Jurnal Universal Technic
Volume 5 Nomor 2 Bulan Oktober 2026
e-ISSN : 2962-6897; p-ISSN : 2962-6900, Hal. 209-225
DOI: https://doi.org/10.58192/unitech.v5i2.5263
Available online at: https://journal.unimar-amni.ac.id/index.php/UNITECH

Implementasi Ekstraksi Fitur Vektor dan Multi Algoritma Klasifikasi Untuk Sentimen Analisis Ulasan Aplikasi Marine Traffic

Harirestio Nurkolis Fadillah[1], Astrid Novita Putri[2*], Setiangga Fachrurrozi[3]

[1*]Universitas Semarang, Indonesia
[2]Universitas Semarang, Indonesia
[3]Universitas Maritim AMNI, Indonesia

Alamat : Jl. Soekarno Hatta, Semarang
Jl. Soekarno Hatta, Semarang
Jl. Soekarno Hatta 180, Semarang
**Korespondensi penulis : astrid@usm.ac.id*

Abstract*. MarineTraffic is a digital-based vessel monitoring application that provides real-time vessel position and movement tracking services. User reviews on the Google Play Store contain important information regarding user satisfaction levels, complaints, and experiences. However, the lack of an automated sentiment analysis system in MarineTraffic application reviews means that complaints and satisfaction cannot be identified quickly, objectively, and measurably, thus requiring automated sentiment analysis. This study aims to implement vector feature extraction and compare the performance of several classification algorithms in analyzing MarineTraffic application review sentiment. Data were obtained through a process of scraping user reviews on the Google Play Store, followed by preprocessing including cleaning, case folding, word normalization, tokenization, and stopword removal. Sentiment labeling was performed using the Lexicon Based method, while text representation was formed using vector feature extraction. The classification process was carried out using the Support Vector Machine (SVM), Naïve Bayes, Decision Tree, and Neural Network algorithms. Model evaluation was carried out using a Confusion Matrix based on accuracy, precision, recall, and F1-score values. From an initial 1,283 scraped reviews, preprocessing and duplicate removal produced a final dataset of 845 labeled reviews. Among the four classifiers, the Support Vector Machine (SVM) with a linear kernel achieved the best performance, with an accuracy of 93.5%, outperforming Neural Network, Decision Tree, and Naive Bayes. The results of this study are expected to provide an overview of user sentiment towards the MarineTraffic application and produce an effective classification model to support application service development*.

Keywords: *Sentiment Analysis, MarineTraffic, Vector Feature Extraction, Lexicon-Based, Machine Learning, Text Classification.*

Abstrak. MarineTraffic merupakan aplikasi pemantauan kapal berbasis digital yang menyediakan layanan pelacakan posisi dan pergerakan kapal secara real-time. Ulasan pengguna pada Google Play Store mengandung informasi penting mengenai tingkat kepuasan, keluhan, dan pengalaman pengguna. Namun, Kurangnya sistem analisis sentimen otomatis pada ulasan aplikasi MarineTraffic menyebabkan keluhan dan kepuasan tersebut belum dapat diidentifikasi secara cepat, objektif, dan terukur sehingga diperlukan analisis sentimen otomatis. Penelitian ini bertujuan untuk mengimplementasikan ekstraksi fitur vektor dan membandingkan kinerja beberapa algoritma klasifikasi dalam menganalisis sentimen ulasan aplikasi. Data diperoleh melalui proses scraping ulasan pengguna pada Google Play Store, kemudian dilakukan preprocessing yang meliputi cleaning, case folding, normalisasi kata, tokenisasi, dan stopword removal. Pelabelan sentimen dilakukan menggunakan Lexicon Based, sedangkan representasi teks dibentuk menggunakan ekstraksi fitur vektor. Proses klasifikasi dilakukan menggunakan algoritma Support Vector Machine (SVM), Naïve Bayes, Decision Tree, dan Neural Network. Evaluasi model dilakukan menggunakan Confusion Matrix berdasarkan nilai accuracy, precision, recall, dan F1-score. Hasil penelitian menghasilkan 845 ulasan berlabel dari 1.283 data mentah yang dikumpulkan setelah melalui tahap penghapusan duplikat dan preprocessing. Dari keempat algoritma yang diuji, Support Vector Machine (SVM) dengan kernel linear memberikan kinerja terbaik dengan akurasi sebesar 93,5%, mengungguli Neural Network, Decision Tree, dan Naive Bayes. Hasil penelitian diharapkan dapat memberikan gambaran sentimen pengguna terhadap aplikasi MarineTraffic serta menghasilkan model klasifikasi yang efektif untuk mendukung pengembangan layanan aplikasi.

Kata kunci: Analisis Sentimen, MarineTraffic, Ekstraksi Fitur Vektor, Lexicon Based, Machine Learning, Klasifikasi Teks.

Received: Juli 14, 2026; Revised: Juli 23, 2026; Accepted: Agustus 1, 2026;
Online Available: Oktober 29, 2026; Published: Oktober 29, 2026;

1. LATAR BELAKANG

Perkembangan era teknologi informasi dan komunikasi telah mendorong transformasi digital pada berbagai sektor, termasuk sektor maritim. Salah satu aplikasi yang banyak digunakan dalam bidang pelayaran dan logistik maritim adalah MarineTraffic, yaitu platform yang menyediakan layanan pemantauan posisi kapal secara real time. Melalui aplikasi ini, pengguna dapat memperoleh informasi mengenai lokasi kapal, pelabuhan, rute pelayaran, serta berbagai aktivitas maritim lainnya. Seiring meningkatnya penggunaan aplikasi digital, pengguna secara aktif memberikan ulasan dan penilaian pada platform distribusi aplikasi seperti Google Play Store. Ulasan tersebut mengandung informasi yang sangat berharga mengenai tingkat kepuasan pengguna, kualitas layanan, performa sistem, maupun berbagai permasalahan yang dialami pengguna.

Meningkatnya jumlah ulasan pengguna menyebabkan analisis secara manual menjadi kurang efektif dan memerlukan waktu yang lama. Oleh karena itu, diperlukan teknologi untuk mengolah data ulasan secara otomatis. Salah satu pendekatan yang banyak digunakan adalah analisis sentimen, yaitu teknik *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi opini atau persepsi pengguna berdasarkan data teks.. Analisis sentimen memungkinkan developer aplikasi memperoleh informasi (Nawulansih et al., 2025) yang lebih cepat dan akurat mengenai ulasan pengguna sehingga dapat digunakan sebagai dasar pengambilan keputusan dalam pengembangan produk.

