ANALISIS SENTIMEN TERHADAP JUAL BELI PRODUK SEPATU *PRELOVED* BERDASARKAN ULASAN X DENGAN PERBANDINGAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) DAN *LONG SHORT-TERM MEMORY* (LSTM)

SKRIPSI

DISUSUN OLEH

SETYO HARRY NUGROHO

2109010013

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

MEDAN

2025

ANALISIS SENTIMEN TERHADAP JUAL BELI PRODUK SEPATU *PRELOVED* BERDASARKAN ULASAN X DENGAN PERBANDINGAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) DAN *LONG SHORT-TERM MEMORY* (LSTM)

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer (S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara

SETYO HARRY NUGROHO
NPM. 2109010013

PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2025

LEMBAR PENGESAHAN

Judul Skripsi	:	ANALISIS SENTIMEN TERHADAP JUAL BELI PRODUK SEPATU *PRELOVED* BERDASARKAN ULASAN X DENGAN PERBANDINGAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) DAN *LONG SHORT-TERM MEMORY* (LSTM)
Nama Mahasiswa	:	SETYO HARRY NUGROHO
NPM	:	2109010013
Program Studi	:	SISTEM INFORMASI

Menyetujui
Komisi Pembimbing

(Assoc. Prof. Dr. Al-Khowarizmi, M.Kom)
NIDN. 0127099201

Ketua Program Studi

(Dr. Firahmi Rizky, S.Kom., M.Kom)
NIDN. 0116079201

Dekan

(Assoc. Prof. Dr. Al-Khowarizmi, M.Kom)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

ANALISIS SENTIMEN TERHADAP JUAL BELI PRODUK SEPATU *PRELOVED* BERDASARKAN ULASAN X DENGAN PERBANDINGAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) DAN *LONG SHORT-TERM MEMORY* (LSTM)

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 21 Agustus 2025

Yang membuat pernyataan

SETYO HARRY NUGROHO

NPM. 2109010013

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama	: Setyo Harry Nugroho
NPM	: 2109010013
Program Studi	: Sistem Informasi
Karya Ilmiah	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif *(Non-Exclusive Royalty free Right)* atas penelitian skripsi saya yang berjudul:

ANALISIS SENTIMEN TERHADAP JUAL BELI PRODUK SEPATU *PRELOVED* BERDASARKAN ULASAN X DENGAN PERBANDINGAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) DAN *LONG SHORT-TERM MEMORY* (LSTM)

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 21 Agustus 2026

Yang membuat pernyataan

SETYO HARRY NUGROHO

NPM. 2109010013

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap	: Setyo Harry Nugroho
Tempat dan Tanggal Lahir	: Berastagi, 01 April 2002
Alamat Rumah	: Villa Bukit Mas Berastagi
Telepon/Faks/HP	: 089646667066
E-mail	: setyoharrynugroho@gmail.com
Instansi Tempat Kerja	: -
Alamat Kantor	: -

DATA PENDIDIKAN

SD	: MIS AL-KAROMAH	TAMAT: 2015
SMP	: MTsN KARO	TAMAT: 2018
SMA	: MAN KARO	TAMAT: 2021

KATA PENGANTAR

Alhamdulillah, puji syukur kepada Allah SWT atas segala rahmat dan hidayah nya, sehingga penulis dapat menyelesaikan skripsi berjudul "(**ANALISIS SENTIMEN TERHADAP JUAL BELI PRODUK SEPATU *PRELOVED* BERDASARKAN ULASAN X DENGAN PERBANDINGAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) DAN *LONG SHORT-TERM MEMORY* (LSTM)**)". Skripsi ini disusun untuk memenuhi syarat mendapatkan gelar sarjana dalam program studi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Universitas Muhammadiyah Sumatra Utara.

Dalam proses penelitian dan penyusunan skripsi ini, banyak pelajaran dan tantangan yang dihadapi, yang semuanya memberikan manfaat di masa depan. Semua pencapaian ini tidak lepas dari dukungan dan motivasi dari banyak pihak. Oleh karena itu, penulis menyampaikan terima kasih sebesar-besarnya kepada :

1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Dr. Firahmi Rizky, S.Kom., M.Kom Ketua Program Studi Sistem Informasi
4. Bapak Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom Sekretaris Program Studi Sistem Informasi

5. Dosen Pembimbing Bapak Assoc. Prof. Dr. Al-Khowarizmi, M.Kom terimakasih sudah membimbing penulis dan memberikan penulis kemudahan untuk bimbingan dengan sangat baik, terimakasih juga atas ilmu yang diberikan kepada penulis sehingga penulis bisa sampai ke tahap ini.
6. Dosen Pembahas sekaligus mentor peneliti Bapak Halim Maulana, ST, M,Kom terimakasih sudah membimbing penulis dan memberikan penulis kemudahan untuk bimbingan dengan baik, terimakasih juga atas ilmu yang diberikan kepada penulis sehingga penulis bisa sampai tahap ini.
7. Orang tua tercinta, Ayah dan Ibu terimakasih untuk semua usaha dan perjuangan, yang selalu mengusahakan apa yang penulis mau, ingin, dan butuhkan. Penulis mengucapkan terima kasih yang tak terhingga kepada kedua orangtua tersayang, yang telah menjadi sumber kekuatan, doa, dan semangat di setiap langkah perjalanan ini. Dukungan tanpa henti, kasih sayang yang tulus, serta pengorbanan yang tiada batas menjadi fondasi utama hingga skripsi ini dapat diselesaikan.
8. Sepupu-sepupu keluarga Alm.Wiratman terimakasih untuk motivasi dan semangat yang telah disampaikan. Penulis akan selalu mengingat kasih sayang yang telah diberikan.
9. Untuk Aca, seseorang yang mendampingi penulis selama masa penelitian ini, penulis mengucapkan terimakasih atas perhatian yang selalu diberikan dan mendengar keluh kesah yang tiada hentinya, terimakasih atas segala bentuk usaha yang dilakukan selama ini, semoga selalu begitu.

10. Terimakasih sahabat-sahabat APL teman seperjuangan, yang memberikan semangat dan kebersamaan dalam proses perkuliahan hingga penelitian ini selesai.
11. Terimakasih untuk diri sendiri telah mampu melewati masa perkualiahan ini dengan baik yang sesuai dengan harapan orang tua penulis harapkan.
12. Serta seluruh pihak yang terlibat yang nnamanya tidak penulis sebutkan satu per satu. Penulis mengucapkan terimakasih sudah suka rela membantu dan memberikan arahan dalam menyelesaikan penelitian ini

ANALISIS SENTIMEN TERHADAP JUAL BELI PRODUK SEPATU *PRELOVED* BERDASARKAN ULASAN X DENGAN PERBANDINGAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) DAN *LONG SHORT-TERM MEMORY* (LSTM)

ABSTRAK

Perkembangan media sosial memungkinkan konsumen menyampaikan opini terhadap produk secara terbuka, termasuk pada sepatu preloved. Ulasan ini penting untuk dipahami karena dapat memengaruhi minat beli dan citra produk. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna di platform X (Twitter) menggunakan algoritma Support Vector Machine (SVM) dan Long Short-Term Memory (LSTM). Dataset yang digunakan berjumlah 1.005 ulasan, yang setelah proses preprocessing dan balancing menjadi 738 data, terbagi menjadi sentimen positif dan negatif. Hasil pengujian menunjukkan bahwa algoritma SVM mencapai akurasi sebesar 68%, sedangkan LSTM menghasilkan akurasi sebesar 61,49% pada konfigurasi terbaik. Dengan demikian, SVM terbukti lebih efisien dalam klasifikasi teks sederhana, sementara LSTM membutuhkan parameter yang lebih kompleks agar optimal. Penelitian ini diharapkan dapat menjadi acuan dalam pemanfaatan analisis sentimen untuk mendukung pengambilan keputusan bisnis pada produk preloved.

Kata Kunci: Analisis Sentimen, Sepatu Preloved, SVM, LSTM, X (Twitter).

SENTIMENT ANALYSIS OF BUYING AND SELLING OF PRELOVED SHOES BASED ON REVIEWS X WITH A COMPARISON OF SUPPORT VECTOR MACHINE (SVM) AND LONG SHORT-TERM MEMORY (LSTM) ALGORITHMS

ABSTRACT

The rapid growth of social media enables consumers to express opinions about products openly, including preloved shoes. These reviews are crucial as they can influence purchase intentions and brand perception. This study aims to analyze user reviews on the X (Twitter) platform using Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) algorithms. A total of 1,005 reviews were collected, then preprocessed and balanced into 738 data consisting of positive and negative sentiments. The results show that SVM achieved an accuracy of 68%, while LSTM obtained 61.49% in its best configuration. Thus, SVM demonstrates better efficiency in classifying simple text, whereas LSTM requires more complex parameters to achieve optimal performance. This research is expected to serve as a reference for utilizing sentiment analysis to support business decision-making in the preloved product market.

Keywords: Sentiment Analysis, Preloved Shoes, SVM, LSTM, X (Twitter).

DAFTAR ISI

Halaman

DAFTAR TABEL

DAFTAR GAMBAR

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan internet di Indonesia semakin meningkat dari tahun ke tahun, hal ini membuktikan bahwa internet memiliki peran yang penting terhadap berbagai aspek yang ada. Awalnya internet dikembangkan dan digunakan untuk kepentingan militer di abad 19, namun perkembangan dan kebutuhan akan teknologi terus menuntut internet digunakan di berbagai banyak sektor seperti perdagangan, sosial, hingga sistem pemerintahan (Akmala, 2018). Twitter adalah media sosial yang memungkinkan penggunanya membuat dan membagikan pesan pendek yang disebut “tweet” , yang mencakup hingga 280 karakter yang panjang, teks, video, atau tautan ke situs web lain (Wulandari & Hasan, 2024). Orang-orang menggunakan twitter untuk berkomunikasi, bertanya, meminta petunjuk arah, dukungan, saran, dan memvalidasi interpretasi atau ide terbuka dengan berdiskusi. Twitter telah menggabungkan penerbitan pribadi dan komunikasi, menghasilkan jenis baru penerbitan waktu nyata (real-time) (Grosseck & Holotescu, n.d.)

Sepatu merupakan produk fashion yang paling diminati, berperan sebagai pelindung kaki namun kini menjadi salah satu penunjang penampilan dan mencerminkan kepribadian pemakainya. Beberapa konsumen rela mengeluarkan biaya lebih untuk mendapatkan sepatu dengan brand terbaik, desain yang menarik dan kenyamanan dalam pemakaian. Namun, banyaknya ulasan yang ditinggalkan pembeli mulai dari pujian hingga keluhan menimbulkan kesulitan bagi pelaku bisnis dalam memonitor respons konsumen. Ulasan produk sangat krusial karna

berdampak langsung pada minat beli konsumen dan citra brand. Ulasan tidak hanya mencerminkan pengalaman individu terhadap produk, tetapi juga menyimpan informasi mengenai presepsi konsumen yang dapat diolah melalui analisis sentimen.

Melalui pendekatan ini, opini yang tersebar dalam bentuk teks dapat diidentifikasi dan diklasifikasi menjadi kategori sentimen positif atau negatif. Dengan hal tersebut, dapat memberikan gambaran kepada pembeli terhadap suatu kualitas produk maupun pelayanan. Analisis ini memungkinkan produsen dan penjual memahami ekspetasi konsumen dengan memperbaiki kekurangan produk dan meyusun strategi pemasaran yang tepat. Selain itu analisis ini dapat menjadi alat dalam mengantisipasi penurunan minat pasar yang timbul dari sentimen negatif. Untuk mendapatkan hasil analisis yang akurat dalam klasifikasi, peneliti menggunakan algoritma *Support Vector Machine* (SVM) dan *Long Short-Term Memory* (LSTM). Peneliti membandingkan kedua algoritma tersebut yang bertujuan untuk mengetahui metode yang lebih akurat dan efisien dalam mengklasifikasi sentimen, sehingga dapat memberikan kontribusi dalam pengambilan keputusan bisnis yang lebih tepat sasaran, baik bagi penjual maupun pembeli.

1.2. Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan, peneliti dapat merumuskan permasalahan yaitu Bagaimana metode Support Vector Machine (SVM) dan Long Short-Term Memory (LSTM) dalam mengklasifikasi komentar berdasarkan ulasan sepatu *preloved* di X?

1.3. Batasan Masalah

Batasan lingkup penelitian ini ditetapkan untuk menghindari perluasan pembahasan sehingga tujuan penelitian dapat tercapai. Pertama, Penelitian ini akan membatasi analisis sentimen pada ulasan-ulasan yang diberikan oleh pengguna X (Twitter) terkait produk sepatu *preloved*. Kedua,. Penelitian ini akan berfokus pada penggunaan metode Support Vector Machine dan Long Short Term-Memory dalam mengklasifikasi sentimen pengguna menjadi positif atau negatif.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan, tujuan penelitian ini adalah untuk menganalisis sentimen pengguna X terhadap produk sepatu yang diulas pada platform X. Penelitian ini bertujuan untuk mengidentifikasi dan memahami persepsi serta opini pengguna mengenai produk tersebut melalui analisis ulasan yang diberikan. Selain itu, penelitian ini juga bertujuan untuk mengeksplorasi penerapan dua metode klasifikasi, yaitu *Support Vector Machine* dan *Long Short-Term Memory* dalam mengklasifikasi sentimen pengguna berdasarkan ulasan di X. Dengan demikian, penelitian dapat memberikan wawasan tentang efektivitas kedua algoritma dalam menganalisis sentimen. Terakhir, penelitian ini bertujuan untuk membandingkan tingkat akurasi antara algoritma *Support Vector Machine* dan *Long Short-Term Memory* dalam melakukan klasifikasi sentimen ulasan pada platform X, sehingga dapat memberikan rekomendasi yang lebih baik dalam pemilihan metode analisis sentimen yang tepat untuk di masa depan.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini dibagi menjadi beberapa aspek yang signifikan. Pertama, penelitian ini akan memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang analisis sentimen dan pemrosesan bahasa alami, khususnya dalam konteks media sosial. Hasil penelitian ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan analisis sentimen di platform digital lainnya. Kedua, penelitian ini dapat memberikan wawasan yang berharga bagi pemilik bisnis, khususnya yang bergerak di bidang penjualan produk sepatu, untuk memahami persepsi dan preferensi konsumen terhadap produk mereka. Dengan mengetahui sentimen pengguna, pemilik bisnis dapat mengambil keputusan yang lebih baik dalam strategi pemasaran, pengembangan produk, dan peningkatan layanan pelanggan. Ketiga, penelitian ini juga akan memberikan informasi yang berguna bagi pengembang algoritma, khususnya dalam penerapan metode *Support Vector Machine* dan *Long Short-Term Memory* untuk melakukan analisis sentimen. Dengan membandingkan kedua algoritma , penelitian dapat membantu pemilihan metode yang paling efektif untuk klasifikasi sentimen dalam analisis data. Akhir dari penelitian ini diharapkan dapat meningkatkan kesadaran masyarakat tentang pentingnya ulasan dan sentimen pengguna di media sosial, serta mendorong interaksi yang lebih positif antara konsumen dan penjual dalam ekosistem e-commerce.

BAB II

LANDASAN TEORI

2.1. Analisis Sentimen

Analisis sentimen adalah proses yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini yang diekspresikan dalam teks. Proses ini sering kali dilakukan melalui analisis teks, yang dikenal juga sebagai penambangan opini. Analisis sentimen bertujuan untuk menentukan sikap atau opini seseorang terhadap suatu entitas, yang biasanya dikategorikan sebagai positif dan negatif

Analisis ini adalah salah satu cabang ilmu komputer yang memuat pemrosesan bahasa alami (*Natural Language Processing*) dan pembelajaran mesin (*Machine Learning*) yang mampu mengidentifikasi dan mengklasifikasikan opini atau emosi yang terkandung dalam teks, seperti ulasan, komentar, cuitan (tweet) atau artikel. Sehingga, analisis sentimen hadir sebagai solusi untuk memahami kepuasan pelanggan, mengevaluasi kinerja produk, mendeteksi perubahan tren, dan merekomendasikan konten (Nur Adhan et al., 2024).

2.2. Text Mining

Text Mining dapat didefinisikan secara luas sebagai proses intensif pengetahuan di mana pengguna berinteraksi dengan kumpulan dokumen dari waktu ke waktu dengan menggunakan seperaangkat alat analisis. Text mining mencoba untuk mengekstrak informasi yang berguna dari sumber data melalui identifikasi dan eksplorasi dari suatu pola menarik (Ardiada et al., 2019). Dalam kasus penambangan teks, bagaimanapun, sumber data adalah kumpulan dokumen, dan pola yang menarik ditemukan bukan di antara catatan database yang diformalkan

tetapi dalam data tekstual yang tidak terstruktur dalam koleksi dokumen. Oleh karena itu, penambangan teks dan sistem penambangan data menunjukkan banyak kesamaan arsitektur tingkat tinggi. Contohnya, kedua jenis sistem ini bergantung pada rutinitas *preprocessing*, algoritma penemuan pola, dan elemen lapisan presentasi seperti alat visualisai untuk meningkatkan penelusuran set jawaban.

Text mining adalah salah satu bidang khusus dalam data mining yang memiliki definisi menambang data berupa teks dimana sumber data biasanya didapatkan dari dokumen dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen. Text mining dapat menganalisa dokumen, mengelompokkan dokumen berdasarkan kata-kata yang terkandung didalamnya, serta menentukan kesamaan di antara dokumen untuk mengetahui bagaimana mereka berhubungan dengan variabel lainnya. Penerapan yang paling umum dilakukan text mining saat ini misalnya penyaringan spam, analisa sentimen, mengukur preferensi pelanggan, meringkas dokumen, pengelompokan topik penelitian, dan banyak lainnya. (Findawati et al., n.d.)

Text Preprocessing merupakan tahapan dari proses awal terhadap teks untuk mempersiapkan teks menjadi data yang akan diolah lebih lanjut. Suatu teks tidak dapat diproses langsung oleh algoritma pencarian, oleh karena itu dibutuhkan *preprocessing text* untuk mengubah teks menjadi data numerik. Proses ini terdiri dari beberapa tahapan seperti berikut:

1. *Case Folding, Case Folding* adalah salah satu tahap awal dalam text *preprocessing*. *Case Folding* bertujuan untuk menyeragamkan bentuk huruf dengan cara mengubah semua huruf kapital menjadi huruf kecil.

