

PENERAPAN METODE *STACKING ENSEMBLE* UNTUK ANALISIS SENTIMEN PADA ULASAN APLIKASI RUANGGURU

Detin Sofia^{1*}, Pina Sekarpuji²

¹Teknik Informatika, Fakultas Teknologi Informasi dan Komunikasi, Institut Teknologi dan Bisnis Bina Sarana Global, Tangerang, Indonesia

²Direktorat Umum - Kelembagaan, Televisi Republik Indonesia (TVRI), Jakarta, Indonesia

Email: ^{1*} detinsofia@global.ac.id, ² pina.skrpj@gmail.com

(* : corresponding author)

Abstrak-Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi Ruangguru berdasarkan ulasan di *Google Play Store*. Studi ini berupaya untuk menggambarkan persepsi masyarakat terhadap layanan pembelajaran digital tersebut. Data berupa 99.000 ulasan pengguna dalam bahasa Indonesia dikumpulkan melalui teknik *web scraping*. Analisis sentiment dilakukan menggunakan pendekatan klasifikasi sentimen berbasis pembelajaran mesin. Penelitian mengikuti tahapan metodologi *Cross Industry Standard Process for Data Mining* (CRISP-DM), yang terdiri dari *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation* dan *Deployment*. Sentimen dikategorikan menjadi tiga kelas: positif, negatif dan netral. Proses pra-pemrosesan data melibatkan tahapan seperti pembersihan, tokenisasi, normalisasi. Ekstraksi fitur menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), serta pelabelan data yang dilakukan dengan pendekatan *lexicon-based* menggunakan kamus kata positif dan negatif yang diunduh dari sumber terbuka. Model klasifikasi yang digunakan dalam penelitian ini adalah *Stacking Ensemble* dengan *Random Forest*, *Support Vector Machine* (SVM), dan *Extreme Gradient Boosting* (XGBoost) sebagai *base learner*, serta *Logistic Regression* sebagai *meta learner*. Evaluasi performa model menggunakan *confusion matrix* dan *Area Under the Curve* (AUC). Model ini menghasilkan Akurasi 88%, *Precision* 87%, *Recall* 88% dan *F1-Score* 87% serta AUC 0,945. Hasil penelitian menunjukkan bahwa mayoritas ulasan memiliki sentimen positif terhadap aplikasi Ruangguru. Temuan ini dapat menjadi masukan strategis bagi pengembang dalam meningkatkan kualitas layanan dan kepuasan pengguna.

Kata Kunci: *Analisis Sentimen, CRISP-DM, Pembelajaran Mesin, Stacking Ensemble, TF-IDF*

Abstract-This research aims to analyze user sentiment towards the Ruangguru application based on reviews from the *Google Play Store*. The study seeks to illustrate public perception of this digital learning service. A total of 99,000 user reviews in Indonesian were collected through *web scraping* techniques. Sentiment analysis was conducted using a sentiment classification approach based on machine learning. The research use the *Cross Industry Standard Process for Data Mining* (CRISP-DM) methodology, which includes the stages of *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, and *Deployment*. Sentiments were categorized into three classes: positive, negative, and neutral. The data preprocessing process involved steps such as cleaning, tokenization, and normalization. Feature extraction was performed using *Term Frequency-Inverse Document Frequency* (TF-IDF), and data labeling was carried out using a *lexicon-based* approach with positive and negative word dictionaries downloaded from open sources. The classification model used in this study was a *Stacking Ensemble* with *Random Forest*, *Support Vector Machine* (SVM), and *Extreme Gradient Boosting* (XGBoost) as base learners, and *Logistic Regression* as the meta-learner. Model performance was evaluated using a *confusion matrix* and *Area Under the Curve* (AUC). The model achieved an Accuracy of 88%, Precision of 87%, Recall of 88%, F1-Score of 87%, and an AUC of 0.945. The results indicate that the majority of reviews expressed a positive sentiment toward the Ruangguru application. These findings can serve as strategic input for developers to improve service quality and user satisfaction.

Keywords: *CRISP-DM, Machine Learning, Sentiment Analysis, Stacking Ensemble, TF-IDF*

1. PENDAHULUAN

Transformasi digital dalam bidang pendidikan telah memberikan dampak yang signifikan, terutama melalui kemunculan berbagai *platform* pembelajaran daring. Saat ini terdapat berbagai pilihan *platform* untuk belajar daring seperti Aplikasi Ruangguru. Berdasarkan *Annual Report* Ruangguru 2022 [1] terdapat lebih dari 42 juta pengguna yang terdaftar mengikut program pembelajaran dalam aplikasi tersebut. Hal itu membuktikan bahwa Ruangguru merupakan aplikasi yang memiliki banyak pengguna di Indonesia. Aplikasi ini menyediakan layanan belajar daring bagi pelajar dari berbagai jenjang pendidikan dengan pilihan program pembelajaran. Namun, seperti aplikasi lain yang berbasis teknologi, kualitas dan keberhasilan layanan sangat ditentukan oleh respons dan kepuasan pengguna. Ulasan dan komentar pengguna yang dapat diakses secara terbuka di *platform Google Play Store* saat mengunduh aplikasi Ruangguru dapat dijadikan sumber data untuk mengevaluasi persepsi dan sentimen terhadap layanan tersebut.

Urgensi dari penelitian ini terletak pada perlunya pemahaman yang lebih mendalam terhadap masukan dan opini pengguna, yang jika dianalisis secara sistematis dapat memberikan informasi penting bagi pengembang aplikasi dalam melakukan perbaikan berkelanjutan. Analisis sentimen pengguna tidak dapat dilakukan secara manual, sehingga diperlukan pendekatan berbasis komputasi yang efisien dan akurat untuk mengekstraksi informasi dari data teks. Penelitian terdahulu terkait analisis sentimen pada aplikasi ruangguru juga sudah banyak dilakukan memanfaatkan

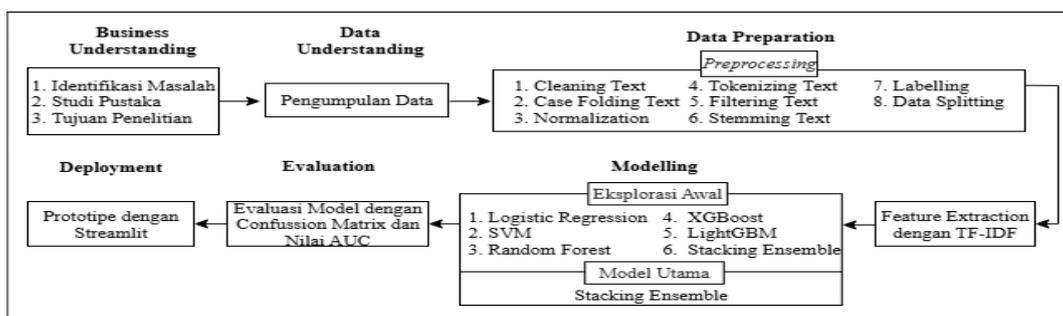
berbagai algoritma yaitu *Support Vector Machine* (SVM) [2], [3] *Naive Bayes* [4], [5]. Terdapat pula penelitian serupa dengan perbandingan beberapa algoritma seperti SVM dan *Decision Tree* [6] kemudian *Naive Bayes* dan SVM [7] kedua penelitian tersebut juga membahas analisis sentimen dengan data serupa yaitu Ulasan Aplikasi Ruangguru yang didapatkan di *Google Playstore*, hasil perbandingan algoritma SVM memiliki tingkat akurasi yang lebih tinggi dibanding algoritma yang lainnya. Penelitian dengan topik analisis sentimen menggunakan algoritma SVM juga dilakukan oleh [8], [9].

