Journal of Information Technology and Computer Science (INTECOMS)
Volume 8 Nomor 3, Tahun 2025
e-ISSN : 2614-1574
p-ISSN : 2621-3249

KLASIFIKASI DAN PREDIKSI PADA DATA ULASAN TRAVELOKA MENGGUNAKAN ALGORITMA LSTM

CLASSIFICATION AND PREDICTION IN TRAVELOKA REVIEW DATA USING LSTM ALGORITHM

Novandra Hanifah Najlawarni[1], Tukino Paryono[2], Fitria Nurapriani[3], Baenil Huda[4]
Universitas Buana Karawang[1,2,3,4]
si21.novandranajlawar@mhs.ubpkarawang.ac.id

ABSTRACT

This study aims to classify user reviews of the Traveloka application using the Long Short-Term Memory (LSTM) algorithm. The dataset consists of 1,102 reviews that were cleaned and manually labeled into four categories: application, technical issues, pricing, and service. The LSTM model achieved a classification accuracy of 89%, with a macro average precision of 0.90, recall of 0.89, and F1-score of 0.90. These results indicate that the LSTM algorithm is effective in handling text-based opinion data and provides reliable classification outcomes. This research supports the applicability of deep learning methods in sentiment-based text classification tasks.

Keywords: *Traveloka, LSTM, sentiment classification, variable prediction, deep learning*

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasikan ulasan pengguna aplikasi Traveloka menggunakan algoritma Long Short-Term Memory (LSTM). Dataset terdiri dari 1102 ulasan yang telah diproses melalui tahap pembersihan dan pelabelan manual ke dalam empat kategori, yaitu aplikasi, teknis, harga, dan pelayanan. Hasil pelatihan menunjukkan bahwa model LSTM mampu mengklasifikasikan data dengan akurasi sebesar 89%, precision rata-rata 0,90, recall 0,89, dan f1-score 0,90. Capaian ini menunjukkan bahwa LSTM efektif dalam memproses data teks ulasan dan memberikan prediksi yang andal. Penelitian ini mendukung penerapan deep learning dalam klasifikasi teks berbasis opini pengguna.

Kata Kunci: Traveloka, LSTM, klasifikasi sentimen, prediksi variabel, deep learning

PENDAHULUAN

Pesatnya perkembangan teknologi informasi dan komunikasi telah membawa dampak signifikan terhadap berbagai aspek kehidupan, termasuk dalam hal bagaimana konsumen berinteraksi dengan layanan digital. Di era digital saat ini, aplikasi mobile menjadi sarana utama masyarakat dalam mengakses berbagai layanan, mulai dari transportasi, akomodasi, hingga pemesanan tiket perjalanan. Salah satu platform yang banyak digunakan oleh masyarakat Indonesia adalah Traveloka. Aplikasi ini menyediakan berbagai layanan perjalanan yang terintegrasi, seperti pemesanan tiket pesawat, hotel, hingga penyewaan kendaraan(Sovia Pramudita & Agustia, n.d.). Dalam penggunaannya, pelanggan secara aktif memberikan umpan balik melalui ulasan yang tersedia di Google Play Store sebagai bentuk ekspresi terhadap pengalaman mereka menggunakan aplikasi tersebut.

Ulasan dari pelanggan ini tidak hanya mencerminkan kepuasan terhadap fitur-fitur yang ditawarkan, tetapi juga dapat mencakup keluhan terkait kendala teknis, kritik terhadap kualitas layanan pelanggan, serta opini mengenai harga atau promo yang tersedia. Melalui data ulasan ini, perusahaan sebenarnya dapat menggali informasi yang sangat bermanfaat untuk menyusun strategi perbaikan layanan, meningkatkan loyalitas pelanggan, hingga memahami tren dan kebutuhan pasar secara lebih mendalam. Namun demikian, tantangan muncul ketika volume data

ulasan yang tersedia sangat besar dan terus bertambah setiap harinya(Fernandes Andry et al., 2025). Dalam kondisi ini, metode analisis manual tidak lagi memadai karena memerlukan waktu dan tenaga yang tidak efisien.

Sebagai solusi, pemanfaatan teknologi kecerdasan buatan *(Artificial Intelligence)* dan teknik data mining menjadi pendekatan yang semakin relevan. Salah satu teknik yang banyak digunakan dalam pengolahan data berbasis teks adalah klasifikasi, yaitu proses mengelompokkan data ke dalam kategori tertentu berdasarkan pola yang teridentifikasi. Klasifikasi teks dapat membantu dalam memilah ulasan konsumen secara otomatis ke dalam topik atau kategori yang telah ditentukan sebelumnya, seperti fitur aplikasi, masalah teknis, pelayanan pelanggan, dan harga(Putra et al., 2022). Hal ini akan sangat membantu perusahaan dalam menilai fokus utama dari setiap keluhan atau pujian yang diterima.