2. *Data Cleaning*, Pembersihan data atau *Data Cleaning* dilakukan untuk memenuhi syarat pemodelan dengan cara menghapus simbol dan tanda baca yang tidak diperlukan. Proses ini termasuk penghapusan angka, URL, emotikon, hashtag, dan *mention* untuk mengurangi *noise* dalam data.
3. *Tokenisasi*, *Tokenisasi* adalah proses memecah teks menjadi potongan-potongan kecil yang disebut token. Token biasanya berupa kata, tapi bisa juga berupa karakter atau frasa, tergantung jenis tokenisasi yang digunakan, dan keputusan yang di ambil dalam tokenisasi dapat berdampak signifikan pada analisis berikutnya.
4. *Normalisasi*, *Normalisasi* adalah proses mengubah kata tidak baku atau kata tidak standar menjadi bentuk baku atau standar, mengacu pada proses mengembalikan bentuk penulisan ke bahasa yang sesuai dengan Kamus Besar Bahasa Indonesia (KKBI). Karena teks di media social sering berisi kata-kata gaul, singkatan, atau typo.
5. *Stopword*, Tahap ini mencakup penghapusan kata-kata umum atau yang sering muncul dalam teks tetapi tidak membawa makna. Contoh kata yang dihapus meliputi kata depan dan sejenisnya.
6. *Stemming*, *Stemming* adalah proses untuk mengubah kata berimbuhan seperti awalan, akhiran, atau sisipan menjadi bentuk dasar dari sebuah kata. Tujuannya adalah untuk menyederhanakan variasi kata yang memiliki makna dasar yang sama. Sebagai contoh “ini pertemuan diundur lagi” akan diubah menjadi “ini temu undur lagi?”.

2.3. Support Vector Machine

Support Vector Machine (SVM) adalah algoritma *Machine Learning* yang menerapkan fungsi *hyperplane* pada data sehingga terbentuk daerah-daerah tiap kelas. *Hyperplane* sendiri merupakan sebuah fungsi yang digunakan sebagai pemisah antar kelas yang ada. Dalam memprediksi suatu kelas dari data, SVM akan melabelinya berdasarkan daerah kelas mana yang merupakan tempat dari data tersebut. SVM biasanya digunakan pada dataset besar yang diambil dari situs online dan menjadi popular karena penerapannya dalam klasifikasi teks (Fikri et al., n.d.).

Dalam SVM, data direpresentasikan sebagai titik-titik dalam ruang multidimensi, yang mana setiap dimensi merepresentasikan fitur-fitur data. Selain itu, SVM juga dapat menggunakan fungsi kernel untuk mengubah ruang fitur menjadi ruang berdimensi lebih tinggi, yang memungkinkan pemisahan yang lebih baik jika data tidak dapat dipisahkan secara linier. SVM memiliki beberapa keunggulan, antara lain kemampuan untuk bekerja dengan baik pada dataset berdimensi tinggi, kemampuan untuk mengatasi masalah overfitting, dan efektif dalam kasus-kasus di mana pemisahan kelas tidak linier. Namun, SVM juga memiliki kelemahan berupa kinerja yang lambat untuk dataset yang sangat besar. Pada dasarnya SVM diukur dengan hyperline berdasarkan Persamaan (Al-Khowarizmi et al., 2023)

2.4. Long Short-Term Memory

Long Short Term-Memory (LSTM) merupakan algoritma jenis *Recurrent Neural Network* (RNN) yang dirancang untuk memproses dan memprediksi data. LSTM dapat diterapkan di analisis sentiment sebagai klasifikasi opini dalam teks seperti positif atau negatif. Dalam struktur LSTM, simpul-simpul pada lapisan

tersembunyi (*hidden layer*) digantikan oleh sel LSTM yang dirancang khusus untuk menyimpan informasi dari waktu sebelumnya (Ardian Pradana et al., 2023). LSTM memiliki tiga gerbang utama yaitu, *forget gate*, *input gate* dan *output gate*. *Forget* berguna untuk menentukan informasi yang harus dilupakan, input berguna untuk memutuskan informasi baru yang akan disimpan dan *output* sebagai penentuan bagian informasi yang akan dikeluarkan atau dihapus dari cell state.

2.5. Sepatu Preloved

Sepatu merupakan salah satu jenis alas kaki yang dirancang untuk melindungi kaki pada saat beraktivitas. Sepatu memiliki jenis alas kaki (*footwear*) yang biasanya terdiri dari bagian-bagian, sol, kap, tali sepatu, *outsole*, *vamp* atau *upper*, *welt*, *tongue* (lidah sepatu), lancing (tempat mengikatkan tali sepatu) dan lain-lain (jurnal adm,+). Selain berfungsi sebagai pelindung, sepatu menjadi elemen penting dalam dunia fashion yang mampu mencerminkan gaya hidup dan preferensi estetika saat dipakai. Sebagai produk yang memiliki nilai estetika dan fungsional, sepatu cukup diperhatikan oleh konsumen dalam proses pemilihan maupun penggunaanya. Dalam hal tersebut konsumen seringkali membentuk opini berdasarkan pengalaman dari segi kualitas, kenyamanan, desain hingga pelayanan saat melakukan pembelian. Opini ini umumnya diekspresikan melalui ulasan dalam secara digital karna konsumen dapat menyuarakan opininya secara terbuka. Melalui ulasan tersebut dapat memberikan wawasan dan memahami kebutuhan sebelum melakukan pembelian. Dengan infromasi ulasan tersebut dapat dimanfaatkan sebagai data yang relevan dalam analisis sentimen.

Preloved menurut Hansen & Le Zotte (2022) diartikan sebagai produk yang telah dijual kembali dalam kondisi prima. Barang *preloved* berbeda dengan barang

bekas karena umumnya dijual dengan harga lebih murah dengan tetap dalam kondisi yang sebanding dengan barang baru. Produk kelas atas dari berbagai perusahaan terkemuka, seperti sepatu atau tas, sering kali dijual sebagai barang *preloved* agar lebih mudah dijangkau dengan harga yang relatif lebih murah bagi konsumen yang lebih luas. Evans (2022) mengemukakan barang yang dijadikan *preloved* berarti merupakan produk dengan variasi yang langka dipasar karena ada beberapa merek yang hanya memproduksi dengan kuantitas yang terbatas (Yuliana et al., n.d.)

Mayoritas konsumen membeli produk *preloved* salah satunya karena faktor harga yang terjangkau. Dengan memperoleh produk dari merek yang menjadi pilihan konsumen sudah menjadi keuntungan bagi konsumen. Selain itu, menurut Yeo (2022), konsumen dapat menambah koleksi *outfit* tanpa khawatir dengan harga yang mahal. Kedua, apabila konsumen jeli maka dapat menemukan koleksi yang tidak diperoleh secara pasaran. Jika konsumen beruntung memperoleh barang dengan kuantitas yang langka dengan kualitas yang baik, maka nilai jual dari produk tersebut akan meningkat. Ini menjadi keuntungan ekonomis bagi konsumen. Ketiga, membantu mengurangi jumlah sampah. Menjadikan barang *preloved* sebagai alternatif berbelanja produk *fashion* dapat membantu mengurangi jumlah pakaian yang tidak dibutuhkan yang berakhir di tempat pembuangan sampah (Yuliana et al., n.d.)

2.6. X (Twitter)

X merupakan platform media sosial berbasis *microblog* yang memungkinkan penggunanya dalam berbagi pesan singkat yang biasanya disebut "*tweets*". Selain itu X juga dapat melakukan komunikasi, berbagi gambar, video, dan konten

lainnya. Platform ini dikenal sebagai sumber informasi real-time untuk berita, tren serta diskusi publik. Sejak peluncurannya pada tahun 2006, platform ini telah menjadi salah satu media sosial paling populer di dunia, termasuk di Indonesia. Pengguna dapat menyampaikan opini, informasi, dan berinteraksi dengan fitur-fitur seperti balasan, retweet, dan tanda suka. X berisi memiliki sumber data yang sangat luas sehingga tak jarang datanya banyak dimanfaatkan untuk analisis penelitian.

2.7. Google Colab

Google Colaboratory (alias Colab) adalah proyek yang bertujuan untuk menyebarluaskan pendidikan dan penelitian pembelajaran mesin. Buku catatan Colaboratory didasarkan pada Jupyter dan berfungsi sebagai objek Google Docs: dapat dibagikan dan pengguna dapat berkolaborasi pada buku catatan yang sama. Colaboratory menyediakan runtime Python 2 dan 3 yang telah di konfigurasi sebelumnya dengan pustaka pembelajaran mesin dan kecerdasan buatan yang penting, seperti TensorFlow, Matplotlib, dan Keras. Mesin virtual di bawah runtime (VM) dinonaktifkan setelah jangka waktu tertentu, dan semua data dan kofigurasi pengguna hilang. Namun, buku catatan tersebut dipertahankan, dan dimungkinkan juga untuk mentransfer file dari hard disk VM ke akun google Drive pengguna. Layanan Google ini menyediakan runtime yang dipercepat GPU, juga dikonfigurasi sepenuhnya dengan perangkat lunak yang diuraikan sebelumnya. Infrastruktur Google Colaboratory dihosting di platfrom Google Cloud (Carneiro et al., 2018)

Salah satu keunggulan utama Google Colab adalah tersedianya akses gratis ke GPU (*Graphics Processing Unit*) dan TPU (*Tensor Processing Unit*), yang sangat membantu dalam mempercepat proses komputasi, khususnya untuk proyek *machine learning*, *deep learning*, dan analisis data berskala besar. Selain itu,

Google Colab juga mendukung integrasi langsung dengan Google Drive, sehingga pengguna bisa menyimpan, membuka, dan berbagi file notebook (.ipynb) secara mudah dan fleksibel.

Google Colab sangat popular dikalangan akademisi, peneliti, dan mahasiswa karena kemudahan penggunaannya serta dukungannya terhadap berbagai pustaka Python yang umum digunakan dalam analisis data, seperti *NumPy*, *Pandas*, *Matplotlib*, *TensorFlow*, dan *Scikit-learn*. Tidak hanya itu, fitur kolaborasi waktu nyata juga menjadi nilai tambah karena memungkinkan beberapa pengguna mengakses dan mengedit notebook yang sama secara bersamaan, mirip seperti Google Docs.

2.8. Python

Perangkat lunak ini bersifat umum dan sangat popular karna fleksibilitas penggunaanya. Python adalah bahasa pemrograman komputer yang umum digunakan untuk membuat situs web, software, dan aplikasi, serta mengotomatiskan tugas dan analisis data (Kencana Putri & Ichsanuddin Nur, 2023). Python dapat digunakan dalam berbagai platform termasuk windows, macOS dan linux tanpa perlu modifikasi khusus. Selain itu software ini dapat membantu dalam pengembangan web, analisis data, kecerdasan buatan (*AI*) dan pengembangan perangkat lunak. Python digunakan sebagai pra-pemrosesan data hingga evaluasi model dan visualiasi hasil. Python sebagai alat utama dalam proses data seperti normalisasi teks, tokenisasi, penghapusan stopwords hingga stemming. Selanjutnya tahap pemodelan untuk membangun dan melatih kedua mode klasifikasinya yaitu SVM dan LSTM, hingga proses evaluasi perfoma model dengan perhitungan metrik seperti *akurasi*, *presisi*, *recall* dan *F1-Score*.

2.9. Term Frequency – Inverse Document Frequency (TF-IDF)

TF-IDF adalah teknik statistik untuk alat hitung yang mengubah teks menjadi bentuk angka. Teknik ini memiliki dua tahapan yaitu, Tahap pertama adalah menghitung frekuensi kemunculan sebuah kata pada setiap dokumen. Tahap kedua adalah menghitung bobot term tersebut di seluruh dokumen (jurnal wahyuni). *TF-IDF* berfungsi untuk menilai tingkat kepentingan suatu kata dalam sebuah dokumen dibandingkan dengan kemunculannya di keseluruhan kumpulan dokumen. Sebuah kata akan memiliki bobot yang tinggi apabila sering digunakan dalam satu dokumen tertentu, namun jarang ditemukan di dokumen-dokumen lainnya, sehingga kata tersebut dianggap lebih bermakna dan relevan terhadap isi dokumen.

Terdapat 2 komponen perhitungan *TF-IDF* yaitu:

1. *Term Frequency* (*TF*)

Mengukur seberapa sering kata tersebut muncul dalam suatu dokumen.

$$TF(t,d) = \frac{\text{Jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{Total jumlah kata dalam dokumen } d}$$

2. *Inverse Document Frequency* (*IDF*)

Mengukur sebarapa jarang sebuah kata diseluruh korpus.

$$IDF(t,D) = \log\left(\frac{\text{Total jumlah dokumen dalam korpus } D}{\text{Jumlah dokumen yang mengandung kata } t}\right)$$

3. *TF-IDF*

Yaitu gabungan *TF* dan *IDF* untuk memberi bobot kata yang relevan.

$$TF - IDF(t,d,D) = TF(t,d) \times IDF(t,D)$$

Keterangan :

TF (t, d) : menghitung frekuensi relatif dari istilah *t* dalam dokumen *d.*

IDF(t, D): mengukur istilah *t* dalam seluruh koleksi dokumen *D*.

t : *term* atau kata yang sedang dianalisis

d : dokumen yang sedang dianalisis

2.10. Word Cloud

Gambar 2.1. Contoh Visualiasi

Word Cloud adalah Teknik visualisasi yang menampilkan kata dari kumpulan dokumen dengan ukuran yang bervariasi, ukuran tersebut mencerminkan frekuensi yang muncul dalam dokumen tersebut. Apabila ukuran yang muncul lebih besar menunjukann bahwa kata tersebut sering digunakan atau muncul, namun sebaliknya kata yang muncu lebih kecil menandakan bahwa kata tersebut jarang digunakan. Dalam penelitian ini, word cloud untuk menggambarkan secara visual kata-kata yang sering muncul dalam komentar pengguna di Instagram. Dari hasil visualiasi nantinya peneliti dengan mudah mengidentifikasi pola yang mencerminkan opini pengguna berdasarkan komentar yang diberikan.

Tabel 2.1. Penelitian Terdahulu

No	Nama Peneliti	Judul Penelitian	Hasil Penelitian
1.	Rahmi Safitri, Irfan Ali, Nining Rahaningsih (Safitri et al., 2024)	ANALISIS SENTIMEN TERHADAP TREN FASHION DI MEDIA SOSIAL DENGAN METODE SUPPORT VECTOR MACHINE (SVM)	Hasil penelitian menunjukkan bahwa metode Support Vector Machine (SVM) mampu mengklasifikasikan sentimen tren fashion di media sosial dengan akurasi sekitar 80%. Mayoritas respons masyarakat bersentimen positif, sementara sentimen netral dan negatif juga terdeteksi. Analisis ini membantu memahami respons masyarakat dan

			mendukung pengambilan keputusan strategis di industri fashion, meskipun terdapat tantangan dalam menganalisis bahasa informal dan variasi opini.
2.	Rona Guines Purnasiwi , Kusrini,Muhammad Hanafi (Purnasiwi et al., 2023)	Analisis Sentimen Pada Review Produk Skincare Menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM)	Penelitian ini berhasil mengembangkan model analisis sentimen untuk review produk skincare menggunakan teknik LSTM dan Word2Vec. Dengan membagi data ke dalam tiga skenario split (70:30, 80:20, dan 90:10), model

			menunjukkan akurasi tertinggi sekitar 74% pada split 90:10. Hasil evaluasi menunjukkan bahwa penggunaan Word2Vec secara signifikan meningkatkan performa model dibanding tanpa Word2Vec, terutama dalam mengidentifikasi review negatif dan positif. Secara keseluruhan, penelitian ini membuktikan bahwa kombinasi LSTM dan Word2Vec efektif dalam otomatisasi

			analisis opini pengguna di platform review skincare.
3.	Faisal Reza Pradhana, Aziz Musthafa, Indah Fitria (Pradhana et al., 2024)	ANALISIS SENTIMEN ULASAN PRODUK DAVIENA DI SHOPEE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE	Penelitian ini berhasil mengembangkan model analisis sentimen ulasan produk Daviena di platform Shopee menggunakan algoritma Support Vector Machine (SVM). Dengan data sebanyak 4.000 ulasan yang telah diproses dan diberi label sentimen positif, netral, dan negatif, model SVM mencapai akurasi tertinggi sebesar

			94% saat data uji berproporsi 90:10. Hasil ini menunjukkan bahwa metode ini efektif dalam mengklasifikasikan sentimen ulasan secara akurat, membantu produsen dan pemasar memahami persepsi konsumen terhadap produk skincare lokal di Indonesia. Keunggulan utama dari pendekatan ini adalah tingkat keakuratan yang tinggi dan kemampuannya mengolah data

			dalam jumlah besar, meskipun memiliki keterbatasan pada platform dan kategori sentimen yang sederhana.
4.	Jesica Emarapenta Br Sinulingga, Hizkya Cesar Kayika Sitorus (Br Sinulingga & Sitorus, 2024)	Analisis Sentimen Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF	Hasil penelitian menunjukkan bahwa model Support Vector Machine (SVM) dengan fitur TF-IDF untuk analisis sentimen opini masyarakat tentang film horor Indonesia dari Twitter mencapai akurasi sebesar 82.51%. Namun, nilai presisi, recall, dan F1-score masih rendah (sekitar 5-

			7%), menunjukkan performa klasifikasi positif dan negatif yang kurang optimal. Evaluasi dilakukan melalui confusion matrix dan visualisasi seperti word cloud dan pie chart. Penelitian ini berhasil membangun model dasar, tetapi perlu peningkatan fitur, preprocessing, dan parameter untuk hasil yang lebih baik.
5.	Alyssa Rasheedah Cahaya Bintang, Hanny Hafiar, dan Centurion	ANALISIS MEDIA MONITORING TERHADAP PRODUK SEPATU ADIDAS BALI PADA BULAN	Hasil analisis menunjukkan bahwa perhatian terhadap produk Adidas Bali

	Chandratama Priyatna	MARET HINGGA APRIL 2024	meningkat sejak Februari 2024, dengan diskusi yang dimulai dari media sosial dan berita sebelum peluncuran resmi pada 29 Februari 2024. Media berita cenderung bersifat netral dan positif serta memiliki jangkauan yang luas, sementara media sosial lebih aktif menyampaikan berbagai sentimen dan opini dari pengguna, meskipun jangkauannya terbatas. Secara umum, media

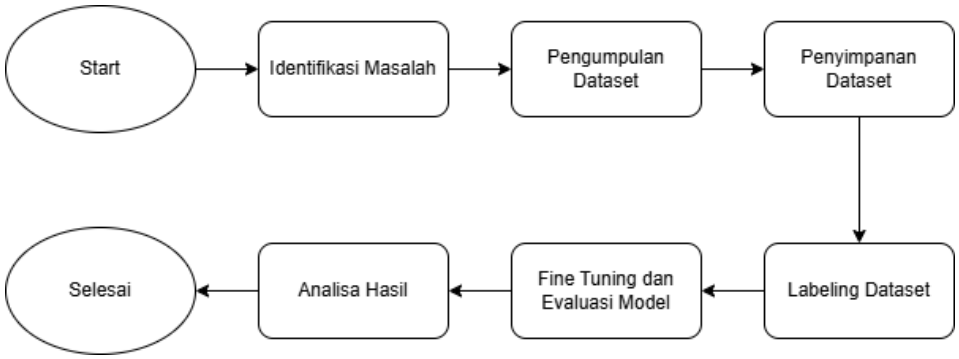
			massa berfungsi sebagai saluran penyebaran informasi resmi, sedangkan media sosial menjadi platform diskusi masyarakat. Peningkatan penyebutan dan jangkauan terjadi menjelang peluncuran produk baru.

