahapan, Studi Kasus, Kelebihan, dan Kekurangan*. Dicoding Blog. <https://www.dicoding.com/blog/crisp-dm-tahapan-studi-kasus-kelebihan-dan-kekurangan/>
- Arya, S. M., Pratama, I. A., Iqbal, I., Siddik, R., & Wahid, R. M. (2022). Strategi viral di TikTok: Panduan Pemasaran melalui media sosial untuk bisnis. *Jurnal Ilmiah Ekonomi Global Masa Kini*, 13(1), 59-67.
- Arifin, N., Enri, U., & Sulistiyowati, N. (2021). Penerapan Algoritma Support Vector Machine (SVM) dengan TF-IDF N-Gram untuk Text Classification. *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, 6(2), 129-136.
- Assyifa, M. S. (2025). *Analisis Sentimen Terhadap STMIK Widya Cipta Dharma Menggunakan Pendekatan Lexicon* (Doctoral dissertation, STMIK Widya Cipta Dharma).
- Asriana, A., Pratiwi, O. N., & Fakhurroja, H. (2024). Analisis Sentimen Pada Media Sosial Twitter Menggunakan Support Vector Machine (SVM) Terhadap Kata Kunci#

- TheLinkinJKT. *eProceedings of Engineering*, 11(4).
- Asyrofi, R. R., & Asyrofi, R. (2023). Implementasi Aplikasi Jupyter Notebook Sebagai Analisis Kreteria Plagiasi Dengan Teknik Simantik. *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 8(2), 627-637.
- Budiman, Y. U., Santoso, M. F., Pattiasina, T., & Wahidin, A. J. (2022). Aplikasi TikTok Sebagai Media Promosi Kelompok Masyarakat Peduli Iklim Bumiku Satu. *J-ABDI: Jurnal Pengabdian kepada Masyarakat*, 2(5), 5085-5092.
- Buku Ajar Dasar Pemrograman (Python). (n.d.). Retrieved May 1, 2025, from https://books.Google.co.id/books/about/Buku_Ajar_Dasar_Pemrograman_Python.html?hl=id&id=zA08EAAQBAJ&redir_esc=y
- Caesar, F. X. B., & Somya, R. (2021). Analisis Minat Beli Produk pada Toko Oleh-Oleh Khas Surabaya dengan Algoritme FP-Growth. *Seri Prosiding Seminar Nasional Dinamika Informatika*, 5(1).
- Crisp DM methodology. (2020, June 17). Smart Vision Europe. <https://www.sv-europe.com/crisp-dm-methodology/>
- Christian, C., & Voutama, A. (2024). Implementasi Aplikasi Antrian Pencucian Mobil Berbasis Web Menggunakan Php, Javascript, Html, Css Dan Uml. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 2243-2248.
- Eunike, E. (2024). *Analisis Sentimen Kepuasan Mahasiswa Terhadap Laboratorium Komputer STMIK Widya Cipta Dharma Menggunakan Algoritma Naive Bayes* (Doctoral dissertation, STMIK Widya Cipta Dharma).
- Faturahman, R. (2024, March 19). *Pengertian Python: dan Sintaksnya*. Artikel Tentang IT. <https://blogs.powercode.id/pengertian-Python-dan-sintaksnya/>
- Gifari, O. I., Adha, M., Hendrawan, I. R., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, 2(1), 36-40.
- Hingis, K., Karolina, U., & Priyadi, R. (2024). Sentimen Analisis Overclaim Skincare Skintific Menggunakan Algoritma BERT. *Applied Information Technology and Computer Science (AICOMS)*, 3(2), 1-5.
- Ichsanudin, M. N., Yusuf, M., & Suraya, S. (2022). Pengujian Fungsional Perangkat Lunak Sistem Informasi Perpustakaan Dengan Metode Black Box Testing Bagi Pemula. *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, 1(2), 1-8.
- Isnain, A. R., Sihabuddin, A., & Suyanto, Y. (2020). Bidirectional long short term memory method and Word2vec extraction approach for hate speech detection. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 14(2), 169-178.
- JUMA PRISTIKA, F. E. B. I. E. Sentimen Komentar Netizen dalam Postingan Pelantikan Menteri ATR/BPN pada Akun Instagram@agusyudhoyono. *Jurnal Ilmu Komunikasi*.