Untuk mendukung proses klasifikasi tersebut, algoritma *Long Short-Term Memory* (LSTM) digunakan karena kemampuannya dalam memproses data berurutan. LSTM merupakan salah satu varian dari *Recurrent Neural Network* (RNN) yang dirancang untuk mengatasi permasalahan *long-term dependencies* dalam pemrosesan teks (Tukino et al., 2024). Algoritma ini dapat menangkap hubungan temporal dan makna kontekstual dalam teks ulasan yang cenderung memiliki struktur kalimat yang tidak selalu teratur. Keunggulan LSTM dalam menjaga informasi yang relevan dari input sebelumnya menjadikannya sangat cocok digunakan dalam kasus analisis ulasan pelanggan, di mana konteks kalimat memiliki peran penting dalam penentuan kategori maupun sentimen(Pasaribu et al., 2020).

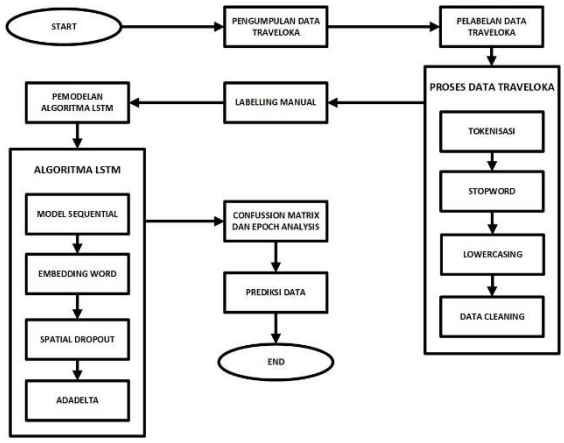
Penelitian ini bertujuan untuk menerapkan algoritma LSTM dalam melakukan klasifikasi terhadap data ulasan pelanggan Traveloka. Ulasan yang telah dikumpulkan akan diproses melalui beberapa tahap pra-pemrosesan (*preprocessing*), termasuk pembersihan teks dari angka dan simbol, penghapusan kata-kata umum (*stopword*), serta konversi ke huruf kecil(“Prediksi Mata Uang Bitcoin Menggunakan LSTM Dan Sentiment Analisis Pada Sosial Media,” 2020). Setelah melalui tahap *preprocessing*, data kemudian akan diberi label secara manual sesuai dengan kategori utama yang telah ditentukan, yakni aplikasi, teknis, harga, dan pelayanan. Tahap pelabelan ini menjadi dasar dalam proses pelatihan model untuk mendeteksi pola dalam ulasan yang ada.

Dengan pemodelan LSTM yang dibangun, sistem diharapkan mampu melakukan prediksi kategori secara otomatis terhadap data baru yang belum terklasifikasi.(Widhiyasana et al., 2021) Evaluasi kinerja model dilakukan menggunakan metrik seperti *confusion matrix* dan akurasi untuk memastikan keandalan sistem dalam mengidentifikasi kategori ulasan secara tepat. Hasil dari penelitian ini tidak hanya memberikan gambaran tentang efektivitas penerapan LSTM dalam pengolahan data ulasan berbasis teks, tetapi juga memberikan manfaat praktis bagi pihak Traveloka atau perusahaan digital lainnya dalam menangani *feedback* pelanggan secara lebih cepat, akurat, dan berbasis data.

Secara keseluruhan, penelitian ini diharapkan mampu memberikan kontribusi dalam pengembangan sistem klasifikasi ulasan berbasis *deep learning* serta membuka peluang penerapan yang lebih luas dalam bidang analisis sentimen dan *text mining*(Pemilihan Penggunaan Jenis Di Bidan Swasta Henok Hayati Novia Cahya Utami Et Al., 2023). Pendekatan ini juga dapat menjadi acuan bagi penelitian selanjutnya dalam mengeksplorasi berbagai metode kecerdasan buatan untuk pengambilan keputusan yang berbasis data konsumen.

METODE

Gambar berikut menunjukkan alur kerja penelitian yang terdiri dari tahapan pengumpulan data, *preprocessing*, pelabelan, pemodelan klasifikasi menggunakan LSTM, hingga evaluasi hasil prediksi.



Gambar 1. Flowchart Penelitian

Penelitian ini dilakukan melalui serangkaian tahapan yang sistematis, dimulai dari pengumpulan data, pemrosesan, pelabelan, hingga pemodelan menggunakan algoritma *Long Short-Term Memory* (LSTM)(Amelia & Santoso, 2023). Bahan yang digunakan dalam penelitian ini berupa data teks mentah dari ulasan pengguna, sedangkan peralatan yang digunakan meliputi perangkat lunak Python dan pustaka pendukung seperti *TensorFlow*, Keras, *NumPy*, dan Pandas. Proses pra-pemrosesan mencakup tokenisasi, normalisasi huruf (*case folding*), pembersihan kata tidak penting (*stopword removal*), dan penghapusan karakter tidak relevan.

Pelabelan dilakukan dalam dua tahap: manual oleh peneliti untuk sebagian data sebagai data validasi, dan sistem semi-otomatis untuk sebagian besar data lainnya. Model LSTM dibangun dengan pendekatan sequential yang mencakup lapisan *embedding* kata, unit LSTM, dan fungsi aktivasi *sigmoid*(Gede et al., n.d.). Proses pelatihan dan pengujian model dilakukan menggunakan parameter tertentu seperti jumlah *epoch*, *batch size*, dan *learning rate* yang disesuaikan.