BAB III

METODOLOGI PENELITIAN

3.1. Jenis Penelitian

Pada gambar 3.1 tahapan dari metode penelitian ini dimulai dari studi literatur, pengumpulan dataset, penyiapan data, *labeling dataset*, *fine tuning* & evaluasi model, dan analisa hasil.



Gambar 3.1 Tahapan Penelitian

3.2. Identifikasi Masalah

Identifikasi masalah membantu penelitian memahami dengan lebih baik tantangan atau permasalahan yang ada dalam domain yang diteliti. Ini memungkinkan untuk memilih masalah yang relevan dan penting untuk dipecahkan.

3.3. Pengumpulan Dataset

Untuk pengumpulan data menggunakan teknik web *scrapping* atau crawling. Dibandingkan dengan survei secara manual, *crawling* atau *scrapping* lebih efisien karena dapat mengambil data dalam jumlah yang besar dengan cepat. Dengan menggunkan *crawling*, peneliti dapat mengumpulakan data yang diperlukan untuk penelitian dengan mudah dan efektif.

```
import snscrape.modules.twitter as sntwitter
tweets = []
for tweet in sntwitter.TwitterSearchScraper("#sepatupreloved lang:id since:2024-01-01").get_
items():
    tweets.append(tweet.content)
```

Gambar 3.2 API Based Crawling Dataset

Proses *crawling* dimulai dengan mengimpor modul *snscrape.modules.twitter* yang diperlukan untuk mengakses fungsi-fungsi terkait Twitter. Selanjutnya, sebuah list kosong bernama *tweets* dibuat untuk menyimpan konten tweet yang berhasil diambil. *Loop for* kemudian digunakan untuk iterasi melalui setiap tweet yang ditemukan oleh *TwitterSearchScraper* dengan parameter pencarian yang telah ditentukan. Setiap tweet yang memenuhi kriteria tersebut akan diambil kontennya (*tweet.content*) dan dimasukkan ke dalam *list tweets*. Kode ini merupakan contoh sederhana namun efektif untuk mengumpulkan data tweet secara otomatis, yang nantinya dapat digunakan untuk berbagai tujuan analisis, seperti analisis sentimen dalam penelitian tentang produk sepatu *preloved.* Dengan menggunakan parameter *lang:id*, kode memastikan bahwa hanya tweet dalam bahasa Indonesia yang dikumpulkan, sehingga relevan untuk target analisis yang spesifik.

3.4. Penyimpanan Data

Pada gambar 3.3, terlihat contoh kode Python yang menggunakan *library Pandas* untuk menyimpan data hasil analisis sentimen ke dalam file CSV.

```
import pandas as pd
df = pd.DataFrame(tweets, columns=['text'])
df.to_csv('reviews.csv')
```

Gambar 3.3 Penyimpanan Data Crawling CSV

Untuk mengimpor *library Pandas* dengan alias 'pd', yang merupakan konvensi umum dalam komunitas *data science* Python. Selanjutnya, data mentah berupa tweet dimasukkan ke dalam sebuah *DataFrame*, yaitu struktur data tabular dua dimensi yang menjadi inti dari Pandas. *Data frame* ini secara spesifik hanya memiliki satu kolom bernama 'text', yang menunjukkan bahwa fokus penyimpanan adalah pada konten teks dari tweet tersebut.

3.5. Labeling Dataset

Setelah melewati tahapan pengumpulan dataset, proses selanjutnya dilakukan pelabelan dataset untuk menentukan apakah sebuah kalimat, teks atau komentar termasuk ke dalam sentimen positif, negatif atau netral, ini menggunakan data kamus negatif dan positif berbahasa indonesia.

Tabel 3.1 Hasil Sampel Labeling Dataset

No	Nama Barang	Nama Pengguna	Komentar	Hasil
1	Sepatu Tennis UbberSonic 5	E***e	Produk sesuai dgn yg dipromosikan, kualitas ori	Positif
2	Sepatu Trail Running Tracefinder	B***u	Ukuran pas 42 Cuman pas di pake ke siksa banget jempol saya, bagian depan nya agak kecil, jadi klo abis di pake setengah harian itu pegel sakit di bagian kaki depan, Untuk segala macem nya dah bagus banget, kece sih.	Negatif
3	Aerostreet 37-44 Brooklyn Putih	L***r	Bagus/keren sepatu sesuai dengan yang di gambar,kece dah pokok nya	Positif
4	DF 690 Sepatu	L***a	Bentuk dan ukuran: kebesaran Aroma:	Positif

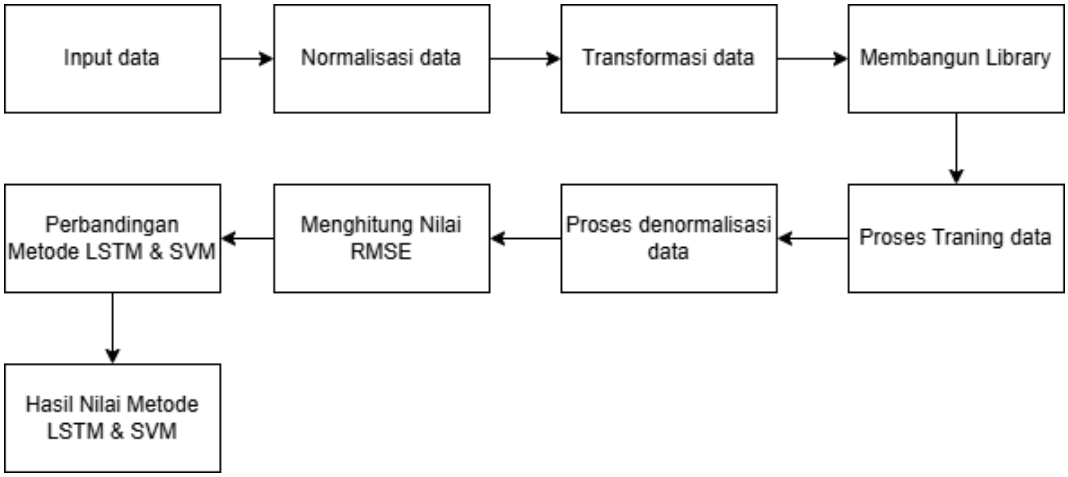
	Wanita Sneakers		wangi Daya tahan pakai: belum di coba Warna: sesuai harapan	
5	Gio Saverino	K***n	Sepatu harga segitu yang gagal yang pernah di beli , bentukan gak jelas , antara kaki kanan Dan kiri gak jelas.	Negatif

Struktur Data *Crawling* yang akan diperoleh terdiri dari Nomor urut, Nama Barang, Nama pengguna, Komentar, Label Sentimen(Positif atau Negatif). Yang memiliki sebuah pola sentimen Teridentifikasi yaitu:

1. Ulasan Positif: Kata kunci "Produk Sesuai", "Kualitas","Sesuai", "Bagus"
2. Ulasan Negatif: Kata kunci "Tidak Sesuai","Packing Kardus", "Kualitas Kurang", "Tidak Amanah", "Insole Lepas"

3.6. *Fine Tuning* dan Evaluasi Model

Fine-tuning adalah proses penyesuaian atau pelatihan kembali (*retraining*) model. *Fine tuning* melibatkan penyesuaian parameter model menggunakan data *ztuning*, model dievaluasi menggunakan data uji untuk mengukur seberapa baik model dapat memprediksi sentimen dengan akurat menggunakan *confusion matrix*.

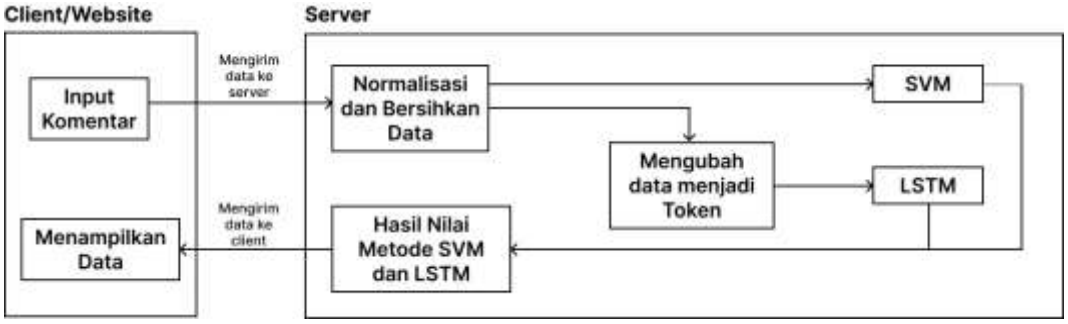


Gambar 3.4 General Architecture LSTM dan SVM

Proses dimulai dari pengumpulan data teks mentah, dilanjutkan dengan tahap pra-pemrosesan seperti normalisasi dan transformasi data untuk mempersiapkan input yang sesuai bagi model. Sistem kemudian membangun dan melatih dua model berbeda, LSTM yang cocok untuk data sekuensial seperti teks, dan SVM yang efektif untuk klasifikasi sebelum melakukan pengujian untuk mengevaluasi performanya. Hasil prediksi melalui proses denormalisasi agar dapat diinterpretasi, sementara evaluasi akhir menggunakan RMSE sebagai metrik akurasi. Alur ini tidak hanya mencakup pelatihan dan pengujian model, tetapi juga memungkinkan perbandingan kinerja antara pendekatan *deep learning* (LSTM) dan *machine learning* tradisional (SVM), sehingga memberikan landasan yang komprehensif untuk pengambilan keputusan dalam memilih model analisis sentimen yang optimal.

3.7. Perancangan Sistem

Alur kerja sebuah sistem analisis komentar berbasis web yang dirancang untuk memanfaatkan dua metode pemrosesan data, yaitu *Support Vector Machine* (SVM) dan *Long Short-Term Memory* (LSTM). Proses dimulai dari sisi client atau website, di mana pengguna memberikan masukan berupa komentar melalui fitur input yang tersedia. Komentar tersebut kemudian dikirimkan ke server untuk menjalani proses pengolahan. Tahap awal yang dilakukan pada server adalah normalisasi dan pembersihan data, yang bertujuan menghilangkan karakter-karakter yang tidak relevan, kesalahan penulisan, atau informasi yang tidak diperlukan. Proses ini penting agar data yang masuk memiliki format yang konsisten dan siap untuk dianalisis. Setelah proses pembersihan selesai, data diproses ke dalam dua jalur berbeda. Jalur pertama langsung mengirimkan data ke metode SVM untuk dianalisis, sedangkan jalur kedua mengubah data terlebih dahulu menjadi token, yang merupakan representasi numerik dari kata atau kalimat, sehingga dapat dibaca dan diproses oleh metode LSTM.



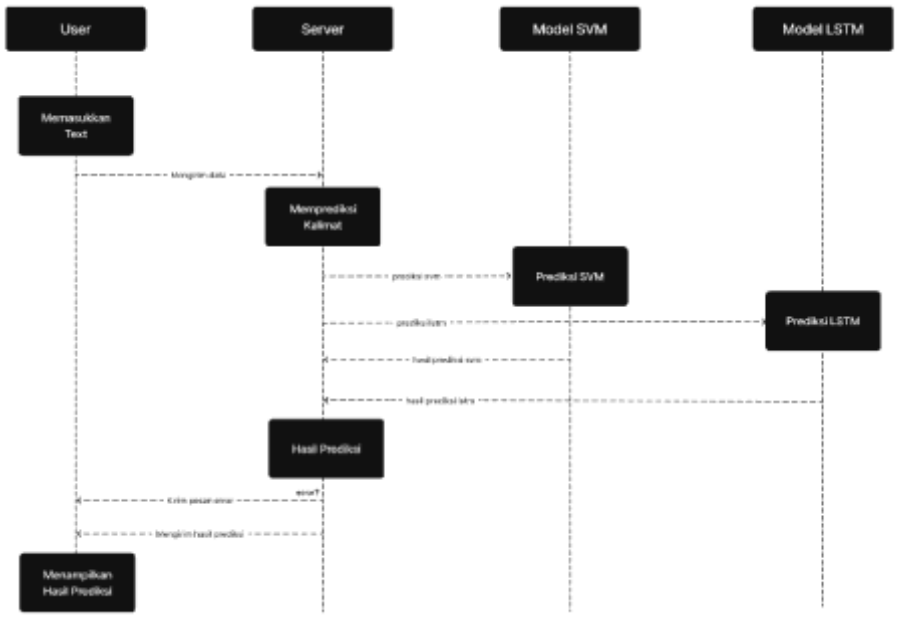
Gambar 3.5 Perancangan Sistem

Tahap selanjutnya adalah pelaksanaan analisis oleh kedua metode tersebut. SVM akan mengolah data dengan memisahkan atau mengklasifikasikan informasi berdasarkan pola tertentu yang telah dipelajari dari data sebelumnya, sedangkan

LSTM, yang merupakan bagian dari jaringan saraf tiruan, akan memanfaatkan kemampuan memahami konteks dan urutan kata untuk menghasilkan analisis yang lebih mendalam. Hasil dari kedua metode ini kemudian digabungkan untuk menghasilkan keluaran yang lebih akurat dan menyeluruh. Data hasil analisis tersebut dikirim kembali ke website dan ditampilkan kepada pengguna secara langsung. Dengan mekanisme kerja seperti ini, sistem tidak hanya mampu memberikan hasil analisis komentar secara otomatis dan cepat, tetapi juga memastikan kualitas informasi yang dihasilkan tetap tinggi, berkat kombinasi kekuatan analisis dari dua algoritma yang berbeda namun saling melengkapi.

3.8. Model Perancangan Sistem

Model perancangan sistem berfungsi untuk menyederhanakan suatu rangkaian proses atau prosedur sehingga lebih mudah dipahami dan divisualisasikan berdasarkan urutan langkah dari suatu proses. Berikut ini adalah flowchart keseluruhan yang digunakan sebagai acuan dalam penelitian.



Gambar 3.6 Model Perancangan Sistem

Pada diagram di atas, alur proses prediksi sentimen dimulai dari User yang memasukkan teks. Teks yang dimasukkan kemudian dikirimkan ke Server untuk diproses. Server akan melakukan tahap Memproses Kalimat dengan mengirimkan data tersebut ke dua model yang digunakan, yaitu Model SVM dan Model LSTM.

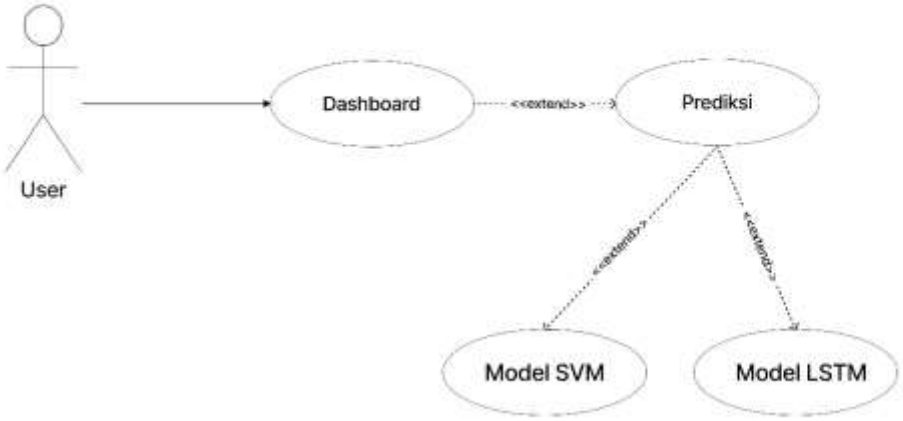
Pertama, server mengirimkan data ke Model SVM untuk dilakukan Prediksi SVM. Setelah hasil prediksi diperoleh, data kemudian juga diproses oleh Model LSTM untuk melakukan Prediksi LSTM. Kedua hasil prediksi (SVM dan LSTM) dikembalikan ke server.

Selanjutnya, server akan mengelola kedua hasil tersebut pada tahap Hasil Prediksi. Jika terjadi kesalahan dalam proses prediksi, server akan mengirimkan pesan error kembali kepada user. Namun jika berhasil, server akan mengirimkan hasil prediksi yang kemudian ditampilkan kepada User pada tahap Menampilkan Hasil Prediksi.

3.8.1 Use Case Diagram

Use case diagram tersebut menggambarkan sebuah sistem yang dirancang untuk memberikan kemudahan kepada pengguna dalam melakukan analisis dan prediksi berbasis machine learning melalui antarmuka Dashboard. Pada tahap awal, pengguna berinteraksi dengan sistem melalui Dashboard yang berfungsi sebagai pusat kontrol utama, tempat seluruh informasi dan fitur dapat diakses secara terintegrasi. Dashboard tidak hanya menampilkan data atau informasi statis, tetapi juga diperluas dengan kemampuan untuk melakukan Prediksi, yang menjadi inti dari sistem ini. Prediksi yang dilakukan sistem tidak berdiri sendiri, melainkan bergantung pada model algoritma machine learning yang ditanamkan di dalamnya. Hubungan *extend* pada diagram menunjukkan bahwa setiap proses prediksi dapat

dikembangkan lebih lanjut dengan memanfaatkan model yang berbeda sesuai kebutuhan atau karakteristik data yang dianalisis.