Kemudian analisis sentimen menggunakan algoritma lainnya seperti perbandingan *Convolutional Neural Network* (CNN), *K-Nearest Neighbor* (KNN) dan *Logistic Regression* [10]. Penelitian lain juga membandingkan KNN dan *Logistic Regression* [11] hasilnya *Logistic Regression* memperoleh akurasi yang lebih unggul dibanding algoritma lainnya. Pada penelitian [12] terkait analisis sentimen membandingkan algoritma *Random Forest*, *Extreme Gradient Boosting* (XGBoost) and *Light Gradient Boosting Machine* (LightGBM) mendapatkan hasil algoritma terbaik yaitu *Random Forest*, sedangkan pada penelitian [13] membandingkan algoritma *Multi Layer Perceptron* (MLP), XGBoost, dan LightGBM mendapatkan algoritma terbaik XGBoost dengan teknik tambahan *Synthetic Minority Over-Sampling Technique* (SMOTE) untuk penanganan dataset tidak seimbang. Penelitian ini tidak menggunakan teknik tersebut, hanya menggunakan data asli dari hasil *web scrapping*. Berdasarkan penelitian terdahulu terkait analisis sentimen pada aplikasi ruangguru sebagian besar penelitian masih menggunakan pendekatan klasifikasi tunggal yang memiliki keterbatasan dalam hal akurasi dan kemampuan generalisasi. Oleh karena itu, dibutuhkan pendekatan yang lebih kuat dan adaptif, seperti metode *Stacked Generalization ensemble* atau dikenal dengan metode *Stacking Ensemble*, untuk menghasilkan analisis sentimen yang lebih akurat dan dapat diandalkan sebagai dasar pengambilan keputusan berbasis data. Hal tersebut didasarkan pada penelitian yang dilakukan para peneliti yang sudah menggunakan metode *Stacking Ensemble* seperti pada penelitian [14], [15], [16], [17], [18] yang menunjukkan bahwa metode *Stacking Ensemble* merupakan pendekatan yang efektif dalam analisis sentimen, khususnya pada data yang berasal dari jaringan sosial dengan menggabungkan beberapa model klasifikasi, model ini mampu meningkatkan akurasi prediksi secara signifikan dibandingkan model tunggal. Selain itu, *Stacking Ensemble* memberikan wawasan yang lebih mendalam terhadap opini pengguna. Maka penggunaan teknik ini tidak hanya memperkuat keandalan model klasifikasi sentimen, tetapi juga memperluas pemahaman terhadap persepsi publik secara lebih komprehensif dan akurat.

Beberapa penelitian terdahulu menunjukkan bahwa salah satu metode pemrosesan bahasa alami adalah analisis sentimen, yang bertujuan untuk mengklasifikasikan pendapat menjadi sentimen positif, negatif dan netral. Berbagai metode telah digunakan dalam analisis sentimen, termasuk *Logistic Regression*, SVM, *Random Forest*, XGBoost, LightGBM serta *Stacking Ensemble*. Penelitian ini juga akan melakukan proses *hyperparameter tuning* untuk mendapatkan hasil akurasi terbaik. Proses penelitian dimulai dengan pengumpulan ulasan pengguna aplikasi Ruangguru dari Google Play Store, dilanjutkan dengan praproses data seperti pembersihan teks, tokenisasi, dan pelabelan sentimen. Data kemudian dianalisis menggunakan beberapa algoritma untuk membangun model klasifikasi terbaik. Evaluasi dilakukan dengan confusion matrix, dan prototipe sederhana dirancang untuk mengimplementasikan model terbaik. Tujuan penelitian ini adalah memahami pola persepsi pengguna terhadap aplikasi Ruangguru serta mengevaluasi performa algoritma dalam klasifikasi sentimen. Hasilnya diharapkan dapat membantu pengembang meningkatkan pengalaman dan layanan pengguna.

2. METODE PENELITIAN

Metode penelitian memberikan penjelasan tahapan yang dilakukan secara sistematis dalam menyelesaikan masalah yang diteliti. Kerangka kerja utama yang digunakan adalah *Cross-Industry Standard Process for Data Mining* (CRISP-DM). CRISP-DM disusun dalam enam fase yang bersifat iteratif dan fleksibel, sehingga banyak digunakan dalam berbagai proyek data mining lintas industri. Menurut [19] CRISP-DM masih menjadi kerangka kerja paling populer dalam pelaksanaan proyek data science saat ini. Gambar 1 memperlihatkan implementasi CRISP-DM dalam penelitian ini, yang dijelaskan lebih lanjut pada bagian berikutnya.



Gambar 1. Implementasi Metode CRISP-DM

2.1. Business Understanding

Tahapan yang dimulai dengan identifikasi permasalahan melalui studi pustaka terhadap penelitian-penelitian terdahulu untuk merumuskan tujuan penelitian. Permasalahan utama yang dianalisis adalah bahwa ulasan pengguna tentang aplikasi Ruangguru yang tersedia di *Google Play Store* bersifat masif dan tidak terstruktur, sehingga sulit dianalisis secara manual untuk mendukung pengambilan keputusan strategis. Oleh karena itu, tujuan dari penelitian ini adalah untuk membuat model analisis sentimen berbasis *machine learning* yang dapat mengkategorikan pendapat pengguna ke dalam kategori positif, negatif, dan netral. Hasil dari model ini diharapkan dapat membantu pengembang meningkatkan kualitas layanan aplikasi secara berkelanjutan.

2.2. Data Understanding

Proses ini diawali dengan pengumpulan data dari *Google Play Store* terkait ulasan pengguna Aplikasi Ruangguru, dengan menggunakan teknik *web scraping*. Proses *scraping* dilakukan melalui *Google Colaboratory* menggunakan bahasa pemrograman *Python*, serta *library google-play-scrap*, dengan memasukkan ID aplikasi Ruangguru, yaitu *id.co.ruangguru.livestudents*. Teknik ini digunakan untuk memperoleh data ulasan berupa teks beserta atribut pendukung lainnya seperti rating, tanggal ulasan, dan nama pengguna. Hasil *scraping* menghasilkan sebanyak 99.000 baris data dengan total 11 kolom, yang selanjutnya digunakan dalam proses analisis sentimen.