Evaluasi model dilakukan dengan mengukur akurasi, *precision*, *recall*, dan *F1-score* menggunakan *confusion matrix*. Selain itu, dilakukan pula analisis *epoch* untuk mengevaluasi kinerja model sepanjang pelatihan, memastikan tidak terjadi *overfitting*.

1. *Scrapping Data*

Scraping data merupakan proses pengambilan informasi dari suatu sumber digital secara otomatis menggunakan program tertentu. Dalam penelitian ini, data diperoleh dari aplikasi Traveloka yang tersedia di *Google Play Store*, dengan memanfaatkan pustaka *google-play-scraper* pada Python.

Teknik ini memungkinkan peneliti untuk mengakses dan mengumpulkan data ulasan pengguna dalam jumlah besar secara efisien(Saputri et al., 2024). Proses scraping dilakukan dengan memperhatikan struktur elemen ulasan yang terdiri dari nama pengguna, isi komentar, rating, dan waktu unggahan, sehingga informasi yang diperoleh dapat digunakan sebagai bahan analisis yang relevan terhadap persepsi pengguna terhadap layanan aplikasi tersebut.

2. *Preprocessing Data*

Dalam penelitian ini, proses *preprocessing* dilakukan dengan membersihkan ulasan dari elemen-elemen yang tidak relevan seperti angka, simbol, dan atribut kata yang tidak memberikan kontribusi terhadap pemahaman konteks (Kurniawan & Husodo Bisowarno, 2024).

Selain itu, seluruh teks ulasan diubah ke dalam huruf kecil (*lowercase*) agar model tidak membedakan kata berdasarkan kapitalisasi, misalnya "Traveloka" dan "traveloka" dianggap sebagai kata yang sama.

Langkah-langkah ini penting untuk menyederhanakan struktur data, mengurangi kompleksitas kata, dan membantu model dalam mengenali pola yang konsisten pada data ulasan yang ada

3. Labelling Manual

Setelah melalui tahapan preprocessing, data ulasan yang telah dibersihkan kemudian dilanjutkan ke proses pelabelan secara manual. Labelling manual dilakukan dengan cara membaca dan memahami isi setiap ulasan untuk menentukan kategori yang paling sesuai berdasarkan konteks kalimat yang disampaikan pengguna.

Proses pelabelan ini dilakukan secara cermat oleh peneliti agar data dapat digunakan sebagai dasar pelatihan model klasifikasi yang lebih akurat dan relevan dengan konteks sebenarnya.

4. Proses Klasifikasi Data dan Pemodelan LSTM

Setelah data ulasan melalui proses pembersihan dan pelabelan manual berdasarkan kategori aplikasi, teknis, harga, dan pelayanan, tahap selanjutnya adalah proses klasifikasi dan pembuatan model prediksi. . LSTM memiliki struktur internal berupa sel memori dan tiga gerbang utama: forget gate, input gate, dan output gate. Sebagaimana dijelaskan sebagai berikut, perincian rumus LSTM

1. Forget Gate : $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
   Forget gate berfungsi untuk menentukan informasi mana dari cell state sebelumnya (C_{t-1}) yang akan "dilupakan" atau disimpan. Nilai f_t berada di antara 0 dan 1, dihasilkan oleh fungsi sigmoid (σ). Jika mendekati 0, maka informasi akan dilupakan, dan jika mendekati 1, maka informasi akan dipertahankan. W_f adalah bobot, $[h_{t-1}, x_t]$ adalah gabungan dari hidden state sebelumnya dan input saat ini, dan b_f adalah bias.
2. Input Gate : $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$
   Input gate bertugas untuk menentukan informasi baru apa yang akan ditambahkan ke cell state saat ini (C_t). Fungsi sigmoid σ akan memilih seberapa besar nilai dari input x_t akan memengaruhi pembaruan cell state. Semakin tinggi nilainya, semakin besar pengaruh input baru tersebut.
3. Output Gate : $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$
   Output gate mengatur informasi mana dari cell state yang akan dikeluarkan sebagai hidden state (h_t) untuk waktu ke-t. Nilai o_t akan dikalikan dengan hasil aktivasi dari cell state saat ini (biasanya menggunakan fungsi tanh), dan hasil akhirnya adalah output LSTM pada waktu tersebut.

Langkah pertama yang dilakukan adalah mengonversi data teks ke dalam bentuk numerik menggunakan Tokenizer, yaitu alat dari pustaka Keras yang digunakan untuk memetakan setiap kata unik ke dalam indeks numerik. Tokenizer ini juga disesuaikan agar hanya mengambil kata-kata yang paling sering muncul berdasarkan jumlah maksimum kata (maksimum *number of words*) (Slamet Riyadi et al., 2023). Setelah itu, seluruh teks diproses melalui tahap *padding sequences* untuk menyamakan panjang input, karena LSTM membutuhkan input dalam bentuk urutan dengan panjang yang seragam.