Penelitian temtang analisis sentimen pada ulasan aplikasi berbasis Google Play Store telah banyak dilakukan. Penelitian oleh (Nawulansih et al., 2025) menunjukkan bahwa penerapan Support Vector Machine (SVM) dengan ekstraksi fitur TF-IDF mampu menghasilkan akurasi yang tinggi dalam mengklasifikasikan sentimen ulasan aplikasi DANA. Hasil penelitian tersebut menunjukkan bahwa kombinasi preprocessing teks, ekstraksi fitur, dan algoritma klasifikasi mampu mengidentifikasi opini pengguna secara efektif.

Penelitian yang dilakukan oleh (Nurfauzi et al., 2025) membandingkan algoritma Naïve Bayes dan Support Vector Machine pada analisis sentimen ulasan aplikasi Grab Indonesia. Hasil penelitian tersebut menunjukkan bahwa SVM memperoleh performa yang lebih baik dibandingkan Naïve Bayes dalam proses klasifikasi sentimen. Temuan tersebut menunjukkan bahwa pemilihan algoritma klasifikasi memiliki pengaruh signifikan terhadap kualitas hasil analisis sentimen.

Selain itu, penelitian (Setyani et al., 2026) pada ulasan aplikasi Jamsostek Mobile menggunakan pendekatan lexicon-based untuk pelabelan sentimen dan TF-IDF sebagai metode pembobotan fitur. Penelitian tersebut membuktikan bahwa kombinasi metode pelabelan otomatis dan representasi fitur vektor dapat meningkatkan efisiensi proses klasifikasi sentimen pada dataset berukuran besar.

Penelitian komparatif juga dilakukan (Dometian & Jatmiko, 2025) yang membandingkan algoritma SVM, Naïve Bayes, dan Logistic Regression pada data ulasan Google Play Store. Hasil penelitian menunjukkan bahwa setiap algoritma memiliki karakteristik dan tingkat performa yang berbeda sehingga diperlukan pengujian multi-algoritma untuk memperoleh model terbaik pada suatu kasus tertentu.

Penelitian (No et al., 2025) membandingkan algoritma SVM, Neural Network, dan Decision Tree pada analisis sentimen ulasan aplikasi Co-Pilot. Penelitian tersebut menunjukkan bahwa Neural Network memberikan akurasi tertinggi, namun SVM juga menunjukkan performa yang sangat kompetitif. Hasil ini mengindikasikan bahwa penggunaan lebih dari satu algoritma klasifikasi dapat memberikan gambaran yang lebih komprehensif mengenai kemampuan model dalam memahami pola sentimen pengguna.

Berdasarkan berbagai penelitian terdahulu, sebagian besar penelitian masih berfokus pada satu atau dua algoritma klasifikasi dan umumnya diterapkan pada aplikasi AI, e-commerce, maupun aplikasi kesehatan. Sementara itu, penelitian yang membahas analisis sentimen pada aplikasi berbasis mairitm masih sangat terbatas. Padahal aplikasi ini memiliki karakteristik pengguna yang berbeda karena digunakan oleh praktisi maritim, operator pelabuhan, perusahaan pelayaran, akademisi, maupun masyarakat umum yang membutuhkan informasi lalu lintas kapal.

Banyaknya ulasan pengguna pada aplikasi MarineTraffic belum diimbangi dengan adanya mekanisme analisis sentimen otomatis yang mampu mengidentifikasi kepuasan, keluhan, dan kebutuhan pengguna secara cepat, objektif, dan terukur. Akibatnya, informasi yang terkandung dalam ulasan pengguna belum dimanfaatkan secara optimal sebagai bahan evaluasi dan pengembangan aplikasi. Selain itu, masih terbatasnya penelitian yang membandingkan beberapa algoritma machine learning pada kasus analisis sentimen aplikasi MarineTraffic menimbulkan kebutuhan untuk mengetahui algoritma klasifikasi yang memiliki performa terbaik. Oleh karena itu, penelitian ini mengusulkan implementasi ekstraksi fitur vektor dan multi-algoritma klasifikasi untuk menganalisis sentimen ulasan aplikasi MarineTraffic guna menghasilkan informasi yang lebih akurat dan mendukung pengambilan keputusan berbasis data.

Berdasarkan uraian penelitian terdahulu di atas, dapat diidentifikasi tiga celah penelitian (research gap) utama. Pertama, dari sisi domain aplikasi, seluruh penelitian terdahulu yang ditinjau berfokus pada sektor finansial (DANA, Jamsostek Mobile), transportasi (Grab), maupun aplikasi berbasis AI (Co-Pilot), sementara aplikasi pada sektor maritim seperti MarineTraffic belum pernah menjadi objek penelitian analisis sentimen, meskipun karakteristik penggunanya (praktisi maritim, operator pelabuhan, perusahaan pelayaran) berbeda secara signifikan dari pengguna aplikasi pada domain lain. Kedua, dari sisi metodologi, sebagian besar penelitian terdahulu hanya membandingkan dua hingga tiga algoritma klasifikasi (misalnya SVM dengan Naïve Bayes, atau SVM dengan Logistic Regression), sehingga belum memberikan gambaran komprehensif mengenai perbandingan algoritma linear (SVM), probabilistik (Naïve Bayes), berbasis aturan (Decision Tree), dan berbasis jaringan saraf (Neural Network) secara bersamaan pada satu dataset yang sama. Ketiga, kombinasi antara pelabelan otomatis berbasis Lexicon Based (InSet) dengan ekstraksi fitur Bag-of-Words melalui CountVectorizer belum banyak dievaluasi secara khusus pada data ulasan berbahasa Indonesia di sektor maritim, sehingga efektivitas kombinasi tersebut pada domain ini masih belum diketahui.

