KLASIFIKASI SURAT MASUK DI KANTOR PENGADILAN MILITER III-15 KUPANG MENGGUNAKAN (LSTM) LONG SHORT-TERM MEMORY

Asmawati Tuto[1*], Sumarlin[2], Yohanis Malelak[3]

[1,2,3] STIKOM Uyelindo Kupang

[1*]asmawati30102003@gmail.com, [2]sumarlin@gmail.com, [3]yohanismalelak32@gmail.com

DOI: https://doi.org/10.46880/mtk.v12i2.5924

ABSTRACT

The increasing volume of incoming correspondence at the Kupang Military Court III-15 Office has made the conventional letter classification process complex and time-consuming. Purpose: this study aims to develop an incoming letter classification system that categorizes letters into four classes (regular letters, circulars, decrees, and orders) using the Long Short-Term Memory (LSTM) method; the main contribution of this study is the application of LSTM to a local military-court correspondence dataset that has not been widely studied, together with a replicable preprocessing pipeline and K-Fold evaluation protocol, providing practical implications for accelerating correspondence administration in military judicial offices. Methods: the dataset consists of 500 incoming letter records in Excel format that underwent a preprocessing stage including cleaning, case folding, normalization, stopword removal, stemming, tokenizing, encoding, and padding, and was then divided into 80% training data and 20% testing data, evaluated using 5-Fold Cross Validation with accuracy, precision, recall, and F1-score as performance metrics. Results: the average model performance results were Accuracy 42.04%, Precision 42.39%, Recall 42.04%, and F1-Score 39.93%, with the highest accuracy obtained in Fold 2 (56.44%) and the lowest in Fold 5 (32.00%), while the model's main difficulty lay in distinguishing between the Circular and Decree categories, which share similar text patterns. Conclusion: the LSTM method is capable of recognizing textual patterns in letters and performing classification; however, its performance remains variable and relatively low on this dataset, so the developed system has the potential to improve the efficiency of incoming letter management at the Kupang Military Court III-15 Office, although further optimization is still needed before independent deployment.

Keywords: LSTM, Text Classification, Incoming Letters, Natural Language Processing, K-Fold Cross Validation, Military Court

I. PENDAHULUAN

Seiring dengan berkembangnya zaman, teknologi berkembang begitu pesat dengan banyak bermunculnya berbagai alat telekomunikasi atau penghubung yang canggih seperti; internet, namun masih ada komunikasi tertulis yang tidak dapat dilupakan keberadaannya, bahkan masih tetap kokoh terpakai seolah tak bisa tergantikan oleh berbagai peralatan komunikasi yang canggih [1]. Surat merupakan salah satu sarana komunikasi yang kehadirannya tidak dapat dilupakan dan bahkan masih digunakan secara kokoh, seakan-akan tidak bisa digantikan oleh berbagai perangkat komunikasi modern yang terus bermunculan. Untuk lebih memahami peran penting surat dalam suatu organisasi, perlu kita tinjau lebih jauh mengenai jenis-jenis surat yang ada, khususnya surat masuk yang menjadi salah satu elemen utama dalam proses komunikasi dan administrasi organisasi. Surat masuk merupakan semua jenis surat yang diterima oleh instansi maupun yang di terima dari pihak [1]. Oleh karena itu suarat tetap memiliki peran yang penting dalam dokumentasi resmi instansi.

Dalam konteks pengelolaan surat yang lebih canggih, setelah proses clustering untuk pengelompokan berdasarkan kesamaan, klasifikasi surat masuk dapat ditingkatkan melalui penerapan model seperti Long Short-Term Memory (LSTM), yang memungkinkan analisis urutan teks secara efektif untuk mengkategorikan surat ke dalam jenis atau fungsi yang tepat. Penelitian yang telah dilakukan sebelumnya menunjukan bahawa penerapan LSTM dalam klasifikasi teks memberikan keuntungan. Sebagai contoh, sebuah studi oleh (Harahap, 2024) menganalisis pengaruh fungsi aktivasi pada model Long Short Term Memory LSTM dalam klasifikasi teks hukum, hasil penelitian menunjukkan bahwa pemilihan fungsi aktivasi yang tepat sangat mempengaruhi akurasi model. Penelitian-penelitian terkini turut menegaskan keunggulan LSTM untuk klasifikasi teks berbahasa Indonesia maupun umum: [18] menerapkan LSTM untuk klasifikasi teks berita dan menunjukkan performa yang kompetitif dibanding metode konvensional; [19] menggunakan Bidirectional LSTM untuk klasifikasi emosi pada teks Twitter berbahasa Indonesia; [20] mengombinasikan TF-IDF dengan LSTM untuk meningkatkan representasi fitur teks; serta [21] menerapkan LSTM pada klasifikasi teks SMS spam berbahasa Indonesia dengan hasil yang baik. Di sisi lain, pendekatan berbasis transformer seperti BERT juga mulai banyak digunakan untuk klasifikasi teks berbahasa Indonesia, misalnya pada klasifikasi ujaran toksik [22], meskipun umumnya membutuhkan sumber daya komputasi dan volume data yang jauh lebih besar dibandingkan LSTM. Berbeda dari penelitian-penelitian tersebut yang sebagian besar berfokus pada teks berita, media sosial, atau data pengaduan umum, penelitian ini mengisi kesenjangan (gap) berupa penerapan LSTM secara spesifik pada dataset surat dinas internal instansi peradilan militer, yang memiliki karakteristik bahasa formal-birokratis dan jumlah data yang jauh lebih terbatas dibandingkan korpus teks berita

atau media sosial pada umumnya.

Meskipun surat masih menjadi elemen krusial dalam komunikasi organisasi, proses pengelolaan surat masuk di Kantor Pengadilan Militer III-15 Kupang hingga saat ini masih dijalankan secara konvensional dengan metode pencatatan dan pengelompokan manual. Kondisi tersebut menimbulkan berbagai kendala, seperti meningkatnya volume surat yang harus ditangani setiap harinya, baik dalam bentuk fisik maupun non fisik, sehingga proses pengelompokan dan klasifikasi menjadi lambat, kurang efisien, serta rawan ketidaktepatan dalam pencatatan dan penyampaian surat. Hal ini berdampak pada keterlambatan pada keterlambatan distribusi informasi kepada pihak terkait dan menurunnya efektivitas proses administrasi.

