TESIS

PERBANDINGAN ALGORITMA DALAM ANALISIS SENTIMEN PADA ULASAN APLIKASI DUKCAPIL DI PLAY STORE MENGGUNAKAN METODE *ENSEMBLE LEARNING*

disusun oleh

Nama : **PUTU PUTRAYASA**
NIM : **23.55.1425**
Konsentrasi : **Business Intelligence**

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2026

TESIS

PERBANDINGAN ALGORITMA DALAM ANALISIS SENTIMEN PADA ULASAN APLIKASI DUKCAPIL DI PLAY STORE MENGGUNAKAN METODE *ENSEMBLE LEARNING*

COMPARISON OF ALGORITHMS IN SENTIMENT ANALYSIS ON DUKCAPIL APPLICATION REVIEWS IN THE PLAY STORE USING ENSEMBLE LEARNING METHODS

Diajukan untuk memenuhi salah satu syarat mencapai derajat Pascasarjana
Program Studi S2 Teknik Informatika

disusun oleh

Nama : **PUTU PUTRAYASA**
NIM : **23.55.1425**
Konsentrasi : **Business Intelligence**

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026

HALAMAN PERSETUJUAN

PERBANDINGAN ALGORITMA DALAM ANALISIS SENTIMEN PADA ULASAN APLIKASI DUKCAPIL DI PLAY STORE MENGGUNAKAN METODE *ENSEMBLE LEARNING*

COMPARISON OF ALGORITHMS IN SENTIMENT ANALYSIS ON DUKCAPIL APPLICATION REVIEWS IN THE PLAY STORE USING ENSEMBLE LEARNING METHODS

yang disusun dan diajukan oleh

Putu Putrayasa

23.55.1425

telah disetujui oleh Dosen Pembimbing Tesis
pada tanggal 02 Juni 2025

Dosen Pembimbing Utama,

Prof. Dr. Ema Utami, S.Si., M.Kom.
NIK. 190302037

Dosen Pembimbing Pendamping,

Robert Marco, S.T., M.T., Ph.D.
NIK. 190302228

HALAMAN PENGESAHAN

PERBANDINGAN ALGORITMA DALAM ANALISIS SENTIMEN PADA ULASAN APLIKASI DUKCAPIL DI *PLAY STORE* MENGGUNAKAN METODE *ENSEMBLE LEARNING*

COMPARISON OF ALGORITHMS IN SENTIMENT ANALYSIS ON DUKCAPIL APPLICATION REVIEWS IN THE PLAY STORE USING ENSEMBLE LEARNING METHODS

yang disusun dan diajukan oleh

Putu Putrayasa

23.55.1425

Telah dipertahankan di depan Dewan Penguji
pada tanggal 02 Juni 2025

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Dhani Ariatmanto, S.Kom., M.Kom., Ph.D. **NIK. 190302197**	
Tonny Hidayat, S.Kom., M.Kom., Ph.D. **NIK. 190302182**	
Prof. Dr. Ema Utami, S.Si., M.Kom. **NIK. 190302037**	

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer
Tanggal 02 Juni 2025

DEKAN FAKULTAS ILMU KOMPUTER

Prof. Dr. Kusrini, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama : **Putu Putrayasa**
NIM : **23.55.1425**

Menyatakan bahwa Tesis dengan judul berikut:

PERBANDINGAN ALGORITMA DALAM ANALISIS SENTIMEN PADA ULASAN APLIKASI DUKCAPIL DI *PLAY STORE* MENGGUNAKAN METODE *ENSEMBLE LEARNING*

Pembimbing Utama : Prof. Dr. Ema Utami, S.Si., M.Kom.
Pembimbing Pendamping : Robert Marco, S.T., M.T., Ph.D.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 02 Juni 2025

Yang Menyatakan,

Putu Putrayasa

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya, saya dapat menyelesaikan tesis ini sebagai bagian dari pemenuhan syarat meraih gelar Magister Komputer di Universitas AMIKOM Yogyakarta.

Tesis ini saya persembahkan kepada kedua orang tua yang selalu memberikan kasih sayang, doa, dan dukungan tanpa henti. Terima kasih juga kepada Orang tua, istri dan anak-anak tercinta yang menjadi sumber semangat saya untuk menyelesaikan studi ini.

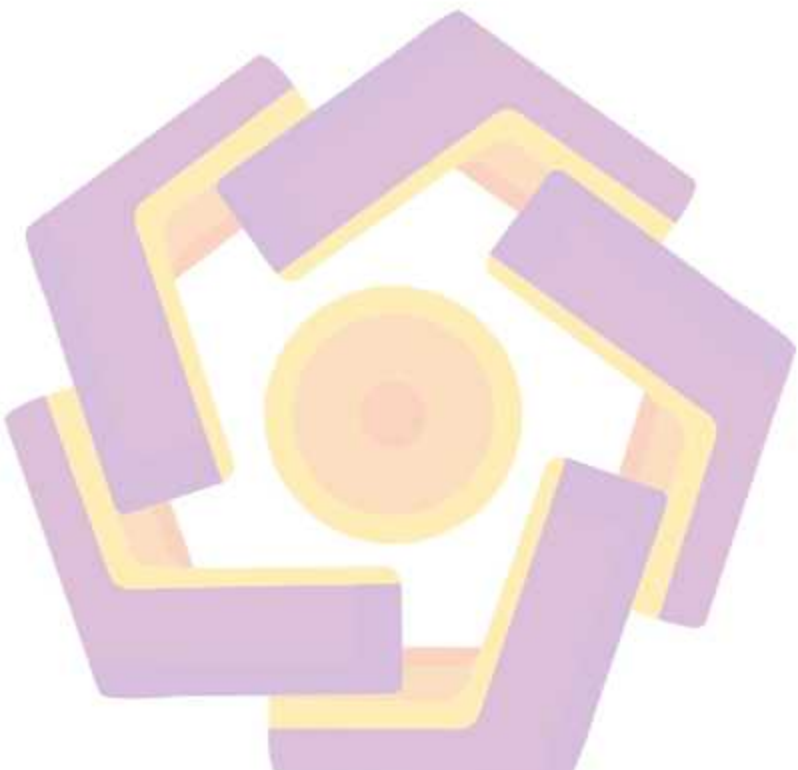
Penghargaan yang tulus saya sampaikan kepada **Prof. Dr. Ema Utami, S.Si., M.Kom.**, selaku Pembimbing Utama, dan **Robert Marco, S.T., M.T., Ph.D.**, selaku Pembimbing Pendamping, atas arahan, bimbingan, serta dukungannya selama proses penyusunan tesis ini. Saya juga menyampaikan terima kasih kepada **Prof. Dr. M. Suyanto, M.M.**, Rektor Universitas AMIKOM Yogyakarta, dan **Prof. Dr. Kusrini, M. Kom.**, Direktur Program Pascasarjana Universitas AMIKOM Yogyakarta, atas motivasi dan dukungan akademik yang diberikan selama masa studi.

Ucapan terima kasih saya tujukan pula kepada seluruh dosen di Program Magister Teknik Informatika, kawan-kawan kuliah, serta rekan kerja di Universitas Mahakarya Asia atas segala dukungan dan kebersamaan yang telah terjalin.

Semoga tesis ini bermanfaat dan dapat memberikan kontribusi positif bagi dunia pendidikan dan masyarakat.

MOTTO

Hidup adalah pembelajaran, belajar dari kesalahan adalah keniscayaan. Semangat untuk terus menjadi lebih baik dari hari kemarin, untuk menjadi pribadi yang bermanfaat bagi orang lain dan memberi arti bagi kehidupan.



HALAMAN PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas limpahan rahmat, karunia, dan kesempatan yang diberikan sehingga penulis dapat menyelesaikan tesis ini dengan judul *"Perbandingan Algoritma dalam Analisis Sentimen pada Ulasan Aplikasi DUKCAPIL di Play Store Menggunakan Metode Ensemble Learning"*.

Tesis ini disusun sebagai bagian dari pemenuhan syarat akademik untuk memperoleh gelar Magister Komputer di Program Magister Teknik Informatika Universitas AMIKOM Yogyakarta, dengan konsentrasi Business Intelligence.

Penulis berharap hasil penelitian ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan, khususnya dalam bidang analisis data dan teknologi informasi. Penulis juga menyadari bahwa penelitian ini masih memiliki keterbatasan, sehingga kritik dan saran yang membangun sangat diharapkan.

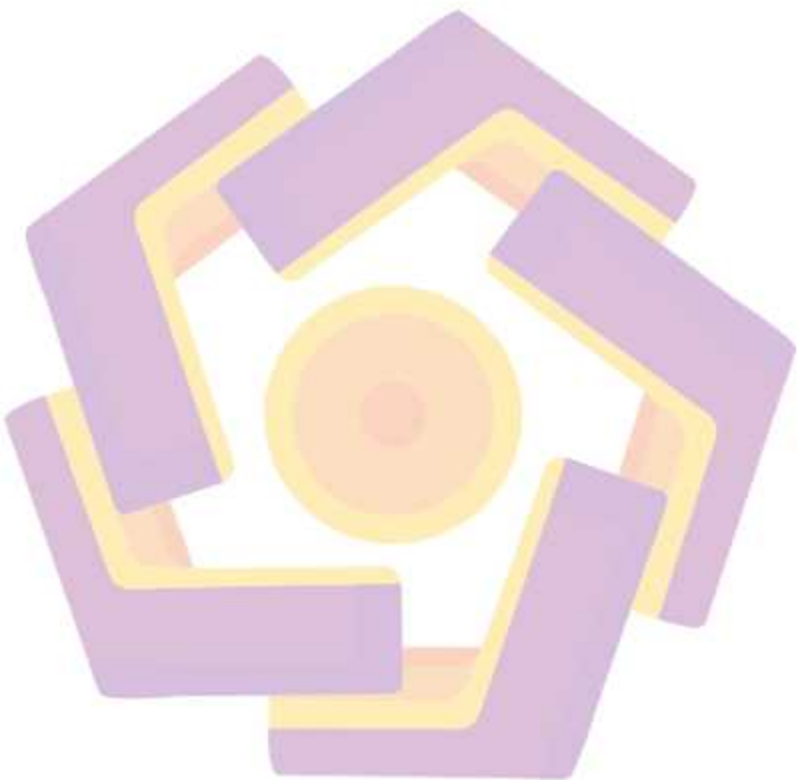
Akhir kata, semoga tesis ini dapat memberikan kontribusi positif bagi dunia pendidikan dan masyarakat luas.

Yogyakarta, 16 Juni 2025
Penulis

Putu Putrayasa

DAFTAR ISI

DAFTAR TABEL



DAFTAR GAMBAR

INTISARI

Penelitian ini bertujuan untuk menganalisis perbandingan kinerja antara model individual dan model ensemble dalam analisis sentimen ulasan aplikasi DUKCAPIL yang diperoleh dari Google Play Store. Selain itu, penelitian ini juga mengevaluasi pengaruh teknik SMOTE dalam mengatasi masalah ketidakseimbangan data terhadap kinerja model. Model ensemble yang digunakan meliputi Extra Trees, Random Forest, dan XGBoost, yang dibandingkan dengan model individual seperti Logistic Regression dan Support Vector Machine (SVM). Hasil penelitian menunjukkan bahwa model ensemble memiliki kinerja yang lebih baik dalam hal akurasi, presisi, recall, dan F1-Score dibandingkan model individual. Penggunaan teknik SMOTE secara signifikan meningkatkan performa model dengan memperbaiki distribusi data ulasan yang tidak seimbang. Model Extra Trees dengan SMOTE memberikan hasil terbaik dengan akurasi mencapai 95,85%, presisi 95,93%, recall 95,85%, dan F1-Score 95,85%. Penelitian ini menyimpulkan bahwa kombinasi model ensemble dan teknik SMOTE efektif dalam meningkatkan kinerja analisis sentimen pada data ulasan yang tidak seimbang.

Kata Kunci: Analisis Sentimen, Ensemble Learning, SMOTE, Extra Trees, Random Forest, XGBoost, Google Play Store, Data Tidak Seimbang

ABSTRACT

This study aims to compare the performance of individual models and ensemble models in sentiment analysis of DUKCAPIL application reviews obtained from Google Play Store. Furthermore, it evaluates the impact of the SMOTE technique in addressing data imbalance on model performance. The ensemble models used in this study include Extra Trees, Random Forest, and XGBoost, which are compared with individual models such as Logistic Regression and Support Vector Machine (SVM). The results indicate that ensemble models outperform individual models in terms of accuracy, precision, recall, and F1-Score. The use of SMOTE significantly improves model performance by balancing the distribution of review data. The Extra Trees model with SMOTE achieved the best results with an accuracy of 95.85%, precision of 95.93%, recall of 95.85%, and F1-Score of 95.85%. The study concludes that combining ensemble models and the SMOTE technique is effective in enhancing sentiment analysis performance on imbalanced review data.

Keywords: Sentiment Analysis, Ensemble Learning, SMOTE, Extra Trees, Random Forest, XGBoost, Data Imbalance, Google Play Store

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Dalam era digital yang berkembang pesat, analisis sentimen telah menjadi salah satu alat utama dalam memahami opini publik terhadap suatu layanan atau aplikasi. Berbagai algoritma machine learning telah digunakan untuk meningkatkan akurasi dalam mengklasifikasikan sentimen, termasuk **Extra Trees, Random Forest, Gradient Boosting, Logistic Regression, dan XGBoost**. Namun, tidak ada satu algoritma yang secara universal unggul dalam semua kondisi. Oleh karena itu, diperlukan studi komprehensif untuk memahami bagaimana performa masing-masing algoritma dalam berbagai skenario pengujian.

Penelitian ini bertujuan untuk **membandingkan performa beberapa algoritma dalam analisis sentimen ulasan aplikasi DUKCAPIL**. Aplikasi DUKCAPIL (Dinas Kependudukan dan Catatan Sipil) merupakan salah satu contoh aplikasi yang menyediakan berbagai layanan kependudukan secara Online. Melalui aplikasi ini, masyarakat dapat mengurus berbagai keperluan administrasi kependudukan seperti pembuatan KTP, KK, akta kelahiran, dan lain-lain.

Penelitian ini fokus pada **identifikasi keunggulan dan keterbatasan algoritma dalam kondisi pengujian yang berbeda**. Evaluasi dilakukan berdasarkan faktor-faktor utama yang mempengaruhi performa algoritma, yaitu:

1. **Kompleksitas data** meliputi data dengan jumlah fitur yang sedikit vs. jumlah fitur yang banyak serta panjang teks ulasan (teks pendek vs. teks panjang).
2. **Distribusi data dan ketidakseimbangan kelas** meliputi performa algoritma pada dataset seimbang vs. tidak seimbang dan ketahanan model terhadap bias kelas mayoritas.
3. **Efisiensi komputasi meliputi** waktu pelatihan dan inferensi model serta penggunaan sumber daya komputasi.

Meskipun berbagai algoritma telah digunakan dalam analisis sentimen, masih terdapat **kesenjangan penelitian** dalam memahami bagaimana algoritma bekerja dalam kondisi data yang berbeda. Banyak studi sebelumnya hanya menilai akurasi tanpa mempertimbangkan **faktor ketidakseimbangan data, ukuran dataset, dan efisiensi komputasi.** Oleh karena itu, penelitian ini menjadi penting dalam memberikan wawasan tentang bagaimana memilih algoritma yang paling sesuai untuk analisis sentimen berdasarkan karakteristik dataset.

Dengan mempertimbangkan faktor-faktor di atas, penelitian ini akan mengevaluasi **kelebihan dan kekurangan masing-masing algoritma dalam berbagai kondisi pengujian** untuk memberikan rekomendasi algoritma terbaik berdasarkan karakteristik dataset.

Kemajuan terkini dalam kekuatan komputasi dan pembelajaran mesin (ML) telah memunculkan beberapa inovasi dan perkembangan di banyak bidang penelitian dan kehidupan manusia (Mienye dan Sun, 2022). Inovasi ini memungkinkan layanan publik menjadi lebih efisien dan responsif terhadap kebutuhan masyarakat (Jim et al., 2024). Aplikasi mobile memberikan akses yang lebih mudah dan cepat bagi masyarakat untuk mendapatkan layanan yang mereka butuhkan (Huang et al., 2024). Penggunaan aplikasi mobile dalam pelayanan publik dapat meningkatkan kualitas hidup masyarakat dengan menyediakan layanan yang lebih terjangkau dan mudah diakses (Matrane et al., 2023).

Analisis sentimen adalah proses menggunakan teknik pemrosesan bahasa alami (*Natural Language Processing*), analisis teks, dan linguistik komputasional untuk mengidentifikasi dan mengekstrak informasi subjektif dari teks. Teknik ini sering digunakan untuk mengetahui pendapat atau sentimen pengguna terhadap produk atau layanan tertentu. Namun, analisis sentimen memiliki tantangan tersendiri, seperti ketidakseimbangan data ulasan di mana jumlah ulasan positif atau negatif jauh lebih banyak dibandingkan kategori lainnya, serta keragaman bahasa yang digunakan oleh pengguna. Untuk mengatasi tantangan ini, teknik ensemble learning dan metode penyeimbangan data seperti SMOTE digunakan. Teknik ensemble learning dibangun untuk membuat kinerja yang lebih dan lebih stabil, sementara SMOTE membantu menyeimbangkan data dengan membuat contoh sintetik dari kelas minoritas, sehingga meningkatkan akurasi dan keandalan prediksi model analisis sentimen.

Analisis sentimen adalah proses menggunakan pemrosesan bahasa alami (*Natural Language Processing*), analisis teks, dan linguistik komputasional untuk mengidentifikasi dan mengekstrak informasi subjektif dalam teks sumber (Mao et al., 2024). Analisis sentimen memungkinkan bisnis untuk memahami sentimen pelanggan, membuat keputusan yang tepat, dan meningkatkan penawaran mereka (Jim et al., 2024). Proses umum dari sistem analisis sentimen mencakup tahap-tahap seperti pengumpulan data, prapemrosesan teks, ekstraksi fitur, dan pelatihan model pembelajaran mesin atau pembelajaran mendalam. Dalam penelitian ini, teknik ensemble learning seperti Extra Trees digunakan untuk mengurangi overfitting dan menghasilkan prediksi yang lebih stabil. Selain itu, teknik SMOTE digunakan untuk mengatasi ketidakseimbangan data dengan membuat contoh sintetik dari kelas minoritas, sehingga meningkatkan akurasi dan keandalan model dalam tugas klasifikasi sentimen (Jim et al., 2024).

Penelitian ini menggunakan algoritma Extra Trees sebagai kandidat algoritma utama untuk analisis sentimen dan membandingkannya dengan sembilan algoritma lain: Random Forest, Gradient Boosting, Logistic Regression, LSTM, SVM, Naive Bayes, KNN, AdaBoost, dan XGBoost. Pemilihan Extra Trees didasarkan pada kemampuannya mengurangi overfitting melalui pohon keputusan yang sepenuhnya terbentuk dan pembagian acak fitur pada setiap split, menghasilkan hasil yang lebih stabil dan akurat (Miguéis et al., 2023). Extra Trees juga lebih efisien secara komputasi karena tidak memerlukan pengoptimalan bertahap seperti pada Gradient Boosting, serta memberikan prediksi yang lebih stabil karena pengacakan fitur yang lebih kuat.

Random Forest efektif tetapi rentan terhadap overfitting dan memerlukan banyak memori (Mighan & Kahani, 2020). Gradient Boosting meningkatkan akurasi tetapi sensitif terhadap outlier dan lambat (Mao et al., 2021). Logistic Regression cepat dan sederhana namun tidak cocok untuk hubungan non-linear (Ahmad et al., 2022). LSTM baik untuk data sekuensial tetapi memerlukan sumber daya besar (Kim et al., 2023). SVM kuat dalam klasifikasi tetapi kurang efisien untuk dataset besar (Choi et al., 2021). Naive Bayes cepat namun asumsi independensi sering tidak realistis (Lee et al., 2022). KNN mudah tetapi tidak efisien untuk dataset besar (Park et al., 2022). AdaBoost meningkatkan akurasi tetapi rentan terhadap noise (Zhao et al., 2021). XGBoost sangat akurat namun kompleks dalam tuning (Liu et al., 2021).

Ensemble learning dapat mengurangi overfitting, meningkatkan akurasi, dan memberikan prediksi yang lebih stabil dengan menggabungkan beberapa model (Mienye dan Sun, 2022). Metode ansambel seperti Extra Trees mengurangi kesalahan varians dan bias yang terkait dengan model pembelajaran mesin tunggal. Misalnya, bagging dalam Extra Trees mengurangi varians tanpa meningkatkan bias, sehingga menghasilkan model yang lebih *robust* dan akurat (Mienye dan Sun, 2022). Pendekatan ansambel ini sangat penting untuk meningkatkan performa analisis sentimen, yang mendasari pilihan algoritma Extra Trees sebagai fokus utama penelitian ini (Vidyashree et al., 2024).

Salah satu tantangan utama dalam analisis sentimen adalah ketidakseimbangan data, di mana jumlah ulasan positif atau negatif mungkin jauh lebih banyak dibandingkan dengan kategori lainnya. Untuk mengatasi masalah ini, digunakan teknik SMOTE (Synthetic Minority Over-sampling Technique). SMOTE adalah teknik yang efektif

dalam menangani ketidakseimbangan data dengan menghasilkan contoh sintetik dari kelas minoritas, sehingga meningkatkan performa model pembelajaran mesin (Mienye dan Sun, 2022). Penggunaan SMOTE dalam kombinasi dengan teknik ensemble learning dapat meningkatkan akurasi dan keandalan prediksi model, terutama dalam tugas-tugas klasifikasi yang kompleks seperti analisis sentimen (Vidyashree et al., 2024).

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut:

a. Bagaimana perbandingan kinerja algoritma model individual dengan model ensemble dalam analisis sentimen ulasan aplikasi DUKCAPIL? Model yang mana yang memberikan hasil yang lebih akurat dan andal?
b. Bagaimana pengaruh penggunaan teknik SMOTE untuk mengatasi ketidakseimbangan data dalam meningkatkan kinerja model analisis sentimen dalam hal akurasi, presisi, recall, dan F1-Score?