Kontribusi penelitian ini terletak pada tiga aspek. Pertama, penelitian ini merupakan penerapan pertama analisis sentimen multi-algoritma pada ulasan aplikasi MarineTraffic, sehingga memperluas cakupan penelitian analisis sentimen ke domain maritim yang selama ini masih sangat terbatas dieksplorasi. Kedua, penelitian ini membandingkan secara sistematis dan setara empat algoritma klasifikasi (SVM, Naïve Bayes, Decision Tree, dan Neural Network) di atas representasi fitur Bag-of-Words yang identik, dengan pembagian data latih dan data uji yang sama untuk seluruh model, sehingga hasil perbandingan performa antar algoritma dapat dipertanggungjawabkan secara adil (fair comparison). Ketiga, penelitian ini mengintegrasikan pelabelan otomatis berbasis Lexicon Based InSet dengan pipeline preprocessing khusus domain maritim (termasuk kamus normalisasi istilah teknis), sehingga menghasilkan kerangka kerja analisis sentimen yang dapat direplikasi untuk aplikasi maritim sejenis.

2. KAJIAN TEORITIS

Analisis Sentimen

Analisis sentimen (*sentiment analysis*) merupakan salah satu bidang dalam *Natural Language Processing* (NLP) yang berfokus pada proses mengidentifikasi, mengklasifikasikan, dan mengevaluasi opini pengguna terhadap suatu produk, layanan, atau topik berdasarkan data

berbentuk teks. Tujuan utama analisis sentimen adalah menentukan polaritas opini ke dalam kategori positif, negatif, atau netral sehingga hasilnya dapat dimanfaatkan sebagai dasar dalam mendukung pengambilan keputusan. (Julianto et al., 2023).

Text Mining

Text mining merupakan proses penggalian informasi dari data teks yang tidak terstruktur untuk memperoleh pola, pengetahuan, atau informasi yang bermanfaat. Dalam penelitian analisis sentimen, text mining digunakan untuk mengolah data ulasan pengguna sehingga dapat dianalisis menggunakan metode machine learning. Tahapan text mining umumnya meliputi pengumpulan data, preprocessing, ekstraksi fitur, klasifikasi, dan evaluasi hasil(Arifiyanti et al., 2022).

Google Play Store Review

Google Play Store merupakan platform distribusi aplikasi Android yang menyediakan fasilitas bagi pengguna untuk memberikan ulasan dan penilaian terhadap aplikasi yang digunakan. Ulasan pengguna menjadi sumber informasi yang penting karena menggambarkan pengalaman, kepuasan, kritik, maupun saran terhadap suatu aplikasi. Oleh karena itu, data ulasan pada Google Play Store sering digunakan sebagai objek penelitian (Dometian & Jatmiko, 2025).

Data Scrapping

Data scraping merupakan teknik pengambilan data secara otomatis dari suatu platform digital menggunakan program komputer atau library tertentu. Dalam penelitian ini, scraping digunakan untuk memperoleh data ulasan pengguna aplikasi MarineTraffic dari Google Play Store sehingga proses pengumpulan data dapat dilakukan secara cepat dan efisien dibandingkan dengan metode manual.

Text Preprocessing

Text preprocessing merupakan langkah awal dalam pengolahan data teks yang bertujuan untuk membersihkan data dan meningkatkan kualitas informasi sebelum dilakukan proses klasifikasi. Tahapan pada preprocessing yang umum digunakan yaitu cleaning, case folding, stopword removal, tokenization, dan stemming (Maharani et al., 2026).

Lexicon Based Sentiment Analysis

Lexicon-Based Sentiment Analysis merupakan metode penentuan sentimen berdasarkan kamus kata positif dan negatif. Setiap kata pada dokumen diberikan skor tertentu sesuai polaritasnya, kemudian skor tersebut digunakan untuk menentukan kategori sentimen suatu ulasan. Metode ini banyak digunakan untuk pelabelan data otomatis karena tidak memerlukan proses anotasi manual dalam jumlah besar (Kumar et al., 2019)

Secara teknis, InSet (Indonesian Sentiment Lexicon) menyimpan setiap entri kata beserta bobot polaritasnya dalam rentang -5 hingga +5, di mana nilai negatif menunjukkan konotasi negatif dan nilai positif menunjukkan konotasi positif. Skor sentimen suatu dokumen dihitung dengan menjumlahkan bobot seluruh token hasil preprocessing yang cocok dengan entri leksikon, sebagaimana dinyatakan pada persamaan $\text{Skor}(d) = \sum w \in d \text{ bobot}(w)$. Ambang batas (threshold) yang digunakan pada penelitian ini adalah nol, sehingga dokumen dengan skor akumulasi ≥ 0 diklasifikasikan sebagai Positif, sedangkan skor < 0 diklasifikasikan sebagai Negatif. Pendekatan ini dipilih karena bersifat unsupervised sehingga tidak bergantung pada data berlabel awal, namun tetap memiliki keterbatasan dalam menangani negasi, sarkasme, dan kata di luar cakupan leksikon (out-of-vocabulary).

Ekstraksi Fitur Vektor

Ekstraksi fitur vektor merupakan proses mengubah data teks menjadi representasi numerik sehingga dapat diproses oleh algoritma machine learning. Proses ini sangat penting karena algoritma klasifikasi tidak dapat mengolah data teks secara langsung. Representasi numerik yang baik akan meningkatkan kemampuan model dalam mengenali pola sentimen pada data ulasan (Dometian & Jatmiko, 2025).

Bag of Words (BoW)

Bag of Words (BoW) merupakan model representasi teks yang memandang suatu dokumen sebagai kumpulan (bag) kata tanpa memperhatikan urutan maupun struktur gramatikal kalimat, sehingga hanya informasi kemunculan dan frekuensi kata yang dipertahankan. Setiap dokumen direpresentasikan sebagai vektor berdimensi V, dengan V merupakan ukuran kosakata (vocabulary) hasil gabungan seluruh kata unik pada korpus data latih, dan setiap elemen vektor menyatakan frekuensi kemunculan kata tertentu pada dokumen tersebut. Model ini dipilih dalam penelitian ini karena kesederhanaannya, kompleksitas komputasi yang rendah, serta kemudahan interpretasi dibandingkan model representasi yang lebih kompleks seperti TF-IDF, Word2Vec, atau model berbasis transformer, sehingga sesuai untuk data ulasan aplikasi yang relatif pendek dan bersifat sparse.

