

TESIS
PENERAPAN DAN EVALUASI MODEL *MACHINE*
***LEARNING* TERLATIH ANTAR DOMAIN PADA ANALISIS**
SENTIMEN ULASAN APLIKASI KEUANGAN DIGITAL



disusun oleh

RAMA ASHARI

23.55.1410

Konsentrasi : Business Intelligence

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025

TESIS
PENERAPAN DAN EVALUASI MODEL *MACHINE*
***LEARNING* TERLATIH ANTAR DOMAIN PADA ANALISIS**
SENTIMEN ULASAN APLIKASI KEUANGAN DIGITAL

CROSS-DOMAIN IMPLEMENTATION AND EVALUATION
OF TRAINED MACHINE LEARNING MODELS FOR
SENTIMENT ANALYSIS OF DIGITAL FINANCIAL
APPLICATION REVIEWS



disusun oleh

RAMA ASHARI

23.55.1410

Konsentrasi : Business Intelligence

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025

HALAMAN PERSETUJUAN

**PENERAPAN DAN EVALUASI MODEL MACHINE LEARNING
TERLATIH ANTAR DOMAIN PADA ANALISIS SENTIMEN ULASAN
APLIKASI KEUANGAN DIGITAL**

**CROSS-DOMAIN IMPLEMENTATION AND EVALUATION OF
TRAINED MACHINE LEARNING MODELS FOR SENTIMENT
ANALYSIS OF DIGITAL FINANCIAL APPLICATION REVIEWS**

yang disusun dan diajukan oleh

Rama Ashari
23.55.1410

telah disetujui oleh Dosen Pembimbing Tesis
pada tanggal 08 Oktober 2025

Dosen Pembimbing,



Dr. Arief Setyanto, S.Si., M.T.
NIK. 190302036

HALAMAN PENGESAHAN

**PENERAPAN DAN EVALUASI MODEL MACHINE LEARNING
TERLATIH ANTAR DOMAIN PADA ANALISIS SENTIMEN ULASAN
APLIKASI KEUANGAN DIGITAL**

**CROSS-DOMAIN IMPLEMENTATION AND EVALUATION OF
TRAINED MACHINE LEARNING MODELS FOR SENTIMENT
ANALYSIS OF DIGITAL FINANCIAL APPLICATION REVIEWS**

yang disusun dan diajukan oleh

Rama Ashari

23.55.1410

Telah dipertahankan di depan Dewan Penguji
pada tanggal 08 Oktober 2025

Susunan Dewan Penguji

Nama Penguji

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

Dr. Andi Sunvoto, M.Kom.
NIK. 190302052

Dr. Arlef Setyanto, S.Si., M.T.
NIK. 190302036

Tanda Tangan



Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer
Tanggal 08 Oktober 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertandatangan di bawah ini,

Nama mahasiswa : Rama Ashari
NIM : 23.55.1410

Menyatakan bahwa Tesis dengan judul berikut:

Penerapan dan Evaluasi Model Machine Learning Terlatih Antar Domain pada Analisis Sentimen Ulasan Aplikasi Keuangan Digital

Dosen Pembimbing : Dr. Arief Setyanto, S.Si., M.T.

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, 08 Oktober 2025

Yang Menyatakan,



Rama Ashari

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur ke hadirat Allah SWT atas segala rahmat, petunjuk, dan kekuatan yang diberikan, karya tesis ini saya persembahkan kepada:

1. Kedua orang tua tersayang, atas doa, kasih sayang, dan pengorbanan yang tiada henti yang menjadi sumber semangat dalam setiap langkah.
2. Istri tercinta Mar'atunnafsiah, yang senantiasa memberikan dukungan moral, motivasi, dan doa dalam menyelesaikan studi ini.
3. Para dosen dan pembimbing, atas bimbingan, ilmu, dan keteladanan yang menjadi dasar bagi pengembangan diri dan karya ilmiah ini.
4. Sahabat dan rekan seperjuangan, yang selalu memberi semangat dan kebersamaan selama proses penelitian dan penulisan tesis ini.

Semoga karya sederhana ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan dan menjadi langkah kecil menuju kemajuan di bidang teknologi informasi.

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah Subhanahu wa Ta'ala atas segala rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan tesis dengan judul:

“Penerapan dan Evaluasi Model Machine Learning Terlatih Antar Domain pada Analisis Sentimen Ulasan Aplikasi Keuangan Digital.”

Tesis ini disusun sebagai salah satu syarat untuk memperoleh gelar Magister pada Program Studi Magister Teknik Informatika.

Penulis menyampaikan penghargaan dan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. Arief Setyanto, S.T., M.Kom., selaku dosen pembimbing utama yang telah memberikan bimbingan, arahan, dan masukan yang berharga selama proses penyusunan tesis ini.
2. Bapak Dr. Robert Marco, S.T., M.Kom., dan Bapak Dr. Andi Sunyoto, S.T., M.T., selaku dosen penguji yang telah memberikan saran dan koreksi konstruktif demi penyempurnaan penelitian ini.
3. Seluruh dosen dan staf Program Studi Magister Informatika, atas ilmu dan bantuan yang telah diberikan selama masa studi.
4. Kedua orang tua tercinta, atas doa, kasih sayang, dan dukungan moral yang tiada henti.
5. Keluarga dan sahabat, yang selalu memberikan motivasi, semangat, dan doa dalam setiap tahap penyelesaian tesis ini.

Penulis menyadari bahwa karya ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa mendatang. Semoga tesis ini dapat memberikan manfaat dan kontribusi bagi pengembangan ilmu pengetahuan, khususnya di bidang analisis sentimen dan pembelajaran mesin.

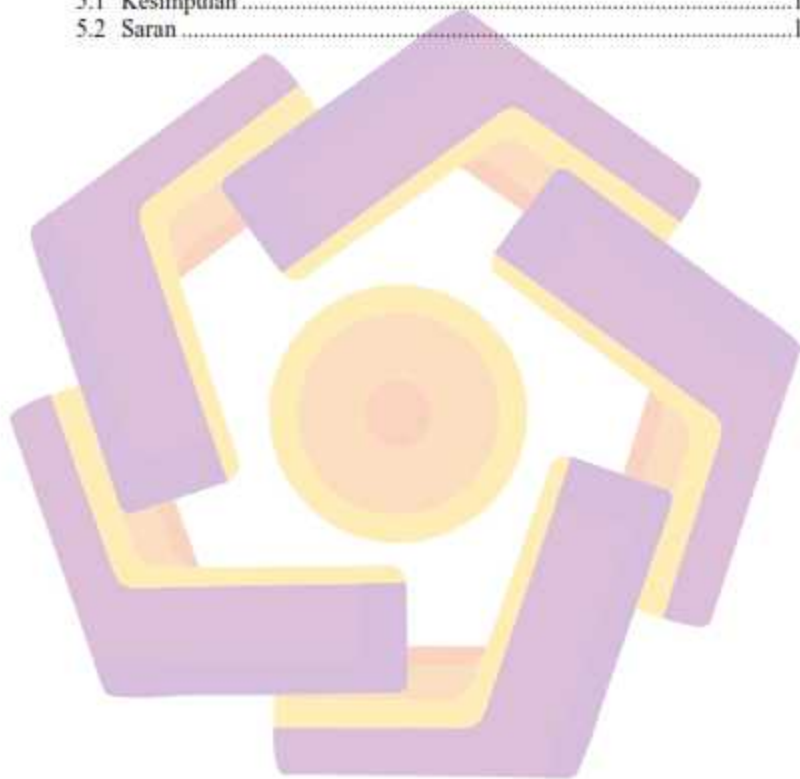
Yogyakarta, 08 Oktober 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN TESIS	Error! Bookmark not defined.
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMBANG DAN SINGKATAN	xi
DAFTAR ISTILAH	xiii
INTISARI.....	xiv
ABSTRACT.....	xv
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian.....	6
1.5 Manfaat Penelitian	6
BAB 2 TINJAUAN PUSTAKA	9
2.1 Tjauan Pustaka.....	9
2.2 Keaslian Penelitian	13
2.3 Landasan Teori	15
BAB 3 METODE PENELITIAN	26
3.1 Jenis, Sifat, dan Pendekatan Penelitian.....	26
3.2 Metode Pengumpulan Data.....	27
3.3 Metode Analisis Data.....	33
3.4 Alur Penelitian	42
BAB 4 HASIL PENELITIAN DAN PEMBAHASAN	43
4.1 Deskripsi Penelitian	43
4.2 Preprocessing Data	44
4.3 Pembobotan Kata.....	58
4.4 Pelatihan Model Dataset Publik.....	61
4.5 Hyperparameter Tuning.....	69
4.6 Analisis Signifikansi Statistik Antar Model pada Dataset Publik	75

4.7 Penerapan Model Terlatih pada Data Ulasan Aplikasi Livin' by Mandiri 79	
4.8 Pelatihan Model Dataset Livin' by Mandiri	89
4.9 Analisis Signifikansi Statistik Antar Model pada Dataset Livin'	91
4.10 Penerapan Model Livin' by Mandiri pada Dataset Publik	94
4.11 Perbandingan Hasil dengan Penelitian Terdahulu	104
BAB 5 PENUTUP	107
5.1 Kesimpulan	107
5.2 Saran	109



DAFTAR TABEL

Tabel 4.1 Sampel Hasil Pembersihan Data Ulasan Livin By Mandiri.....	45
Tabel 4.2 Sampel Hasil Pembersihan Data Publik.....	46
Tabel 4.3 Sampel Hasil Tokenisasi Data Ulasan Livin By Mandiri.....	47
Tabel 4.4 Sampel Hasil Tokenisasi Data Publik.....	48
Tabel 4.5 Sampel Hasil Stopword Removal Data Ulasan Livin By Mandiri.....	49
Tabel 4.6 Sampel Hasil Stopword Removal Data Publik.....	49
Tabel 4.7 Sampel Hasil Stemming Data Ulasan Livin By Mandiri.....	51
Tabel 4.8 Sampel Hasil Stemming Data Publik.....	51
Tabel 4.9 Sampel Hasil Lemmatization Data Ulasan Livin By Mandiri.....	52
Tabel 4.10 Sampel Hasil Lemmatization Data Publik.....	53
Tabel 4.11 Sampel Hasil Vektorisasi Kalimat.....	59
Tabel 4.12 Hasil Evaluasi Awal Model SVM Sebelum Tuning.....	61
Tabel 4.13 Hasil Evaluasi Awal Model ANN Sebelum Tuning.....	64
Tabel 4.14 Hasil Evaluasi Awal Model LSTM Sebelum Tuning.....	66
Tabel 4.15 Perbandingan Evaluasi Model Setelah Tuning.....	74
Tabel 4.16 Hasil 5-fold cross-validation Model.....	75
Tabel 4.17 SVM vs LSTM.....	76
Tabel 4.18 SVM vs ANN.....	77
Tabel 4.19 ANN vs LSTM.....	78
Tabel 4.20 Hasil Uji Signifikansi Statistik antar Model pada Dataset Publik.....	79
Tabel 4.21 Sampel Hasil Prediksi Sentimen Model Terlatih pada Ulasan Livin' by Mandiri.....	80
Tabel 4.22 Perbandingan Hasil Evaluasi Model Terlatih pada Dataset Livin' by Mandiri.....	87
Tabel 4.23 Jumlah Kesalahan Prediksi antar Model.....	88
Tabel 4.24 Perbandingan Hasil Evaluasi Model yang Dilatih pada Dataset Livin by Mandiri.....	90
Tabel 4.25 Hasil 5-fold cross-validation Model.....	91
Tabel 4.26 SVM vs LSTM.....	92
Tabel 4.27 SVM vs ANN.....	93
Tabel 4.28 ANN vs LSTM.....	93
Tabel 4.29 Hasil Uji Signifikansi Statistik antar Model pada Dataset Livin.....	94
Tabel 4.30 Sampel Hasil Prediksi Model Livin pada Dataset Publik.....	95
Tabel 4.31 Perbandingan Hasil Evaluasi Model Livin by Mandiri pada Dataset Publik.....	102
Tabel 4.32 Jumlah Kesalahan Prediksi antar Model.....	103
Tabel 4.33 Perbandingan Akurasi dengan Penelitian Terdahulu pada Domain Spesifik.....	104

DAFTAR GAMBAR

Gambar 4.1 Word Cloud Sentimen Positif Dataset Publik.....	54
Gambar 4.2 Word Cloud Sentimen Negatif Dataset Publik.....	55
Gambar 4.3 Word Cloud Sentimen Positif Dataset Livin by Mandiri.....	56
Gambar 4.4 Word Cloud Sentimen Negatif Dataset Livin by Mandiri.....	57
Gambar 4.5 Distribusi Rata-rata Nilai dalam Vektor FastText.....	60
Gambar 4.6 Confusion Matrix Model SVM Sebelum Tuning.....	63
Gambar 4.7 Confusion Matrix Model ANN Sebelum Tuning.....	65
Gambar 4.8 Confusion Matrix Model LSTM Sebelum Tuning.....	68
Gambar 4.9 Word Cloud Hasil Prediksi Positif Model SVM.....	82
Gambar 4.10 Word Cloud Hasil Prediksi Positif Model ANN.....	83
Gambar 4.11 Word Cloud Hasil Prediksi Positif Model LSTM.....	83
Gambar 4.12 Word Cloud Hasil Prediksi Negatif Model SVM.....	84
Gambar 4.13 Word Cloud Hasil Prediksi Negatif Model ANN.....	85
Gambar 4.14 Word Cloud Hasil Prediksi Negatif Model LSTM.....	85
Gambar 4.15 Jumlah Prediksi per Kelas Sentimen dari Setiap Model.....	86
Gambar 4.16 Word Cloud Hasil Prediksi Positif Model SVM.....	97
Gambar 4.17 Word Cloud Hasil Prediksi Positif Model ANN.....	97
Gambar 4.18 Word Cloud Hasil Prediksi Positif Model LSTM.....	98
Gambar 4.19 Word Cloud Hasil Prediksi Negatif Model SVM.....	99
Gambar 4.20 Word Cloud Hasil Prediksi Negatif Model ANN.....	99
Gambar 4.21 Word Cloud Hasil Prediksi Negatif Model LSTM.....	100
Gambar 4.22 Jumlah Prediksi per Kelas Sentimen dari Setiap Model.....	101

DAFTAR LAMBANG DAN SINGKATAN

x	Vektor fitur dari data input
y	Label atau kelas dari data (positif/negatif)
w	Vektor bobot (weight vector) pada model
b	Nilai bias pada model
$f(x)$	Fungsi keputusan (decision function) pada SVM
C	Parameter regularisasi pada SVM
γ	Parameter kernel RBF pada SVM
σ	Fungsi aktivasi (activation function)
η	Learning rate dalam proses pelatihan
L	Fungsi loss atau error model
t	Waktu atau langkah dalam urutan data sekuensial
h_t	Hidden state pada LSTM pada waktu ke- t
c_t	Cell state pada LSTM pada waktu ke- t
i_t	Input gate pada LSTM
f_t	Forget gate pada LSTM
o_t	Output gate pada LSTM
\tilde{c}_t	Candidate cell state pada LSTM
W	Matriks bobot (weight matrix)
U	Matriks bobot tambahan (recurrent weight matrix)
ReLU	Rectified Linear Unit – fungsi aktivasi non-linier
α	Tingkat signifikansi (significance level) pada uji-t
μ	Nilai rata-rata (mean)
σ^2	Varians data
n	Jumlah sampel (jumlah observasi)
k	Jumlah lipatan (fold) dalam k-fold cross-validation
AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
CSV	Comma-Separated Values
DL	Deep Learning
FN	False Negative
FP	False Positive
LSTM	Long Short-Term Memory
MLP	Multi-Layer Perceptron
NLP	Natural Language Processing
OOV	Out-of-Vocabulary

RNN	Recurrent Neural Network
SVM	Support Vector Machine
SMOTE	Synthetic Minority Over-sampling Technique
TF-IDF	Term Frequency - Inverse Document Frequency
TN	True Negative
TP	True Positive
URL	Uniform Resource Locator
ReLU	Rectified Linear Unit
SGD	Stochastic Gradient Descent
CSV	Comma Separated Values
GPU	Graphics Processing Unit
K-Fold CV	K-Fold Cross Validation
MSE	Mean Squared Error
SRM	Structural Risk Minimization



DAFTAR ISTILAH

Analisis Sentimen	Proses menganalisis teks untuk menentukan kecenderungan opini, apakah positif, negatif, atau netral.
Cross-Domain Sentiment Analysis	Pendekatan untuk menguji model analisis sentimen pada domain berbeda dari data pelatihannya, guna menilai kemampuan generalisasi model.
Dataset Publik	Kumpulan data yang tersedia secara terbuka (open-source) dan digunakan untuk melatih model klasifikasi sentimen.
FastText	Metode pembelajaran representasi kata (word embedding) yang memperhitungkan sub-kata (subword) sehingga mampu memproses bahasa dengan morfologi kompleks.
Hyperparameter Tuning	Proses penyesuaian parameter eksternal model untuk memperoleh performa terbaik.
Klasifikasi Sentimen	Proses pengelompokan teks ke dalam kategori berdasarkan polaritas emosi atau opini.
Machine Learning	Cabang kecerdasan buatan yang menggunakan algoritma untuk mempelajari pola dari data.
Overfitting	Kondisi ketika model terlalu menyesuaikan data latih sehingga gagal melakukan generalisasi pada data baru.
Preprocessing Teks	Tahap pembersihan dan normalisasi teks sebelum dimasukkan ke model pembelajaran mesin.
Word Embedding	Representasi numerik dari kata-kata dalam bentuk vektor yang menggambarkan makna semantik.
Accuracy	Persentase prediksi yang benar dibandingkan dengan total prediksi.
Precision	Proporsi prediksi positif yang benar-benar positif.
Recall	Proporsi data positif yang berhasil diprediksi dengan benar oleh model.
F1-Score	Rata-rata harmonis antara precision dan recall.
Grid Search	Teknik pencarian sistematis kombinasi hyperparameter terbaik berdasarkan evaluasi performa model.
Ulasan (Review)	Pendapat atau tanggapan pengguna terhadap aplikasi atau layanan yang digunakan.
Livin' by Mandiri	Aplikasi mobile banking yang dikembangkan oleh Bank Mandiri sebagai sarana layanan keuangan digital.

INTISARI

Penelitian ini mengevaluasi model machine learning (SVM, ANN, dan LSTM) untuk analisis sentimen lintas domain (cross-domain) pada ulasan aplikasi keuangan digital Livin' by Mandiri. Tujuan penelitian adalah (1) membandingkan performa model sebelum dan sesudah hyperparameter tuning pada dataset publik, (2) mengukur kemampuan generalisasi model dari dataset publik ke dataset spesifik Livin' by Mandiri, dan (3) menentukan model yang paling andal. Metode penelitian menggunakan supervised learning yang mencakup preprocessing teks, pembobotan kata FastText, penanganan data tidak seimbang menggunakan SMOTE, dan hyperparameter tuning.

Hasil penelitian pada dataset publik menunjukkan hyperparameter tuning berhasil meningkatkan akurasi ketiga model hingga di atas 94%. LSTM (Tuned) mencapai akurasi tertinggi (0.9422), namun analisis statistik menunjukkan tidak ada perbedaan performa yang signifikan ($p\text{-value} > 0.05$) di antara ketiga model. Saat model publik diuji pada data Livin', terjadi penurunan performa drastis (akurasi 39-42%), yang membuktikan adanya masalah domain shift parah. Sebaliknya, saat model dilatih secara langsung pada dataset Livin', model ANN (Artificial Neural Network) menunjukkan performa terbaik dengan akurasi 91.21%, mengungguli LSTM (90.89%), SVM (86.35%), dan penelitian terdahulu.

Kesimpulannya, hyperparameter tuning krusial untuk data publik, namun ketiga model memiliki performa yang setara. Model yang dilatih pada data publik gagal digeneralisasi ke domain spesifik. Untuk domain aplikasi Livin' by Mandiri, ANN terbukti menjadi model yang paling optimal dan unggul secara signifikan.

Kata kunci: Analisis Sentimen, Cross-Domain, SVM, ANN, LSTM.

ABSTRACT

This research evaluates machine learning models (SVM, ANN, and LSTM) for cross-domain sentiment analysis on reviews of the digital financial application Livin' by Mandiri. The objectives are (1) to compare model performance before and after hyperparameter tuning on a public dataset, (2) to measure the generalization capability of models from the public dataset to the specific Livin' by Mandiri dataset, and (3) to determine the most reliable model. The research method uses supervised learning which includes text preprocessing, FastText word embedding, imbalanced data handling using SMOTE, and hyperparameter tuning.

The results on the public dataset show that hyperparameter tuning successfully increased the accuracy of all three models to over 94%. LSTM (Tuned) achieved the highest accuracy (0.9422), but statistical analysis showed no significant performance difference (p -value > 0.05) among the three models. When the public models were tested on the Livin' data, a drastic performance drop occurred (39-42% accuracy), proving a severe domain shift problem. Conversely, when the models were trained directly on the Livin' dataset, the ANN (Artificial Neural Network) model demonstrated the best performance with an accuracy of 91.21%, outperforming LSTM (90.89%), SVM (86.35%), and previous studies.

In conclusion, hyperparameter tuning is crucial for the public dataset, yet all three models perform equivalently. Models trained on public data fail to generalize to a specific domain. For the Livin' by Mandiri application domain, ANN proved to be the most optimal and significantly superior model.

Keyword: *Sentiment Analysis, Cross-Domain, SVM, ANN, LSTM.*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis sentimen merupakan salah satu bidang yang mendapatkan perhatian besar dalam domain *Natural Language Processing* (NLP) dan *machine learning*. Di era digital, ulasan pelanggan yang tersebar di berbagai platform online memberikan data yang sangat berharga untuk perusahaan dalam memahami kepuasan dan persepsi pelanggan. Namun, tantangan utama yang muncul adalah bagaimana cara mengekstraksi informasi bermakna dari teks yang tidak terstruktur dan beragam ini. Ulasan pelanggan tidak hanya berperan sebagai referensi bagi calon pelanggan lainnya, tetapi juga berfungsi sebagai umpan balik yang kritis bagi penyedia layanan. Ulasan membantu dalam pemahaman pasar, perbaikan produk, dan pengambilan keputusan bisnis yang lebih tepat (Yang et al., 2020). Kepuasan pelanggan, sebagai faktor utama dalam menilai sejauh mana produk atau layanan memenuhi harapan pelanggan, juga menjadi indikator penting (Chepukaka & Kirugi, 2019). Pertumbuhan eksponensial data teks yang dihasilkan oleh pengguna di berbagai platform digital, seperti ulasan produk, media sosial, dan forum diskusi, menciptakan tantangan baru bagi para peneliti untuk mengekstraksi informasi yang bermakna dari data teks yang tidak terstruktur ini (Liu, 2020).

Berbagai algoritma *machine learning* telah diterapkan dalam analisis sentimen, termasuk *Decision Tree*, *Random Forest*, *Naïve Bayes*, *Artificial Neural Networks* (ANN), dan *Support Vector Machine* (SVM). *Decision Tree* dikenal karena interpretabilitasnya yang tinggi, tetapi sering mengalami *overfitting*, terutama ketika bekerja dengan dataset besar dan tidak seimbang (Rokach & Maimon, 2005). Untuk mengatasi masalah *overfitting*, *Random Forest* dibentuk sebagai ansambel dari banyak pohon keputusan, yang umumnya meningkatkan

akurasi namun membutuhkan biaya komputasi yang lebih tinggi (Breiman, 2001). Naïve Bayes adalah algoritma yang efisien untuk data teks, namun asumsi independensi fitur sering kali tidak realistis, yang dapat menyebabkan penurunan akurasi dalam beberapa kasus (Zhang, 2004).

Artificial Neural Networks (ANN) dan *Support Vector Machine (SVM)* adalah algoritma yang lebih kompleks dan kuat. ANN, terutama dalam bentuk deep learning, mampu menangkap pola non-linear dan hubungan kompleks dalam data, tetapi memerlukan jumlah data yang besar serta penyesuaian parameter yang tepat untuk mencapai performa optimal (Bengio, 2009; Schmidhuber, 2015). SVM sangat efektif dalam ruang fitur yang tinggi dan biasanya memberikan hasil yang baik, tetapi kurang efisien ketika diterapkan pada dataset yang sangat besar (Cortes & Vapnik, 1995).