Untuk mengatasi permasalahan tersebut, penerapan model Long Short-Term Memory (LSTM) dapat menjadi solusi inovatif dan efektif dalam klasifikasi surat masuk di Kantor Pengadilan Militer III-15 Kupang. Pendekatan ini tidak hanya mempercepat proses administrasi, tetapi juga mengklasifikasi sesuai kategori surat, seperti Surat Keputusan, Surat Perintah, Surat Edaran, dan Surat Bebas, sehingga Organisasi atau Insatansi dapat memanfaatkan surat sebagai alat komunikasi yang lebih optimal. Penelitian ini bertujuan untuk membuat Klasifikasi Surat Masuk di Kantor Pengadilan Militer III-15 Kupang menggunakan (LSTM) Long Short-Term Memory. Secara terperinci, kontribusi penelitian ini mencakup: (1) penyediaan dataset lokal berupa 500 surat masuk dari Kantor Pengadilan Militer III-15 Kupang yang terbagi ke dalam empat kategori (Biasa, Edaran, Keputusan, Perintah); (2) perancangan pipeline preprocessing teks berbahasa Indonesia formal-birokratis yang meliputi cleaning, normalisasi, stemming berbasis Sastrawi, tokenizing, dan padding; serta (3) evaluasi model LSTM secara sistematis menggunakan skema 5-Fold Cross Validation dengan pelaporan metrik accuracy, precision, recall, dan F1-score pada tiap fold, sehingga memberikan gambaran kelayakan sekaligus keterbatasan pendekatan LSTM untuk klasifikasi surat dinas pada instansi peradilan militer.

II. KAJIAN TERKAIT

Surat

Surat merupakan media komunikasi tertulis yang digunakan untuk menyampaikan pesan, informasi dari satu pihak kepada pihak. Surat merupakan catatan tertulis yang digunakan sebagai alat komunikasi untuk menyampaikan informasi dari satu pihak ke pihak lain. Surat memiliki fungsi penting dalam dokumentasi, pengingat, dan bukti historis dalam berbagai kegiatan organisasi Surat masuk adalah surat atau komunikasi tertulis yang diterima oleh sebuah organisasi atau individu dari pihak eksternal, seperti pengirim yang tidak terkait secara langsung dengan organisasi tersebut [3].[4] Penelitian ini mendefinisikan surat itu sendiri merupakan salah satu sarana sangat penting didalam sebuah organisasi dikarenakan banyak sekali informasi sangat penting yang terkandung didalamnya, maka diperlukan pengelolaan yang benar agar tidak ada kesalahan seperti adanya surat yang hilang ataupun rusak yang bisa mengakibatkan kerugian pada instansi-intsansi yang bersangkutan.

Surat merupakan media tertulis yang berguna untuk melakukan koordinasi kerja, baik antar sesama rekan kerja dibagian atau departemen atau divisi yang sama atau antar bagian yangberbeda dilevel organisasi yang sama maupun dilevel organisasi yang berbeda [5]. [1] berpendapat bahwa surat adalah alat komunikasi tertulis untuk menyampaikan pesan kepada pihak lain, yang memiliki persyaratan khusus yaitu penggunaan kode dan notasi (lampiran dan perihal), Penggunaan kertas, model dan bentuk yang diterapkan, pemilihan bahasa khas, serta penambahan tanda tangan. Studi yang dilakukan [6] mendefinisikan Surat merupakan suatu selembar kertas atau lebih yang tertulis suatu pernyataan atau informasi yang ingin disampaikan, diberitakan atau diutarakan kepada orang lain.

Menurut [7] Surat itu bertindak sebagai pembawa pesan atau pesan atas nama penulis. Surat juga berfungsi sebagai bukti hukum yang dapat digunakan sebagai bukti dalam situasi bisnis, seperti kuitansi tagihan, dan perjanjian. Surat adalah alat komunikasi tertulis yang berasal dari satu pihak dan ditujukan kepada pihak lain untuk menyampaikan warta[8]. Menurut Gie, surat didefinisikan sebagai segala bentuk catatan yang ditulis atau digambarkan, yang berisi informasi tentang suatu hal atau kejadian, dan dibuat oleh seseorang untuk memfasilitasi daya ingat. Fungsi surat adalah sebagai sarana dalam penyampaian pesan secara tertulis, surat berperan dalam mencapai tujuan suatu instansi atau organisasi dalam menjalin kerja sama antar organisasi.

Data Mining

Temuan ini mendefinisikan Data Mining adalah satu proses logis yang digunakan untuk menemukan data yang diharapkan dari keseluruhan data yang berjumlah besar [9]. [10] Data mining adalah proses yang menggunakan berbagai perangkat analisis data untuk menemukan pola dan hubungan dalam data yang mungkin dapat digunakan untuk membuat prediksi yang valid. Langkah pertama dan paling sederhana dalam data mining yaitu menggambarkan data – menyimpulkan atribut statistik (seperti rata-rata dan standar deviasi), mereview secara visual menggunakan diagram dan grafik, serta mencari relasi berarti yang potensial antar variabel (misalnya nilai yang sering muncul bersamaan). Mengumpulkan, mengeksplor, dan memilih data yang tepat adalah sangat penting.

Klasifikasi

Klasifikasi adalah jenis analisis data yang membantu orang memprediksi label kelas mana yang harus diklasifikasikan ke dalam sampel [13]. Salah satu metode yang dimiliki oleh data mining adalah klasifikasi. Klasifikasi merupakan teknik supervised dalam data mining. Klasifikasi sendiri adalah cara pengelompokkan benda berdasarkan ciri – ciri yang dimiliki oleh objek yang akan diklasifikasikan. Dalam prosesnya, klasifikasi dapat dilakukan dengan banyak cara baik secara manual ataupun dengan bantuan teknologi [9].

Untuk mengidentifikasi item mana yang masuk ke dalam kategori tertentu, klasifikasi adalah strategi untuk mengidentifikasi pola yang dapat membagi satu kelas data dari yang lain dengan memperhatikan karakteristik dan perilaku pengelompokan yang telah ditentukan [11]. Klasifikasi data adalah proses yang menemukan properti yang sama dalam sekumpulan objek dalam database dan mengklasifikasikannya ke dalam kelas yang berbeda

sesuai dengan model klasifikasi yang ditentukan. Tujuan dari klasifikasi adalah untuk menemukan model dari training set yang membedakan atribut ke dalam kategori atau kelas yang sesuai, model tersebut kemudian digunakan untuk mengklasifikasikan atribut yang kelasnya belum diketahui sebelumnya [12].

Long Short-Term Memory (LSTM) merupakan variasi dari pengembangan Recurrent Neural Network (RNN) yang dibuat untuk menghindari masalah ketergantungan jangka panjang (long-term dependency). digunakan kembali dalam prediksi pada urutan yang panjang. LSTM memiliki struktur yang lebih kompleks dengan adanya komponen-komponen khusus seperti forget gate, input gate, cell state, dan output gate [14][15] Long Short-Term Memory (LSTM) yaitu jenis model dari Recurrent Neural Network (RNN). LSTM dikembangkan karena kemampuannya menangani ketergantungan jangka panjang dalam data, yang memungkinkan penyimpanan memori dari waktu yang lama.