2.3. Data Preparation

Pada tahap ini, proses pembersihan data dilakukan karena data yang diperoleh merupakan data mentah yang tidak dapat digunakan secara langsung dalam proses klasifikasi analisis sentimen. *Preprocessing*, tahap penting dalam data mining, berfungsi untuk membersihkan dan memperbaiki data mentah, mempersiapkannya untuk memenuhi standar metode analisis, dan menghasilkan analisis yang akurat dan dapat diandalkan. [20] Pada penelitian ini tahap *preprocessing* yang dilakukan adalah sebagai berikut:

- Cleaning Text*, proses pembersihan teks seperti penghapusan simbol, tautan, dan token tidak penting secara langsung memengaruhi kualitas hasil klasifikasi sentimen. [21]
- Case Folding Text*, setiap huruf dalam teks diubah menjadi huruf kecil. Ini membantu menyamakan format kata, mencegah duplikasi fitur dan membuat langkah berikutnya lebih mudah. [22]
- Normalization* merupakan proses normalisasi dengan fungsi *slangwords* yang dilakukan dengan mengganti kata-kata slang atau tidak formal dalam kalimat menjadi kata-kata yang lebih umum atau sesuai dengan kamus yang tersedia. Tahapan ini bertujuan untuk memastikan konsistensi dalam gaya penulisan kata atau frasa spesifik. [23]
- Tokenizing Text*, langkah penting dalam analisis sentimen karena membagi teks menjadi bagian yang lebih kecil yang disebut token, yang biasanya terdiri dari kata atau sub-kata. Proses ini mengubah data teks yang tidak terstruktur menjadi format yang terstruktur yang dapat dianalisis dengan algoritma *machine learning*. [24]
- Filtering Text* merupakan proses penghapusan *stopwords* yang dimaksudkan untuk menghilangkan kata-kata umum yang tidak memberikan informasi yang signifikan untuk analisis teks. [25]
- Stemming Text* merupakan langkah tambahan yang dilakukan setelah penyaringan untuk mengumpulkan kata kunci dari setiap token, ini dapat meningkatkan akurasi model klasifikasi. [26]
- Labelling* merupakan salah satu bagian dari proses pembuatan data latihan dan data uji untuk analisis sentimen menggunakan metode *machine learning*. [27]
- Data Splitting* merupakan pembagian dataset antara data pelatihan dan data pengujian. Membagi data dengan benar antara data pelatihan dan data pengujian sangat penting untuk memastikan bahwa model yang dibuat bekerja dengan baik dan tidak *overfitting*. [28]
- Feature Extraction* dengan *Term Frequency-Inverse Document Frequency* (TF-IDF), metode pemrosesan teks digunakan untuk menilai tingkat pentingnya suatu kata dalam kumpulan dokumen. [29]

2.4. Modelling

Pada tahap ini dilakukan proses eksplorasi awal *modelling* dengan komparasi beberapa algoritma yaitu *Logistic Regression*, *Support Vector Machine* (SVM), *Random Forest*, *Extreme Gradient Boosting* (XGBoost), *Light Gradient Boosting Machine* (LightGBM). Setelah eksplorasi awal dilakukan dengan membandingkan model algoritma tunggal, selanjutnya dilakukan proses *hyperparameter tuning* namun pada proses ini tidak mendapatkan perubahan akurasi yang signifikan maka dilanjutkan dengan metode *Stacking Ensemble*. *Stacking* merupakan teknik yang diusulkan oleh Wolpert (1992) memiliki konsep menggabungkan beberapa model dalam proses pelatihan. Konsep ini bertujuan untuk meningkatkan akurasi suatu model dalam melakukan estimasi. [30] Skema *stacking* terdiri dari dua tingkat: tingkat 0 dan tingkat 1. Pada tingkat 0 disebut juga *base learner* atau model dasar dilatih menggunakan data pelatihan, dan keluaran dari model-model tersebut disimpan sebagai fitur baru dalam dataset baru. Selanjutnya, data validasi diberikan ke model dasar, dan prediksi mereka digunakan sebagai input untuk *meta-model*. Ini membentuk tingkat 1, disebut juga *meta learner* atau model penggabung di mana *meta-model* dilatih menggunakan data validasi dengan keluaran dari model dasar sebagai fitur input dengan tujuan menggabungkan prediksi tersebut menjadi prediksi akhir yang lebih baik. [31]

2.5. Evaluation

Pada tahap ini, model dievaluasi menggunakan *confusion matrix* yang dihasilkan dari hasil prediksi pada data uji. Evaluasi dilakukan setelah model dilatih dengan membagi data latih dan data uji menjadi dua bagian 80:20. Metode ini biasa digunakan untuk menilai kemampuan model untuk generalisasi terhadap data baru. Hasil prediksi sentimen, baik yang benar maupun yang salah, dibagi menjadi tiga klasifikasi yaitu positif, negative dan netral. Semua hasil ini digambarkan dalam *confusion matrix*. *Confusion matrix* adalah tabel $m \times m$ yang menggambarkan kinerja pengklasifikasi dengan membandingkan label prediksi dan label aktual [20]. Pada *confusion matrix* terdapat empat istilah yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)* dan *False Negative (FN)*. Nilai TP merupakan data positif yang terdeteksi benar positif, sedangkan TN adalah data negatif yang terdeteksi benar negatif. FP merupakan data negatif yang terdeteksi sebagai data positif, sedangkan FN adalah data positif yang terdeteksi sebagai negatif. Berdasarkan nilai TP, TN, FP dan FN diperoleh nilai akurasi, precision dan *recall* kemudian gabungan antara *precision* dan *recall* menghasilkan nilai *F1-score*.

Selain itu, evaluasi model juga dilakukan dengan metrik *Area Under the Curve (AUC)* yaitu ukuran numerik yang dipakai sebagai pembandingan kinerja model dan sebagai tolak ukur seberapa berhasil sekaligus benar peringkat dari model yang dibandingkan dengan menggunakan pemisahan antara rate positif dan negatif [32]. Nilai AUC yang tinggi menggambarkan bahwa model memiliki kinerja klasifikasi yang cukup baik dan stabil terhadap data uji.

2.6. Deployment

Pada tahap ini, proses perancangan prototipe sederhana dilakukan untuk menampilkan hasil dari proses *modeling* yang telah dilakukan dengan data baru, dengan tujuan untuk memungkinkan pengguna melihat hasil tampilan secara langsung. Aplikasi prototipe ini dirancang khusus untuk menampilkan hasil analisis sentimen berdasarkan ulasan pengguna aplikasi Ruangguru, sehingga pengguna dapat melihat secara langsung apakah sentimen positif, negatif, atau netral dari model. Untuk keperluan ini digunakan *Streamlit*, sebuah *framework Python* berbasis web yang memungkinkan pembuatan antarmuka interaktif dengan cepat. Prototipe menampilkan hasil klasifikasi sentimen dari teks input pengguna dengan tampilan visual sederhana, sehingga mempermudah proses evaluasi hasil model secara praktis dan aplikatif.

