



IMPLEMENTASI ALGORITMA *LONG SHORT-TERM MEMORY* PADA ISU KENAIKAN UANG KULIAH TUNGGAL TERHADAP MINAT KULIAH MAHASISWA

Muhammad Saharullah Raiya¹, Muhammad Raihan Putra Khamil², Nur Fadillah³, Rizal Adi Saputra⁴

^{1,2,3,4} Teknik Informatika, Universitas Halu Oleo
Kendari, Sulawesi Tenggara, Indonesia 93232

mosahrulraiya.22123@gmail.com, muhraihanputrakhamil@gmail.com, nurfadillahaskari@gmail.com,
rizaladisaputra@uho.ac.id

Abstract

The increase in Single Tuition Fee (UKT) has become increasingly concerning for many prospective students and parents in Indonesia. The significant rise in UKT has led many prospective students to decide not to pursue higher education, resulting in a decline in college enrollment interest. This study aims to analyze the impact of the UKT increase on the declining interest in higher education among prospective students using the Long Short-Term Memory (LSTM) algorithm in Machine Learning. Based on sentiment data from YouTube and Instagram regarding the UKT increase issue, the LSTM model is implemented to predict college enrollment interest based on changes in UKT. The results of this study indicate a significant correlation between the UKT increase and the decline in new student enrollment. The Long Short-Term Memory (LSTM) algorithm model effectively predicts the impact of tuition fee increases on prospective students' interest in higher education. This is supported by an accuracy of 84%, showcasing the model's ability to recognize historical patterns in the data used optimally.

Keywords: Data mining, Long Short-Term Memory, Machine Learning, Student, UKT

Abstrak

Kenaikan Uang Kuliah Tunggal (UKT) menjadi isu yang semakin memprihatinkan bagi banyak calon mahasiswa dan orang tua di Indonesia. Keputusan kenaikan UKT secara signifikan membuat banyak calon mahasiswa memutuskan untuk tidak melanjutkan pendidikan ke perguruan tinggi. Sehingga berdampak pada penurunan minat kuliah. Penelitian ini bertujuan untuk menganalisis dampak kenaikan Uang Kuliah Tunggal (UKT) terhadap penurunan minat kuliah calon mahasiswa baru menggunakan algoritma *Long Short-Term Memory* (LSTM) dalam *Machine Learning*. Berdasarkan data sentimen dari platform youtube dan instagram terkait isu kenaikan UKT, model LSTM diimplementasikan untuk memprediksi minat kuliah berdasarkan isu perubahan UKT. Hasil penelitian menunjukkan bahwa terdapat korelasi yang signifikan antara kenaikan UKT dan penurunan jumlah pendaftaran mahasiswa baru. Model algoritma *Long Short-Term Memory* (LSTM) menunjukkan efektivitas dalam memprediksi dampak kenaikan uang kuliah tunggal terhadap minat kuliah calon mahasiswa. Hal ini dibuktikan dengan akurasi sebesar 84%, yang menunjukkan kemampuan model dalam mengenali pola historis dari data yang digunakan secara optimal.

Kata kunci: Data mining, Long Short-Term Memory, Machine Learning, Mahasiswa, UKT

1. PENDAHULUAN

Uang Kuliah Tunggal (UKT) adalah sistem yang digunakan dalam membiayai perkuliahan yang harus dibayar oleh mahasiswa Perguruan Tinggi Negeri yang berada di bawah Kementerian Riset, Teknologi dan Pendidikan Tinggi (Kemristek Dikti)[1]. Dalam hal ini, pembiayaan biaya pendidikan tidak sama dengan masa lalu, dimana biaya pendidikan mahasiswa terdiri dari beberapa komponen

tersendiri, seperti biaya pendidikan, biaya pelatihan, biaya orang tua mahasiswa (IOM), biaya skripsi, biaya wisuda, biaya kuliah, dan biaya (SPKP). Dengan diberlakukannya UKT, mahasiswa akan membayar satu jenis biaya pendidikan saja yaitu UKT [2].

Munculnya isu kenaikan Uang Kuliah Tunggal (UKT) di perguruan tinggi saat ini menjadi topik hangat pembicaraan yang menimbulkan berbagai reaksi di kalangan masyarakat.

UKT adalah sistem pembayaran yang diperkenalkan oleh Kementerian Pendidikan dan Kebudayaan untuk meringankan beban biaya pendidikan mahasiswa berdasarkan kemampuan ekonomi keluarga mereka. Namun, kebijakan ini tidak terlepas dari kendala serta tantangan, terutama ketika munculnya isu kenaikan UKT yang signifikan. Dimana kenaikan ini dianggap memberatkan dan tidak sejalan dengan kondisi ekonomi sebagian besar masyarakat. Sehingga menjadi sebuah hambatan bagi calon mahasiswa baru untuk melanjutkan kependidikan perguruan tinggi [3].