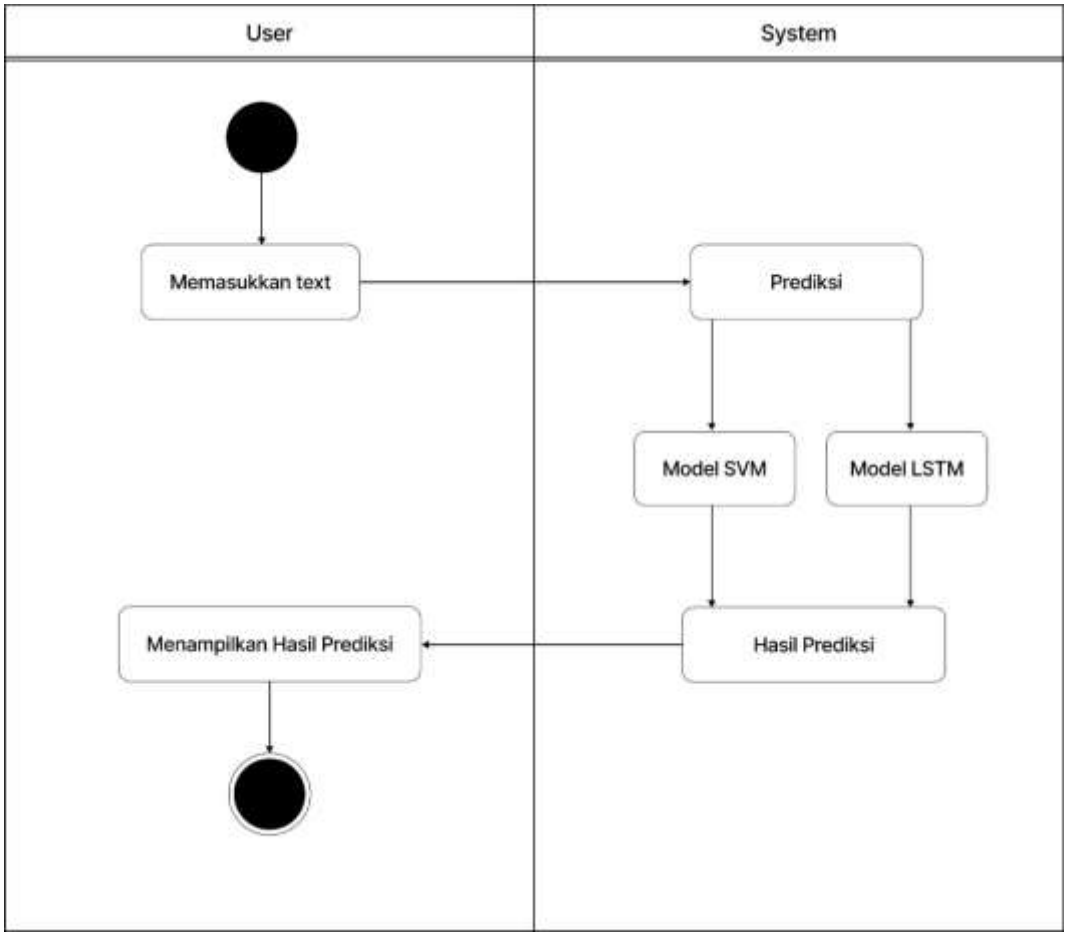
Gambar 3.7 Use Case Diagram

Dalam implementasinya, sistem menyediakan dua jalur pemodelan utama, yaitu Model SVM (Support Vector Machine) dan Model LSTM (Long Short-Term Memory), yang masing-masing memiliki keunggulan dan peran tersendiri. Model SVM biasanya digunakan untuk data dengan dimensi tinggi dan cocok dalam memisahkan kelas secara jelas menggunakan hyperplane, sehingga sering diterapkan untuk masalah klasifikasi dengan data yang relatif terstruktur.

3.8.2 Activitiy Diagram

Activity yang ditampilkan secara komprehensif menggambarkan bagaimana interaksi antara pengguna dan sistem terjalin dalam sebuah alur yang runtut dan terstruktur, khususnya dalam konteks pemrosesan teks untuk menghasilkan prediksi berbasis machine learning. Proses dimulai ketika pengguna

memasukkan teks sebagai input utama yang berperan sebagai titik awal dari keseluruhan aktivitas.



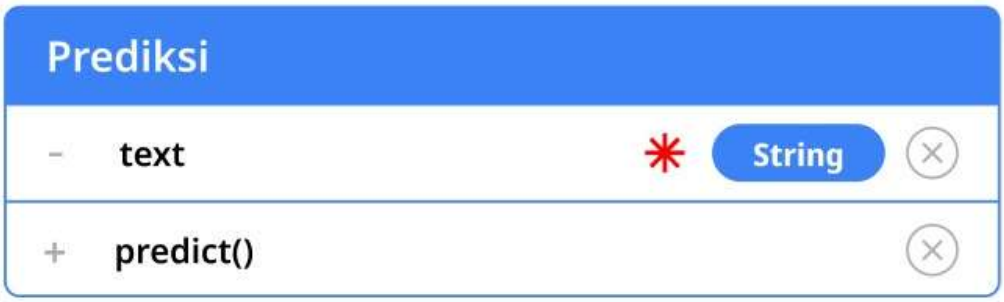
Gambar 3.8 Activitiy Diagram

Input tersebut kemudian diproses oleh sistem melalui tahap prediksi, yang merupakan inti pemrosesan dan berfungsi sebagai jembatan antara data mentah dengan hasil yang dapat diinterpretasikan. Pada tahap ini, sistem tidak bekerja secara tunggal, melainkan menyediakan dua jalur pemodelan berbeda yang saling melengkapi, yaitu Model SVM dan Model LSTM.

3.8.3 Klasifikasi

kelas ini memiliki properti text yang bertipe data String dan ditandai dengan simbol bintang merah yang menandakan bahwa atribut ini bersifat wajib atau

mandatory, sehingga kelas tidak dapat dijalankan tanpa adanya input teks. Atribut ini berfungsi sebagai penampung data masukan berupa teks yang nantinya akan diproses dalam metode prediksi.



Gambar 3.9 Klasifikasi

Selain itu, kelas ini juga memiliki sebuah metode bernama predict() yang menjadi inti dari fungsionalitas kelas, di mana metode tersebut akan memanfaatkan nilai yang tersimpan pada atribut text untuk menghasilkan keluaran berupa hasil analisis atau prediksi tertentu.

BAB IV

HASIL DAN PEMBAHASAN

4.1. Deskripsi Umum Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari kumpulan teks yang diambil dari platform media sosial X (sebelumnya dikenal sebagai Twitter). Platform ini dipilih karena sering digunakan masyarakat untuk menyampaikan pendapat, perasaan, dan penilaian mengenai berbagai topik, termasuk produk seperti sepatu preloved. Jumlah data teks yang berhasil dikumpulkan adalah 1. 005, masing-masing mewakili satu opini atau pernyataan pengguna yang secara eksplisit atau implisit menyampaikan sentimen terhadap sepatu preloved.

Proses pelabelan ini mempertimbangkan konteks keseluruhan kalimat, gaya bahasa, serta kata-kata yang mengungkapkan emosi atau sikap pengguna terhadap topik yang dibahas. Label yang digunakan terbagi menjadi dua kategori utama, yaitu sentimen positif dan negatif. Sentimen positif menunjukkan penilaian atau perasaan yang baik terhadap sepatu preloved, sedangkan sentimen negatif menunjukkan ketidakpuasan, penolakan, atau kritik terhadapnya.

Tabel 4.1 Dataset

No	Komentar	Nama Barang	Label
1	Sol lem nya Ngelupas	Sneaker Casual	Negatif
2	model sesuai bahan berasa bagus dan lembut empuk agak berat ukuran lbh kecil dr sepatu dgn ukuran yg sama,	Aerostreat Casual 37	Positif
3	Size Sepatu nya malah kecil banget	Onit Onsu Tiger	Negatif
4	Murah Bagus Bingitss	Tuta Casual Shoes	Positif

Label sentimen ini berfungsi sebagai target atau ground truth dalam pelatihan model klasifikasi sentimen berbasis pembelajaran mesin. Akurasi dan kemampuan model sangat bergantung pada kualitas dan keakuratan pelabelan. Oleh karena itu, proses penge labelan dilakukan secara hati-hati untuk mengurangi kesalahan dan subjektivitas dalam klasifikasi.

Sebelum digunakan untuk pelatihan dan evaluasi model, dataset harus melalui beberapa tahap pra-pemrosesan. Proses ini mencakup pembersihan teks dari karakter tidak relevan, penghapusan tanda baca dan kata-kata umum (stopwords), normalisasi teks, serta tokenisasi. Tahapan pra-pemrosesan ini penting agar kualitas data meningkat dan model dapat memahami pola dalam teks secara optimal. Dengan demikian, hasil klasifikasi yang dihasilkan akan lebih akurat dan dapat diandalkan untuk pengambilan keputusan selanjutnya.

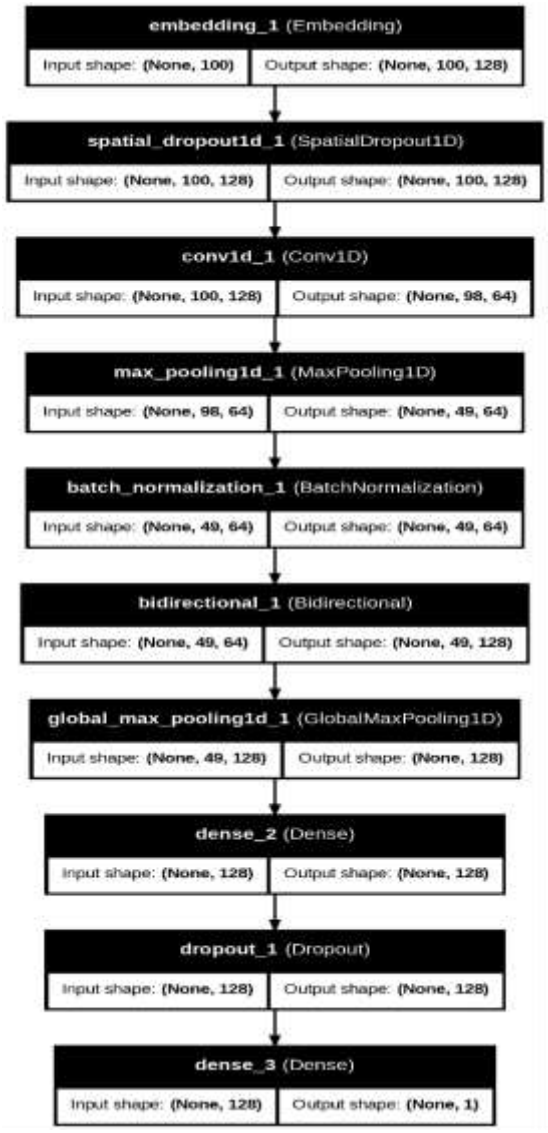
4.2. Prepocessing Dataset

Data yang digunakan dalam tahapan pemrosesan sebanyak 1.005 data. Kemudian, data tersebut diberi label secara manual dengan nilai positif dan negatif. Dari proses pelabelan tersebut didapatlah data kelas positif sebanyak 636 dan data kelas negatif sebanyak 369 data. Keseimbangan antara kelas positif dan negatif sangat berpengaruh terhadap akurasi yang didapatkan. Oleh karena itu jumlah dari kelas positif dikurangi menjadi 369 data agar seimbang terhadap kelas negatif. Total dataset yang digunakan pada akhirnya berjumlah 738 data. Dataset tersebut akan dibagi menjadi data training dan data testing dengan pembagian data training sebesar 80% dan data testing sebesar 20%. Pembagian yang telah dilakukan dapat dilihat pada Tabel 4.2.

Tabel 4.2. Pembagian Data Acak

Data	Jumlah
Training	591
Testing	147

4.3. Arsitektur Model LSTM



Gambar 4.1 Arsitektur LSTM

a. Layer Embedding (embedding_1)

Fungsi: Mengubah input teks (yang sudah di-tokenisasi dan dikonversi menjadi angka) ke dalam representasi vektor berdimensi 128.

Input shape: (None, 100) → 100 token per input (maksimal panjang teks).

Output shape: (None, 100, 128) → Tiap token diubah menjadi vektor 128 dimensi.

b. Layer SpatialDropout1D (spatial_dropout1d_1)

Fungsi: Mengurangi overfitting dengan secara acak menghilangkan seluruh fitur embedding pada posisi tertentu.

Output shape: Tetap (None, 100, 128).

c. Layer Conv1D (conv1d_1)

Fungsi: Mendeteksi fitur lokal (seperti n-gram) dari teks.

Output shape: (None, 98, 64) → Panjang urutan berkurang karena penggunaan kernel (misal ukuran 3), menghasilkan 64 filter.

d. Layer MaxPooling1D (max_pooling1d_1)

Fungsi: Mengurangi dimensi fitur dengan mengambil nilai maksimum (downsampling).

Output shape: (None, 49, 64) → Panjang sequence berkurang.

e. Layer BatchNormalization (batch_normalization_1)

Fungsi: Menormalkan output agar pelatihan lebih stabil dan cepat.

Output shape: Tetap (None, 49, 64).

f. Layer Bidirectional (bidirectional_1)

Fungsi: Layer RNN dua arah (kemungkinan LSTM/GRU) untuk

menangkap konteks dari dua arah (maju & mundur).

Output shape: (None, 49, 128) → Gabungan output forward dan backward RNN (masing-masing 64 unit).

g. Layer GlobalMaxPooling1D (global_max_pooling1d_1)

Fungsi: Mengambil nilai maksimum dari setiap fitur di seluruh time step, mereduksi dimensi.

Output shape: (None, 128) → 1 vektor per teks.

h. Layer Dense (dense_2)

Fungsi: Layer fully connected, mungkin dengan aktivasi ReLU.

Output shape: (None, 128).

i. Layer Dropout (dropout_1)

Fungsi: Regularisasi untuk mencegah overfitting dengan cara menonaktifkan neuron secara acak saat training.

Output shape: (None, 128).

j. Layer Dense (dense_3)

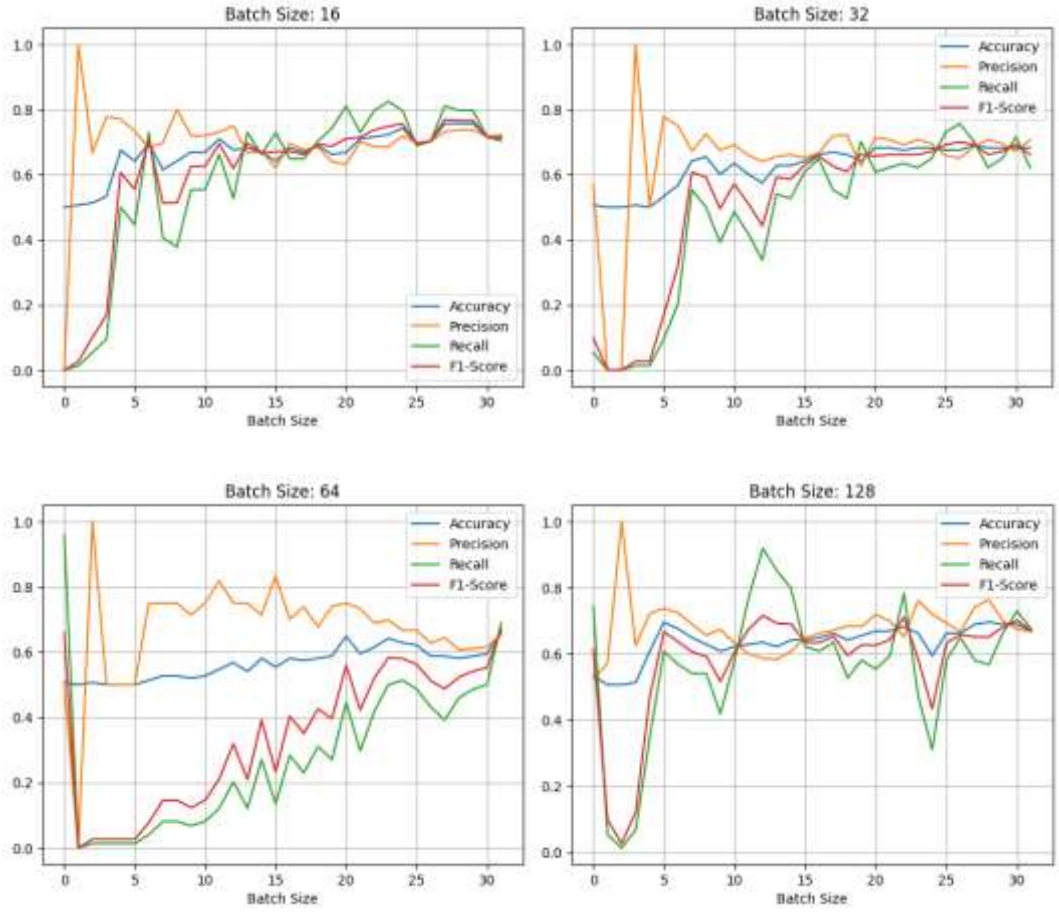
Fungsi: Layer output, kemungkinan besar dengan aktivasi sigmoid (karena output hanya 1).

Output shape: (None, 1) → Output probabilitas untuk klasifikasi biner (positif/negatif, spam/tidak, dsb.).

4.4. Pengujian LSTM Berdasarkan Nilai Batch Size

Pengujian batch size dilakukan dengan mengubah besar nilai batch size dalam rentang tertentu untuk mendapatkan nilai optimal. Batch size sendiri adalah pembagian jumlah sampel data yang disebarkan pada sebuah neural network. Pada pengujian ini, ditetapkan neuron sebanyak 64, epoch sebanyak 32 dan nilai batch

size yang akan diujikan, yaitu 16, 32, 64, dan 128. Dari pengujian tersebut didapatlah hasil yang ditunjukkan pada Gambar 4.2 dan Tabel 4.3.



Gambar 4.2 History Pengujian LSTM Berdasarkan Batch Size

a. Batch Size: 16

1. Terdapat fluktuasi yang cukup tajam, khususnya pada Recall (garis hijau) yang menunjukkan naik-turun signifikan pada awal hingga pertengahan batch.
2. Precision (garis oranye) cenderung tinggi namun sedikit tidak stabil di awal.
3. Accuracy dan F1-Score menunjukkan tren yang stabil namun tetap berfluktuasi.