CountVectorizer

CountVectorizer merupakan metode ekstraksi fitur yang mengubah teks menjadi matriks frekuensi kata berdasarkan jumlah kemunculan kata dalam dokumen. Metode ini termasuk pendekatan Bag of Words (BoW) yang banyak digunakan pada klasifikasi teks karena sederhana dan memiliki kompleksitas komputasi yang relatif rendah (K & Thilagam, 2022).
Secara matematis, untuk korpus data latih $D = \{d1, d2, ..., dn\}$ dengan kosakata unik $V = \{t1, t2, ..., tm\}$, CountVectorizer menghasilkan matriks dokumen-term X berukuran $n \times m$, dengan

Xij menyatakan jumlah kemunculan term tj pada dokumen di. Proses ini terdiri atas dua tahap, yaitu pembentukan kosakata (fit) dari seluruh data latih dan transformasi (transform) setiap dokumen ke dalam vektor frekuensi berdasarkan kosakata tersebut. Pada penelitian ini, CountVectorizer dipilih sebagai metode ekstraksi fitur karena tiga pertimbangan utama: (1) kompleksitas komputasi yang rendah sehingga sesuai untuk data ulasan berukuran menengah, (2) interpretabilitas hasil yang tinggi karena setiap dimensi vektor merepresentasikan frekuensi kata secara langsung, dan (3) kesesuaiannya sebagai baseline representasi teks sebelum dibandingkan dengan metode pembobotan yang lebih kompleks seperti TF-IDF pada penelitian lanjutan.

Support Vector Machine

Support Vector Machine (SVM) adalah algoritma supervised learning yang bekerja dengan cara mencari hyperplane optimal untuk memisahkan data dari kelas yang berbeda. SVM banyak digunakan dalam klasifikasi teks karena mempunyai kemampuan yang baik dalam menangani data berdimensi tinggi dan menghasilkan performa klasifikasi yang tinggi(Fachrurrozi et al., 2021)

Naive Bayes

Naïve Bayes merupakan algoritma klasifikasi probabilistik yang menggunakan Teorema Bayes dengan asumsi bahwa setiap fitur memiliki sifat independen. Algoritma ini dikenal memiliki proses komputasi yang cepat dan sering digunakan dalam penelitian analisis sentimen berbasis teks (Hartmann et al., 2019)

Decision Tree

Decision Tree merupakan algoritma klasifikasi yang membangun struktur pohon keputusan berdasarkan atribut yang dimiliki data. Keunggulan utama algoritma ini adalah kemampuannya dalam menghasilkan aturan keputusan yang mudah dimengerti dan diinterpretasikan (Cahyani & Muslim, 2020).

Neural Network

Neural Network merupakan model machine learning yang meniru cara kerja jaringan saraf manusia dalam mempelajari pola data. Algoritma ini mampu mengenali pola yang kompleks sehingga sering digunakan pada berbagai permasalahan klasifikasi, termasuk analisis sentiment (Zablith & Osman, 2020).

Confusion Matrix

Confusion Matrix merupakan metode evaluasi yang dipergunakan untuk mengukur performa model klasifikasi dengan cara membandingkan hasil prediksi model terhadap data aktual. Melalui confusion matrix dapat dihitung berbagai metrik evaluasi seperti akurasi,

precision, recall, dan F1-score yang digunakan untuk menentukan kualitas model klasifikasi (Fachrurrozi et al., 2021).

Empat komponen utama confusion matrix pada klasifikasi biner adalah True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Berdasarkan keempat komponen tersebut, metrik evaluasi klasifikasi dihitung menggunakan persamaan berikut: Accuracy = (TP + TN) / (TP + TN + FP + FN), Precision = TP / (TP + FP), Recall = TP / (TP + FN), dan F1-Score = 2 × (Precision × Recall) / (Precision + Recall). Accuracy mengukur proporsi prediksi benar secara keseluruhan, Precision mengukur proporsi prediksi positif yang benar-benar positif, Recall mengukur proporsi data positif aktual yang berhasil dikenali model, sedangkan F1-Score merupakan rata-rata harmonis antara Precision dan Recall yang relevan digunakan pada dataset dengan distribusi kelas tidak seimbang seperti pada penelitian ini.

3. METODE PENELITIAN

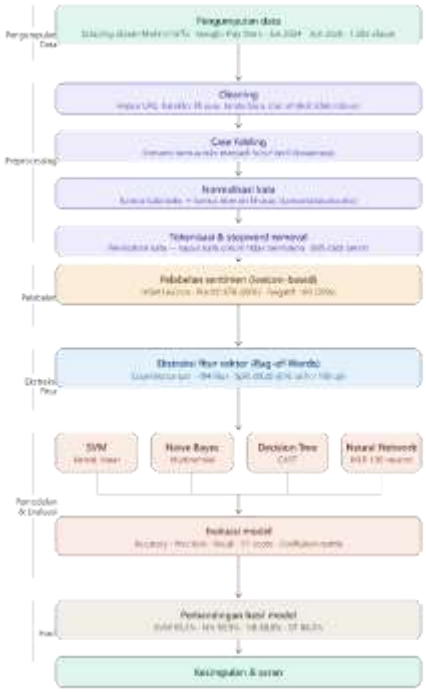
Metode penelitian yang digunakan dalam penelitian ini adalah metode eksperimen (experimental research) dengan pendekatan kuantitatif. Metode eksperimen digunakan untuk menguji dan membandingkan performa beberapa algoritma klasifikasi dalam melakukan analisis sentimen terhadap ulasan pengguna aplikasi MarineTraffic. Penelitian dilakukan melalui serangkaian tahapan yang meliputi pengumpulan data, preprocessing teks, pelabelan sentimen, ekstraksi fitur vektor, proses pelatihan model, pengujian model, evaluasi performa, serta analisis hasil klasifikasi. Pendekatan kuantitatif digunakan karena hasil penelitian diukur menggunakan nilai numerik berupa Accuracy, Precision, Recall, dan F1-Score. Pengumpulan data pada penelitian ini dilakukan melalui proses scraping ulasan pengguna aplikasi MarineTraffic di platform Google Play Store. Proses scraping dilakukan menggunakan library google-play-scraper pada bahasa pemrograman Python dengan parameter identifikasi paket aplikasi com.marinetraffic.android. Rentang waktu pengumpulan data mencakup periode 1 Januari 2024 hingga 10 Juni 2026, dengan pengambilan data berbahasa Indonesia (lang='id') dari pengguna yang berlokasi di Indonesia (country='id'). Pengambilan data dilakukan secara otomatis berdasarkan urutan ulasan terbaru (Sort.NEWEST) dengan batas maksimum 10.000 data per permintaan. Pembagian data dilakukan menggunakan teknik *stratified sampling* (stratify=y) untuk menjaga agar proporsi kelas Positif dan Negatif tetap seimbang antara data latih dan data uji, sehingga distribusi kelas pada data uji merepresentasikan distribusi kelas pada keseluruhan dataset dan mengurangi risiko bias evaluasi akibat ketidakseimbangan kelas.