1.3. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

a. Penelitian ini akan memfokuskan pada ulasan pengguna yang terdapat di Aplikasi DUKCAPIL.
b. Data ulasan yang digunakan akan berasal dari Google Play Store pada Aplikasi DUKCAPIL.
c. Penelitian ini akan memfokuskan pada sentimen positif dan sentimen negatif dalam ulasan.

d. Sentimen netral atau ambigu mungkin tidak akan dianalisis secara mendalam.

e. Model yang digunakan terbatas pada beberapa algoritma antara lain Extra Trees, Random Forest, Gradient Boosting, Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Gaussian Naive Bayes, AdaBoost dan XGBoost.

f. Fitur yang akan digunakan dalam analisis sentimen mencakup teks dan ulasan, yang akan diproses menggunakan teknik prapemrosesan seperti penghapusan tanda baca dan angka, tokenisasi, penghapusan stopwords, dan stemming.

g. Penelitian ini akan berfokus pada bahasa Indonesia karena aplikasi ini ditujukan untuk pengguna di Indonesia.

h. Data ulasan mencakup periode Januari 2021 sampai dengan periode Desember 2023.

i. Perubahan tren sentimen seiring waktu mungkin tidak akan dianalisis secara mendalam.

j. Penggunaan Google Colab dalam pengolahan data.

k. Asumsi bahwa data ulasan yang digunakan mencakup berbagai jenis ulasan, baik positif maupun negatif.

1.4. Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

a. Mengembangkan metode yang efektif dan efisien untuk mengoptimalkan analisis sentimen dengan memanfaatkan teknik prapemrosesan data, teknik vektorisasi teks, dan berbagai algoritma pembelajaran mesin.

b. Mengevaluasi secara menyeluruh pengaruh penggunaan teknik balancing data, seperti SMOTE, terhadap kinerja model analisis sentimen, dengan mempertimbangkan metrik evaluasi seperti akurasi, presisi, recall, dan F1-Score.

c. Mengidentifikasi tantangan utama dalam implementasi metode ensemble learning untuk analisis sentimen pada data ulasan yang tidak terstruktur, serta mengusulkan solusi untuk mengatasi tantangan tersebut guna meningkatkan akurasi dan ketahanan model ensemble dibandingkan dengan model individual.

1.5. Manfaat Penelitian

Manfaat penelitian ini adalah sebagai berikut:

a. Menambah wawasan dan pengetahuan dalam bidang analisis sentimen, khususnya dalam penerapan metode ensemble learning. Penelitian ini diharapkan dapat memberikan kontribusi signifikan terhadap literatur ilmiah mengenai optimalisasi analisis sentimen dengan berbagai teknik pembelajaran mesin dan prapemrosesan data.

b. Membantu pengembang aplikasi DUKCAPIL untuk memahami sentimen dan persepsi pengguna dengan lebih baik. Hasil analisis sentimen yang dioptimalkan dapat digunakan untuk mengidentifikasi kekuatan dan kelemahan aplikasi, sehingga memungkinkan pengembang untuk melakukan perbaikan dan penyesuaian yang tepat guna meningkatkan kualitas layanan.

c. Memberikan rekomendasi yang lebih spesifik dan berbasis data untuk perbaikan aplikasi DUKCAPIL. Rekomendasi ini dapat digunakan oleh pengembang aplikasi untuk mengimplementasikan perubahan yang berfokus pada

peningkatan kepuasan pengguna, sehingga secara keseluruhan dapat meningkatkan efisiensi dan efektivitas layanan yang diberikan melalui aplikasi DUKCAPIL.

1.6. Hipotesis

Meskipun berbagai algoritma telah digunakan dalam analisis sentimen, masih terdapat kesenjangan penelitian dalam memahami bagaimana algoritma bekerja dalam kondisi data yang berbeda. Banyak studi sebelumnya hanya menilai akurasi tanpa mempertimbangkan faktor ketidakseimbangan data, ukuran dataset, dan efisiensi komputasi. Oleh karena itu, penelitian ini menjadi penting dalam memberikan wawasan tentang bagaimana memilih algoritma yang paling sesuai untuk analisis sentimen berdasarkan karakteristik dataset.

Hipotesis penelitian ini adalah sebagai berikut:

1. Algoritma ensemble seperti Extra Trees, Random Forest, dan XGBoost memiliki performa lebih baik dalam dataset dengan jumlah fitur yang banyak dan kondisi kelas tidak seimbang dibandingkan Logistic Regression.
2. Gradient Boosting dan XGBoost dapat menghasilkan performa yang baik, tetapi membutuhkan tuning parameter yang lebih kompleks dibandingkan dengan model lainnya.

Penelitian ini akan menguji hipotesis ini melalui serangkaian eksperimen untuk memberikan wawasan dalam pemilihan algoritma yang optimal.

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Ulasan pengguna aplikasi android di Play Store memberikan wawasan berharga tentang pengalaman mereka dalam menggunakan aplikasi DUKCAPIL. Ulasan ini mencakup berbagai aspek seperti kemudahan penggunaan, kecepatan layanan, dan masalah teknis yang dihadapi. Dengan melakukan analisis sentimen terhadap ulasan-ulasan ini, kita dapat memahami persepsi dan kepuasan pengguna terhadap aplikasi DUKCAPIL, serta mengidentifikasi area yang perlu diperbaiki. Analisis sentimen memungkinkan bisnis untuk memahami sentimen pelanggan, membuat keputusan yang tepat, dan meningkatkan penawaran mereka (Jim et al., 2024). Selain itu, analisis sentimen telah diterapkan pada banyak bidang, termasuk pemasaran, politik, dan psikologi, untuk memahami dan memprediksi reaksi pengguna (Mao et al., 2024). Dengan demikian, analisis sentimen terhadap ulasan pengguna dapat membantu pemerintah dan pengembang aplikasi untuk terus meningkatkan kualitas layanan yang diberikan melalui aplikasi DUKCAPIL. Pendekatan ansambel dalam analisis sentimen dapat meningkatkan akurasi prediksi dan memberikan wawasan yang lebih mendalam tentang persepsi pengguna (Mienye dan Sun, 2022). Teknik ensemble learning membantu dalam mengatasi tantangan

ketidakseimbangan data dan meningkatkan kinerja model analisis sentimen (Vidyashree et al., 2024).

Penelitian terkait metode ensemble learning yang sudah dilakukan pada penelitian sebelumnya yaitu Sentiment Analysis Using Machine Learning Approach Based on Random Forest (RF). Kesimpulan dalam penelitian ini adalah Random Forest efektif dalam analisis sentimen, tetapi perlu metode tambahan untuk menangani data sekuensial dan meningkatkan akurasi. Random Forest (RF) adalah metode terkenal yang digunakan untuk masalah klasifikasi dan regresi. Ini telah berkembang menjadi metode pembelajaran ansambel yang dibangun di atas beberapa pohon keputusan (Vidyashree et al., 2024) . Pada penelitian ini masih diperlukan adanya kombinasi dengan model sekuensial dan metode boosting untuk peningkatan performa. Sedangkan penelitian yang akan penulis lakukan adalah menggabungkan Random Forest dengan Gradient Boosting dan SMOTE untuk menangani data sekuensial dan meningkatkan akurasi analisis sentimen. Gradient Boosting adalah teknik boosting yang membangun model secara bertahap dan mengoptimalkan fungsi loss, yang membuatnya sangat efektif dalam berbagai tugas klasifikasi (Mienye dan Sun, 2022).

Ensemble learning meningkatkan akurasi dengan menggabungkan beberapa algoritma, tetapi tidak mengeksplorasi penggunaan model sekuensial seperti LSTM. LSTM adalah jenis RNN yang dirancang untuk mengingat ketergantungan jangka panjang dengan menggunakan sel memori untuk menyimpan informasi dalam jangka waktu yang lama (Ganaie et al., 2022). Kesimpulan tersebut ada pada penelitian

dengan judul A Survey of Ensemble Learning: Concepts, Algorithms, Applications. Kebutuhan untuk menggabungkan ensemble learning dengan model sekuensial untuk analisis teks yang lebih kompleks sangatlah penting. Pendekatan ansambel mengurangi kesalahan varians dan bias yang terkait dengan model pembelajaran mesin tunggal; misalnya, bagging mengurangi varians tanpa meningkatkan bias, sementara boosting mengurangi bias (Mienye dan Sun, 2022). Penggunaan SMOTE dalam kombinasi dengan teknik ensemble learning dapat meningkatkan akurasi dan keandalan prediksi model, terutama dalam tugas-tugas klasifikasi yang kompleks seperti analisis sentimen (Vidyashree et al., 2024).

Berikut adalah tinjauan terhadap 10 algoritma yang akan diteliti :

1. Extra Trees: Extra Trees classifier mengurangi korelasi antara pohon yang dipilih secara acak, sehingga dapat mengatasi masalah overfitting dan membantu meningkatkan akurasi prediksi (Vidyashree et al., 2024).
2. Random Forest: Random Forest (RF) adalah metode terkenal yang digunakan untuk masalah klasifikasi dan regresi. Ini telah berkembang menjadi metode pembelajaran ansambel yang dibangun di atas beberapa pohon keputusan (Vidyashree et al., 2024).
3. Gradient Boosting: Gradient Boosting adalah teknik boosting yang membangun model secara bertahap dan mengoptimalkan fungsi loss, yang membuatnya sangat efektif dalam berbagai tugas klasifikasi (Mienye dan Sun, 2022).
4. Logistic Regression: Logistic regression adalah model klasifikasi yang menggunakan fungsi logistik untuk variabel biner (Vidyashree et al., 2024).

5. Support Vector Machine (SVM): Support Vector Machines (SVM) adalah model pembelajaran terawasi dengan algoritma pembelajaran yang terkait yang menganalisis data untuk klasifikasi dan regresi (Diekson et al., 2023).
6. K-Nearest Neighbors (KNN): Algoritma K-Nearest Neighbors (KNN) dikenal karena kinerjanya yang sangat baik dalam tugas klasifikasi dan regresi, terutama saat berurusan dengan input multimodal (Ghosh et al., 2023).
7. Gaussian Naive Bayes: Naive Bayes adalah algoritma yang sederhana dan efisien untuk tugas klasifikasi, terutama cocok untuk klasifikasi teks dan analisis sentimen (Jim et al., 2024).
8. AdaBoost: Teknik AdaBoost digunakan untuk rekayasa fitur guna memperoleh ruang fitur yang sesuai. Setelah itu, model pembelajaran dikembangkan menggunakan classifier extra tree dan random forest (Vidyashree et al., 2024).
9. XGBoost: Algoritma XGBoost adalah ansambel berbasis pohon keputusan yang menggunakan kerangka kerja boosting gradient. Ini adalah algoritma yang skalabel dan sangat akurat yang digunakan untuk aplikasi klasifikasi dan regresi (Mienye dan Sun, 2022).
10. Long Short-Term Memory (LSTM): LSTM adalah jenis RNN yang dirancang untuk mengingat ketergantungan jangka panjang dengan menggunakan sel memori untuk menyimpan informasi dalam jangka waktu yang lama (Ganaie et al., 2022).

Dengan menggabungkan teknik ensemble learning seperti Extra Trees dan metode penyeimbangan data seperti SMOTE, penelitian ini bertujuan untuk mengeksplorasi potensi maksimal dari analisis sentimen ulasan aplikasi DUKCAPIL. Penggunaan Extra Trees diharapkan dapat mengurangi overfitting dan memberikan prediksi yang lebih stabil, sementara SMOTE akan membantu mengatasi ketidakseimbangan data dengan membuat contoh sintetik dari kelas minoritas. Kombinasi ini diharapkan dapat menangani data yang tidak terstruktur dengan lebih baik dan memberikan hasil yang lebih akurat dan andal.

2.2. Landasan Teori

2.2.1. Sentimen Analisis

Sentimen analisis adalah proses menggunakan metode otomatis untuk mengekstraksi, mengidentifikasi, dan mengklasifikasikan opini atau sentimen yang dinyatakan dalam teks menjadi kategori seperti positif, negatif, atau netral. Teknik ini sering digunakan dalam berbagai aplikasi, termasuk analisis ulasan produk, survei pelanggan, dan pengawasan media sosial (Mao et al., 2024). Dengan berkembangnya model berbasis machine learning dan deep learning, analisis sentimen telah menjadi alat yang lebih andal untuk menangkap nuansa emosi dalam teks (Fairhall, 2024).

Teknik-teknik dalam Sentimen Analisis

Berikut merupakan teknik yang biasa digunakan dalam sentiment analisis seperti yang ditampilkan pada Gambar 2.1 (Y. Mao, t ai., 2024):

a. **Lexicon-based Approach**: Pendekatan berbasis leksikon dalam analisis sentimen memanfaatkan kamus sentimen yang telah ditentukan sebelumnya untuk mengklasifikasikan teks (Huang et al., 2023).

b. **Machine Learning Approach**: Teknik ini melibatkan penggunaan algoritma pembelajaran mesin untuk mengklasifikasikan teks berdasarkan sentimen. Model yang umum digunakan termasuk Support Vector Machines (SVM), Naive Bayes,

dan Random Forest. Teknik pembelajaran mesin dalam analisis sentimen menggunakan model klasifikasi untuk memprediksi sentimen berdasarkan fitur yang diekstraksi dari teks (Mienye dan Sun, 2022).

c. **Deep Learning Approach**: Menggunakan jaringan saraf dalam seperti Recurrent Neural Networks (RNN) dan Long Short-Term Memory (LSTM) untuk menangani urutan data teks dan menangkap konteks yang lebih dalam. LSTM adalah jenis RNN yang dirancang untuk mengingat ketergantungan jangka panjang dengan menggunakan sel memori untuk menyimpan informasi dalam jangka waktu yang lama (Ganaie et al., 2022).

Gambar 2. 1 (Y. Mao, t ai., 2024)

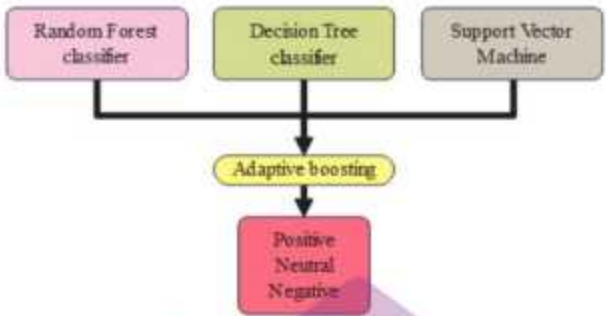
2.2.2. Metode *Ensemble Learning*

Ensemble learning adalah teknik pembelajaran mesin yang menggabungkan beberapa model untuk meningkatkan kinerja prediksi. Dengan menggabungkan beberapa model, ensemble learning dapat mengurangi overfitting, meningkatkan akurasi, dan memberikan prediksi yang lebih stabil.

2.2.2.1. Definisi dan Konsep Dasar Ensemble Learning

Ensemble learning seperti ditampilkan dalam Gambar 2.2 (V. KP. et al., 2024) menggabungkan beberapa model untuk meningkatkan kinerja prediksi (Mienye dan Sun, 2022). Teknik-teknik ensemble learning meliputi:

1. **Bagging (Bootstrap Aggregating):** Teknik ini melibatkan pelatihan beberapa model pada subset data yang berbeda dan menggabungkan hasilnya dengan mengambil rata-rata atau voting. Bagging mengurangi varians tanpa meningkatkan bias (Mienye dan Sun, 2022).
2. **Boosting:** Teknik ini melibatkan pelatihan model secara berurutan di mana setiap model berusaha untuk memperbaiki kesalahan yang dibuat oleh model sebelumnya. Gradient Boosting adalah teknik boosting yang membangun model secara bertahap dan mengoptimalkan fungsi loss (Mienye dan Sun, 2022).
3. **Stacking:** Teknik ini menggabungkan beberapa model dasar dan menggunakan model meta untuk membuat prediksi akhir berdasarkan output dari model dasar. Stacking menggabungkan berbagai model untuk meningkatkan akurasi dan keandalan prediksi (Mohammed dan Kora, 2022).



Gambar 2. 2 (V. KP. et al., 2024)

2.2.2.3. Implementasi Ensemble Learning

Berikut adalah beberapa penerapan ensemble learning:

a. **Random Forest** menggunakan beberapa pohon keputusan untuk meningkatkan akurasi prediksi. Setiap pohon keputusan dilatih pada subset data yang berbeda, dan hasil akhirnya diambil berdasarkan voting dari semua pohon. Random Forest (RF) adalah metode terkenal yang digunakan untuk masalah klasifikasi dan regresi. Ini telah berkembang menjadi metode pembelajaran ansambel yang dibangun di atas beberapa pohon keputusan (Vidyashree et al., 2024). Penelitian oleh Mao et al. (2024) menunjukkan bahwa Random Forest efektif dalam analisis sentimen karena kemampuannya untuk menangani data yang bervariasi dan mengurangi overfitting. Penggunaan Random Forest dalam analisis sentimen memungkinkan model untuk belajar dari berbagai pola data, meningkatkan akurasi prediksi dan ketahanan terhadap data yang bervariasi (Mienye dan Sun, 2022).

b. Gradient Boosting menggunakan pendekatan boosting untuk meningkatkan kinerja model dalam analisis sentimen. Teknik ini bekerja dengan melatih model secara berurutan, di mana setiap model berusaha untuk memperbaiki kesalahan yang dibuat oleh model sebelumnya. Gradient Boosting adalah teknik boosting yang membangun model secara bertahap dan mengoptimalkan fungsi loss, yang membuatnya sangat efektif dalam berbagai tugas klasifikasi (Mienye dan Sun, 2022). Penelitian oleh Huang et al. (2023) menemukan bahwa Gradient Boosting dapat meningkatkan akurasi analisis sentimen dengan menangkap hubungan kompleks dalam data teks. Gradient Boosting menangkap pola kompleks dalam data teks, yang membantu dalam meningkatkan akurasi prediksi dalam analisis sentimen (Jim et al., 2024).

c. Stacking menggabungkan beberapa model pembelajaran mesin untuk memberikan prediksi yang lebih akurat. Teknik ini menggunakan model meta untuk mengombinasikan prediksi dari berbagai model dasar, sehingga meningkatkan kinerja keseluruhan. Stacking menggabungkan berbagai model untuk meningkatkan akurasi dan keandalan prediksi (Mohammed dan Kora, 2022). Penelitian oleh Ganaie et al. (2022) menunjukkan bahwa pendekatan stacking dapat memberikan hasil yang lebih baik dibandingkan dengan model tunggal, terutama dalam tugas-tugas yang kompleks seperti analisis sentimen. Pendekatan stacking dalam analisis sentimen memungkinkan kombinasi kekuatan dari berbagai model, menghasilkan prediksi yang lebih akurat dan andal (Mao et al., 2024).

Dengan mengimplementasikan berbagai teknik ensemble learning seperti Random Forest, Gradient Boosting, dan Stacking, penelitian ini bertujuan untuk meningkatkan akurasi dan keandalan analisis sentimen ulasan aplikasi DUKCAPIL. Teknik-teknik ini diharapkan dapat memberikan pemahaman yang lebih baik tentang persepsi pengguna dan membantu pengembang dalam meningkatkan kualitas layanan aplikasi.

Algoritma yang digunakan antara lain:

1. **Random Forest**: Random Forest (RF) adalah metode terkenal yang digunakan untuk masalah klasifikasi dan regresi. Ini telah berkembang menjadi metode pembelajaran ansambel yang dibangun di atas beberapa pohon keputusan (Vidyashree et al., 2024).
2. **Extra Trees**: Extra Trees classifier, yang kemudian menyediakan mode representasi untuk klasifikasi. Dengan mengurangi korelasi antara pohon yang dipilih secara acak, classifier ini dapat mengatasi masalah overfitting dan membantu meningkatkan akurasi prediksi (Vidyashree et al., 2024).
3. **Gradient Boosting:** Gradient Boosting adalah teknik boosting yang membangun model secara bertahap dan mengoptimalkan fungsi loss, yang membuatnya sangat efektif dalam berbagai tugas klasifikasi (Mienye dan Sun, 2022).
4. **Logistic Regression:** Logistic regression adalah model klasifikasi yang menggunakan fungsi logistik untuk variabel biner (Vidyashree et al., 2024).
5. **Support Vector Machine (SVM):** Support Vector Machines (SVM) adalah model pembelajaran terawasi dengan algoritma pembelajaran yang terkait yang menganalisis data untuk klasifikasi dan regresi (Diekson et al., 2023).

6. **K-Nearest Neighbors (KNN):** Algoritma K-Nearest Neighbors (KNN) dikenal karena kinerjanya yang sangat baik dalam tugas klasifikasi dan regresi, terutama saat berurusan dengan input multimodal (Ghosh et al., 2023).
7. **Gaussian Naive Bayes:** Naive Bayes adalah algoritma yang sederhana dan efisien untuk tugas klasifikasi, terutama cocok untuk klasifikasi teks dan analisis sentimen (Jim et al., 2024).
8. **AdaBoost:** Teknik AdaBoost digunakan untuk rekayasa fitur guna memperoleh ruang fitur yang sesuai. Setelah itu, model pembelajaran dikembangkan menggunakan classifier extra tree dan random forest (Vidyashree et al., 2024).
9. **XGBoost**: Algoritma XGBoost adalah ansambel berbasis pohon keputusan yang menggunakan kerangka kerja boosting gradient. Ini adalah algoritma yang skalabel dan sangat akurat yang digunakan untuk aplikasi klasifikasi dan regresi (Mienye dan Sun, 2022).
10. **Long Short-Term Memory (LSTM)**: LSTM adalah jenis RNN yang dirancang untuk mengingat ketergantungan jangka panjang dengan menggunakan sel memori untuk menyimpan informasi dalam jangka waktu yang lama (Ganaie et al., 2022).

Pendekatan utama dalam **ensemble learning**, yaitu teknik yang menggabungkan prediksi dari beberapa model untuk meningkatkan kinerja keseluruhan. Berikut penjelasan singkat masing-masing:

2.2.3. Teknik *Preprocessing* Teks

Tahapan *preprocessing* teks adalah langkah penting dalam analisis teks untuk menghilangkan noise dan mempersiapkan data untuk analisis lebih lanjut. Berikut adalah tahapan-tahapan yang dilakukan dalam *preprocessing* teks:

2.2.3.1. Informasi awal data

Tahap awal dalam preprocessing data adalah memahami struktur dan karakteristik dataset. Informasi awal data melibatkan pengumpulan metadata, seperti jumlah sampel, jumlah fitur, tipe data setiap kolom, dan distribusi kelas dalam dataset. Tahap ini bertujuan untuk mengidentifikasi potensi masalah, seperti ketidakseimbangan kelas, outlier, atau inkonsistensi data.

Penelitian oleh Mienye dan Sun (2022) menekankan pentingnya tahap ini untuk memastikan bahwa data sesuai dengan kebutuhan analisis dan pemodelan. Sebagai contoh, distribusi data yang tidak seimbang dapat memengaruhi akurasi model, sehingga memerlukan langkah-langkah khusus seperti SMOTE untuk mengatasinya. Selain itu, analisis deskriptif seperti statistik ringkasan (mean, median, standar deviasi) membantu mengidentifikasi pola awal dalam dataset, seperti rentang nilai dan outlier yang signifikan (Vidyashree et al., 2024).