3. HASIL DAN PEMBAHASAN

3.1. Business Understanding

Berdasarkan hasil klasifikasi sentimen terhadap ulasan pengguna aplikasi Ruangguru, diperoleh bahwa mayoritas ulasan tergolong dalam kategori positif, diikuti oleh sentimen netral dan negatif. Temuan ini mendukung tujuan awal penelitian untuk memahami sudut pandang pengguna terhadap layanan aplikasi secara sistematis. Model analisis sentimen yang dibangun tidak hanya mampu mengelompokkan opini pengguna, tetapi juga memberikan informasi yang dapat dimanfaatkan oleh tim pengembang untuk mengidentifikasi aspek yang paling diapresiasi maupun yang perlu diperbaiki. Dengan demikian, pendekatan ini memberikan solusi terhadap permasalahan awal, yaitu bagaimana mengekstrak wawasan bermakna dari data ulasan yang tidak terstruktur dan dalam jumlah besar.

3.2. Data Understanding

Ulasan pengguna aplikasi Ruangguru yang diperoleh melalui *platform Google Play Store* adalah sumber data yang digunakan dalam penelitian ini. Dataset awal terdiri dari beberapa atribut seperti *reviewId*, *userName*, *userImage*, *content*, *Score*, *thumbsUpCount*, *reviewCreatedVersion*, *at*, *replyContent*, *repliedAt*, dan *appVersion*. Namun, atribut-atribut lain yang tidak relevan akan dihapus karena fokus analisis berada di atribut *content*, yang berisi teks ulasan. Hasil eksplorasi awal menunjukkan bahwa sebagian besar ulasan ditulis dalam bahasa Indonesia informal, dengan panjang teks yang bervariasi. Beberapa ulasan mengandung pujian eksplisit seperti “bagus banget”, “membantu”, dan “bermanfaat”, sementara yang lain memuat kritik dalam bahasa halus. Data tersebut belum memiliki label sentimen, maka analisis belum dapat mengaitkan pola kata dengan kategori tertentu secara pasti. Oleh karena itu, diperlukan tahapan *preprocessing* seperti *case folding*, penghapusan tanda baca, *tokenization* dan *stemming* sebelum proses *labeling* dan pemodelan dilakukan. Temuan ini menunjukkan bahwa data bersifat tidak terstruktur dan membutuhkan pengolahan lanjutan agar dapat digunakan secara efektif dalam model klasifikasi sentimen. Tabel 1 menyajikan beberapa atribut gambaran data mentah hasil *web scraping*.

Tabel 1. Data Mentah Hasil Scrapping

userName	Content	Score	at
Pengguna Google	animasinya lucu banget, cocok buat saya yang gaya belajarnya visual, penjelasan kakaknya juga mudah dipahami, apa lagi ada fitur ujian,..	5	31/07/2024 12:41
Pengguna Google	Apk nya baguss bangettt,, tapi aku ada sedikit kritik untuk tampilan layar nya kalau di web kan untuk yg utbk, artikel ² , dan beberapa fitur itu ada..	5	13/07/2024 16:33

3.3. Data Preparation

Proses pembersihan data dilakukan karena data yang diperoleh merupakan data mentah yang tidak dapat digunakan secara langsung dalam proses analisis sentimen. Tahap *preprocessing* dalam pengolahan data sangat penting untuk membersihkan dan memperbaiki data mentah, membuatnya sesuai dengan persyaratan metode analisis dan menghasilkan analisis yang akurat dan dapat diandalkan[20].

a. Preprocessing

Di awal fase ini, teks dibersihkan dari elemen yang tidak penting seperti *mention*, *hashtag*, tautan *link*, angka, simbol dan tanda baca. Selanjutnya, *case folding* dilakukan untuk memotong seluruh huruf menjadi huruf kecil agar konsisten dalam pemrosesan. serta *normalization* adalah proses mengubah kata-kata yang tidak baku atau *slang* menjadi kata baku yang sesuai dengan kamus referensi. Tabel 2 menyajikan hasil dari proses tersebut.

Tabel 2. Data Hasil Cleaning Text, Case Folding dan Normalization

Content	Cleaned Text	Case Folding	Normalization
animasinya lucu banget,	animasinya lucu banget	animasinya lucu banget	animasinya lucu banget
cocok buat saya yang ...	cocok buat saya yang g...	cocok buat saya yang g...	cocok buat saya yang g...
Apk nya baguss banget,,	Apk nya baguss bangettt	apk nya baguss bangettt	apk nya baguss bangettt
tapi aku ada sedikit...	tapi aku ada sedikit k...	tapi aku ada sedikit k...	tapi aku ada sedikit k...

Kemudian Tabel 3 merupakan hasil dari Proses *tokenizing*, yaitu memecah kalimat menjadi daftar kata. Kemudian proses *filtering* untuk menghilangkan kata-kata *stopwords* biasa yang tidak penting untuk analisis sentimen. Selanjutnya, *stemming* dilakukan untuk mengembalikan setiap kata ke bentuk aslinya.

Tabel 3. Data Hasil Tokenizing Text, Filtering dan Stemming

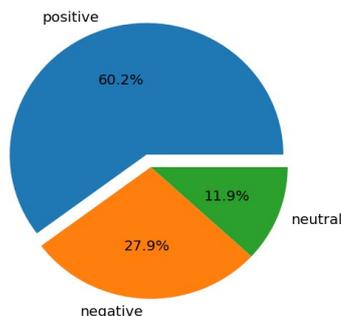
Tokenizing Text	Filtering	Stemmed	Text Akhir
[animasinya, lucu, banget,	[animasinya, lucu, banget,	animasi lucu banget	animasinya lucu banget
cocok, buat, saya,...	cocok, gaya, belaj...	gaya ajar visual jelas...	cocok gaya belajarnya ...
[apk, nya, baguss, banget,	[apk, baguss, banget, kritik,	apk baguss banget kritik	apk baguss banget kritik
tapi, aku, ada, s...	tampilan, laya...	tampil layar web utbk..	tampilan layar web ...

Setiap tahapan *preprocessing* disimpan ke dalam kolom baru agar hasil setiap transformasi dapat dilacak dan dievaluasi. Setelah seluruh proses selesai, jumlah atribut pada dataset menjadi 8 kolom utama dari 11 kolom awal. Berdasarkan hasil tersebut hanya kolom *text_akhir* saja yang akan digunakan pada proses *modelling*, selanjutnya perlu dilakukan proses *labelling* untuk dapat mengklasifikasikan ulasan kalimat.

b. Labelling

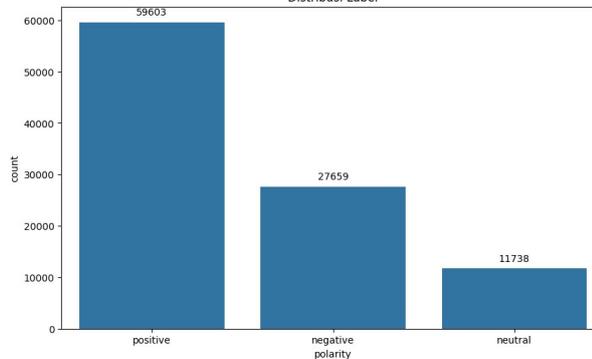
Data yang didapatkan belum memiliki label, maka pada tahap ini dilakukan proses *labelling* untuk memberikan label positif, negatif dan netral. Penelitian ini menggunakan pendekatan berbasis lexicon untuk mendukung proses klasifikasi sentimen. Salah satu komponen pentingnya adalah kamus kata-kata positif dan negatif, yang memuat daftar kosakata bernuansa positif dan negatif beserta bobot nilainya. Kamus ini diperoleh dari sumber terbuka yang diunggah pada platform *GitHub* dalam format *Comma-Separated Values* (CSV). Proses pengambilan data dilakukan melalui permintaan *HTTP* ke alamat *URL* file tersebut. Setelah data berhasil diakses, file CSV kemudian diparsing sebagai struktur tabular, di mana setiap baris terdiri dari dua elemen: kata dan skor sentimennya. Kemudian dilakukan klasifikasi sentimen dengan *polarity* dengan penentuan skor untuk label positif dan negatif, apabila tidak sesuai ketentuan skor maka termasuk ke dalam label netral. Gambar 2 dan Gambar 3 menunjukkan hasil klasifikasi positif, negatif dan netral.