4. Ini mengindikasikan bahwa batch size kecil membuat model sangat sensitif terhadap perbedaan kecil pada data, sehingga menghasilkan performa yang lebih variatif antar batch.

b. Batch Size: 32

1. Fluktuasi pada seluruh metrik mulai berkurang, khususnya pada Recall dan F1-Score.
2. Keempat metrik terlihat lebih konvergen dan stabil seiring bertambahnya batch, menunjukkan proses pelatihan yang lebih seimbang.
3. Precision tetap tinggi secara konsisten, dan Accuracy menunjukkan perbaikan yang stabil.
4. Batch size 32 merupakan ukuran yang cukup ideal untuk menyeimbangkan sensitivitas model dan kestabilan pelatihan.

c. Batch Size: 64

1. Fluktuasi kembali meningkat, terutama pada Recall dan F1-Score yang bergerak secara lambat dan tidak konvergen secara cepat.
2. Precision menunjukkan nilai yang tinggi namun sangat fluktuatif pada batch awal.
3. Accuracy cenderung stabil namun sedikit lebih rendah dibanding batch size 32.
4. Semakin besarnya batch, model memerlukan waktu lebih lama untuk konvergen, karena pembaruan bobot yang dilakukan lebih lambat dan kurang responsif terhadap data minoritas

d. Batch Size: 128

1. Fluktuasi sangat terlihat jelas di awal batch, terutama pada Recall yang sangat tidak stabil.
2. Seiring bertambahnya batch, metrik mulai stabil dan berkumpul, namun performa secara umum cenderung rata-rata (tidak ada yang menonjol).
3. Precision tetap cukup tinggi dan stabil, tapi Accuracy dan F1-Score memerlukan waktu lebih lama untuk mencapai konsistensi.
4. Hal ini menunjukkan bahwa batch terlalu besar bisa menyebabkan kurangnya kemampuan adaptasi terhadap detail-detail penting dari data, sehingga memperlambat proses pembelajaran.

e. Kesimpulan

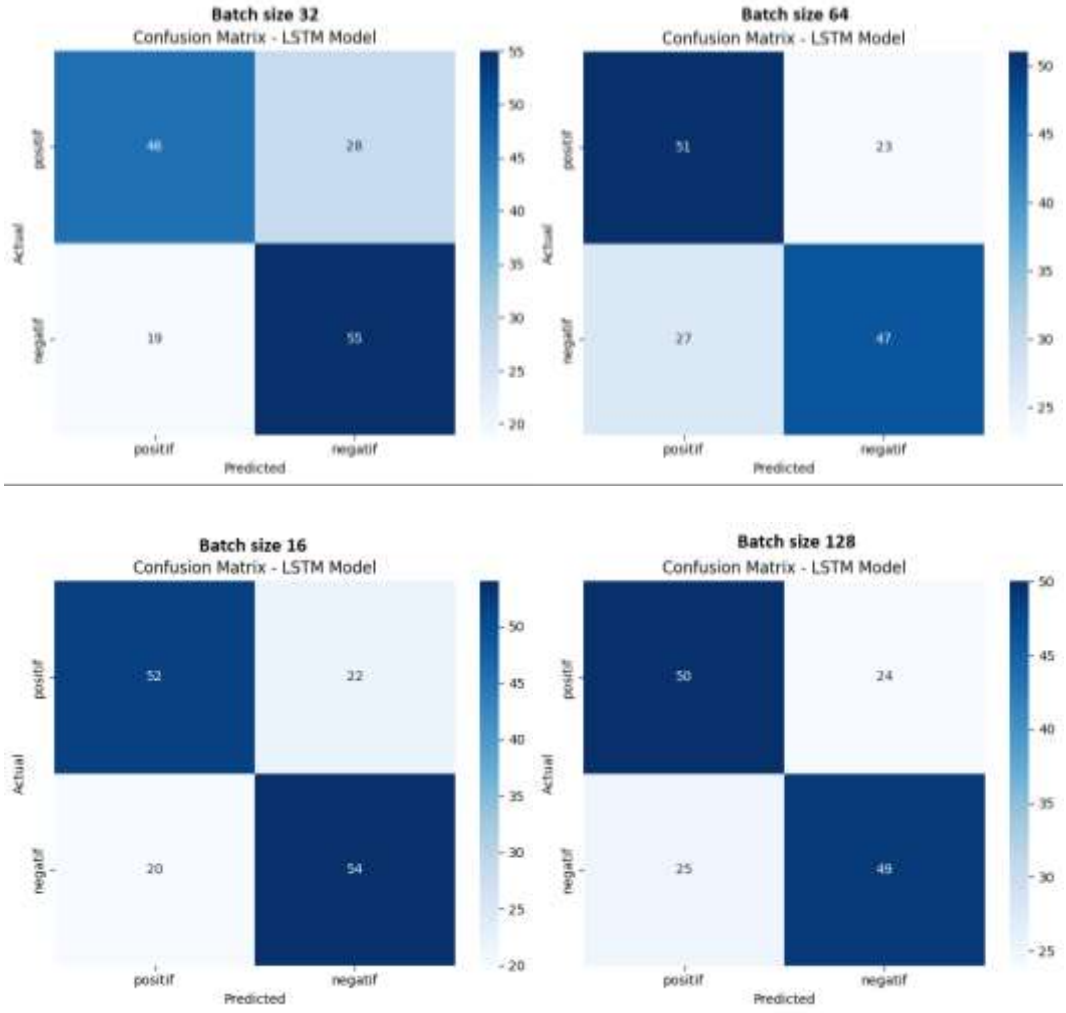
1. Batch size 16 responsif, tapi sangat fluktuatif.
2. Batch size 32 memberikan kombinasi terbaik antara stabilitas dan performa.
3. Batch size 64 dan 128 stabil tapi cenderung lambat belajar, dan bisa kehilangan variasi data penting.

Tabel 4.3 Pengujian LSTM Berdasarkan Batch Size

Batch Size	Akurasi (%)	Recall (%)	Presisi (%)	F-measure (%)	Waktu (detik)
16	71.62	70.27	72.22	71.23	73
32	68.24	62.16	70.77	66.19	54
64	66.22	68.92	65.38	67.11	40
128	66.89	67.57	66.67	67.11	40

Dapat dilihat dari Tabel 4.3 bahwa hasil akurasi tertinggi diperoleh pada batch size 16 dengan akurasi sebesar 71.62%. Nilai recall tertinggi juga didapat pada batch size yang sama, yaitu sebesar 70.27%. Nilai presisi tertinggi juga didapat

pada batch size yang sama, yaitu sebesar 72.22%. Nilai f-measure tertinggi juga didapat pada batch size yang sama, yaitu sebesar 71.23%. Batch size 16 memerlukan waktu paling lama yaitu 73 detik, sedangkan batch size 128 hanya membutuhkan waktu 40 detik. Hal ini menunjukkan bahwa peningkatan ukuran batch dapat mempercepat proses pelatihan, meskipun perlu diperhatikan bahwa akurasi dan metrik evaluasi lainnya tidak selalu meningkat seiring dengan bertambahnya batch size.



Gambar 4.3 Confusion Matrix Pengujian LSTM Berdasarkan Batch Size

Batch Size 16 mengklasifikasikan True Positif (TP) 52, True Negatif (TN) 54, False Positif (FP) 22, False Negatif (FN) 20. Model menunjukkan kinerja cukup

baik pada batch size ini dengan ketepatan prediksi yang seimbang antara kedua kelas.

Batch Size 32 mengklasifikasikan True Positif (TP) 46, True Negatif (TN) 55, False Positif (FP) 28, False Negatif (FN) 13. Model menunjukkan peningkatan dalam memprediksi data negatif secara benar (TN meningkat), namun terjadi peningkatan kesalahan pada data negatif yang diklasifikasikan sebagai positif (FP meningkat). Ini mengindikasikan bahwa model cenderung lebih bias terhadap kelas positif.

Batch Size 64 mengklasifikasikan True Positif (TP) 51, True Negatif (TN) 47, False Positif (FP) 27, False Negatif (FN) 23. Pada batch 64, terjadi penurunan pada jumlah prediksi benar baik pada kelas positif maupun negatif. Ini menunjukkan adanya ketidakseimbangan kinerja model, dan tingkat kesalahan klasifikasi meningkat dibandingkan batch size sebelumnya.

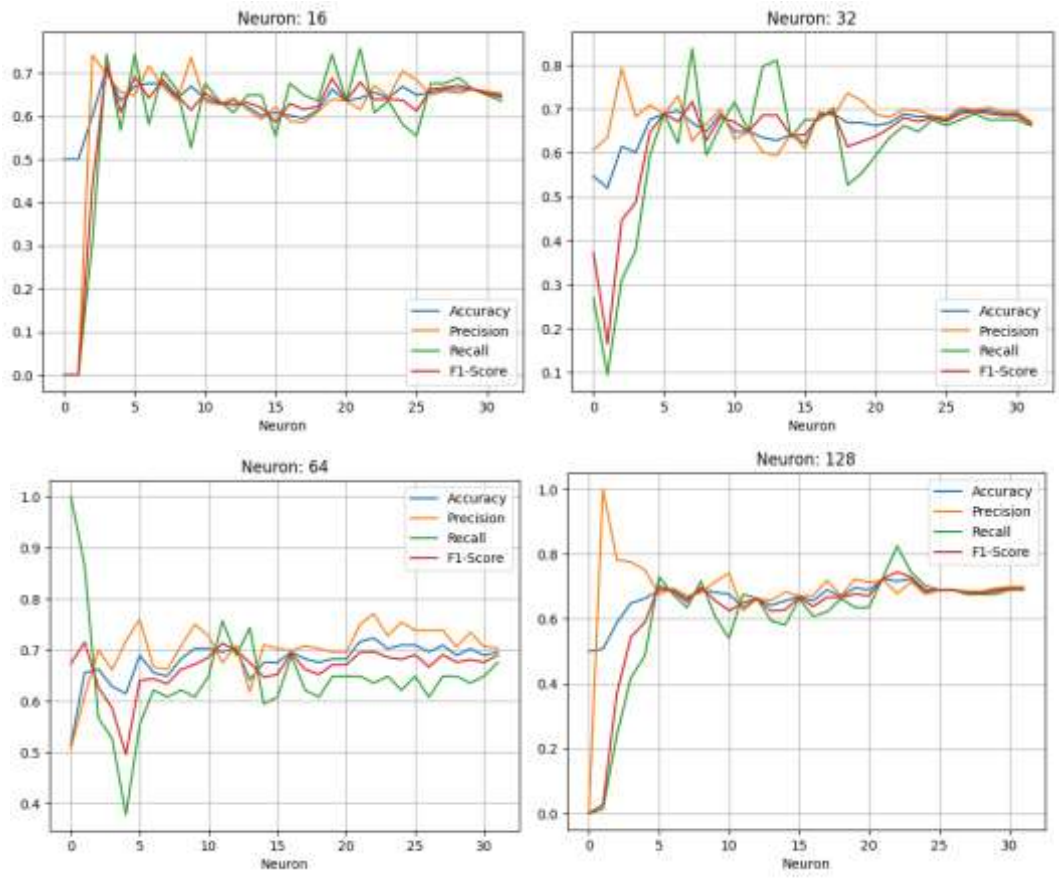
Batch Size 128 mengklasifikasikan True Positif (TP) 50, True Negatif (TN) 49, False Positif (FP) 25, False Negatif (FN) 24. Model menunjukkan kinerja yang relatif stabil namun tidak sebaik batch size 16 atau 32. Nilai TP dan TN sedikit menurun, sedangkan FN relatif tinggi, yang berarti masih cukup banyak data positif yang tidak dikenali dengan benar.

Kesimpulan batch size 16 menghasilkan performa terbaik secara keseluruhan dengan nilai TP dan TN yang tinggi serta kesalahan yang relatif rendah. Batch size 32 juga memiliki performa yang baik, khususnya dalam mengenali data true negatif (TN) tinggi, namun cenderung menghasilkan lebih banyak False Positif (FP). Peningkatan batch size di atas 64 justru menurunkan performa model,

mengindikasikan bahwa ukuran batch yang terlalu besar dapat mengurangi sensitivitas model terhadap pola pada data pelatihan.

4.5. Pengujian LSTM Berdasarkan Nilai Neuron

Pengujian neuron dilakukan dengan mengubah besar nilai neuron dalam rentang tertentu untuk mendapatkan nilai optimal. Pada pengujian ini, ditetapkan batch size sebesar 16 dan epoch sebanyak 32 karena berdasarkan pengujian sebelumnya mendapatkan nilai akurasi tertinggi pada neuron 64. Pada pengujian ini, nilai neuron akan ditetapkan sebanyak 16, 32, 64, 128. Dari pengujian tersebut didapatlah hasil yang ditunjukkan pada Gambar 4.4 dan Tabel 4.4.



Gambar 4.4 History Pengujian LSTM Berdasarkan Nilai Neuron

a. Neuron: 16

1. Model menunjukkan stabilitas yang baik setelah batch ke-5.

2. Semua metrik (Accuracy, Precision, Recall, F1-Score) berkumpul cukup rapat dan menunjukkan performa yang seimbang.
3. Terdapat sedikit fluktuasi, terutama pada Recall (hijau) yang masihnaik-turun secara ringan.
4. Jumlah neuron yang sedikit ini menunjukkan bahwa model tetap mampu belajar dengan cukup baik, namun mungkin belum optimal untuk menangkap kompleksitas data sepenuhnya.

b. Neuron: 32

1. Performa model meningkat dibanding neuron 16.
2. Recall dan Precision menunjukkan nilai yang lebih tinggi dan relatif stabil.
3. Accuracy dan F1-Score juga menunjukkan stabilitas yang lebih baik, meskipun Recall masih agak fluktuatif di beberapa batch awal.
4. Ini menunjukkan bahwa dengan jumlah neuron yang cukup (tidak terlalu kecil), model mampu menangkap pola-pola penting dari data dengan baik.

c. Neuron: 64

1. Terlihat fluktuasi yang signifikan, terutama pada Recall dan Precision.
2. Meskipun Accuracy dan F1-Score tidak terlalu turun, nilai keduanya cenderung stagnan.
3. Hal ini mengindikasikan bahwa terlalu banyak neuron bisa menyebabkan overfitting atau kesulitan dalam generalisasi, terutama jika data atau arsitektur tidak mendukung kompleksitas tersebut.
4. Model terlihat tidak stabil, sehingga 64 neuron mungkin terlalu besar

untuk arsitektur atau dataset ini.

d. Neuron: 128

1. Fluktuasi di awal sangat jelas, terutama pada Recall.
2. Namun setelah batch ke-10, seluruh metrik mulai stabil dan konvergen.
3. Precision terlihat konsisten tinggi, dan F1-Score menunjukkan performa yang baik.
4. Ini mengindikasikan bahwa meskipun 128 neuron memerlukan lebih banyak waktu untuk konvergen, namun setelah stabil, performanya sangat kompetitif dan seimbang.
5. Dengan catatan: waktu pelatihan dan beban komputasi juga meningkat.

e. Kesimpulan

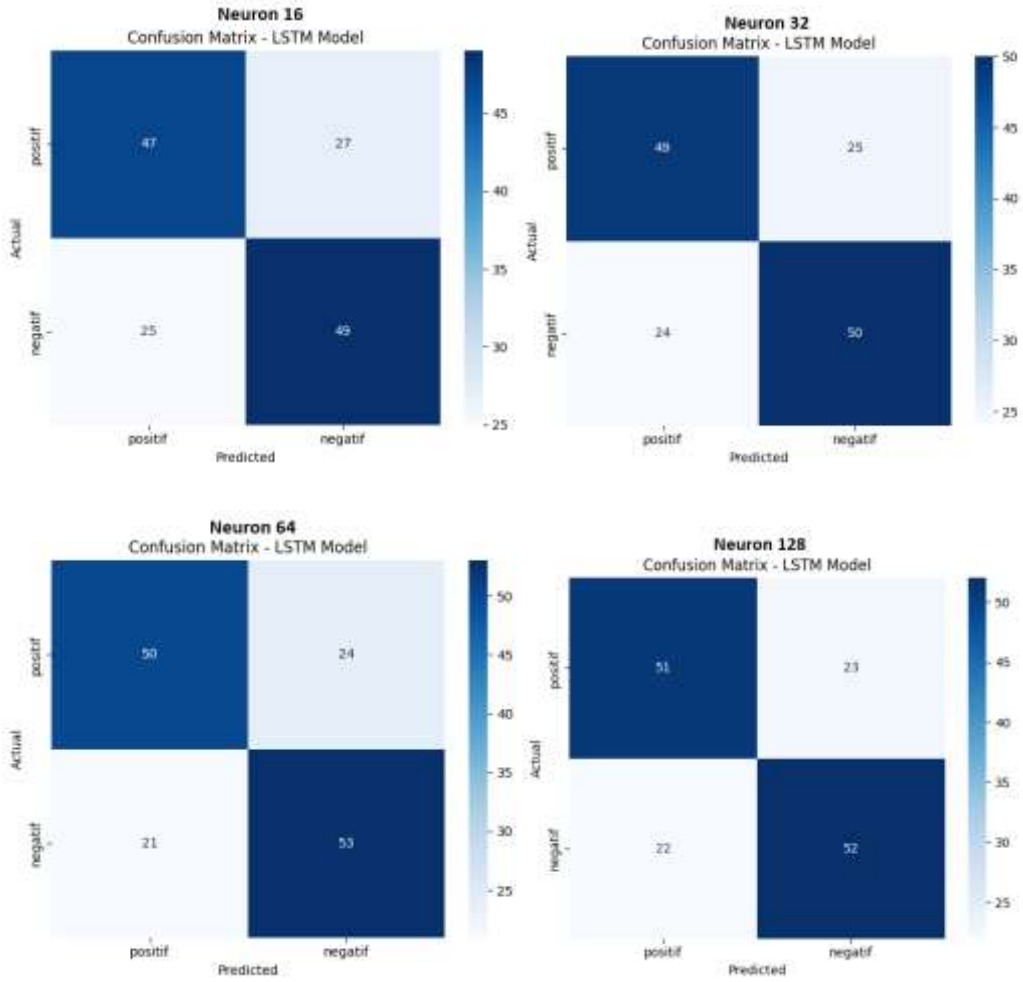
1. Neuron 16 stabil dan ringan, namun performa kurang maksimal.
2. Neuron 32 keseimbangan terbaik antara performa dan stabilitas.
3. Neuron 64 fluktuatif, performa cenderung stagnan, kemungkinan overfitting ringan.
4. Neuron 128 performa baik setelah stabil, tapi memerlukan waktu konvergensi lebih lama.