Lebih lanjut, pendekatan deep learning berbasis arsitektur Recurrent Neural Network (RNN) dan variannya, Long Short-Term Memory (LSTM), menunjukkan performa yang lebih baik pada data sekuensial. LSTM terbukti efektif dalam menangani masalah vanishing gradient pada RNN konvensional serta mampu memodelkan hubungan temporal dalam teks (Hochreiter & Schmidhuber, 1997a). Meski demikian, kebutuhan komputasi yang tinggi dan waktu pelatihan yang relatif lama tetap menjadi tantangan yang perlu dipertimbangkan (Gers et al., 2000).

Selain itu, performa model seperti SVM, ANN, dan LSTM sangat dipengaruhi oleh pemilihan hyperparameter. Proses optimasi hyperparameter (tuning) terbukti mampu meningkatkan akurasi dan kemampuan generalisasi model, sementara konfigurasi default berisiko menghasilkan performa yang suboptimal (Bergstra & Bengio, 2012). Penelitian terdahulu di Indonesia telah banyak menerapkan metode machine learning untuk analisis sentimen, misalnya oleh (Handayani et al., 2020), (Haque et al., 2018), dan (Prananda & Thalib, 2020). Namun, sebagian besar penelitian masih berfokus pada perbandingan algoritma dengan hyperparameter default, tanpa mengeksplorasi secara mendalam aspek tuning dan evaluasi lintas domain. Padahal, tantangan utama dalam analisis

sentimen adalah kemampuan generalisasi model ketika diaplikasikan pada domain berbeda, yang dikenal dengan cross-domain sentiment analysis. Sebuah model yang dilatih pada data ulasan umum sering kali mengalami penurunan performa ketika diterapkan pada domain spesifik, seperti aplikasi keuangan digital, karena adanya perbedaan kosakata, gaya bahasa, serta konteks penggunaan (Glorot et al., 2011).

Berdasarkan kondisi tersebut, penelitian ini menawarkan kontribusi dalam mengevaluasi model machine learning terlatih antar domain (cross-domain) pada analisis sentimen ulasan aplikasi keuangan digital di Indonesia. Penelitian ini secara khusus membandingkan model SVM, ANN, dan LSTM dengan menerapkan hyperparameter tuning secara sistematis. Selain itu, evaluasi dilakukan dua arah, yakni dengan melatih model pada dataset publik dan mengujinya pada dataset spesifik, serta sebaliknya. Pendekatan ini diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai daya tahan (robustness) dan kemampuan generalisasi model analisis sentimen berbahasa Indonesia di berbagai domain.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, tantangan utama dalam analisis sentimen pada domain spesifik seperti aplikasi keuangan digital adalah kelangkaan dataset berlabel. Hal ini menimbulkan pertanyaan mengenai kemampuan generalisasi model yang dilatih pada domain data yang berbeda. Oleh karena itu, penelitian ini merumuskan tiga masalah utama sebagai berikut:

1. Bagaimana performa perbandingan model Support Vector Machine (SVM), Artificial Neural Network (ANN), dan Long Short-Term Memory (LSTM) sebelum dan sesudah optimasi menggunakan hyperparameter

tuning dalam tugas klasifikasi sentimen pada dataset publik berbahasa Indonesia?

2. Se jauh mana kapabilitas generalisasi model yang telah dilatih pada dataset publik ketika diuji secara lintas domain (cross-domain) pada dataset ulasan aplikasi keuangan digital (Livin' by Mandiri) yang memiliki karakteristik data berbeda?
3. Model manakah yang menunjukkan performa paling optimal dan stabil dalam skenario pengujian lintas domain, baik dari domain publik ke spesifik maupun sebaliknya, untuk memberikan wawasan mengenai model yang paling adaptif untuk analisis sentimen berbahasa Indonesia?

1.3 Batasan Masalah

Berdasarkan latar belakang masalah dan rumusan masalah yang telah disajikan, berikut adalah batasan-batasan masalah penelitian ini:

1. Dataset publik yang digunakan dalam proses pelatihan model merupakan dataset sentimen berbahasa Indonesia yang tersedia secara open-source, dengan jumlah dan distribusi data sesuai yang tersedia pada repository publik, tanpa dilakukan modifikasi label.
2. Proses pelatihan dan hyperparameter tuning hanya dilakukan menggunakan dataset publik. Dataset ulasan aplikasi Livin' by Mandiri hasil scraping tidak digunakan dalam proses pelatihan model, melainkan hanya digunakan sebagai dataset uji (testing) untuk memperoleh prediksi sentimen.

3. Dataset ulasan aplikasi Livin' by Mandiri tidak memiliki label ground truth sentimen, sehingga hasil prediksi model pada dataset tersebut hanya dapat dianalisis melalui distribusi jumlah prediksi sentimen (positif dan negatif) dan tidak dilakukan evaluasi performa menggunakan metrik akurasi, precision, recall, maupun F1-score.
4. Proses preprocessing teks yang diterapkan pada dataset publik dan dataset ulasan aplikasi Livin' by Mandiri disamakan, yaitu melalui tahapan normalisasi, pembersihan karakter non-alfabet, tokenisasi, stopword removal, stemming, dan pembobotan kata menggunakan FastText embedding.
5. Model klasifikasi yang digunakan dalam penelitian ini terbatas pada tiga algoritma, yaitu; Support Vector Machine (SVM), Artificial Neural Network (ANN), dan Long Short-Term Memory (LSTM)
6. Hyperparameter tuning dilakukan secara terpisah untuk masing-masing model menggunakan metode GridSearchCV untuk SVM dan search space eksplorasi trial-based tuning untuk ANN dan LSTM, terbatas pada kombinasi parameter yang telah ditentukan dalam penelitian ini.
7. Evaluasi performa model hanya dilakukan pada dataset publik yang memiliki label valid, sedangkan pada dataset ulasan aplikasi Livin' by Mandiri hanya dilakukan analisis deskriptif terhadap distribusi prediksi sentimen.

Dengan membatasi cakupan penelitian pada aspek-aspek tersebut, diharapkan penelitian dapat memberikan hasil yang lebih fokus dan relevan terkait dengan tujuan dan rumusan masalah yang telah dijabarkan.

1.4 Tujuan Penelitian

Fokus utama penelitian ini adalah mengevaluasi kemampuan generalisasi model machine learning dan deep learning dalam skenario lintas domain (*cross-domain sentiment analysis*) antara dataset publik dan dataset ulasan aplikasi keuangan digital. Selaras dengan hal tersebut, maka tujuan dari penelitian ini adalah:

1. Mengevaluasi dan membandingkan performa model SVM, ANN, dan LSTM sebelum dan sesudah melalui proses hyperparameter tuning untuk menemukan konfigurasi parameter terbaik pada tugas klasifikasi sentimen ulasan berbahasa Indonesia.
2. Mengukur kemampuan generalisasi model yang dilatih menggunakan dataset publik saat diterapkan pada dataset ulasan aplikasi Livin' by Mandiri untuk mengidentifikasi penurunan performa akibat pergeseran domain (*domain shift*).
3. Menentukan model klasifikasi yang paling andal dan adaptif berdasarkan hasil evaluasi komprehensif dari skenario pengujian lintas domain dua arah, sebagai rekomendasi untuk aplikasi analisis sentimen di dunia nyata.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat baik secara teoretis maupun praktis dalam bidang klasifikasi sentimen berbasis machine learning dan

deep learning, khususnya pada data berbahasa Indonesia. Adapun manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Memberikan kontribusi ilmiah dalam pengembangan metode klasifikasi sentimen untuk Bahasa Indonesia melalui penerapan tiga algoritma populer, yaitu Support Vector Machine (SVM), Artificial Neural Network (ANN), dan Long Short-Term Memory (LSTM) dengan skema pelatihan menggunakan dataset publik yang telah tervalidasi.
2. Menghadirkan pendekatan penerapan model hasil pelatihan dari dataset publik ke dataset domain spesifik (aplikasi Livin' by Mandiri) yang tidak memiliki label sentimen. Konsep ini menjadi bentuk sederhana dari cross-domain sentiment classification, yang jarang diaplikasikan pada konteks Bahasa Indonesia, khususnya untuk aplikasi keuangan digital.
3. Memperkaya literatur tentang efektivitas hyperparameter tuning dalam meningkatkan performa klasifikasi sentimen pada ketiga algoritma yang dibandingkan, sekaligus menunjukkan bahwa tuning parameter dapat berdampak signifikan terhadap hasil klasifikasi, terutama dalam konteks data teks ulasan pengguna.
4. Menyediakan gambaran distribusi kecenderungan sentimen pengguna aplikasi Livin' by Mandiri berdasarkan hasil prediksi model machine learning dan deep learning, sehingga dapat menjadi masukan awal bagi pengembang aplikasi dalam memahami persepsi pengguna terhadap aplikasi yang dikelola.
5. Menawarkan alur metodologi sistematis untuk penelitian serupa yang ingin menerapkan model hasil pelatihan dari dataset publik ke dataset spesifik.

tanpa label, sebagai solusi praktis dalam menghadapi keterbatasan ketersediaan dataset berlabel di ranah industri atau aplikasi digital nasional.

6. Mendorong pemanfaatan dataset publik open-source Bahasa Indonesia yang telah tervalidasi sebagai basis pelatihan model, yang dapat menghemat waktu, biaya, dan sumber daya dalam proses pengembangan sistem analisis sentimen berbasis AI di berbagai sektor.
7. Menghasilkan prototipe model analisis sentimen yang telah tervalidasi, yang nantinya dapat diimplementasikan oleh pengembang atau perusahaan untuk memonitor sentimen pelanggan secara *real-time* dari berbagai platform di luar Google Play Store, seperti media sosial (contohnya Twitter) atau portal berita, di mana data ulasan tidak memiliki label eksplisit seperti rating bintang.

BAB 2

TINJAUAN PUSTAKA

2.1 Tujuan Pustaka

Sentimen analisis memiliki dampak yang cukup signifikan dalam kemajuan dan kelayakan pada layanan digital suatu perusahaan. Oleh karena itu, tidak sedikit peneliti yang melakukan penelitian tentang hal tersebut. Beberapa penelitian mengklasifikasikan ulasan yang dituliskan oleh pengguna berdasarkan kelas Positif dan Negatif.

Menganalisis sentimen pada sebuah kalimat memerlukan pembersihan kata dari imbuhan yang tidak memiliki arti. Proses ini dinamakan proses stemming, dimana kata diolah dan dibuang imbuhan dan menormalkan kalimat untuk pemahaman yang lebih baik (Divya et al., 2020). Selain itu, hal yang penting juga adalah penghapusan stopwords. Stopwods adalah kata yang sering muncul dan kurang memiliki makna. Manfaat utama penghapusan stopword adalah untuk meningkatkan kinerja sistem dalam hal pengurangan memori dan menghemat waktu pemrosesan selama pengindeksan dan pencarian (Rani & Lobiyal, 2022).

FastText adalah sebuah metode untuk pembelajaran representasi kata (word embedding) dan klasifikasi teks yang dikembangkan oleh tim peneliti di Facebook AI Research (FAIR). Metode ini merupakan perluasan dari model Word2Vec (Mikolov et al., 2013), namun dirancang untuk mengatasi beberapa

keterbatasan fundamentalnya, terutama dalam menangani kata-kata yang jarang muncul dan kata-kata di luar kamus (Out-of-Vocabulary atau OOV).

Inovasi utama dari FastText terletak pada pemanfaatan informasi sub-kata (subword information). Berbeda dengan Word2Vec yang memperlakukan setiap kata sebagai satu unit tunggal, FastText merepresentasikan sebuah kata sebagai sekumpulan karakter n -gram (Bojanowski et al., 2017). Sebagai contoh, untuk kata "belajar", dengan $n=3$ (trigram), FastText akan memecahnya menjadi beberapa bagian seperti <be, bel, ela, laj, aja, jar, ar> (dengan < dan > sebagai penanda awal dan akhir kata), beserta kata utuh "belajar" itu sendiri. Setiap n -gram ini memiliki representasi vektornya sendiri, dan vektor untuk kata "belajar" dihitung sebagai jumlah dari semua vektor n -gram penyusunnya.

Selain untuk menghasilkan word embedding, pustaka FastText juga menyediakan arsitektur yang sangat efisien untuk klasifikasi teks. Model ini menggunakan pendekatan bag-of-words yang diperkaya dengan bag of n -grams, kemudian melatih model klasifikasi linear sederhana (Joulin et al., 2017). Meskipun sederhana, metode ini terbukti sangat cepat dan mampu mencapai performa yang setara dengan model deep learning yang jauh lebih kompleks pada banyak tugas standar.

Penelitian (Obiedat et al., 2022) mengenai sentimen analisis pada ulasan pelanggan di sebuah restoran telah menunjukkan betapa pentingnya informasi ini dalam meningkatkan kualitas layanan dan memahami persepsi pelanggan. Di lain pihak, analisis sentimen juga digunakan dalam berbagai platform, termasuk

aplikasi mobile banking yang mengklasifikasikan ulasan dari penggunanya berdasarkan kelas positif dan negatif dengan menggunakan Support Vector Machine (SVM) (Handayani et al., 2020). selain itu, sentimen analisis digunakan untuk mendeteksi ulasan pada e-commerce (Haque et al., 2018), layanan transportasi online (Prananda & Thalib, 2020). Pada penelitian (Patel et al., 2022), menganalisis sentimen feedback pelanggan dan review layanan maskapai menggunakan BERT dan menghasilkan akurasi tertinggi sebesar 83%. Penelitian analisis sentimen dilakukan untuk menganalisis review dari pelanggan terhadap aplikasi Uber (S. Ali et al., 2020). menganalisis review yang diberikan penonton pada konten YouTube lagu India dan Pakistan, komentar yang diambil merupakan komentar yang ditulis dalam aksara latin berbahasa Urdu, yang kemudian diklasifikasikan menggunakan beberapa algoritma dan menghasilkan akurasi tertinggi sebesar 92,25% menggunakan LR (Qureshi et al., 2022).

Pada penelitian terdahulu, beberapa peneliti melakukan perbandingan pada beberapa model. Setiap model yang digunakan memiliki hasil akurasi yang berbeda, algoritma KNN lebih baik dibandingkan Naïve Bayes pada klasifikasi sentimen analisis Digital Payment (Wisnu et al., 2020). Disisi lain, algoritma Decision Tree memiliki kinerja klasifikasi terbaik dibandingkan dengan Neural Network, NB, dan SVM (Prananda & Thalib, 2020). LSTM, yang merupakan jenis RNN, sangat efektif dalam menangani dependensi jangka panjang dalam data sekuensial (Hochreiter & Schmidhuber, 1997b).

Hyperparameter tuning adalah proses krusial dalam optimasi model pembelajaran mesin. Menurut (Bergstra & Bengio, 2012), optimasi

hyperparameter dapat meningkatkan performa model secara signifikan, terutama pada algoritma yang memiliki banyak parameter seperti ANN dan LSTM. Pendekatan seperti *grid search* dan *random search* adalah metode yang umum digunakan untuk menemukan kombinasi hyperparameter yang optimal. Penelitian oleh (Claesen & De Moor, 2015) juga menunjukkan bahwa *hyperparameter tuning* dapat mengurangi overfitting dan meningkatkan generalisasi model.



2.2 Keaslian Penelitian

Tabel 2.1 Matriks literatur review dan posisi penelitian

Penerapan dan Evaluasi Model Machine Learning Terlatih Antar Domain pada Analisis Sentimen Ulasan Aplikasi Keuangan Digital

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Kelebihan	Kelemahan dan Perbandingan
1	Analisis Sentimen Ulasan Aplikasi Perbankan di Google Play Store menggunakan Algoritma Support Vector Machine	Martinius Juan Prasetyo, I Made Artha Agastya, <i>Sistemasi: Jurnal Sistem Informasi</i> Volume 13, Nomor 6, 2024: 2386-2400, 2024	Menganalisis sentimen pengguna terhadap beberapa aplikasi perbankan digital di Indonesia.	SVM memberikan hasil akurasi sebesar 87,46% untuk dataset Livin by Mandiri	Menggunakan dataset multi-aplikasi, menunjukkan generalisasi model	Belum menggunakan regularisasi atau metode oversampling; fokus hanya pada SVM.
2	Analisis Sentimen Pengguna Mobile Banking Livin' by Mandiri Menggunakan Algoritma Support Vector Machine	Faizal Riza, Bagas Kurniawan, <i>Jurnal Ismetek</i> , Vol. 17 No.2, 2024	Mengklasifikasikan sentimen pengguna aplikasi Livin' by Mandiri menggunakan metode SVM dengan data yang tidak seimbang.	SVM efektif untuk data teks finansial dengan peningkatan melalui SMOTE dengan akurasi sebesar 81,68%	Menggunakan SMOTE untuk penyeimbangan kelas	Tidak melakukan perbandingan dengan algoritma lain dan tidak menguji lintas domain
3	Analisis Sentimen Aplikasi Livin dengan Algoritma Neural Network, Random	Muhammad Hafizul Ali, Juni Ismail, <i>Jurnal JPILKOM</i> , 3	Membandingkan kinerja tiga algoritma klasifikasi pada data ulasan	SVM memberikan hasil akurasi sebesar 84,9%, unggul tipis dibanding NN sebesar 83,9% dan	Melakukan komparasi antar algoritma klasik.	Belum ada teknik peningkatan performa (SMOTE, regularisasi, atau cross-domain).

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Kelebihan	Kelemahan dan Perbandingan
	Forest, dan SVM	(2), 2025	Livin by Mandiri.	RF 76,2% pad. dataset yang sama Livin by Mandiri.		
4	Impact of Feature Extraction on Multi-Aspect Sentiment Classification for Livin' by Mandiri Using BiLSTM	Balqis Sayyidahtul Atikah, Yuliant Sibaroni, Diyas Puspendari, <i>JOURNAL LA MULTIAPP</i> , 2024	Menilai pengaruh metode ekstraksi fitur terhadap hasil klasifikasi sentimen multi-aspek.	Akurasi BiLSTM tertinggi sebesar 87,3%	Menggunakan deep learning dan analisis multi-aspek.	Fokus pada single model, tidak dibandingkan dengan algoritma machine learning klasik.
5	Analisis Sentimen Pengguna Mobile Banking Livin' By Mandiri Menggunakan Algoritma Support Vector Machine	Faizal Riza, Bagus Kurniawan <i>Jurnal Ismetek</i> , Vol. 17 No.2 2024	Mengklasifikasikan sentimen pengguna aplikasi Livin' by Mandiri menjadi positif dan negatif menggunakan algoritma SVM.	Algoritma SVM mampu mengklasifikasikan sentimen dengan nilai akurasi 81,68%.	Menggunakan SMOTE dan cross validation untuk optimalisasi model.	Fokus hanya pada satu algoritma (SVM). Tidak melakukan perbandingan dengan model deep learning (ANN/LSTM) atau pengujian lintas domain.
6.	Perbandingan Performa Algoritma Klasifikasi Machine Learning Untuk Analisis Sentimen Pengguna Mobile Banking Livin' By Mandiri Pada Google Play Store	Fajzal Riza, Bagus Kurniawan <i>Jurnal Ismetek</i> , Vol. 18 No.2 2024	Membandingkan performa empat algoritma machine learning (SVM, LGBM, XGBM, CatBoost) untuk analisis sentimen ulasan Livin' by Mandiri.	XGBM memiliki akurasi tertinggi (86,8%) , diikuti oleh LGBM (86,2%), SVM (85,6%) , dan CatBoost (85,6%).	Melakukan komparasi antar beberapa algoritma machine learning. Menganalisis waktu eksekusi sebagai metrik evaluasi.	Akurasi tertinggi masih dibawah 90%

2.3 Landasan Teori

1. Analisis sentimen

Analisis sentimen adalah proses menggunakan *Natural Language Processing* (NLP), analisis teks, dan linguistik komputasional untuk mengidentifikasi dan mengekstrak informasi subjektif dari sumber teks. Analisis sentimen umumnya digunakan untuk menentukan apakah teks tertentu (seperti ulasan produk atau komentar media sosial) memiliki nada positif, negatif, atau netral. Teknik ini sangat berguna dalam berbagai aplikasi, termasuk pemasaran, riset pasar, dan layanan pelanggan (Yadav & Vishwakarma, 2020).

2. *Natural Language Processing* (NLP)

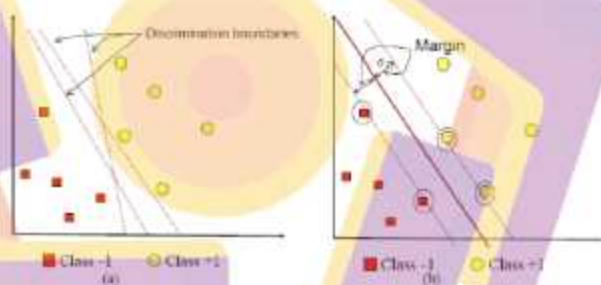
Natural Language Processing (NLP) adalah cabang kecerdasan buatan yang berfokus pada interaksi antara komputer dan manusia menggunakan bahasa alami. NLP mencakup berbagai teknik dan metode untuk mengubah teks atau ucapan manusia menjadi data yang dapat dipahami dan diproses oleh komputer. Dalam konteks analisis sentimen, beberapa teknik yang sering digunakan termasuk tokenisasi, stemming, lemmatization, dan transformasi teks ke vektor (seperti TF-IDF atau *word embeddings*) (Devlin et al., 2018).

3. Algoritma *Machine Learning* untuk Analisis Sentimen

Dalam analisis sentimen, berbagai algoritma pembelajaran mesin dapat digunakan untuk mengklasifikasikan teks berdasarkan sentimen yang terkandung di dalamnya. Beberapa algoritma populer termasuk:

A. *Support Vector Machine (SVM)*

Algoritma pembelajaran mesin yang diawasi yang dapat digunakan untuk regresi dan klasifikasi dikenal sebagai *Support Vector Machine (SVM)*. Teori SVM didasarkan pada prinsip *Structural Risk Minimization (SRM)*, yang bertujuan untuk menemukan hyperplane optimal yang dapat memisahkan data ke dalam kelas-kelas yang berbeda dengan margin terbesar (Cortes & Vapnik, 1995).



Gambar 2.1 Ilustrasi Metode Support Vector Machine (Sumber :

(Buntoro, 2017))

Dalam kasus klasifikasi biner, SVM mencari fungsi *Structural Risk Minimization (SRM)*, yang bertujuan untuk menemukan hyperplane optimal yang dapat memisahkan data ke dalam kelas-kelas yang berbeda dengan margin terbesar (Cortes & Vapnik, 1995). Dalam kasus klasifikasi biner, SVM mencari fungsi sebagai berikut:

$$f(x) = w^T x + b \quad (1)$$

Formula ini adalah fungsi keputusan yang digunakan oleh SVM untuk melakukan prediksi pada data baru. Di sini, x adalah vektor fitur dari data yang akan diklasifikasikan (hasil vektorisasi), w adalah vektor bobot, dan b adalah bias. Vektor bobot w dan bias b adalah parameter yang dipelajari oleh model selama proses pelatihan. Hasil dari fungsi ini, $f(x)$ akan menentukan kelas dari data tersebut; misalnya, jika hasilnya positif, data diklasifikasikan sebagai sentimen 'positif', dan sebaliknya.

$$w^T x + b = 0 \quad (2)$$

Persamaan ini mendefinisikan hyperplane pemisah, yaitu batas keputusan yang memisahkan data ke dalam kelas yang berbeda. Dalam ruang dua dimensi, hyperplane ini berupa sebuah garis. Dalam ruang tiga dimensi, ia berupa sebuah bidang datar, dan dalam ruang berdimensi lebih tinggi, ia disebut hyperplane. Semua titik data x yang memenuhi persamaan ini dianggap berada tepat di perbatasan antara dua kelas.

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (3)$$

Formula ini merepresentasikan tujuan utama dari algoritma SVM. Tujuan SVM adalah menemukan hyperplane yang memiliki margin atau jarak terbesar antara hyperplane itu sendiri dengan titik data terdekat dari masing-masing kelas. Secara matematis, memaksimalkan margin ini setara dengan meminimalkan nilai dari $\|w\|^2$ (norma kuadrat

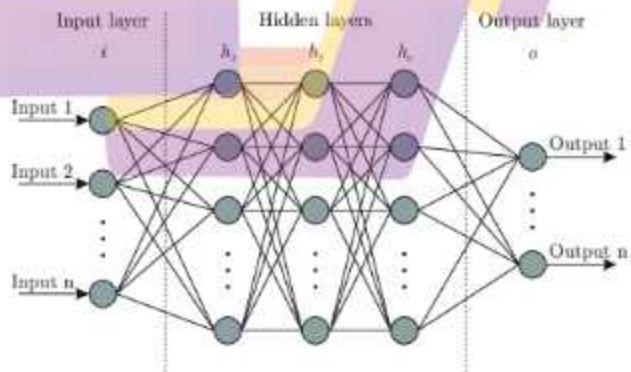
dari vektor bobot). Dengan meminimalkan nilai ini, algoritma secara efektif menemukan hyperplane pemisah yang paling optimal.

$$y_i (w^T x_i + b) \geq 1 \quad (4)$$

Formula ini adalah batasan atau syarat (constraint) yang harus dipenuhi oleh setiap data latih (x_i, y_i) . Syarat ini memastikan bahwa semua titik data tidak hanya diklasifikasikan dengan benar tetapi juga berada di sisi margin yang tepat. Di sini, y_i adalah label kelas dari data ke- i . Ketidakesetaraan ini memaksa setiap titik data untuk berada pada atau di luar batas margin dari kelasnya masing-masing, sehingga menciptakan "zona larangan" di antara kedua kelas dan memperkuat daya generalisasi model.