K-Fold Cross Validation

K-fold cross validation merupakan cara statistikal yang digunakan dalam evaluasi performan dari model yang akan dibangun. Selain itu k-fold cross validation adalah cara yang efektif dalam menggabungkan atribut dan setting parameter machine learning dalam melatih model prediction yang lebih baik [16]. K-fold cross validation adalah Teknik yang digunakan untuk mengoptimalkan kinerja algoritma klasifikasi dengan partisi dataset menjadi k lipatan dengan (k – 1) sebagai data pelatihan dan sisanya sebagai data uji. Kinerja model menjadi sensitif terhadap k jika nilai k yang ditentukan sangat kecil. Namun dibutuhkan waktu yang lama untuk menghitung ketika dihadapkan dengan nilai k yang terlalu besar. Gambar berikut menampilakan proses tersebut [17].

III. METODE

Data Surat Masuk

Pengumpulan data dilakukan di Kantor Pengadilan Militer III-15 Kupang melalui observasi langsung terhadap proses pengelolaan surat masuk, serta pendataan mengenai jumlah surat masuk dan kategori surat yang diterima oleh instansi. Data yang dikumpulkan sebanyak 500 surat masuk. Dengan data tersebut digunakan sebagai dasar untuk membangun dan menguji model klasifikasi berbasis Long Short-Term Memory (LSTM) dalam mengkategori surat masuk sesuai dengan jenis.

Praproses Data

Preprocessing Data adalah tahap penting dalam analisis sentimen yang bertujuan mengubah data tidak terstruktur menjadi data terstruktur yang siap dianalisis. Data yang diperoleh biasanya masih dalam bentuk mentah dan tidak dapat langsung digunakan. Oleh karena itu, tahap ini dilakukan untuk mempersiapkan data agar lebih berkualitas dan siap untuk dianalisis dengan baik. Terdapat serangkaian langkah yang dilakukan yaitu membersihkan data, menghilangkan noise, dan mengatur data sesuai dengan kebutuhan analisis. Berikut beberapa proses yang dilakukan dalam preprocessing data:

1. Cleaning merupakan proses penghapusan elemen elemen yang tidak diperlukan dalam teks seperti emoticon, spasi berlebih, tanda baca, angka, url atau karakter khusus yang tidak berkontribusi pada analisis.
2. Case folding merupakan proses mengubah semua huruf kapital dalam teks menjadi huruf kecil. Tujuannya agar memastikan konsistensi, ketika kata yang sama dapat diperlakukan sebagai entitas yang sama dalam analisis.
3. Normalisasi Data merupakan proses transformasi atau konversi kata- kata tidak baku (*slang words*) yang sering digunakan dalam komunikasi informal menjadi bentuk kata baku atau standar. Proses ini bertujuan untuk meningkatkan konsistensi dan akurasi dalam analisis teks, khususnya pada aplikasi seperti klasifikasi sentimen.
4. Stopwords Removal merupakan proses penghilangan kata yang sering muncul dalam bahasa tetapi tidak memiliki makna penting untuk analisis.
5. Stemming merupakan proses mengubah kata menjadi kata dasar menggunakan pustaka Sastrawi untuk Bahasa Indonesia.
6. Tokenizing merupakan proses memecah teks menjadi unit-unit terkecil berupa kata yang disebut token agar memudahkan proses analisis elemen-elemen teks satu per satu.
7. Encoding token merupakan proses mengubah token menjadi representasi angka menggunakan Keras Tokenizer dengan ukuran vocabulary dibatasi hingga 20.000 kata (20.000-*word vocabulary*) yang paling sering muncul pada korpus. Teknik encoding yang digunakan adalah Word Embedding dengan dimensi 128. Setiap token direpresentasikan sebagai vektor berdimensi 128 yang dipelajari selama proses pelatihan (trainable embedding), bukan One-Hot Encoding.
8. Padding

   Padding merupakan proses menambahkan token khusus [PAD] pada akhir urutan token (*post-padding*) apabila jumlah token dalam teks lebih pendek dari panjang maksimum yang telah ditetapkan, yaitu maxlen = 200 token. Melalui proses padding, panjang semua data teks dapat diseragamkan yang memungkinkan model dapat melakukan proses pelatihan dan mencegah terjadinya perbedaan dimensi di antara data.

   Dari penjelasan di atas dpat dilihat tabel sebelum dan sesudah preprocessing sebagai berikut.

Tabel 1. Surat Sebelum Dan Sesudah Preprocessing

Tahap	Text Asli	Sesudah Preprocessing
Cleaning	KENAIKAN PANGKAT PERWIRA TNI DARI LETTU KE KAPTEN DI LINGKUNGAN PERADILAN MILITER PERIODE 1 OKTOBER 2025	KENAIKAN PANGKAT PERWIRA TNI DARI LETTU KE KAPTEN DI LINGKUNGAN PERADILAN MILITER PERIODE OKTOBER
Case Folding	KENAIKAN PANGKAT PERWIRA TNI DARI LETTU KE KAPTEN DI LINGKUNGAN PERADILAN MILITER PERIODE OKTOBER	kenaikan pangkat perwira tni dari lettu ke kapten di lingkungan peradilan militer periode oktober
Normalisasi	kenaikan pangkat perwira tni dari lettu ke kapten di lingkungan peradilan militer periode oktober	kenaikan pangkat perwira tni dari lettu ke kapten di lingkungan peradilan militer periode oktober
Stopword	kenaikan pangkat	kenaikan pangkat perwira tni

	perwira tni dari lettu ke kapten di lingkungan peradilan militer periode oktober	lettu kapten lingkungan peradilan militer periode oktober
Stemming	kenaikan pangkat perwira tni lettu kapten lingkungan peradilan militer periode oktober	naik pangkat perwira tni lettu kapten lingkung adil militer periode oktober
Tokenizing	naik pangkat perwira tni lettu kapten lingkung adil militer periode oktober	['naik', 'pangkat', 'perwira', 'tni', 'lettu', 'kapten', 'lingkung', 'adil', 'militer', 'periode', 'oktober']
Encoding	['naik', 'pangkat', 'perwira', 'tni', 'lettu', 'kapten', 'lingkung', 'adil', 'militer', 'periode', 'oktober']	[121, 95, 300, 62, 55, 301, 19, 5, 9, 96, 220]
Padding	[121, 95, 300, 62, 55, 301, 19, 5, 9, 96, 220]	[121, 95, 300, 62, 55, 301, 19, 5,9, 96, 220, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,0, 0, 0, 0]

Pembagian Data

Tahap pembagian data dilakukan dengan metode K-Fold Cross Validation dengan dibagai menjadi 2 proposisi antara lain: Data Latih (*Training Set*) sebesar 80%, digunakan untuk melatih model LSTM agar dapat mengenali pola dari isi surat dan Data Uji (Testing Set) sebesar 20%, digunakan untuk menguji performa model.