- Kusumo, S., & Somya, R. (2022). Penerapan Web Scraping Deskripsi Produk Menggunakan Selenium Python Dan Framework Laravel. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(4), 3426-3435.
- Nadia, P. H., Wahyuni, S. S., & Sari, A. P. (2024). Penerapan Naive Bayes Classifier untuk Analisis sentimen Ulasan Pelanggan pada Frenz Accessories Handphone. *Jurnal Mahasiswa Teknik Informatika*, 3(2), 19-29.
- Nadira, A., Setiawan, N. Y., & Purnomo, W. (2023). Analisis Sentimen Pada Ulasan Aplikasi Mobile Banking Menggunakan Metode Naïve Bayes Dengan Kamus InSet. *Indexia*, 5(01), 35-47
- Nawar, R. (2024, January 4). What is Black Box Testing: Techniques, and Best Practices. *LambdaTest*. <https://www.lambdatest.com/learning-hub/black-box-testing>
- Nazar, R. (2024). Implementasi Pemrograman Python Menggunakan Google Colab. *JIK: Jurnal Informatika dan Komputer*, 15(1), 50-56.
- Manullang, O., Prianto, C., & Harani, N. H. (2023). Analisis Sentimen Untuk MemPrediksi Hasil Calon Pemilu Presiden Menggunakan Lexicon Based Dan Random Forest. *Jurnal Ilmiah Informatika*, 11(02), 159-169.
- Merdiansah, R., Siska, S., & Ridha, A. A. (2024). Analisis sentimen pengguna X Indonesia terkait kendaraan listrik menggunakan IndoBERT. *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, 7(1), 221-228.
- Muttaqin, A. A., Alam, S., & Komara, M. A. (2024). Analisis Sentimen Isu Kecurangan Pemilu 2024 Berdasarkan Opini Pada Media Sosial Twitter Menggunakan Metode Crisp-Dm Dengan Algoritma Naïve Bayes Classifier. *Jati (Jurnal Mahasiswa Teknik Informatika)*, 8(5), 8764-8772.
- Peter. (2025, April 8). *Web Scraping Explained: From Fundamentals to Scalable Solutions*. Making Data Meaningful. <https://makingdatameaningful.com/Web-scraping/Project-Jupyter>. (n.d.). Home. Retrieved May 1, 2025, from <https://jupyter.org/>
- Qudsya, A. L., & Habibah, S. M. (2023). KEBEBASAN MAHASISWA DALAM BERPENDAPAT DARI PRESPEKTIF HAM. *Bureaucracy Journal: Indonesia Journal of Law and Social-Political Governance*, 3(3), 3035-3049.
- Rachman, F. F., & Pramana, S. (2020). Analisis sentimen pro dan kontra masyarakat Indonesia tentang vaksin COVID-19 pada media sosial

- Twitter. *Indonesian of Health Information Management Journal (INOHIM)*, 8(2), 100-109.
- RATAMA, A. F. S., & RATAMA, K. H. (2025). PENERAPAN METODE NATURAL LANGUAGE PROCESSING (NLP) DALAM IMPLEMENTASI ASISTEN VIRTUAL CHATBOT DENGAN MEMANFAATKAN API CHATGPT DAN GRADIO APP. *Journal of Research and Publication Innovation*, 3(1), 1-15
- Rumaisa, F., Puspitarani, Y., Rosita, A., Zakiah, A., & Violina, S. (2021). Penerapan Natural Language Processing (NLP) Di Bidang Pendidikan. *Jurnal Inovasi Masyarakat*, 1(3), 232-235.
- Romadhon, M. H., Yudhistira, Y., & Mukrodin, M. (2021). Sistem Informasi Rental Mobil Berbasis Android Dan Website Menggunakan Framework Codeigniter 3 Studi Kasus: CV Kopja Mandiri: Array. *Jurnal Sistem Informasi dan Teknologi Peradaban*, 2(1), 30-36.
- Saurina, N., Rahayuningsih, T., & Retnawati, L. (2022). Analisis Sentimen Ulasan Pelanggan Batik Ecoprint Menggunakan Naïve Bayes dan KNN Classifier. *Jatiji (Jurnal Teknik Informatika Dan Sistem Informasi)*, 9(2), 1532-1542.