B. *Artificial Neural Networks (ANN)*

Artificial Neural Networks (ANN) adalah model komputasi yang terinspirasi oleh jaringan saraf biologis. ANN terdiri dari lapisan input, satu atau lebih lapisan tersembunyi, dan lapisan output.



Gambar 2.2 Ilustrasi Metode Artificial Neural Networks (Sumber :
(Bre et al., 2018))

Setiap lapisan terdiri dari neuron yang saling terhubung melalui bobot yang dapat disesuaikan. pada feedforward neural network seperti Multi-Layer Perceptron (MLP), informasi bergerak dalam satu arah dari input ke output. Output dari sebuah neuron dihitung menggunakan fungsi aktivasi sigma dari jumlah terbobot inputnya sebagai berikut:

$$a = \sigma \left(\sum_{i=1}^n w_i x_i + b \right) \quad (5)$$

Formula ini menjelaskan bagaimana output atau aktivasi (a) dari sebuah neuron dihitung. Dimana :

- a : Merupakan nilai output (aktivasi) dari neuron.
- σ : Merupakan fungsi aktivasi (misalnya, ReLU, Sigmoid, atau Tanh) yang mengubah hasil penjumlahan linear menjadi output non-linear.
- $w_i x_i$: Merupakan hasil perkalian antara bobot koneksi ke- i dengan nilai input ke- i .
- b : Merupakan bias.

Proses pelatihan ANN melibatkan penyesuaian bobot w_i dan bias b menggunakan algoritma optimasi seperti stochastic gradient descent (SGD) untuk meminimalkan fungsi kerugian (Goodfellow et al., 2016).

C. *Long Short-Term Memory (LSTM)*

Long Short-Term Memory (LSTM) adalah jenis jaringan saraf berulang (Recurrent Neural Network, RNN) yang dirancang khusus untuk mengatasi masalah dependensi jangka panjang (long-term dependencies) yang sering terjadi pada data sekuensial. Meskipun LSTM termasuk dalam keluarga *deep learning*, pada penelitian ini LSTM dikategorikan bersama algoritma machine learning karena digunakan dalam konteks yang sama untuk tugas klasifikasi supervisi, tanpa proses *transfer learning* atau arsitektur multi-layer kompleks seperti transformer. Tujuannya adalah memastikan perbandingan adil antar model berdasarkan kompleksitas pelatihan dan bukan kedalaman arsitektur. Penelitian ini tidak menggunakan model deep learning yang lebih baru seperti BERT atau RoBERTa karena fokusnya adalah pada model yang dapat dilatih dari awal (*from scratch*) dengan sumber daya komputasi terbatas dan tetap representatif untuk bahasa Indonesia. Struktur LSTM dirancang untuk mengatur aliran informasi menggunakan tiga gerbang utama dan sebuah cell state dan tiga gerbang (gates) (Hochreiter & Schmidhuber, 1997a) yang dirumuskan sebagai berikut :

$$f_t = \sigma(W_f x_t + U_f h_t + b_f) \quad (6)$$

dimana :

- f_t : Merupakan forget gate.
- x_t : Merupakan Vektor input pada waktu saat ini.
- h_{t-1} : Hidden state atau output dari waktu sebelumnya.

C_{t-1} : Cell state dari waktu sebelumnya.

W, U : Matriks bobot yang dipelajari selama pelatihan

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (7)$$

dimana :

i_t : Merupakan input gate.

$$C_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (8)$$

dimana :

C_t : Merupakan candidate state.

$$C_t = f_t \circ C_{t-1} + i_t \circ C_t \quad (9)$$

dimana :

C_t : Merupakan cell state.

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (10)$$

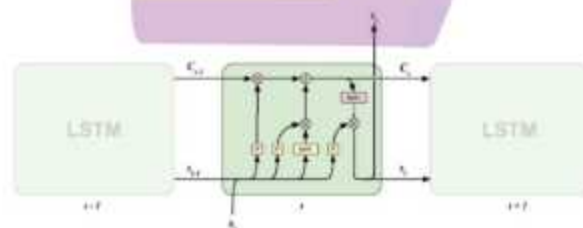
dimana :

o_t : Merupakan output gate.

$$h_t = o_t \circ \tanh(C_t) \quad (11)$$

dimana :

h_t : Merupakan final gate.



Gambar 2.3 Ilustrasi Metode Long Short-Term Memory (Sumber:
(Enriko et al., 2023))

4. Hyperparameter Tuning

Hyperparameter tuning adalah proses penyesuaian parameter-parameter eksternal yang mengontrol proses pembelajaran model. Hyperparameter seperti learning rate, jumlah lapisan dalam ANN, atau jenis kernel dalam SVM dapat memiliki dampak besar pada kinerja model. Tuning hyperparameter dilakukan untuk menemukan kombinasi parameter yang optimal, yang dapat meningkatkan performa dan kemampuan generalisasi model (Bergstra & Bengio, 2012).

Salah satu metode yang paling umum digunakan untuk hyperparameter tuning adalah Grid Search. Metode ini bekerja dengan cara mendefinisikan sebuah "kisi" (grid) dari kombinasi hyperparameter yang mungkin, lalu melatih dan mengevaluasi model untuk setiap kombinasi tersebut secara sistematis. Misalnya, untuk model SVM, kita dapat mendefinisikan kisi untuk parameter C [0.1, 1, 10] dan kernel ['linear', 'rbf']. Grid Search akan menguji semua $3 \times 2 = 6$ kombinasi. Untuk mengevaluasi setiap kombinasi, sering kali digunakan teknik cross-validation untuk memastikan bahwa skor performa yang dihasilkan stabil dan tidak bergantung pada pembagian data latih-uji tertentu (Pedregosa et al., 2011). Meskipun komprehensif, Grid Search bisa menjadi sangat lambat jika ruang pencarian parameternya besar.

5. Metode Evaluasi Model

Evaluasi kinerja model pembelajaran mesin adalah langkah penting untuk memastikan bahwa model yang dikembangkan memiliki kinerja yang baik dan dapat diandalkan. Dalam klasifikasi teks, beberapa metrik evaluasi yang umum digunakan meliputi:

A. Akurasi (*Accuracy*)

Proporsi dari prediksi yang benar terhadap total prediksi yang dibuat. Akurasi adalah metrik dasar yang mengukur seberapa baik model mengklasifikasikan data secara keseluruhan.

B. Precision dan Recall

Precision adalah proporsi dari prediksi positif yang benar-benar positif, sedangkan recall adalah proporsi dari kasus positif yang terprediksi dengan benar. Precision dan recall sangat penting ketika ada ketidakseimbangan kelas dalam data (Manning et al., 2009).

C. F1-score

F1-score adalah rata-rata harmonis dari precision dan recall, memberikan satu metrik yang menggambarkan keseimbangan antara keduanya. Ini berguna dalam situasi di mana penting untuk mempertimbangkan baik false positives maupun false negatives (Sasaki, 2015).

6. SMOTE (Synthetic Minority Over-sampling Technique)

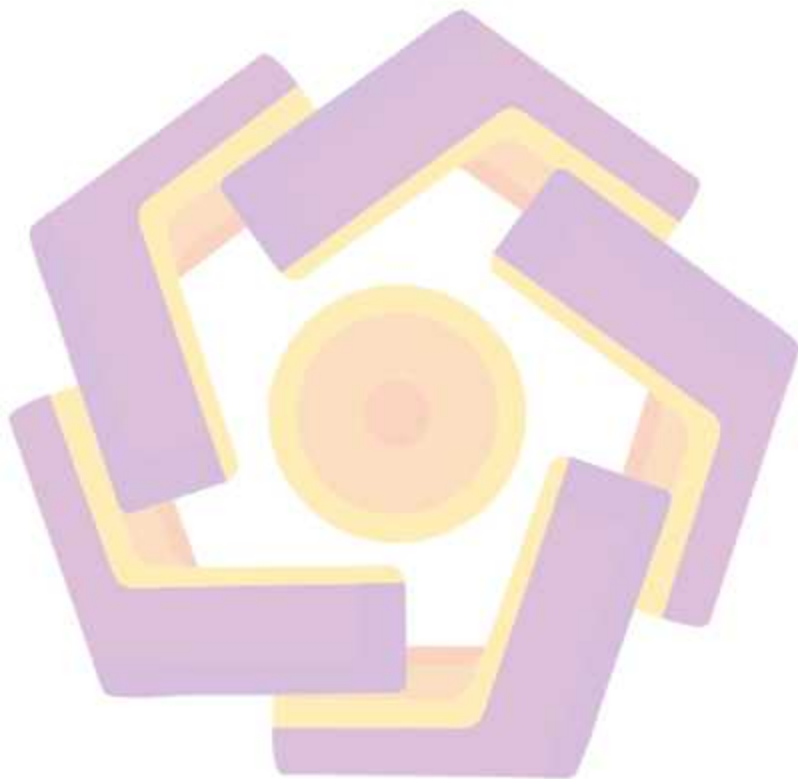
Dalam banyak aplikasi machine learning di dunia nyata, sering kali ditemui masalah ketidakseimbangan kelas (imbalanced data), di mana jumlah sampel pada satu kelas (kelas mayoritas) jauh lebih banyak

daripada kelas lainnya (kelas minoritas). Kondisi ini dapat menyebabkan model prediktif menjadi bias, cenderung mengabaikan kelas minoritas dan menghasilkan performa yang buruk pada kelas tersebut, meskipun akurasi keseluruhannya mungkin terlihat tinggi. Untuk mengatasi masalah ini, berbagai teknik resampling telah dikembangkan, salah satunya adalah SMOTE (Synthetic Minority Over-sampling Technique), yang diperkenalkan (Chawla et al., 2002) pada tahun 2002. SMOTE adalah metode oversampling yang bertujuan untuk menyeimbangkan distribusi kelas dengan cara menambah jumlah sampel pada kelas minoritas. Berbeda dengan teknik oversampling sederhana yang hanya menduplikasi sampel minoritas secara acak (yang berisiko menyebabkan overfitting), SMOTE bekerja dengan cara membuat sampel sintetis atau buatan (synthetic samples).

7. Analisis T-test

Setelah melatih dan mengevaluasi beberapa model machine learning, sering kali perlu untuk menentukan apakah perbedaan performa yang diamati (misalnya, perbedaan akurasi) antara dua model bersifat signifikan secara statistik atau hanya karena variasi acak. Uji-t (t-test) adalah salah satu uji statistik parametrik yang paling umum digunakan untuk tujuan ini. Uji-t berfungsi untuk membandingkan nilai rata-rata (mean) dari dua kelompok untuk menentukan apakah ada perbedaan yang nyata di antara keduanya. Dalam konteks evaluasi model machine learning, jenis uji-t yang paling relevan adalah uji-t berpasangan (paired samples t-test). Uji

ini digunakan ketika ada hubungan atau pasangan alami antara pengamatan di kedua kelompok, seperti saat membandingkan performa dua model pada set data lipatan yang sama dalam k-fold cross-validation (Kim, 2013).



BAB 3

METODE PENELITIAN

3.1 Jenis, Sifat, dan Pendekatan Penelitian

Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimen komputasi berbasis supervised learning, yang bertujuan untuk membangun dan menguji model klasifikasi sentimen terhadap data ulasan pelanggan aplikasi Livin' by Mandiri. Proses penelitian dilakukan dalam beberapa tahap sistematis, dimulai dari pengumpulan dataset publik, pelatihan model, hyperparameter tuning, hingga penerapan model ke dataset ulasan aplikasi Livin' by Mandiri.

Penelitian ini menggunakan dua sumber dataset, yaitu dataset publik berbahasa Indonesia yang telah memiliki label sentimen dan dataset hasil scraping dari Google Play Store aplikasi Livin' by Mandiri yang tidak memiliki label sentimen. Proses pelatihan model dilakukan sepenuhnya menggunakan dataset publik, sedangkan dataset Livin' by Mandiri hanya digunakan sebagai data uji (testing) untuk memperoleh hasil prediksi sentimen. Dataset publik yang digunakan merupakan dataset analisis sentimen berbahasa Indonesia yang juga telah digunakan dalam penelitian terdahulu, sehingga memungkinkan validasi eksternal yang lebih kuat terhadap performa model.

Desain penelitian ini juga menerapkan hyperparameter tuning terhadap masing-masing model klasifikasi untuk memperoleh konfigurasi parameter

terbaik yang mampu meningkatkan performa model pada dataset publik, sebelum model diterapkan pada dataset *Livin' by Mandiri*.

3.2 Metode Pengumpulan Data

Tahap pengumpulan data merupakan langkah awal yang dilakukan dalam penelitian ini, dengan tujuan untuk memperoleh data ulasan yang akan digunakan dalam proses pelatihan, pengujian, serta analisis klasifikasi sentimen. Data yang digunakan dalam penelitian ini berasal dari dua sumber berbeda, yaitu dataset publik dan dataset hasil scraping dari Google Play Store. Adapun deskripsi masing-masing sumber data dijelaskan sebagai berikut:

1. Dataset Publik

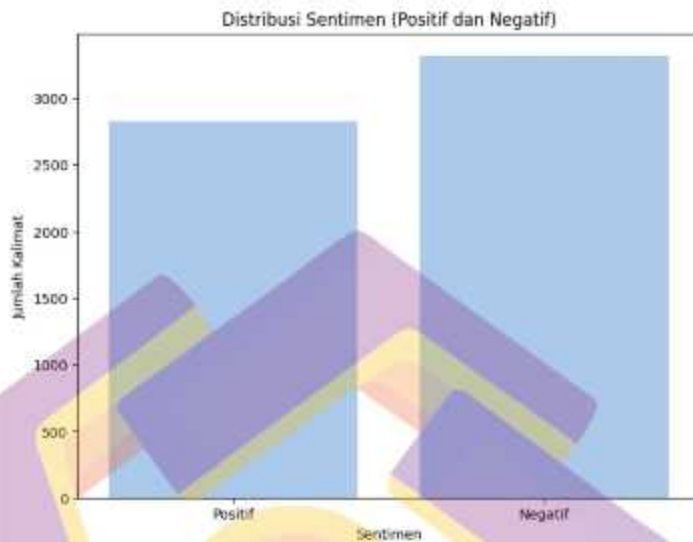
Dataset publik yang digunakan dalam penelitian ini diperoleh melalui platform Kaggle, dengan judul "Corpus Bahasa Indonesia Label" yang dapat diakses pada tautan <https://www.kaggle.com/datasets/achmadshidiq/corpus-bahasa-indonesia-label>. Dataset ini berisi kumpulan kalimat atau frasa berbahasa Indonesia yang telah diberikan label sentimen. Data yang digunakan adalah data yang berlabel positif dan negatif, sedangkan data berlabel netral tidak digunakan pada penelitian ini.

Dataset ini digunakan sebagai dasar pelatihan model klasifikasi sentimen karena memiliki keunggulan label sentimen yang telah divalidasi, bahasa yang sesuai dengan konteks pengguna aplikasi di Indonesia, dan Format dataset yang terstruktur sehingga memudahkan proses preprocessing dan pelatihan model.

Tabel 3.1 Sampel Data Corpus Bahasa Indonesia Label.

kallmat	sentiment
Manusia membutuhkan makanan dan air supaya menjadi kuat dan sehat, begitu juga tanaman	Positif
Tanaman yang tidak mendapat air akan layu dan menjadi kering	Positif
Keberhasilan belajar murid tidak hanya bergantung pada jumlah dan mutu guru yang memadai	Positif
Namun, di sekolah yang jumlah dan mutu gurunya kurang, diduga potensi belajar murid akan sulit mencapai hasil maksimal	Negatif
Secara umum terdapat kecenderungan bahwa murid yang memperoleh nilai rendah berasal dari sekolah yang gurunya lebih banyak absen	Negatif
Dalam kaitan itu, banyak orang tua murid tidak mempunyai cukup kemampuan untuk menjelaskan proses penalaran tersebut	Negatif
tetapi sampai batas tertentu, yaitu dengan rasio 25:1,	Positif
kenaikan angka rasionya mulai berkorelasi negatif terhadap nilai tes murid	Negatif
Sekolah yang memiliki lapangan bermain (olahraga) berkorelasi negatif dengan hasil tes murid	Negatif
Persoalan ini perlu dipelajari lebih lanjut, karena studi menunjukkan efek positif dari aktivitas fisik pada anak-anak	Positif

Jumlah data yang tersedia pada dataset publik ini sebanyak 6.145 kalimat dengan komposisi distribusi masing-masing sentimen positif sebanyak 2.826 kalimat dan sentimen negatif sebanyak 3.319 kalimat.



Gambar 3.1 Grafik Sebaran Sentimen

2. Dataset Ulasan Aplikasi Livin' by Mandiri

Selain menggunakan dataset publik, penelitian ini juga memanfaatkan data ulasan aplikasi Livin' by Mandiri yang diperoleh melalui proses web scraping dari halaman aplikasi di Google Play Store. Proses pengambilan data dilakukan menggunakan library Google Play Scraper berbasis Python, yang mampu mengambil data ulasan pelanggan secara otomatis. Data yang diambil meliputi "at," "content," "score," "thumbUpCount," dan "appVersion."

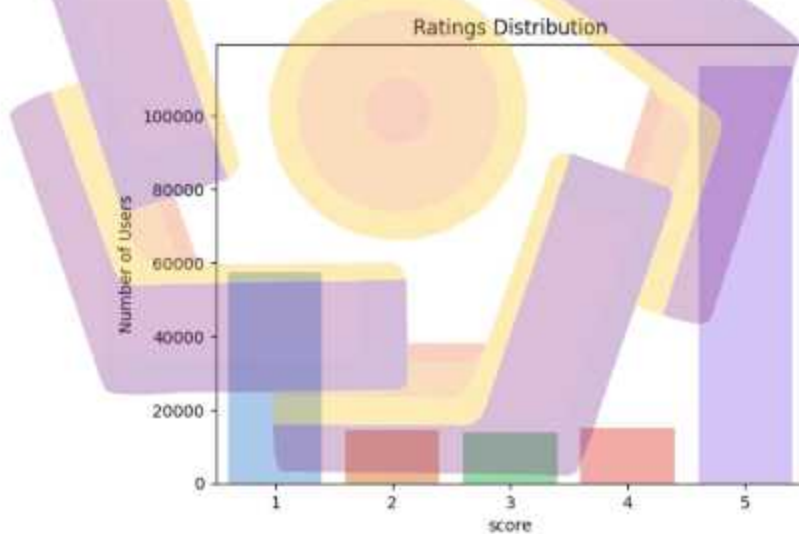
Tabel 3.2 Sampel Data Ulasan Hasil Scrapping.

at	content	score	thumbUp Count	App Version
17/12/2023 06:58	Sangat membantu	5	0	

17/12/2023 06:54	Susah dipakenya, setiap masuk selalu ada di halaman Selamat Datang, dan tdk ada tombol buat login or skip.. Caranya gmn kl mau login? Pas perlu mo pake, jd dipersulit kek gini ðŸ™€	2	0	1.7.0
17/12/2023 06:53	kenapa kalo tiap mau buka aplikasinya, harus di update terus kalo ga di update aplikasi ga bisa kebuka walau cuma intip saldo doang...dr pertama menggunakan livin mandiri setahun yg lalu...entah sdh brp kali diupdate terus utk bisa dipake, memakan memory yg tdk sedikit jgðŸ™€	3	0	
17/12/2023 06:48	Kenapa sudah beberapa hari ini saya mau upgrate livin saya...tpi malah status tertunda...jdinya pengen lakukan transaksi jd terkendali.	3	0	
17/12/2023 06:47	Aplikasi ini kok sulit diupdate yaa,,	1	0	1.6.0
17/12/2023 06:47	Baru di instal sudah minta updat	2	0	
17/12/2023 06:45	Bener2 aplikasi sampah. Sudah 3 hari nglakuin segala cara buat ngakalin transaksi mulai dari restart hp, ubah time scr otomatis, sampe reinstall aplikasi malah jd gak bisa masuk ke aplikasinya. Uninstall saja aplikasi sampah ini. Gak berguna, diupgrade ke versi terbaru malah jd jelek. Gimana dah developernya.	1	0	1.7.0
17/12/2023 06:45	terlalu sering update namun kualitas semakin amburadul	1	0	1.6.0

17/12/2023 06:40	Pemeliharaan rutin di saat perlu. nya? Ayolalh...sedari siang kemarin untuk bisa login perlu menunggu lamaaaa sekali. Banyak gangguannya	3	0	
17/12/2023 06:38	Livin sering lemot kalo transaksi ga loding	1	0	1.7.0

Selama proses scraping, data yang berhasil dikumpulkan berjumlah 214.612 ulasan pelanggan dalam rentang waktu 30 September 2021 hingga 17 Desember 2023. Selanjutnya, data hasil scraping disimpan dalam format CSV untuk memudahkan proses preprocessing dan analisis.



Gambar 3.2 Distribusi Rating Ulasan Pengguna.

Gambar di atas menunjukkan distribusi jumlah ulasan pengguna berdasarkan rating yang diberikan terhadap aplikasi Livin' by Mandiri

pada platform Google Play Store. Rating yang ditampilkan berkisar dari 1 hingga 5, di mana angka 1 merepresentasikan ulasan sangat negatif, dan angka 5 merepresentasikan ulasan sangat positif.

Berdasarkan gambar tersebut, dapat diketahui bahwa rating 5 merupakan kategori yang paling dominan, dengan jumlah ulasan mencapai 113.784 ulasan, atau sekitar 53.02% dari total data. Hal ini menunjukkan bahwa mayoritas pengguna merasa sangat puas terhadap layanan aplikasi tersebut. Rating 1 menempati posisi kedua terbanyak, dengan jumlah ulasan sekitar 57.342 ulasan, menunjukkan adanya cukup banyak keluhan atau ketidakpuasan dari pengguna. Sedangkan rating 2, 3, dan 4 memiliki jumlah yang relatif seimbang, masing-masing berjumlah 14.617, 13.914, dan 14.955 ulasan, jauh di bawah jumlah ulasan pada rating 1 dan 5.

3. Pelabelan Dataset Livin' by Mandiri

Dataset ulasan aplikasi Livin' by Mandiri yang diperoleh dari proses scraping tidak memiliki label sentimen eksplisit. Untuk memungkinkan pelatihan dan evaluasi model, dilakukan proses pelabelan otomatis berdasarkan rating bintang yang diberikan pengguna. Dalam penelitian ini, ulasan dengan rating bintang 5 diberi label 'positif', sementara ulasan dengan rating bintang 1 diberi label 'negatif'. Rating 2, 3, dan 4 dihapus karena dianggap ambigu secara polaritas, rating 3 cenderung netral, sementara rating 2 dan 4 tidak memiliki kecenderungan sentimen yang kuat, sehingga dapat menurunkan kejelasan label dan menyebabkan kebingungan model dalam proses pembelajaran. Pendekatan ini sesuai

dengan praktik umum analisis sentimen di media sosial, di mana data ekstrem (sangat puas dan sangat tidak puas) lebih representatif untuk pelatihan model klasifikasi biner. Berdasarkan proses tersebut, dari total 214.612 ulasan, diperoleh 113.784 ulasan berlabel positif dan 57.342 ulasan berlabel negatif.

3.3 Metode Analisis Data

1. Preprocessing Data

Seluruh data ulasan yang diperoleh, baik dari dataset publik maupun dataset *Living by Mandiri*, terlebih dahulu melalui proses preprocessing untuk memastikan data dalam kondisi bersih dan siap untuk diproses oleh algoritma machine learning. Tahapan preprocessing yang dilakukan meliputi:

- a) Pembersihan kata untuk menghapus angka, simbol, tanda baca, dan karakter khusus serta mengubah seluruh teks ulasan menjadi huruf kecil.
- b) Tokenizing, yakni memecah teks ulasan menjadi kata-kata tunggal.
- c) Stopword Removal, yaitu menghapus kata-kata umum yang tidak memiliki makna sentimen, menggunakan daftar stopwords Bahasa Indonesia.
- d) Stemming dan Lemmatization, untuk mengembalikan kata ke bentuk dasarnya, menggunakan library Sastrawi dan Stanza.