Pemilihan CountVectorizer sebagai metode ekstraksi fitur didasarkan pada tiga pertimbangan: kompleksitas komputasi yang rendah sehingga proses pelatihan pada keempat

model tetap efisien, kemampuannya menjadi baseline representasi teks yang hasilnya dapat dibandingkan secara objektif dengan metode pembobotan lain pada penelitian lanjutan, dan kesesuaiannya dengan karakteristik data ulasan aplikasi yang relatif pendek sehingga representasi frekuensi kata sederhana sudah cukup informatif untuk membedakan sentimen.

Pemilihan empat algoritma klasifikasi (SVM, Naive Bayes, Decision Tree, dan Neural Network) dilakukan agar perbandingan mencakup empat paradigma pembelajaran yang berbeda, yaitu algoritma berbasis margin/hyperplane (SVM), algoritma probabilistik (Naive Bayes), algoritma berbasis aturan pohon keputusan (Decision Tree), dan algoritma berbasis jaringan saraf tiruan (Neural Network/MLP). Keragaman paradigma ini memungkinkan penelitian untuk menilai algoritma mana yang paling sesuai dengan karakteristik data BoW yang berdimensi tinggi dan sparse, sekaligus memberikan gambaran komprehensif yang belum tersedia pada penelitian terdahulu yang umumnya hanya membandingkan dua hingga tiga algoritma.

SVM dengan kernel linear secara khusus dipilih sebagai salah satu algoritma pembanding karena keunggulannya yang telah dibuktikan pada berbagai penelitian terdahulu (Nawulansih et al., 2025; Nurfauzi et al., 2025) dalam menangani data teks berdimensi tinggi hasil ekstraksi fitur BoW/TF-IDF. Kernel linear dipilih dibandingkan kernel non-linear (RBF, polynomial) karena data BoW pada umumnya sudah bersifat linearly separable pada ruang berdimensi tinggi, sehingga kernel linear dapat memberikan performa yang optimal dengan kompleksitas komputasi yang lebih rendah dan risiko overfitting yang lebih kecil dibandingkan kernel non-linear pada dataset berukuran menengah seperti pada penelitian ini.



Gambar 1. Alur Penelitian

4. HASIL DAN PEMBAHASAN

Total data mentah yang berhasil dikumpulkan sebanyak 1.283 ulasan. Setiap ulasan mencakup informasi seperti identitas pengguna, skor penilaian (rating), waktu ulasan, konten teks ulasan, jumlah thumbsUp, serta balasan dari pengembang aplikasi. Selanjutnya, data ini melewati proses verifikasi duplikat menggunakan metode hashing SHA-1 dan kesamaan teks (SequenceMatcher) untuk menghindari penghitungan ganda terhadap ulasan yang berasal dari pengguna yang sama.

Preprocessing Data

Dilakukan serangkaian tahapan preprocessing teks untuk membersihkan dan menyeragamkan data. Proses ini terdiri atas 6 tahapan utama yang dilaksanakan secara berurutan, sebagaimana dirangkum pada Tabel 1.

Tabel 1. Tahapan Preprocessing Teks

No.	Tahapan	Deskripsi Proses
1	Cleaning	Penghapusan URL, karakter khusus, tanda baca, dan simbol tidak relevan dari teks ulasan
2	Case Folding	Konversi seluruh karakter teks menjadi huruf kecil (lowercase) untuk menyeragamkan bentuk kata
3	Normalisasi Kata	Penggantian kata tidak baku menggunakan kamus kata baku yang diunduh dari repositori GitHub (kamuskatabaku.xlsx)

4	Tokenisasi	Pemisahan teks menjadi unit kata individual (token) menggunakan metode pemisahan spasi
5	Stopword Removal	Penghapusan kata-kata umum tidak bermakna (stopword) menggunakan daftar stopword Bahasa Indonesia dari pustaka NLTK, dilengkapi kata tambahan kustom
6	Pembuatan Kamus Khusus	Normalisasi tambahan menggunakan kamus khusus domain (mis. 'apk' → 'aplikasi', 'hp' → 'handphone') untuk menangani singkatan teknis

Proses deteksi duplikat mengidentifikasi 481 baris data yang termasuk dalam kelompok konten identik (mencakup data asli beserta salinannya). Setelah dilakukan penghapusan duplikat dengan mempertahankan kemunculan pertama (*drop_duplicates*), sebanyak 411 data duplikat dihapus, sehingga tersisa 872 ulasan unik yang digunakan sebagai dasar proses preprocessing lebih lanjut. Pada tahap preprocessing lanjutan (cleaning, case folding, normalisasi, tokenisasi, dan stopword removal), ditemukan 27 ulasan yang menghasilkan teks kosong setelah seluruh proses pembersihan yang umumnya berasal dari ulasan yang hanya berisi simbol, emotikon, angka, atau kata-kata stopword. Data tersebut dihapus, sehingga jumlah data akhir yang siap dilabeli adalah sebanyak 845 ulasan.

Pelabelan Data dengan metode Lexicon-Based

Pelabelan sentimen dilakukan secara otomatis menggunakan pendekatan lexicon-based dengan dua kelas (2-class), yaitu Positif dan Negatif. Leksikon sentimen yang digunakan adalah InSet (Indonesian Sentiment Lexicon) yang terdiri atas kamus kata positif dan kata negatif yang dapat diakses secara publik melalui repositori GitHub. Proses pelabelan dilakukan dengan menghitung skor sentimen berdasarkan kecocokan token dalam teks yang telah dipreproses dengan entri di dalam leksikon. Apabila skor akumulasi bernilai positif atau nol, ulasan diberi label 'Positif', sedangkan apabila skor bernilai negatif, ulasan diberi label 'Negatif'. Distribusi hasil pelabelan dari 845 ulasan disajikan pada Tabel 2.