- Syah, A., Nurdiansyah, F., & Rahman, A. Y. (2024). Analisis sentimen aplikasi Shopee, Tokopedia, Lazada dan Bibli menggunakan leksikon dan Random Forest. *Jurnal Informatika dan Teknik Elektro Terapan*, 12(3S1).
- Sa'ad, M. I., & Bryan, D. (2020, November). Decision Support System for Covid19 Affected Family Cash Aid Recipients Using the Naïve Bayes Algorithm and the Weight Product Method. In *2020 3rd International Conference on Information and Communications Technology (ICOIAC)* (pp. 120-125). IEEE.
- Salam, A., Zeniarja, J., & Khasanah, R. S. U. (2018). Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekspres Indonesia).
- Sanjaya, G., & Lhaksana, K. M. (2020). Analisis Sentimen Komentar YouTube tentang Terpilihnya Menteri Kabinet Indonesia Maju Menggunakan Lexicon Based. *eProceedings of Engineering*, 7(3).
- Sari, I. P., Azzahrah, A., Qathrunada, I. F., Lubis, N., & Anggraini, T. (2022). Perancangan sistem absensi pegawai kantor secara online pada Website berbasis HTML dan CSS. *Blend sains jurnal teknik*, 1(1), 8-15.
- Setiawan, R. (2021, November 16). *Black Box Testing Untuk Menguji Perangkat Lunak*. Dicoding Blog. <https://www.dicoding.com/blog/black-box-testing/>
- Sihombing, D. O. (2022). Implementasi natural language processing (nlp) dan algoritma cosine similarity dalam penilaian ujian esai otomatis. *Jurnal Sistem Komputer dan Informatika (JSON) Hal*, 396, 406.
- Singgale, Y. A. (2023). Penerapan Metode CRISP-DM dalam Klasifikasi Data Ulasan Pengunjung Destinasi Danau Toba Menggunakan Algoritma Naïve Bayes Classifier (NBC) dan Decision Tree (DT). *J. Media Inform. Budidarma*, 7(3), 1551-1562.
- Sirati, A. A., & Nasution, M. I. P. (2024). Efektivitas Platform TikTok Sebagai Media Pembelajaran PAI Berbasis Literasi Digital. *Dirosat: Journal of Islamic Studies*, 9(1), 83-96.
- Sulistati, T., Yuliansyah, F., Romzi, M., & Aryani, R. (2020). Membangun Website toko online pempek nthree menggunakan PHP dan MYSQL. *JTIM: Jurnal Teknik Informatika Mahakarya*, 3(1), 35-44.
- Syafnidawaty. (2020, October 19). *White Box Testing*. UR. <https://raharja.ac.id/2020/10/19/white-box-testing/>
- Team. (2023, October 24). *Black Box Testing Adalah: Tipe Dan Contoh Pengujiannya*. Coding Studio. <https://codingstudio.id/blog/black-box-testing-adalah/>
- Uchenna, E. (2024, March 17). Automating Data Collection with Apify: From Script to Deployment. *DEV Community*. <https://dev.to/eunit/automating-data-collection-with-Apify-from-script-to-deployment-3d6o>
- Wattimena, F. Y., Renyaan, A. S., Koibur, R., Sumanik, E. D., & Nahakleky, T. J. (2022). Pelatihan Desain Web Bagi UMKM Menggunakan Google Site, Wix, HTML dan CSS. *Community Development Journal: Jurnal Pengabdian Masyarakat*, 3(3), 1466-1472.
- Widhyaestoeti, D., Iqam, S., Mutiyah, S. N., & Khairunnisa, Y. (2021). *Black Box Testing Equivalence Partitions Untuk Pengujian Front-End Pada Sistem Akademik Sitoda*. *Jurnal Ilmiah Teknologi Infomasi Terapan*, 7(3), 211-216.
- Yudiana, Y., Agustina, A. Y., & Khofifah, N. (2023). *Prediksi Customer Churn Menggunakan Metode CRISP-DM Pada Industri Telekomunikasi Sebagai Implementasi Mempertahankan Pelanggan*. *Indonesian Journal of Islamic Economics and Business*, 8(1), 1-20.