2. Pembobotan Kata

Setelah data teks ulasan melalui proses preprocessing, tahap selanjutnya dalam penelitian ini adalah pembobotan kata (word embedding). Tahapan ini bertujuan untuk mengubah data teks hasil preprocessing yang berbentuk string menjadi representasi numerik berbasis vektor, sehingga dapat diproses oleh algoritma machine learning dan deep learning.

Pembobotan kata penting dilakukan karena data teks bersifat tidak terstruktur dan tidak dapat langsung diproses oleh model klasifikasi. Representasi numerik hasil pembobotan kata memungkinkan model untuk mengenali pola makna antar kata dalam ruang vektor multidimensi.

Pada penelitian ini, metode pembobotan kata yang digunakan adalah FastText embedding. FastText dipilih karena memiliki kemampuan untuk mempertimbangkan representasi kata berdasarkan subword (karakter n-gram), sehingga lebih efektif dalam memproses data teks berbahasa Indonesia yang memiliki banyak bentuk imbuhan dan kata turunan.

Adapun tahapan pembobotan kata dalam penelitian ini adalah sebagai berikut:

- a) Model FastText terlebih dahulu dilatih menggunakan dataset publik yang telah melalui tahap preprocessing. Proses ini menghasilkan model embedding yang mampu menghasilkan vektor representasi untuk setiap kata yang terdapat dalam dataset.
- b) Setelah model FastText terbentuk, seluruh data ulasan, baik dari dataset publik maupun dataset ulasan aplikasi Livin' by Mandiri,

diproses menggunakan model embedding yang sama. Setiap teks ulasan diubah menjadi sentence vector dengan cara menghitung rata-rata vektor dari seluruh kata dalam satu ulasan.

- c) Seluruh data teks ulasan diubah menjadi vektor berdimensi tetap sesuai hasil model FastText yang telah ditentukan pada saat pelatihan. Hal ini bertujuan agar seluruh data memiliki format fitur yang seragam saat dimasukkan ke dalam algoritma klasifikasi.
- d) Vektor representasi hasil pembobotan kata selanjutnya disimpan dalam format array, yang kemudian digunakan sebagai input dalam proses pelatihan model klasifikasi sentimen maupun prediksi sentimen terhadap dataset Livin' by Mandiri.

3. Pembagian Data dan Parameter Awal Model

Untuk setiap skenario pelatihan, dataset publik dan dataset Livin' by Mandiri dibagi menjadi data latih dan data uji dengan rasio 80:20. Pembagian ini dilakukan secara acak namun terstratifikasi untuk memastikan proporsi kelas sentimen pada data latih dan data uji tetap sama dengan proporsi pada dataset asli. Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk evaluasi akhir setelah model selesai dilatih.

Pelatihan awal dilakukan menggunakan parameter default dari library scikit-learn dan Keras untuk menetapkan baseline performa. Parameter yang digunakan untuk setiap model sebelum proses tuning diperlihatkan pada Tabel 4.

Tabel 3.3 Parameter Awal Algoritma Sebelum Hyperparameter Tuning.

Model	Library	Parameter
SVM	Scikit-learn	kernel = 'linear', C = 1.0
ANN	Scikit-learn	hidden_layer_sizes = (100,), activation = 'relu', solver = 'adam', max_iter = 200
LSTM	Keras	LSTM units = 128, Optimizer = 'adam', Loss = 'binary_crossentropy', Dropout = 0.2

4. Penanganan Ketidakseimbangan Data

Setelah proses preprocessing data, jumlah data dengan sentimen positif dan sentiment positif tidak seimbang. Kondisi ini berisiko menyebabkan model klasifikasi menjadi bias, di mana model cenderung lebih akurat dalam memprediksi kelas mayoritas dan memiliki performa yang buruk pada kelas minoritas.

Untuk mengatasi masalah ini, penelitian ini menerapkan teknik oversampling yang disebut SMOTE (Synthetic Minority Over-sampling Technique). SMOTE adalah metode yang populer untuk menyeimbangkan distribusi kelas dengan cara membuat sampel sintetis untuk kelas minoritas, alih-alih hanya menduplikasi data yang sudah ada (Chawla et al., 2002). Pendekatan ini membantu mencegah overfitting dan memungkinkan model untuk mempelajari batas keputusan yang lebih baik.

Dalam penelitian ini, SMOTE diterapkan hanya pada data latih setelah proses pembagian data. Hal ini merupakan praktik penting untuk

mencegah kebocoran data, di mana informasi dari data uji secara tidak sengaja memengaruhi proses pelatihan. Dengan menerapkan SMOTE, data latih menjadi seimbang, sehingga ketiga model dapat mempelajari karakteristik dari kedua kelas sentimen secara lebih efektif dan menghasilkan model klasifikasi yang lebih adil dan tidak bias.

5. Pelatihan Model

Setelah data dari dataset publik selesai diproses, tahap selanjutnya adalah pelatihan model klasifikasi sentimen. Tiga algoritma machine learning dan deep learning yang digunakan dalam penelitian ini meliputi:

- a) Support Vector Machine (SVM)
- b) Artificial Neural Network (ANN)
- c) Long Short-Term Memory (LSTM)

Proses pelatihan dilakukan menggunakan dataset publik Corpus Bahasa Indonesia Label hasil preprocessing dan vectorizing. Setiap model dilatih untuk mengenali pola sentimen positif dan negatif berdasarkan fitur hasil FastText embedding.

6. Hyperparameter Tuning

Untuk memperoleh performa optimal dari masing-masing model, dilakukan proses hyperparameter tuning. Proses ini bertujuan mencari kombinasi parameter terbaik yang dapat meningkatkan performa klasifikasi sentimen pada dataset publik. Pada model SVM, tuning dilakukan menggunakan metode GridSearchCV dengan parameter C, gamma, dan kernel. Sementara Untuk model ANN dan LSTM, tuning

dilakukan menggunakan metode exploratory trial-based tuning untuk menentukan nilai optimal pada parameter seperti embedding dimension, jumlah units, dan learning rate.

Hasil tuning digunakan untuk melakukan pelatihan ulang ketiga model agar memperoleh performa terbaik sebelum diterapkan ke dataset Livin' by Mandiri.

Tabel 3.4 Parameter Algoritma Setelah Hyperparameter Tuning.

Model	Library	Parameter
SVM	Scikit-learn	C : 0.1, kernel : linear
ANN	Scikit-learn	activation : tanh, hidden_layer_sizes : (50,), max_iter : 200, solver : sgd
LSTM	Keras	model_lstm_units : 128, model_dense_units : 64, model_dropout_rate : 0.5, batch_size : 16, epochs : 20

7. Penerapan Model ke Dataset Livin' by Mandiri

Model-model hasil pelatihan dan hyperparameter tuning kemudian diterapkan untuk memprediksi sentimen terhadap dataset ulasan aplikasi Livin' by Mandiri. Sebelumnya, dataset ulasan ini diproses menggunakan tahapan preprocessing yang sama dengan dataset publik, agar fitur yang dihasilkan memiliki kesesuaian format dan dimensi.

Data hasil preprocessing kemudian diubah menjadi vektor menggunakan model FastText yang telah dilatih sebelumnya. Hasil vektorisasi tersebut digunakan sebagai input ke model SVM, ANN, dan LSTM untuk menghasilkan prediksi sentimen.

8. Pelatihan Model Menggunakan Dataset Livin' by Mandiri

Pada tahap ini, dilakukan pelatihan model klasifikasi sentimen menggunakan dataset ulasan aplikasi Livin' by Mandiri yang telah diberi label otomatis berdasarkan rating pengguna, di mana ulasan dengan rating bintang 1 dikategorikan sebagai sentimen negatif dan bintang 5 sebagai sentimen positif. Dataset yang telah melalui proses pra-proses dan representasi vektor menggunakan FastText kemudian dibagi menjadi data latih dan data uji menggunakan teknik *train-test split* dengan rasio 80:20 untuk memastikan evaluasi yang adil terhadap model. Tiga algoritma pembelajaran yang digunakan dalam pelatihan ini adalah SVM, ANN, LSTM.

Untuk model SVM, digunakan pendekatan klasifikasi linier dengan parameter default pada pustaka *scikit-learn*. Model ANN dibangun menggunakan *MLPClassifier* dari *sklearn.neural network* dengan arsitektur yang terdiri dari satu lapisan tersembunyi berisi 100 neuron, fungsi aktivasi ReLU, dan solver *stochastic gradient descent (SGD)*. Sementara itu, model LSTM dibangun menggunakan *TensorFlow/Keras* dalam bentuk *Sequential Model* yang terdiri dari satu lapisan LSTM dengan 128 unit, diikuti oleh lapisan *Dense* dan *Dropout* untuk mencegah *overfitting*. Input vektor kalimat hasil *FastText* diubah ke dalam bentuk tiga dimensi agar sesuai dengan kebutuhan input jaringan LSTM.

Setelah pelatihan, ketiga model dievaluasi menggunakan data uji untuk menghitung metrik performa seperti akurasi, presisi, recall, dan *F1-score*. Evaluasi dilakukan dengan membandingkan label prediksi dan label

sebenarnya, dan digunakan pula *classification report* untuk memperoleh pemahaman yang lebih mendalam terhadap performa model dalam mengklasifikasikan sentimen positif dan negatif. Seluruh proses pelatihan dan evaluasi dilakukan menggunakan bahasa pemrograman Python dan pustaka-pustaka pendukung seperti *scikit-learn*, TensorFlow, NumPy, dan Pandas.

9. Penerapan Model Livin' by Mandiri pada Dataset Publik

Setelah model klasifikasi sentimen dilatih menggunakan dataset Livin' by Mandiri yang telah dilabeli otomatis berdasarkan rating bintang, langkah selanjutnya adalah menerapkan model tersebut pada dataset publik untuk menguji kemampuan generalisasi model antar domain data.

Ketiga model terlatih, yaitu Support Vector Machine (SVM), Artificial Neural Network (ANN), dan Long Short-Term Memory (LSTM), dimuat dari penyimpanan permanen dan digunakan untuk memprediksi label sentimen pada setiap ulasan dalam dataset publik. Karena model dilatih dengan format label {0, 1} dan dataset publik memiliki label {1, 2}, dilakukan penyesuaian konversi label untuk memastikan kesesuaian dalam evaluasi. Evaluasi performa dilakukan dengan membandingkan hasil prediksi terhadap label sebenarnya menggunakan metrik akurasi, presisi, recall, dan F1-score, serta ditampilkan dalam bentuk *classification report* untuk memberikan gambaran menyeluruh atas kinerja model lintas domain.

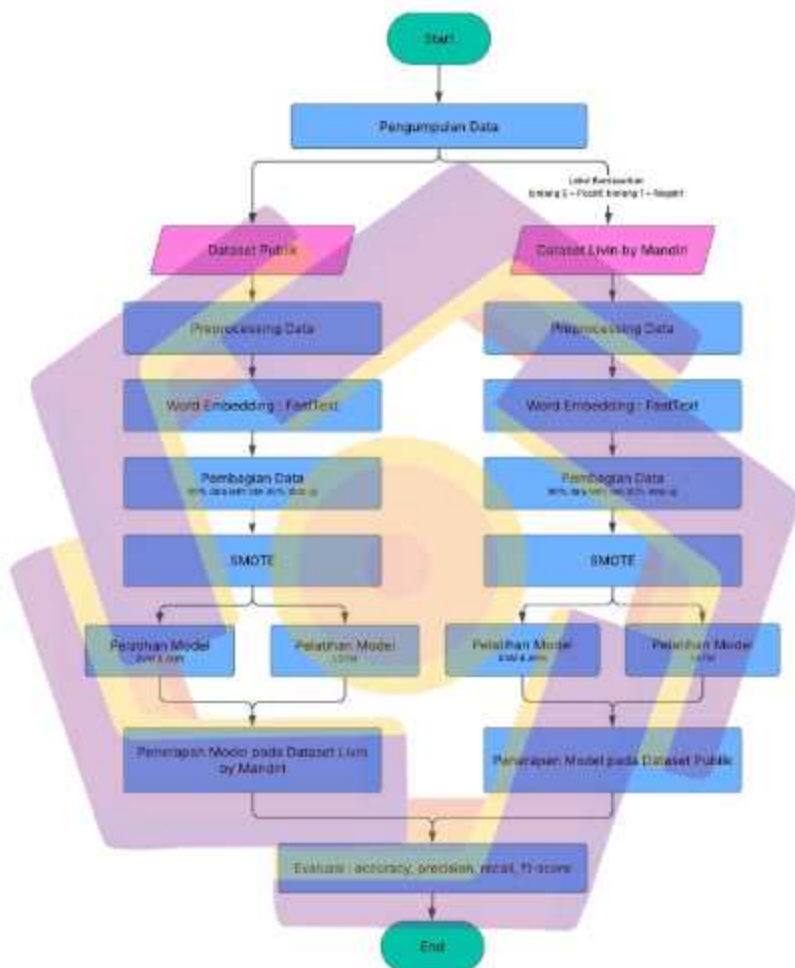
Langkah ini bertujuan untuk menilai sejauh mana model yang dilatih dalam domain aplikasi finansial (Livin' by Mandiri) dapat diaplikasikan pada data beragam dari domain umum. Hasil dari proses ini memberikan informasi penting mengenai kemampuan adaptasi model terhadap distribusi data yang berbeda serta potensi penggunaan model dalam skenario real-world yang lebih luas.

10. Evaluasi Statistik

Selain menggunakan metrik standar seperti akurasi, presisi, recall, dan F1-score, perbandingan performa antar model juga perlu divalidasi secara statistik untuk menentukan apakah perbedaan skor yang diamati signifikan secara statistik atau hanya karena kebetulan. Meskipun dalam penelitian ini eksperimen dilakukan satu kali, pendekatan yang lebih robas adalah melakukan eksperimen berulang dengan pembagian data acak yang berbeda untuk setiap iterasi (Dietterich, 1998).

Setelah mendapatkan serangkaian skor akurasi dari eksperimen berulang, uji hipotesis statistik seperti uji-t berpasangan dapat digunakan untuk membandingkan dua model. Uji ini akan menentukan apakah rata-rata skor akurasi dari model A secara signifikan berbeda dari model B. Hasil uji ini akan memberikan keyakinan yang lebih kuat dalam menyimpulkan model mana yang superior. Implementasi evaluasi statistik ini disarankan untuk pengembangan penelitian di masa depan.

3.4 Alur Penelitian



Gambar 3.3 Alur Penelitian

BAB 4 HASIL PENELITIAN DAN PEMBAHASAN

4.1 Deskripsi Penelitian

Penelitian ini dilakukan untuk membangun dan menguji model klasifikasi sentimen berbasis machine learning terhadap data ulasan aplikasi Livin' by Mandiri. Proses penelitian dilaksanakan secara bertahap dan sistematis, dimulai dari pengumpulan data, preprocessing, pembobotan kata, pelatihan model, hyperparameter tuning, hingga penerapan model terhadap dataset aplikasi Livin' by Mandiri.

Adapun data yang digunakan dalam penelitian ini terdiri dari dua sumber utama. Dataset pertama berasal dari dataset publik berbahasa Indonesia dengan label sentimen yang diperoleh melalui platform Kaggle dengan judul "Corpus Bahasa Indonesia Label" pada tautan <https://www.kaggle.com/datasets/achmadshidiq/corpus-bahasa-indonesia-label>. Dataset ini digunakan sebagai dasar pelatihan model klasifikasi sentimen.

Dataset kedua diperoleh melalui proses scraping dari halaman aplikasi Livin' by Mandiri di Google Play Store menggunakan library Google Play Scraper. Data yang diambil berupa teks ulasan pelanggan beserta nilai rating. Namun demikian, dataset ini tidak memiliki label sentimen, sehingga hanya digunakan sebagai data uji untuk memperoleh prediksi sentimen dari model yang telah dilatih.

Seluruh data ulasan, baik dari dataset publik maupun hasil scraping, terlebih dahulu diproses menggunakan tahapan preprocessing teks, yang meliputi normalisasi huruf kecil, pembersihan karakter non-alfabet, tokenisasi, stopwords removal, stemming, dan lemmatization. Setelah itu, dilakukan tahap pembobotan kata menggunakan FastText embedding untuk mengubah data teks menjadi representasi vektor berdimensi tetap.

Model klasifikasi yang digunakan dalam penelitian ini meliputi Support Vector Machine (SVM), Artificial Neural Network (ANN), dan Long Short-Term Memory (LSTM). Ketiga model dilatih menggunakan dataset publik, kemudian dilakukan hyperparameter tuning untuk memperoleh konfigurasi parameter terbaik pada masing-masing model. Model hasil tuning selanjutnya diterapkan untuk memprediksi sentimen pada dataset ulasan aplikasi Livin' by Mandiri.

Hasil prediksi sentimen dari masing-masing model dianalisis melalui distribusi jumlah prediksi kategori positif dan negatif. Analisis dilakukan untuk mengetahui pola kecenderungan prediksi antar model serta sebaran sentimen pada dataset ulasan aplikasi Livin' by Mandiri. Karena dataset hasil scraping tidak memiliki label sentimen, maka evaluasi performa model hanya dilakukan pada dataset publik, sedangkan pada dataset Livin' by Mandiri analisis dibatasi pada deskripsi hasil distribusi prediksi sentiment.

4.2 Preprocessing Data

Setelah proses pengumpulan data selesai dilakukan, tahap selanjutnya adalah preprocessing data. Preprocessing merupakan langkah penting untuk

memastikan bahwa data teks yang diperoleh siap digunakan dalam proses ekstraksi fitur dan pelatihan model machine learning. Dalam penelitian ini, preprocessing dilakukan secara bertahap melalui beberapa teknik berikut:

1. Pembersihan Data

Pada tahap ini, data teks ulasan dibersihkan dari karakter-karakter yang tidak diperlukan, seperti tanda baca, angka, emoji, URL, dan simbol-simbol khusus. Selain itu, dilakukan konversi seluruh teks menjadi huruf kecil (lowercase) untuk memastikan konsistensi data. Hasil dari tahap ini diperlihatkan pada tabel di bawah.

Tabel 4.1 Sampel Hasil Pembersihan Data Ulasan Livin By Mandiri.

content	data_cleaned
Tolong timbulkan kan donk ujakan pinjaman	tolong timbulkan kan donk ujakan pinjaman
Izin kan saya mendonlowd	izin kan saya mendonlowd
Mantap pokoknya semenjak saya pake lifin saya jadi pahlawan power rangers	mantap pokoknya semenjak saya pake lifin saya jadi pahlawan power rangers
update terus	update terus
Livin' by mandiri ok	livin by mandiri ok
Mengecewakan	mengecewakan
Terlalu banyak update,malah bingung	terlalu banyak update malah bingung
Tanpa solusi, akhirnya tutup rekening	tanpa solusi akhirnya tutup rekening
Aplikasi saya malah nggak bisa dibuka karna update yang sekarang , kalau yang dulu malah enak buka nya .	aplikasi saya malah nggak bisa dibuka karna update yang sekarang kalau yang dulu malah enak buka nya
Selalu gagal di verifikasi wajah, padahal sudah memenuhi syarat pencahay aan	selalu gagal di verifikasi wajah padahal sudah memenuhi syarat pencahay aan

Tabel 4.2 Sampel Hasil Pembersihan Data Publik.

kallmat	sentiment	data_cleaned
Manusia membutuhkan makanan dan air supaya menjadi kuat dan sehat, begitu juga tanaman	Positif	manusia membutuhkan makanan dan air supaya menjadi kuat dan sehat begitu juga tanaman
Tanaman yang tidak mendapat air akan layu dan menjadi kering	Positif	tanaman yang tidak mendapat air akan layu dan menjadi kering
Keberhasilan belajar murid tidak hanya bergantung pada jumlah dan mutu guru yang memadai	Positif	keberhasilan belajar murid tidak hanya bergantung pada jumlah dan mutu guru yang memadai
Namun, di sekolah yang jumlah dan mutu gurunya kurang, diduga potensi belajar murid akan sulit mencapai hasil maksimal	Negatif	namun di sekolah yang jumlah dan mutu gurunya kurang diduga potensi belajar murid akan sulit mencapai hasil maksimal
Secara umum terdapat kecenderungan bahwa murid yang memperoleh nilai rendah berasal dari sekolah yang gurunya lebih banyak absen	Negatif	secara umum terdapat kecenderungan bahwa murid yang memperoleh nilai rendah berasal dari sekolah yang gurunya lebih banyak absen
Dalam kaitan itu, banyak orang tua murid tidak mempunyai cukup kemampuan untuk menjelaskan proses penalaran tersebut	Negatif	dalam kaitan itu banyak orang tua murid tidak mempunyai cukup kemampuan untuk menjelaskan proses penalaran tersebut
tetapi sampai batas tertentu, yaitu dengan rasio 25:1,	Positif	tetapi sampai batas tertentu yaitu dengan rasio
kenaikan angka rasionya mulai berkorelasi negatif terhadap nilai tes murid	Negatif	kenaikan angka rasionya mulai berkorelasi negatif terhadap nilai tes murid
Sekolah yang memiliki lapangan bermain (olahraga) berkorelasi negatif dengan hasil tes	Negatif	sekolah yang memiliki lapangan bermain olahraga berkorelasi negatif dengan hasil tes murid

murid Persoalan ini perlu dipelajari lebih lanjut, karena studi menunjukkan efek positif dari aktivitas fisik pada anak-anak	Positif	persoalan ini perlu dipelajari lebih lanjut karena studi menunjukkan efek positif dari aktivitas fisik pada anak anak
---	---------	---

2. Tokenisasi

Hasil pembersihan data kemudian dipecah menjadi potongan kata (token) menggunakan fungsi `word_tokenize`. Proses ini penting untuk memudahkan analisis dan manipulasi kata per kata pada tahap berikutnya.