Langkah ini juga melibatkan eksplorasi visual menggunakan histogram, boxplot, atau heatmap untuk memahami pola hubungan antar fitur, yang dapat memandu langkah preprocessing selanjutnya.

Ketidaksempurnaan data, seperti nilai yang hilang (missing values) dan data duplikat, sering kali ditemukan dalam dataset mentah. Missing values dapat terjadi karena kesalahan input, pengambilan data yang tidak lengkap, atau kerusakan sistem. Sementara itu, duplikasi data dapat disebabkan oleh kesalahan penggabungan dataset atau pencatatan berulang. Penelitian oleh Fairhall (2024) menunjukkan bahwa membersihkan data dari missing values dan duplikasi secara signifikan meningkatkan kualitas fitur, yang berkontribusi pada akurasi model pembelajaran mesin.

Menghapus tanda baca dan angka dari teks untuk membersihkan data dan memfokuskan analisis pada kata-kata yang relevan. Penghapusan tanda baca dan angka merupakan langkah awal dalam preprocessing teks untuk memastikan bahwa hanya informasi yang relevan yang dianalisis (Mao et al., 2024). Memecah teks menjadi unit-unit kecil yang disebut token. Token ini dapat berupa kata, frasa, atau karakter tergantung pada kebutuhan analisis. Tokenisasi adalah proses memecah teks

menjadi unit-unit kecil yang disebut token, yang dapat berupa kata, frasa, atau karakter (Huang et al., 2023). Menghapus kata-kata umum yang tidak memberikan informasi penting, seperti dan, atau, tetapi, dll. Penghapusan stopwords adalah langkah penting dalam preprocessing teks untuk menghilangkan kata-kata yang tidak memberikan informasi penting, sehingga meningkatkan akurasi analisis (Jim et al., 2024).

Stemming mengubah kata ke bentuk dasarnya dengan menghapus akhiran, sedangkan lematisasi mengubah kata ke bentuk dasarnya berdasarkan kamus. Stemming dan lematisasi adalah teknik untuk mengubah kata ke bentuk dasarnya, dengan stemming menghapus akhiran dan lematisasi menggunakan bentuk dasar kata dari kamus (Ganaie et al., 2022).

2.2.3.7. Validasi sentiment oleh manusia

Validasi sentimen oleh manusia adalah proses di mana teks atau komentar dievaluasi secara manual untuk menentukan polaritas sentimen (positif, negatif, atau netral) berdasarkan interpretasi manusia. Pendekatan ini sering dianggap sebagai langkah tambahan untuk meningkatkan akurasi dan kualitas label sentimen dalam dataset, terutama jika ada ambiguitas dalam data atau ketidaksesuaian antara teks dan skor sentimen yang dihasilkan secara otomatis (Mao et al., 2024).

Meskipun skor bintang yang diberikan pengguna sering kali digunakan sebagai label sentimen, studi menunjukkan bahwa skor ini tidak selalu mencerminkan isi teks dengan akurat. Sebagai contoh, seorang pengguna mungkin memberikan skor bintang tinggi tetapi menyertakan komentar kritis dalam ulasannya, atau sebaliknya. Dalam konteks ini, validasi manual dapat digunakan untuk mengidentifikasi dan memperbaiki ketidaksesuaian tersebut (Vidyashree et al., 2024).

2.2.4. Data Encoding

Encoding memastikan bahwa data mentah dapat diolah secara efisien tanpa kehilangan informasi yang relevan, terutama dalam dataset yang berisi variabel kategori, teks, atau data tidak terstruktur lainnya (Mao et al., 2024). Dengan encoding yang tepat, model dapat menangkap pola dan hubungan dalam data secara lebih akurat, meningkatkan performa prediksi secara keseluruhan (Fairhall, 2024).

2.2.5. Feature Extraction dengan TF-IDF

Feature extraction adalah proses mengubah data mentah menjadi representasi fitur yang lebih bermakna bagi algoritma pembelajaran mesin. Dalam analisis data teks, feature extraction memainkan peran penting dalam mengurangi dimensi data dan meningkatkan efisiensi proses pembelajaran mesin. Pendekatan seperti TF-IDF, Word Embedding, dan Contextual Embedding telah digunakan secara luas untuk meningkatkan performa model dalam berbagai bidang, termasuk pendidikan, keuangan, dan deteksi penipuan (Priyambada et al., 2023; Barongo et al., 2024).

TF-IDF adalah metode yang digunakan untuk mengukur seberapa penting sebuah kata dalam sebuah dokumen relatif terhadap kumpulan dokumen (Mao et al., 2024). Teknik ini mengkombinasikan frekuensi kemunculan kata (Term Frequency) dengan frekuensi terbalik dokumen (Inverse Document Frequency) untuk memberikan bobot yang lebih besar pada kata-kata yang unik dalam dokumen tertentu (Huang et al., 2023).

Prinsip Dasar TF-IDF

1. **Term Frequency (TF):** Mengukur seberapa sering sebuah kata muncul dalam sebuah dokumen. Frekuensi kata (TF) digunakan untuk menghitung seberapa sering kata tertentu muncul dalam sebuah dokumen, yang membantu dalam menentukan relevansi kata tersebut dalam konteks dokumen tersebut (Jim et al., 2024).
2. **Inverse Document Frequency (IDF):** Mengukur seberapa jarang sebuah kata muncul di seluruh dokumen dalam korpus. Frekuensi terbalik dokumen (IDF) memberikan bobot lebih besar pada kata-kata yang jarang muncul di seluruh dokumen, sehingga membantu mengidentifikasi kata-kata yang lebih spesifik dan informatif (Ganaie et al., 2022).

2.2.6. SMOTE (Synthetic Minority Over-sampling Technique)

Ketidakseimbangan data adalah masalah umum dalam pembelajaran mesin, di mana kelas mayoritas mendominasi dataset, sehingga mengakibatkan model cenderung mengabaikan kelas minoritas. Hal ini dapat berdampak negatif pada kinerja model, terutama dalam tugas klasifikasi di mana kelas minoritas sering kali menjadi fokus utama. SMOTE (Synthetic Minority Over-sampling Technique) adalah salah satu solusi yang paling efektif untuk mengatasi masalah ini dengan menciptakan sampel sintetik untuk kelas minoritas, sehingga membantu menyeimbangkan data dan meningkatkan performa model (Mao et al., 2024).

Untuk mengatasi ketidakseimbangan data, digunakan teknik SMOTE (Synthetic Minority Over-sampling Technique). SMOTE adalah teknik yang digunakan untuk mengatasi masalah ketidakseimbangan data dengan membuat contoh sintetik dari kelas minoritas. Penggunaan SMOTE dalam kombinasi dengan teknik ensemble learning dapat meningkatkan akurasi dan keandalan prediksi model, terutama dalam tugas-tugas klasifikasi yang kompleks seperti analisis sentimen (Vidyashree et al., 2024).

2.2.6.1 Over-sampling dan Under-sampling

Ketidakseimbangan data dapat diatasi dengan dua pendekatan utama:

1. **Over-sampling**: Menambah jumlah sampel kelas minoritas untuk menyeimbangkan proporsi dataset. SMOTE adalah contoh over-sampling yang efektif, karena tidak hanya menduplikasi data, tetapi juga menciptakan sampel sintetik baru. Namun, over-sampling dapat meningkatkan risiko overfitting jika tidak diterapkan dengan hati-hati.

2. **Under-sampling**: Mengurangi jumlah sampel dari kelas mayoritas sehingga distribusi data menjadi seimbang. Teknik ini dapat menghilangkan informasi penting dari kelas mayoritas, tetapi dengan algoritma seperti NearMiss, sampel mayoritas yang paling relevan dengan kelas minoritas dapat dipilih secara selektif (Mienye dan Sun, 2022).

2.2.6.2. Penerapan SMOTE dalam Analisis Data

SMOTE telah membuktikan efektivitasnya dalam berbagai domain analisis data, terutama dalam mengatasi ketidakseimbangan kelas yang dapat mengurangi keandalan model pembelajaran mesin. Dalam tugas analisis sentimen, ketidakseimbangan data sering kali terjadi karena ulasan positif cenderung mendominasi dibandingkan ulasan negatif. Hal ini dapat menyebabkan model bias terhadap kelas mayoritas, sehingga pola-pola unik dalam ulasan negatif sulit dikenali. Dengan menerapkan SMOTE, model mampu menghasilkan representasi kelas minoritas yang lebih baik, memungkinkan analisis sentimen menjadi lebih akurat dan sensitif terhadap variasi emosi pengguna (Mao et al., 2024; Ganaie et al)

2.2.7. Membangun Model, Pelatihan dan Pengujian

Model dalam sentiment analysis dalam penelitian ini, menggunakan perbandingan kinerja antara algoritma individual dan algoritma ensemble learning. Perhatian utama, ditujukan kepada algoritma-algoritma ensemble learning dengan beberapa pendekatan seperti bagging, boosting, stacking, voting dan kombinasi bagging dengan metode lainnya. Setelah model didefinisikan, dilanjutkan dengan pelatihan model dari dataset yang dibagi dua yang menjadi data untuk pelatihan (training) dan data untuk pengujian (test). Umumnya data komposisinya data pelatihan berbanding data pengujian adalah 80:20.

2.2.8. Evaluasi Kinerja Model

Evaluasi kinerja model bertujuan untuk mengukur seberapa baik model pembelajaran mesin atau deep learning dalam menyelesaikan tugas tertentu. Langkah

ini penting untuk memastikan model tidak hanya memiliki performa baik pada data pelatihan, tetapi juga mampu melakukan generalisasi dengan baik pada data baru. Dalam analisis sentimen, evaluasi bertujuan untuk mengukur kemampuan model dalam mengklasifikasikan teks ke dalam kategori sentimen seperti positif, negatif, atau netral (Mao et al., 2024). Confusion matrix adalah representasi tabel yang menunjukkan prediksi model terhadap data uji dibandingkan dengan label sebenarnya. Matriks ini digunakan untuk menghitung metrik evaluasi lainnya, seperti precision, recall, dan F1-score.

Data dapat dibagi menjadi dua subset: data pelatihan dan data pengujian. Pembagian ini memungkinkan model dilatih pada satu subset dan diuji pada subset lainnya untuk mengukur performa generalisasi (Fairhall, 2024). Cross-validation dapat digunakan untuk membagi dataset ke dalam beberapa lipatan (fold) dan melatih serta menguji model secara bergantian. Teknik ini membantu mengurangi bias evaluasi yang mungkin timbul dari pembagian data yang tidak representatif.

2.2.9. Hyperparameter Tuning

Hyperparameter tuning adalah proses mengoptimalkan nilai-nilai hyperparameter dalam model pembelajaran mesin untuk meningkatkan kinerja model. Hyperparameter adalah parameter yang tidak dipelajari oleh model selama proses pelatihan, tetapi ditentukan sebelum pelatihan dimulai. Contoh hyperparameter termasuk jumlah lapisan dalam jaringan saraf, nilai learning rate, dan jumlah pohon dalam Random Forest (Mao et al., 2024).

Proses tuning hyperparameter bertujuan untuk mencari kombinasi nilai yang menghasilkan model dengan performa terbaik pada data validasi. Teknik ini sangat penting karena nilai hyperparameter yang dipilih secara suboptimal dapat menyebabkan model underfitting atau overfitting (Vidyashree et al., 2024).

2.3. Keaslian Penelitian

Beberapa penelitian sebelumnya yang relevan dengan penelitian yang sedang dilakukan di antaranya disajikan dalam tabel 2. Matriks literature review dan posisi penelitian dibawah.

Tabel 2.1 (Matriks literatur review dan posisi penelitian)

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	An effective ensemble deep learning framework for text classification	A Mohammed dan R Kora, Journal of King Saud University - Computer and Information Sciences, 2022	Mengembangkan kerangka kerja pembelajaran mendalam ensemble untuk klasifikasi teks	Random Subspace mengungguli metode ensemble lainnya dalam analisis sentimen	Memerlukan metode tambahan untuk menangani data sekuensial dan meningkatkan akurasi	Penelitian ini menggabungkan Extra Trees dengan SMOTE untuk menangani ketidakseimbangan data dan meningkatkan akurasi analisis sentimen.
2	Multilayer Hybrid Ensemble Machine Learning Model for Analysis of Covid-19 Vaccine Sentiments	Jim et al., Natural Language Processing Journal, 2024	Menggunakan model ensemble untuk analisis sentimen vaksin Covid-19	Kombinasi berbagai model meningkatkan akurasi analisis sentimen	Tidak menggabungkan teknik SMOTE	Penelitian ini menggabungkan Extra Trees dengan SMOTE untuk meningkatkan akurasi dan ketahanan model dalam analisis sentimen.
3	Efficient State of Charge Estimation in Electric Vehicles Batteries	K. Perumal, T. Sudhamathi, Measurement Sensors, 2024	Menggunakan model ansambel untuk estimasi state of charge pada baterai kendaraan listrik	Model ansambel meningkatkan akurasi estimasi state of charge	Memerlukan tuning parameter yang kompleks	Penelitian ini menggunakan Extra Trees dengan SMOTE untuk menangani ketidakseimbangan data dan meningkatkan akurasi prediksi.
4	Application of Artificial Intelligence in Predicting	Khan et al., Journal of Computational Science, 2024	Fokus pada penggunaan AI dan teknik ensemble untuk	AI meningkatkan prediksi dinamika	Memerlukan metode tambahan untuk penanganan	Penelitian ini menggabungkan Extra Trees dengan SMOTE

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	the Dynamics of Various Factors Using AI		memprediksi dinamika berbagai faktor	berbagai faktor	data tidak terstruktur	untuk meningkatkan akurasi dan ketahanan model dalam analisis sentimen.
5	Another Use of SMOTE for Interpretable Data Coll	Bunkhumpornpat et al., Expert Systems with Applications, 2023	Menggunakan SMOTE untuk menangani ketidakseimbangan data	SMOTE meningkatkan kinerja model pembelajaran mesin dengan membuat contoh sintetik dari kelas minoritas	Tidak menggabungkan teknik ensemble	Penelitian ini menggabungkan SMOTE dengan Extra Trees untuk meningkatkan akurasi analisis sentimen pada ulasan aplikasi DUKCAPIL.
6	Machine Learning Based Intrusion Detection on Multi-Classification	Dhanabal et al., Procedia Computer Science, 2024	Menggunakan model ensemble untuk deteksi intrusi pada multi-klasifikasi	Model ensemble meningkatkan akurasi deteksi intrusi	Memerlukan tuning parameter yang kompleks	Penelitian ini menggunakan Extra Trees dengan SMOTE untuk meningkatkan akurasi dan ketahanan model dalam analisis sentimen.

BAB III

METODE PENELITIAN

3.1. Jenis, Sifat dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk mengembangkan dan mengevaluasi model analisis sentimen. Pendekatan kuantitatif digunakan dalam penelitian ini karena memungkinkan pengukuran yang objektif dan analisis statistik dari data ulasan pengguna. Melalui pendekatan ini, diharapkan dapat mengevaluasi kinerja model secara numerik dan membandingkan berbagai teknik yang digunakan dalam analisis sentimen. Penelitian ini melibatkan beberapa tahapan penting, termasuk pengumpulan data, preprocessing, vektorisasi teks, penyeimbangan data, pembentukan model, dan evaluasi model.

3.2. Metode Pengumpulan Data

Metode pengumpulan data dalam penelitian ini dilakukan melalui proses web scraping yang menggunakan Google Play Scraper dan diimplementasikan menggunakan lingkungan pengembangan (IDE) Google Colab. Web scraping dilakukan untuk mengumpulkan ulasan pengguna aplikasi DUKCAPIL yang tersedia di Google Play Store selama periode waktu dari tahun 2021 hingga Desember 2023 yang kemudian disimpan dalam bentuk format csv sebanyak 10000 data ulasan. Proses web scraping ini dilakukan dengan tujuan untuk mengakses dan mengumpulkan data ulasan pengguna secara otomatis.

Penggunaan platform Google Colab yang menyediakan lingkungan pemrograman Python dengan akses ke GPU. Google Colab dipilih karena kemudahan penggunaannya dan kemampuannya untuk menangani komputasi yang intensif. Beberapa pustaka yang digunakan dalam penelitian ini antara lain *imbalanced-learn, xgboost, scikit-learn, nltk*, dan Sastrawi.

3.3. Metode Analisis Data

Beberapa langkah-langkah yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Meng*install* pustaka yang diperlukan dalam proses penelitian ini. Pustaka yang dibutuhkan, berkembang sesuai kebutuhan penelitian ini.
2. Mengunduh stopwords bahasa Indonesia menggunakan menggunakan pustaka Sastrawi. Pustaka Sastrawi adalah library dalam Bahasa pemrograman Python. Pustaka ini menjadi penting karena data yang akan dilakukan proses stemming adalah dataset dalam Bahasa Indonesia.
3. Menyiapkan dataset ulasan yang akan digunakan dalam analisis sentiment. Dataset ini terdiri dari teks ulasan dan skor yang diberikan oleh pengguna. Skor ulasan digunakan untuk menentukan sentimen pengguna terhadap aplikasi, dengan skor tinggi menunjukkan sentimen positif dan skor rendah menunjukkan sentimen negatif.
4. Proses *preprocessing* data. *Preprocessing* data adalah langkah penting dalam analisis teks untuk menghilangkan noise dan mempersiapkan data untuk analisis lebih lanjut. Langkah-langkah preprocessing yang dilakukan dalam penelitian ini meliputi penghapusan tanda baca dan angka, tokenisasi, penghapusan stopwords, dan stemming.
5. Vektorisasi Teks dengan TF-IDF. Setelah melakukan *preprocessing*, langkah selanjutnya adalah mengubah teks ulasan menjadi fitur numerik menggunakan teknik TF-IDF. TF-IDF (Term Frequency-Inverse Document Frequency) adalah teknik yang mengukur pentingnya kata dalam dokumen berdasarkan

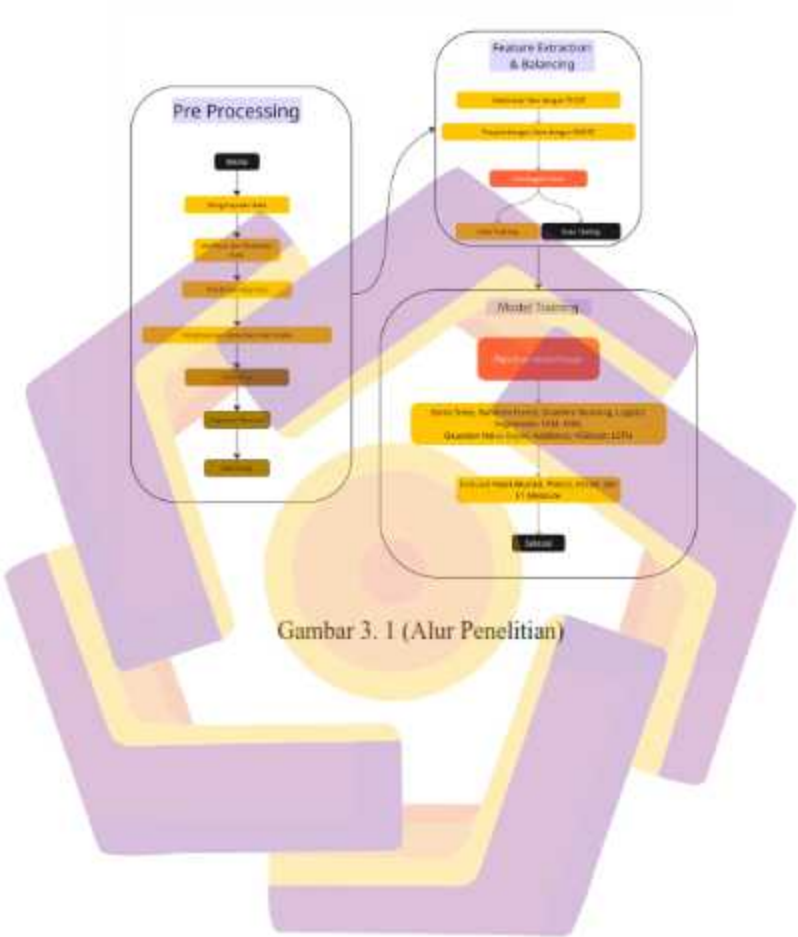
frekuensi kemunculannya dan seberapa jarang kata tersebut muncul di seluruh dokumen.

6. Jika data yang terdapat dalam data yang telah diverifikasi ternyata tidak seimbang, maka dilakukan penyeimbangan data dengan SMOTE. Ketidakseimbangan data ulasan, di mana jumlah ulasan positif, netral, dan negatif tidak seimbang. Untuk mengatasi masalah ini, penggunaan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) untuk menyeimbangkan data.
7. Pembagian data. Setelah *preprocessing* dan penyeimbangan data, langkah selanjutnya adalah membagi data menjadi set pelatihan dan set pengujian. Pembagian ini penting untuk mengevaluasi kinerja model secara objektif.
8. Pelatihan model dasar. Model dasar yang digunakan dalam penelitian ini adalah 10 Algoritma yang telah dijelaskan dibagian awal penelitian ini.

Evaluasi model dilakukan untuk mengukur kinerja model dalam analisis sentimen. Metrik yang digunakan dalam evaluasi meliputi akurasi, presisi, recall, dan F1-Score.

3.4. Alur Penelitian

Alur penelitiannya seperti yang ditampilkan pada Gambar 3.1 (Alur Penelitian).



Gambar 3. 1 (Alur Penelitian)

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Pengumpulan Data

Dalam penelitian ini, pengumpulan data dilakukan dengan metologi scraping, dikerjakan di google colab, menggunakan bahasa python dan pustaka yang diperlukan.

Proses pengumpulan data dilakukan dengan teknik scraping menggunakan pustaka Python bernama google-play-scraper. Data yang diambil mencakup 10.000 ulasan periode Januari 2021 hingga Desember 2023 pada URL seperti yang ditampilkan pada Gambar 4.1 (Google Maps Scraper):
https://play.google.com/store/apps/details?id=gov.dukcapil.mobile_id&hl=en-US

Gambar 4. 1 (Google Maps Scraper)

Pada screenshoot yang disajikan dalam gambar 4.1 nampak bahwa aplikasi identitas kependudukan digital yang dipublikasikan oleh DITJEN DUKCAPIL KEMENDAGRI telah di download lebih dari 100 juta kali (100M+). Hal tersebut menandakan aplikasi ini memang digunakan secara luas oleh masyarakat Indonesia.