Model klasifikasi yang digunakan adalah *Long Short-Term Memory* (LSTM), yaitu arsitektur dari jaringan saraf dalam *kategori Recurrent Neural Network* (RNN) yang efektif dalam memahami urutan kata atau konteks dalam data teks.(Huda et al., n.d.) Dalam proses pelatihan, data dibagi menjadi dua bagian, yaitu data latih (*training set*) dan data uji (*test set*), dengan rasio tertentu agar model dapat belajar dan kemudian diuji akurasinya terhadap data yang belum pernah dilihat sebelumnya (*Klasifikasi Hasil Penjualan Minuman Ringan Pada Koperasi Berdasarkan Jenis Barang Menggunakan Algoritma K-Means Clustering*, N.D.). Selain itu, data label dikonversi ke dalam bentuk biner melalui teknik *one-hot encoding*, agar bisa diproses oleh jaringan saraf. Pemodelan ini bertujuan agar sistem dapat mempelajari

pola-pola tertentu dari setiap kategori ulasan, sehingga nantinya mampu melakukan prediksi klasifikasi terhadap data ulasan baru secara otomatis berdasarkan pola yang telah dipelajari.

5. Prediksi Data

Prediksi dilakukan dengan memasukkan data teks yang telah melalui proses tokenisasi dan padding ke dalam model yang telah dilatih (Wibiksana, 2023). Hasil dari proses ini berupa output berupa skor probabilitas yang kemudian dikonversi menjadi kelas kategori dengan nilai tertinggi. Proses ini bertujuan untuk mengidentifikasi kecenderungan kategori dari masing-masing ulasan pengguna, apakah termasuk dalam kategori aplikasi, teknis, harga, atau pelayanan. Hasil prediksi inilah yang selanjutnya dibandingkan dengan label asli untuk dievaluasi menggunakan metrik klasifikasi seperti akurasi, *precision, recall,* dan *confusion matrix*.

HASIL DAN PEMBAHASAN

Dataset

Data yang digunakan dalam penelitian ini merupakan data ulasan pengguna aplikasi Traveloka yang diperoleh melalui teknik *web scraping* menggunakan library *Google Play Scraper*. Total data yang berhasil dikumpulkan berjumlah 1.500 ulasan yang diambil dari ulasan terbaru guna memastikan bahwa data yang dianalisis merepresentasikan kondisi terkini dari pengalaman pengguna.

Tabel 1. Hasil Scrapping Data

Index	Content
	koq tampilan di dashboard saya hanya peawat dan hotel. pemesanan bus dan lainnya tidak ada
	aku lgi buka playstore buat nyari aplikasi di pencarian tpi malah lgsg arah ke traveloka.
	Klo orang mau check-in harus pada saat tanggal hari ini juga dong, masa orang check-in sekarang kepesanya buat Minggu besok, kocak nih aplikasi, jadi ribet di hotel pake acara reschedule dulu ke hari sekarang, mana lama lagi prosesnya, bikin emosi
	tertulis "bisa reschedule sampai..." tapi kelihatannya sengaja dibuat error terus, katanya proses reschedule mengikuti regulasi hotel, terus kenapa ditampilkan bisa reschedule kocak, nipu
	harga pas mau bayar tidak sesuai dengan yang tertera di Traveloka awal walaupun sudah potong asuransi & biaya layanan bisa selisih 200k

Untuk memastikan kualitas data, dilakukan proses seleksi dengan menghapus ulasan yang dinilai tidak informatif, seperti ulasan yang hanya terdiri dari satu hingga tiga kata, serta ulasan yang hanya mengandung ucapan terima kasih tanpa adanya kritik, saran, atau informasi tambahan terkait pengalaman pengguna. Setelah dilakukan penyaringan, diperoleh sebanyak 1.102 data ulasan bersih yang kemudian digunakan sebagai dataset utama dalam penelitian ini.

Setiap ulasan yang telah dibersihkan kemudian dilakukan proses pelabelan secara manual berdasarkan konteks yang terkandung di dalamnya. Pelabelan ini dilakukan oleh peneliti dengan mempertimbangkan aspek utama yang paling dominan dalam setiap ulasan. Terdapat empat kategori label yang digunakan, yaitu: (1) aplikasi, untuk ulasan yang membahas fitur, tampilan, atau performa aplikasi; (2) teknis, untuk ulasan yang berkaitan dengan kendala sistem seperti error, crash, atau loading lambat; (3) harga, untuk ulasan yang membahas harga, diskon, atau promo; dan (4) pelayanan, untuk ulasan yang berkaitan dengan respons atau layanan pelanggan. Proses pelabelan ini penting untuk mendukung proses klasifikasi selanjutnya menggunakan algoritma LSTM, sehingga setiap data memiliki konteks yang jelas dalam pengelompokan.