Tabel 4.3 Sampel Hasil Tokenisasi Data Ulasan Livin By Mandiri.

data_cleaned	tokenized
tolong timbulkan kan donk ujakan pinjaman	['tolong', 'timbulkan', 'kan', 'donk', 'ujakan', 'pinjaman']
izin kan saya mendoanlowd	['izin', 'kan', 'saya', 'mendonlowd']
mantap pokoknya semenjak saya pake lifin saya jadi pahlawan power rangers	['mantap', 'pokoknya', 'semenjak', 'saya', 'pake', 'lifin', 'saya', 'jadi', 'pahlawan', 'power', 'rangers']
update terus	['update', 'terus']
livin by mandiri ok	['livin', 'by', 'mandiri', 'ok']
mengecewakan	['mengecewakan']
terlalu banyak update malah bingung	['terlalu', 'banyak', 'update', 'malah', 'bingung']
tanpa solusi akhirnya tutup rekening	['tanpa', 'solusi', 'akhirnya', 'tutup', 'rekening']
aplikasi saya malah nggak bisa dibuka karna update yang sekarang kalau yang dulu malah enak buka nya	['aplikasi', 'saya', 'malah', 'nggak', 'bisa', 'dibuka', 'karna', 'update', 'yang', 'sekarang', 'kalau', 'yang', 'dulu', 'malah', 'enak', 'buka', 'nya']
selalu gagal di verifikasi wajah padahal sudah memenuhi syarat pencahay aan	['selalu', 'gagal', 'di', 'verifikasi', 'wajah', 'padahal', 'sudah', 'memenuhi', 'syarat', 'pencahay', 'aan']

Tabel 4.4 Sampel Hasil Tokenisasi Data Publik.

data_cleaned	tokenized
manusia membutuhkan makanan dan air supaya menjadi kuat dan sehat begitu juga tanaman	['manusia', 'membutuhkan', 'makanan', 'dan', 'air', 'supaya', 'menjadi', 'kuat', 'dan', 'sehat', 'begitu', 'juga', 'tanaman']
tanaman yang tidak mendapat air akan layu dan menjadi kering	['tanaman', 'yang', 'tidak', 'mendapat', 'air', 'akan', 'layu', 'dan', 'menjadi', 'kering']
keberhasilan belajar murid tidak hanya bergantung pada jumlah dan mutu guru yang memadai	['keberhasilan', 'belajar', 'murid', 'tidak', 'hanya', 'bergantung', 'pada', 'jumlah', 'dan', 'mutu', 'guru', 'yang', 'memadai']
namun di sekolah yang jumlah dan mutu gurunya kurang diduga potensi belajar murid akan sulit mencapai hasil maksimal	['namun', 'di', 'sekolah', 'yang', 'jumlah', 'dan', 'mutu', 'gurunya', 'kurang', 'diduga', 'potensi', 'belajar', 'murid', 'akan', 'sulit', 'mencapai', 'hasil', 'maksimal']
secara umum terdapat kecenderungan bahwa murid yang memperoleh nilai rendah berasal dari sekolah yang gurunya lebih banyak absen dalam kaitan itu banyak orang tua murid tidak mempunyai cukup kemampuan untuk menjelaskan proses penalaran tersebut	['secara', 'umum', 'terdapat', 'kecenderungan', 'bahwa', 'murid', 'yang', 'memperoleh', 'nilai', 'rendah', 'berasal', 'dari', 'sekolah', 'yang', 'gurunya', 'lebih', 'banyak', 'absen'] ['dalam', 'kaitan', 'itu', 'banyak', 'orang', 'tua', 'murid', 'tidak', 'mempunyai', 'cukup', 'kemampuan', 'untuk', 'menjelaskan', 'proses', 'penalaran', 'tersebut']
tetapi sampai batas tertentu yaitu dengan rasio kenaikan angka rasionya mulai berkorelasi negatif terhadap nilai tes murid	['tetapi', 'sampai', 'batas', 'tertentu', 'yaitu', 'dengan', 'rasio'] ['kenaikan', 'angka', 'rasionya', 'mulai', 'berkorelasi', 'negatif', 'terhadap', 'nilai', 'tes', 'murid']
sekolah yang memiliki lapangan bermain olahraga berkorelasi negatif dengan hasil tes murid	['sekolah', 'yang', 'memiliki', 'lapangan', 'bermain', 'olahraga', 'berkorelasi', 'negatif', 'dengan', 'hasil', 'tes', 'murid']
persoalan ini perlu dipelajari lebih lanjut karena studi menunjukkan efek positif dari aktivitas fisik pada anak anak	['persoalan', 'ini', 'perlu', 'dipelajari', 'lebih', 'lanjut', 'karena', 'studi', 'menunjukkan', 'efek', 'positif', 'dari', 'aktivitas', 'fisik', 'pada', 'anak', 'anak']

3. Stopword Removal

Kata-kata umum yang sering muncul namun tidak memiliki makna penting dalam konteks analisis sentiment. Kata-kata seperti "saya", "kan", "terus", "terlalu", "banyak, dst, dihapus menggunakan daftar stopwords Bahasa Indonesia dari library Sastrawi. Penghapusan stopwords ini bertujuan untuk mengurangi noise dan meningkatkan efisiensi pemrosesan data. Hasil tahap ini diperlihatkan pada tabel di bawah.

Tabel 4.5 Sampel Hasil Stopword Removal Data Ulasan Livin By Mandiri.

tokenized	stopword_removal
['tolong', 'timbulkan', 'kan', 'donk', 'ujakan', 'pinjaman']	['tolong', 'timbulkan', 'donk', 'ujakan', 'pinjaman']
['izin', 'kan', 'saya', 'mendonlowd']	['izin', 'mendonlowd']
['mantap', 'pokoknya', 'semenjak', 'saya', 'pake', 'lifin', 'saya', 'jadi', 'pahlawan', 'power', 'rangers']	['mantap', 'pokoknya', 'semenjak', 'pake', 'lifin', 'pahlawan', 'power', 'rangers']
['update', 'terus']	['update']
['livin', 'by', 'mandiri', 'ok']	['livin', 'by', 'mandiri', 'ok']
['mengecewakan']	['mengecewakan']
['terlalu', 'banyak', 'update', 'malah', 'bingung']	['update', 'bingung']
['tanpa', 'solusi', 'akhirnya', 'tutup', 'rekening']	['solusi', 'tutup', 'rekening']
['aplikasi', 'saya', 'malah', 'nggak', 'bisa', 'dibuka', 'karna', 'update', 'yang', 'sekarang', 'kalau', 'yang', 'dulu', 'malah', 'enak', 'buka', 'nya']	['aplikasi', 'nggak', 'dibuka', 'karna', 'update', 'enak', 'buka', 'nya']
['selalu', 'gagal', 'di', 'verifikasi', 'wajah', 'padahal', 'sudah', 'memenuhi', 'syarat', 'pencahay', 'aan']	['gagal', 'verifikasi', 'wajah', 'memenuhi', 'syarat', 'pencahay', 'aan']

Tabel 4.6 Sampel Hasil Stopword Removal Data Publik.

tokenized	stopword
['manusia', 'membutuhkan',	['manusia', 'membutuhkan',

'makanan', 'dan', 'air', 'supaya', 'menjadi', 'kuat', 'dan', 'sehat', 'begitu', 'juga', 'tanaman']	'makanan', 'air', 'kuat', 'sehat', 'tanaman']
['tanaman', 'yang', 'tidak', 'mendapat', 'air', 'akan', 'layu', 'dan', 'menjadi', 'kering']	['tanaman', 'air', 'layu', 'kering']
['keberhasilan', 'belajar', 'murid', 'tidak', 'hanya', 'bergantung', 'pada', 'jumlah', 'dan', 'mutu', 'guru', 'yang', 'memadai']	['keberhasilan', 'belajar', 'murid', 'bergantung', 'mutu', 'guru', 'memadai']
['namun', 'di', 'sekolah', 'yang', 'jumlah', 'dan', 'mutu', 'gurunya', 'kurang', 'diduga', 'potensi', 'belajar', 'murid', 'akan', 'sulit', 'mencapai', 'hasil', 'maksimal']	['sekolah', 'mutu', 'gurunya', 'diduga', 'potensi', 'belajar', 'murid', 'sulit', 'mencapai', 'hasil', 'maksimal']
['secara', 'umum', 'terdapat', 'kecenderungan', 'bahwa', 'murid', 'yang', 'memperoleh', 'nilai', 'rendah', 'berasal', 'dari', 'sekolah', 'yang', 'gurunya', 'lebih', 'banyak', 'absen']	['kecenderungan', 'murid', 'memperoleh', 'nilai', 'rendah', 'berasal', 'sekolah', 'gurunya', 'absen']
['dalam', 'kaitan', 'itu', 'banyak', 'orang', 'tua', 'murid', 'tidak', 'mempunyai', 'cukup', 'kemampuan', 'untuk', 'menjelaskan', 'proses', 'penalaran', 'tersebut']	['kaitan', 'orang', 'tua', 'murid', 'kemampuan', 'proses', 'penalaran']
['tetapi', 'sampai', 'batas', 'tertentu', 'yaitu', 'dengan', 'rasio']	['batas', 'rasio']
['kenaikan', 'angka', 'rasionya', 'mulai', 'berkorelasi', 'negatif', 'terhadap', 'nilai', 'tes', 'murid']	['kenaikan', 'angka', 'rasionya', 'berkorelasi', 'negatif', 'nilai', 'tes', 'murid']
['sekolah', 'yang', 'memiliki', 'lapangan', 'bermain', 'olahraga', 'berkorelasi', 'negatif', 'dengan', 'hasil', 'tes', 'murid']	['sekolah', 'memiliki', 'lapangan', 'bermain', 'olahraga', 'berkorelasi', 'negatif', 'hasil', 'tes', 'murid']
['persoalan', 'ini', 'perlu', 'dipelajari', 'lebih', 'lanjut', 'karena', 'studi', 'menunjukkan', 'efek', 'positif', 'dari', 'aktivitas', 'fisik', 'pada', 'anak', 'anak']	['dipelajari', 'studi', 'efek', 'positif', 'aktivitas', 'fisik', 'anak', 'anak']

4. Stemming

Stemming dilakukan untuk mengembalikan kata ke bentuk dasarnya. Proses ini dilakukan menggunakan library Sastrawi Stemmer. Seperti yang terlihat pada tabel di bawah, kata “timbulkan” dikembalikan menjadi “timbul” dan kata “pinjaman” dikembalikan menjadi “pinjam”.

Tabel 4.7 Sampel Hasil Stemming Data Ulasan Livin By Mandiri.

stopword_removal	stemming
['tolong', 'timbulkan', 'donk', 'ujakan', 'pinjaman']	['tolong', 'timbul', 'donk', 'ujakan', 'pinjam']
['izin', 'mendonlowd']	['izin', 'mendonlowd']
['mantap', 'pokoknya', 'semenjak', 'pake', 'lifin', 'pahlawan', 'power', 'rangers']	['mantap', 'pokok', 'semenjak', 'pake', 'lifin', 'pahlawan', 'power', 'rangers']
['update']	['update']
['livin', 'by', 'mandiri', 'ok']	['livin', 'by', 'mandiri', 'ok']
['mengecewakan']	['kecewa']
['update', 'bingung']	['update', 'bingung']
['solusi', 'tutup', 'rekening']	['solusi', 'tutup', 'rekening']
['aplikasi', 'nggak', 'dibuka', 'karna', 'update', 'enak', 'buka', 'nya']	['aplikasi', 'nggak', 'buka', 'karna', 'update', 'enak', 'buka', 'nya']
['gagal', 'verifikasi', 'wajah', 'memenuhi', 'syarat', 'pencahay', 'aan']	['gagal', 'verifikasi', 'wajah', 'penuh', 'syarat', 'pencahay', 'aan']

Tabel 4.8 Sampel Hasil Stemming Data Publik.

stopword	stemming
['manusia', 'membutuhkan', 'makanan', 'air', 'kuat', 'sehat', 'tanaman']	['manusia', 'butuh', 'makan', 'air', 'kuat', 'sehat', 'tanam']
['tanaman', 'air', 'layu', 'kering']	['tanam', 'air', 'layu', 'kering']
['keberhasilan', 'belajar', 'murid', 'bergantung', 'mutu', 'guru', 'memadai']	['hasil', 'ajar', 'murid', 'gantung', 'mutu', 'guru', 'pada']
['sekolah', 'mutu', 'gurunya', 'diduga', 'potensi', 'belajar', 'murid', 'sulit', 'mencapai', 'hasil', 'maksimal']	['sekolah', 'mutu', 'guru', 'duga', 'potensi', 'ajar', 'murid', 'sulit', 'capai', 'hasil', 'maksimal']
['kecenderungan', 'murid']	['cenderung', 'murid', 'oleh', 'nilai']

'memperoleh', 'nilai', 'rendah', 'berasal', 'sekolah', 'gurunya', 'absen']	'rendah', 'asal', 'sekolah', 'guru', 'absen']
['kaitan', 'orang', 'tua', 'murid', 'kemampuan', 'proses', 'penalaran']	['kait', 'orang', 'tua', 'murid', 'mampu', 'proses', 'nalar']
['batas', 'rasio']	['batas', 'rasio']
['kenaikan', 'angka', 'rasionya', 'berkorelasi', 'negatif', 'nilai', 'tes', 'murid']	['naik', 'angka', 'rasio', 'korelasi', 'negatif', 'nilai', 'tes', 'murid']
['sekolah', 'memiliki', 'lapangan', 'bermain', 'olahraga', 'berkorelasi', 'negatif', 'hasil', 'tes', 'murid']	['sekolah', 'milik', 'lapang', 'main', 'olahraga', 'korelasi', 'negatif', 'hasil', 'tes', 'murid']
['dipelajari', 'studi', 'efek', 'positif', 'aktivitas', 'fisik', 'anak', 'anak']	['ajar', 'studi', 'efek', 'positif', 'aktivitas', 'fisik', 'anak', 'anak']

5. Lemmatization

Meskipun stemming sudah cukup untuk Bahasa Indonesia, dalam penelitian ini juga dilakukan lemmatization terhadap beberapa kata kunci yang rentan ambigu setelah stemming, agar maknanya tetap sesuai dalam konteks ulasan pelanggan. Hasil lemmatization pada tabel 10 yaitu terdapat pada kata “ujakan” yang berubah menjadi “ujak” dan kata “pencahay” berubah menjadi “cahay”.

Tabel 4.9 Sampel Hasil Lemmatization Data Ulasan Livin By Mandiri.

stemming	review_lemmatized
['tolong', 'timbul', 'donk', 'ujakan', 'pinjam']	tolong timbul donk ujak pinjam
['izin', 'mendonlowd']	izin mendonlowd
['mantap', 'pokok', 'semenjak', 'pake', 'lifin', 'pahlawan', 'power', 'rangers']	mantap pokok semenjak pake lifin pahlawan power rangers
['update']	update
['livin', 'by', 'mandiri', 'ok']	livin by mandiri ok
['kecewa']	kecewa
['update', 'bingung']	update bingung
['solusi', 'tutup', 'rekening']	solusi tutup rekening
['aplikasi', 'nggak', 'buka', 'karna', 'update']	aplikasi nggak buka karna update

'update', 'enak', 'buka', 'nya']	enak buka dia
['gagal', 'verifikasi', 'wajah', 'penuh', 'syarat', 'pencahay', 'aan']	gagal verifikasi wajah penuh syarat cahay aan

Tabel 4.10 Sampel Hasil Lemmatization Data Publik.

stemming	review_lemmatized
['manusia', 'butuh', 'makan', 'air', 'kuat', 'sehat', 'tanam']	manusia butuh makan air kuat sehat tanam
['tanam', 'air', 'layu', 'kering']	tanam air layu kering
['hasil', 'ajar', 'murid', 'gantung', 'mutu', 'guru', 'pada']	hasil ajar murid gantung mutu guru pada
['sekolah', 'mutu', 'guru', 'duga', 'potensi', 'ajar', 'murid', 'sulit', 'capai', 'hasil', 'maksimal']	sekolah mutu guru duga potensi ajar murid sulit capai hasil maksimal
['cenderung', 'murid', 'oleh', 'nilai', 'rendah', 'asal', 'sekolah', 'guru', 'absen']	cenderung murid oleh nilai rendah asal sekolah guru absen
['kait', 'orang', 'tua', 'murid', 'mampu', 'proses', 'nalar']	kait orang tua murid mampu proses nalar
['batas', 'rasio']	batas rasio
['naik', 'angka', 'rasio', 'korelasi', 'negatif', 'nilai', 'tes', 'murid']	naik angka rasio korelasi negatif nilai tes murid
['sekolah', 'milik', 'lapang', 'main', 'olahraga', 'korelasi', 'negatif', 'hasil', 'tes', 'murid']	sekolah milik lapang main olahraga korelasi negatif hasil tes murid
['ajar', 'studi', 'efek', 'positif', 'aktivitas', 'fisik', 'anak', 'anak']	ajar studi efek positif aktivitas fisik anak anak

6. Visualisasi Frekuensi Data

Untuk mendapatkan pemahaman kualitatif mengenai karakteristik data yang digunakan untuk melatih model, dilakukan visualisasi word cloud. Visualisasi ini bertujuan untuk menampilkan kata-kata dengan frekuensi kemunculan tertinggi untuk masing-masing kelas sentimen (positif dan negatif) setelah melalui tahap preprocessing.

Mandiri, dibuatlah visualisasi word cloud yang dipisahkan berdasarkan sentimen positif dan negatif. Analisis ini memberikan wawasan kualitatif mengenai aspek-aspek aplikasi yang paling diapresiasi dan yang paling sering dikeluhkan oleh pengguna.



Gambar 4.3 Word Cloud Sentimen Positif Dataset Livin by Mandiri.

Gambar 4.3 menyajikan kata-kata yang paling sering muncul dalam ulasan bernada positif. Terlihat dengan jelas bahwa kata-kata seperti "bagus", "bantu", "mudah", "cepat", "aplikasi", dan "transaksi" mendominasi visualisasi. Selain itu, kata kunci seperti "livin" dan "mandiri" juga menonjol, menunjukkan bahwa ulasan ini sangat relevan dengan aplikasi yang diteliti.

Distribusi kata ini mengindikasikan bahwa kepuasan pengguna sebagian besar didorong oleh kemudahan dan efisiensi fungsionalitas aplikasi. Pengguna merasa sangat terbantu (bantu) karena aplikasi ini

dapat mengakses atau menggunakan layanan dasar. Kata susah, ribet, dan lemot juga memperkuat narasi tentang pengalaman pengguna yang buruk.

Salah satu temuan paling signifikan dari word cloud ini adalah dominasi kata "update". Ukurannya yang sangat besar menandakan bahwa pembaruan aplikasi merupakan sumber utama keluhan. Hal ini mengimplikasikan bahwa setiap kali versi baru dirilis, sering kali muncul bug, error, atau gangguan baru yang menyebabkan aplikasi menjadi tidak stabil atau bahkan tidak dapat digunakan sama sekali. Kata-kata seperti kecewa, parah, dan buruk adalah ekspresi langsung dari kekecewaan pengguna terhadap kondisi ini.

4.3 Pembobotan Kata

Tahap pembobotan kata merupakan proses penting dalam pengolahan data teks, karena pada tahap ini data teks yang telah dipreproses diubah menjadi representasi numerik berupa vektor berdimensi tetap. Representasi numerik ini diperlukan agar data teks dapat diolah dan dianalisis menggunakan algoritma machine learning dan deep learning.

Dalam penelitian ini, metode pembobotan kata yang digunakan adalah FastText embedding. FastText dipilih karena memiliki kemampuan menghasilkan representasi kata berbasis karakter n-gram, sehingga mampu mengatasi permasalahan out-of-vocabulary (OOV) serta mempertahankan informasi morfologi kata, yang sangat penting dalam Bahasa Indonesia.

Dataset publik yang telah dipreprocessing diubah ke dalam format standar FastText, yaitu dengan menambahkan label sentimen di awal kalimat menggunakan format `__label__<sentimen>`. Data kemudian disimpan ke dalam file .txt untuk digunakan pada proses pelatihan model embedding FastText. Model FastText dilatih menggunakan dataset yang telah diformat sebelumnya. Parameter pelatihan yang digunakan dalam penelitian ini yaitu epoch: 25, Dimensi Vektor (dim): 100. Proses pelatihan ini menghasilkan model embedding yang dapat menghasilkan vektor representasi untuk setiap kalimat atau kata dalam dataset. Setelah model embedding terbentuk, dilakukan proses ekstraksi vektor untuk setiap kalimat ulasan menggunakan fungsi `model.get_sentence_vector()`. Setiap kalimat diubah menjadi sebuah vektor berdimensi 100 yang merepresentasikan makna dari kalimat tersebut secara numerik.

Tabel 4.11 Sampel Hasil Vektorisasi Kalimat.

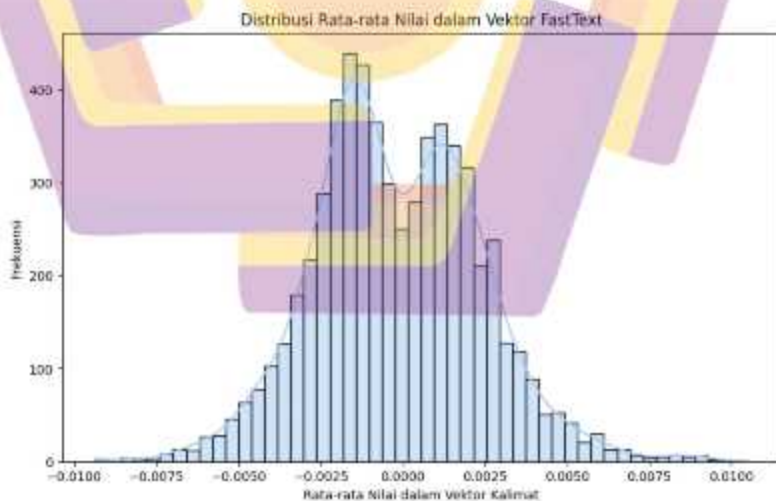
kalimat	sentiment	sentence_vector
Manusia membutuhkan makanan dan air supaya menjadi kuat dan sehat, begitu juga tanaman	Positif	[0.02492376 -0.02076756 -0.00541622 0.00401366 0.00266338 -0.0074173 -0.00326526 0.01204691 -0.01472791 - 0.02349015 0.0072611 0.011812 -0.0094352 -0.01438379 0.02905061 0.00938808 0.01084798 -0.00987781 0.00129599 0.0116742 -0.00871094 - 0.01576739 -0.01104972 0.00047064 -0.01971692 0.00527723 0.00295026 - 0.00361578 0.01150762 -0.0146739 0.01399044 0.00435264 0.01249241 0.0107606 0.01196892 -0.01235288 -0.00480816 0.03001042 0.00388176 0.01077852 -0.0091496 -0.0141182 0.00132459 -0.01328696 0.00843064 0.00996713 0.00918595 -0.00422829 0.02342113 0.01342709 0.00778391 - 0.00956122 0.00623868 -0.00060387

```

-0.00476711 0.01408543 -0.0142306 -
0.00628406 0.00419204 0.00120413
0.0099591 0.00356264 -0.00103507 -
0.01534386 -0.00194317 -0.02135227
0.0141285 -0.02009686 -0.0160515 -
0.00599122 0.02334252 -0.01080518
0.00025703 0.01880915 0.01021054
0.00322251 -0.01263238 -0.01067551
0.0061204 0.01335881 0.00310504
0.01018198 -0.01478565 0.00188918
0.02596896 0.02107049 0.00830601
0.00512903 -0.00387233 0.01448681
-0.02288934 0.00277181 -0.01622383
0.01017986 -0.00967374 0.01583315
-0.01895444 -0.00360958 0.02133182
0.00835942]

```

Seluruh vektor kalimat hasil embedding kemudian dikonversi menjadi array numpy 2D sebagai fitur (X) dan array label sentimen sebagai target (y). Proses ini menghasilkan X shape: (6059, 100) dan y shape: (6059,).



Gambar 4.5 Distribusi Rata-rata Nilai dalam Vektor FastText.

4.4 Pelatihan Model Dataset Publik

Tahap pelatihan model merupakan proses inti dalam penelitian ini. Tujuannya adalah untuk membangun model klasifikasi sentimen berdasarkan dataset publik yang telah dipreprocessing dan dibobotkan menggunakan FastText. Model-model yang digunakan meliputi Support Vector Machine (SVM), Artificial Neural Network (ANN), dan Long Short-Term Memory (LSTM).

Pelatihan dilakukan sepenuhnya menggunakan dataset publik yang telah memiliki label sentimen, agar model dapat mempelajari pola-pola linguistik yang terkait dengan sentimen positif dan negatif. Model yang telah dilatih kemudian digunakan untuk melakukan prediksi pada data ulasan aplikasi Livin' by Mandiri.

1. Model Support Vector Machine (SVM)

Pada tahap awal pelatihan model, algoritma Support Vector Machine (SVM) dilatih menggunakan dataset publik yang telah melalui tahapan preprocessing dan pembobotan kata. Model SVM digunakan sebagai baseline awal tanpa dilakukan proses hyperparameter tuning, dengan menggunakan parameter default library scikit-learn, yaitu kernel linear dan parameter regularisasi $C=1$.

Tabel 4.12 Hasil Evaluasi Awal Model SVM Sebelum Tuning.

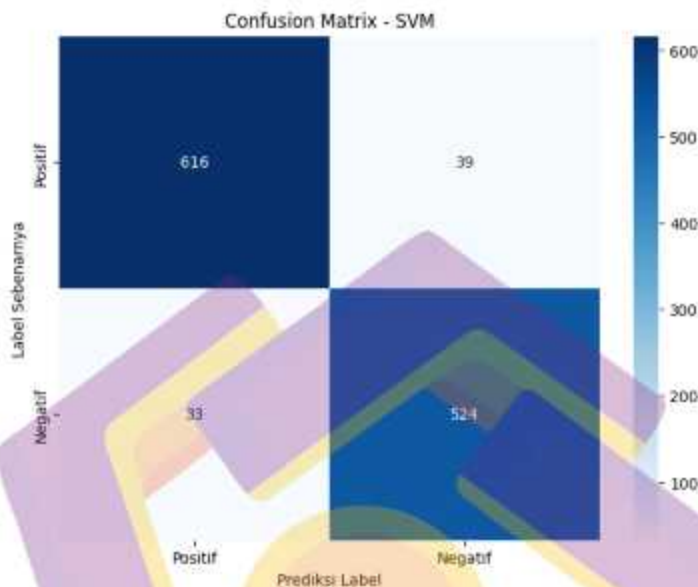
Model	Accuracy	Precision	Recall	F1-score
SVM	0.9406	0.9407	0.9406	0.9406

Berdasarkan hasil evaluasi yang ditunjukkan pada tabel tersebut, dapat disimpulkan bahwa model SVM memiliki performa yang sangat baik meskipun masih menggunakan parameter default. Nilai accuracy sebesar 0,9406 menunjukkan bahwa model mampu mengklasifikasikan data ulasan dengan benar pada 94,06% dari seluruh data uji.