Polaritas Sentimen pada Data Ulasan



Gambar 2. Polaritas Sentimen Data Ulasan

Distribusi Label



Gambar 3. Distribusi Label

Berdasarkan hasil *labelling* dapat dilihat bahwa jumlah data ulasan dengan sentimen positif memiliki persentase tertinggi yaitu sebesar 60,2% dengan jumlah distribusi label 59.603. Hal tersebut menandakan bahwa mayoritas

pengguna memberikan tanggapan yang positif terhadap aplikasi Ruangguru. Ulasan positif ini umumnya mencerminkan kepuasan terhadap fitur-fitur pembelajaran yang disediakan, kualitas materi, atau kemudahan penggunaan aplikasi. Di sisi lain, persentase sentimen negatif dan netral yang lebih rendah menunjukkan bahwa keluhan atau ketidakpuasan pengguna tidak terlalu dominan, meskipun tetap perlu dianalisis untuk perbaikan layanan di masa mendatang. Temuan ini dapat menjadi indikator awal bahwa aplikasi Ruangguru telah memenuhi sebagian besar ekspektasi penggunanya, sekaligus menjadi dasar bagi pengembang untuk terus meningkatkan aspek-aspek yang masih dinilai kurang oleh sebagian pengguna. Pada Gambar 4 disajikan hasil visualisasi *WordCloud* dari data Ulasan Positif, Negatif dan Netral.



Gambar 4. *Word Cloud* dari Data Ulasan Positif, Negatif dan Netral

Setelah melalui tahapan *preprocessing* seperti pembersihan teks, normalisasi dan *stemming*, diperoleh data bersih pada kolom *text_akhir* yang siap untuk dimodelkan. Kolom ini merepresentasikan ulasan pengguna dalam bentuk teks yang sudah disederhanakan, sementara kolom *polarity* menunjukkan label sentimen yaitu positif, negatif dan netral. Sebagaimana ditampilkan pada Tabel 4, ulasan dengan kata-kata positif seperti "bagus banget" dan "cocok juara kelas" diklasifikasikan sebagai positif, sedangkan ulasan berisi keluhan seperti masalah koneksi internet dikategorikan sebagai negatif. Hasil ini menunjukkan bahwa data telah siap digunakan dalam tahap pemodelan sentimen.

Tabel 4. Data Bersih Hasil *Preprocessing*

Text Akhir	Polarity
animasinya lucu bangett cocok gaya belajarnya visual penjelasan kakaknya mudah dipahami fitur ujian pembahasan kecepatan ngomongnya disesuaikan diulang kalipun bagus banget cocok juara kelas.	positif
pakai aplikasi ruangguru mempermudah belajar terkadang connection internetnya lambat nonton live teaching connection terdeteksi jelek cek terdeteksi connection internet bagus mohon bantuannya..	negatif

c. *Data Splitting*

Setelah proses *labelling* selesai, data dibagi menjadi dua bagian yaitu data latihan 80% dan data uji 20%. Kemudian dilakukan proses ekstraksi fitur yaitu proses transformasi data teks menjadi representasi numerik, digunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Tujuan dari teknik ini adalah untuk mengetahui seberapa signifikan suatu kata dalam dokumen dibandingkan dengan seluruh korpus, memberikan bobot yang lebih proporsional daripada frekuensi biasa.

3.4. *Modelling*

Pada tahap *modeling*, dilakukan proses pembangunan dan pelatihan model klasifikasi sentimen dengan memanfaatkan algoritma *machine learning*. Beberapa algoritma klasifikasi yang digunakan dalam penelitian ini untuk komparasi yaitu *Logistic Regression*, *Support Vector Machine* (SVM), *Random Forest*, *Extreme Gradient Boosting* (XGBoost), *Light Gradient Boosting Machine* (LightGBM), serta *Stacking Ensemble* diimplementasikan untuk mengidentifikasi sentimen ulasan pengguna dikategori menjadi positif, negatif, dan netral.

Gambar 5 merupakan hasil perbandingan komparasi model, *Logistic Regression* menunjukkan kinerja terbaik dengan akurasi data uji sebesar 86,8% dan nilai AUC 0,931, mencerminkan hasil yang stabil tanpa *overfitting* karena akurasi data latihnya juga cukup dekat yaitu 87,3%. LightGBM mencatat nilai AUC tertinggi yaitu 0,934, namun akurasi data uji sedikit lebih rendah sebesar 86,2%, meskipun akurasi data latihnya cukup tinggi 87,8%. SVM dengan *kernel Linear* juga memberikan hasil yang kompetitif dengan akurasi data uji 86,6% dan AUC 0,919, serta akurasi data latih sebesar 87%, yang menunjukkan keseimbangan cukup baik. Di sisi lain, *Random Forest* dan XGBoost menunjukkan selisih yang cukup signifikan antara akurasi data latih dan uji, dengan *Random Forest* memiliki akurasi latih sangat tinggi yaitu 96,2% tetapi akurasi uji hanya 85,2%, sedangkan XGBoost dengan akurasi latih 88,3% dan akurasi uji 85,7%, yang mengindikasikan *overfitting* dan kurangnya kemampuan generalisasi terhadap data baru.

Rangkuman Akurasi dan AUC Semua Model:

Model Algoritma	Akurasi Data Latih	Akurasi Data Uji	Nilai AUC
Logistic Regression	0,873	0,868	0,931
SVM (LinearSVC)	0,870	0,866	0,919
LightGBM	0,878	0,862	0,934
XGBoost	0,883	0,857	0,932
Random Forest	0,962	0,852	0,926

Gambar 5. Komparasi Model

Penelitian ini juga melakukan teknik *hyperparameter tuning* untuk memperoleh akurasi terbaik berdasarkan kombinasi parameter. Namun pada penelitian ini hasil akurasi yang diperoleh tidak menunjukkan perbedaan signifikan dibandingkan dengan model tanpa *hyperparameter tuning*, bahkan memerlukan waktu eksekusi yang lebih lama karena proses pencarian parameter terbaik tersebut. Oleh karena itu, dilakukan tinjauan terhadap beberapa studi pustaka yang membahas metode peningkatan akurasi model. Berdasarkan hasil tinjauan tersebut, ditemukan bahwa beberapa penelitian menggunakan metode *stacking ensemble*, yaitu metode yang menggabungkan beberapa algoritma untuk meningkatkan performa prediksi. Gambar 6 menunjukkan hasil penerapan metode *stacking ensemble* dalam penelitian ini