Screenshoot aplikasi ini, dilakukan saat melakukan revisi penelitian ini pada tanggal 16 Februari 2025, pasca Seminar Hasil Penelitian Thesis (SHPT). Dalam gambar tersaji bahwa ada 57.6K review dan score bintang 2.8 sebagai rerata dari score yang telah diberikan dalam 57.6K Review tersebut. Score diberikan dalam bentuk bintang 1 hingga bintang 5. Bintang 1-2 sebagai negatif, bintang 3 sebagai netral dan bintang 4-5 sebagai positif, maka rating 2.8 berada dibawah netral dan cenderung pada negatif. Diharapkan melalui penelitian ini, bisa memberikan temuan berharga yang bermanfaat bagi pengembangan aplikasi ini di masa mendatang.

Selanjutnya, dari aplikasi tersebut berbagai komentar review terhadap aplikasi tersebut diunduh dengan cara melakukan scraping yang dilakukan di google colabs, dan pustaka google play scraper sebanyak 10.000 data.

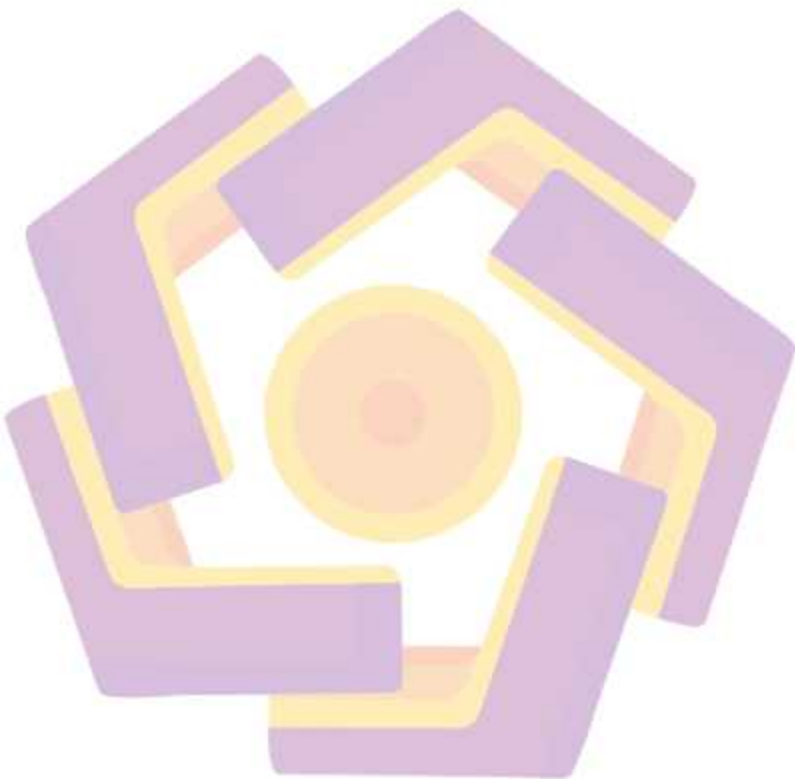
Langkah-langkah scraping adalah sebagai berikut:

1. Instalasi pustaka google-play-scraper.
2. Mengambil data ulasan aplikasi DUKCAPIL dengan menentukan ID aplikasi.
3. Menyimpan hasil scraping ke dalam file CSV untuk analisis lebih lanjut.

Teknik ini memungkinkan pengumpulan data ulasan secara otomatis dan efisien, dengan atribut utama yang diperoleh antara lain:

- **reviewId**: ID unik ulasan.
- **UserName** : Nama pengguna aplikasi
- **Userimage** : url yang mengandung image pengguna
- **content**: Teks ulasan pengguna.
- **score**: Skor ulasan dalam skala 1-5 bintang.
- **thumsUpCount** : jumlah tanda jempol yang diberikan oleh pengguna
- **revisewCreatedVersion** : mengandung versi dari review diberikan

- **at**: Tanggal ulasan diberikan.
- **replyContent** : menunjukkan berapa kali ulasan dibalas
- **replyAt** : menunjukkan waktu ulasan dibalas
- **appVersion** : menunjukkan versi dari aplikasi



Data hasil scraping dapat dilihat pada gambar yang disajikan pada gambar 4.2 (Hasil Data Scraping).

index	reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	appVersion
0	[illegible]	[illegible]	https://play-lh.googleusercontent.com/a/ALV-[illegible]	[illegible]	2	[illegible]	1.2.2	2023-12-10 22:04:51	NaN	NaN	1.2.2
1	[illegible]	[illegible]	https://play-lh.googleusercontent.com/a/ALV-[illegible]	[illegible]	3	308	1.2.2	2023-12-12 16:20:04	NaN	NaN	1.2.2
2	[illegible]	[illegible]	https://play-lh.googleusercontent.com/a/ALV-[illegible]	[illegible]	4	200	1.2.2	2023-12-21 [illegible]	NaN	NaN	1.2.2
3	[illegible]	[illegible]	https://play-lh.googleusercontent.com/a/ALV-[illegible]	[illegible]	[illegible]	36	1.2.2	2023-11-30 [illegible]	NaN	NaN	1.2.2
4	[illegible]	[illegible]	https://play-lh.googleusercontent.com/a/ALV-[illegible]	[illegible]	4	[illegible]	1.2.2	2023-10-24 [illegible]	NaN	NaN	1.2.2

Gambar 4. 2 (Hasil Data Scraping)

Adapun metadata dari dataset tersebut, disajikan dalam gambar 4.3 (Metadata)

```
reviewId                object
userName                object
userImage               object
content                 object
score                    int64
thumbsUpCount            int64
reviewCreatedVersion    object
at                      object
replyContent           float64
repliedAt              float64
appVersion              object
dtype: object
```

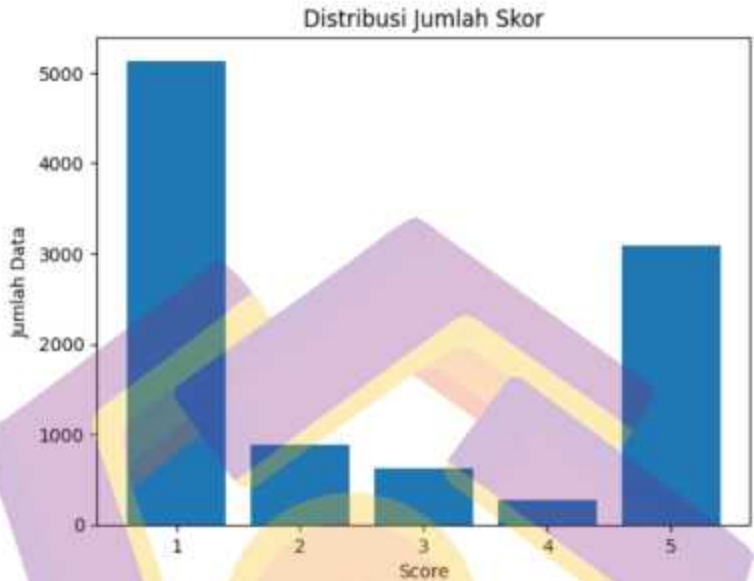
Gambar 4. 3 (Metadata)

dataset yang diperoleh berisi total 10.000 ulasan dengan distribusi skor sesuai dengan tabel 4.1:

Tabel 4.1 (Distribusi Skor)

Score	Jumlah data
1	5.133
2	881
3	625
4	278
5	3.083

Jika divisualisasikan dalam format grafik batang, distribusi dataset yang dihasilkan dari proses scraping dapat disajikan seperti gambar 4.4 (Distribusi Dataset):



Gambar 4. 4 (Distribusi Dataset)

4.2. Preprocessing Data

Tahap preprocessing bertujuan untuk mempersiapkan data mentah menjadi data yang siap dianalisis. Langkah-langkah yang dilakukan meliputi:

1. Penghapusan Tanda Baca dan Angka: Teks ulasan dibersihkan dari tanda baca dan angka untuk mengurangi noise.
2. Tokenisasi: Setiap ulasan dipecah menjadi kata-kata individual menggunakan teknik tokenisasi.
3. Penghapusan Stopwords: Kata-kata umum yang tidak memiliki makna spesifik, seperti "dan," "yang," "di," dihapus menggunakan pustaka NLTK.

4. Stemming: Kata-kata diubah ke bentuk dasarnya menggunakan library Sastrawi.

Hasil preprocessing menunjukkan bahwa teks ulasan menjadi lebih bersih dan terstruktur, memudahkan proses analisis selanjutnya. Proses diatas, menghasilkan dataframe seperti ditampilkan dalam gambar 4.5 (Dataframe).

Gambar 4. 5 (Dataframe)

4.2.1. Penanganan Data Missing dan Duplicates

Sebelum melakukan preprocessing, data diperiksa untuk nilai yang hilang (missing values) dan duplikasi.

Hasil pemeriksaan menunjukkan bahwa data kosong dan duplikat dihapus agar tidak memengaruhi hasil analisis, menghasilkan jumlah data yang berkurang. Selain menghilangkan data kosong dan duplikat, terlebih dahulu telah dipilih hanya kolom-kolom yang akan digunakan dalam penelitian ini, yaitu kolom content dan score. Dataset yang terdiri atas fitur content dan score dengan menghilangkan data duplikat dan data kosong ditunjukkan pada gambar 4.6 (Content dan Score).

```
Dataset setelah dibersihkan (menghapus duplikat dan data kosong):
                                                content  score
0  Tolong di tingkatkan performa yh biar support ...      2
1  Perlu dikembangkan lagi dan perkuat keamanan d...      1
2  VERIFIKASI DIDEPAN PETUGAS DUKCAPIL... KUNO ! ...      1
3  Masuk ke aplikasi saja susah, bagaimana pakai ...      1
4  Kenapa mau verifikasi kok selalu token selalu ...      1
Dataset berhasil disimpan sebagai ulasan_rapi.csv
```

Gambar 4. 6 (Content dan Score)

Setelah menghilangkan data duplikat dan data kosong, terlihat terjadi pengurangan jumlah data dari semula 10.000 baris data ulasan, menjadi 9.599 baris data ulasan seperti ditampilkan dalam Gambar 4.7 (Baris data ulasan).

```
Jumlah baris dan kolom: (9599, 2)

Tipe data setiap kolom:

content    object
score       int64
dtype: object

Jumlah nilai yang hilang di setiap kolom:

content    0
score      0
dtype: int64
```

Gambar 4. 7 (Baris data ulasan)

4.2.2. Tanpa Validasi Sentimen oleh manusia

Berdasarkan referensi yang telah disampaikan pada landasan teori, maka peneliti memilih untuk tidak melakukan validasi dengan sentimen manusia. Jadi dalam penelitian ini, peneliti memilih mengolah data yang diperoleh dalam proses scraping, dilanjutkan dengan preprocessing, memilih data pada kolom yang akan digunakan yaitu kolom content dan score serta menghilangkan data kosong dan data duplikat.

Pertimbangan utama dari peneliti adalah masalah waktu dan tidak praktis, serta bahwa sekiranya model dideploy dan diimplementasikan di dunia nyata, maka rasanya tidak praktis, jika setiap aplikasi DUKCAPIL mendapatkan komentar, akan terlebih dahulu dianalisa sentimen dengan persepsi manusia terlebih dahulu. Jauh lebih mudah, praktis dan orisinil, jika model yang telah dideploy bekerja langsung memberikan output yang diinginkan tanpa melalui campur tangan manusia.

4.2.3. Data Encoding

Data encoding adalah proses mentransformasikan data mentah ke dalam format numerik yang dapat diolah oleh algoritma pembelajaran mesin. Peneliti memilih Label encoding, yaitu mengubah data kategoris menjadi nilai numerik unik. Label encoding dirasa cocok dengan dataset yang dimiliki oleh peneliti. Sentimen negatif, netral dan positif diubah menjadi 0,1,2 dalam proses ini. Dimana sebelumnya, telah dibuat kolom sentiment, yang didefinisikan dari score 1-2 sebagai sentiment negatif, score 3 sebagai sentiment netral. Score 4-5 sebagai sentiment positif seperti yang ditampilkan pada Gambar 4.8 (Data Encoding).

```
[10] print(df_dataset.head())

                                             content  score sentiment
0  Tolong di tingkatkan performa yh biar support ...      2   negatif
1  Perlu dikembangkan lagi dan perkuat keamanan d...      1   negatif
2  VERIFIKASI DIDEPAN PETUGAS DUKCAPIL... KUNO ! ...      1   negatif
3  Masuk ke aplikasi saja susah, bagaimana pakai ...      1   negatif
4  Kenapa mau verifikasi kok selalu token selalu ...      1   negatif
```

Gambar 4. 8 (Data Encoding)

Pada gambar 4.9 (Text Length dan Word Count), kolom sentiment yang semula berisi kata negatif, telah berubah menjadi angka 0. Hal ini terjadi karena sentiment negatif, netral dan positif telah encoding menjadi 0, 1, 2. Kebetulan data yang ditampilkan, semuanya adalah sentiment negatif, sehingga semua data menunjukkan sentiment 0 , setelah proses label encoding. Sebagai penjelasan tambahan, bersamaan dalam gambar 4.9 (Text Length dan Word Count),

ditampilkan fitur baru yaitu text_length dan word_count , yang akan digunakan dalam proses selanjutnya dalam penelitian ini.

	content	score	sentiment	text_length	word_count
0	Tolong di tingkatkan performa yh biar support ...	2	0	57	10
1	Perlu dikembangkan lagi dan perkuat keamanan d...	1	0	61	8
2	VERIFIKASI DIDEPAN PETUGAS DUKCAPIL.... KUNO !	1	0	74	10
3	Masuk ke aplikasi saja susah, bagaimana pakai ...	1	0	333	56
4	Kenapa mau verifikasi kok selalu token selalu ...	1	0	152	22

Gambar 4. 9 (Text Length dan Word Count)

4.2.4 Feature Extraction

Dalam penelitian kali ini, dengan berbagai pilihan teknik feature extraction yang ada, peneliti memilih Teknik TF-IDF (Term Frequency-Inverse Document Frequency) digunakan untuk mengonversi teks ulasan menjadi fitur numerik. TF-IDF menghitung pentingnya setiap kata dalam dokumen relatif terhadap seluruh dataset. Proses ini menghasilkan matriks numerik yang merepresentasikan teks ulasan, yang kemudian digunakan sebagai input untuk model machine learning.

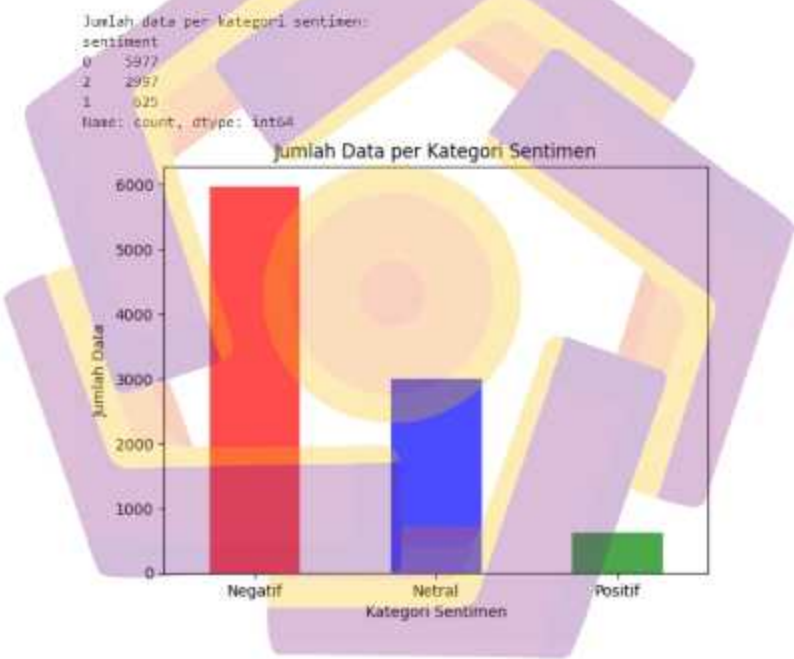
Penggunaan TF-IDF efektif dalam menangkap kata-kata yang memiliki kontribusi signifikan terhadap analisis sentimen, pada dataset ini. Hal ini dapat dibuktikan dari kinerja berbagai algoritma ensemble learning yang rata-rata memiliki metrik kinerja di atas 90%. Tentu menarik untuk mencoba berbagai teknik feature extraction yang lain seperti word embeddings, contextual embeddings maupun sentence atau document embeddings, namun dengan keterbatasan waktu, rasanya tidak bijak jika mencoba semua hal dan akhirnya mengakibatkan tidak terpenuhinya tenggat waktu yang telah ditetapkan dalam menyelesaikan penelitian thesis ini.

4.3. Penyeimbangan Data

Ketidakseimbangan data menjadi salah satu tantangan utama dalam penelitian ini, di mana ulasan negatif (bintang 1-2) lebih dominan dibandingkan ulasan positif (bintang 4-5). Teknik SMOTE (Synthetic Minority Over-sampling Technique) diterapkan untuk menghasilkan sampel sintetis dari kelas minoritas.

4.3.1. Distribusi Data tanpa SMOTE

Untuk menjawab rumusan masalah yang telah ditetapkan dalam penelitian ini, maka model yang dibangun akan dilatih dan diuji dengan dua macam data. Data yang pertama adalah data yang tidak seimbang, sementara yang kedua adalah data yang telah melalui teknik SMOTE agar terjadi keseimbangan data. Dataset ulasan terdiri dari berbagai skor yang mencerminkan persepsi pengguna terhadap aplikasi DUKCAPIL. Distribusi data yang tidak seimbang dapat dilihat pada gambar 4.10 (Distribusi Data tanpa SMOTE).



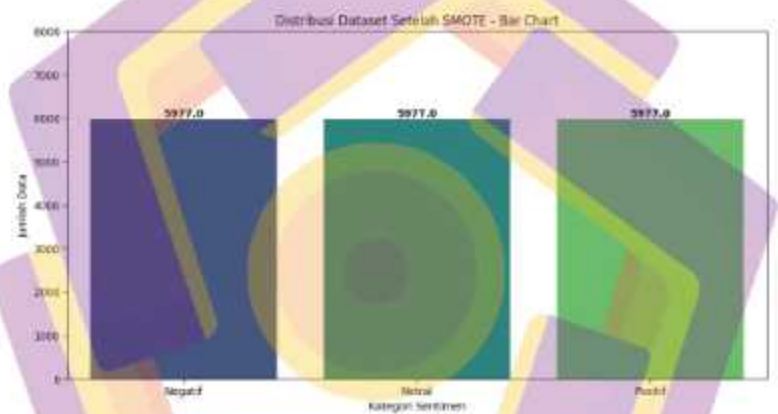
Gambar 4. 10 (Distribusi Data tanpa SMOTE)

Total ulasan: 9,599 (setelah penghapusan data kosong dan duplikasi), Kolom utama: content, score. Dalam grafik tersebut, nampak sekali bahwa data sangat tidak seimbang, dimana data berlabel positif, jauh lebih sedikit dibandingkan

dengan data berlabel netral dan negatif. Dimana data positif hanya 625, dibandingkan data netral 2997 dan data negatif 5977.

4.3.2. Teknik SMOTE untuk penyeimbangan data

Penerapan SMOTE berhasil meningkatkan distribusi data menjadi lebih seimbang, memungkinkan model untuk belajar pola dari setiap kelas secara adil. Grafik distribusi sebelum dan sesudah SMOTE menunjukkan perubahan yang signifikan, di mana jumlah ulasan di setiap kelas menjadi proporsional, yang disajikan dalam gambar 4.11 (Distribusi Dataset setelah SMOTE).



Gambar 4. 11 (Distribusi Dataset setelah SMOTE)

Dataset yang semulai tidak seimbang dengan didominasi label negatif serta sangat minimnya label netral, diseimbangkan sesuai dengan jumlah data label negatif yaitu 5977 data. Seimbangnya data, sangat berpengaruh terhadap kinerja model yang dibangun dalam penelitian ini.

4.4. Membangun Model, Pelatihan dan Pengujian

Selanjutnya penelitian dilakukan dengan membangun model dengan 10 algoritma yang telah dipilih dalam bab metodologi penelitian. Dataset dibagi menjadi data train (pelatihan) dan test (pengujian) dengan komposisi 80% data

pelatihan dan 20% data pengujian. Model dilatih dengan dataset yang tidak seimbang, sesuai dengan data yang telah melalui preprocessing dengan menghilangkan data duplikat dan data kosong pada dataset. Setelah itu, model dilatih dengan dataset yang telah melalui proses penyeimbangan data dengan teknik SMOTE.

Dari lima algoritma ensemble yang digunakan dalam penelitian (Random Forest, Extra Trees, AdaBoost, Gradient Boosting, dan XGBoost), masing-masing algoritma dapat diklasifikasikan berdasarkan metode ensemble yang digunakan:

- **Bagging:** Random Forest, Extra Trees
- **Boosting:** AdaBoost, Gradient Boosting, XGBoost
- **Voting dan Stacking:** Tidak ada algoritma dari kategori ini dalam penelitian ini.

Klasifikasi ini menunjukkan bahwa algoritma yang digunakan mencakup dua pendekatan ensemble utama, yaitu **Bagging** untuk meningkatkan stabilitas dan **Boosting** untuk meningkatkan akurasi dengan memperbaiki kesalahan secara iteratif.

4.4.1.1. Bagging

a. **Random Forest**

Dalam penelitian ini, kinerja algoritma ini terlihat unggul setelah kinerja algoritma Extra Trees yang menduduki peringkat pertama. Dengan Akurasi 93,3% , Presisi 93,5%, Recall 93,3% dan F1-Score 93,3% nampak bahwa Random Forest yang mewakili ensemble learning kategori bagging, cukup unggul dibandingkan kategori boosting, stacking maupun voting.

b. **Extra Trees (Extremely Randomized Trees)**

Dalam penelitian ini, Extra Trees menjadi jawara dengan menduduki peringkat pertama, dengan Accuracy 95,8% , Presisi 95,9%, Recalll 95,8%

dan F1-Score 95,8% nampak bahwa algoritma ini cukup akurat dan stabil dalam memprediksi label pada dataset DUKCAPIL ini.