Evaluasi model dilakukan menggunakan metode Stratified K-Fold Cross Validation dengan jumlah fold sebanyak 5 (n_splits = 5). Sebelum proses pembagian data, seluruh dataset terlebih dahulu diacak (*shuffle* = True) dengan nilai random seed yang ditetapkan sebesar 42 (*random_state* = 42) guna memastikan hasil pengujian dapat direproduksi secara konsisten. Metode Stratified K-Fold dipilih karena pembagian data dilakukan secara proporsional berdasarkan distribusi kelas (*stratified*), sehingga setiap fold memiliki proporsi jumlah data dari masing-masing kategori surat (Biasa, Edaran, Keputusan, dan Perintah) yang seimbang dan representatif. Pada setiap iterasi, 80% data digunakan sebagai data latih dan 20% data digunakan sebagai data uji, sehingga dalam 5 fold seluruh data mendapat kesempatan menjadi data uji tepat satu kali. Nilai rata-rata (*mean*) dari kelima fold pengujian, bukan nilai dari satu fold tunggal. Perhitungan rata-rata dilakukan untuk memberikan gambaran performa model yang lebih objektif dan tidak bergantung pada satu pembagian data tertentu. Selain nilai rata-rata, standar deviasi (*standard deviation*) antar fold juga dihitung untuk mengukur konsistensi dan stabilitas performa model di setiap fold pengujian.

LSTM

Long Short-Term Memory (LSTM) digunakan untuk mengklasifikasikan surat masuk di Pengadilan Militer III-15 Kupang ke dalam kategori yang telah ditentukan. Setelah melalui tahap preprocessing, data kemudian diproses oleh model LSTM yang mampu mempelajari urutan serta pola kata dalam teks secara lebih mendalam. Model LSTM dibangun dengan satu layer LSTM berisi 128 unit, diikuti oleh Dense layer dengan 64 unit dan fungsi aktivasi ReLU, serta output layer dengan fungsi aktivasi Softmax sesuai jumlah kelas. Dropout tidak diterapkan pada model ini (rate = 0.0). Model dikompilasi menggunakan optimizer Adam dengan learning rate default sebesar 0.001 dan fungsi loss sparse categorical crossentropy. Proses pelatihan dilakukan dengan maksimum 30 epoch dan batch size sebesar 32, disertai mekanisme *Early Stopping* dengan nilai patience sebesar 3 yang memantau validation loss dan mengembalikan bobot terbaik secara otomatis *(restore_best_weights=True*). Untuk menangani ketidakseimbangan distribusi kelas, diterapkan dua pendekatan secara bersamaan, yaitu *Random Oversampling* pada data sebelum proses pembagian fold, serta pembobotan kelas (*class weighting*) dengan metode balanced yang diterapkan selama proses pelatihan. Evaluasi model dilakukan menggunakan metode K-Fold Cross Validation dengan lima fold. Seluruh eksperimen menggunakan random seed 42 yang diterapkan secara konsisten pada seluruh komponen, termasuk Python, NumPy, TensorFlow, pembagian data, dan oversampling, guna menjamin reproducibility hasil.

Pengujian

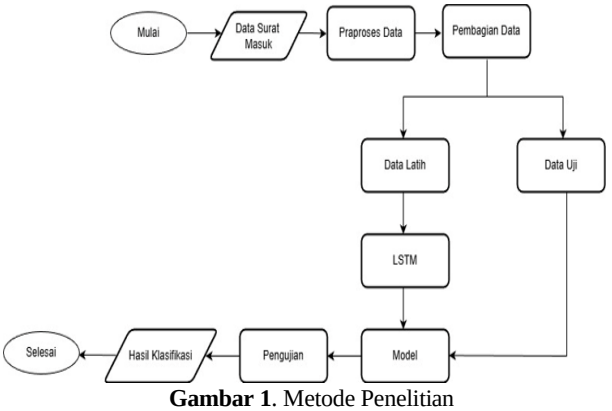
Pada tahap pengujian dilakukan untuk mengevaluasi kinerja model LSTM dalam mengklasifikasikan surat masuk. Evaluasi menggunakan skema 5-Fold Cross Validation, yaitu keseluruhan data dibagi menjadi lima bagian (*fold*) secara acak (*random shuffle*), dengan empat bagian digunakan sebagai data latih dan satu bagian tersisa sebagai data uji pada setiap iterasi, sehingga setiap fold memperoleh giliran menjadi data uji. Data uji dimasukkan ke dalam model untuk menghasilkan klasifikasi surat, kemudian hasil klasifikasi dibandingkan dengan label sebenarnya menggunakan confusion matrix. Metrik evaluasi dihitung per kelas pada setiap fold kemudian dirata-ratakan (*macro-average*) antar fold untuk memperoleh nilai evaluasi akhir. Berdasarkan matriks tersebut, dihitung nilai sebagai berikut: Accuracy (Persamaan 1), Precision (Persamaan 2), Recall (Persamaan 3), F1-Score (Persamaan 4).

Tahap ini bertujuan untuk memastikan bahwa model LSTM mampu mengklasifikasikan surat masuk ke dalam kategori yang tepat.

Hasil Klasifikasi

Tahap ini merupakan keluaran (*output*) akhir dari keseluruhan sistem, yaitu label kategori yang diprediksi oleh model untuk setiap surat masuk yang diuji. Setelah surat melalui proses Pengujian menggunakan data uji (20%), model akan menghasilkan prediksi berupa probabilitas untuk masing-masing dari empat kelas (surat biasa, edaran, keputusan, perintah).

Dari penjelasan tersebut dapat dilihat pada gambar 1 berikut.