Tabel 3. Hasil Labelling Manual

Label_manual	Cleaned_content
Aplikasi	koq tampilan di dashboard saya hanya pesawat dan hotel pemesanan bus dan lainnya tidak ada

Teknis	aku lagi buka playstore buat nyari aplikasi di pencarian tapi malah langsung arah ke traveloka
Teknis	klo orang mau checkin harus pada saat tanggal hari ini juga dong masa orang checkin sekarang kepesannya buat minggu besok kocak nih aplikasi jadi riber di hotel pake acara reschedule dulu ke hari sekarang mana lama lagi prosesnya bikin emosi
Pelayanan	tertulis bisa reschedule sampai tapi kelihatannya sengaja dibuat error terus katanya proses reschedule mengikuti regulasi hotel terus kenapa ditampilkan bisa reschedule kocak nipu
Harga	harga pas mau bayar tidak sesuai dengan yang tertera di traveloka awal walaupun sudah potong asuransi biaya layanan bisa selisih

Setelah proses labelling, ulasan yang telah dibersihkan dan dilabeli diproses lebih lanjut menggunakan teknik tokenisasi. Dalam tahap ini, setiap kata dalam ulasan dipecah menjadi token menggunakan metode Tokenizer dari Keras. Tokenisasi dilakukan dengan filter khusus yang menghapus tanda baca dan mengubah seluruh teks ke dalam huruf kecil (*lowercase*) untuk meningkatkan konsistensi data.

Selanjutnya, hasil tokenisasi dikonversi menjadi urutan numerik berdasarkan indeks kata yang telah dibentuk oleh tokenizer. Proses ini disebut *text-to-sequence*, yang kemudian dipadatkan ke dalam panjang maksimum 250 token menggunakan *pad_sequences*. Proses ini penting agar model LSTM dapat menerima input dengan dimensi yang seragam.

LSTM

Model klasifikasi dibangun menggunakan arsitektur *Long Short-Term Memory* (LSTM), yang dikembangkan dalam bentuk model sekuensial. Arsitektur model terdiri dari beberapa lapisan utama, yakni: *Embedding Layer* untuk menerjemahkan token menjadi vektor representasi, *SpatialDropout1D* untuk mencegah *overfitting*, dan LSTM Layer yang menangkap hubungan temporal antar kata dalam ulasan. Di bagian akhir, model menggunakan lapisan Dense dengan aktivasi softmax untuk menghasilkan klasifikasi ke dalam empat kelas.

Tabel 4. SpatialDropOut dan Dense

No.	SpatialDropOut	Dense
1	0.5	128
2	0.2	100
3	0.4	64

Optimasi dilakukan menggunakan algoritma Adadelta, yang dikenal adaptif dalam menyesuaikan learning rate secara dinamis. Untuk meningkatkan performa dan mencegah *overfitting*, digunakan pula *callback* seperti *EarlyStopping* dan *ReduceLROnPlateau*.

Evaluasi

Setelah model LSTM selesai dibangun dan dilatih menggunakan data ulasan Traveloka yang telah diproses, dilakukan evaluasi kinerja model untuk mengetahui sejauh mana efektivitas model dalam mengklasifikasikan data ke dalam empat kategori utama: aplikasi, teknis, harga, dan pelayanan. Evaluasi dilakukan melalui dua pendekatan utama, yaitu melalui perhitungan metrik akurasi selama proses pelatihan dan pengujian, serta melalui analisis *confusion matrix* untuk mengamati distribusi prediksi terhadap label aktual.

Tabel 5. LSTM dan Adadelta

No.	LR	Batch Size	Epoch
1.	0.5	32	100
2.	1.0	64	100
3.	0.1	64	100

Selama proses pelatihan, akurasi model menunjukkan tren peningkatan yang konsisten pada setiap *epoch*, yang mengindikasikan bahwa model berhasil mempelajari pola-pola dari data pelatihan secara progresif. Akurasi validasi terbaik yang diperoleh mencapai sekitar 88.7%,

dengan nilai *loss* yang cenderung menurun hingga stabil mendekati akhir pelatihan. Hal ini menandakan bahwa model memiliki kemampuan generalisasi yang cukup baik dan tidak mengalami gejala *overfitting* secara ekstrem. Meskipun demikian, perbedaan kecil antara akurasi pelatihan dan validasi tetap menjadi indikator bahwa model perlu diuji lebih lanjut menggunakan data di luar sampel.