Tabel 4.4 Pengujian LSTM Berdasarkan Nilai Neuron

Neuron	Akurasi (%)	Recall (%)	Presisi (%)	F-measure (%)	Waktu (detik)
16	64.86	63.51	65.28	64.38	51
32	66.89	66.22	67.12	66.67	66
64	69.59	67.57	70.42	68.97	74
128	69.59	68.92	69.86	69.39	131

Dapat dilihat dari Tabel 4.4 di atas bahwa hasil akurasi dan presisi tertinggi diperoleh pada neuron 64, dengan akurasi sebesar 69.59% dan presisi sebesar

70.42%. Sedangkan recall dan f-measure tertinggi didapat pada neuron 128, dengan nilai recall sebesar 68.92% dan f-measure sebesar 69.39%. Berdasarkan waktu yang diperlukan, semakin besar jumlah neuron cenderung membutuhkan waktu pemrosesan yang lebih lama. Neuron 16 memerlukan waktu 51 detik dan neuron 128 memerlukan waktu paling lama yaitu 131 detik. Hal ini menunjukkan bahwa peningkatan jumlah neuron dapat meningkatkan kompleksitas model dan berdampak pada waktu pelatihan yang lebih panjang, meskipun tidak selalu menjamin peningkatan performa secara signifikan.



Gambar 4.5 Confusion Matrix Pengujian LSTM Berdasarkan Neuron

Neuron 16 mengklasifikasikan True Positif (TP) 47, True Negatif (TN) 49, False Positif (FP) 23, False Negatif (FN) 25. Model dengan 16 neuron menunjukkan kinerja awal yang cukup stabil, terutama dalam mengenali kelas negatif (TN tinggi). Namun, masih terdapat cukup banyak kesalahan dalam mengklasifikasikan positif (FN dan FP masih cukup tinggi), menunjukkan ruang perbaikan dalam sensitivitas terhadap kelas positif.

Neuron 32 mengklasifikasikan True Positif (TP) 49, True Negatif (TN) 50, False Positif (FP) 24, False Negatif (FN) 23. Model ini menunjukkan peningkatan akurasi dibanding sebelumnya, dengan nilai TP dan TN yang sedikit meningkat dan FN yang menurun. Hal ini menandakan model mulai belajar mengenali pola lebih baik, terutama dalam mengurangi kesalahan pada prediksi kelas positif.

Neuron 64 mengklasifikasikan True Positif (TP) 50, True Negatif (TN) 50, False Positif (FP) 24, False Negatif (FN) 21. Model pada neuron 64 memberikan hasil paling seimbang dan optimal di antara konfigurasi lainnya. Nilai TP dan TN tinggi, dan kesalahan (FN dan FP) mulai menurun. Ini menunjukkan bahwa dengan jumlah neuron yang lebih banyak, model mampu menangkap representasi fitur yang lebih baik tanpa overfitting.

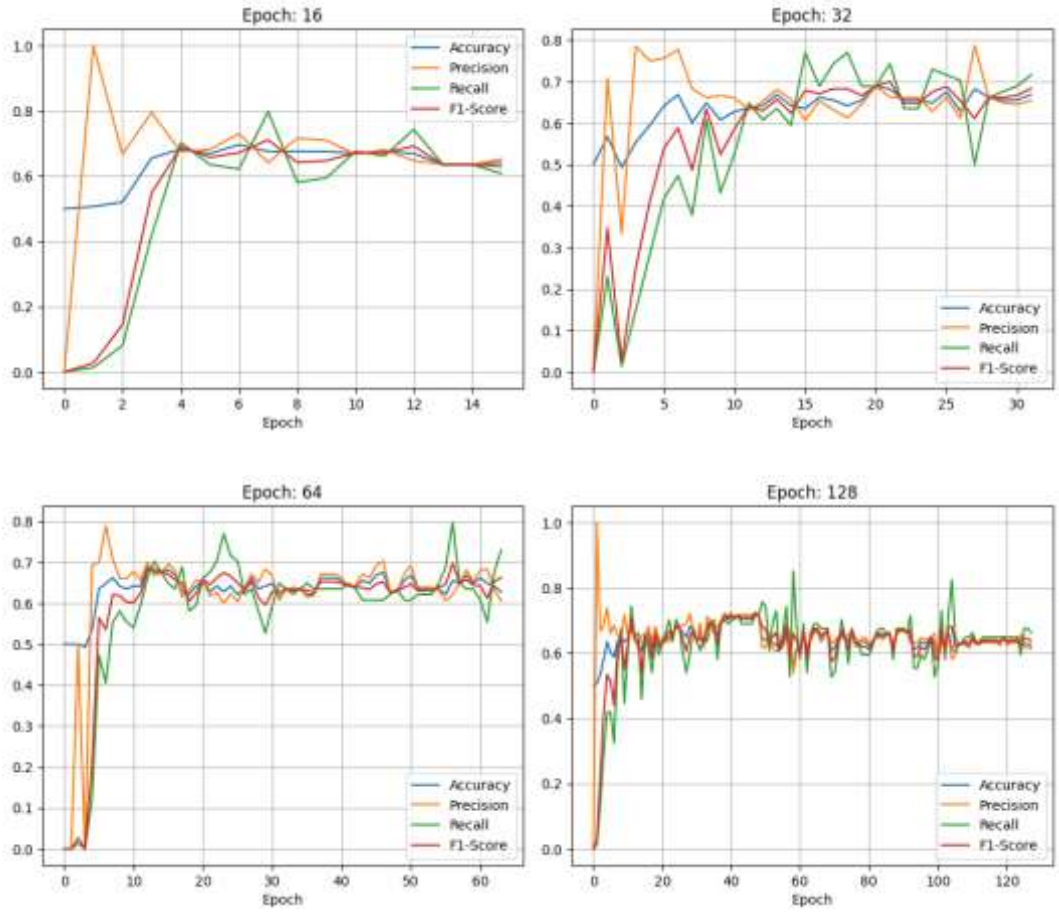
Neuron 128 mengklasifikasikan True Positif (TP) 51, True Negatif (TN) 52, False Positif (FP) 22, False Negatif (FN) 23. Model pada neuron 128 menghasilkan TP dan TN tertinggi, serta FP paling rendah. Namun, nilai FN sedikit naik dibandingkan neuron 64. Ini bisa mengindikasikan awal mula overfitting, di mana model terlalu fokus pada data pelatihan dan kehilangan generalisasi pada data uji.

Kesimpulan model dengan neuron 64 memberikan performa paling optimal, dengan keseimbangan antara TP dan TN serta tingkat kesalahan klasifikasi yang

rendah. Neuron 128 menunjukkan performa tinggi juga, tetapi mulai menunjukkan gejala overfitting dengan naiknya FN. Model dengan 16 dan 32 neuron masih berada dalam tahap awal pembelajaran, dengan performa yang cenderung stabil namun belum optimal.

4.6. Pengujian LSTM Berdasarkan Nilai Epoch

Pengujian berdasarkan nilai epoch dilakukan dengan mengubah besar nilai epoch dalam rentang tertentu untuk mendapatkan nilai optimal. Pada pengujian ini, ditetapkan neuron sebanyak 16, batch size sebanyak 16 dan nilai epoch akan dimulai dari 16, 32, 64, dan 128. Dari pengujian tersebut didapatlah hasil yang ditunjukkan pada Gambar 4.6 dan Tabel 4.5.



Gambar 4.6 History Pengujian LSTM Berdasarkan Nilai Epoch

a. Epoch: 16

1. Fluktuasi tinggi di awal pelatihan, terutama pada Precision (oranye) dan Recall (hijau).
2. Setelah sekitar epoch ke-5, semua metrik mulai stabil namun belum optimal.
3. F1-Score (merah) dan Accuracy (biru) menunjukkan tren mendatar tanpa banyak peningkatan setelah awal pelatihan.
4. Jumlah epoch ini terlalu sedikit, sehingga model belum mencapai performa optimal (underfitting).

b. Epoch: 32

1. Fluktuasi masih terlihat di awal, tetapi metrik mulai meningkat dan stabil pada pertengahan epoch.
2. Semua metrik terutama Recall dan F1-Score meningkat lebih signifikan dibanding epoch 16.
3. Precision juga stabil di atas metrik lain, menunjukkan model bisa memprediksi dengan benar kelas positif.
4. Ini mengindikasikan bahwa epoch 32 memberikan performa yang lebih seimbang dan optimal dibanding sebelumnya.

c. Epoch: 64

1. Semua metrik sangat stabil setelah epoch ke-10.
2. Recall dan F1-Score mengalami kenaikan signifikan dan tetap stabil.
3. Fluktuasi sangat kecil, menunjukkan model telah belajar dengan cukup baik.
4. Ini menandakan epoch 64 memberikan waktu pelatihan cukup untuk

mencapai stabilitas dan performa tinggi

d. Epoch: 128

1. Fluktuasi di awal sangat jelas terutama pada Recall.
2. Namun setelah sekitar epoch ke-30, seluruh metrik stabil dan berkumpul.
3. Precision tetap tinggi, dan F1-Score serta Accuracy menunjukkan kestabilan jangka panjang.
4. Jumlah epoch ini mungkin berlebihan jika model sudah konvergen lebih awal, tetapi tetap menghasilkan performa yang sangat baik.

e. Kesimpulan

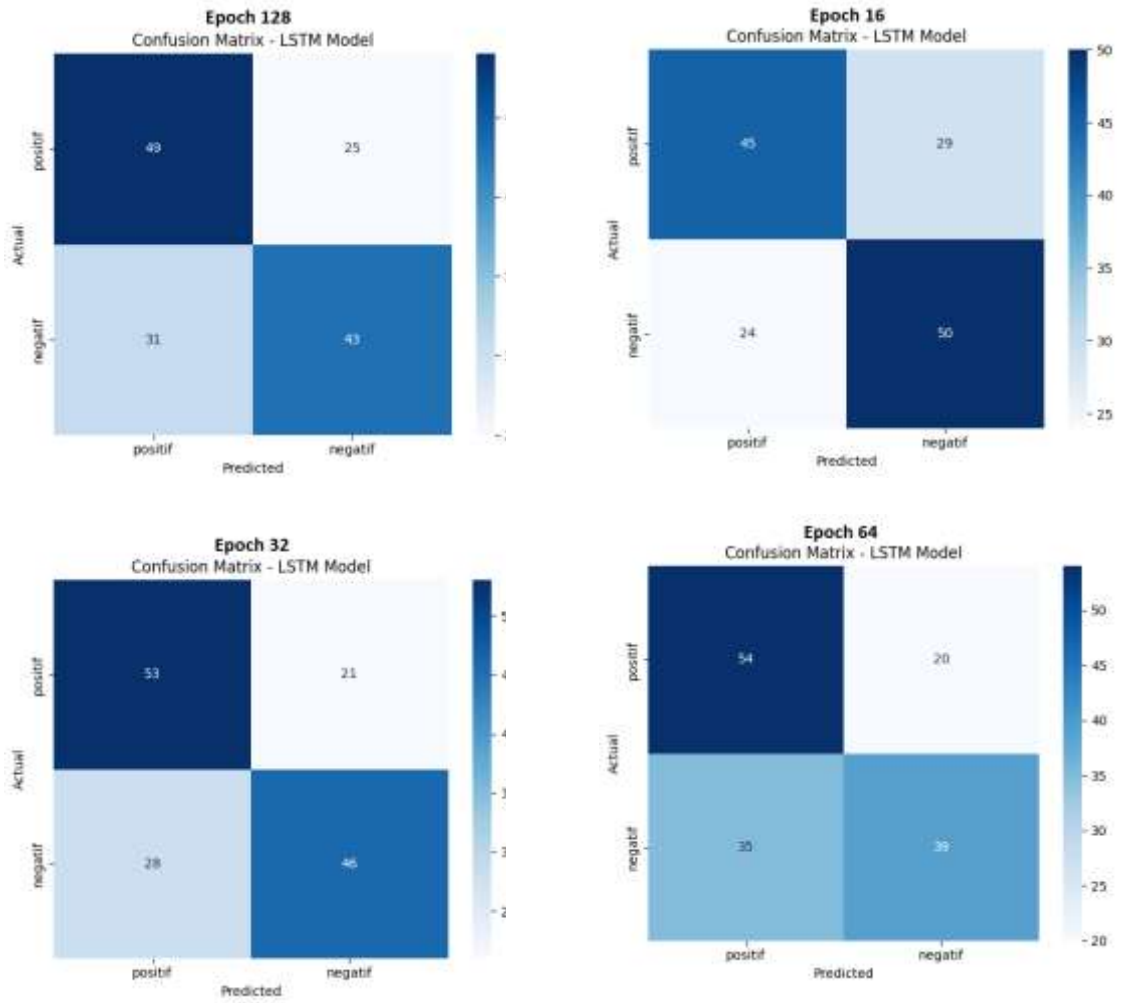
1. Epoch 16 terlalu singkat, model belum konvergen.
2. Epoch 32 meningkatkan performa secara signifikan dibanding 16, cocok untuk pelatihan awal.
3. Epoch 64 kombinasi optimal antara kestabilan dan performa.
4. Epoch 128 memberikan performa terbaik, tapi efisiensi pelatihan menurun karena model sudah stabil sebelum akhir.

Tabel 4.5 Pengujian LSTM Berdasarkan Nilai Epoch

Epoch	Akurasi (%)	Recall (%)	Presisi (%)	F-measure (%)	Waktu (detik)
16	64.19	60.81	65.22	62.94	41
32	66.89	71.62	65.43	68.39	73
64	62.84	72.97	60.67	66.26	141
128	62.16	66.22	61.25	63.64	286

Tabel 4.5 Pengujian LSTM Berdasarkan Nilai Epoch menunjukkan hasil evaluasi model berdasarkan variasi jumlah epoch. Dapat dilihat bahwa nilai akurasi dan presisi tertinggi diperoleh pada epoch 32, dengan akurasi sebesar 66,89% dan

presisi sebesar 65,43%. Sementara itu, nilai recall dan f-measure tertinggi juga dicapai pada epoch 32, masing-masing sebesar 71,62% dan 68,39%. Dari sisi waktu pelatihan, terlihat bahwa semakin besar jumlah epoch, maka semakin lama waktu yang dibutuhkan. Epoch 16 hanya memerlukan waktu 41 detik, sedangkan epoch 128 membutuhkan waktu paling lama yaitu 286 detik. Hal ini menunjukkan bahwa penambahan epoch memang dapat meningkatkan performa model hingga titik tertentu, namun juga menyebabkan peningkatan waktu pelatihan secara signifikan. Oleh karena itu, pemilihan jumlah epoch perlu mempertimbangkan trade-off antara performa model dan efisiensi waktu pelatihan.



Gambar 4.7 Confusion Matrix Pengujian LSTM Berdasarkan Epoch

Epoch 16 mengklasifikasikan True Positif (TP) 45, True Negatif (TN) 50, False Positif (FP) 21, False Negatif (FN) 24. Model pada epoch 16 menunjukkan performa yang cukup stabil dengan nilai TN yang tinggi, menandakan model cukup baik dalam mengenali kelas negatif. Namun, masih terdapat kesalahan yang cukup banyak dalam mengenali kelas positif (FN cukup tinggi).

Epoch 32 mengklasifikasikan True Positif (TP) 48, True Negatif (TN) 53, False Positif (FP) 21, False Negatif (FN) 18. Terdapat peningkatan performa dibanding epoch 16, baik pada kelas positif maupun negatif. Model lebih seimbang dalam mengenali kedua kelas, dan jumlah kesalahan (FN dan FP) cenderung menurun. Ini menunjukkan model semakin baik dalam proses pembelajaran.

Epoch 64 mengklasifikasikan True Positif (TP) 54, True Negatif (TN) 35, False Positif (FP) 29, False Negatif (FN) 20. Pada epoch 64, meskipun nilai TP meningkat, namun nilai TN justru menurun cukup drastis. Hal ini menunjukkan model menjadi lebih condong terhadap kelas positif dan cenderung lebih sering salah dalam mengenali data negatif (FP meningkat signifikan).

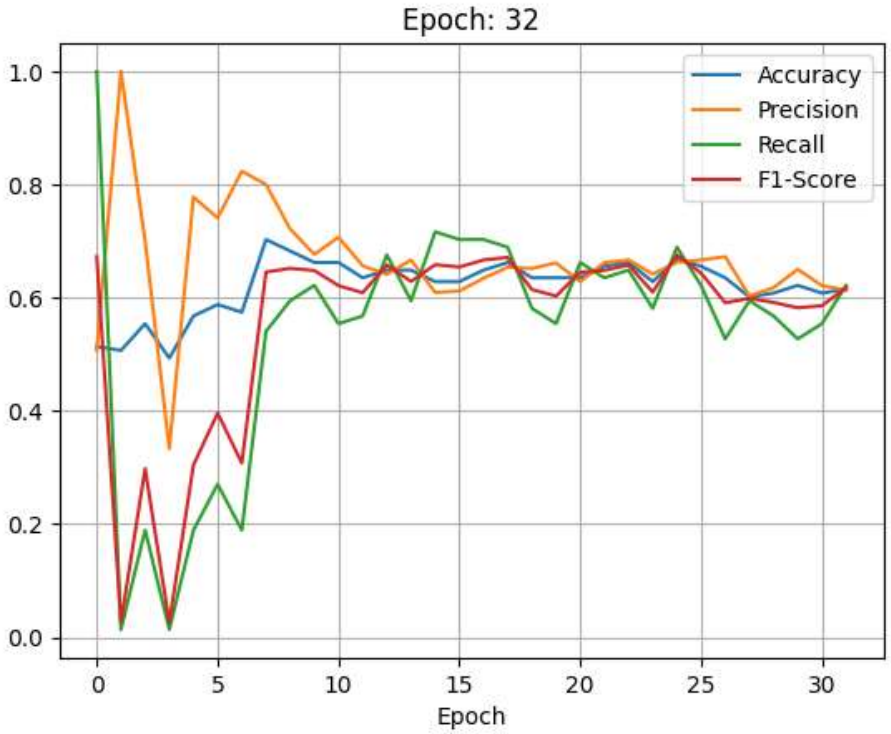
Epoch 128 mengklasifikasikan True Positif (TP) 49, True Negatif (TN) 41, False Positif (FP) 25, False Negatif (FN) 23 Model pada epoch 128 mengalami penurunan performa dibanding epoch sebelumnya. Baik TP maupun TN sedikit menurun, sementara FN dan FP cukup tinggi. Ini mengindikasikan overfitting atau model kehilangan kemampuan generalisasi setelah terlalu banyak epoch.

Kesimpulan epoch 32 memberikan performa terbaik secara keseluruhan, dengan keseimbangan antara TP dan TN serta kesalahan klasifikasi (FP dan FN) yang relatif rendah. Sementara itu, epoch 64 dan 128 menunjukkan adanya

penurunan performa akibat ketidakseimbangan klasifikasi atau overfitting. Epoch 16 cukup stabil, namun belum sebaik epoch 32.