Gambar 1. Metode Penelitian

IV. HASIL DAN PENELITIAN

Data Surat Masuk

Dataset yang digunakan dalam penelitian ini diperoleh dari Pengadilan Militer III-15 Kupang pada tahun 2026. Data yang dikumpulkan adalah data surat masuk yang digunakan sebagai bahan utama dalam proses klasifikasi surat menggunakan metode Long Short-Term Memori (LSTM). Dataset tersebut diperoleh dalam bentuk file Microsoft Excel sehingga memudahkan peroses pengolahan data menggunakan bahasa pemrograman Python. Jumlah dataset yang digunakan dalam penelitian ini sebanyak 500 data surat. Dataset tersebut kemudian dikelompokan ke dalam empat kategori surat: Biasa, Edaran, Keputusan, dan perintah. Dataset tersebut dapat dilihat sebagai berikut:

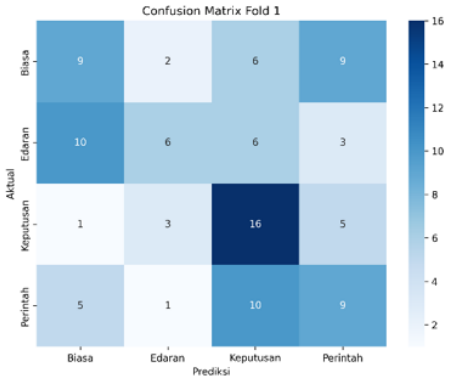
Table 2. Sampel Dataset Surat Masuk (Total 500 Surat)

No	Perihal	Kategori
0	KENAIKAN PANGKAT PERWIRA TNI DARI LETTU KE KAPTEN DI LINGKUNGAN PERADILAN MILITER PERIODE 1 OKTOBER 2025	Biasa
1	Pemanggilan Orientasi Pegawai Pemerintah dengan Perjanjian Kerja (PPPK) Tahap II Batch V dan VI Peserta Tahun 2025	Edaran
2	Panggilan sidang bagi para saksi dan Terdakwa an. Prada Irvan R. Syahputra	Perintah
.		
499	Pelaksanaan Pembangunan Zona Integritas Menuju WBK/WBBM di Lingkungan Peradilan Militer dan Peradilan Tata Usaha Negara Tahun 2026	Keputusan

Distribusi jumlah sampel per kategori dari total 500 surat masuk adalah sebagai berikut: kategori Biasa 125 surat, Edaran 125 surat, Keputusan 125 surat, dan Perintah 125 surat. Distribusi ini menunjukkan sebaran kelas yang seimbang (balanced), di mana setiap kategori memiliki jumlah sampel yang sama sehingga tidak terdapat dominasi satu kelas terhadap kelas lainnya.

Pengujian

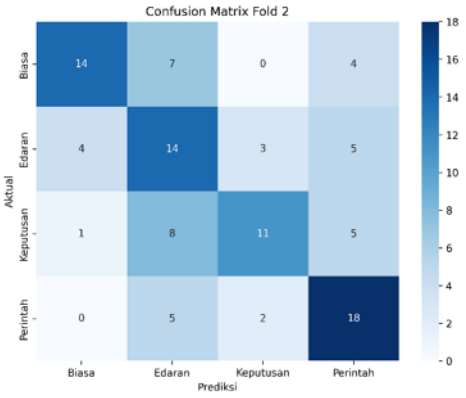
Pada tahap ini dilakukan pengujian sistem untuk mengevaluasi kinerja metod LSTM yang telah dibangun menggunakan 5-Fold Cros Validation terhadap data testing (testing data). Proses pengujian ini diukur dan ditampilkan melalui Confusion Matrix untuk mengetahi sejauh mana kemampuan sistem dalam mengklasifikasi surat masuk ke dalam kategori yang tepat. Metode evaluasi ini dipilih karena mampu menjelsakan secara detail perbandingan antara kelas yang sebenarnya (Actual Label) dengan kelas hasil prediksi sistem (Predicted Label). Melalui pengujian ini, tidak hanya nilai akurasi keseluruhan yang didapatkan, tetapi juga dapat diidentifikasi kategori surat masuk mana yang paling berhasil dikenali serta bagaimana pola kesalahan klasifikasi (misclassification) yang terjadi pada model. Hasil visualisasi Confusion Matrix dari kelima iterasi tersebut dipaparkan secara lengkap melalui Gambar 7 (Fold 1), Gambar 8 (Fold 2), Gambar 9 (Fold 3), Gambar 10 (Fold 4), serta Gambar 11 (Fold 5) di bawah ini.



Gambar 2. Fold 1

Berdasarkan hasil confusion matrix fold 1 Pada kategori Biasa, dari keseluruhan data aktual kategori Biasa, model berhasil memprediksi dengan benar sebanyak 9 surat, sedangkan sisanya mengalami kesalahan klasifikasi yaitu 2 surat diprediksi sebagai Edaran, 6 surat diprediksi sebagai Keputusan, dan 9 surat diprediksi sebagai Perintah. Selanjutnya kategori Edaran, model hanya mampu memprediksi dengan benar sebanyak 6 surat, sementara terjadi kesalahan klasifikasi sebanyak 10 surat diprediksi sebagai Biasa, 6 surat diprediksi sebagai Keputusan, dan 3 surat diprediksi sebagai Perintah.

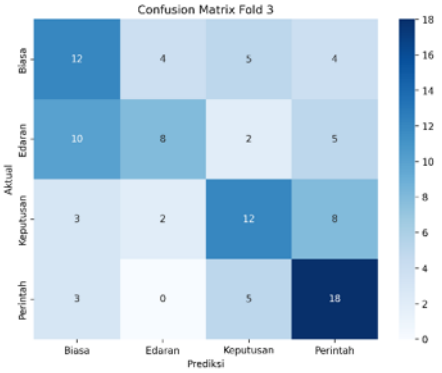
Kemudian kategori Keputusan, model menunjukkan performa terbaik dibandingkan kategori lainnya dengan berhasil memprediksi dengan benar sebanyak 16 surat. Namun demikian, masih terdapat kesalahan klasifikasi yaitu 1 surat diprediksi sebagai Biasa, 3 surat diprediksi sebagai Edaran, dan 5 surat diprediksi sebagai Perintah. Pada kategori Perintah, model berhasil memprediksi dengan benar sebanyak 9 surat, sedangkan kesalahan klasifikasi terjadi pada 5 surat yang diprediksi sebagai Biasa, 1 surat diprediksi sebagai Edaran, dan 10 surat diprediksi sebagai Keputusan. sehingga menghasilkan nilai akurasi sebesar 39,60%. Kategori Keputusan menjadi kelas dengan performa klasifikasi terbaik, sedangkan kategori Edaran dan Biasa menunjukkan tingkat kesalahan klasifikasi yang paling tinggi pada Fold 1 ini.



Gambar 3. Fold 2

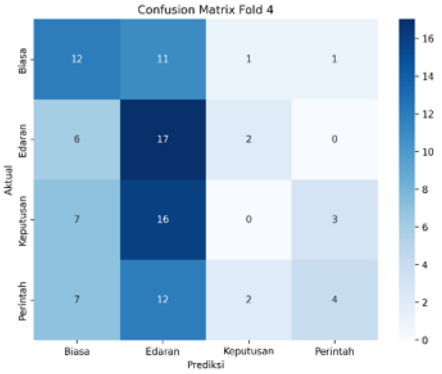
pada confusion matrix Fold 2, dapat diketahui bahwa model klasifikasi menunjukkan performa terbaik dalam mengidentifikasi kategori surat dengan menghasilkan nilai akurasi sebesar 56,44%. Fold 2 merupakan fold dengan performa terbaik dibandingkan fold-fold lainnya, di mana kategori Perintah menjadi kelas dengan jumlah prediksi benar tertinggi, sedangkan kategori Keputusan masih menunjukkan tingkat kesalahan klasifikasi yang cukup

tinggi terutama karena banyak data yang salah diprediksi sebagai kategori Edaran.