Tabel 2. Distribusi Kelas Sentimen Hasil Pelabelan

Kelas Sentimen	Jumlah Data	Persentase (%)
Positif	601	71,12
Negatif	244	28,88
Total	845	100

Kondisi ini mengindikasikan bahwa mayoritas pengguna memberikan penilaian positif terhadap aplikasi MarineTraffic. Distribusi ini berkaitan langsung dengan kualitas kinerja

model klasifikasi, terutama pada kemampuan model dalam mengenali kelas minoritas (Negatif)

Ekstraksi Fitur Vektor (Bag of Words)

Untuk mengubah data teks menjadi representasi numerik yang dapat diterima oleh algoritma machine learning, digunakan metode Bag-of-Words (BoW) melalui CountVectorizer dari pustaka scikit-learn. CountVectorizer bekerja dengan membangun kamus kata (vocabulary) dari seluruh data latih, kemudian merepresentasikan setiap dokumen sebagai vektor frekuensi kemunculan kata.

Sebelum proses vektorisasi, dataset dibagi menjadi data latih (training set) dan data uji (testing set) dengan rasio 80:20. Dari total 845 data, sebanyak 676 data (80%) digunakan untuk pelatihan dan 169 data (20%) digunakan untuk pengujian. Pembagian ini dilakukan menggunakan fungsi train_test_split dari scikit-learn dengan parameter random_state=42 untuk memastikan reproduktifitas hasil.

Tabel 3. Statistik Pembagian Dataset dan Ekstraksi Fitur

Jenis Data	Nilai
Total Data	845
Data Latih (80%)	676
Data Uji (20%)	169

Berdasarkan Tabel 3, CountVectorizer menghasilkan 784 fitur unik dari data latih. Nilai ini mencerminkan ukuran kosakata efektif yang terbentuk setelah proses preprocessing, di mana jumlah fitur yang lebih kecil umumnya berdampak pada efisiensi komputasi yang lebih baik. Representasi vektor yang dihasilkan bersifat sparse (jarang terisi), yang merupakan karakteristik umum dari metode BoW pada teks pendek seperti ulasan aplikasi.

Hasil Klasifikasi Multi Algoritma

Pada tahap ini, empat algoritma klasifikasi diimplementasikan dan dievaluasi, yaitu Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT), dan Neural Network (NN/MLP). Setiap model dilatih menggunakan data latih dan dievaluasi pada data uji. Parameter yang digunakan untuk masing-masing model adalah: SVM dengan kernel linear random_state=42), Naive Bayes Multinomial, Decision Tree dengan random_state=42, dan Multilayer Perceptron dengan hidden_layer_sizes=(100,) dan max_iter=500

Tabel 4. Kinerja Klasifikasi per Kelas Sentimen – Kelas Negatif

Algoritma	Precision	Recall	F1-Score	Support
SVM	0,852	0,939	0,893	49
Naive Bayes	0,789	0,612	0,690	49
Decision Tree	0,812	0,796	0,804	49
Neural Network	0,843	0,878	0,860	49

Tabel 5. Kinerja Klasifikasi per Kelas Sentimen – Kelas Positif

Algoritma	Precision	Recall	F1-Score	Support
SVM	0,974	0,933	0,953	120
Naive Bayes	0,855	0,933	0,892	120
Decision Tree	0,917	0,925	0,921	120
Neural Network	0,949	0,933	0,941	120

Berdasarkan Tabel 4 dan Tabel 5, SVM secara konsisten menunjukkan kinerja terbaik di kedua kelas sentimen. Pada kelas Negatif, SVM mencapai F1-score sebesar 0,893, melampaui Naive Bayes (0,690), dan Decision Tree (0,804) dan Neural Network (0,860). Pada kelas Positif, SVM juga unggul dengan F1-score (0,953), dibandingkan Naive Bayes (0,892), Decision Tree (0,921) dan Neural Network (0,941).

Tabel 6. Perbandingan Akurasi dan Weighted Average F1-Score Keempat Model

Algoritma	Akurasi	Macro Avg Precision	Macro Avg Recall	Macro Avg F1-Score	Weighted Avg F1-Score
SVM	93,5%	0,913	0,936	0,923	0,936
Neural Network	91,7%	0,896	0,905	0,901	0,918
Naive Bayes	84%	0,822	0,773	0,791	0,834
Decision Tree	88,8%	0,865	0,860	0,863	0,887

Berdasarkan Tabel 6, SVM dengan kernel linear menunjukkan akurasi tertinggi sebesar 93,5%, diikuti oleh Neural Network (91,7%), Decision Tree (88,8%) dan Naive Bayes (84%). Rentang perbedaan akurasi antara model terbaik dan terburuk adalah sebesar 9,5 poin persentase, yang mengindikasikan bahwa pemilihan algoritma memberikan dampak yang cukup signifikan terhadap kualitas prediksi

Analisis Confuison Matrix

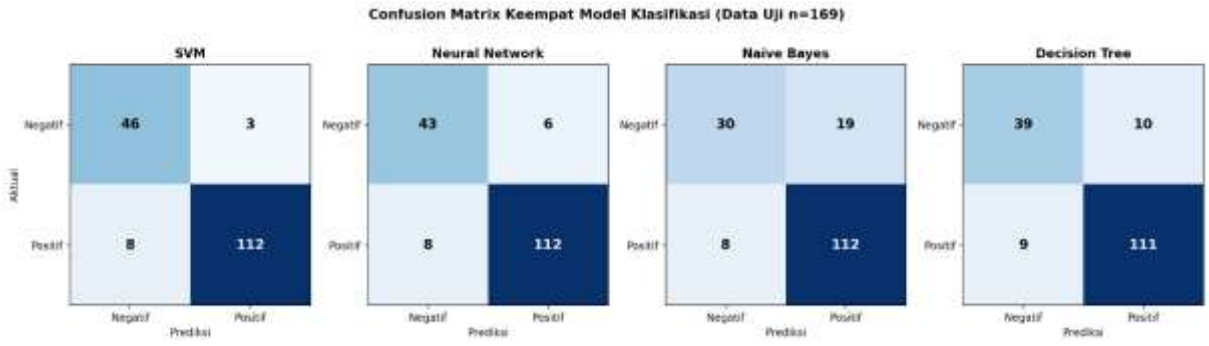
Untuk memperoleh pemahaman yang lebih mendalam mengenai pola kesalahan klasifikasi, dilakukan analisis confusion matrix terhadap setiap model. Confusion matrix menyajikan jumlah prediksi benar (True Positive dan True Negative) serta kesalahan klasifikasi (False Positive dan False Negative) secara terperinci. Tabel 7 merangkum komponen confusion matrix dari keempat model.