Rangkuman Akurasi dan AUC Semua Model dengan Stacking Ensemble:

Model Algoritma	Akurasi Data Latih	Akurasi Data Uji	Nilai AUC
Stacking Ensemble 1	0,915	0,875	0,945
Stacking Ensemble 2	0,893	0,870	0,943
Logistic Regression	0,873	0,868	0,931
SVM (LinearSVC)	0,870	0,866	0,919
LightGBM	0,878	0,862	0,934
XGBoost	0,883	0,857	0,932
Random Forest	0,962	0,852	0,926

Gambar 6. Hasil Penerapan Metode *Stacking Ensemble*

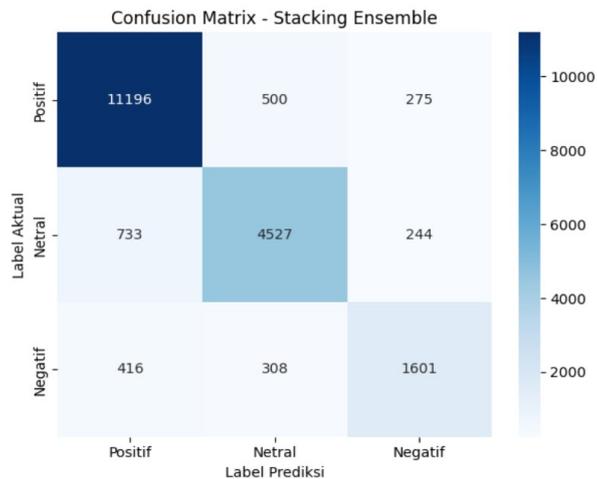
Stacking Ensemble 1 dibangun dengan menggabungkan *Random Forest*, SVM, dan XGBoost sebagai *base learner*, serta *Logistic Regression* sebagai *meta learner*. Sementara itu, *Stacking Ensemble 2* menggunakan SVM, LightGBM, dan XGBoost sebagai *base learner*, dengan *meta learner* yang sama. Hasil evaluasi menggunakan akurasi dan *Area Under the Curve (AUC)* menunjukkan bahwa *Stacking Ensemble 1* memberikan kinerja terbaik pada akurasi data uji yaitu 87,5% dan nilai AUC 0,945. Meskipun hasil ini tidak berbeda jauh dari *Stacking Ensemble 2*, *Stacking Ensemble 1* menunjukkan akurasi data latih yang lebih tinggi yaitu 91,5%. Hasil tersebut memberikan bukti bahwa metode *stacking ensemble* dapat meningkatkan performa model, baik dari segi akurasi maupun AUC, dibandingkan dengan model individual. Namun, metode *stacking ensemble* juga memiliki beberapa kelemahan, seperti peningkatan kompleksitas model, waktu pelatihan yang lebih lama dan potensi *overfitting* pada *meta learner* jika tidak diatur dengan benar. Selain itu, interpretasi hasil prediksi menjadi lebih sulit dibandingkan model yang lebih sederhana.

3.5. Evaluation

Model *Stacking Ensemble 1* dibangun dengan menggabungkan beberapa *base learner*, yaitu *Random Forest*, *Support Vector Machine (SVM)* dan XGBoost dengan *Logistic Regression* sebagai *meta learner*. Pendekatan *stacking* ini bertujuan untuk mengoptimalkan prediksi dengan mengombinasikan keunggulan dari masing-masing algoritma. Model ini dilatih menggunakan skema pembagian data 80:20 dan dievaluasi menggunakan *5-fold cross validation* untuk meningkatkan keandalan hasil.

Berdasarkan hasil evaluasi, *Stacking Ensemble 1* menunjukkan performa paling optimal dengan akurasi uji yaitu 87,5% dan AUC sebesar 0,945, menjadikannya model dengan akurasi unggul dalam penelitian ini. Selain metrik umum seperti akurasi dan AUC, performa model juga dievaluasi melalui *confusion matrix*, yang memberikan gambaran detail terkait distribusi prediksi untuk tiap kelas sentimen yaitu positif, negatif, dan netral.

Gambar 7 menunjukkan kemampuan model untuk mengkategorika ulasan berlabel positif dengan sangat baik, ditandai dengan jumlah *True Positive (TP)* yang tinggi. Untuk ulasan netral, masih terdapat kesalahan klasifikasi ke kelas negatif, yang dapat disebabkan oleh ambiguitas dalam ekspresi bahasa pengguna. Sementara itu, sebagian ulasan negatif juga diprediksi sebagai netral, mengindikasikan bahwa model terkadang kesulitan membedakan ekspresi negatif ringan dengan netral.



Gambar 7. Confusion Matrix - Stacking Ensemble

Pada Gambar 8 dapat dilihat nilai keseluruhan metrik yang diperoleh, yaitu Akurasi 88%, *Precision* 87%, *Recall* 88% dan *F1-Score* 87%. Secara keseluruhan, *Stacking Ensemble* 1 tidak hanya unggul dalam metrik kuantitatif, tetapi juga menunjukkan distribusi klasifikasi yang cukup seimbang dan minim kesalahan signifikan, sebagaimana terlihat pada *confusion matrix*. Oleh karena itu, model ini dinilai paling tepat untuk digunakan dalam mengklasifikasikan sentimen ulasan pengguna terhadap aplikasi Ruangguru.

Classification Report:

	precision	recall	f1-score	support
positive	0.91	0.94	0.92	11971
negative	0.85	0.82	0.84	5504
neutral	0.75	0.69	0.72	2325
accuracy			0.88	19800
macro avg	0.84	0.82	0.83	19800
weighted avg	0.87	0.88	0.87	19800

Gambar 8. Classification Report

3.6. Deployment

Tahap ini bertujuan untuk mengimplementasikan model terbaik, yaitu *Stacking Ensemble* 1, ke dalam aplikasi sederhana yang dapat digunakan secara interaktif. Proses ini dilakukan menggunakan *Streamlit*, sebuah *framework Python* yang memungkinkan pembuatan aplikasi *web* secara cepat dan sederhana. Model beserta *pipeline preprocessing* disimpan menggunakan fungsi *joblib*, kemudian diintegrasikan ke dalam aplikasi *Streamlit*.

Pengguna cukup memasukkan ulasan teks dan sistem akan memproses input tersebut serta menampilkan prediksi sentimen secara langsung, apakah termasuk sentimen positif, negatif, dan netral, berdasarkan kalimat yang diberikan.

Prediksi Sentimen Kalimat

Masukkan kalimat:

Aplikasinya sangat membantu belajar di rumah, materi mudah dipahami.

Prediksi Sentimen

Sentimen kalimat adalah POSITIF.

Gambar 9. Hasil Sentimen Positif

Prediksi Sentimen Kalimat ⇄

Masukkan kalimat:

Aplikasinya cukup oke, tapi masih bisa dikembangkan.

Prediksi Sentimen

Sentimen kalimat adalah NETRAL.