Nilai precision sebesar 0,9407 menunjukkan bahwa dari seluruh prediksi yang dilakukan sebagai sentimen positif, sebesar 94,07% benar-benar merupakan sentimen positif. Nilai ini menunjukkan kemampuan model dalam meminimalisasi kesalahan prediksi positif palsu (false positive).

Nilai recall sebesar 0,9406 mengindikasikan bahwa dari seluruh data yang seharusnya diklasifikasikan sebagai sentimen positif, sebesar 94,06% berhasil diprediksi dengan benar oleh model. Nilai recall yang tinggi menunjukkan bahwa model memiliki sensitivitas yang baik dalam mengenali sentimen positif.

Adapun nilai F1-score sebesar 0,9406 yang merupakan rata-rata harmonis dari precision dan recall, menunjukkan bahwa model memiliki keseimbangan performa yang sangat baik antara kemampuan mendeteksi sentimen positif dan menghindari kesalahan prediksi.



Gambar 4.6 Confusion Matrix Model SVM Sebelum Tuning.

Berdasarkan Gambar 4.6, dapat diketahui bahwa sebanyak 616 data ulasan dengan label sentimen positif berhasil diprediksi dengan benar sebagai positif. Sebanyak 524 data ulasan dengan label sentimen negatif berhasil diprediksi dengan benar sebagai negatif. Sebanyak 39 data ulasan positif salah diprediksi sebagai negatif. Sebanyak 33 data ulasan negatif salah diprediksi sebagai positif. Dari nilai-nilai tersebut, dapat disimpulkan bahwa model memiliki kecenderungan yang seimbang dalam mengenali kedua kelas sentimen, baik positif maupun negatif.

2. Model Artificial Neural Network (ANN)

Setelah dilakukan pelatihan awal terhadap model Artificial Neural Network (ANN) menggunakan dataset publik yang telah diproses,

diperoleh hasil evaluasi performa sebagaimana disajikan pada Tabel 14 berikut:

Tabel 4.13 Hasil Evaluasi Awal Model ANN Sebelum Tuning.

Model	Accuracy	Precision	Recall	F1-score
ANN	0.9414	0.9415	0.9414	0.9414

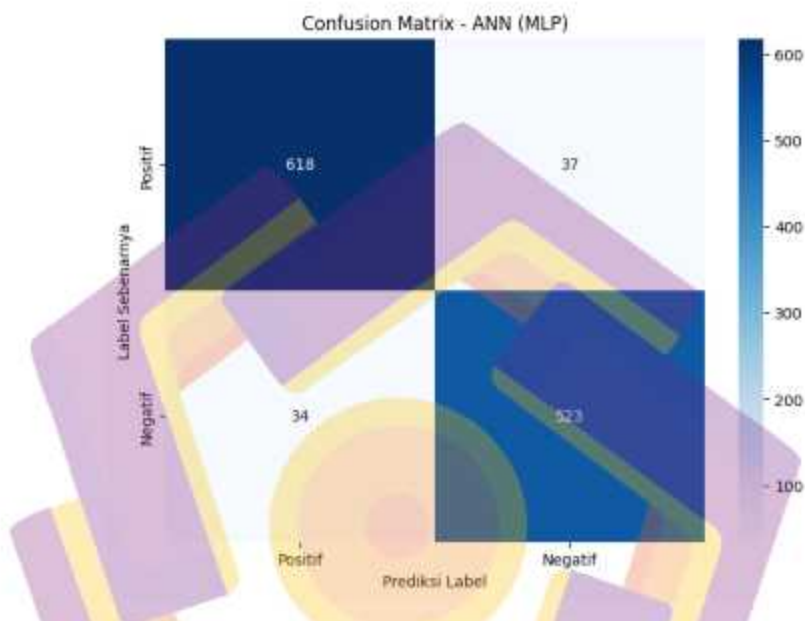
Berdasarkan tabel tersebut, dapat dilihat bahwa performa awal model ANN cukup kompetitif dibandingkan model SVM, dengan nilai accuracy sebesar 0,9414. Artinya, model ANN mampu mengklasifikasikan 94,14% data ulasan secara benar ke dalam kategori sentimen positif dan negatif.

Nilai precision sebesar 0,9415 menunjukkan bahwa dari seluruh prediksi yang dinyatakan sebagai sentimen positif, sebanyak 94,15% prediksi tersebut benar-benar sesuai dengan label aslinya. Nilai precision yang tinggi ini menandakan kemampuan model dalam menekan jumlah kesalahan prediksi positif palsu pada tahap awal pelatihan.

Nilai recall sebesar 0,9414 menunjukkan bahwa model ANN dapat mendeteksi 94,14% dari seluruh data ulasan yang seharusnya diklasifikasikan sebagai sentimen positif. Nilai recall yang tinggi ini menunjukkan sensitivitas model yang baik terhadap kategori sentimen positif.

Nilai F1-score sebesar 0,9414, yang merupakan rata-rata harmonis dari precision dan recall, menunjukkan keseimbangan performa model

antara ketepatan prediksi dan sensitivitas terhadap kategori sentimen positif.



Gambar 4.7 Confusion Matrix Model ANN Sebelum Tuning.

Berdasarkan Gambar 4.7, dapat dijelaskan hasil prediksi model ANN terhadap data uji bahwa sebanyak 618 data ulasan dengan label sentimen positif berhasil diprediksi dengan benar sebagai positif. Sebanyak 523 data ulasan dengan label sentimen negatif berhasil diprediksi dengan benar sebagai negatif. Sebanyak 37 data ulasan positif salah diprediksi sebagai negatif. Sebanyak 34 data ulasan negatif salah diprediksi sebagai positif. Dari hasil ini, dapat terlihat bahwa model ANN cukup seimbang dalam mengenali kedua kelas sentimen. Walaupun jumlah prediksi benar untuk

kelas negatif sedikit lebih tinggi, selisihnya sangat kecil, sehingga performa model dalam mengklasifikasikan kedua kelas dapat dikatakan stabil.

3. Model Long Short-Term Memory (LSTM)

Pada tahap awal penelitian, model Long Short-Term Memory (LSTM) dilatih menggunakan dataset publik hasil preprocessing dan pembobotan kata dengan FastText embedding. Model LSTM dipilih karena kemampuannya dalam menangkap pola sekuensial pada data teks, sehingga diharapkan mampu mengenali konteks sentimen secara lebih akurat.

Hasil evaluasi performa awal model LSTM tanpa dilakukan proses hyperparameter tuning ditunjukkan pada Tabel 4.14 berikut:

Tabel 4.14 Hasil Evaluasi Awal Model LSTM Sebelum Tuning.

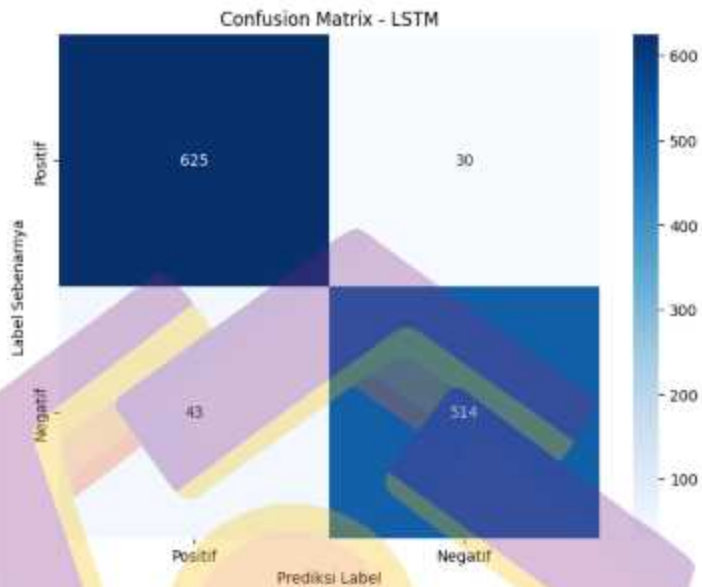
Model	Accuracy	Precision	Recall	F1-score
LSTM	0.9398	0.9399	0.9398	0.9397

Dari tabel di atas dapat dilihat bahwa model LSTM menunjukkan performa awal yang sangat baik, bahkan menjadi model dengan akurasi tertinggi dibandingkan model SVM dan ANN pada percobaan awal. Nilai accuracy sebesar 0,9398 menunjukkan bahwa model mampu mengklasifikasikan 93,98% data ulasan dengan benar ke dalam kategori sentimen positif dan negatif.

Nilai precision sebesar 0,9399 menunjukkan bahwa dari seluruh prediksi sentimen positif yang dilakukan oleh model, sebesar 93,99% benar-benar merupakan data dengan label sentimen positif. Nilai precision yang tinggi ini menandakan bahwa model LSTM memiliki kemampuan yang sangat baik dalam meminimalkan kesalahan prediksi positif palsu.

Nilai recall sebesar 0,9398 menunjukkan bahwa model berhasil mengenali 93,98% dari seluruh data ulasan yang seharusnya diklasifikasikan sebagai sentimen positif. Tingginya nilai recall ini menunjukkan sensitivitas model yang baik dalam mengenali kategori positif.

Adapun nilai F1-score sebesar 0,9397, yang merupakan rata-rata harmonis dari precision dan recall, menegaskan bahwa performa model LSTM sangat seimbang, baik dalam hal ketepatan prediksi maupun sensitivitas terhadap masing-masing kelas sentimen.



Gambar 4.8 Confusion Matrix Model LSTM Sebelum Tuning.

Berdasarkan Gambar 14, diperoleh hasil sebanyak 625 data ulasan dengan label sentimen positif berhasil diprediksi dengan benar sebagai positif. Sebanyak 514 data ulasan dengan label sentimen negatif berhasil diprediksi dengan benar sebagai negatif. Sebanyak 0 data ulasan positif salah diprediksi sebagai negatif. Sebanyak 43 data ulasan negatif salah diprediksi sebagai positif. Nilai prediksi benar (true positive dan true negative) yang tinggi menunjukkan bahwa model LSTM memiliki kemampuan klasifikasi yang sangat baik terhadap kedua kelas sentimen. Selisih antara prediksi benar pada kelas positif dan negatif juga relatif kecil, yang menandakan kestabilan performa model dalam menangani distribusi kelas.

4.5 Hyperparameter Tuning

1. Hyperparameter Tuning Support Vector Machine (SVM)

Setelah melakukan pelatihan awal model Support Vector Machine (SVM) menggunakan parameter default, tahap selanjutnya dalam penelitian ini adalah melakukan proses hyperparameter tuning. Tujuan dari tahap ini adalah untuk mencari konfigurasi parameter optimal yang dapat meningkatkan performa model dalam melakukan klasifikasi sentimen ulasan.

Proses hyperparameter tuning dilakukan menggunakan metode GridSearchCV, dengan kombinasi parameter yaitu $C : [0.1, 1, 10, 100]$, kernel : ['linear', 'rbf', 'poly', 'sigmoid']. Proses tuning dilakukan dengan menggunakan teknik 5-fold cross-validation untuk memastikan hasil tuning bersifat generalizable dan tidak bias terhadap subset data tertentu.

Berdasarkan hasil tuning, diperoleh kombinasi parameter terbaik yaitu $C : 10$, kernel : rbf. Dengan kombinasi parameter tersebut, model SVM berhasil memperoleh skor akurasi cross-validation terbaik sebesar 0,9422.

Nilai parameter C yang optimal pada angka 10 menunjukkan bahwa model SVM lebih memilih batas keputusan yang lebih kompleks dan ketat agar dapat menyesuaikan diri dengan data latih secara akurat. Untuk dataset ini, pendekatan tersebut terbukti lebih efektif dan mampu melakukan generalisasi dengan baik tanpa mengalami overfitting yang parah.

Sementara itu, pilihan kernel `rbf` sebagai parameter terbaik menegaskan bahwa distribusi data sentimen dalam ruang fitur yang dihasilkan oleh `FastText` tidak dapat dipisahkan secara linear. Artinya, tidak ada garis lurus atau bidang datar sederhana yang mampu memisahkan kelas sentimen positif dan negatif secara efektif. Kernel `'rbf'` bekerja dengan memetakan data ke ruang berdimensi lebih tinggi, di mana ia dapat menemukan batas keputusan yang bersifat non-linear dan lebih kompleks. Keberhasilan kernel `'rbf'` menunjukkan bahwa ia mampu menangkap pola hubungan yang rumit antara vektor kalimat dan label sentimennya, yang gagal dilakukan oleh kernel linear.

2. Hyperparameter Artificial Neural Network (ANN)

Setelah melakukan pelatihan awal pada model Artificial Neural Network (ANN) menggunakan parameter default, tahapan selanjutnya adalah menerapkan hyperparameter tuning guna memperoleh konfigurasi parameter terbaik yang dapat meningkatkan performa klasifikasi sentimen.

Proses tuning parameter dilakukan menggunakan `GridSearchCV` dengan kombinasi parameter yaitu `activation : ['relu', 'tanh', 'logistic']`, `hidden_layer_sizes : [(50,), (100,), (50,50)]`, `max_iter : [100, 200, 300]`, `solver : ['adam', 'sgd']`. Proses tuning dilakukan dengan metode 5-fold cross-validation untuk memastikan validitas hasil tuning terhadap berbagai subset data.

Berdasarkan hasil tuning, diperoleh kombinasi parameter terbaik yaitu `activation : tanh`, `hidden_layer_sizes : (50,50)`, `max_iter : 200`, `solver :`

adam. Dengan kombinasi parameter tersebut, model ANN berhasil mencapai skor akurasi cross-validation terbaik sebesar 0,9418.

Beberapa poin penting yang dapat dianalisis dari hasil tuning ini antara lain fungsi aktivasi tanh menjadi pilihan optimal dalam penelitian ini. Fungsi aktivasi tanh memiliki output nilai antara -1 hingga 1, sehingga mampu memberikan distribusi nilai output yang lebih stabil dan simetris dibandingkan fungsi relu atau logistic, terutama pada dataset teks yang telah diembedd ke dalam vektor numerik. Arsitektur optimal yang ditemukan terdiri dari dua hidden layer, dengan masing-masing layer memiliki 50 neuron, hal ini menunjukkan bahwa data memiliki tingkat kompleksitas yang mendapat manfaat dari pembelajaran fitur secara hirarkis, Layer pertama kemungkinan besar belajar untuk mengenali pola-pola dasar dari vektor kalimat, sementara layer kedua belajar untuk mengkombinasikan pola-pola tersebut menjadi representasi yang lebih abstrak dan kompleks sebelum membuat keputusan akhir mengenai sentimen. Parameter `max_iter = 200` menunjukkan bahwa model memerlukan iterasi pelatihan yang moderat agar mencapai konvergensi optimal. Adam terpilih sebagai solver atau algoritma optimasi terbaik. Adam adalah optimizer yang sangat efisien dan populer karena kemampuannya menyesuaikan learning rate secara adaptif untuk setiap parameter (bobot) dalam jaringan. Hal ini membuatnya mampu melakukan konvergensi lebih cepat dan lebih andal dibandingkan solver tradisional seperti sgd pada banyak kasus. Pemilihannya di sini menegaskan

reputasinya sebagai solver yang robas dan cocok untuk dataset dengan dimensi tinggi seperti vektor fitur dari teks.

3. Hyperparameter Long Short-Term Memory (LSTM)

Setelah pelatihan awal menggunakan konfigurasi default, proses berikutnya adalah melakukan hyperparameter tuning terhadap model LSTM untuk memperoleh kombinasi parameter optimal yang dapat meningkatkan performa klasifikasi sentimen. Pada penelitian ini, tuning dilakukan dengan menggunakan metode GridSearchCV yang membungkus model LSTM melalui Scikeras KerasClassifier sebagai estimator, sehingga memungkinkan integrasi penuh antara model Keras dan pustaka scikit-learn.

Proses tuning dilakukan dengan menentukan kombinasi beberapa parameter kunci yang berpengaruh terhadap performa model, yaitu jumlah unit LSTM (`model_lstm_units`) : [64, 128], Jumlah unit Dense (`model_dense_units`) : [32, 64], Dropout rate (`model_dropout_rate`) : [0.3, 0.5], `batch_size` : [16, 32], `epochs` : [10, 20]. Seluruh kombinasi parameter tersebut diuji menggunakan skema 5-fold cross-validation dan skor evaluasi menggunakan metrik accuracy.

Berdasarkan hasil tuning yang dilakukan melalui GridSearchCV, diperoleh kombinasi parameter terbaik yaitu `model_lstm_units` : 128, `model_dense_units` : 64, `model_dropout_rate` : 0.5, `batch_size` : 16, `epochs` : 20. Kombinasi ini menghasilkan skor akurasi cross-validation terbaik sebesar 0,9422.

Beberapa hal penting yang dapat dianalisis dari hasil tuning ini sesuai adalah jumlah unit LSTM sebanyak 64 unit menunjukkan bahwa kapasitas memori sel LSTM untuk menangkap dependensi antar kata dalam kalimat optimal berada pada jumlah ini. Jumlah 128 unit pada layer LSTM menunjukkan bahwa model memerlukan kapasitas memori yang cukup besar untuk menangkap dependensi dan konteks sekuensial dalam data ulasan. Angka ini memberikan keseimbangan yang baik, memungkinkan model untuk mempelajari pola kalimat yang kompleks tanpa menjadi terlalu besar yang berisiko overfitting atau sangat lambat untuk dilatih.. Jumlah unit Dense sebanyak 64 juga terbukti memberikan keseimbangan antara kapasitas representasi fitur dan kemampuan generalisasi model. Pemilihan dropout_rate sebesar 0.5 adalah temuan yang sangat penting. Dropout adalah teknik regularisasi yang kuat untuk mencegah overfitting pada jaringan saraf. Tingkat 0.5 berarti bahwa selama pelatihan, 50% neuron akan dinonaktifkan secara acak pada setiap iterasi. Nilai yang relatif tinggi ini mengindikasikan bahwa model LSTM dengan 128 unit sangat rentan terhadap overfitting. Dengan menerapkan dropout yang kuat, model dipaksa untuk belajar representasi yang lebih robust dan tidak terlalu bergantung pada neuron tunggal, sehingga kemampuan generalisasinya pada data baru meningkat secara signifikan. Batch_size 16 dipilih optimal karena batch yang lebih kecil memungkinkan update bobot lebih sering, mempercepat konvergensi, dan meningkatkan generalisasi model, sesuai konteks dataset embedding FastText yang dimensinya relatif sedang.

Jumlah epoch sebanyak 20 terbukti cukup untuk mencapai konvergensi tanpa overfitting, didukung oleh dropout yang stabil.

Sebagai bagian dari analisis akhir dalam penelitian ini, dilakukan perbandingan performa ketiga model klasifikasi, yaitu Support Vector Machine (SVM), Artificial Neural Network (ANN), dan Long Short-Term Memory (LSTM) sebelum dan sesudah proses hyperparameter tuning. Perbandingan ini bertujuan untuk mengukur sejauh mana proses tuning berpengaruh terhadap peningkatan akurasi model dalam melakukan klasifikasi sentimen ulasan berbahasa Indonesia.

Tabel 4.15 Perbandingan Evaluasi Model Setelah Tuning.

Model	Parameter	Accuracy
SVM	default	0.9406
ANN	default	0.9414
LSTM	default	0.9398
SVM (Tuned)	C : 0.1, kernel : linear	0.9421
ANN (Tuned)	activation : tanh, hidden_layer_sizes : (50,)	0.9418
LSTM (Tuned)	max_iter : 200, solver : sgd model_lstm_units : 128, model_dense_units : 64, model_dropout_rate : 0.5, batch_size : 16, epochs : 20	0.9422

Pada kondisi sebelum tuning, model ANN memperoleh akurasi tertinggi yaitu sebesar 0,9414, diikuti oleh SVM dengan 0,9406, dan LSTM dengan 0,9398. Hasil ini menunjukkan bahwa ANN secara alami memiliki performa awal yang lebih unggul dibandingkan model lain karena kemampuannya dalam memproses data sekuensial serta menangkap konteks kata dalam kalimat. Skor

yang tinggi ini mengindikasikan bahwa representasi fitur menggunakan FastText embedding sangat efektif dan mampu menyediakan dasar yang kuat bagi semua arsitektur model untuk melakukan klasifikasi dengan baik bahkan dengan konfigurasi default.

Setelah proses hyperparameter tuning diterapkan, semua model mengalami peningkatan akurasi, yang menegaskan bahwa optimasi parameter merupakan langkah krusial untuk memaksimalkan potensi setiap model. LSTM menunjukkan peningkatan paling signifikan, naik dari 0.9398 menjadi 0.9422 yaitu peningkatan sebesar 0.0024. Hal ini wajar, mengingat arsitektur deep learning seperti LSTM memiliki lebih banyak parameter yang sangat sensitif terhadap penyesuaian. SVM juga mengalami peningkatan yang solid, dari 0.9406 menjadi 0.9421, menunjukkan bahwa pemilihan kernel dan parameter regularisasi yang tepat berhasil menemukan batas keputusan yang lebih optimal. ANN mengalami peningkatan yang paling kecil, dari 0.9414 menjadi 0.9418. Ini mengindikasikan bahwa konfigurasi default dari library scikit-learn untuk MLP sudah cukup mendekati kondisi optimal untuk dataset ini.

4.6 Analisis Signifikansi Statistik Antar Model pada Dataset Publik

Tabel 4.16 menyajikan perbandingan rinci dari hasil 5-fold cross-validation untuk model SVM, ANN, dan LSTM yang telah dioptimalkan. Analisis ini tidak hanya berfokus pada performa rata-rata tetapi juga pada stabilitas atau konsistensi kinerja setiap model di berbagai subset data.

Tabel 4.16 Hasil 5-fold cross-validation Model.

Fold	SVM	ANN	LSTM
1	0.9495	0.9495	0.9523
2	0.9447	0.9438	0.9428
3	0.9246	0.9265	0.9265
4	0.9532	0.9494	0.9437
5	0.9389	0.9399	0.9456

Perbandingan selanjutnya dilakukan antara SVM dan Long Short-Term Memory (LSTM) untuk menguji apakah perbedaan performa kedua model signifikan. Berdasarkan hasil uji-t, rata-rata selisih akurasi sebesar 0.0000 dengan standar deviasi 0.0055, menghasilkan nilai t-statistic sebesar 0.00 dan p-value lebih besar dari 0.05.

Nilai t-statistic yang sangat kecil menunjukkan bahwa kedua model menghasilkan performa yang hampir identik pada seluruh fold pengujian. Hal ini mengindikasikan bahwa meskipun LSTM secara teoritis memiliki kemampuan lebih baik dalam memahami konteks sekuensial pada teks, pada dataset yang relatif pendek dan sederhana seperti ulasan aplikasi keuangan, keunggulan tersebut belum memberikan perbedaan signifikan dibandingkan SVM.

Dengan demikian, dari sisi efisiensi, SVM dapat dianggap lebih optimal karena menghasilkan performa setara dengan LSTM namun dengan waktu pelatihan yang jauh lebih singkat dan kebutuhan komputasi yang lebih rendah.

Tabel 4.17 SVM vs LSTM.

Fold	SVM	LSTM	d = SVM - LSTM
1	0.9495	0.9523	-0.0028
2	0.9447	0.9428	0.0019

3	0.9246	0.9265	-0.0019
4	0.9532	0.9437	0.0095
5	0.9389	0.9456	-0.0067

Uji paired samples t-test antara model Support Vector Machine (SVM) dan Artificial Neural Network (ANN) dilakukan untuk mengetahui apakah perbedaan rata-rata akurasi keduanya signifikan secara statistik. Berdasarkan hasil cross-validation pada lima fold, diperoleh nilai rata-rata selisih akurasi sebesar 0.00036 dengan standar deviasi 0.00213. Nilai t-statistic yang diperoleh sebesar 0.38 dengan p-value lebih besar dari 0.05.

Hasil ini menunjukkan bahwa tidak terdapat perbedaan yang signifikan antara performa model SVM dan ANN dalam melakukan klasifikasi sentimen pada dataset ulasan aplikasi Livin' by Mandiri. Dengan kata lain, meskipun ANN menunjukkan kecenderungan akurasi sedikit lebih tinggi pada beberapa fold, secara statistik kinerjanya tidak berbeda bermakna dibandingkan dengan SVM.

Tabel 4.18 SVM vs ANN.