Tabel 7. Rekapitulasi Confusion Matrix Keempat Model Klasifikasi

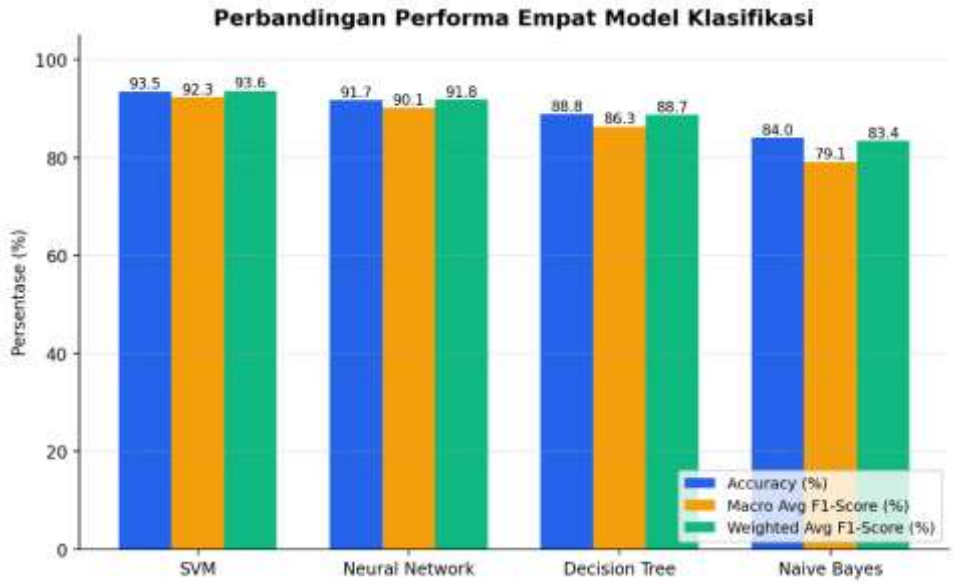
Algoritma	TP (Neg→Neg)	FP (Neg→Pos)	FN (Pos→Neg)	TN (Pos→Pos)
SVM	46	3	8	112
Neural Network	43	6	8	112
Naive Bayes	30	19	8	112
Decision Tree	39	10	9	111

Keterangan: TP = True Positive (Negatif diprediksi benar sebagai Negatif); FP = False Positive (Negatif salah diprediksi sebagai Positif); FN = False Negative (Positif salah diprediksi sebagai Negatif); TN = True Negative (Positif diprediksi benar sebagai Positif).

Berdasarkan Tabel 7, SVM menghasilkan kesalahan klasifikasi paling sedikit dengan total 11 kesalahan (3 FP + 8 FN) dari 169 data uji. Sebaliknya, Naive Bayes menghasilkan 27 kesalahan, diikuti oleh Decision Tree dengan 19 kesalahan klasifikasi, dan Neural Network dengan 14 kesalahan klasifikasi. Selain itu, terlihat adanya perbedaan pola kesalahan antar algoritma. SVM dan Neural Network menghasilkan jumlah false negative (FN) lebih tinggi dibandingkan false positive (FP), sedangkan Naive Bayes dan Decision Tree menghasilkan jumlah false positive (FP) yang lebih tinggi dibandingkan false negative (FN). Perbedaan ini menunjukkan bahwa setiap algoritma memiliki karakteristik kesalahan klasifikasi yang berbeda dalam membedakan ulasan negatif dan positif pada data yang memiliki distribusi kelas tidak seimbang.



Gambar 2. Visualisasi Confusion Matrix Keempat Model Klasifikasi



Gambar 3. Grafik Perbandingan Accuracy, Macro Avg F1-Score, dan Weighted Avg F1-Score Keempat Model

Gambar 2 menegaskan secara visual bahwa SVM dan Neural Network memiliki pola sel diagonal (TP dan TN) yang lebih pekat dibandingkan Naive Bayes, yang selnya menyebar lebih merata ke kuadran False Positive, mengindikasikan kecenderungan Naive Bayes untuk salah mengklasifikasikan ulasan Negatif sebagai Positif. Gambar 3 memperlihatkan bahwa keunggulan SVM konsisten pada ketiga metrik (Accuracy, Macro Avg F1-Score, dan Weighted Avg F1-Score), dengan selisih terhadap model peringkat kedua (Neural Network) yang relatif kecil (kurang dari 2 poin persentase), sedangkan jarak terhadap Naive Bayes jauh lebih lebar, khususnya pada Macro Avg F1-Score yang sensitif terhadap performa kelas minoritas.

Di luar temuan numerik di atas, penelitian ini memberikan tiga kontribusi ilmiah. Pertama, penelitian ini menghasilkan bukti empiris pertama mengenai kinerja komparatif SVM, Naive Bayes, Decision Tree, dan Neural Network pada data ulasan aplikasi maritim, sehingga mengisi celah literatur yang selama ini didominasi oleh studi pada domain finansial dan e-commerce. Kedua, dengan menerapkan pembagian data latih/uji yang identik dan representasi fitur BoW yang sama pada keempat model, penelitian ini menunjukkan bahwa keunggulan SVM tidak semata-mata disebabkan oleh perbedaan preprocessing atau ekstraksi fitur, melainkan oleh kesesuaian karakteristik algoritma margin-based dengan struktur data BoW yang berdimensi tinggi dan sparse, sebuah temuan yang memperkuat dan memperluas hasil penelitian sejenis pada domain lain (Nurfauzi et al., 2025). Ketiga, pola kesalahan klasifikasi yang teridentifikasi pada Gambar 2, yaitu kecenderungan seluruh model untuk lebih sering keliru mengenali ulasan Negatif dibandingkan ulasan Positif, memberikan kontribusi praktis berupa indikasi bahwa penelitian lanjutan pada domain maritim perlu

mempertimbangkan teknik penanganan ketidakseimbangan kelas secara eksplisit, bukan sekadar sebagai catatan keterbatasan, melainkan sebagai arah pengembangan model yang terukur.