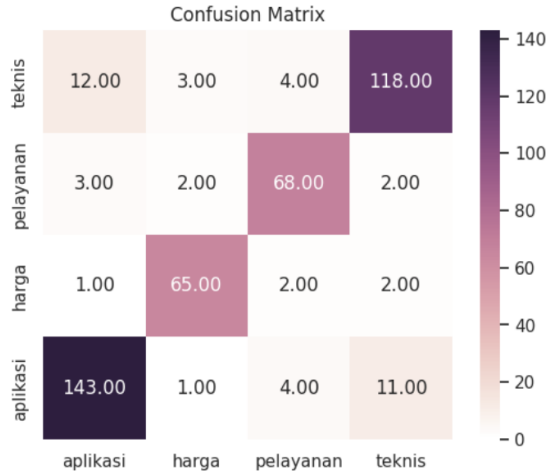
Tabel 6. Hasil Training

No	Training	
	Accuracy	Val Accuracy
1	98.73	88.70
2	99.89	88.14
3	85.25	76.27

Confusion matrix digunakan untuk melihat distribusi prediksi model terhadap setiap kategori secara lebih mendetail. Hasilnya menunjukkan bahwa prediksi terhadap kategori "aplikasi" dan "harga" memiliki jumlah klasifikasi benar (true positive) yang tinggi. Ini menunjukkan bahwa model mampu mengenali kata kunci atau pola kalimat tertentu yang khas pada ulasan terkait fitur aplikasi atau harga dengan cukup akurat. Sebaliknya, ditemukan bahwa beberapa data dengan label "teknis" sering kali salah diklasifikasikan sebagai "aplikasi", yang kemungkinan besar disebabkan oleh tumpang tindih konteks dalam penggunaan kata-kata yang berkaitan dengan performa aplikasi secara umum. Berikut *syntax* yang digunakan oleh peneliti untuk mendapatkan hasil *confussion matrix* yang sesuai dengan *output.*

```
plt.figure(figsize=(8, 6))

plt.plot(historya3.history['accuracy'], label='TrainingAccuracy',color='blue',linestyle='dashed')

plt.plot(historya3.history['val_accuracy'], label='Validation Accuracy', color='orange')

plt.xlabel('Epochs')

plt.ylabel('Accuracy')

plt.title('Training vs Validation Accuracy')

plt.legend()

plt.grid(True)

plt.show()
```

Hasil yang dikeluarkan berupa *confussion matrix* berikut



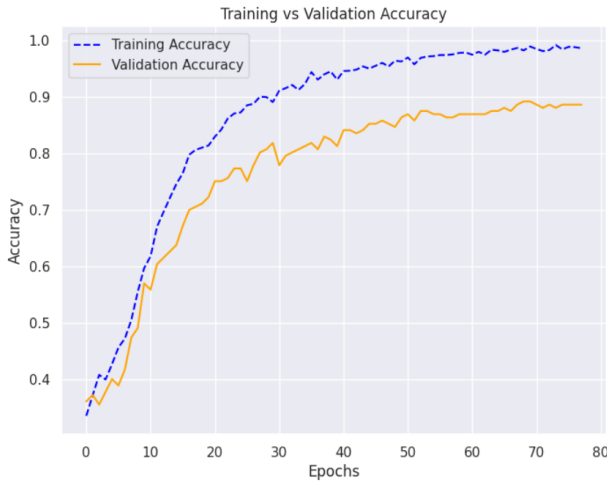
Gambar 2. Confussion Matrix

Gambar grafik akurasi menunjukkan peningkatan performa model secara bertahap selama proses pelatihan yang berlangsung selama 78 *epoch*, dengan kurva akurasi pelatihan (*training accuracy*) dan akurasi validasi (*validation accuracy*) yang meningkat secara konsisten. Pada awal pelatihan, akurasi kedua kurva masih rendah, namun mulai mengalami kenaikan signifikan setelah melewati sekitar 10 epoch pertama. Akurasi pelatihan terus meningkat hingga mendekati angka sempurna (mendekati 1.00), sementara akurasi validasi stabil di angka sekitar 0.89. Hal ini mengindikasikan bahwa model *Long Short-Term Memory* (LSTM) mampu mempelajari pola data dengan baik tanpa mengalami *overfitting* yang berarti, karena selisih antara akurasi pelatihan dan validasi tetap kecil. Stabilitas kurva validasi yang tidak mengalami penurunan drastis juga menandakan bahwa pemilihan parameter model, seperti penggunaan *dropout* dan *optimizer* Adadelta, telah berhasil menjaga generalisasi model terhadap data baru. Dengan demikian,

model LSTM yang digunakan dalam penelitian ini menunjukkan kinerja yang baik dalam mengklasifikasikan data ulasan pengguna Traveloka secara akurat dan konsisten. Berikut *syntax* yang digunakan oleh peneliti untuk mendapatkan hasil *epoch* yang sesuai dengan *output*

```
new_Y_pred = np.argmax(Y_pred, axis=1)

new_Y_test = np.argmax(Y_test, axis=1)

print(classification_report(new_Y_test, new_Y_pred))

sns.set_theme()

conf_arr = np.zeros((num_classes, num_classes))

for i in range(len(new_Y_pred)):

conf_arr[new_Y_pred[i]][new_Y_test[i]] += 1
```

Hasil yang dikeluarkan berupa hasil *epoch* berikut



Gambar 3. Diagram Hasil

Model juga menghasilkan beberapa kesalahan klasifikasi pada ulasan dengan konteks pelayanan, terutama ketika pengguna membahas masalah pelayanan teknis atau respon *customer service* yang terkait dengan masalah teknis. Hal ini menunjukkan bahwa meskipun model sudah cukup baik dalam memahami struktur kalimat, diperlukan pendekatan tambahan untuk meningkatkan sensitivitas model terhadap konteks kalimat yang bersifat ambigu atau mengandung lebih dari satu topik pembahasan.

Lebih lanjut, dilakukan pengujian dengan metrik klasifikasi lain seperti *precision*, *recall*, dan *f1-score*. Hasil menunjukkan bahwa meskipun akurasi keseluruhan cukup tinggi, nilai *precision* dan *recall* antar kelas masih bervariasi. Kelas dengan jumlah data yang lebih banyak cenderung memiliki metrik yang lebih stabil, sedangkan kelas dengan jumlah data lebih sedikit mengalami fluktuasi nilai evaluasi. Hal ini memperkuat pentingnya keseimbangan data dan perlunya eksplorasi terhadap teknik data augmentation atau *re-weighting* untuk mengatasi ketimpangan distribusi label.