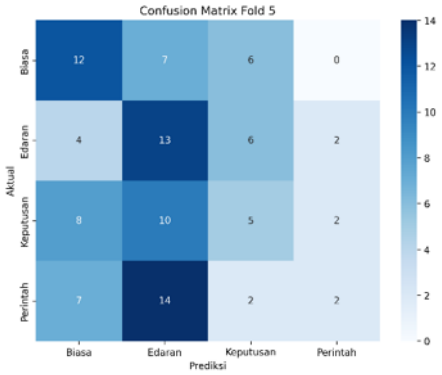
Gambar 4. Fold 3

Difold 3, dapat dilihat bahwa model Klasifikasi menghasilkan nilai akurasi sebesar 49,50%. Fold 3 menunjukkan peningkatan performa dibandingkan Fold 1, di mana kategori Perintah kembali menjadi kelas dengan jumlah prediksi benar tertinggi. Sementara itu, kategori Edaran masih menunjukkan tingkat kesalahan klasifikasi yang paling tinggi, terutama karena banyak data Edaran yang salah diprediksi sebagai kategori Biasa sebanyak 10 surat, yang mengindikasikan adanya kemiripan pola teks antara kedua kategori tersebut.



Gambar 5. Fold 4

Hasil dari confusion matrix Fold 4, menunjukan bahwa kinerja model dalam mengkalsifikasi mengalami penurunan dibandingkan Fold sebelumnya. Ini dilihat dari rendahnya nilai akurasi sebesar 32,67%. Fold 4 merupakan fold dengan performa yang menurun signifikan dibandingkan fold-fold sebelumnya. Hal yang paling mencolok adalah model sama sekali tidak mampu mengklasifikasikan kategori Keputusan dengan benar, di mana seluruh data Keputusan salah diklasifikasikan ke kategori lain, terutama ke kategori Edaran sebanyak 16 surat. Kondisi ini mengindikasikan bahwa distribusi data pada Fold 4 memiliki karakteristik yang lebih sulit dipelajari oleh model, sehingga berdampak pada penurnan kinerja.



Gambar 6. Fold 5

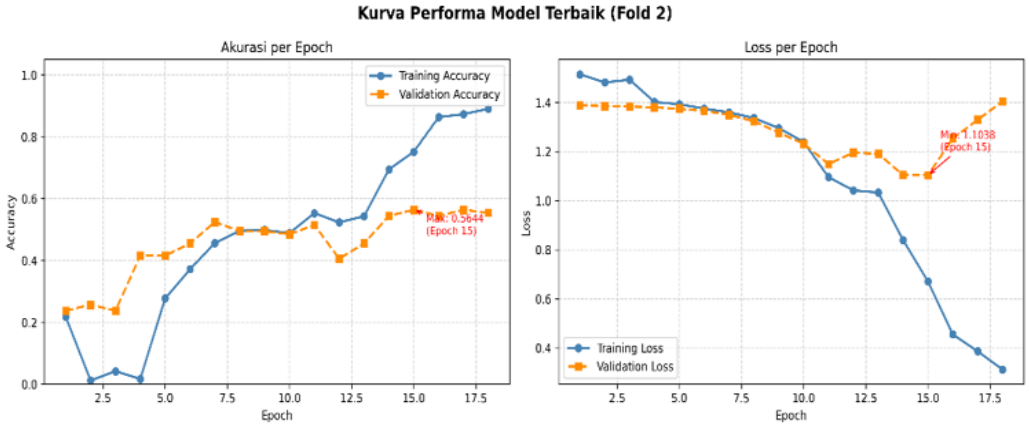
Berdasarkan hasil evaluasi pada Fold 5, nilai akurasi sebesar 32,00%. Fold 5 merupakan fold dengan performa terendah dibandingkan seluruh fold lainnya. Kategori Edaran menjadi kelas dengan jumlah prediksi benar tertinggi sebanyak 13 surat, sedangkan kategori Perintah menunjukkan performa terburuk dengan hanya 2 surat terprediksi benar dari keseluruhan data aktualnya. Kecenderungan model yang sangat dominan memprediksi data ke kategori Edaran pada Fold 5 ini, terlihat dari tingginya jumlah kesalahan klasifikasi yang mengarah ke kategori tersebut dari hampir semua kelas, mengindikasikan bahwa model mengalami kesulitan dalam membedakan pola teks antar kategori surat pada distribusi data di Fold 5.

Dari hasil evaluasi kinerja model tersebut dapat dilihat table evaluasi tiap fold sebagai berikut:

Table 3. Hasil Evaluasi Model LSTM Tiap Fold (5-Fold Cross Validation)

Fold	Accuracy	precision	Recall	F1-Score
1	0.3960	0.4063	0.3960	0.3842
2	0.5644	0.5978	0.5644	0.5668
3	0.4950	0.5037	0.4950	0.4868
4	0.3267	0.2917	0.3267	0.2681
5	0.3200	0.3198	0.3200	0.2904

Variabilitas antar fold juga cukup tinggi: nilai accuracy pada kelima fold memiliki rata-rata 42,04% dengan standar deviasi (SD) sebesar ±10,70%, yang mengindikasikan performa model belum stabil antar pembagian data. Sebagai catatan pengembangan lanjutan, penelitian ini belum menyertakan perbandingan langsung dengan metode baseline (mis. Naive Bayes, SVM, atau TF-IDF+Logistic Regression) maupun laporan precision/recall/F1 per kelas dan confusion matrix gabungan antar fold; penulis perlu menjalankan eksperimen tambahan tersebut sebelum submission final agar signifikansi peningkatan performa LSTM terhadap metode konvensional dapat dinilai secara kuantitatif. Dari hasil evaluasi tersebut dapat disimpulkan bahwa fold ke 2 memiliki performa terbaik. Berikut kurva performa terbaik pada fold 2 antara lain:

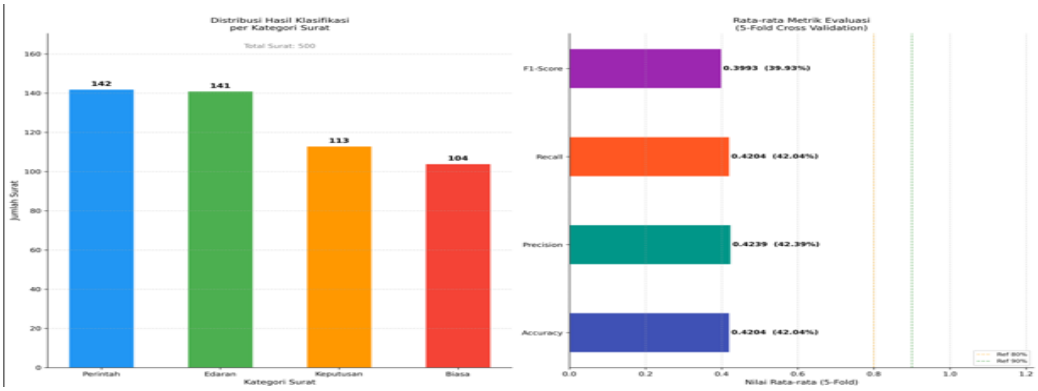


Gambar 7. Kurva Fold Terbaik

Berdasarkan hasil pengujian menggunakan metode K-Fold Cross Validation dengan lima fold, model terbaik diperoleh pada Fold 2 dengan nilai akurasi validasi tertinggi sebesar 56,44%. Kurva performa model terbaik tersebut ditampilkan dalam dua grafik, yaitu kurva akurasi per epoch dan kurva loss per epoch.

Hasil Klasifikasi

Berikut adalah hasil klasifikasi data surat masuk antara lain dapat ditunjukkan pada Gambar 8.



Gambar 8. Hasil Klasifikasi Surat

Berdasarkan grafik hasil klasifikasi yang ditampilkan, terdapat dua panel utama yaitu distribusi hasil klasifikasi per kategori surat dan rata-rata metrik evaluasi menggunakan metode 5-Fold Cross Validation. Pada panel kiri, distribusi hasil klasifikasi dari total 500 data surat masuk menunjukkan bahwa kategori Perintah memperoleh jumlah klasifikasi terbanyak dengan 142 surat, diikuti oleh kategori Edaran sebanyak 141 surat, kemudian kategori Keputusan sebanyak 113 surat, dan kategori Biasa sebanyak 104 surat. Distribusi hasil klasifikasi tersebut menunjukkan bahwa model cenderung lebih dominan dalam mengklasifikasikan surat ke dalam kategori Perintah dan Edaran dibandingkan kategori lainnya.

Pada panel kanan, hasil rata-rata metrik evaluasi dari 5-Fold Cross Validation menunjukkan bahwa model memperoleh nilai Accuracy sebesar 42,04%, nilai Precision sebesar 42,39%, nilai Recall sebesar 42,04%, dan nilai F1-Score sebesar 39,93%. Keseluruhan nilai metrik evaluasi tersebut masih berada di bawah garis referensi 80% maupun 90%, yang mengindikasikan bahwa performa model Bidirectional LSTM dalam mengklasifikasikan surat masuk di Kantor Pengadilan Militer III-15 Kupang masih tergolong rendah dan belum mencapai tingkat akurasi yang optimal. Secara linguistik, kesalahan klasifikasi yang paling sering terjadi antara kategori Edaran dan Keputusan (serta Edaran dan Biasa) diduga disebabkan oleh kemiripan struktur kalimat formal-birokratis pada ketiga jenis surat tersebut, seperti penggunaan frasa pembuka dan penutup baku (mis. "sehubungan dengan", "menindaklanjuti", "disampaikan dengan hormat") yang muncul lintas kategori, sehingga model kesulitan menangkap kata kunci pembeda yang bersifat spesifik-kategori hanya dari 100 token pertama teks. Verifikasi menyeluruh terhadap contoh-contoh surat yang salah diklasifikasikan (beserta kutipan perihal aslinya) perlu dilakukan penulis secara langsung pada data mentah sebagai pelengkap pembahasan ini.

V. KESIMPULAN

Penelitian ini mengenai Sistem Klasifikasi Surat Masuk Di Kantor Pengadilan Militer III-15 Kupang Menggunakan Metode Long Short-Term Memory (LSTM), dapat disimpulkan bahwa sistem yang dibangun mampu melakukan klasifikasi surat masuk ke dalam kategori surat biasa, surat edaran, surat keputusan, dan surat perintah secara otomatis. Hasil pengujian menggunakan metode K-Fold Cross Validation menunjukkan bahwa nilai rata-rata Accuracy 42,04%, Precision 42,39%, Recall 42,04%, F1-Score 39,93%. Hasil tersebut menunjukkan bahwa metode LSTM memiliki kemampuan yang cukup baik dalam mengenali pola teks surat masuk, meskipun tingkat akurasi yang diperoleh masih perlu ditingkatkan. Penerapan sistem ini dapat membantu proses pengelolaan surat masuk yang sebelumnya dilakukan secara manual menjadi lebih cepat, efisien, dan terstruktur, sehingga dapat mendukung

peningkatan pelayanan administrasi di Kantor Pengadilan Militer III-15 Kupang.

Meskipun demikian, hasil evaluasi juga menunjukkan bahwa performa model belum konsisten di seluruh skenario pengujian, dengan rentang akurasi antar fold yang cukup lebar (32,00% hingga 56,44%) dan kecenderungan kesalahan klasifikasi pada kategori surat dengan karakteristik bahasa yang mirip, seperti Edaran dan Keputusan. Temuan ini menegaskan bahwa sistem yang dibangun masih bersifat purwarupa (*prototipe*) yang menunjukkan kelayakan pendekatan LSTM untuk tugas klasifikasi surat, namun belum dapat diandalkan sepenuhnya untuk operasional harian tanpa adanya proses verifikasi manual serta upaya lanjutan untuk meningkatkan stabilitas dan akurasi model. Berdasarkan hasil penelitian yang didapatkan, ada beberapa saran yang diberikan untuk pengembangan sistem selanjutnya yaitu sebagai berikut: 1). Menambah jumlah dataset agar model dapat belajar lebih banyak pola surat sehingga hasil klasifikasi menjadi lebih akurat.2). Melakukan pengujian dengan metode lain, seperti CNN, RNN, atau SVM sebagai perbandingan performa dengan metode LSTM. 3). Mengembangkan sistem agar dapat terintegrasi langsung dengan sistem administrasi surat di Kantor Pengadilan Militer III-15 Kupang. 4). Mencoba arsitektur berbasis Transformer/BERT (mis. IndoBERT) sebagai pembanding LSTM, mengingat pendekatan ini terbukti kompetitif untuk klasifikasi teks berbahasa Indonesia pada penelitian terkini, meski perlu mempertimbangkan kebutuhan data dan sumber daya komputasi yang lebih besar. 5). Melakukan fine-tuning embedding kata yang bersifat domain-specific (dilatih dari korpus surat dinas/hukum) sebagai pengganti embedding yang dilatih dari awal, untuk membantu model membedakan kosakata formal-birokratis yang tumpang tindih antar kategori.