5. KESIMPULAN DAN SARAN

Dari keempat algoritma yang diuji, SVM dengan kernel linear terbukti memberikan kinerja terbaik dengan akurasi 93,5%, precision 0,852, recall 0,939, dan F1-score 0,889 untuk kelas Negatif, serta precision 0,974, recall 0,933, dan F1-score 0,953 untuk kelas Positif. Hasil ini menjawab tujuan penelitian bahwa kombinasi representasi BoW dengan SVM merupakan pendekatan yang paling efektif di antara algoritma yang dibandingkan. Neural Network menempati posisi kedua dengan akurasi 91,7%, diikuti Decision Tree 88,8% dan Naive Bayes 84%. Seluruh model menunjukkan performa yang lebih baik dalam mengklasifikasikan kelas Positif dibandingkan kelas Negatif. hasil dari penelitian ini menunjukkan klasifikasi kelas Positif lebih banyak dibandingkan kelas Negatif.

Untuk penelitian selanjutnya, disarankan untuk eksplorasi metode representasi teks yang lebih kaya konteks seperti TF-IDF, Word2Vec, FastText, atau model berbasis transformer seperti IndoBERT diterapkan guna membandingkan efektivitasnya dengan BoW. Penerapan evaluasi menggunakan k-fold cross-validation untuk penelitian lanjutan juga dapat dilakukan. Penerapan teknik penanganan ketidakseimbangan kelas seperti SMOTE atau class weighting juga direkomendasikan untuk meningkatkan kemampuan deteksi kelas minoritas. Selain itu, validasi label melalui anotasi manual oleh ahli bahasa atau domain dapat memperkuat reliabilitas dataset. Perluasan cakupan analisis ke ulasan berbahasa Inggris atau bahasa lainnya dan penelitian lintas aplikasi maritim sejenis juga dapat memperluas penelitian selanjutnya.

DAFTAR REFERENSI

Arifiyanti, A. A., Shantika, N. R., & Syafira, A. O. (2022). *ANALISIS SENTIMEN ULASAN PENGGUNA BSI MOBILE PADA*. 283–288. https://doi.org/10.33795/jip.v9i3.1003

Cahyani, N., & Muslim, M. A. (2020). Increasing Accuracy of C4 . 5 Algorithm by Applying Discretization and Correlation-based Feature Selection for Chronic Kidney Disease Diagnosis. *Scientific Journal of Informatics*, *2018*(March), 1–10. https://jtec.utem.edu.my/jtec/article/view/4922

Dometian, C., & Jatmiko, S. (2025). *Analisis Sentimen Ulasan Google Play Store : Studi Komparatif Algoritma SVM , Naïve Bayes , dan Logistic Regression*. *15*(3), 610–620. https://doi.org/10.37859/jf.v15i3.10016

Fachrurrozi, S., Muljono, Shidik, G. F., Fanani, A. Z., Purwanto, & Zami, F. Al. (2021). Increasing Accuracy of Support Vector Machine (SVM) by Applying N-Gram and Chi-

Square Feature Selection for Text Classification. *Proceedings - 2021 International Seminar on Application for Technology of Information and Communication: IT Opportunities and Creativities for Digital Innovation and Communication within Global Pandemic, ISemantic 2021*, 42–47. https://doi.org/10.1109/iSemantic52711.2021.9573210

Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, *36*(1), 20–38. https://doi.org/10.1016/j.ijresmar.2018.09.009

Julianto, I. T., Kurniadi, D., Septiana, Y., & Sutedi, A. (2023). *ALTERNATIVE TEXT PRE-PROCESSING USING CHAT GPT OPEN AI*. *12*(1), 67–77. https://doi.org/10.23887/janapati.v12i1.59746

K, S., & Thilagam, P. S. (2022). Multi-layer perceptron based fake news classification using knowledge base triples. *Applied Intelligence*. https://doi.org/10.1007/s10489-022-03627-9

Kumar, K., Harish, B. S., & Darshan, H. K. (2019). Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method. *International Journal of Interactive Multimedia and Artificial Intelligence*, *5*(5), 109. https://doi.org/10.9781/ijimai.2018.12.005

Maharani, V. P., Harvanny, K., Samudra, D., Muthoharoh, L., Satria, A., Clinton, M., & Manullang, T. (2026). *S ENTIMENT A NALYSIS OF M OBILE L EGENDS A PP R EVIEWS U SING M ACHINE L EARNING AND LSTM-B ASED D EEP*. arXiv preprint arXiv:2605.01317. https://doi.org/10.48550/arXiv.2605.01317

Nawulansih, D. F., Santi, N. C., Aristia, I., Informatika, T., Nahdlatul, U., Sunan, U., Bojonegoro, G., Informasi, S., Nahdlatul, U., Sunan, U., & Bojonegoro, G. (2025). *Analisis Sentimen Ulasan Aplikasi DANA di Google Play Store : Penerapan Support Vector Machine dan Synthetic Minority Over-sampling Technique Sentiment Analysis of DANA Application Reviews on Google Play Store : Implementation of Support Vector Machine an*. *5*(9), 2660–2671. https://doi.org/10.52436/1.jpti.1053

No, V., Najibulloh, I. K., Intan, D., & Saputra, S. (2025). *Edumatic : Jurnal Pendidikan Informatika Analisis Sentimen Ulasan Co-Pilot Google Play dengan SVM , Neural Network , dan Decision Tree*. *9*(1), 275–283. https://doi.org/10.29408/edumatic.v9i1.29673

Nurfauzi, O. M., Hilabi, S. S., Nurapriani, F., & Huda, B. (2025). Analisis Sentimen, Grab Indonesi Analisis Sentimen Grab Indonesia Pada Ulasan Google Play Store Menggunakan Algoritma Naïve Bayes Dan SVM. *SMARTICS Journal*, *11*(1), 8–13. https://doi.org/10.21067/smartics.v11i1.11789

Setyani, T., Sari, K., Heidy, H. N., & Suryono, R. R. (2026). *Sentiment Analysis of Jamsostek Mobile Application Reviews on Google Play Store Using Support Vector Machine and Naive Bayes Algorithms Analisis Sentimen Pengguna Aplikasi Jamsostek Mobile Berdasarkan Ulasan Google Play Store Menggunakan Algoritma Support*. *6*(January), 373–384. https://doi.org/10.57152/malcom.v6i1.2526

Zablith, F., & Osman, I. H. (2020). ReviewModus : Text classification and sentiment prediction of unstructured reviews using a hybrid combination of machine learning and evaluation models. *Applied Mathematical Modelling*, *71*(2019), 569–583. https://doi.org/10.1016/j.apm.2019.02.032