Fold	SVM	ANN	d = SVM - ANN
1	0.9495	0.9495	0.0000
2	0.9447	0.9438	0.0009
3	0.9246	0.9265	-0.0019
4	0.9532	0.9494	0.0038
5	0.9389	0.9399	-0.0010

Uji paired samples t-test terakhir dilakukan antara Artificial Neural Network (ANN) dan LSTM. Berdasarkan hasil perhitungan, rata-rata selisih akurasi sebesar -0.00036 dengan standar deviasi 0.00439, menghasilkan nilai t-statistic sebesar -0.18 dengan p-value > 0.05 .

Hasil tersebut menunjukkan bahwa tidak ada perbedaan signifikan antara model ANN dan LSTM dalam mengklasifikasikan sentimen ulasan pengguna. Meskipun LSTM memiliki struktur berlapis dengan memori jangka panjang yang secara teori dapat menangkap konteks antar kata lebih baik, dalam kasus ini ANN mampu mencapai performa yang setara. Salah satu faktor penyebabnya adalah karena sebagian besar ulasan aplikasi Livin' by Mandiri bersifat pendek dan langsung, sehingga konteks temporal antar kata tidak banyak memengaruhi hasil prediksi sentimen.

Hal ini memperkuat argumentasi bahwa ANN yang dioptimalkan dengan teknik regularisasi dan data seimbang melalui SMOTE mampu menjadi alternatif efisien untuk dataset berbahasa Indonesia dengan panjang kalimat pendek hingga menengah.

Tabel 4.19 ANN vs LSTM.

Fold	ANN	LSTM	d = ANN - LSTM
1	0.9495	0.9523	-0.0028
2	0.9438	0.9428	0.0010
3	0.9265	0.9265	0.0000
4	0.9494	0.9437	0.0057
5	0.9399	0.9456	-0.0057

Secara keseluruhan, hasil uji paired t-test terhadap ketiga pasangan model menunjukkan bahwa tidak terdapat perbedaan performa yang signifikan secara statistik di antara SVM, ANN, dan LSTM ($p\text{-value} > 0.05$ pada seluruh pengujian). Ringkasan hasil uji disajikan Tabel 25.

Tabel 4.20 Hasil Uji Signifikansi Statistik antar Model pada Dataset Publik.

Perbandingan	t-value	p-value (approx)	Kesimpulan
SVM vs ANN	0.38	> 0.05	Tidak signifikan
SVM vs LSTM	0.00	> 0.05	Tidak signifikan
ANN vs LSTM	-0.18	> 0.05	Tidak signifikan

Temuan ini mengindikasikan bahwa ketiga model memiliki performa yang relatif setara dalam menganalisis sentimen ulasan pengguna Livin' by Mandiri. Meskipun tidak signifikan secara statistik, ANN menunjukkan rata-rata akurasi tertinggi pada hasil cross-validation, yang menandakan bahwa arsitektur jaringan saraf sederhana dengan regularisasi dan data yang seimbang mampu mencapai hasil yang optimal tanpa kompleksitas berlebih.

4.7 Penerapan Model Terlatih pada Data Ulasan Aplikasi Livin' by Mandiri

Sebagai bagian dari validasi kemampuan generalisasi model, ketiga model yang telah dilatih dan dioptimasi, yakni Support Vector Machine (SVM), Artificial Neural Network (ANN), dan Long Short-Term Memory (LSTM), diterapkan pada dataset ulasan aplikasi Livin' by Mandiri yang diperoleh dari Google Play Store. Data ulasan ini telah melalui tahapan preprocessing dan pembobotan kata menggunakan skema yang sama dengan data pelatihan, untuk memastikan keseragaman format input antar model. Tabel 16 berikut menyajikan sebagian hasil prediksi sentimen dari ketiga model terhadap sejumlah sampel ulasan.

Tabel 4.21 Sampel Hasil Prediksi Sentimen Model Terlatih pada Ulasan Livin' by Mandiri.

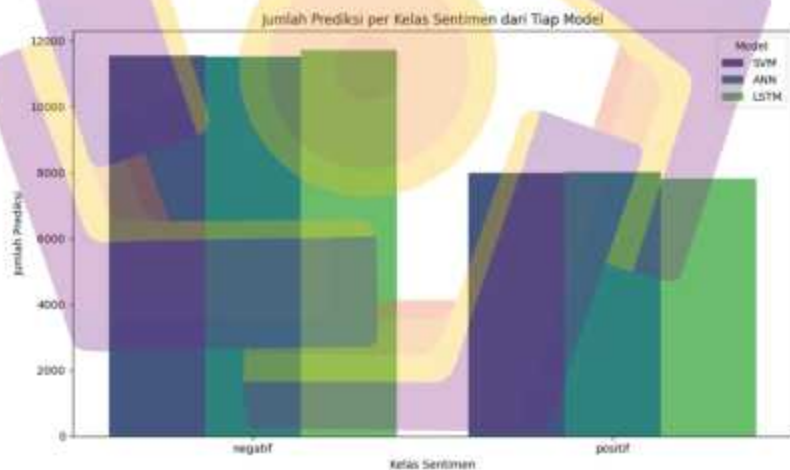
content	label otomatis	prediksi		
		svm	ann	lstm
Tolong timbulkan kan donk ujakan pinjaman	positif	negatif	negatif	positif
Izin kan saya mendonlowd	positif	negatif	negatif	positif
Mantap pokoknya semenjak saya pake lfin saya jadi pahlawan power rangers	positif	positif	positif	positif
update terus	positif	negatif	negatif	positif
Livin' by mandiri ok	positif	negatif	negatif	positif
Saya kembalikan bintangnya jd 5 lagi, terimakasih sudah menyelesaikan kendala saya.	positif	negatif	positif	positif
Good	positif	negatif	negatif	positif
Mantap	positif	positif	positif	positif
Mau coba buka rekening mandiri lewat aplikasi ini ternyata gampang banget, dan proses nya juga cepat ...	positif	negatif	positif	positif
Mantap sanagat puas	positif	negatif	positif	positif

Untuk mendapatkan pemahaman yang lebih dalam mengenai perilaku dan kapabilitas masing-masing model, dilakukan analisis kualitatif terhadap beberapa sampel hasil prediksi pada ulasan Livin' by Mandiri, seperti yang disajikan pada Tabel 18. Analisis ini bertujuan untuk mengidentifikasi kekuatan dan kelemahan setiap model dalam menangani berbagai jenis kalimat.

Pada ulasan yang tidak mengandung ekspresi kepuasan eksplisit, seperti kalimat permintaan atau pernyataan netral, terlihat perbedaan perilaku yang jelas antara model-model. SVM dan ANN secara konsisten memberikan prediksi negatif pada ulasan seperti "Tolong timbulkan kan donk ujakan pinjaman", "Izin kan saya mendonlowd", "update terus", "Livin' by mandiri ok", dan "Good". Perilaku ini menunjukkan bahwa SVM dan ANN telah belajar untuk memiliki ambang batas yang lebih tinggi untuk sentimen positif. Keduanya efektif dalam menyaring ulasan yang bukan merupakan pujian atau ekspresi kepuasan yang tulus. Sebaliknya, model LSTM menunjukkan kelemahan pada kasus-kasus ini dengan secara konsisten memberikan prediksi positif. Hal ini mengindikasikan bahwa LSTM memiliki kecenderungan terlalu optimis dan kurang mampu membedakan antara kalimat yang benar-benar positif dengan kalimat yang hanya berupa permintaan atau pernyataan netral.

ANN dan LSTM yang berbasis jaringan saraf menunjukkan keunggulan. Pada ulasan "Saya kembalikan bintangnya jd 5 lagi, terimakasih sudah menyelesaikan kendala saya," keduanya berhasil memprediksi sentimen positif. Ini menunjukkan kemampuan mereka untuk memahami konteks bahwa sebuah "kendala" telah "diselesaikan", yang menjadikan sentimen akhir kalimat menjadi positif. Keduanya juga berhasil mengklasifikasikan ulasan "Mantap sanagat puas" dan " Mau coba buka rekening mandiri lewat aplikasi ini ternyata gampang banget, dan proses nya juga cepat ..." sebagai positif, menunjukkan robustisitas terhadap salah ketik dan kemampuan memahami narasi.

“login”, “mandiri”, “ga bisa”, dan “login”. Namun, yang menarik adalah kehadiran kata-kata yang jelas-jelas positif seperti “bagus”, “mudah”, “baik”, dan “bantu” di dalam word cloud negatif pada ketiga model. Munculnya kata-kata positif ini menandakan adanya kasus false negative, di mana model salah mengklasifikasikan ulasan yang sebenarnya positif atau setidaknya memiliki elemen positif sebagai ulasan negatif. Hal ini kemungkinan terjadi pada ulasan dengan sentimen campuran misalnya, “Aplikasinya bagus tapi sering error saat transaksi”, di mana model terlalu fokus pada sinyal negatif dan mengabaikan pujian yang ada.



Gambar 4.15 Jumlah Prediksi per Kelas Sentimen dari Setiap Model.

Untuk memahami kecenderungan prediktif dari setiap model, dilakukan analisis terhadap distribusi jumlah prediksi untuk kelas sentimen positif dan

negatif pada keseluruhan dataset ulasan Livin' by Mandiri. Hasil dari analisis ini disajikan pada Gambar 21, yang mengungkapkan perbedaan perilaku yang signifikan antar model.

Temuan paling signifikan dari analisis ini adalah tingkat konsistensi yang tinggi dalam pola prediksi di antara ketiga model. Berbeda dengan beberapa iterasi awal, hasil akhir menunjukkan bahwa SVM, ANN, dan LSTM, setelah melalui proses tuning, kini memiliki perilaku prediktif yang sangat mirip. Secara spesifik, ketiga model menunjukkan kecenderungan yang jelas untuk mengklasifikasikan lebih banyak ulasan sebagai negatif daripada positif. SVM menghasilkan 11.552 prediksi negatif dan 7.978 prediksi positif, sedangkan ANN menghasilkan 11.517 prediksi negatif dan 8.013 prediksi positif, dan LSTM menghasilkan 11.710 prediksi negatif dan 7.820 prediksi positif. Meskipun terdapat perbedaan minor dengan LSTM menjadi yang paling cenderung negatif dan ANN yang paling seimbang secara keseluruhan, ketiganya sepakat bahwa jumlah ulasan negatif lebih dominan dalam dataset ini.

Tabel 4.22 Perbandingan Hasil Evaluasi Model Terlatih pada Dataset Livin' by Mandiri.

Model	Accuracy	Precision	Recall	F1-score
SVM	0.3936	0.3886	0.3936	0.3885
ANN	0.3937	0.3888	0.3937	0.3888
LSTM	0.4241	0.4193	0.4241	0.4184

Seluruh model mengalami penurunan performa yang cukup signifikan dibandingkan saat diuji pada dataset publik. Misalnya, model LSTM yang

sebelumnya mencatat akurasi di atas 94% pada data publik, hanya memperoleh 42.41% akurasi pada data Livin'. Hal ini mengindikasikan bahwa model yang dilatih pada data terstruktur mengalami kesulitan dalam menggeneralisasi ke data riil yang lebih beragam secara linguistik dan kontekstual.

SVM dan ANN menunjukkan performa yang hampir identik, dengan akurasi 0.3936 untuk SVM dan 0.3937 untuk ANN. F1-score keduanya juga sangat mirip. Ini mengindikasikan bahwa dalam skenario domain shift yang ekstrem ini, tidak ada keunggulan signifikan antara arsitektur SVM dan ANN. Model LSTM, meskipun juga berkinerja buruk, secara konsisten sedikit lebih unggul dibandingkan dua model lainnya di semua metrik. Dengan akurasi 42.41% dan F1-score 0.4184, LSTM menunjukkan daya tahan yang sedikit lebih baik terhadap pergeseran domain. Keunggulan ini kemungkinan disebabkan oleh kemampuannya untuk menangkap pola sekuensial dan konteks kalimat, yang mungkin masih memberikan sedikit sinyal yang berguna meskipun banyak kosakata yang tidak dikenal.

Tabel 4.23 Jumlah Kesalahan Prediksi antar Model

Model	Jumlah Kesalahan Prediksi
SVM	11,843
ANN	11,842
LSTM	11,247

Untuk memberikan perspektif evaluasi yang lebih langsung dan praktis, dilakukan perbandingan jumlah absolut prediksi yang salah dari setiap model pada

dataset ulasan Livin' by Mandiri. Tabel 20 menyajikan hasil perhitungan ini dan mengungkapkan perbedaan kinerja yang jelas antar model.

Berdasarkan data pada tabel, model SVM dan ANN menunjukkan tingkat kesalahan yang hampir identik. Model SVM tercatat melakukan 11.843 prediksi yang salah, sementara model ANN melakukan 11.842 kesalahan. Dengan selisih hanya satu prediksi, performa kedua model ini dalam hal jumlah kesalahan dapat dianggap setara atau sebanding. Ini mengindikasikan bahwa, meskipun memiliki arsitektur yang sangat berbeda, kedua model ini menghadapi kesulitan yang sama dalam mengklasifikasikan ulasan pada dataset target. Temuan yang paling signifikan dari analisis ini adalah keunggulan yang jelas dari model LSTM. Model ini tercatat hanya melakukan 11.247 prediksi yang salah. Angka ini secara kuantitatif lebih rendah secara signifikan dibandingkan dua model lainnya yaitu LSTM membuat 596 lebih sedikit kesalahan dibandingkan SVM dan 595 dibandingkan dengan ANN. Pengurangan jumlah kesalahan yang mendekati 600 prediksi ini merupakan bukti kuat akan superioritas performa LSTM pada dataset yang lebih menantang dan bemuansa seperti ulasan pengguna di dunia nyata

4.8 Pelatihan Model Dataset Livin' by Mandiri

Pada tahap ini, dilakukan pelatihan model machine learning secara langsung menggunakan dataset Livin' by Mandiri, yaitu kumpulan ulasan pengguna yang telah diberi label sentimen berdasarkan rating bintang. Label "positif" diberikan pada ulasan dengan rating bintang lima, sedangkan label "negatif" diberikan pada ulasan dengan rating bintang satu. Pemilihan dua label

ekstrem ini bertujuan untuk mengurangi ambiguitas dalam sentimen dan menghasilkan data yang lebih bersih dalam proses klasifikasi.

Model yang dilatih terdiri dari tiga jenis algoritma, yaitu SVM, ANN, dan LSTM. Sebelum proses pelatihan dimulai, dataset telah melalui serangkaian tahapan preprocessing dan pembobotan menggunakan representasi vektor FastText, yang memungkinkan teks ulasan dikonversi ke dalam bentuk numerik yang dapat diproses oleh model.

Tabel 4.24 Perbandingan Hasil Evaluasi Model yang Dilatih pada Dataset Livin by Mandiri.

Model	Accuracy	Precision	Recall	F1-score
SVM	0.8635	0.8771	0.8635	0.8620
ANN	0.9121	0.9130	0.9127	0.9127
LSTM	0.9089	0.9089	0.9089	0.9088

Hasil pelatihan menunjukkan bahwa model ANN memberikan performa terbaik dibandingkan dua model lainnya. ANN menghasilkan akurasi sebesar 91,21%, precision 91,30%, recall 91,27%, dan f1-score 91,27%. Model LSTM menempati posisi kedua dengan akurasi 90,89%, sementara model SVM mencatatkan akurasi 86,35%. Performa model ANN yang unggul dapat dikaitkan dengan kemampuannya dalam mempelajari pola nonlinier dari representasi data, serta sensitivitas terhadap fitur yang relevan dalam teks.

Secara umum, ketiga model menunjukkan performa yang cukup baik saat dilatih menggunakan dataset Livin'. Hal ini menunjukkan bahwa informasi sentimen yang terkandung dalam ulasan Livin' dapat dipelajari secara efektif oleh

model machine learning, terutama jika didukung oleh fitur representasi kata yang kuat seperti FastText. Selain itu, hasil ini juga memperkuat temuan sebelumnya bahwa model yang dilatih pada data target secara langsung cenderung memberikan performa yang lebih baik dibandingkan model yang hanya mengandalkan pelatihan dari domain berbeda.

Pelatihan langsung menggunakan data dari domain target seperti ini penting dilakukan dalam konteks adaptasi domain dan penerapan nyata, mengingat perbedaan bahasa, gaya penulisan, serta konteks penggunaan aplikasi dapat sangat memengaruhi performa klasifikasi. Dengan demikian, pelatihan model menggunakan dataset Livin' merupakan langkah strategis dalam meningkatkan akurasi sistem analisis sentimen untuk aplikasi perbankan digital di Indonesia.

4.9 Analisis Signifikansi Statistik Antar Model pada Dataset Livin'

Tabel 4.25 menyajikan hasil evaluasi yang lebih robas dan mendalam menggunakan skema 5-fold cross-validation pada dataset Livin' by Mandiri. Analisis ini tidak hanya mengukur performa rata-rata, tetapi juga stabilitas kinerja setiap model, yang merupakan faktor krusial dalam menentukan keandalannya.

Tabel 4.25 Hasil 5-fold cross-validation Model.

Fold	SVM	ANN	LSTM
1	0.8661	0.9096	0.9042
2	0.8525	0.9050	0.9012
3	0.8520	0.9063	0.9017
4	0.8566	0.9099	0.9078
5	0.8625	0.9053	0.9030

Nilai t-statistic sebesar -19.65 ($p < 0.001$) menunjukkan bahwa LSTM memberikan hasil yang signifikan lebih baik dibandingkan SVM. Rata-rata akurasi LSTM (0.9036) lebih tinggi daripada SVM (0.8579). Perbedaan ini disebabkan oleh kemampuan LSTM untuk memanfaatkan konteks urutan kata, yang sangat berguna dalam mendeteksi nuansa sentimen dalam kalimat. Namun demikian, margin perbedaan performa ($-4,6\%$) menunjukkan bahwa meskipun LSTM unggul, SVM tetap kompetitif dengan kompleksitas yang jauh lebih rendah.

Tabel 4.26 SVM vs LSTM.

Fold	SVM	LSTM	$d = \text{SVM} - \text{LSTM}$
1	0.8661	0.9042	-0.0381
2	0.8525	0.9012	-0.0487
3	0.8520	0.9017	-0.0497
4	0.8566	0.9078	-0.0512
5	0.8625	0.9030	-0.0405

Uji paired t-test menunjukkan nilai t-statistic sebesar -22.04 dengan p-value < 0.001 , yang berarti terdapat perbedaan signifikan antara model SVM dan ANN. Nilai rata-rata akurasi ANN (0.9072) jauh lebih tinggi dibandingkan SVM (0.8579). Hasil ini menunjukkan bahwa ANN memiliki kemampuan yang lebih baik dalam menangkap pola non-linear pada data ulasan aplikasi Livin' by Mandiri. ANN juga diuntungkan oleh penerapan regularisasi dan SMOTE, yang membantu menyeimbangkan kelas dan mengurangi bias terhadap kelas mayoritas. Sebaliknya, SVM yang berbasis margin linear cenderung kurang fleksibel dalam

menangani data dengan struktur semantik kompleks seperti ulasan teks berbahasa Indonesia.

Tabel 4.27 SVM vs ANN.

Fold	SVM	ANN	$d = \text{SVM} - \text{ANN}$
1	0.8661	0.9096	-0.0435
2	0.8525	0.9050	-0.0525
3	0.8520	0.9063	-0.0543
4	0.8566	0.9099	-0.0533
5	0.8625	0.9053	-0.0428

Perbandingan antara ANN dan LSTM menghasilkan nilai t-statistic sebesar 6.06 dengan p-value = 0.004, menunjukkan perbedaan yang signifikan secara statistik. Meskipun keduanya merupakan model berbasis jaringan saraf, ANN menghasilkan akurasi rata-rata lebih tinggi (0.9072) dibandingkan LSTM (0.9036). Hal ini dapat dijelaskan oleh karakteristik dataset Livin yang sebagian besar terdiri atas kalimat pendek dan langsung pada inti opini, sehingga model berbasis urutan seperti LSTM tidak memperoleh keuntungan yang besar dari struktur temporal kata. ANN justru lebih efisien dalam mempelajari representasi pola sederhana tanpa memerlukan pemrosesan sekuensial.

Tabel 4.28 ANN vs LSTM.

Fold	ANN	LSTM	$d = \text{ANN} - \text{LSTM}$
1	0.9096	0.9042	0.0054
2	0.9050	0.9012	0.0038
3	0.9063	0.9017	0.0046
4	0.9099	0.9078	0.0021
5	0.9053	0.9030	0.0023

Hasil uji paired t-test menunjukkan bahwa ketiga model memiliki perbedaan performa yang signifikan secara statistik, dengan urutan performa rata-rata sebagai sesuai dengan yang ditunjukkan pada tabel 4.29.

Tabel 4.29 Hasil Uji Signifikansi Statistik antar Model pada Dataset Livin.

Perbandingan	t-value	p-value (approx)	Kesimpulan
SVM vs ANN	-22.04	< 0.001	Signifikan
SVM vs LSTM	-19.65	< 0.001	Signifikan
ANN vs LSTM	6.06	0.004	Signifikan

ANN unggul karena kombinasi regularisasi, SMOTE, dan representasi fitur FastText, yang membuat model lebih adaptif terhadap variasi bahasa dalam ulasan pengguna aplikasi keuangan digital. Temuan ini menegaskan bahwa model ANN dapat menjadi alternatif yang efisien untuk klasifikasi sentimen berbahasa Indonesia, memberikan hasil akurasi tinggi tanpa memerlukan arsitektur kompleks seperti LSTM.

4.10 Penerapan Model Livin' by Mandiri pada Dataset Publik

Pada tahap ini, dilakukan penerapan model yang telah dilatih menggunakan dataset Livin' by Mandiri terhadap dataset publik. Tujuan dari eksperimen ini adalah untuk mengevaluasi kemampuan generalisasi dari model yang dibangun berdasarkan data spesifik aplikasi Livin' by Mandiri ketika dihadapkan pada domain data yang berbeda, yaitu dataset publik berbahasa Indonesia dari Kaggle yang terdiri atas ulasan pengguna dari berbagai topik umum.

Setelah model berhasil disimpan, langkah selanjutnya adalah melakukan prediksi terhadap data pada dataset publik yang telah melalui tahap preprocessing dan pembobotan kata menggunakan FastText, sehingga memiliki representasi vektor yang sesuai dengan model.

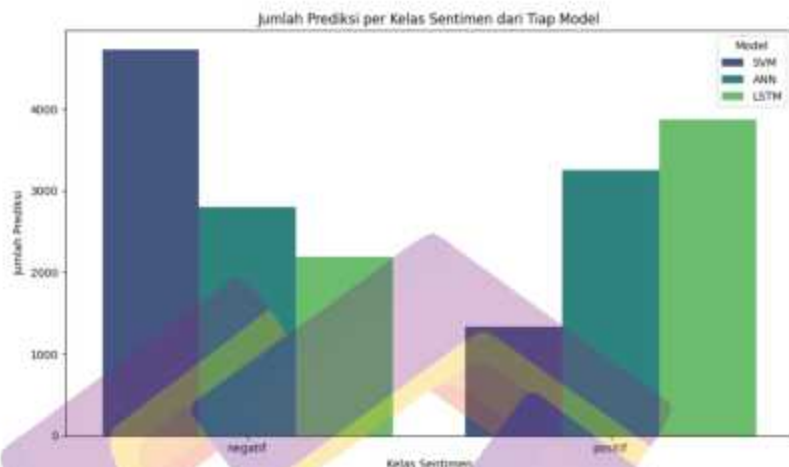
Tabel 4.30 Sampel Hasil Prediksi Model Livin pada Dataset Publik.

content	label otomatis	prediksi		
		svm	ann	lstm
Manusia membutuhkan makanan dan air supaya menjadi kuat dan sehat, begitu juga tanaman	positif	positif	negatif	positif
Tanaman yang tidak mendapat air akan layu dan menjadi kering	positif	negatif	positif	positif
Keberhasilan belajar murid tidak hanya bergantung pada jumlah dan mutu guru yang memadal	positif	negatif	negatif	negatif
Namun, di sekolah yang jumlah dan mutu gurunya kurang, diduga potensi belajar murid akan sulit mencapai hasil maksimal	negatif	negatif	negatif	negatif
Secara umum terdapat kecenderungan bahwa murid yang memperoleh nilai rendah berasal dari sekolah yang gurunya lebih banyak absen	negatif	negatif	negatif	negatif
Dalam kaitan itu, banyak orang tua murid tidak mempunyai cukup kemampuan untuk menjelaskan proses penalaran tersebut	negatif	negatif	negatif	positif
tetapi sampai batas tertentu, yaitu dengan raslo 25:1,	positif	negatif	positif	positif
kenalkan angka raslonya mulai berkorelasi negatif terhadap nilai tes murid	negatif	negatif	negatif	negatif
Sekolah yang memiliki lapangan bermain (olahraga) berkorelasi negatif dengan hasil tes murid	negatif	negatif	negatif	negatif
Persoalan ini perlu dipelajari lebih lanjut, karena studi menunjukkan	positif	negatif	negatif	negatif

efek positif dari aktivitas fisik pada anak-anak

Tabel 4.30 menyajikan analisis kualitatif untuk mengevaluasi kemampuan generalisasi dari model-model yang telah dilatih secara eksklusif pada dataset *Living' by Mandiri* ketika dihadapkan pada data dari domain publik yang umum. Analisis ini memberikan wawasan tentang sejauh mana pengetahuan sentimen yang dipelajari dari domain spesifik dapat ditransfer ke domain yang lebih luas. Pada contoh seperti "Namun, di sekolah yang jumlah dan mutu gurunya kurang, diduga potensi belajar murid akan sulit mencapai hasil maksimal" dan "kenaikan angka rasionya mulai berkorelasi negatif terhadap nilai tes murid", Ketiga model secara serempak dan benar memberikan prediksi negatif. Hal ini menunjukkan bahwa pola-pola linguistik dasar yang menandakan sentimen negatif yang dipelajari dari ulasan aplikasi ternyata cukup fundamental dan dapat diterapkan secara efektif pada konteks kalimat yang sangat berbeda di domain publik.