Tabel 7. Evaluasi

	PRE	REC	F1	SUP
1	0.90	0.90	0.90	159
2	0.93	0.92	0.92	71
3	0.91	0.87	0.89	78
4	0.86	0.89	0.87	133
Accuracy			0.89	441
Macro	0.90	0.89	0.90	441
Weighted	0.89	0.89	0.89	441

Secara umum, hasil dari evaluasi ini menunjukkan bahwa algoritma LSTM memiliki performa yang cukup menjanjikan dalam klasifikasi teks ulasan pengguna Traveloka. Kemampuan LSTM dalam mengenali pola berurutan dan konteks kalimat terbukti bermanfaat dalam memahami makna tersembunyi dari ulasan yang kompleks. Namun, terdapat ruang untuk perbaikan, terutama dalam menangani konteks yang bersifat tumpang tindih antar kategori. Penelitian selanjutnya dapat memperkaya hasil dengan menerapkan arsitektur *hybrid* atau *attention mechanism* untuk meningkatkan akurasi model dalam menangkap makna semantik yang lebih dalam.

Dengan demikian, penerapan LSTM dalam klasifikasi teks ulasan pelanggan tidak hanya memberikan hasil yang akurat,

tetapi juga dapat dijadikan solusi praktis untuk membantu perusahaan dalam menyaring dan memahami umpan balik pelanggan secara otomatis. Model ini dapat dikembangkan lebih lanjut untuk mendukung sistem rekomendasi, peringatan dini terhadap isu layanan, hingga penilaian kualitas produk atau fitur secara berkala.

Prediksi

Model LSTM yang telah dibangun kemudian digunakan untuk melakukan prediksi terhadap data uji yang telah dipisahkan dari data latih. Hasil prediksi menunjukkan bahwa model mampu mengklasifikasikan sebagian besar ulasan pengguna Traveloka dengan cukup baik ke dalam empat kategori utama, yaitu aplikasi, harga, pelayanan, dan teknis. Hal ini terlihat dari nilai prediksi yang konsisten dengan label aktual pada sebagian besar data. Misalnya, untuk kategori aplikasi, model berhasil memprediksi dengan benar sebanyak 143 dari total 159 data, sementara pada kategori teknis juga menunjukkan kinerja yang baik dengan 118 prediksi benar dari 133 data. Untuk kategori harga dan pelayanan, meskipun terdapat sedikit kesalahan klasifikasi, hasil prediksi masih menunjukkan kecenderungan model mengenali pola yang tepat dengan akurasi yang cukup tinggi, masing-masing 65 dari 71 dan 68 dari 78 prediksi yang benar. Beberapa kesalahan klasifikasi terjadi karena kemiripan konteks antar kategori, seperti antara harga dan pelayanan, atau antara aplikasi dan teknis, yang kadang memiliki makna tumpang tindih dalam ulasan pengguna. Secara keseluruhan, hasil prediksi ini menunjukkan bahwa model LSTM yang digunakan mampu mengidentifikasi kategori ulasan dengan cukup akurat dan konsisten, memperkuat nilai evaluasi akurasi sebesar 89% yang telah diperoleh sebelumnya. Hasil ini menjadi indikator kuat bahwa model dapat digunakan untuk membantu proses klasifikasi otomatis terhadap ulasan pelanggan dalam skala besar.

SIMPULAN

Berdasarkan hasil yang diperoleh, algoritma LSTM terbukti mampu melakukan klasifikasi terhadap ulasan pengguna aplikasi Traveloka dengan performa yang tinggi. Model mencapai akurasi sebesar 89%, *precision* rata-rata 0,90, *recall* 0,89, dan *f1-score* 0,90. Nilai-nilai tersebut menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik dalam mengenali pola bahasa pada data ulasan. Dengan demikian, metode ini dapat dijadikan alternatif solusi dalam pengolahan data teks, khususnya untuk analisis opini pengguna dalam aplikasi digital. Hasil ini menunjukkan bahwa pendekatan *deep learning* berbasis LSTM sangat potensial untuk digunakan dalam analisis teks ulasan pengguna sebagai dasar pengambilan keputusan strategis di masa mendatang. Untuk penelitian selanjutnya, disarankan untuk mengeksplorasi model hibrida atau penambahan lapisan attention untuk meningkatkan akurasi serta sensitivitas terhadap nuansa makna dalam ulasan yang lebih kompleks.

DAFTAR PUSTAKA

Amelia, R., & Santoso, D. B. (2023). Prediksi Genre Film Dengan Klasifikasi Multi Kelas Sinopsis Menggunakan Jaringan Lstm Film Genre Prediction With Multi-Class Classification Synopsis Using Lstm Network. *Journal Of Information Technology And Computer Science (Intecoms)*, *6*(2).