Gambar 10. Hasil Sentimen Netral

Gambar 9 dan Gambar 10 menunjukkan hasil pengecekan terhadap model yang telah dibuat dan disimpan, untuk memastikan bahwa model tersebut memiliki kemampuan untuk membuat prediksi yang akurat. Tahap *deployment* ini memungkinkan pemanfaatan hasil analisis sentimen secara langsung, serta memberikan kemudahan bagi pengembang atau pihak terkait untuk memantau persepsi pengguna terhadap aplikasi Ruangguru secara berkelanjutan.

4. KESIMPULAN

Penelitian ini dilakukan untuk mengetahui pandangan pengguna tentang aplikasi Ruangguru melalui analisis sentimen berbasis *machine learning*. Seluruh proses mengikuti tahapan CRISP-DM, mulai dari pemahaman bisnis, pengolahan data teks ulasan, hingga pemodelan dan deployment. Berdasarkan hasil evaluasi, model *Stacking Ensemble 1* yang menggabungkan algoritma *Random Forest*, *Support Vector Machine* (SVM) dan *XGBoost* dengan *Logistic Regression* sebagai meta learner menunjukkan performa terbaik dengan akurasi uji sebesar 88% *Precision* 87%, *Recall* 88% dan *F1-Score* 87% dan AUC sebesar 0,945. Evaluasi menggunakan *confusion matrix* menunjukkan kemampuan model mengklasifikasikan sentimen dengan cukup seimbang, terutama pada kelas positif. Penerapan model ke dalam aplikasi berbasis *Streamlit* membuat pengguna dapat melakukan prediksi sentimen secara langsung dan interaktif. Hasil ini menjawab persoalan utama dalam penelitian, yakni bagaimana membangun sistem klasifikasi sentimen yang akurat dan dapat digunakan untuk mendukung evaluasi layanan aplikasi pembelajaran digital seperti Ruangguru.

Sebagai pengembangan pada penelitian selanjutnya, penelitian ini dapat dilanjutkan dengan menggunakan algoritma *deep learning* yang umumnya dapat memberikan hasil akurasi lebih baik dengan dataset yang kompleks, selain itu penelitian ini juga dapat dilakukan lebih lanjut dengan analisis aspek sentimen untuk mengidentifikasi elemen spesifik aplikasi yang paling banyak mendapat respons positif atau negatif, seperti fitur, layanan guru atau harga. Selain itu, pendekatan *multimodal sentiment analysis* yang menggabungkan teks dengan ulasan suara atau gambar pengguna dapat menjadi arah penelitian potensial dalam konteks platform pembelajaran berbasis digital. Sehingga penelitian ini menghasilkan model yang efektif, namun juga membuka peluang kajian lebih lanjut dalam ranah analisis opini digital yang lebih mendalam dan aplikatif.

DAFTAR PUSTAKA

- [1] Ruangguru, "Annual Impact Report Ruangguru 2022." Accessed: Jun. 14, 2025. [Online]. Available: <https://www.ruangpeduli.org/impact-report>
- [2] R. Darmawan and A. Surahmat, "Optimalisasi Support Vector Machine (SVM) Berbasis Particle Swarm Optimization (PSO) Pada Analisis Sentimen Terhadap Official Account Ruang Guru Di Twitter," *Jurnal Kajian Ilmiah*, vol. 22, no. 2, pp. 1410–9794, 2022, doi: <https://doi.org/10.31599/g0dv0y21>.
- [3] B. Haryanto, "Penerapan Algoritma Support Vector Machine Untuk Analisis Sentimen Review Pengguna Pada Twitter Ruang Guru," Universitas Pendidikan Indonesia, 2023. Accessed: Jun. 25, 2025. [Online]. Available: <https://repository.upi.edu/99721/>
- [4] I. Novitasari, T. B. Kurniawan, D. A. Dewi, and Misinem, "Analisis Sentimen Masyarakat Terhadap Tweet Ruang Guru Menggunakan Algoritma Naive Bayes Classifier (NBC)," *Jurnal Mantik*, vol. 6, no. 3, pp. 2685–4236, 2022, Accessed: Jun. 22, 2025. [Online]. Available: <https://iocscience.org/ejournal/index.php/mantik/article/view/2887>
- [5] H. H. Mubaroroh, H. Yasin, and A. Rusgiyono, "Analisis Sentimen Data Ulasan Aplikasi Ruangguru Pada Situs Google Play Menggunakan Algoritma Naive Bayes Classifier Dengan Normalisasi Kata Levenshtein Distance," *Jurnal Gaussian*, vol. 11, no. 2, pp. 248–257, 2022, doi: <https://doi.org/10.14710/j.gauss.v11i2.35472>.
- [6] I. Nurul Hassanah, S. Faisal, A. Mutoi Siregar, U. Buana Perjuangan Karawang Jl HSRonggo Waluyo, T. Timur, and J. Barat, "Perbandingan Algoritma Support Vector Machine dengan Decision Tree Pada Aplikasi Ruang Guru," *Jurnal Ilmiah Kumpulan jurnal Ilmu Komputer (KLIK)*, vol. 10 No. 01, pp. 39–50, 2023, doi: <http://dx.doi.org/10.20527/klik.v10i1.602>.
- [7] E. Fernando Baharsyah, T. Hasanah Bimastari Aviani, and C. Wulandari, "Analisis Sentimen Pengguna Terhadap Aplikasi Belajar Online Ruang Guru Pada Ulasan Google Play Store Menggunakan Algoritma Naive Bayes dan Support Vector Machine," *INNOVATIVE: Journal of Social Science Research*, vol. 4 No. 3, pp. 2965–2979, 2024, doi: <https://doi.org/10.31004/innovative.v4i3.10769>.
- [8] A. B. Muzayyanah, R. E. Pawening, and Z. Arifin, "Analisis Sentimen Pada Ulasan Aplikasi Ehadrah Di Google Playstore Menggunakan Support Vector Machine," *IDEALIS: Indonesia Journal Information System*, vol. 7, no. 2, pp. 258–266, 2024, doi: <https://doi.org/10.36080/ideal.v7i2.3250>.
- [9] A. Sasi Kirana, R. Roeswidiah, and A. Pudoli, "Analisis Sentimen Pada Media Sosial Terhadap Layanan Samsat Digital Nasional Dengan Support Vector Machine," *IDEALIS: Indonesia Journal Information System*, vol. 8, no. 1, pp. 53–63, 2025, doi: <https://doi.org/10.36080/ideal.v8i1.3276>.
- [10] E. Caroline Dwi Wijaya, K. Satya Wacana, and J. Diponegoro, "Analisis Sentimen Terhadap Film Sri Asih Dengan CNN, KNN, dan Logistic Regression," *CSRID Journal*, vol. 15, no. 3, pp. 2085–1367, 2023, Accessed: Jun. 25, 2025. [Online]. Available: <https://csridjournal.potensi-utama.org/index.php/CSRIDjournal/article/view/44>
- [11] F. Fadhilah, N. P. Oktariani, and A. Rachmadita, "Perbandingan Algoritma K-Nearest Neighbor dan Logistic Regression pada Analisis Sentimen terhadap Vaksinasi Covid-19 pada Media Sosial Twitter dengan Pelabelan Vader dan Textblob," *eProceedings of Engineering*, vol. Vol. 10 No. 2, 2023, Accessed: Jun. 25, 2025. [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/19960>
- [12] N. Anggraini, S. J. Putra, L. K. Wardhani, F. D. U. Arif, N. Hakiem, and I. M. Shofi, "A Comparative Analysis of Random Forest, XGBoost, and LightGBM Algorithms for Emotion Classification in Reddit Comments," *Jurnal Teknik Informatika*, vol. 17, no. 1, pp. 88–97, May 2024, doi: [10.15408/jti.v17i1.38651](https://doi.org/10.15408/jti.v17i1.38651).
- [13] I. R. Yunita, W. Maulana Baihaqi, A. Shafira, T. Damayanti, and L. Akhaerunnisa, "Analisis Performa Algoritma Klasifikasi pada Sentimen Ulasan Pengguna terhadap Aplikasi Muamalat DIN," *Cogito Smart Journal* |, vol. 9, no. 2, 2023, doi: <https://doi.org/10.31154/cogito.v9i2.511.241-251>.