4.7. Pengujian Algoritma LSTM Dengan Parameter Terbaik

Setelah dilakukan serangkaian pengujian terhadap model Long Short-Term Memory (LSTM) dengan berbagai kombinasi parameter, diperoleh konfigurasi terbaik yang menghasilkan performa paling optimal dalam mengklasifikasikan data. Adapun parameter terbaik yang digunakan dalam pengujian akhir ini adalah sebagai berikut: Batch Size: 16, Jumlah Neuron: 64, dan Epoch: 32.



Gambar 4.8 Pengujian Algoritma LSTM Dengan Parameter Terbaik

Tabel 4.6 Pengujian Algoritma LSTM

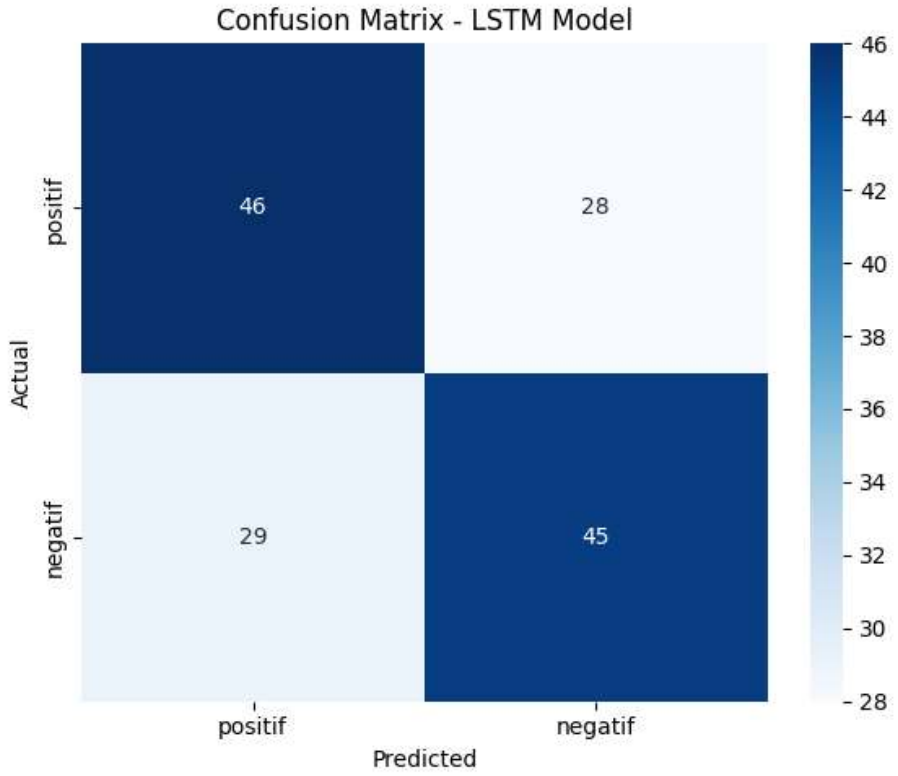
Akurasi (%)	Recall (%)	Presisi (%)	F-measure (%)	Waktu (Detik)
61.49	62.16	61.33	61.74	73

Stabilitas Model, pada awal pelatihan (epoch 0–5), keempat metrik menunjukkan fluktuasi yang sangat tajam, terutama precision dan recall yang sempat mendekati angka 0. Hal ini wajar karena model masih dalam tahap awal pembelajaran dan bobot-bobot jaringan belum optimal. Setelah melewati epoch ke-5, grafik mulai menunjukkan pola yang lebih stabil, menandakan bahwa model mulai mengenali pola data secara lebih baik.

Akurasi (Accuracy), nilai akurasi meningkat tajam pada awal pelatihan dan cenderung stabil di kisaran 0.6 hingga 0.65 setelah epoch ke-10. Hal ini menunjukkan bahwa model secara konsisten mampu mengklasifikasikan sekitar 60–65% data dengan benar.

Presisi (Precision), precision sempat sangat tinggi di awal, namun itu bisa jadi disebabkan oleh ketidakseimbangan prediksi di kelas tertentu (model hanya menebak satu kelas). Setelah stabil, precision bergerak di kisaran 0.6–0.7, menandakan bahwa dari semua data yang diprediksi sebagai positif, sekitar 60–70% benar-benar positif.

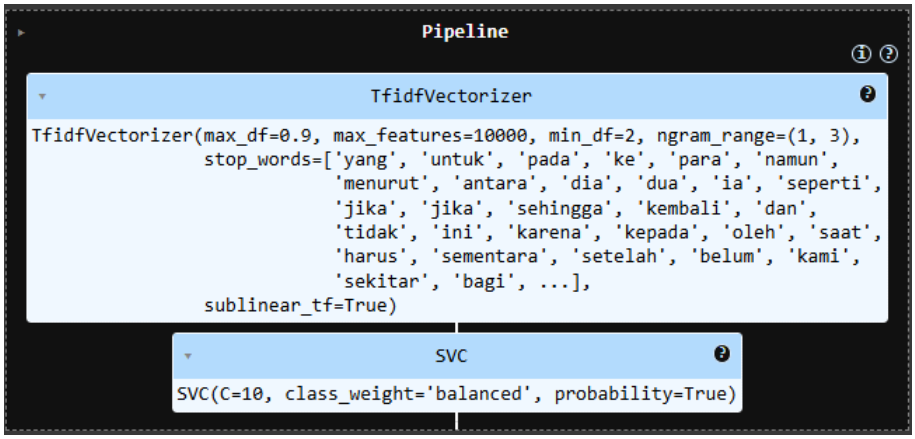
Recall, nilai recall juga mengalami kenaikan drastis setelah beberapa epoch dan cenderung fluktuatif tapi relatif stabil di sekitar 0.6. Ini menunjukkan bahwa model mampu menemukan sekitar 60% dari semua data yang benar-benar positif. F1-Score merupakan rata-rata harmonis antara precision dan recall, dan menunjukkan stabilitas yang cukup baik mulai dari epoch ke-10. Pergerakannya yang mendekati precision dan recall menandakan bahwa keduanya relatif seimbang.



Gambar 4.9 Confusion Matrix Pengujian LSTM Berdasarkan Parameter Terbaik

Model dengan parameter terbaik dapat mengklasifikasikan True Positif (TP) 46, True Negatif (TN) 49, False Positif (FP) 29, False Negatif (FN) 28. Model memiliki kinerja yang relatif seimbang antara kedua kelas, dengan jumlah TP (46) dan TN (45) yang cukup tinggi. Namun, nilai FN (28) dan FP (29) juga cukup signifikan, menunjukkan bahwa masih terdapat cukup banyak kesalahan klasifikasi. Performa model dalam mengenali data positif dan negatif hampir seimbang, namun masih ada ruang untuk perbaikan, terutama dalam menurunkan kesalahan prediksi silang antar kelas.

4.8. Arsitektur Model SVM



Gambar 4.10 Pipeline SVM

a. TfidfVectorizer

Merupakan fitur ekstraksi teks yang mengubah data teks menjadi representasi numerik berbasis TF-IDF (Term Frequency - Inverse Document Frequency). TF-IDF menilai seberapa penting suatu kata dalam sebuah dokumen relatif terhadap korpus keseluruhan. Parameter yang digunakan**:**

1. max_df=0.9. Mengabaikan kata-kata yang muncul di lebih dari 90% dokumen (umumnya stop words yang tidak informatif).
2. min_df=2. Mengabaikan kata-kata yang hanya muncul di 1 dokumen saja (tidak cukup representatif).
3. max_features=10000. Hanya menyimpan 10.000 fitur (kata/ngram) teratas berdasarkan skor TF-IDF.
4. ngram_range=(1, 3). Menggunakan unigram (kata tunggal), bigram (dua kata), dan trigram (tiga kata) sebagai fitur.
5. stop_words=[...]. Daftar kata umum dalam bahasa Indonesia yang dihapus karena tidak membawa informasi penting (contoh: "yang",

"untuk", "pada", "karena", dll.).

6. sublinear_tf=True. Menggunakan sublinear term frequency scaling (1 + log(tf)) agar kata yang sering muncul tidak terlalu mendominasi.

b. SVC (Support Vector Classification)

1. Merupakan algoritma klasifikasi berbasis Support Vector Machine (SVM) yang efektif untuk klasifikasi teks dan data dengan dimensi tinggi.
2. Parameter yang digunakan:
3. C=10. Parameter regularisasi. Nilai yang lebih tinggi menandakan model akan berusaha lebih keras untuk mengklasifikasikan data pelatihan dengan benar
4. (resiko overfitting sedikit lebih tinggi).
5. class_weight='balanced'. Menyesuaikan bobot kelas secara otomatis berdasarkan frekuensi kemunculan di dataset. Berguna jika data tidak seimbang (misal: data positif lebih sedikit daripada negatif).
6. probability=True. Mengaktifkan prediksi probabilitas (menggunakan metode seperti Platt scaling), berguna untuk ROC, thresholding, dsb.

4.9. Pengujian Algoritma Support Vector Machine

Pengujian Support Vector Machine dilakukan untuk memanfaatkan library sklearn pada python. Data yang digunakan pada klasifikasi ini dibagi menjadi data training sebanyak 80% dan data testing sebanyak 20%. Skenario pembagian data tersebut mampu menghasilkan nilai akurasi yang tinggi pada metode Support Vector Machine. Selanjutnya, metode yang digunakan dalam ekstraksi fitur adalah TF-IDF. Pada pengujian ini didapatlah hasil yang ditunjukkan pada Tabel 4.7**.**

Tabel 4.7 Pengujian Algoritma Support Vector Machine

Akurasi (%)	Recall (%)	Presisi (%)	F-measure (%)	Waktu (Detik)
68	68	68	68	17

Dapat dilihat pada Tabel 4.7, bahwa akurasi yang didapat sebesar 68%, recall 68%, presisi 68%, dan f-measure 68% dengan waktu pelatihan 17 detik.

4.10. Perbandingan Hasil Sentimen

Dalam pengujian model Long Short-Term Memory (LSTM), telah dilakukan berbagai percobaan dengan mengubah parameter-parameter penting seperti batch size, jumlah neuron, dan jumlah epoch. Berdasarkan hasil evaluasi menggunakan confusion matrix serta metrik evaluasi lainnya, diperoleh bahwa kombinasi terbaik untuk model LSTM terdapat pada: Batch Size: 16, Neuron: 64, dan Epoch: 32. Kombinasi ini menghasilkan keseimbangan performa terbaik, baik dalam mengklasifikasikan data positif maupun negatif secara akurat. Berikut uraian mendalamnya**.**

Tabel 4.8 Perbandingan Hasil Sentimen

Metode	Akurasi (%)	Recall (%)	Presisi (%)	F-measure (%)	Waktu (Detik)
LSTM	61.49	62.16	61.33	61.74	73
SVM	68	68	68	68	17

Pada Tabel 4.8, ditunjukkan hasil dari pengujian tiap metode. Berdasarkan nilai akurasinya, metode Support Vector Machine (SVM) memiliki nilai akurasi yang

lebih tinggi, yaitu 68%, dibandingkan dengan metode LSTM yang memiliki akurasi sebesar 61.49%. Hal ini menunjukkan bahwa dalam hal akurasi, SVM memiliki performa yang lebih baik dibandingkan LSTM dalam mengklasifikasikan data sentimen pada pengujian ini.

4.11. Website

Website berfungsi sebagai antarmuka (interface) untuk melakukan pengujian dan visualisasi hasil perbandingan sentimen menggunakan SVM dan LSTM. Website ini dibangun dengan memanfaatkan framework Flask (python) sebagai backend dan ReactJS sebagai tampilan antar muka dan terintegrasi langsung dengan model machine learning yang telah dilatih sebelumnya.

Melalui website ini, pengguna dapat memasukkan komentar atau kalimat yang ingin dianalisis pada sebuah kolom input. Setelah komentar dikirim, sistem akan memproses data tersebut menggunakan dua metode yang telah dibandingkan dalam penelitian ini, yaitu Long Short-Term Memory (LSTM) dan Support Vector Machine (SVM).

Hasil prediksi akan ditampilkan dalam bentuk label sentimen, misalnya positif atau negatif, beserta nilai probabilitas atau tingkat kepercayaan prediksi dari masing-masing metode. Selain itu, website ini juga dilengkapi dengan fitur perbandingan hasil prediksi antara kedua metode sehingga pengguna dapat melihat perbedaan kinerja keduanya secara langsung.

Desain website dibuat sederhana, responsif, dan mudah digunakan agar dapat diakses melalui berbagai perangkat, termasuk komputer dan smartphone. Tampilan visual juga dibuat jelas dan informatif untuk memudahkan interpretasi hasil analisis sentimen oleh pengguna.

Gambar 4.11 Website Analisis Sentimen SVM dan LSTM

4.12. Testing

Pada tahap ini, dilakukan pengujian terhadap beberapa komentar baru yang belum pernah digunakan dalam proses pelatihan model. Tujuan dari pengujian ini adalah untuk melihat seberapa baik model dalam mengklasifikasikan sentimen komentar secara aktual, serta untuk membandingkan hasil prediksi antara dua algoritma yang digunakan, yaitu LSTM dan SVM. Masing-masing model memberikan prediksi label sentimen beserta nilai skor keyakinannya (confidence score) dalam bentuk persentase. Berikut adalah hasil pengujian terhadap lima komentar pengguna.

Komentar	Aktual	LSTM		SVM	
		Prediksi	Skor Keyakinan (%)	Prediksi	Skor Keyakinan (%)
Walau preloved, kualitas sepatu ini masih oke banget. Nggak nyesel beli di sini!	Positif	Negatif	99.98	Positif	18.00
Sepatunya keren, bagus banget sumpah jadi nyesel gua beli sepatu konve**e asli harganya mahal, ternyata kualitasnya hampir sama dengan yang harga terjangkau.	Positif	Negatif	58.71	Positif	18.00
Bau sepatunya menyengat, kayak belum dicuci sama sekali.	Negatif	Negatif	12.49	Negatif	18.00
Sangat kecewa, sepatunya kotor dan tidak layak pakai.	Negatif	Negatif	0	Negatif	18.00
Sepatunya masih kelihatan baru banget, padahal preloved. Harganya juga ramah di kantong!	Positif	Negatif	76.48	Positif	18.00

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan dalam klasifikasi sentimen menggunakan algoritma **Long Short-Term Memory (LSTM)** dan **Support Vector Machine (SVM)** terhadap data sentimen berbahasa Indonesia, diperoleh beberapa temuan penting yang menjadi dasar dalam penarikan kesimpulan. Model LSTM memberikan hasil klasifikasi terbaik ketika menggunakan kombinasi parameter **batch size 16**, jumlah neuron **64**, dan **32 epoch**. Pada konfigurasi tersebut, model LSTM menghasilkan akurasi sebesar **61,49%**, dengan nilai **recall 62,16%**, **presisi 61,33%**, dan **F-measure 61,74%**. Proses pelatihan dan pengujian model LSTM memerlukan waktu sekitar **73 detik**.

Sementara itu, model SVM menunjukkan performa yang lebih unggul dibandingkan LSTM. SVM berhasil mencapai akurasi sebesar **68%**, dengan nilai recall, presisi, dan F-measure yang masing-masing juga sebesar **68%**, menunjukkan konsistensi yang baik antar metrik evaluasi. Selain itu, dari segi efisiensi, SVM hanya membutuhkan waktu **17 detik** untuk keseluruhan proses pelatihan dan pengujian.

Secara keseluruhan, SVM terbukti lebih optimal dalam konteks penelitian ini. Selain memiliki akurasi yang lebih tinggi, model ini juga lebih efisien dalam hal waktu dan menunjukkan kestabilan kinerja dalam berbagai metrik evaluasi. Meski demikian, algoritma LSTM tetap memiliki potensi besar untuk mencapai hasil yang lebih baik, terutama apabila dilakukan tuning parameter secara lebih mendalam. Mengingat LSTM dirancang khusus untuk menangani data

sekuensial seperti teks, model ini memiliki keunggulan dalam mempelajari konteks dan hubungan antar kata yang lebih kompleks dibandingkan SVM.

5.2. Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk pengembangan dan penelitian selanjutnya dalam bidang klasifikasi sentimen.

Pertama, peningkatan kualitas dataset menjadi hal yang sangat disarankan. Hal ini mencakup penambahan jumlah data serta memperluas variasi kata dan struktur kalimat, sehingga dapat memberikan representasi yang lebih baik terhadap berbagai ekspresi sentimen dalam bahasa Indonesia. Kualitas dataset yang baik akan sangat mendukung kinerja model, terutama pada arsitektur deep learning seperti LSTM yang memerlukan data dalam jumlah besar untuk dapat belajar secara optimal.

Kedua, eksplorasi terhadap arsitektur deep learning lainnya, seperti **Bidirectional LSTM**, **GRU,** atau kombinasi model seperti **CNN-LSTM**, sangat layak untuk dilakukan. Arsitektur-arsitektur ini memiliki potensi untuk menangkap konteks dan pola dalam data secara lebih efektif, yang pada akhirnya dapat meningkatkan akurasi klasifikasi. Selain itu, penggunaan teknik praproses lanjutan juga dapat menjadi faktor penting dalam meningkatkan performa model. Beberapa pendekatan yang dapat dieksplorasi meliputi penggunaan **stemming** dan **lemmatization** untuk Bahasa Indonesia, serta penerapan **word embedding** yang lebih sesuai secara linguistik, seperti **IndoBERT** atau **FastText Bahasa Indonesia**, yang dirancang khusus untuk menangani nuansa bahasa local. Proses **optimasi parameter (hyperparameter tuning)** juga disarankan untuk dilakukan secara

lebih menyeluruh pada kedua model. Dengan mencari kombinasi parameter yang optimal, diharapkan model dapat mencapai performa klasifikasi yang lebih tinggi dan stabil.

Terakhir, untuk kebutuhan implementasi pada sistem produksi atau aplikasi nyata, model **SVM** lebih direkomendasikan. Hal ini disebabkan oleh kestabilannya dalam menghasilkan prediksi serta efisiensi waktu pemrosesan yang jauh lebih cepat dibandingkan dengan model berbasis deep learning, sehingga lebih cocok diterapkan dalam sistem yang membutuhkan respon cepat dan akurat.