4.4.1.2. Boosting

a. **AdaBoost (Adaptive Boosting)**

Berbeda dengan algoritma kategori boosting yang lain, kinerja algoritma ini cenderung buruk dalam mengolah dataset dalam penelitian ini. Akurasi 69,1% dibarengi dengan precision 70,6%, Recall 69,1% dan F1-Score 69,3% adalah cerminan dari rendahnya kinerja algoritma kategori boosting ini, pada penelitian kali ini. Dengan kinerja buruk ini, AdaBoost menjadi satu-satunya algoritma, yang kinerjanya dibawah algoritma individual seperti Logistic Regression dan Decision Tree, yang berkinerja dikisaran 83% di semua metrik pengukuran kinerja.

b. **Gradient Boosting**

Dibandingkan dengan AdaBoost yang telah kita ulas duluan, algoritma ini cukup lumayan kinerjanya. Dengan Accuracy 85,6%, Presisi 86,1%, Recall 85,6% dan F1-Score 85,6%, kinerja algoritma ensemble ini, setidaknya sudah lebih baik dibandingkan dengan algoritma individual seperti Decision Tree dan Logistic Regression yang berkinerja dikisaran 83%.

c. **XGBoost (Extreme Gradient Boosting)**

Dalam penelitian kali ini, XGBoost berada diperingkat ketiga dengan kinerja di atas 90% hampir semua metrik kinerja. Accuracy 90,9%, Precision 91,3% , Recall 90,9% dan F1-Score 90,9% membuat algoritma ini layak dibanggakan, karena turut berkontribusi atas keunggulan kinerja algoritma ensemble dalam kontes ini.

4.4.1.3. Stacking

Hasil evaluasi performa model **Stacking Ensemble** disajikan pada gambar 4.12 (Stacking Ensemble). Model ini menggabungkan kekuatan

beberapa algoritma dasar dengan menggunakan model meta-learner untuk meningkatkan akurasi prediksi. Berikut adalah penjelasan rinci dari hasil tersebut:

```
Training Stacking Ensemble...
Training completed in 3059.91 seconds.

Stacking Ensemble Metrics:
Accuracy: 0.9358
Precision: 0.9365
Recall: 0.9358
F1-Score: 0.9360
              precision    recall  f1-score   support

           0       0.90      0.93      0.91      1165
           1       0.99      0.97      0.98      1181
           2       0.92      0.90      0.91      1191

    accuracy                           0.94      3537
   macro avg       0.94      0.94      0.94      3537
weighted avg       0.94      0.94      0.94      3537
```

Gambar 4. 12 (Stacking Ensemble)

Akurasi model menunjukkan bahwa sebanyak 93,58% , presisi 93,65% , Recall 93,58% dan F1-Score 93,60% menunjukkan bahwa, Model bekerja dengan sangat baik pada dataset. Meskipun demikian, pendekatan stacking ini, belum mampu mengungguli kinerja dengan metode bagging.

Kelas sentimen netral memiliki performa terbaik pada angka 99%, sementara sentimen negatif dan positif sedikit lebih rendah diangka 90% dan 92%, tetapi tetap kompetitif. Model ini layak dipertimbangkan sebagai pendekatan utama untuk tugas analisis sentimen karena kemampuannya dalam menangkap pola yang kompleks melalui penggabungan beberapa algoritma dasar.

4.4.1.4. Voting

Hasil evaluasi model **Voting Ensemble** disajikan pada gambar 4.13 (Voting Ensemble). Model ini menggabungkan prediksi dari beberapa algoritma dasar

menggunakan pendekatan voting untuk meningkatkan akurasi dan stabilitas prediksi. Berikut adalah rincian hasil evaluasi:

```
Voting Ensemble Metrics:
Accuracy: 0.9141
Precision: 0.9162
Recall: 0.9141
F1-Score: 0.9143
              precision    recall  f1-score   support

           0       0.86      0.93      0.89      1165
           1       0.97      0.94      0.95      1181
           2       0.92      0.87      0.90      1191

    accuracy                           0.91      3537
   macro avg       0.92      0.91      0.91      3537
weighted avg       0.92      0.91      0.91      3537
```

Gambar 4. 13 (Voting Ensemble)

Metrik Keseluruhan

Akurasi model menunjukkan bahwa sebanyak 91,41% prediksi yang dihasilkan oleh model sesuai dengan label sebenarnya dalam dataset. Ini mencerminkan performa yang sangat baik dalam menangkap pola data. Presisi 91,62%, menunjukkan kemampuan model untuk meminimalkan jumlah prediksi positif palsu (false positives), memastikan prediksi yang lebih tepat. Recall 91,41% menunjukkan bahwa model ini mampu mendeteksi data positif yang sebenarnya dengan sangat baik. F1-Score 91,43%, yang menggabungkan presisi dan recall, menunjukkan keseimbangan performa model dalam membuat prediksi akurat.

Performa Berdasarkan Kelas

Kelas 0 (Sentimen Negatif): Presisi: 86%, Recall: 93%, F1-Score: 89% menunjukkan model cukup baik dalam mendeteksi sentimen negatif, dengan recall yang tinggi meskipun presisi sedikit lebih rendah.

Kelas 1 (Sentimen Netral): Presisi: 97%, Recall: 94%, F1-Score: 95%. Sentimen netral memiliki performa terbaik di antara ketiga kelas, dengan presisi yang hampir sempurna.

Kelas 2 (Sentimen Positif): Presisi: 92%, Recall: 87%, F1-Score: 90% menunjukkan Model memiliki kinerja yang cukup baik dalam mendeteksi sentimen positif, meskipun recall sedikit lebih rendah dibandingkan presisi.

Statistik Rata-Rata

Macro Average: Presisi: 92%, Recall: 91%, F1-Score: 91%
Statistik ini menunjukkan rata-rata tidak berbobot dari metrik di seluruh kelas, mencerminkan performa yang merata.

Weighted Average:

Presisi: 92%, Recall: 91%, F1-Score: 91% Statistik ini mempertimbangkan proporsi data pada setiap kelas, memberikan gambaran performa keseluruhan model.

Model **Voting Ensemble** menunjukkan performa yang sangat baik, dengan akurasi, presisi, recall, dan F1-Score di atas 91%. Sentimen netral memiliki performa terbaik dengan presisi dan F1-Score yang tinggi, sedangkan sentimen negatif dan positif menunjukkan hasil yang baik namun sedikit lebih rendah. Model ini memberikan solusi yang andal dalam analisis sentimen dengan memanfaatkan pendekatan voting untuk meningkatkan stabilitas prediksi.

4.4.1.5 Voting ensemble dikombinasi bagging + Boosting

Dalam penelitian ini, peneliti mencoba menggunakan pendekatan **Voting Ensemble** dengan kombinasi metode **Bagging** dan **Boosting**, yang diharapkan dapat secara signifikan meningkatkan keakuratan dan stabilitas prediksi dibandingkan dengan model individu.

Dalam gambar 4.14 (Voting Ensemble Bagging dan Boosting) adalah potongan koding, untuk menggambarkan model dengan kombinasi bagging + boosting dalam konsep voting ensemble yang dimaksud.

```
# Definisikan Bagging dan Boosting models
bagging_model = ExtraTreesClassifier(n_estimators=200, random_state=42)
boosting_model_1 = XGBClassifier(use_label_encoder=False, eval_metric='mlogloss', random_state=42)
boosting_model_2 = GradientBoostingClassifier(n_estimators=100, random_state=42)

# Voting Classifier (soft voting untuk probabilitas lebih baik)
voting_clf = VotingClassifier(
    estimators=[
        ('et', bagging_model),
        ('xgb', boosting_model_1),
        ('gb', boosting_model_2)
    ],
    voting='soft'
)
```

Gambar 4. 14 (Voting Ensemble Bagging dan Boosting)

Selanjutnya pada gambar 4.15 (Metrik Kinerja) disajikan metrik kinerja model diatas.

```
Bagging + Boosting Voting Ensemble Metrics:
Accuracy: 0.8908
Precision: 0.8926
Recall: 0.8908
F1-Score: 0.8905
```

Gambar 4. 15 (Metrik Kinerja)

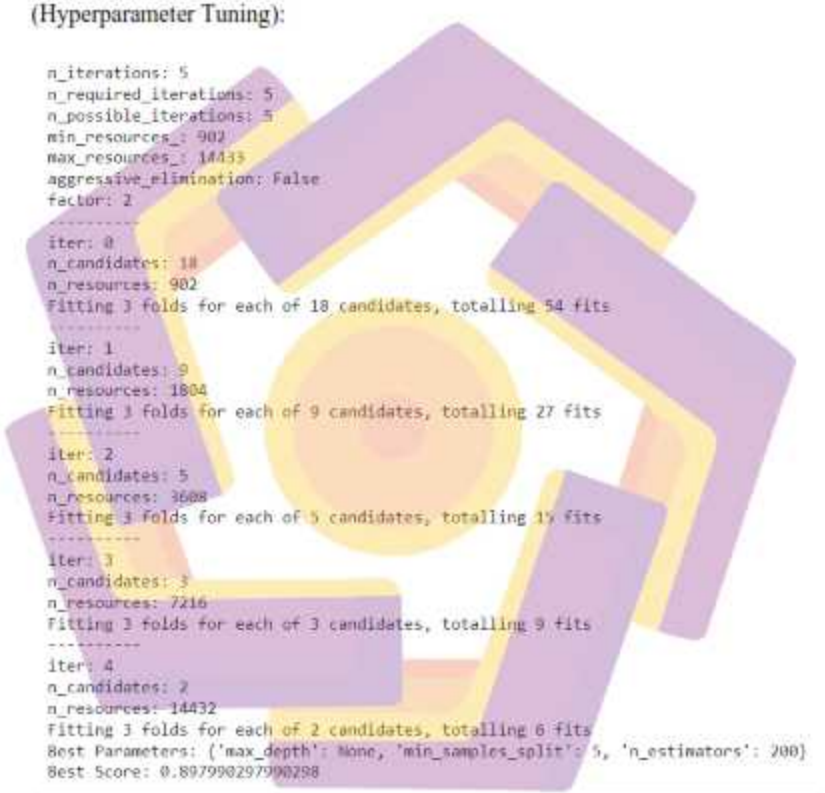
Hasil evaluasi model menunjukkan performa yang cukup baik dengan metrik berikut:

Accuracy (Akurasi): 0.8908, menunjukkan bahwa model berhasil memprediksi dengan benar sekitar 89.08% dari keseluruhan data. Precision 89,26%, Recall 89,08, F1-Score: 89,05% merupakan rata-rata harmonis antara presisi dan recall, mengindikasikan keseimbangan kinerja model.

Meskipun demikian, kinerja dari model kombinasi ini, belum menghasilkan kinerja yang lebih baik, dibandingkan dengan ketika model bekerja secara individual, terutama ketika algoritma Extra Trees maupun Random Forest secara individual.

4.4.2. Hyperparameter Tuning

Proses hyperparameter tuning dilakukan menggunakan Halving Random Search, sebuah metode yang mengoptimalkan waktu komputasi dengan secara bertahap mengeliminasi kandidat yang kurang menjanjikan berdasarkan performa mereka. Penjelasan detail dari hasil proses tersebut dapat dilihat pada Gambar 4.16 (Hyperparameter Tuning):



Gambar 4. 16 (Hyperparameter Tuning)

Proses hyperparameter tuning, dapat dilihat dari gambar 4.16. Proses dilakukan dengan total lima iterasi dilakukan selama proses pencarian parameter terbaik. Setiap kandidat pada iterasi awal menggunakan setidaknya 902 sampel. Pada iterasi terakhir, kandidat terbaik menggunakan total 14.433 sampel untuk

pelatihan. Jumlah kandidat dikurangi separuh pada setiap iterasi, menghasilkan pengurangan secara eksponensial. Kandidat tidak dieliminasi lebih agresif di luar konfigurasi faktor eliminasi.

Proses Iterasi

Iterasi 0:

Jumlah Kandidat: 18, Sumber Daya per Kandidat: 902 sampel, Jumlah Total Fit: 54 (3 fold cross-validation untuk setiap kandidat). Pada iterasi ini, semua kandidat dievaluasi secara awal menggunakan data dengan sumber daya minimum.

Iterasi 1:

Jumlah Kandidat: 9 (pengurangan 50%), Sumber Daya per Kandidat: 1.804 sampel, Jumlah Total Fit: 27. Hanya kandidat yang menjanjikan dari iterasi sebelumnya yang dievaluasi ulang dengan sumber daya yang lebih besar.

Iterasi 2:

Jumlah Kandidat: 5, Sumber Daya per Kandidat: 3.608 sampel, Jumlah Total Fit: 15. Proses ini semakin menyaring kandidat, memastikan hanya model dengan performa terbaik yang melanjutkan ke iterasi berikutnya.

Iterasi 3:

Jumlah Kandidat: 3, Sumber Daya per Kandidat: 7.216 sampel, Jumlah Total Fit: 9. Kandidat yang tersisa semakin difokuskan pada data yang lebih besar untuk mengevaluasi performa mereka dengan lebih akurat.

Iterasi 4:

Jumlah Kandidat: 2 (final), Sumber Daya per Kandidat: 14.433 sampel, Jumlah Total Fit: 6. Dua kandidat terbaik dievaluasi menggunakan seluruh data, dan parameter terbaik dipilih berdasarkan skor validasi.

3. Hasil Akhir

Parameter Terbaik: max_depth: None (tidak ada batasan kedalaman pohon). min_samples_split: 5 (minimum sampel untuk melakukan split). n_estimators: 200 (jumlah pohon keputusan). Best Score: 0.89799 (akurasi terbaik dari kandidat terpilih).

Proses Halving Random Search berhasil mengidentifikasi parameter optimal untuk model yang diuji. Dengan pendekatan bertahap ini: Waktu komputasi dapat dihemat dengan secara efisien mengurangi kandidat yang kurang menjanjikan. Parameter terbaik (max_depth, min_samples_split, dan n_estimators) menunjukkan bahwa model dengan konfigurasi tersebut memberikan performa yang paling optimal dengan akurasi hampir 90%. Pendekatan ini memastikan bahwa model dapat mencapai performa terbaik dengan waktu pelatihan yang lebih singkat dibandingkan metode pencarian parameter konvensional.

4.5. Evaluasi Kinerja Model

Setelah model dijalankan, diperoleh metrik kinerja dari model yang meliputi accuracy, precision, recall dan F1-Score. Semua model dijalankan satu persatu dengan memberikan metrik kinerja masing-masing.

4.5.1. Metrik kinerja model tanpa SMOTE

Pengujian model, dimulai dengan menguji model dengan data yang tidak seimbang dan tanpa menggunakan teknik SMOTE untuk menyeimbangkan data. Setelah melalui pelatihan 80% data, menghasilkan tabel perbandingan kinerja masing-masing model seperti disajikan pada gambar 4.17 (Metrik Kinerja model tanpa SMOTE)

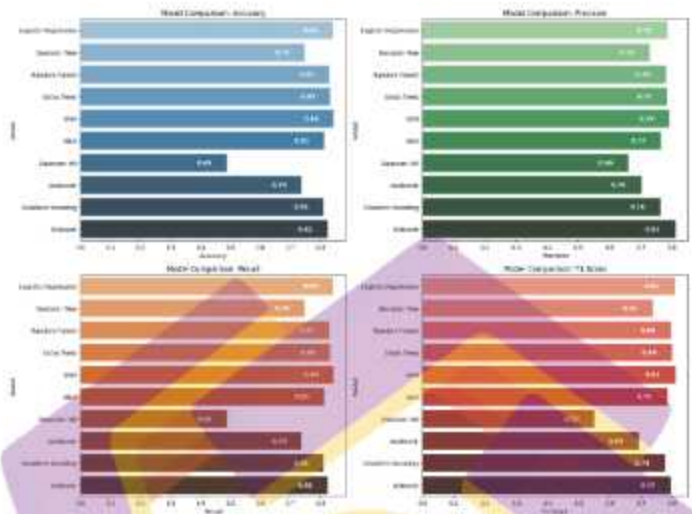
Summary of Results:

	Model	Accuracy	Precision	Recall	F1-Score
0	Logistic Regression	0.841146	0.785617	0.841146	0.807969
1	Decision Tree	0.746354	0.729254	0.746354	0.737456
2	Random Forest	0.829167	0.781872	0.829167	0.795386
3	Extra Trees	0.832292	0.785157	0.832292	0.798802
4	SVM	0.843750	0.793438	0.843750	0.809918
5	KNN	0.812500	0.765702	0.812500	0.784355
6	Gaussian NB	0.487500	0.661421	0.487500	0.548743
7	AdaBoost	0.736979	0.702819	0.736979	0.692284
8	Gradient Boosting	0.810937	0.763589	0.810937	0.776566
9	XGBoost	0.824479	0.812194	0.824479	0.794375

Gambar 4. 17 (Metrik Kinerja model tanpa SMOTE)

Dengan dataset yang tidak seimbang, seperti telah disajikan sebelumnya, kontes 10 algoritma yang terdiri atas 5 algoritma individual (Logistic Regression, Decision Tree, SVM, KNN, Gaussian NB) dengan 5 algoritma Ensemble Learning (Random Forest, Extra Trees, AdaBoost, Gradient Boosting dan XGBoost) dimenangkan oleh Support Vector Machine dengan Accuracy 84,4%, Precision 79,3%, Recall 84,4% dan F1-Score 80,99%. Hal yang menarik adalah Posisi pertama dan kedua ditempati oleh algoritma individual yaitu Support Vector Machine dan Logistic Regression, Dilanjutkan oleh 5 algoritma Ensemble Learning ygn semuanya berkinerja Accuracy di atas 80%.

Perbandingan kinerja keseluruhan algoritma, ditampilkan dalam grafik batang pada gambar 4.18 (Perbandingan kinerja seluruh Algoritma).



Gambar 4. 18 (Perbandingan kinerja seluruh Algoritma)

Pada gambar 4.18 (Perbandingan kinerja seluruh Algoritma) dapat tersaji bahwa Accuracy tertinggi tercapai diangka 84% oleh algoritma Logistic Regression dan SVM, Precision tertinggi dicapai oleh XGBoost dengan angka 81%, Recall tertinggi dicapai oleh SVM dan Logistic Regression serta terakhir F1-Score dipertahankan oleh SVM dan Logistic Regression di angka 81%

4.5.2. Metrik kinerja model dengan SMOTE

Selanjutnya model dijalankan dengan dataset yang telah melalui proses penyeimbangan data dengan teknik SMOTE. Hasil kinerja model disajikan laporan tabel perbandingan kinerja masing-masing model seperti disajikan dalam gambar 4.19 (Teknik SMOTE) yang menunjukkan hasil evaluasi berdasarkan accuracy, precision, recall, dan F1-Score masing-masing model :

Summary of Results:

	Model	Accuracy	Precision	Recall	F1-Score
0	Logistic Regression	0.838863	0.842309	0.838863	0.839737
1	Decision Tree	0.832172	0.832015	0.832172	0.831938
2	Random Forest	0.933092	0.935455	0.933092	0.933098
3	Extra Trees	0.958461	0.959284	0.958461	0.958405
4	SVM	0.439922	0.440190	0.439922	0.416716
5	KNN	0.686925	0.706514	0.686925	0.647555
6	Gaussian NB	0.777530	0.820590	0.777530	0.771917
7	AdaBoost	0.691386	0.706483	0.691386	0.693942
8	Gradient Boosting	0.856147	0.861225	0.856147	0.856020
9	XGBoost	0.909116	0.913898	0.909116	0.909565

Gambar 4. 19 (Teknik SMOTE)

Dalam penelitian ini, berbagai model machine learning diuji untuk mengevaluasi kinerjanya dalam menyelesaikan masalah yang dihadapi. Evaluasi dilakukan berdasarkan empat metrik utama: Akurasi, Presisi, Recall, dan F1-Score. Hasilnya menunjukkan bahwa model **Extra Trees** memberikan performa terbaik dengan akurasi 95,85%, presisi 95,93%, recall 95,85%, dan F1-Score 95,85%. Model ini unggul dibandingkan model lainnya karena mampu menangkap pola dengan lebih baik.

Model **Random Forest** berada di peringkat kedua dengan akurasi 93,31% dan performa metrik lainnya yang juga sangat baik. Model **XGBoost** mengikuti dengan akurasi 90,91%, menunjukkan kemampuan yang kompetitif namun masih di bawah Extra Trees. Di sisi lain, model **Gradient Boosting** mencapai akurasi 85,61%, yang meskipun cukup tinggi, tidak sebanding dengan performa Random Forest dan Extra Trees.

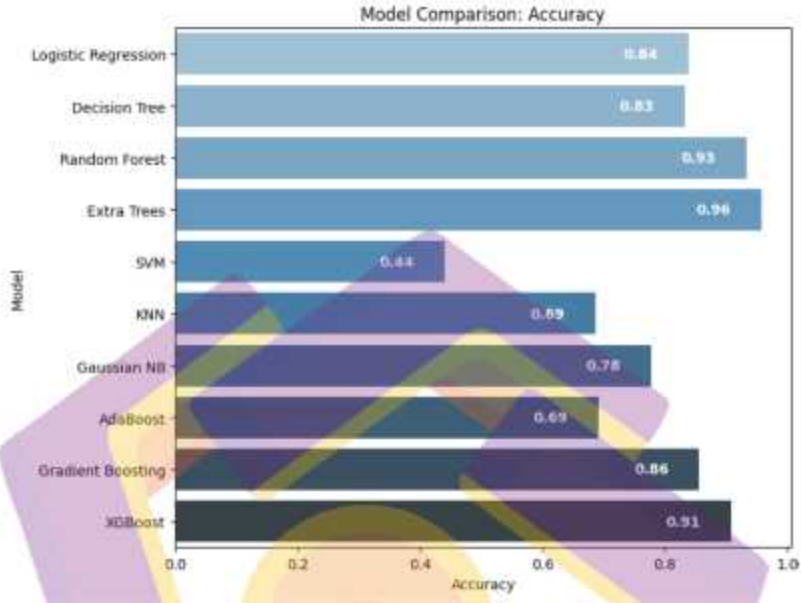
Sebaliknya, model **SVM (Support Vector Machine)** menunjukkan performa terburuk dengan akurasi 43,99%, yang mengindikasikan bahwa model ini tidak cocok untuk data yang digunakan dalam penelitian ini. Model **KNN** dan **AdaBoost** juga memiliki performa yang relatif rendah, masing-masing dengan akurasi 68,69% dan 69,14%.

Untuk model probabilistik seperti **Gaussian Naïve Bayes**, akurasi tercatat sebesar 77,75%, menunjukkan performa yang cukup memadai dalam menangani dataset ini. Model berbasis pohon keputusan seperti **Decision Tree** menghasilkan akurasi sebesar 83,21%, yang cukup kompetitif tetapi masih kalah dibandingkan Random Forest dan Extra Trees.

Keseluruhan hasil ini menunjukkan bahwa model berbasis ensemble seperti Extra Trees dan Random Forest memiliki keunggulan signifikan dalam menangkap pola dan memberikan prediksi yang lebih akurat pada dataset yang digunakan. Temuan ini mendukung penggunaan model ensemble dalam penelitian atau aplikasi nyata dengan karakteristik data serupa.