VI. REFERENSI

[1] H. Kisma and Yohanis, "Pengelolaan Surat Masuk Dan Surat Keluar di Sekretariat Daerah Kabupaten Solok Selatan," *J. Adm. Publik dan Pemerintah.*, vol. 3, no. 1, pp. 17–23, 2024, doi: 10.55850/simbol.v3i1.99.

[2] A. Fungsi and A. Pada, "Long Short Term Memory Untuk Klasifikasi Teks Oleh : Daniel Royhan Suhega Harahap Fakultas Teknik Diajukan sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana di Fakultas Teknik Universitas Medan Area Daniel Royhan Suhega Harahap Fakultas Teknik," 2024.

[3] M. Al Fayed, U. Pembangunan, and P. Budi, "Perancangan Sistem Informasi Administrasi Surat Masuk dan Surat Keluar Pada Dinas Kominfo Serdang Bedagai," *J. Minfo Polgan*, vol. 13, pp. 2049–2055, 2024.

[4] I. Pratiwi, "Implementasi Algoritma Left Corner Parsing Dalam Pencarian Surat Pada Arsip Surat Masuk dan Surat Keluar," *Bull. Data Sci.*, vol. 1, no. 3, pp. 99–102, 2022, doi: 10.47065/bulletinds.v1i3.2147.

[5] Rosalina Kambia and Megananda Cahya Dewi, "Prosedur Penanganan Surat Masuk Dan Surat Keluar Di Kantor Pos Solo 57100 Studi Kasus Pada Penanganan Surat Pt Pos Yang Ada Di Kota Solo," *J. Ilm. Ekon. Dan Manaj.*, vol. 2, no. 9, pp. 437–447, 2024, doi: 10.61722/jiem.v2i9.2564.

[6] R. Amalia, N. Suarna, I. Ali, and D. I. Efendi, "Implementasi Algoritma K-Means Dalam Optimalisasi Pengelompokan Surat Masuk Di Pg. Sindanglaut," *J. Inform. dan Tek. Elektro Terap.*, vol. 13, no. 1, 2025, doi: 10.23960/jitet.v13i1.5876.

[7] D. C. Arsiah and D. Anggita, "Pengelolaan Surat Menyurat Pada Bidang Peerdagangan Luar NegeriDinas Perindustrian Dan Perdagangan Kota Medan," *J. Pengabdi. dan Pemberdaya. Masy.*, vol. 1, no. 2, pp. 70–78, 2024.

[8] M. Rizky Asyari and S. Ramadhani, "Sistem Informasi Arsip Surat Menyurat," *J. Teknol. dan Inf. Bisnis*, vol. 3, no. 1, pp. 31–2021, 2021, [Online]. Available: https://doi.org/10.47233/jteksis.v3i1.172

[9] N. A. Ramadhani and H. A. Rosyid, "Review: Algoritma Data Mining untuk Klasifikasi Data," *J. Inov. Teknol. dan Edukasi Tek.*, vol. 2, no. 12, pp. 550–556, 2022, doi: 10.17977/um068v2i122022p550-556.

[10] K. N. Nababan, "Penerapan Metode K-Means Dalam Menentukan Klasifikasi Produk Pembelian Pelanggan," *BEES Bull. Electr. Electron. Eng.*, vol. 3, no. 2, pp. 81–97, 2022, doi: 10.47065/bees.v3i2.1129.

[11] D. Azlil Huriah and N. Dienwati Nuris, "Klasifikasi Penerima Bantuan Sosial Umkm Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 1, pp. 360–365, 2023, doi: 10.36040/jati.v7i1.6300.

[12] P. A. M. Z. R.W.P.P.Zer and I. Gunawan, "Penerapan Data Mining Naïve Bayes Dalam Klasifikasi Kepuasan Mahasiswa Berlangganan WiFi Indihome," *J. Media Inform.*, vol. 3, no. 2, pp. 112–118, 2022, doi: 10.55338/jumin.v3i2.488.

[13] Jatnika Fahmi Idris, Rafid Ramadhani, and Muhammad Malik Mutoffar, "Klasifikasi Penyakit Kanker Paru Menggunakan Perbandingan Algoritma Machine Learning," *J. Media Akad.*, vol. 2, no. 2, 2024, doi: 10.62281/v2i2.145.

[14] F. Setiawan, "Klasifikasi Tingkat Urgensi Pengaduan Masyarakat Menggunakan Long Short-Term Memory (Lstm) (Studi Kasus : Website Lapak Aduan Banyumas)," vol. 12, no. 4, pp. 6820–6827, 2025.

[15] F. N. Fajri and S. Syaiful, "Klasifikasi Nama Paket Pengadaan Menggunakan Long Short-Term Memory (LSTM) Pada Data Pengadaan," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1625–1633, 2022, doi: 10.47065/bits.v4i3.2635.

[16] A. I. Pradana and V. Atina, "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," pp. 239–248, 2024, doi: 10.33364/algoritma/v.21-1.1618.

[17] M. Anjas, "Optimization of Classification Algorithms Performance with k-Fold Cross Validation," no. 2, 2024.

[18] I. Triyadi, B. Prasetiyo, and T. L. Nikmah, "News text classification using Long-Term Short Memory (LSTM) algorithm," *J. Soft Comput. Explor.*, vol. 4, no. 2, pp. 79–86, 2023.

[19] A. Glenn, P. LaCasse, and B. Cox, "Emotion classification of Indonesian Tweets using Bidirectional LSTM," *Neural Comput. Appl.*, vol. 35, no. 13, pp. 9567–9578, 2023, doi: 10.1007/s00521-022-08186-1.

[20] M. Liang and T. Niu, "Research on Text Classification Techniques Based on Improved TF-IDF Algorithm and LSTM Inputs," *Procedia Comput. Sci.*, vol. 208, pp. 460–470, 2022, doi: 10.1016/j.procs.2022.10.064.

[21] E. D. Pratama, "Implementasi Model Long-Short Term Memory (LSTM) pada Klasifikasi Teks Data SMS Spam Berbahasa Indonesia," *J. Mach. Learn. Comput. Intell.*, vol. 1, no. 2, 2022, doi: 10.26740/vol1iss2y2022id12.

[22] Y. Sagama and A. Alamsyah, "Multi-label Classification of Indonesian Online Toxicity Using BERT and RoBERTa," in *Proc. IAICT*, 2023, doi: 10.1109/IAICT59002.2023.10205892.