Kelemahan utama model-model ini terungkap saat dihadapkan pada kalimat positif, terutama yang memiliki struktur kompleks atau mengandung kata bernuansa negatif. Pada kalimat "Persoalan ini perlu dipelajari lebih lanjut, karena studi menunjukkan efek positif dari aktivitas fisik pada anak-anak", ketiga model secara keliru memberikan prediksi negatif. Ini kemungkinan besar terjadi karena model terlalu fokus pada kata "Persoalan" di awal kalimat dan gagal memproses sisa kalimat yang sebenarnya menyimpulkan adanya "efek positif". Demikian pula pada kalimat "Keberhasilan belajar murid tidak hanya bergantung pada jumlah dan mutu guru yang memadai", semua model gagal dan memprediksi negatif.



Gambar 4.22 Jumlah Prediksi per Kelas Sentimen dari Setiap Model.

Grafik pada Gambar 28 menyajikan perbandingan distribusi jumlah prediksi untuk kelas sentimen positif dan negatif dari ketiga model SVM, ANN, dan LSTM. Analisis ini mengungkapkan adanya perbedaan perilaku atau bias prediktif yang fundamental dan signifikan di antara setiap model, meskipun sebelumnya terbukti memiliki skor akurasi yang mirip. Temuan utama dari grafik ini adalah bahwa setiap model menunjukkan pola prediksi yang unik dan sangat berbeda satu sama lain. Model SVM menunjukkan kecenderungan yang sangat kuat untuk mengklasifikasikan ulasan sebagai negatif. Seperti terlihat pada grafik, jumlah prediksi negatifnya 4.725 jauh melampaui jumlah prediksi positifnya 1.334. Perilaku ini menandakan bahwa SVM memiliki ambang batas yang sangat tinggi untuk menganggap sebuah ulasan sebagai positif dan sangat sensitif terhadap kata atau frasa yang berpotensi negatif. Berkebalikan dengan SVM,

model LSTM menunjukkan bias yang kuat ke arah sentimen positif. LSTM menjadi model yang paling banyak memberikan prediksi positif 3.868 dan paling sedikit memberikan prediksi negatif 2.191. Ini mengindikasikan bahwa arsitektur LSTM, yang mampu memproses konteks kalimat, mungkin memberikan bobot lebih pada frasa-frasa resolusi atau pujian dalam sebuah ulasan. Model ANN berada di antara dua ekstrem tersebut, menampilkan distribusi prediksi yang paling seimbang. Dengan jumlah prediksi negatif 2.803 dan positif 3.256 yang tidak terpaut jauh, model ini menerapkan kriteria yang lebih merata dalam membedakan kedua kelas sentimen.

Tabel 4.31 Perbandingan Hasil Evaluasi Model Livin' by Mandiri pada Dataset Publik.

Model	Accuracy	Precision	Recall	F1-score
SVM	0.5532	0.5456	0.5532	0.5151
ANN	0.5305	0.5372	0.5305	0.5306
LSTM	0.5235	0.5418	0.5235	0.5151

Temuan utama dari tabel ini adalah bahwa ketiga model menunjukkan kemampuan generalisasi yang terbatas namun positif. Dengan skor akurasi yang berada di rentang 52% hingga 55%, semua model berhasil berkinerja sedikit lebih baik daripada sebaliknya. Hasil ini membuktikan bahwa model berhasil mempelajari beberapa pola dan kata kunci sentimen yang fundamental misalnya, "baik", "buruk", "masalah", "terima kasih" dari dataset Livin' by Mandiri, dan pengetahuan dasar ini dapat ditransfer ke domain lain. Meskipun demikian, tingkat akurasi yang tidak terlalu tinggi juga secara jelas menunjukkan batasan dari proses transfer pengetahuan ini. Model-model tersebut masih kesulitan karena

perbedaan kosakata, topik, dan struktur kalimat antara ulasan aplikasi yang spesifik dan teks berbahasa Indonesia yang umum.

Analisis perbandingan antar model tidak menunjukkan adanya satu model yang unggul secara absolut di semua metrik. SVM berhasil mencapai akurasi tertinggi di antara ketiganya, dengan skor 55.32%. Namun, F1-score-nya 0.5151 merupakan yang terendah, yang mengindikasikan kemungkinan adanya ketidakseimbangan antara presisi dan recall pada prediksinya. ANN menjadi model dengan F1-score tertinggi sebesar 0.5306. Hal ini menunjukkan bahwa ANN adalah model yang paling seimbang dalam melakukan prediksi, dengan harmoni terbaik antara presisi dan recall, meskipun akurasi sedikit di bawah SVM. LSTM, yang diharapkan dapat mentransfer pemahaman konteks, secara mengejutkan justru mencatatkan akurasi terendah sebesar 52.35%.

Tabel 4.32 Jumlah Kesalahan Prediksi antar Model

Model	Jumlah Kesalahan Prediksi
SVM	2,707
ANN	2,845
LSTM	2,887

Untuk melengkapi evaluasi performa, dilakukan analisis terhadap jumlah absolut prediksi yang salah dari setiap model pada dataset Livin' by Mandiri. Metrik ini memberikan perspektif yang langsung dan praktis mengenai tingkat keandalan setiap model. Hasil dari analisis ini disajikan pada Tabel 26. Model SVM dengan total 2.707 prediksi yang salah, SVM tercatat sebagai model yang paling sedikit membuat kesalahan. Performa ini secara kuantitatif lebih baik

dibandingkan kedua model berbasis jaringan saraf, Jumlah kesalahan yang lebih rendah ini secara langsung menunjukkan bahwa model SVM memiliki kemampuan generalisasi yang paling baik pada data uji, menjadikannya model yang paling akurat dalam analisis ini. Model ANN dan LSTM, meskipun menunjukkan performa yang kuat, tercatat memiliki jumlah kesalahan yang lebih tinggi. Model ANN melakukan 2.845 kesalahan, sementara LSTM melakukan 2.887 kesalahan, yang merupakan angka tertinggi di antara ketiganya. Perbedaan jumlah kesalahan antara ANN dan LSTM sendiri relatif kecil, menunjukkan bahwa performa keduanya cukup sebanding, namun tetap berada di bawah SVM.

4.11 Perbandingan Hasil dengan Penelitian Terdahulu

Untuk mengontekstualisasikan hasil yang diperoleh, penting untuk membandingkan performa model terbaik dari penelitian ini dengan hasil dari penelitian-penelitian relevan yang telah dibahas pada Bab II. Tabel di bawah ini menyajikan perbandingan tersebut, dengan fokus pada akurasi.

Tabel 4.33 Perbandingan Akurasi dengan Penelitian Terdahulu pada Domain Spesifik

Peneliti (Tahun)	Model Terbaik	Akurasi
(Riza & Kurniawan, 2024a)	SVM	81,68%
(M. H. Ali et al., 2025)	SVM	84.90%
(Riza & Kurniawan, 2024b)	XGBM	86,80%
(Suharman et al., 2025)	SVM	87.00%
(Atikah et al., 2024)	BiLSTM	87.30%
(Prasetyo & Agastya, 2024)	SVM	87,46%
Penelitian ini	ANN	91.21%

Tabel di atas memperlihatkan perbandingan hasil akurasi beberapa penelitian terdahulu yang menggunakan berbagai algoritma untuk analisis sentimen ulasan aplikasi perbankan digital, khususnya Livin' by Mandiri, dengan penelitian ini. Terlihat bahwa secara umum, algoritma Support Vector Machine (SVM) merupakan metode yang paling sering digunakan dalam penelitian terdahulu karena kemampuannya dalam menangani data berdimensi tinggi serta menghasilkan margin pemisah optimal antar kelas.

Penelitian (Riza & Kurniawan, 2024a) yang menerapkan SVM dengan teknik oversampling menggunakan SMOTE memperoleh akurasi sebesar 81,68%, menunjukkan performa yang cukup baik pada dataset ulasan Livin' by Mandiri. Hasil ini kemudian meningkat pada penelitian (M. H. Ali et al., 2025) yang juga menggunakan SVM dengan akurasi 84,90%, diduga karena adanya perbaikan pada proses preprocessing dan jumlah data yang lebih besar. Sementara itu, (Riza & Kurniawan, 2024b) mengusulkan penggunaan algoritma Extreme Gradient Boosting (XGBM) sebagai alternatif model ensemble, yang menunjukkan peningkatan performa hingga 86,80%. Temuan ini memperkuat bukti bahwa metode ensemble mampu menangani non-linearity dan noise dengan lebih baik dibanding algoritma linear klasik seperti SVM. Penelitian (Suharman et al., 2025) kembali menggunakan SVM dan melaporkan akurasi 87,00%, yang sejalan dengan tren peningkatan kualitas data dan metode optimasi parameter (hyperparameter tuning) pada penelitian-penelitian yang lebih baru. Dari sisi pendekatan deep learning, penelitian oleh (Atikah et al., 2024) menggunakan Bidirectional Long Short-Term Memory (BiLSTM) dengan dua metode ekstraksi

fitur (TF-IDF dan Word2Vec) dan mencapai akurasi 87,30%. Hasil ini menunjukkan bahwa model yang mampu memahami konteks urutan kata memberikan keunggulan dalam menangkap makna semantik teks. Kemudian, (Prasetyo & Agastya, 2024) juga menggunakan SVM untuk berbagai aplikasi mobile banking dan memperoleh akurasi 87,46%, yang menjadi salah satu hasil tertinggi di antara model berbasis pembelajaran mesin klasik.

Adapun penelitian ini menghasilkan akurasi tertinggi sebesar 91,21% menggunakan algoritma Artificial Neural Network (ANN), hasil ini menunjukkan bahwa ANN mampu mengungguli model klasik seperti SVM maupun model deep learning BiLSTM pada domain data yang sama, terutama ketika proses pelatihan disertai penyeimbangan data dan regularisasi parameter. Dengan demikian, penelitian ini berkontribusi dalam menunjukkan efektivitas model neural network konvensional yang dioptimalkan terhadap dataset sentimen berbahasa Indonesia, sekaligus memperluas pemahaman mengenai pengaruh metode pra-proses dan strategi pelatihan dalam meningkatkan performa klasifikasi sentimen.

BAB 5 PENUTUP

5.1 Kesimpulan

Berdasarkan analisis dan evaluasi yang telah dilakukan, penelitian ini menghasilkan beberapa kesimpulan utama yang menjawab rumusan masalah:

1. Pada dataset publik berbahasa Indonesia, proses hyperparameter tuning terbukti krusial dan berhasil meningkatkan performa ketiga model (SVM, ANN, dan LSTM) hingga mencapai tingkat performa puncak yang hampir identik di atas 94.2% akurasi. Analisis statistik lebih lanjut mengonfirmasi bahwatidak ada perbedaan yang signifikan secara statistik antara ketiga model yang telah dioptimalkan tersebut, menunjukkan bahwa ketiganya merupakan pilihan yang sama kuatnya untuk tugas klasifikasi sentimen pada domain umum.
2. Penelitian ini secara definitif membuktikan adanya masalah pergeseran domain (domain shift) yang parah saat model yang dilatih pada data publik diuji pada domain spesifik aplikasi keuangan. Terjadi penurunan kapabilitas generalisasi yang drastis, di mana akurasi model anjlok dari di atas 94% menjadi hanya 39-42%. Hal ini menegaskan bahwa model analisis sentimen yang bersifat umumtidak dapat diandalkan untuk diterapkan pada domain spesifik tanpa adanya adaptasi atau pelatihan khusus.

3. Ketika model dilatih dan dievaluasi secara langsung pada domain spesifik (ulasan Livin' by Mandiri), hierarki performa yang jelas muncul. Model ANN (Artificial Neural Network) terbukti menjadi model yang paling optimal dan andal, dengan mencapai akurasi tertinggi 91.21% dan stabilitas terbaik (standar deviasi terendah) dalam cross-validation. Analisis statistik menunjukkan bahwa performa ANN secara signifikan lebih unggul daripada SVM dan memiliki kinerja yang setara secara statistik dengan LSTM, menjadikannya pilihan terbaik untuk aplikasi praktis yang dapat digunakan untuk menilai sentimen pada platform yang tidak menyediakan mekanisme bintang atau rating seperti X/Twitter, Facebook Comments, Instagram, atau forum online seperti Reddit dan Kaskus.

Secara keseluruhan, penelitian ini membuktikan bahwa proses pelatihan dan evaluasi lintas domain sangat penting dalam pengembangan model klasifikasi sentimen yang andal. Penggunaan dataset domain-spesifik seperti Livin' by Mandiri tidak serta-merta membatasi generalisasi model, justru dapat menjadi sumber pelatihan yang efektif jika diolah dan dimodelkan dengan pendekatan yang tepat, seperti LSTM. Model yang telah dilatih dengan data ekstrem (rating 1 dan 5) dapat diterapkan untuk menilai sentimen komentar pengguna di media sosial terhadap layanan perbankan digital yang tidak memiliki label bintang, seperti Twitter atau Instagram

5.2 Saran

Berdasarkan hasil penelitian yang telah diperoleh, terdapat beberapa saran yang dapat diajukan untuk penelitian selanjutnya:

1. Model yang telah dikembangkan dapat diadaptasi untuk menganalisis sentimen pada platform lain yang tidak memiliki label eksplisit seperti Twitter atau Instagram. Penelitian selanjutnya disarankan untuk menambahkan model transformer-based seperti IndoBERT atau RoBERTa Indo untuk menguji sejauh mana peningkatan arsitektur memengaruhi kemampuan cross-domain generalization.
2. Penelitian ini menggunakan model LSTM sebagai perwakilan dari algoritma deep learning. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi arsitektur lain seperti Bidirectional LSTM (Bi-LSTM), GRU (Gated Recurrent Unit), atau transformer-based models seperti BERT untuk melihat perbandingan performa yang lebih komprehensif pada data ulasan berbahasa Indonesia.
3. Penelitian ini berfokus pada klasifikasi sentimen polaritas secara umum. Akan lebih bermanfaat apabila penelitian berikutnya mengembangkan metode aspect-based sentiment analysis agar dapat mengidentifikasi sentimen terhadap aspek-aspek tertentu dalam aplikasi, seperti fitur layanan, user interface, dan kecepatan transaksi.
4. Jumlah data berlabel Bahasa Indonesia masih terbatas, sehingga implementasi teknik data augmentation berbasis NLP seperti synonym

replacement, back translation, atau contextual augmentation dapat menjadi solusi untuk meningkatkan performa model di masa depan.

5. Karena data ulasan Livin' by Mandiri yang digunakan tidak memiliki label sentimen ground truth, disarankan untuk melibatkan labeling manual oleh annotator atau menggunakan semi-supervised learning agar evaluasi model dapat dilakukan secara objektif menggunakan metrik performa standar.



DAFTAR PUSTAKA

- Ali, M. H., Ismail, J., & Info, A. (2025). *Jurnal JPILKOM (Jurnal Penelitian Ilmu Komputer) Analisis Sentimen Aplikasi Livin dengan Algoritma Neural Network , Random Forest , dan SVM*. 3(1).
- Ali, S., Wang, G., & Riaz, S. (2020). Aspect based sentiment analysis of ridesharing platform reviews for kansei engineering. *IEEE Access*, 8, 173186–173196. <https://doi.org/10.1109/ACCESS.2020.3025823>
- Atikah, B. S., Sibaroni, Y., & Puspandari, D. (2024). Impact of Feature Extraction on Multi-Aspect Sentiment Classification for Livin' by Mandiri Using BiLSTM. *JOURNAL LA MULTIAPP*, 05(05). <https://doi.org/10.37899/journallamultiapp.v5i5.1541>
- Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends in Machine Learning*, 2(1), 1–27. <https://doi.org/10.1561/22000000006>
- Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13, 281–305. <https://doi.org/10.1145/3575882.3575900>
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5, 135–146. <https://transacl.org/ojs/index.php/tacl/article/view/999>
- Bre, F., Gimenez, J. M., & Fachinotti, V. D. (2018). Prediction of wind pressure coefficients on building surfaces using artificial neural networks. *Energy and Buildings*, 158(November 2017), 1429–1441. <https://doi.org/10.1016/j.enbuild.2017.11.045>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. https://doi.org/10.1007/978-3-030-62008-0_35
- Buntoro, G. A. (2017). Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter. *INTEGER: Journal of Information Technology*, 2(1). <https://doi.org/10.31284/j.integer.2017.v2i1.95>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16(February), 321–357. <https://doi.org/10.1613/jair.953>
- Chepkukaka, Z. K., & Kirugi, F. K. (2019). Service Quality and Customer Satisfaction At Kenya National Archives and Documentation Service, Nairobi County: Servqual Model Revisited. *Nternational Journal on Customer Relations*, 7(1), 1–14.
- Claesen, M., & De Moor, B. (2015). *Hyperparameter Search in Machine Learning*. February. <http://arxiv.org/abs/1502.02127>
- Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20, 273–297. <https://doi.org/10.1088/1742-6596/628/1/012073>

- Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Naacl-Hlt 2019, Mlm*, 4171–4186. <https://aclanthology.org/N19-1423.pdf>
- Divya, K., Siddhartha, B. S., Niveditha, N. M., & Divya, B. M. (2020). An Interpretation of Lemmatization and Stemming in Natural Language Processing. *Journal of University of Shanghai for Science and Technology*, 22(10), 351. <https://www.researchgate.net/publication/348306833>
- Enriko, I. K. A., Gustiyana, F. N., & Putra, R. H. (2023). Komparasi Hasil Optimasi Pada Prediksi Harga Saham PT. Telkom Indonesia Menggunakan Algoritma Long Short Term Memory. *Jurnal Media Informatika Budidarma*, 7(2), 659. <https://doi.org/10.30865/mib.v7i2.5822>
- Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10), 2451–2471. <https://doi.org/10.1162/089976600300015015>
- Glorot, X., Bordes, A., & Bengio, Y. (2011). Domain adaptation for large-scale sentiment classification: A deep learning approach. *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, 1, 513–520.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. In *Linear Algebra*. <https://doi.org/10.1016/b978-0-12-391420-0.09987-x>
- Handayani, Y., Hakim, A. R., & Muljono. (2020). Sentiment analysis of Bank BNI user comments using the support vector machine method. *Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, ISemantic 2020*, 202–207. <https://doi.org/10.1109/iSemantic50169.2020.9234230>
- Haque, T. U., Saber, N. N., & Shah, F. M. (2018). Sentiment analysis on large scale Amazon product reviews. *2018 IEEE International Conference on Innovative Research and Development, ICIRD 2018*, 1–6. <https://doi.org/10.1109/ICIRD.2018.8376299>
- Hochreiter, S., & Schmidhuber, J. (1997a). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hochreiter, S., & Schmidhuber, J. (1997b). LSTM paper. *Neural Computation*, 9(8), 1735–1780.
- Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2017). Bag of tricks for efficient text classification. *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference*, 2, 427–431. <https://doi.org/10.18653/v1/e17-2068>
- Kim, T. K. (2013). T test as a parametric statistic. *Encyclopedia of Sciences and Religions, Tabel 2*, 2215–2216. https://doi.org/10.1007/978-1-4020-8265-8_101087

- Liu, B. (2020). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Second Edition. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Second Edition*, May, 1–432. <https://doi.org/10.1017/9781108639286>
- Manning, C. D., Raghavan, P., & Schultze, H. (2009). *Introduction to Information Retrieval*. Cambridge University Press.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 1–9.
- Obiedat, R., Qaddoura, R., Al-Zoubi, A. M., Al-Qaisi, L., Harfoushi, O., Alrefai, M., & Faris, H. (2022). Sentiment Analysis of Customers' Reviews Using a Hybrid Evolutionary SVM-Based Approach in an Imbalanced Data Distribution. *IEEE Access*, 10, 22260–22273. <https://doi.org/10.1109/ACCESS.2022.3149482>
- Patel, A., Oza, P., & Agrawal, S. (2022). Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model. *Procedia Computer Science*, 218, 2459–2467. <https://doi.org/10.1016/j.procs.2023.01.221>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12(January), 2825–2830.
- Prananda, A. R., & Thalib, I. (2020). Sentiment Analysis for Customer Review: Case Study of GO-JEK Expansion. *Journal of Information Systems Engineering and Business Intelligence*, 6(1), 1. <https://doi.org/10.20473/jjisebi.6.1.1-8>
- Prasetyo, M. J., & Agastya, I. M. A. (2024). Analisis Sentimen Ulasan Aplikasi Perbankan di Google Play Store menggunakan Algoritma Support Vector Machine. *Sentiment Analysis of Banking Application Reviews on Google Play Store using Support Vector Machine Algorithm*. 13, 2386–2400.
- Qureshi, M. A., Asif, M., Hassan, M. F., Abid, A., Kamal, A., Safdar, S., & Akber, R. (2022). Sentiment Analysis of Reviews in Natural Language: Roman Urdu as a Case Study. *IEEE Access*, 10(December 2021), 24945–24954. <https://doi.org/10.1109/ACCESS.2022.3150172>
- Rani, R., & Lobiyal, D. K. (2022). Performance evaluation of text-mining models with Hindi stopwords lists. *Journal of King Saud University - Computer and Information Sciences*, 34(6), 2771–2786. <https://doi.org/10.1016/j.jksuci.2020.03.003>
- Riza, F., & Kurniawan, B. (2024a). ANALISIS SENTIMEN PENGGUNA MOBILE BANKING LIVIN' BY MANDIRI MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. *Jurnal Ismetek*, 17(2), 70–77.
- Riza, F., & Kurniawan, B. (2024b). PERBANDINGAN PERFORMA ALGORITMA KLASIFIKASI MACHINE LEARNING UNTUK ANALISIS SENTIMEN PENGGUNA MOBILE BANKING LIVIN' BY MANDIRI PADA

GOOGLE PLAY STORE. *Jurnal Ismetek*, 18(2), 116–122.

Rokach, L., & Maimon, O. (2005). Decision Trees. *Data Mining and Knowledge Discovery Handbook, Chapter 9*, 165–192. <https://doi.org/10.1007/0-387-25465-X>

Sasaki, Y. (2015). The truth of the F-measure The truth of the F-measure. *International Conference on Information Retrieval, January 2007*, 1–6.

Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>

Suharman, A., Sulaeman, M. K., Industri, T., Muhammadiyah, U., & Hamka, P. (2025). Analisis Sentimen Pengguna Aplikasi Livin ' by Mandiri Menggunakan Metode Support Vector Machine (SVM) dengan Ekstraksi Fitur TF-IDF dan Word2Vec User Sentiment Analysis of the Livin ' by Mandiri Application Using the Support Vector Machine (SVM) Method with TF-IDF and Word2Vec Feature Extraction. 5(8), 2201–2212.

Wisnu, H., Afif, M., & Ruldeyyani, Y. (2020). Sentiment analysis on customer satisfaction of digital payment in Indonesia: A comparative study using KNN and Naive Bayes. *Journal of Physics: Conference Series*, 1444(1). <https://doi.org/10.1088/1742-6596/1444/1/012034>

Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, 53(6), 4335–4385. <https://doi.org/10.1007/s10462-019-09794-5>

Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning. *IEEE Access*, 8, 23522–23530. <https://doi.org/10.1109/ACCESS.2020.2969854>

Zhang, H. (2004). The optimality of Naive Bayes. *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2004*, 2, 562–567.