4.5.2.1. Komparasi kinerja accuracy model

Salah satu laporan visual dari proses perbandingan model, disajikan dalam bentuk Grafik Model Comparison: Accuracy. Grafik tersebut menyajikan perbandingan akurasi dari sepuluh algoritma yang. Hasil evaluasi menunjukkan perbedaan signifikan antara algoritma berbasis ensemble learning dan non-ensemble learning. Penjabaran rinci berdasarkan hasil grafik seperti dituangkan dalam Gambar 4.20 (Grafik Model Comparison: Accuracy):



Gambar 4. 20 (Grafik Model Comparison: Accuracy)

Algoritma dengan Akurasi Tertinggi

1. **Extra Trees** mencatat akurasi tertinggi sebesar **96%**, menjadikannya algoritma paling unggul dalam penelitian ini. Model ini mampu menangkap pola data dengan sangat baik, sehingga menghasilkan prediksi yang akurat.
2. **Random Forest** berada di posisi kedua dengan akurasi **93%**, menunjukkan performa yang juga sangat kompetitif.
3. **XGBoost** menempati peringkat ketiga dengan akurasi **91%**, mendekati hasil Random Forest dan tetap dalam kategori kinerja tinggi.

Algoritma Non-Ensemble dengan Performa Baik

1. **Logistic Regression** dan **Decision Tree** masing-masing mencatat akurasi sebesar **84%** dan **83%**, menunjukkan bahwa algoritma non-ensemble ini masih dapat bersaing, meskipun tidak sebaik model berbasis ensemble learning.
2. **Gaussian Naïve Bayes** juga memiliki performa yang cukup baik dengan akurasi **78%**, menjadikannya salah satu algoritma probabilistik yang cocok untuk analisis sentimen ini.

Algoritma dengan Akurasi Rendah

1. **Support Vector Machine (SVM)** mencatat akurasi paling rendah, yaitu **44%**, menunjukkan bahwa algoritma ini kurang cocok untuk dataset yang digunakan dalam penelitian ini.
2. **KNN (K-Nearest Neighbors)** dan **AdaBoost** masing-masing mencatat akurasi **69%**, yang tergolong moderat tetapi tidak dapat bersaing dengan model ensemble terbaik.

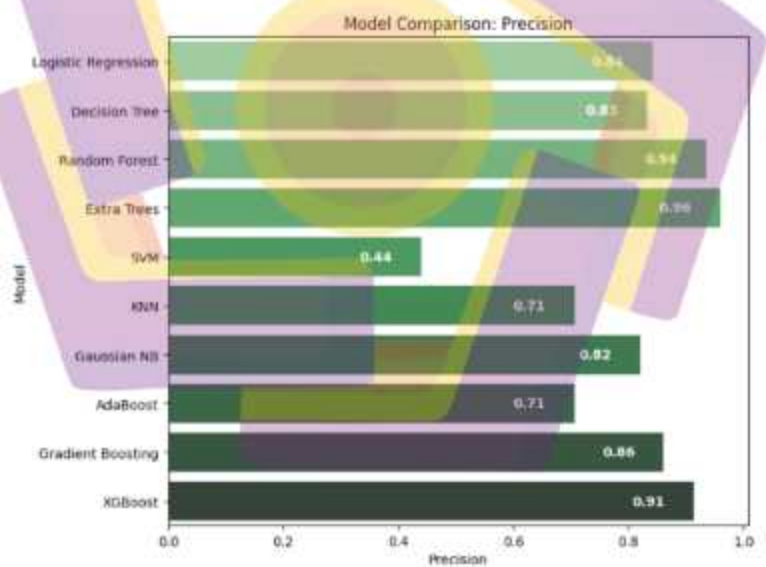
Performa Keseluruhan Algoritma Ensemble Learning

Secara keseluruhan, algoritma ensemble learning (Extra Trees, Random Forest, XGBoost, Gradient Boosting, dan AdaBoost) menunjukkan performa yang jauh lebih baik dibandingkan dengan algoritma non-ensemble learning. Model seperti Extra Trees, Random Forest, dan XGBoost secara konsisten memberikan akurasi yang lebih tinggi, membuktikan efektivitas pendekatan ensemble learning dalam meningkatkan performa prediksi.

Hasil grafik ini menguatkan bahwa algoritma berbasis ensemble learning, khususnya **Extra Trees**, merupakan pilihan terbaik pada dataset yang digunakan. Dengan akurasi tertinggi dan konsistensi yang baik, algoritma ensemble learning memberikan keunggulan signifikan dibandingkan dengan algoritma non-ensemble learning.

4.5.2.2. Komparasi kinerja presisi model

Grafik "Model Comparison: Precision" menggambarkan perbandingan presisi dari sepuluh algoritma yang digunakan. Presisi adalah metrik yang mengukur ketepatan prediksi positif yang dibuat oleh model, dan hasilnya menunjukkan kinerja yang signifikan antara model ensemble dan non-ensemble learning seperti ditampilkan dalam Gambar 4.21 (Grafik Model Comparison: Precision).



Gambar 4. 21 (Grafik Model Comparison: Precision)

Algoritma dengan Presisi Tertinggi

1. **Extra Trees** mencatat presisi tertinggi sebesar **96%**, menegaskan keunggulannya dalam memprediksi data secara akurat dengan meminimalkan kesalahan prediksi positif.
2. **Random Forest** mengikuti di posisi kedua dengan presisi **94%**, menunjukkan kemampuan yang hampir setara dengan Extra Trees.
3. **XGBoost** berada di peringkat ketiga dengan presisi **91%**, mengukuhkan model ini sebagai salah satu yang terbaik dalam penelitian ini.

Algoritma Non-Ensemble dengan Presisi Baik

1. **Logistic Regression** dan **Decision Tree** mencatat presisi masing-masing sebesar **84%** dan **83%**, menunjukkan bahwa meskipun mereka adalah algoritma non-ensemble, hasilnya cukup kompetitif.
2. **Gaussian Naïve Bayes** memiliki presisi sebesar **82%**, yang mencerminkan performa yang solid dalam tugas analisis sentimen.

Algoritma dengan Presisi Rendah

1. **Support Vector Machine (SVM)** mencatat presisi paling rendah sebesar **44%**, menegaskan bahwa algoritma ini tidak cocok untuk dataset yang digunakan.

2. **KNN (K-Nearest Neighbors)** dan **AdaBoost** masing-masing mencatat presisi sebesar **71%**, yang tergolong sedang tetapi jauh dari performa model terbaik.

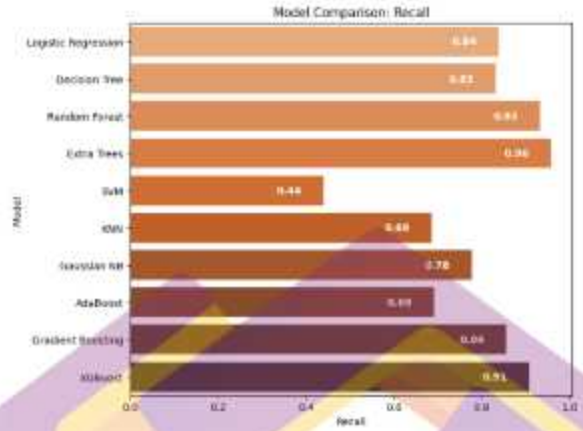
Performa Keseluruhan Algoritma Ensemble Learning

Secara keseluruhan, algoritma ensemble learning (Extra Trees, Random Forest, XGBoost, Gradient Boosting, dan AdaBoost) memberikan hasil presisi yang lebih tinggi dibandingkan dengan algoritma non-ensemble learning. Extra Trees, Random Forest, dan XGBoost secara khusus menonjol dengan presisi di atas 90%, menunjukkan keunggulan mereka dalam meminimalkan prediksi positif palsu.

Hasil grafik ini menguatkan bahwa algoritma berbasis ensemble learning, khususnya **Extra Trees**, adalah pilihan terbaik dalam hal presisi. Dengan presisi tertinggi, model ini menunjukkan kemampuan yang sangat andal dalam membuat prediksi positif yang akurat. Algoritma non-ensemble seperti Logistic Regression dan Decision Tree juga cukup kompetitif, tetapi belum dapat mengungguli model ensemble terbaik.

4.5.2.3 Komparasi kinerja recall model

Grafik "Model Comparison: Recall" menampilkan perbandingan nilai recall dari sepuluh algoritma yang digunakan seperti yang dituangkan pada Gambar 4.22 (Grafik Model Comparison: Recall).



Gambar 4. 22 (Grafik Model Comparison: Recall)

Algoritma dengan Recall Tertinggi

1. **Extra Trees** mencapai nilai recall tertinggi sebesar **96%**, menegaskan kemampuannya untuk mengidentifikasi seluruh data positif dengan sangat baik.
2. **Random Forest** berada di posisi kedua dengan nilai recall **93%**, menunjukkan performa yang sangat baik dalam mendeteksi data positif.
3. **XGBoost** menduduki peringkat ketiga dengan nilai recall **91%**, yang juga sangat kompetitif dan mendukung kekuatannya sebagai salah satu model ensemble terbaik.

Algoritma Non-Ensemble dengan Recall Baik

1. **Logistic Regression** dan **Decision Tree** masing-masing mencatat nilai recall sebesar **84%** dan **83%**, menunjukkan performa yang memadai di antara algoritma non-ensemble.
2. **Gaussian Naïve Bayes** memiliki nilai recall sebesar **78%**, yang mencerminkan kesesuaian algoritma ini untuk mendeteksi data positif dalam tugas analisis sentimen.

Algoritma dengan Recall Rendah

1. **Support Vector Machine (SVM)** mencatat nilai recall terendah sebesar **44%**, mengindikasikan bahwa algoritma ini kurang cocok untuk dataset yang digunakan dalam penelitian ini.
2. **KNN (K-Nearest Neighbors)** dan **AdaBoost** masing-masing mencatat nilai recall sebesar **69%**, yang tergolong sedang tetapi tidak dapat bersaing dengan model ensemble yang lebih kuat.

Performa Keseluruhan Algoritma Ensemble Learning

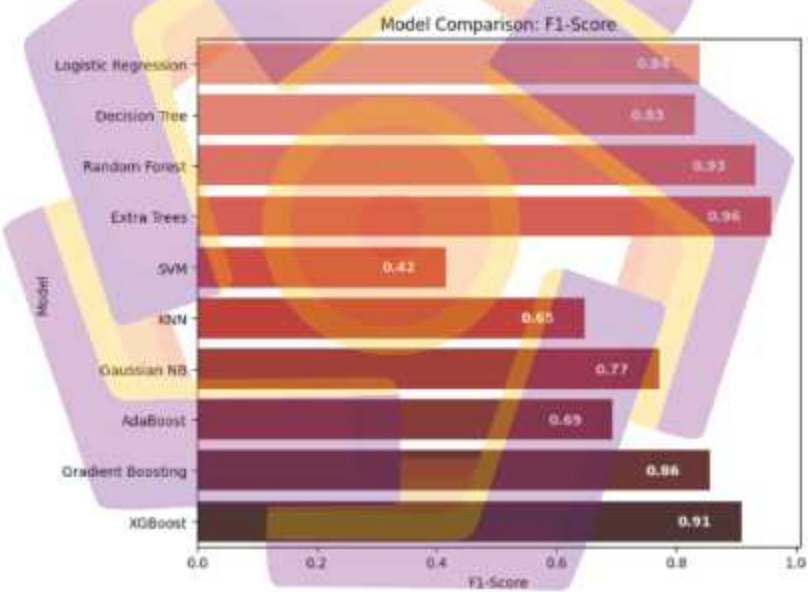
Secara keseluruhan, algoritma ensemble learning (Extra Trees, Random Forest, XGBoost, Gradient Boosting, dan AdaBoost) memiliki nilai recall yang lebih tinggi dibandingkan dengan algoritma non-ensemble. Model seperti Extra Trees, Random Forest, dan XGBoost secara konsisten menunjukkan keunggulan dalam mendeteksi data positif dengan presisi tinggi, menjadikannya pilihan yang lebih andal untuk analisis sentimen.

Hasil grafik ini memperkuat bahwa algoritma berbasis ensemble learning, terutama **Extra Trees**, adalah pilihan terbaik untuk tugas analisis sentimen

berdasarkan nilai recall. Dengan recall tertinggi, model ini menunjukkan kemampuan yang unggul dalam mengidentifikasi data positif yang sebenarnya. Meskipun algoritma non-ensemble seperti Logistic Regression dan Decision Tree memberikan hasil yang memadai, mereka belum dapat menyaingi keunggulan yang ditawarkan oleh model ensemble learning.

4.5.2.4 Komparasi kinerja f1-score model

Grafik "Model Comparison: F1-Score" menggambarkan perbandingan nilai F1-Score dari sepuluh algoritma yang digunakan seperti yang dituangkan pada Gambar 4.23 (Grafik Model Comparison: F1-Score).



Gambar 4. 23 (Grafik Model Comparison: F1-Score)

Algoritma dengan F1-Score Tertinggi

1. **Extra Trees** mencatat F1-Score tertinggi sebesar **96%**, menunjukkan kinerjanya yang sangat unggul dalam memberikan prediksi yang seimbang antara presisi dan recall.

2. **Random Forest** berada di posisi kedua dengan nilai F1-Score **93%**, memperlihatkan performa yang sangat baik untuk tugas analisis sentimen ini.
3. **XGBoost** menempati peringkat ketiga dengan nilai F1-Score **91%**, yang juga mendukung posisinya sebagai salah satu algoritma terbaik.

Algoritma Non-Ensemble dengan F1-Score Baik

1. **Logistic Regression** dan **Decision Tree** masing-masing mencatat F1-Score sebesar **84%** dan **83%**, menunjukkan bahwa algoritma non-ensemble ini mampu memberikan performa yang cukup kompetitif.
2. **Gaussian Naïve Bayes** memiliki F1-Score sebesar **77%**, yang mencerminkan hasil yang baik meskipun tidak sebanding dengan algoritma berbasis ensemble.

Algoritma dengan F1-Score Rendah

1. **Support Vector Machine (SVM)** mencatat nilai F1-Score terendah sebesar **42%**, mengindikasikan bahwa algoritma ini tidak cocok untuk dataset yang digunakan.
2. **KNN (K-Nearest Neighbors)** dan **AdaBoost** masing-masing mencatat nilai F1-Score sebesar **65%** dan **69%**, yang tergolong sedang tetapi jauh dari performa model terbaik.

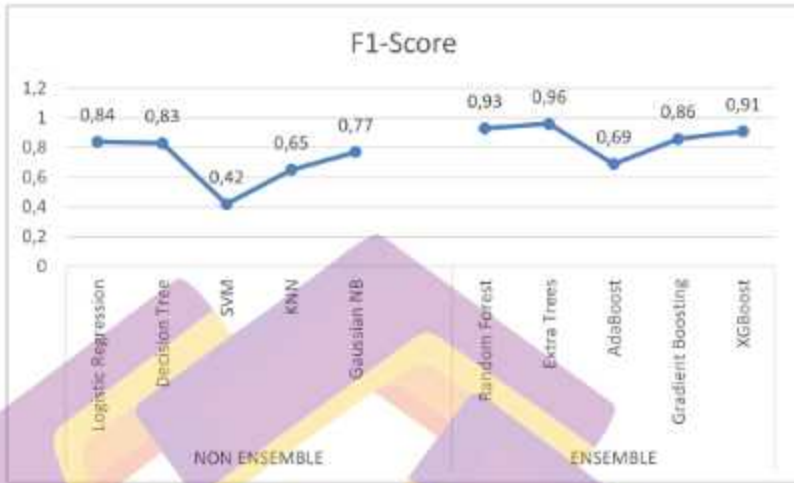
Performa Keseluruhan Algoritma Ensemble Learning

Secara keseluruhan, algoritma berbasis ensemble learning (Extra Trees, Random Forest, XGBoost, Gradient Boosting, dan AdaBoost) memberikan hasil F1-Score yang lebih tinggi dibandingkan algoritma non-ensemble learning. Model seperti Extra Trees dan Random Forest secara konsisten menunjukkan keunggulan dalam menghasilkan prediksi yang akurat dan seimbang, menjadikannya pilihan yang lebih andal untuk analisis sentimen.

Hasil grafik ini menguatkan bahwa algoritma berbasis ensemble learning, terutama **Extra Trees**, adalah model terbaik berdasarkan nilai F1-Score. Dengan F1-Score tertinggi, model ini menunjukkan kemampuannya dalam memberikan prediksi yang seimbang antara presisi dan recall. Algoritma non-ensemble seperti Logistic Regression dan Decision Tree memberikan performa yang cukup baik, tetapi mereka belum dapat menyamai keunggulan yang ditawarkan oleh model ensemble.

4.5.3. Perbandingan kinerja model individual vs Ensemble Learning

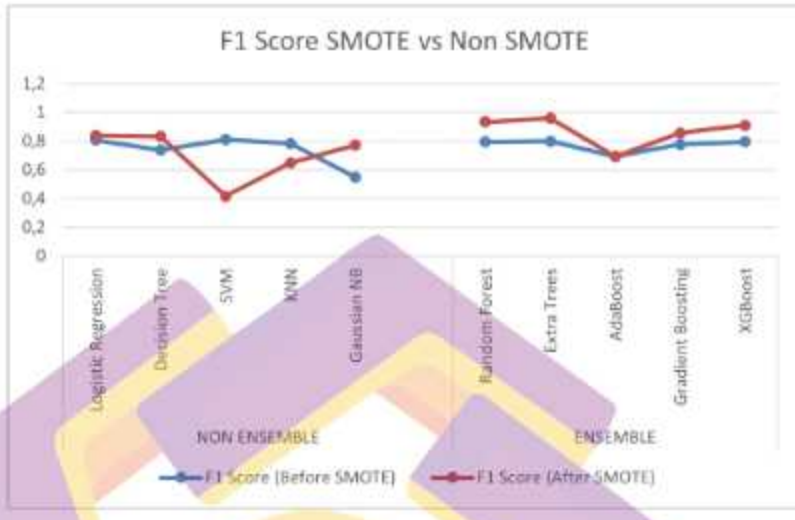
Untuk menjawab rumusan masalah yang telah ditetapkan sebelumnya, dibagian awal penelitian ini yaitu : Bagaimana perbandingan kinerja algoritma model individual dengan model ensemble dalam analisis sentimen ulasan aplikasi DUKCAPIL? Model yang mana yang memberikan hasil yang lebih akurat dan andal?. Dari hasil pelatihan dan pengujian model ditemukan bahwa model Ensemble Learning, setelah menggunakan SMOTE ternyata memiliki kinerja lebih unggul dibandingkan semua model individual, seperti ditampilkan dalam Gambar 4.24 (Grafik Perbandingan Kinerja Model Non Ensemble vs Ensemble).



Gambar 4. 24 (Grafik Perbandingan Kinerja Model Non Ensemble vs Ensemble)

4.5.4. Perbandingan Kinerja model tanpa dan dengan SMOTE

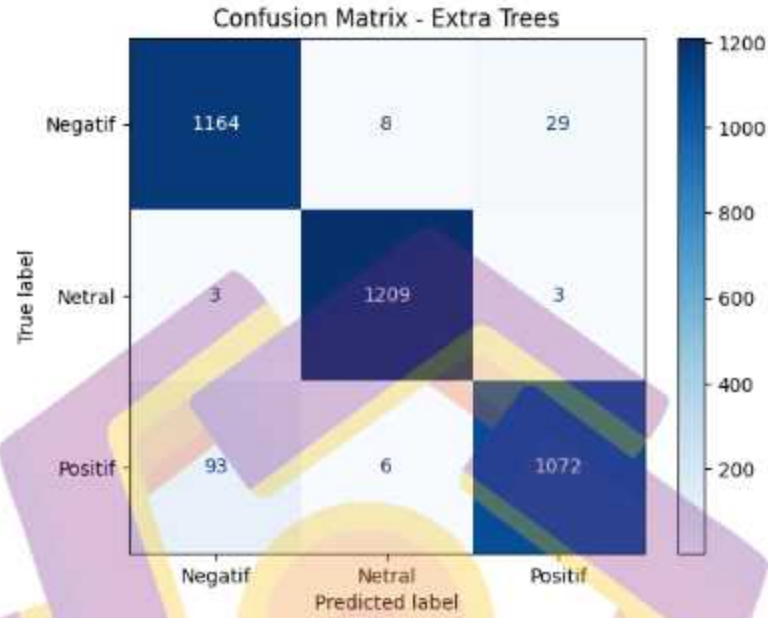
Menjawab rumusan masalah kedua, yaitu : Bagaimana pengaruh penggunaan teknik SMOTE untuk mengatasi ketidakseimbangan data dalam meningkatkan kinerja model analisis sentimen dalam hal akurasi, presisi, recall, dan F1-Score?. Dari hasil penelitian ini, diperoleh bahwa teknik SMOTE memang bekerja dengan sangat baik pada dataset DUKCAPIL ini, ditunjukkan dengan peningkatan yang sangat signifikan, terutama pada model-model berbasis algoritma Ensemble Learning, dari yang semula berkinerja dibawah algoritma individual, hingga jauh meninggalkan kinerja algoritma individual seperti ditampilkan dalam Gambar 4.25 (Grafik Perbandingan Kinerja F1 Score SMOTE vs non SMOTE).



Gambar 4. 25(Grafik Perbandingan Kinerja F1 Score SMOTE vs non SMOTE)

4.5.5. Confusion Matrix Algoritma Extra Trees

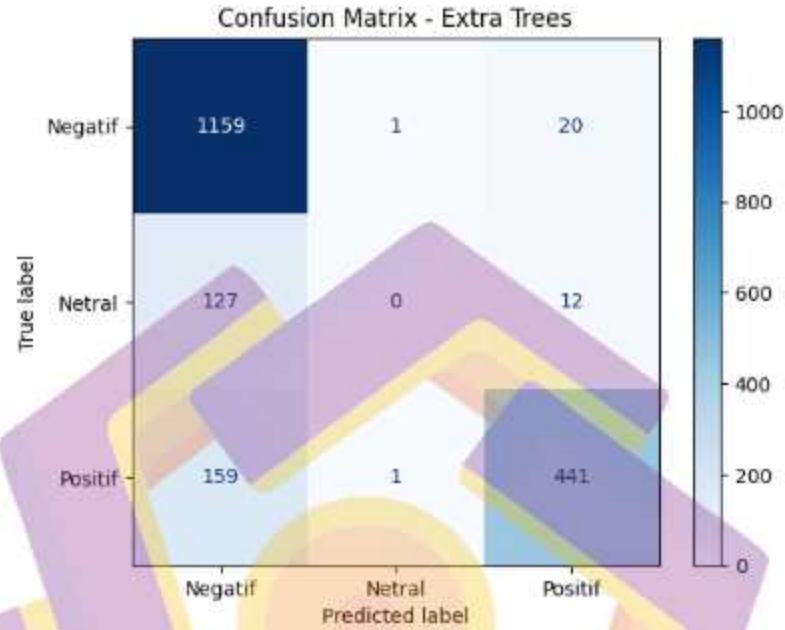
Confusion matrix digunakan untuk menunjukkan performa model terbaik dalam memprediksi kelas dengan benar. Dalam penelitian ini, confusion matrix menunjukkan bahwa model dengan menggunakan algoritma Extra Trees sebagai algoritma berkinerja terbaik, mampu memprediksi kelas dengan tingkat kesalahan minimal, seperti disajikan dalam gambar 4.26 (Confusion Matrix- Extra Trees).