Fernandes Andry, J., Wandy, W., Sakti Lee, F., & Ronaldo Yusup, C. (2025). Implementation Of Mental Health Data Mining Using Association And Clusterization Implementasi Data Mining Kesehatan Mental Menggunakan Asosiasi Dan Klasterisasi. *Journal Of Information Technology And Computer Science*

(Intecoms), *8*(1). Https://Www.Kaggle.Com/Datasets/Cather

Gede, I., Arya Budaya, B., Putu, L., Pratiwi, S., & Agustino, D. P. (N.D.). *Klasifikasi Sentimen Untuk Analisis Kepuasan Pelayanan Puskesmas Berbasis Arsitektur Lstm*.

Huda, B., Sembiring, I., Setiawan, I., Manongga, D., Dwi Purnomo, H., Fauzi, A., Lia Hananto, A., Kristen Satya Wacana, U., & Buana Perjuangan Karawang, U. (N.D.). *Analisis Sentimen E-Learning X Terhadap Antarmuka Pengguna Menggunakan Kombinasi Multinomial Naive Bayes Dan Pendekatan Design Thinking Analysis Of E-Learning X Sentiment On User Interface Using A Combination Of Multinomial Naive Bayes And Design Thinking Approach*. Https://Doi.Org/10.25126/Jtiik.1147678

Klasifikasi Hasil Penjualan Minuman Ringan Pada Koperasi Berdasarkan Jenis Barang Menggunakan Algoritma K-Means Clustering. (N.D.).

Kurniawan, W., & Husodo Bisowarno, B. (2024). Deep Learning Dengan Metode Lstm (Long Short-Term Memory) Untuk Pemodelan Kelakuan Dinamik Sistem Distilasi Menggunakan Real Operating Data. In *Jurnal Integrasi Proses* (Vol. 13, Issue 1). Http://Jurnal.Untirta.Ac.Id/Index.Php/Jip

Pasaribu, D. J. M., Kusrini, K., & Sudarmawan, S. (2020). Peningkatan Akurasi Klasifikasi Sentimen Ulasan Makanan Amazon Dengan Bidirectional Lstm Dan Bert Embedding. *Inspiration: Jurnal Teknologi Informasi Dan Komunikasi*, *10*(1). Https://Doi.Org/10.35585/Inspir.V10i1.2568

Pemilihan Penggunaan Jenis Di Bidan Swasta Henok Hayati Novia Cahya Utami, U. K., Lia Hananto, A., & Shofiah Hilabi, S. (2023). Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naïve Bayes. *Journal Of Social Science Research*, *3*, 503–513.

Prediksi Mata Uang Bitcoin Menggunakan Lstm Dan Sentiment Analisis Pada Sosial Media. (2020). *Jurnal Ilmiah Komputasi*, *19*(4). Https://Doi.Org/10.32409/Jikstik.19.4.370

Putra, T. I. Z. M., Suprapto, S., & Bukhori, A. F. (2022). Model Klasifikasi Berbasis Multiclass Classification Dengan Kombinasi Indobert Embedding Dan Long Short-Term Memory Untuk Tweet Berbahasa Indonesia. *Jurnal Ilmu Siber Dan Teknologi Digital*, *1*(1), 1–28. Https://Doi.Org/10.35912/Jisted.V1i1.1509

Saputri, N. K. T. A., Gunadi, I. G. A., & Sunarya, I. M. G. (2024). Analisis Sentimen Pelayanan Daring Di Fakultas Teknik Dan Kejuruan Universitas Pendidikan Ganesha Menggunakan Algoritma Naïve Bayes Dan Lstm. *Malcom: Indonesian Journal Of Machine Learning And Computer Science*, *4*(3), 1120–1129. Https://Doi.Org/10.57152/Malcom.V4i3.1336

Slamet Riyadi, A., Puspa Wardhani, I., & Andi Perdana, Dan. (2023). Aplikasi Perbandingan Prediksi Harga Bitcoin Menggunakan Deep Learning Dengan Metode Arima, Sarima, Ltsm Dan Gradient Boosting Regressor. *Seminar Nasional Teknologi Informasi Dan Komunikasi Sti&K (Sentik)*, *7*(1).

Sovia Pramudita, A., & Agustia, R. E. (N.D.). *E-Service Quality Dan E-Promotion Terhadap Keputusan Pembelian Konsumen Pada Website Traveloka.Com*.

Http://Ejurnal.Poltekpos.Ac.Id/Index.Php/Competitive|105

Tukino, Hananto, A., Nanda, R. A., Novalia, E., Sediyono, E., & Sanjaya, J. (2024). Lstm And Word Embedding: Classification And Prediction Of Puskesmas Reviews Via Twitter. *E3s Web Of Conferences*, *500*. Https://Doi.Org/10.1051/E3sconf/202450001018

Wibiksana, H. (2023). Prediksi Harga Saham Pada Portofolio Investor Dengan Analisis Time Series Harga Saham Menggunakan Ann. *Bulletin Of Computer Science Research*, *4*(1), 10–17. Https://Doi.Org/10.47065/Bulletincsr.V4i1.298

Widhiyasana, Y., Semiawan, T., Gibran, I., Mudzakir, A., & Noor, M. R. (2021). Penerapan Convolutional Long Short-Term Memory Untuk Klasifikasi Teks Berita Bahasa Indonesia (Convolutional Long Short-Term Memory Implementation For Indonesian News Classification). In *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi* | (Vol. 10, Issue 4).