- [14] M. Salomi; Govind Ashish Kalawate; Yash Talreja, “Enhanced Sentiment Analysis of Twitter(X) Data Using an Ensemble Stacking Model,” *2025 International Conference on Artificial Intelligence and Data Engineering (AIDE)*, May 2025, doi: 0.1109/AIDE64228.2025.10987415.
- [15] Md. Shymon Islam; S.M. Aisha Siddiquea Fariha; Shaharia Sarmin Snaha; M. Raihan; Md. Sharzul Mostafa; SM Fahim Abrar, “An Interpretable Sentiment Recognition Method from Political Bangla Texts Using Stacking Ensemble Model,” *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, May 2025, doi: 10.1109/ECCE64574.2025.11013334.
- [16] M. L. and A. F. W. Chuanmin Mi, “Predicting video views of web series based on comment sentiment analysis and improved stacking ensemble model,” *Electronic Commerce Research*, vol. 24, pp. 2637–2664, Dec. 2024, doi: <https://doi.org/10.1007/s10660-022-09642-9>.
- [17] A. Ali and Anuradha Thakare, “Advancing Human-Computer Interaction: A Stacking Classifier Approach to Textual Sentiment Analysis using Ensemble Machine Learning,” *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, May 2024, doi: 10.1109/ICICV62344.2024.00088.
- [18] D. B. Santoso, A. Munna, and D. H. Untari Ningsih, “Improved playstore review sentiment classification accuracy with stacking ensemble,” *Journal of Soft Computing Exploration*, vol. 5, no. 1, pp. 38–45, Mar. 2024, doi: <https://doi.org/10.52465/josce.v5i1.247>.
- [19] Data Science PM, “CRISP-DM is Still the Most Popular Framework for Executing Data Science Projects.” Accessed: Jun. 26, 2025. [Online]. Available: <https://www.datascience-pm.com/crisp-dm-still-most-popular/>
- [20] N. A. Fadhlurrohmah *et al.*, “Analisis Sentimen Terhadap Skema Student Loan Untuk Biaya Perguruan Tinggi Pada Twitter Menggunakan Algoritma Naïve Bayes,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 2, 2025, doi: <https://doi.org/10.36040/jati.v9i2.13001>.
- [21] M. A. Palomino and F. Aider, “Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis,” *Applied Sciences (Switzerland)*, vol. 12, no. 17, Sep. 2022, doi: 10.3390/app12178765.
- [22] Fatihah Rahmayana and Yuliant Sibaroni, “Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, pp. 936–942, Oct. 2021, doi: 10.29207/resti.v5i5.3457.
- [23] U. Khairani, V. Mutiawani, and H. Ahmadian, “Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indobertweet Untuk Mendeteksi Emosi Pada Komentor Akun Berita Instagram,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 4, pp. 887–894, Aug. 2024, doi: 10.25126/jtiik.1148315.
- [24] S. Elmitwalli and J. Mehegan, “Sentiment analysis of COP9-related tweets: a comparative study of pre-trained models and traditional techniques,” *Front Big Data*, vol. 7, 2024, doi: 10.3389/fdata.2024.1357926.
- [25] A. Miyajiwala, A. Ladkat, S. Jagadale, and R. Joshi, “On Sensitivity of Deep Learning Based Text Classification Algorithms to Practical Input Perturbations,” *Computing Conference 2022*, Jan. 2022, doi: 10.1007/978-3-031-10464-0_42.
- [26] Rianto, A. B. Mutiara, E. P. Wibowo, and P. I. Santosa, “Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation,” *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00413-1.
- [27] F. A. Inayanti, “Perbandingan Pelabelan Manual Dan Vader Pada Analisis Sentimen Tentang Anti-Asian Hate Speech Terkait Covid-19 Di Twitter Dengan Metode Support Vector Machine,” Universitas Pembangunan Nasional “Veteran,” Yogyakarta, 2023. Accessed: Jun. 26, 2025. [Online]. Available: <http://eprints.upnyk.ac.id/36980/>
- [28] Y. A. Prasetyo, E. Utami, and A. Yaqin, “Pengaruh Komposisi Split Data Terhadap Performa Akurasi Analisis Sentimen Algoritma Naïve Bayes dan SVM,” *Journal homepage: Journal of Electrical Engineering and Computer (JEECOM)*, vol. 6, no. 2, 2024, doi: 10.33650/jeeecom.v4i2.
- [29] D. K. L. F. M. S. A. R. Yessy Asri, *Machine Learning & Deep Learning: Analisis Sentimen Menggunakan Ulasan Pengguna Aplikasi*. Ponorogo: Uwais Inspirasi Indonesia, 2024. Accessed: Jun. 26, 2025. [Online]. Available: https://www.google.co.id/books/edition/MACHINE_LEARNING_DEEP_LEARNING_Analisis/Yu7uEAAAQBAJ?hl=en&gbpv=1&dq=definisi+tf+idf&pg=PA60&printsec=frontcover
- [30] D. Al Mahkya, A. Djuraidah, A. H. Wigena, and B. Sartono, “Stacking Ensemble Approach In Statistical Downscaling Using CMIP6-DCP for Rainfall Estimation in Riau,” *Media Statistika*, vol. 17, no. 1, pp. 1–12, Sep. 2024, doi: 10.14710/medstat.17.1.1-12.
- [31] A. Ghasemieh, A. Lloyed, P. Bahrami, P. Vajar, and R. Kashef, “A novel machine learning model with Stacking Ensemble Learner for predicting emergency readmission of heart-disease patients,” *Decision Analytics Journal*, vol. 7, Jun. 2023, doi: 10.1016/j.dajour.2023.100242.
- [32] A. Hardoni, “Integrasi SMOTE pada Naive Bayes dan Logistic Regression Berbasis Particle Swarm Optimization untuk Prediksi Cacat Perangkat Lunak,” *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 9, no. 2, p. 144, Apr. 2021, doi: 10.26418/justin.v9i2.43173.