Gambar 4. 26 (Confusion Matrix- Extra Trees)

Pada confusion matrix yang disajikan dalam gambar 4.26 (Confusion Matrix-Extra Trees), nampak bahwa model Extra Trees sebagai pemenang dalam kontes ini, dengan memprediksi benar 1164 pada label negatif, dibandingkan 8 dan 29 kesalahan. Memprediksi 1209 benar pada sentimen label netral dibandingkan 3 dan 3 kesalahan. Selanjutnya model masih menunjukkan kinerja yang sangat baik dengan memprediksi 1072 benar pada label positif dibandingkan 93 dan 6 kesalahan.

Dengan kinerja yang gemilang ketika menggunakan teknik SMOTE, ternyata Extratrees gagal memprediksi label netral ketika tanpa menggunakan SMOTE. Bisa dimaklumi, minimnya dataset berlabel positif, menyebabkan data training dan test yang minim, dibandingkan dengan label negatif dan positif. Kegagalan algoritma Extratrees pada dataset tanpa SMOTE, digambarkan oleh gambar 4.27 (Extra Trees tanpa SMOTE).



Gambar 4. 27 (Extra Trees tanpa SMOTE)

Nampak sekali pada gambar 4.27 bahwa dari 139 data netral yang ada, tidak satupun yang diprediksi benar oleh model Extra Trees, ketika data belum diseimbangkan dengan SMOTE. Ini menunjukkan pada dataset ini, SMOTE memegang peranan sangat penting dan sangat signifikan menjadi faktor penentu dari model yang dibangun dengan algoritma Extra Trees.

4.5.6. Word Cloud

Dalam konteks analisis sentimen ulasan aplikasi DUKCAPIL, word cloud digunakan untuk mengidentifikasi kata-kata unik yang sering muncul pada kategori sentimen negatif, netral, dan positif. Hal ini membantu memahami aspek apa saja yang menjadi fokus perhatian pengguna aplikasi, baik dalam bentuk kritik, masukan netral, maupun apresiasi.

4.5.6.1 Word Cloud untuk Sentimen Negatif

Gambar 4. 28 (Word Cloud untuk Sentimen Negatif)

Word cloud untuk sentimen negatif seperti yang disajikan pada gambar 4.28 (Word Cloud untuk Sentimen Negatif) menunjukkan kata-kata yang sering digunakan oleh pengguna yang memberikan ulasan kurang puas terhadap aplikasi. Beberapa kata kunci yang dominan adalah:

- **"Aplikasi"**, **"lagi"**, dan **"barcode"**, yang menunjukkan adanya frustrasi terkait fitur atau pengoperasian aplikasi.
- Kata seperti **"ribet"**, **"susah"**, dan **"tidak"** mencerminkan pengalaman pengguna yang sulit atau tidak sesuai harapan.
- Istilah seperti **"KTP"**, **"dukcapil"**, dan **"coba"** menunjukkan fokus masalah pada layanan terkait dokumen kependudukan.

4.6.6.2 Word Cloud untuk Sentimen Netral

Gambar 4. 29 (Word Cloud untuk Sentimen Netral)

Word cloud untuk sentimen netral menampilkan kata-kata yang mengandung nada netral atau masukan yang tidak secara langsung bersifat negatif maupun positif seperti yang dapat dilihat pada gambar 4.29 (Word Cloud untuk Sentimen Netral). Kata-kata dominan meliputi:

- **"Surat"**, **"masuk"**, dan **"ini"**, yang menunjukkan komentar seputar dokumen dan proses masuk aplikasi.
- Istilah seperti **"tolong"**, **"admin"**, dan **"kecamatan"** mengindikasikan adanya masukan dari pengguna terkait layanan administratif dan dukungan teknis.
- Kata-kata seperti **"lagi"** dan **"KTP"** tetap muncul, mencerminkan fokus pada fungsi utama aplikasi.

4.6.6.3. Word Cloud untuk Sentimen Positif

Gambar 4. 30 (Word Cloud untuk Sentimen Positif)

Word cloud untuk sentimen positif menunjukkan kata-kata yang sering digunakan oleh pengguna yang memberikan apresiasi terhadap aplikasi seperti disajikan dalam gambar 4.30 (Word Cloud untuk Sentimen Positif). Kata-kata dominan meliputi:

- **"Mantap"**, **"membantu"**, dan **"bagus"**, yang menunjukkan kepuasan pengguna terhadap manfaat aplikasi.
- Kata seperti **"terima kasih"**, **"keren"**, dan **"praktis"** mengindikasikan penghargaan terhadap kemudahan dan fungsi aplikasi.
- Istilah **"dukcapil"**, **"data"**, dan **"digital"** menunjukkan pengakuan positif terhadap transformasi digital dalam layanan kependudukan.

Word cloud ini memberikan wawasan penting tentang persepsi pengguna terhadap aplikasi DUKCAPIL.

1. **Sentimen negatif** menunjukkan aspek yang perlu diperbaiki, seperti fitur teknis dan kemudahan akses.
2. **Sentimen netral** memberikan masukan spesifik yang dapat dijadikan panduan untuk peningkatan layanan.
3. **Sentimen positif** menyoroti keberhasilan aplikasi dalam memberikan manfaat kepada pengguna.

Data ini dapat digunakan untuk merancang strategi perbaikan dan pengembangan aplikasi lebih lanjut berdasarkan kebutuhan dan harapan pengguna.

4.6. Hasil Penelitian dan diskusi

1. Extra Trees adalah algoritma terbaik dengan akurasi tertinggi, diikuti oleh Random Forest dan XGBoost.
2. Algoritma seperti SVM dan KNN menunjukkan performa yang kurang baik karena ketergantungan mereka pada distribusi data yang seimbang dan pola yang jelas.
3. Teknik ensemble secara konsisten mengungguli algoritma individu, menunjukkan bahwa penggabungan model memberikan stabilitas dan akurasi yang lebih baik.

4.6.1. Hasil Penelitian

Penelitian ini membandingkan kinerja algoritma ensemble learning dengan algoritma non-ensemble learning. Hasil evaluasi kinerja masing-masing algoritma disajikan pada tabel disajikan dalam tabel 4.2:

Tabel 4.2 (evaluasi kinerja masing-masing algoritma)

1. Algoritma Ensemble Learning

Model	Akurasi	Presisi	Recall	F1-Score
Random Forest	93,31%	93,55%	93,31%	93,31%
Extra Trees	95,85%	95,93%	95,85%	95,85%
AdaBoost	69,14%	70,65%	69,14%	69,39%
Gradient Boosting	85,61%	86,12%	85,61%	85,60%
XGBoost	90,91%	91,39%	90,91%	90,96%

2. Algoritma Non-Ensemble Learning

Model	Akurasi	Presisi	Recall	F1-Score
Logistic Regression	83,88%	84,23%	83,88%	83,97%
Decision Tree	83,21%	83,20%	83,21%	83,19%
SVM (Support Vector Machine)	43,99%	44,02%	43,99%	41,67%
KNN (K-Nearest Neighbors)	68,69%	70,65%	68,69%	64,76%
Gaussian Naive Bayes	77,75%	82,05%	77,75%	77,19%

Algoritma Ensemble Learning

Model ensemble learning menunjukkan performa yang lebih baik secara keseluruhan. Extra Trees menempati posisi terbaik dengan akurasi 95,85%, diikuti oleh Random Forest (93,31%) dan XGBoost (90,91%). Model ensemble ini unggul karena mampu menangkap pola data dengan lebih efektif menggunakan metode kombinasi pohon keputusan.

Algoritma Non-Ensemble Learning

Di antara algoritma non-ensemble learning, Logistic Regression dan Decision Tree memiliki performa yang cukup baik, dengan akurasi masing-masing 83,88% dan 83,21%. Namun, algoritma seperti SVM (43,99%) dan KNN (68,69%) menunjukkan performa yang jauh lebih rendah dibandingkan model ensemble, yang mengindikasikan keterbatasan algoritma ini dalam menangani data sentimen.

Hasil penelitian ini menunjukkan bahwa algoritma ensemble learning, khususnya Extra Trees dan Random Forest, memiliki performa yang jauh lebih baik dibandingkan algoritma non-ensemble learning. Temuan ini mendukung

penggunaan metode ensemble learning dalam penelitian analisis sentimen pada dataset ini.

4.6.2. Perbandingan dengan Penelitian Sebelumnya

Penelitian ini dibandingkan dengan studi oleh Shkurti et al. (2022) berjudul *"Performance Comparison of Machine Learning Algorithms for Albanian News Articles"*. Studi tersebut menggunakan berbagai algoritma klasifikasi—termasuk Multinomial Naïve Bayes, Logistic Regression, SVM, k-NN, Perceptron, Passive Aggressive, dan Random Forest—untuk mengklasifikasikan berita dalam 8 kategori. Mereka melibatkan dataset besar (hingga 80.000 artikel) dan menilai performa berdasarkan akurasi serta waktu pelatihan dan pengujian.

Perbedaan Kuantitatif:

Dalam studi Shkurti et al., Passive Aggressive menunjukkan akurasi tertinggi, namun angka akurasi tidak melebihi 90%. Sementara itu, Random Forest adalah model dengan akurasi terendah di bawah 80%. Sebaliknya, dalam penelitian ini, model Extra Trees dengan SMOTE mencapai akurasi 95,85%, precision 95,93%, recall 95,85%, dan F1-score 95,85%, menunjukkan keunggulan kuantitatif signifikan dalam hal akurasi dan stabilitas prediksi.

Perbedaan Kualitatif Algoritma:

1. Jenis Data: Studi Shkurti menggunakan teks berita yang bersifat objektif, formal, dan panjang, cocok untuk algoritma linear seperti Passive Aggressive. Penelitian ini menggunakan data ulasan pengguna aplikasi yang lebih subjektif, pendek, dan tidak terstruktur, yang lebih kompleks dan menantang untuk diprediksi.
2. Tujuan Klasifikasi: Studi Shkurti fokus pada multi-class classification (8 kelas berita), sedangkan penelitian ini fokus pada binary classification

(positif/negatif). Klasifikasi biner umumnya menghasilkan metrik yang lebih tinggi karena kompleksitasnya lebih rendah.

3. Pendekatan Model: Studi Shkurti hanya menguji model-model individual klasik tanpa penguatan, sementara penelitian ini menerapkan ensemble learning (Extra Trees, Random Forest, XGBoost) yang diketahui memiliki keunggulan dalam mengurangi overfitting dan meningkatkan stabilitas prediksi.
4. Teknik Penyeimbangan Data: Penelitian ini juga menggunakan SMOTE untuk menangani ketidakseimbangan data—faktor penting dalam analisis sentimen. Studi Shkurti tidak melibatkan teknik balancing, sehingga model seperti Random Forest menjadi kurang akurat.
5. Evaluasi Kinerja: Penelitian ini menggunakan metrik yang lebih komprehensif—precision, recall, dan F1-score—selain akurasi. Sedangkan Shkurti et al. hanya menggunakan akurasi, yang bisa menyesatkan dalam kasus data tidak seimbang.

Kesimpulan Komparatif:
Hasil ini menunjukkan bahwa kombinasi ensemble learning dan SMOTE lebih efektif dalam menganalisis sentimen berbasis teks pendek dan tidak terstruktur seperti ulasan aplikasi. Di sisi lain, model seperti Passive Aggressive lebih cocok untuk klasifikasi teks panjang dan formal seperti berita. Oleh karena itu, pemilihan algoritma harus disesuaikan dengan karakteristik data dan tujuan analisis, bukan semata-mata berdasarkan performa umum suatu model.

Selain penelitian di atas, berbagai penelitian telah membandingkan kinerja metode machine learning, menunjukkan keunggulan metode ensemble:

a. Studi oleh An Ensemble Learning Approach for Anomaly Detection (2023) menunjukkan bahwa pendekatan ensemble, seperti Random Forest dan Gradient Boosting, menghasilkan performa yang lebih stabil dalam mendeteksi outlier, dengan akurasi hingga 92%.
b. Penelitian lain tentang Meta-learning dan Stacked Regression (2023) menunjukkan bahwa metode stacking ensemble dapat meningkatkan presisi prediksi hingga 5-10% dibandingkan metode individu dalam aplikasi prediksi ekonomi.

Penggunaan SMOTE untuk Ketidakseimbangan Data

Penelitian yang fokus pada ketidakseimbangan data menunjukkan bahwa penggunaan teknik seperti SMOTE dapat secara signifikan meningkatkan kinerja model:

a. Studi tentang Unbalanced Credit Card Fraud Detection (2023) mencatat peningkatan akurasi hingga 99% setelah menerapkan SMOTE untuk meningkatkan distribusi kelas minoritas.
b. Penelitian oleh Improved Hybrid Resampling (2022) menunjukkan bahwa kombinasi SMOTE dengan metode hibrid mampu meningkatkan F1-Score hingga 3% dibandingkan baseline model.

Penerapan Analisis Sentimen di Domain Lain

Penelitian terkait analisis sentimen di berbagai domain juga menunjukkan hasil yang relevan:

a. Yelp Reviews Analysis (2021) memanfaatkan ulasan pelanggan untuk mengevaluasi pengalaman pengguna selama pandemi COVID-19. Dengan pendekatan analisis kata kunci dan multiple correspondence analysis, penelitian ini memberikan wawasan mendalam tentang pengaruh protokol keamanan terhadap kepuasan pelanggan.

b. Penelitian tentang Sentiment Classification in Slovak (2023) mencatat akurasi hingga 90% menggunakan model ensemble, yang membuktikan efektivitas pendekatan ini di berbagai bahasa dan konteks.

Penelitian ini memberikan kontribusi yang signifikan dalam membandingkan algoritma ensemble dan non-ensemble untuk analisis sentimen pada ulasan aplikasi DUKCAPIL. Dengan pendekatan yang melibatkan SMOTE, hyperparameter tuning, dan algoritma ensemble learning, penelitian ini menunjukkan bahwa algoritma seperti Extra Trees dan Random Forest mampu mencapai akurasi lebih dari 95%, yang lebih unggul dibandingkan sebagian besar penelitian sebelumnya. Hasil ini mendukung klaim bahwa metode ensemble memiliki keunggulan signifikan dalam menangkap pola data yang kompleks.

4.6.3. Signifikansi Dampak dan Kontribusi Penelitian

Penelitian ini membandingkan 10 algoritma klasifikasi dalam konteks analisis sentimen terhadap ulasan aplikasi DUKCAPIL. Evaluasi dilakukan dengan mempertimbangkan berbagai metrik kinerja seperti akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa algoritma ensemble—khususnya Extra Trees dengan SMOTE—memiliki performa paling unggul dibandingkan model-model individual.

Eksplorasi mendalam terhadap hasil ini mengindikasikan adanya signifikansi dampak metodologis dan praktis dalam pemilihan algoritma berbasis karakteristik data. Penelitian ini menunjukkan bahwa:

1. Model ensemble dapat menangani kompleksitas dan ketidakseimbangan data lebih baik daripada model individual.
2. Perbedaan performa antar algoritma tidak hanya bersifat teknis, tetapi juga kontekstual terhadap jenis teks (ulasan pendek dan subjektif).
3. Penggunaan SMOTE terbukti meningkatkan performa model secara konsisten.

Dengan demikian, penelitian ini berkontribusi terhadap pengembangan pengetahuan di bidang Ilmu Komputer, khususnya dalam:

1. Memberikan panduan empiris dalam memilih algoritma berdasarkan skenario data nyata.
2. Memperkuat argumen pentingnya preprocessing dan balancing dalam pipeline machine learning.
3. Menunjukkan bahwa kombinasi metode modern seperti ensemble learning dan teknik balancing dapat meningkatkan kualitas hasil analisis sentimen dalam aplikasi dunia nyata.

BAB V

KESIMPULAN DAN SARAN

1.1.Kesimpulan

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan sebagai berikut:

1. **Perbandingan Kinerja Algoritma Individual dengan Model Ensemble**
   Berdasarkan analisis kinerja berbagai algoritma dalam tugas analisis sentimen ulasan aplikasi DUKCAPIL, model ensemble seperti Extra Trees, Random Forest, dan XGBoost menunjukkan hasil yang lebih akurat dan andal dibandingkan model individual seperti Logistic Regression, SVM, dan KNN. Model ensemble menghasilkan performa yang lebih stabil dalam hal metrik evaluasi seperti akurasi, presisi, recall, dan F1-Score. Hal ini menunjukkan bahwa model ensemble lebih mampu menangani kompleksitas data ulasan yang tidak terstruktur dibandingkan dengan model individual.
2. **Pengaruh Penggunaan Teknik SMOTE terhadap Kinerja Model**
   Penggunaan teknik SMOTE (Synthetic Minority Over-sampling Technique) secara signifikan meningkatkan kinerja model analisis sentimen dalam hal akurasi, presisi, recall, dan F1-Score. Teknik SMOTE mampu mengatasi masalah ketidakseimbangan data dengan membuat sampel

sintetik dari kelas minoritas, sehingga model dapat belajar dari data yang lebih seimbang. Hasil evaluasi menunjukkan bahwa model yang dilatih dengan data yang telah diseimbangkan menggunakan SMOTE memiliki performa yang lebih baik dibandingkan dengan model yang dilatih dengan data yang tidak seimbang. Model Extra Trees dengan SMOTE memberikan hasil terbaik dengan akurasi mencapai 95,85%, presisi 95,93%, recall 95,85%, dan F1-Score 95,85%.

1.2.Saran

Berdasarkan hasil penelitian yang telah dilakukan, berikut beberapa saran yang dapat diberikan untuk penelitian selanjutnya:

1. **Penerapan Model Hybrid untuk Peningkatan Kinerja** Penelitian di masa mendatang disarankan untuk menggabungkan model ensemble dengan model deep learning, seperti Long Short-Term Memory (LSTM) atau Bidirectional LSTM, untuk menangani data ulasan yang bersifat sekuensial dan memiliki konteks temporal. Kombinasi ini diharapkan dapat meningkatkan akurasi prediksi sentimen.
2. **Eksplorasi Teknik Preprocessing yang Lebih Lanjut** Disarankan untuk mengeksplorasi teknik preprocessing yang lebih canggih, seperti penggunaan embedding berbasis konteks (misalnya Word2Vec, GloVe, atau BERT), yang dapat meningkatkan kualitas representasi data teks. Teknik ini dapat membantu model untuk menangkap hubungan semantik dalam data ulasan dengan lebih baik.

3. **Penggunaan Dataset yang Lebih Beragam** Penelitian selanjutnya dapat menggunakan dataset ulasan dari berbagai aplikasi layanan publik lainnya untuk menguji generalisasi model yang dikembangkan. Hal ini akan memberikan wawasan yang lebih luas tentang kinerja model dalam konteks yang berbeda.
4. **Penerapan Model dalam Lingkungan Dunia Nyata** Disarankan untuk mengimplementasikan model yang telah dikembangkan dalam lingkungan dunia nyata, seperti pada aplikasi DUKCAPIL yang sebenarnya. Implementasi ini dapat membantu menguji performa model secara langsung dalam menangani data ulasan yang diterima secara real-time.
5. **Pengujian dengan Metrik Evaluasi Tambahan** Selain menggunakan metrik akurasi, presisi, recall, dan F1-Score, penelitian selanjutnya dapat mempertimbangkan penggunaan metrik evaluasi tambahan seperti ROC-AUC, Log Loss, dan Matthews Correlation Coefficient (MCC) untuk memberikan gambaran yang lebih komprehensif tentang kinerja model.

DAFTAR PUSTAKA

Abedin, M. Z., Hajek, P., Sharif, T., Satu, M. S., & Khan, M. I. (2023). Modelling bank customer behaviour using feature engineering and classification techniques. *Research in International Business and Finance*, *65*. https://doi.org/10.1016/j.ribaf.2023.101913

Achitouv, I., Gorduza, D., & Jacquier, A. (2024). Natural Language Processing for Financial Regulation. In *Journal of Artificial Intelligence and Autonomous Intelligence: Research* (Vol. 1, Issue 1). https://jaiai.org/

Adiningtyas, H., & Auliani, A. S. (2024). Sentiment analysis for mobile banking service quality measurement. *Procedia Computer Science*, *234*, 40–50. https://doi.org/10.1016/j.procs.2024.02.150

Akber, Md. A., Ferdousi, T., Ahmed, R., Asfara, R., Rab, R., & Zakia, U. (2024). Personality and emotion—A comprehensive analysis using contextual text embeddings. *Natural Language Processing Journal*, *9*, 100105. https://doi.org/10.1016/j.nlp.2024.100105

Alsayat, A. (2022). Improving Sentiment Analysis for Social Media Applications Using an Ensemble Deep Learning Language Model. *Arabian Journal for Science and Engineering*, *47*(2), 2499–2511. https://doi.org/10.1007/s13369-021-06227-w

Aniceto, M. C., Barboza, F., & Kimura, H. (2020). Machine learning predictivity applied to consumer creditworthiness. *Future Business Journal*, *6*(1). https://doi.org/10.1186/s43093-020-00041-w

Arafa, A., El-Fishawy, N., Badawy, M., & Radad, M. (2022). RN-SMOTE: Reduced Noise SMOTE based on DBSCAN for enhancing imbalanced data classification. *Journal of King Saud University - Computer and Information Sciences*, *34*(8), 5059–5074. https://doi.org/10.1016/j.jksuci.2022.06.005

Barongo, R. I., & Tibangayuka Mbelwa, J. (2024). Using machine learning for detecting liquidity risk in banks. *Machine Learning with Applications*, *15*, 100511. https://doi.org/10.24433/CO.5061127.v1

Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., Modi, K., & Ghayvat, H. (2021). Cnn variants for computer vision: History, architecture, application, challenges and future scope. In *Electronics (Switzerland)* (Vol. 10, Issue 20). MDPI. https://doi.org/10.3390/electronics10202470

Bitetto, A., Cerchiello, P., Filomeni, S., Tanda, A., & Tarantino, B. (2023). Machine learning and credit risk: Empirical evidence from small- and mid-sized businesses. *Socio-Economic Planning Sciences*, *90*. https://doi.org/10.1016/j.seps.2023.101746

Bonechi, S., Palma, G., Caronna, M., Ugolini, M., Massaro, A., & Rizzo, A. (2024). Enhancing Customer Support in Banking: Leveraging AI for Efficient Ticket Classification. *Procedia Computer Science*, *246*, 128–137. https://doi.org/10.1016/j.procs.2024.09.235

Bouteska, A., Abedin, M. Z., Hajek, P., & Yuan, K. (2024). Cryptocurrency price forecasting – A comparative analysis of ensemble learning and deep learning methods. *International Review of Financial Analysis*, *92*. https://doi.org/10.1016/j.irfa.2023.103055

Breiman, L. (2001a). *Random Forests* (Vol. 45).