DAFTAR PUSTAKA

Akmala, S. (2018). *CyberSecurity dan Forensik Digital PERKEMBANGAN INTERNET PADA GENERASI MUDA DI INDONESIA DENGAN KAITAN UNDANG-UNDANG ITE YANG BERLAKU*. *1*(2), 45–49. http://www.hukumonline.com/berita/baca/lt4fb47c3

Al-Khowarizmi, Sari, I. P., & Maulana, H. (2023). Detecting Cyberbullying on Social Media Using Support Vector Machine: A Case Study on Twitter. *International Journal of Safety and Security Engineering*, *13*(4), 709–714. https://doi.org/10.18280/ijsse.130413

Ardiada, I. M. D., Sudarma, M., & Giriantari, D. (2019). Text Mining pada Sosial Media untuk Mendeteksi Emosi Pengguna Menggunakan Metode Support Vector Machine dan K-Nearest Neighbour. *Majalah Ilmiah Teknologi Elektro*, *18*(1), 55. https://doi.org/10.24843/mite.2019.v18i01.p08

Ardian Pradana, Y., Cholissodin, I., & Kurnianingtyas, D. (2023). *Analisis Sentimen Pemindahan Ibu Kota Indonesia pada Media Sosial Twitter menggunakan Metode LSTM dan Word2Vec* (Vol. 7, Issue 5). http://j-ptiik.ub.ac.id

Br Sinulingga, J. E., & Sitorus, H. C. K. (2024). Analisis Sentimen Opini Masyarakat terhadap Film Horor Indonesia Menggunakan Metode SVM dan TF-IDF. *Jurnal Manajemen Informatika (JAMIKA)*, *14*(1), 42–53. https://doi.org/10.34010/jamika.v14i1.11946

Carneiro, T., Da Nobrega, R. V. M., Nepomuceno, T., Bian, G. Bin, De Albuquerque, V. H. C., & Filho, P. P. R. (2018). Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications. *IEEE Access*, *6*, 61677–61685. https://doi.org/10.1109/ACCESS.2018.2874767

Fikri, M. I., Sabrila, T. S., Azhar, Y., & Malang, U. M. (n.d.). *Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter*.

Findawati, O. Y., Muhammad, M. M., Rosid, A., Kom, S., & Kom, M. (n.d.). *BUKU AJAR TEXT MINING Diterbitkan oleh UMSIDA PRESS*.

Grosseck, G., & Holotescu, C. (n.d.). *CAN WE USE TWITTER FOR EDUCATIONAL ACTIVITIES?* http://twitter.com/ggrosseck

Kencana Putri, A., & Ichsanuddin Nur, D. (2023). PENGGUNAAN BAHASA PYTHON UNTUK ANALISIS DAN VISUALISASI DATA PENDUDUK DI DESA SUMBERJO, NGANJUK. In *Jurnal Pengabdian Kepada Masyarakat* (Vol. 3, Issue 3). https://jurnalfkip.samawa-university.ac.id/karya_jpm/index

Nur Adhan, S., Ngurah Adhi Wibawa, G., Christien Arisona, D., Yahya, I., Studi Statistika, P., & Matematika dan Ilmu Pengetahuan Alam, F. (2024). *ANALISIS SENTIMEN ULASAN APLIKASI WATTPAD DI GOOGLE PLAY STORE DENGAN METODE RANDOM FOREST* (Vol. 2, Issue 1).

Pradhana, F. R., Musthafa, A., & Fitria, I. (2024). *ANALISIS SENTIMEN ULASAN PRODUK DAVIENA DI SHOPEE*. *November*, 1–13.

Purnasiwi, R. G., Kusrini, & Hanafi, M. (2023). Analisis Sentimen Pada Review Produk Skincare Menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM). *Innovative: Journal Of Social Science Research*, *3*(2), 11433–11448.

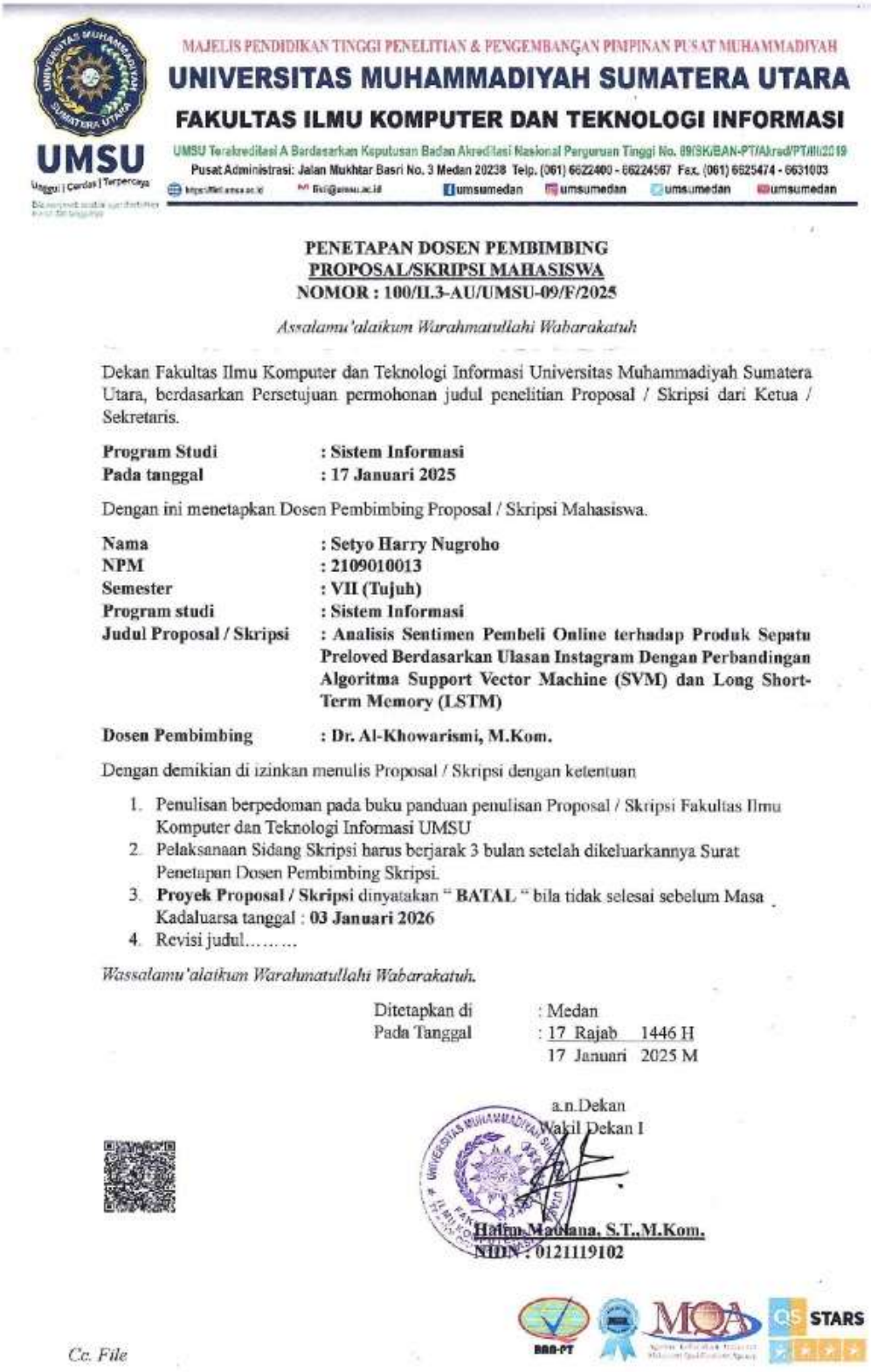
Safitri, R., Ali, I., & Rahaningsih, N. (2024). Analisis Sentimen Terhadap Tren Fashion Di Media Sosial Dengan Metode Support Vector Machine (Svm). *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(2), 1746–1754. https://doi.org/10.36040/jati.v8i2.9045

Wulandari, S., & Hasan, F. N. (2024). Analisis Sentimen Masyarakat Indonesia Terhadap Pengalaman Belanja Thrifting Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *8*(2), 768. https://doi.org/10.30865/mib.v8i2.7520

Yuliana, L., Apriyana, N., Fauzan, R., Larasati, N., Alhazami, L., Eko, I., & Sutopo, B. (n.d.). ANALISIS MINAT PEMBELIAN PRODUK PRELOVED SEBAGAI UPAYA PEDULI LINGKUNGAN. In *Jurnal Keuangan dan Bisnis ISSN* (Vol. 21, Issue 1).

LAMPIRAN

Lampiran 1. Surat Penetapan Dosen Pembimbing



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
umsumedan umsumedan umsumedan umsumedan

PENETAPAN DOSEN PEMBIMBING
PROPOSAL/SKRIPSI MAHASISWA
NOMOR : 100/IL3-AU/UMSU-09/F/2025

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, berdasarkan Persetujuan permohonan judul penelitian Proposal / Skripsi dari Ketua / Sekretaris.

Program Studi	**: Sistem Informasi**
Pada tanggal	**: 17 Januari 2025**

Dengan ini menetapkan Dosen Pembimbing Proposal / Skripsi Mahasiswa.

Nama	**: Setyo Harry Nugroho**
NPM	**: 2109010013**
Semester	**: VII (Tujuh)**
Program studi	**: Sistem Informasi**
Judul Proposal / Skripsi	**: Analisis Sentimen Pembeli Online terhadap Produk Sepatu Preloved Berdasarkan Ulasan Instagram Dengan Perbandingan Algoritma Support Vector Machine (SVM) dan Long Short-Term Memory (LSTM)**
Dosen Pembimbing	**: Dr. Al-Khowarismi, M.Kom.**

Dengan demikian di izinkan menulis Proposal / Skripsi dengan ketentuan

1. Penulisan berpedoman pada buku panduan penulisan Proposal / Skripsi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU
2. Pelaksanaan Sidang Skripsi harus berjarak 3 bulan setelah dikeluarkannya Surat Penetapan Dosen Pembimbing Skripsi.
3. **Proyek Proposal / Skripsi** dinyatakan " **BATAL** " bila tidak selesai sebelum Masa Kadaluarsa tanggal : **03 Januari 2026**
4. Revisi judul.....

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Ditetapkan di : Medan
Pada Tanggal : 17 Rajab 1446 H
17 Januari 2025 M

a.n.Dekan
Wakil Dekan I

Halim Maulana, S.T.,M.Kom.
NIDN : 0121119102

Cc. File

Lampiran 2. Surat Permohonan Seminar Proposal Skripsi

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
UMSU Unggul | Cerdas | Terpercaya
umsumedan umsumedan umsumedan umsumedan

PERMOHONAN
SEMINAR PROPOSAL SKRIPSI

Kepada Yth.
Bapak Dekan FIKTI UMSU
Di
Medan

Medan,.....2025

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dengan hormat, saya yang bertanda tangan di bawah ini mahasiswa Fakultas Ilmu Komputer dan Teknologi Informasi UMSU :

Nama Lengkap	: Setyo Harry Nugroho
NPM	: 2109010013
Program Studi	: Sistem Informasi

Mengajukan permohonan Mengikuti **Seminar Proposal Skripsi** yang ditetapkan dengan Surat Penetapan Judul Skripsi dan Pembimbing Nomor 100/IL3-AU/UMSU-09/F/2024 Tanggal .....dengan judul sebagai berikut :
ANALISIS SENTIMEN TERHADAP JUAL BELI PRODUK SEPATU *PRELOVED* BERDASARKAN ULASAN X DENGAN PERBANDINGAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) DAN *LONG SHORT-TERM MEMORY* (LSTM)

Bersama permohonan ini saya lampirkan :

1. Surat Penetapan Judul Skripsi (SK-1),
2. Surat Penetapan Pembimbing (SK-2),
3. DEKAM yang telah disahkan,
4. Kartu Hasil Studi Semester 1 s/d terakhir **ASLI**,
5. Tanda Bukti Lunas Beban SPP tahap berjalan,
6. Tanda Bukti Lunas Biaya Seminar Proposal Skripsi,
7. Proposal Skripsi yang telah disahkan oleh Pembimbing (rangkap-3),
8. Semua berkas dimasukan ke dalam MAP warna **BIRU**.

Demikian permohonan saya untuk pengurusan selanjutnya. Atas perhatian Bapak saya ucapkan terima kasih.

Wassalamualaikum Warahmatullahi Wabarakatuh.

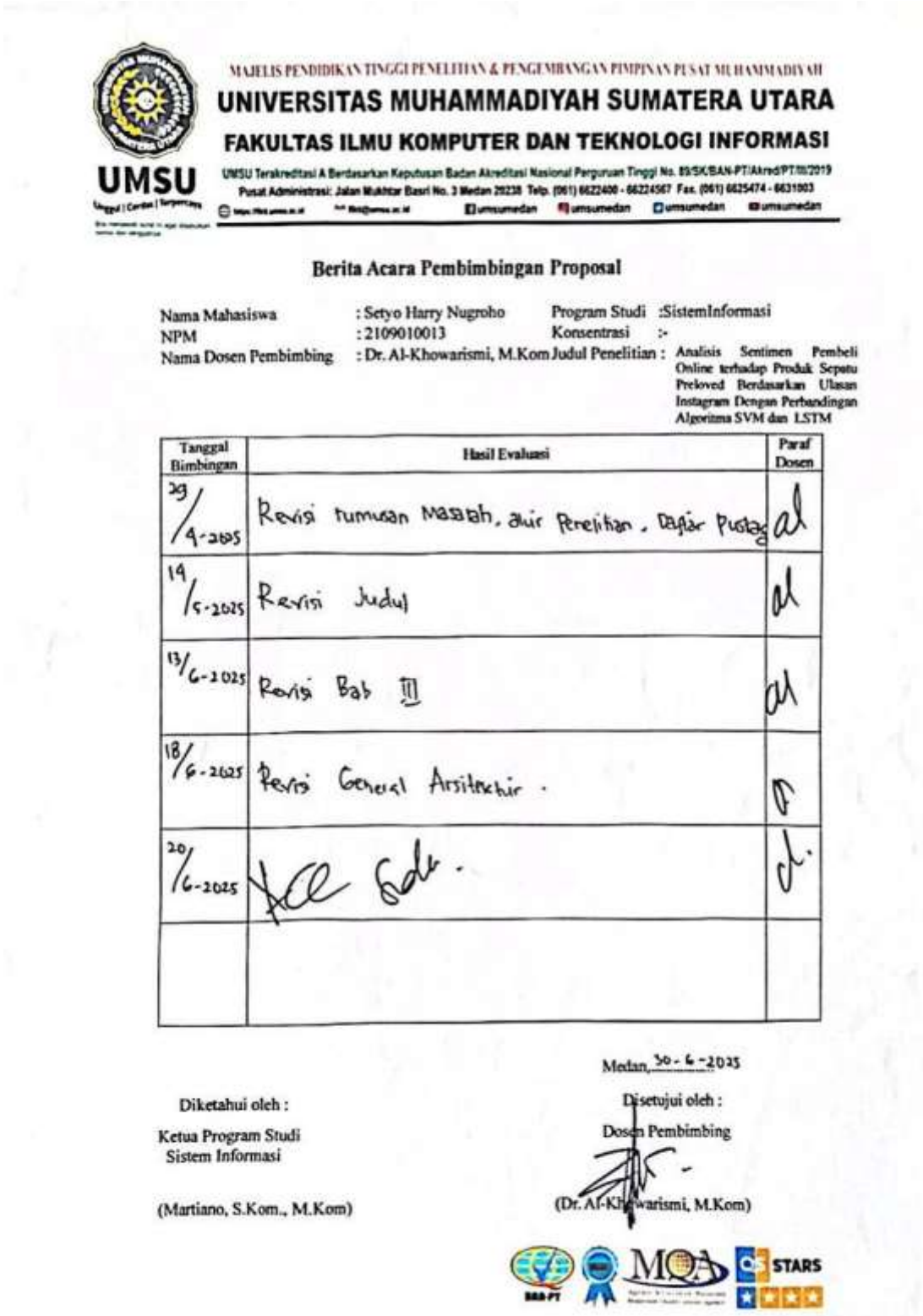
Menyetujui :
Pembimbing

(Dr. Al-Khowarizmi, S.Kom., M.Kom.)

Pemohon

(Setyo Harry Nugroho)

Lampiran 3. Berita Acara imbingan



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019

Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003

umsumedan umsumedan umsumedan umsumedan

Berita Acara Pembimbingan Proposal

Nama Mahasiswa : Setyo Harry Nugroho
NPM : 2109010013
Program Studi : SistemInformasi
Konsentrasi :-
Nama Dosen Pembimbing : Dr. Al-Khowarismi, M.Kom
Judul Penelitian : Analisis Sentimen Pembeli Online terhadap Produk Sepatu Preloved Berdasarkan Ulasan Instagram Dengan Perbandingan Algoritma SVM dan LSTM

Tanggal Bimbingan	Hasil Evaluasi	Paraf Dosen
29/4-2025	Revisi rumusan masalah, alur Penelitian, Daftar Pustaka	
19/5-2025	Revisi Judul	
13/6-2025	Revisi Bab III	
18/6-2025	Revisi General Arsitektur.	
20/6-2025	Acc Sidang.	

Medan, 30-6-2025

Diketahui oleh :
Ketua Program Studi
Sistem Informasi

(Martiano, S.Kom., M.Kom)

Disetujui oleh :
Dosen Pembimbing

(Dr. Al-Khowarismi, M.Kom)

Lampiran 4. Letter Of Acceptance

Website: https://ioinformatic.org/

LETTER OF ACCEPTANCE

Dear Corresponding Author **Hengky Triyo**

We are pleased to inform you that your submission ID**: 1773-JAIEA** titled **Sentiment Analysis of Pre-Loved Shoe Product Sales Based on X Reviews with a Comparison of Support Vector Machine (SVM) and Long Short-Term Memory (LSTM) Algorithms "** having author(s): **Setyo Harry Nugroho, Al-khowarizmi** for publication in **Journal of Artificial Intelligence and Engineering Applications** (E-ISSN **2808-4519), Accredited by Sinta Rank 5.** The acceptance decision was based on the reviewers' evaluation after double-blind peer review and the chief editor's **approval**.

You shall submit the Open Access processing fee (Rp.300,000,- / $20.00) please use transfer/bank draft to:

Bank	: Permata Bank
Account Name	: Akim Manaor Hara Pardede
Account Number (IDR)/(USD)	: 4119651064
Swift Code	: BBBAIDJ10SS

After performing the bank transfer, please email a copy of the transaction to **jaiea@ioinformatic.org** for with subject [Payment for: Your paper id, author's name]. All bank charges are to be covered by the author. Payments made are NOT refundable.

Kindly proceed with registration fee submission for slot allocation in ***15th October 2025. Vol. 5. No. 1***. The final updated copy can be submitted at a later time after slot reservation.

We shall encourage more quality submissions from you and your colleagues in the future.

Please do acknowledge receiving this notification.

Regards,

Dr. Ir. Akim Manaor Hara Pardede, ST., M.Kom
Editor-in-Chief
Journal of Artificial Intelligence and Engineering Applications (JAIEA)

Lampiran 5. Hasil Turnitin