Breiman, L. (2001b). *Random Forests* (Vol. 45).

Chachoui, Y., Azizi, N., Hotte, R., & Bensebaa, T. (2024). Enhancing algorithmic assessment in education: Equi-fused-data-based SMOTE for balanced learning. *Computers and Education: Artificial Intelligence*, *6*. https://doi.org/10.1016/j.caeai.2024.100222

Citterio, A., & King, T. (2023). The role of Environmental, Social, and Governance (ESG) in predicting bank financial distress. *Finance Research Letters*, *51*, 103411. https://doi.org/10.1016/j.frl.2022.103411

Dantas, R. M., Firdaus, R., Jaleel, F., Neves Mata, P., Mata, M. N., & Li, G. (2022). Systemic Acquired Critique of Credit Card Deception Exposure through Machine Learning. *Journal of Open Innovation: Technology, Market, and Complexity*, *8*(4). https://doi.org/10.3390/joitmc8040192

Diekson, Z. A., Prakoso, M. R. B., Putra, M. S. Q., Syaputra, M. S. A. F., Achmad, S., & Sutoyo, R. (2022). Sentiment analysis for customer review: Case study of Traveloka. *Procedia Computer Science*, *216*, 682–690. https://doi.org/10.1016/j.procs.2022.12.184

Domashova, J., & Kripak, E. (2022). Development of a generalized algorithm for identifying atypical bank transactions using machine learning methods. *Procedia Computer Science*, *213*(C), 101–109. https://doi.org/10.1016/j.procs.2022.11.044

Fairhall, S. L. (2024). Sentence-level embeddings reveal dissociable word- and sentence-level cortical representation across coarse- and fine-grained levels of meaning. *Brain and Language*, *250*. https://doi.org/10.1016/j.bandl.2024.105389

Gadgil, K., Gill, S. S., & Abdelmoniem, A. M. (2023). A meta-learning based stacked regression approach for customer lifetime value prediction. *Journal of Economy and Technology*, *1*, 197–207. https://doi.org/10.1016/j.ject.2023.09.001

Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2021). *Ensemble deep learning: A review*. https://doi.org/10.1016/j.engappai.2022.105151

Goudarzi, M., & Bazzana, F. (2023). Identification of high-frequency trading: A machine learning approach. *Research in International Business and Finance*, *66*. https://doi.org/10.1016/j.ribaf.2023.102078

Guerra, P., Castelli, M., & Côrte-Real, N. (2022). Machine learning for liquidity risk modelling: A supervisory perspective. *Economic Analysis and Policy*, *74*, 175–187. https://doi.org/10.1016/j.eap.2022.02.001

Gupta, P., Varshney, A., Khan, M. R., Ahmed, R., Shuaib, M., & Alam, S. (2022). Unbalanced Credit Card Fraud Detection Data: A Machine Learning-Oriented

Comparative Study of Balancing Techniques. *Procedia Computer Science*, *218*, 2575–2584. https://doi.org/10.1016/j.procs.2023.01.231

Hardiman, J. P. W., Thio, D. C., Zakiyyah, A. Y., & Meiliana. (2024). AI-powered dialogues and quests generation in role-playing games using Google's Gemini and Sentence BERT framework. *Procedia Computer Science*, *245*(C), 1111–1119. https://doi.org/10.1016/j.procs.2024.10.340

Hasni, M. J. S., Rehman, M. A., Pontes, N., & Yaqub, M. Z. (2025). Health Star Rating Labels: A systematic review and future research agenda. In *Food Quality and Preference* (Vol. 122). Elsevier Ltd. https://doi.org/10.1016/j.foodqual.2024.105310

Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2017). *Squeeze-and-Excitation Networks*. http://arxiv.org/abs/1709.01507

Huang, H., Zavareh, A. A., & Mustafa, M. B. (2023). Sentiment Analysis in E-Commerce Platforms: A Review of Current Techniques and Future Directions. In *IEEE Access* (Vol. 11, pp. 90367–90382). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2023.3307308

Huang, J., Wen, H., Hu, J., Liu, B., Zhou, X., & Liao, M. (2024). Deciphering decision-making mechanisms for the susceptibility of different slope geohazards: A case study on a SMOTE-RF-SHAP hybrid model. *Journal of Rock Mechanics and Geotechnical Engineering*. https://doi.org/10.1016/j.jrmge.2024.03.008

Huang, S., Wang, Y., Wong, E. Y. C., & Yu, L. (2024). Ensemble learning with soft-prompted pretrained language models for fact checking. *Natural Language Processing Journal*, *7*, 100067. https://doi.org/10.1016/j.nlp.2024.100067

Imakura, A., Kihira, M., Okada, Y., & Sakurai, T. (2023). Another use of SMOTE for interpretable data collaboration analysis. *Expert Systems with Applications*, *228*. https://doi.org/10.1016/j.eswa.2023.120385

Iqbal, J., Saeed, A., & Khan, R. A. (2023). The relative importance of textual indexes in predicting the future performance of banks: A connection weight approach. *Borsa Istanbul Review*, *23*(1), 240–253. https://doi.org/10.1016/j.bir.2022.10.004

Islam, M. A., Uddin, M. A., Aryal, S., & Stea, G. (2023). An ensemble learning approach for anomaly detection in credit card data with imbalanced and overlapped classes. *Journal of Information Security and Applications*, *78*. https://doi.org/10.1016/j.jisa.2023.103618

Jafari, S., & Byun, Y. C. (2024). Efficient state of charge estimation in electric vehicles batteries based on the extra tree regressor: A data-driven approach. *Heliyon*, *10*(4). https://doi.org/10.1016/j.heliyon.2024.e25949

Jim, J. R., Talukder, M. A. R., Malakar, P., Kabir, M. M., Nur, K., & Mridha, M. F. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, *6*, 100059. https://doi.org/10.1016/j.nlp.2024.100059

Kalchev, E., Georgiev, R., & Ivanova, D. (2024). A novel 5-stage visual rating scale for global arterial spin labeling perfusion assessment in the brain: Simplifying evaluation for clinical implementation. *Cerebral Circulation - Cognition and Behavior*, *6*. https://doi.org/10.1016/j.cccb.2024.100200

Karbasi, M., Ali, M., Bateni, S. M., Jun, C., Jamei, M., & Yaseen, Z. M. (2024). Boruta extra tree-bidirectional long short-term memory model development for Pan evaporation forecasting: Investigation of arid climate condition. *Alexandria Engineering Journal*, *86*, 425–442. https://doi.org/10.1016/j.aej.2023.11.061

Kelebercová, L., & Zozuk, N. C. (2023). Sentiment classification of annual reports in Slovak language using symbolic, subsymbolic and statistic methods. *Procedia Computer Science*, *225*, 3508–3516. https://doi.org/10.1016/j.procs.2023.10.346

Kenrick, Ivancio, S. C., Isabellini, E., Adhi, C. G. S., & Prasetyo, A. (2024). Multivariate stock market forecasting using lstm on Indonesian banking stock

market. *Procedia Computer Science*, *245*(C), 1047–1056. https://doi.org/10.1016/j.procs.2024.10.333

Khanam, R., Hussain, M., Hill, R., & Allen, P. (2024). A Comprehensive Review of Convolutional Neural Networks for Defect Detection in Industrial Applications. *IEEE Access*, *12*, 94250–94295. https://doi.org/10.1109/ACCESS.2024.3425166

Khasanah, H. M., Aminuddin, A., Abdulloh, F. F., Rahardi, M., Hairani, H., & Asaddulloh, B. P. (2024). Optimizing mushroom classification through machine learning and hyperparameter tuning. *Engineering and Applied Science Research*, *51*(5), 651–660. https://doi.org/10.14456/easr.2024.61

Kostromitina, M., Keller, D., Cavusoglu, M., & Beloin, K. (2021). "His lack of a mask ruined everything." Restaurant customer satisfaction during the COVID-19 outbreak: An analysis of Yelp review texts and star-ratings. *International Journal of Hospitality Management*, *98*. https://doi.org/10.1016/j.ijhm.2021.103048

Kou, G., Chen, H., & Hefni, M. A. (2022). Improved hybrid resampling and ensemble model for imbalance learning and credit evaluation. *Journal of Management Science and Engineering*, *7*(4), 511–529. https://doi.org/10.1016/j.jmse.2022.06.002

Koyyada, S. P., & Singh, T. P. (2022). A multi stage approach to handle class imbalance:An ensemble method. *Procedia Computer Science*, *218*, 2666–2674. https://doi.org/10.1016/j.procs.2023.01.239

KP, V., AB, R., HL, G., Ravi, V., & Krichen, M. (2024). A tweet sentiment classification approach using an ensemble classifier. *International Journal of Cognitive Computing in Engineering*, *5*, 170–177. https://doi.org/10.1016/j.ijcce.2024.04.001

Kristiyanti, D. A., Pramudya, W. B. N., & Sanjaya, S. A. (2024). How can we predict transportation stock prices using artificial intelligence? Findings from experiments with Long Short-Term Memory based algorithms. *International Journal of*

Information Management Data Insights, *4*(2). https://doi.org/10.1016/j.jjimei.2024.100293

Kristóf, T., & Virág, M. (2022). EU-27 bank failure prediction with C5.0 decision trees and deep learning neural networks. *Research in International Business and Finance*, *61*. https://doi.org/10.1016/j.ribaf.2022.101644

Kummer, A., Ruppert, T., Medvegy, T., & Abonyi, J. (2022). Machine learning-based software sensors for machine state monitoring - The role of SMOTE-based data augmentation. *Results in Engineering*, *16*. https://doi.org/10.1016/j.rineng.2022.100778

Kurniadi, F. I., Paramita, N. L. P. S. P., Sihotang, E. F. A., Anggreainy, M. S., & Zhang, R. (2024). BERT and RoBERTa Models for Enhanced Detection of Depression in Social Media Text. *Procedia Computer Science*, *245*(C), 202–209. https://doi.org/10.1016/j.procs.2024.10.244

Kusuma, I. Y., Visnyovszki, Á., & Bahar, M. A. (2024). Mapping the Mpox discourse: A network and sentiment analysis. *Exploratory Research in Clinical and Social Pharmacy*, *16*. https://doi.org/10.1016/j.rcsop.2024.100521

Liu, Z., Zhang, Z., Yang, H., Wang, G., & Xu, Z. (2023). An innovative model fusion algorithm to improve the recall rate of peer-to-peer lending default customers. *Intelligent Systems with Applications*, *20*. https://doi.org/10.1016/j.iswa.2023.200272

Locarso, G. K. (2022). ANALISIS SENTIMEN REVIEW APLIKASI PEDULILINDUNGI PADA GOOGLE PLAY STORE MENGGUNAKAN NBC. *Jurnal Teknik Informatika Kaputama (JTIK)*, *6*(2).

Malik, H., & Shakshuki, E. M. (2016). Mining Collective Opinions for Comparison of Mobile Apps. *Procedia Computer Science*, *94*, 168–175. https://doi.org/10.1016/j.procs.2016.08.026

Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. In *Journal of King Saud University -*

Computer and Information Sciences (Vol. 36, Issue 4). King Saud bin Abdulaziz University. https://doi.org/10.1016/j.jksuci.2024.102048

Marutho, D., Muljono, Rustad, S., & Purwanto. (2024). Optimizing aspect-based sentiment analysis using sentence embedding transformer, bayesian search clustering, and sparse attention mechanism. *Journal of Open Innovation: Technology, Market, and Complexity*, *10*(1). https://doi.org/10.1016/j.joitmc.2024.100211

Matrane, Y., Benabbou, F., & Sael, N. (2023). A systematic literature review of Arabic dialect sentiment analysis. In *Journal of King Saud University - Computer and Information Sciences* (Vol. 35, Issue 6). King Saud bin Abdulaziz University. https://doi.org/10.1016/j.jksuci.2023.101570

Mensh, B., & Kording, K. (2017). Ten simple rules for structuring papers. In *PLoS Computational Biology* (Vol. 13, Issue 9). Public Library of Science. https://doi.org/10.1371/journal.pcbi.1005619

Mian, Z., Deng, X., Dong, X., Tian, Y., Cao, T., Chen, K., & Jaber, T. Al. (2024). A literature review of fault diagnosis based on ensemble learning. *Engineering Applications of Artificial Intelligence*, *127*. https://doi.org/10.1016/j.engappai.2023.107357

Mienye, I. D., & Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. In *IEEE Access* (Vol. 10, pp. 99129–99149). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ACCESS.2022.3207287

Mohammed, A., & Kora, R. (2022). An effective ensemble deep learning framework for text classification. *Journal of King Saud University - Computer and Information Sciences*, *34*(10), 8825–8837. https://doi.org/10.1016/j.jksuci.2021.11.001

Nabiilah, G. Z., Prasetyo, S. Y., Izdihar, Z. N., & Girsang, A. S. (2022). BERT base model for toxic comment analysis on Indonesian social media. *Procedia Computer Science*, *216*, 714–721. https://doi.org/10.1016/j.procs.2022.12.188

Najem, R., Bahnasse, A., & Talea, M. (2024). Toward an Enhanced Stock Market Forecasting with Machine Learning and Deep Learning Models. *Procedia Computer Science*, *241*, 97–103. https://doi.org/10.1016/j.procs.2024.08.015

Narayanan, & Jayashree. (2024). Implementation of Efficient Machine Learning Techniques for Prediction of Cardiac Disease using SMOTE. *Procedia Computer Science*, *233*, 558–569. https://doi.org/10.1016/j.procs.2024.03.245

Okoro, E. E., Obomanu, T., Sanni, S. E., Olatunji, D. I., & Igbinedion, P. (2022). Application of artificial intelligence in predicting the dynamics of bottom hole pressure for under-balanced drilling: Extra tree compared with feed forward neural network model. *Petroleum*, *8*(2), 227–236. https://doi.org/10.1016/j.petlm.2021.03.001

Onan, A. (2023). Hierarchical graph-based text classification framework with contextual node embedding and BERT-based dynamic fusion. *Journal of King Saud University - Computer and Information Sciences*, *35*(7). https://doi.org/10.1016/j.jksuci.2023.101610

Pattnaik, D., Ray, S., & Raman, R. (2024). Applications of artificial intelligence and machine learning in the financial services industry: A bibliometric review. *Heliyon*, *10*(1). https://doi.org/10.1016/j.heliyon.2023.e23492

Perezhohin, Y., Peres, F., & Castelli, M. (2024). Combining computational linguistics with sentence embedding to create a zero-shot NLIDB. *Array*, *24*. https://doi.org/10.1016/j.array.2024.100368

Polireddi, N. S. A. (2024). An effective role of artificial intelligence and machine learning in banking sector. *Measurement: Sensors*, *33*, 101135. https://doi.org/10.1016/j.measen.2024.101135

Poornima, S., & Mahalakshmi, R. (2024). Automated malware detection using machine learning and deep learning approaches for android applications. *Measurement: Sensors*, *32*. https://doi.org/10.1016/j.measen.2023.100955

Pramana, R., Jonathan, M., Yani, H. S., & Sutoyo, R. (2024). A Comparison of BiLSTM, BERT, and Ensemble Method for Emotion Recognition on Indonesian Product Reviews. *Procedia Computer Science*, *245*(C), 399–408. https://doi.org/10.1016/j.procs.2024.10.266

Priyambada, S. A., Usagawa, T., & ER, M. (2023). Two-layer ensemble prediction of students' performance using learning behavior and domain knowledge. *Computers and Education: Artificial Intelligence*, *5*. https://doi.org/10.1016/j.caeai.2023.100149

Riccosan, & Saputra, K. E. (2023). Multilabel multiclass sentiment and emotion dataset from indonesian mobile application review. *Data in Brief*, *50*. https://doi.org/10.1016/j.dib.2023.109576

Schmitt, M. (2023). Automated machine learning: AI-driven decision making in business analytics. *Intelligent Systems with Applications*, *18*. https://doi.org/10.1016/j.iswa.2023.200188

Shar, L. K., Binh Duong, T. N., Yeo, Y. C., & Fan, J. (2024). Empirical Evaluation of Hyper-parameter Optimization Techniques for Deep Learning-based Malware Detectors. *Procedia Computer Science*, *246*, 2090–2099. https://doi.org/10.1016/j.procs.2024.09.640

Sharma, D., Kumar, R., & Jain, A. (2022). Breast cancer prediction based on neural networks and extra tree classifier using feature ensemble learning. *Measurement: Sensors*, *24*. https://doi.org/10.1016/j.measen.2022.100560

Shi, H. (2024). A short video sentiment analysis model based on multimodal feature fusion. *Systems and Soft Computing*, *6*. https://doi.org/10.1016/j.sasc.2024.200148

Shkurti, L., Kabashi, F., Sofiu, V., & Susuri, A. (2022). Performance Comparison of Machine Learning Algorithms for Albanian News articles. *IFAC-PapersOnLine*, *55*(39), 292–295. https://doi.org/10.1016/j.ifacol.2022.12.037

SUDHAMATHI, T., & Perumal, K. (2024). ENSEMBLE REGRESSION BASED EXTRA TREE REGRESSORFOR HYBRID CROP YIELD PREDICTION SYSTEM. *Measurement: Sensors*, 101277. https://doi.org/10.1016/j.measen.2024.101277

Sun, Z., Harit, A., Cristea, A. I., Wang, J., & Lio, P. (2023). MONEY: Ensemble learning for stock price movement prediction via a convolutional network with adversarial hypergraph model. *AI Open*, *4*, 165–174. https://doi.org/10.1016/j.aiopen.2023.10.002

Syed, A. A., Gaol, F. L., Boediman, A., & Budiharto, W. (2024). Airline reviews processing: Abstractive summarization and rating-based sentiment classification using deep transfer learning. *International Journal of Information Management Data Insights*, *4*(2), 100238. https://doi.org/10.1016/j.jjimei.2024.100238

Tanoto, K., Gunawan, A. A. S., Suhartono, D., Mursitama, T. N., Rahayu, A., & Ariff, M. I. M. (2024). Investigation of challenges in aspect-based sentiment analysis enhanced using softmax function on twitter during the 2024 Indonesian presidential election. *Procedia Computer Science*, *245*(C), 989–997. https://doi.org/10.1016/j.procs.2024.10.327

Tiwari, D., Nagpal, B., Bhati, B. S., Mishra, A., & Kumar, M. (2023). A systematic review of social network sentiment analysis with comparative study of ensemble-based techniques. *Artificial Intelligence Review*, *56*(11), 13407–13461. https://doi.org/10.1007/s10462-023-10472-w

Uddin, N., Uddin Ahamed, M. K., Uddin, M. A., Islam, M. M., Talukder, M. A., & Aryal, S. (2023). An ensemble machine learning based bank loan approval predictions system with a smart application. *International Journal of Cognitive Computing in Engineering*, *4*, 327–339. https://doi.org/10.1016/j.ijcce.2023.09.001

Umer, M., Sadiq, S., karamti, H., Abdulmajid Eshmawi, A., Nappi, M., Usman Sana, M., & Ashraf, I. (2022). ETCNN: Extra Tree and Convolutional Neural Network-based Ensemble Model for COVID-19 Tweets Sentiment Classification: ETCNN: COVID-19 Tweets Sentiment Classification. *Pattern Recognition Letters*, *164*, 224–231. https://doi.org/10.1016/j.patrec.2022.11.012

Villaizán-Vallelado, M., Salvatori, M., Carro, B., & Sanchez-Esguevillas, A. J. (2024). Graph Neural Network contextual embedding for Deep Learning on tabular data. *Neural Networks*, *173*. https://doi.org/10.1016/j.neunet.2024.106180

Wahyudi, R., Kusumawardhana, G., Purwokerto, A., Letjend, J., Soemarto, P., Purwanegara, K., Purwokerto, T., & Banyumas, K. (2021). Analisis Sentimen pada review Aplikasi Grab di Google Play Store Menggunakan Support Vector Machine. *JURNAL INFORMATIKA*, *8*(2). http://ejournal.bsi.ac.id/ejurnal/index.php/ji

Wang, M., & Ji, H. (2022). Research on a Risk Control Model Based on Selective Ensemble Algorithm of Hierarchical Clustering and Simulated Annealing. *Procedia Computer Science*, *202*, 164–171. https://doi.org/10.1016/j.procs.2022.04.023

Wang, Y., Zhang, Y., Liang, M., Yuan, R., Feng, J., & Wu, J. (2023). National student loans default risk prediction: A heterogeneous ensemble learning approach and the SHAP method. *Computers and Education: Artificial Intelligence*, *5*. https://doi.org/10.1016/j.caeai.2023.100166

Widodo, A. O., Setiawan, B., & Indraswari, R. (2024). Machine Learning-Based Intrusion Detection on Multi-Class Imbalanced Dataset Using SMOTE. *Procedia Computer Science*, *234*, 578–583. https://doi.org/10.1016/j.procs.2024.03.042

Xiao, P., Liu, Z., Zhao, G., & Pan, P. (2024). Novel stacking models based on SMOTE for the prediction of rockburst grades at four deep gold mines. *Underground Space*. https://doi.org/10.1016/j.undsp.2024.03.004

Yu, Y., Scheidegger, S., Elliott, J., & Löfgren, Å. (2024). climateBUG [Formula presented]: A data-driven framework for analyzing bank reporting through a climate lens. *Expert Systems with Applications*, *239*. https://doi.org/10.1016/j.eswa.2023.122162

Zeng, Q., Gong, Z., Wu, S., Zhuang, C., & Li, S. (2024). Measuring cyclists' subjective perceptions of the street riding environment using K-means SMOTE-RF model and street view imagery. *International Journal of Applied Earth Observation and Geoinformation*, *128*. https://doi.org/10.1016/j.jag.2024.103739

Zhang, J., Cai, K., & Wen, J. (2024). A survey of deep learning applications in cryptocurrency. In *iScience* (Vol. 27, Issue 1). Elsevier Inc. https://doi.org/10.1016/j.isci.2023.108509