Untuk mengetahui dampak kenaikan UKT terhadap minat kuliah calon mahasiswa, teknologi *analysis* data canggih dapat digunakan. Salah satu metode yang efektif adalah menggunakan algoritma *Long Short-Term Memory* (LSTM) [4]. LSTM merupakan jenis jaringan saraf tiruan yang dirancang untuk mengenali pola dalam data berurutan dan sangat efektif dalam menangani masalah yang melibatkan urutan waktu. Dengan menggunakan algoritma LSTM, kita dapat menganalisis data historis tentang kenaikan UKT dan minat kuliah, serta membuat prediksi yang lebih akurat tentang bagaimana perubahan UKT dimasa depan akan mempengaruhi keputusan calon mahasiswa untuk melanjutkan Pendidikan tinggi [5].

Pendekatan ini memungkinkan pemangku kepentingan Pendidikan untuk membuat keputusan yang lebih terinformasi. Dengan lebih memahami implikasi pertumbuhan UKT, Universitas dan pemerintah dapat merancang kebijakan yang meningkatkan minat belajar meskipun biaya Pendidikan meningkat. Penelitian ini diharapkan dapat memberikan kontribusi penting dalam mencari keseimbangan antara kebutuhan finansial Pendidikan dan ketersediaan calon mahasiswa baru.

Dengan adanya prediksi yang lebih akurat, pemangku kepentingan pendidikan, termasuk universitas dan pemerintah, dapat membuat keputusan yang lebih terinformasi dalam merancang kebijakan yang tetap menjaga minat belajar mahasiswa meskipun biaya pendidikan mengalami kenaikan. Kebijakan yang didukung oleh analisis data yang kuat dapat membantu menciptakan keseimbangan antara kebutuhan finansial pendidikan dan aksesibilitas calon mahasiswa baru. Oleh karena itu, penelitian ini diharapkan dapat memberikan kontribusi penting dalam upaya menjaga kesetaraan akses pendidikan tinggi, terutama bagi kelompok masyarakat yang rentan terhadap dampak kenaikan biaya kuliah.

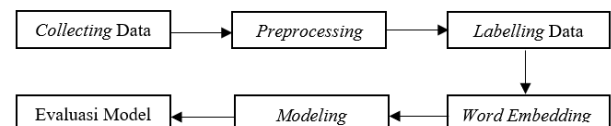
Penelitian Sebelumnya [6] menunjukkan bahwa model LSTM mampu mengklasifikasikan sentimen dengan tingkat akurasi terbaik mencapai 80,42%. Model ini dilatih menggunakan *dataset* berisi 658 *tweet*, yang mencerminkan bagaimana opini masyarakat terhadap kebijakan MBKM dapat diprediksi secara lebih akurat menggunakan *deep learning*. Penelitian lain [7] menunjukkan bahwa model LSTM juga efektif dalam menganalisis sentimen berita

online. Hasil penelitian ini menunjukkan bahwa model LSTM mencapai tingkat akurasi sebesar 81,36%, yang membuktikan keunggulan algoritma ini dalam menangani teks berbahasa Inggris dengan kompleksitas yang lebih tinggi. Penelitian [8] juga membuktikan efektivitas LSTM dalam mendeteksi ujaran kebencian di media sosial. Model ini, setelah melalui pra-pemrosesan, mencapai akurasi 83% dengan nilai rata-rata makro 65%, menunjukkan keunggulan LSTM dalam analisis sentimen.

Berdasarkan efektivitas LSTM dalam penelitian sebelumnya, penelitian ini bertujuan untuk menerapkan model LSTM dalam menganalisis dampak kenaikan uang kuliah tunggal terhadap minat calon mahasiswa. Model ini digunakan untuk memprediksi perubahan minat berdasarkan data yang ada, sehingga dapat membantu institusi pendidikan dalam merancang kebijakan yang lebih tepat.

2. METODE PENELITIAN

Metode penelitian ini terdiri dari beberapa tahapan, yaitu studi literatur, pengumpulan data, *preprocessing*, *labelling* data, *word embedding*, *modeling*, dan yang terakhir evaluasi model. Data dikumpulkan dari platform *YouTube* dan *Instagram*, kemudian diproses menggunakan teknik *preprocessing* sebelum diberi label sentimen. Selanjutnya, dilakukan *word embedding* untuk mengubah teks menjadi bentuk numerik yang dapat dipahami oleh model. Model LSTM dibangun dan dilatih untuk menganalisis serta memprediksi *trend* minat kuliah berdasarkan sentimen publik. Terakhir, model dievaluasi menggunakan metrik seperti akurasi, *precision*, *recall*, dan *F1-score* seperti yang dijelaskan pada Gambar 1.



Gambar 1. Alur Metode dan Perancangan

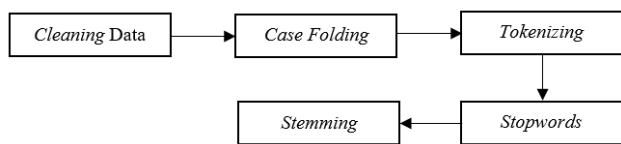
2.1 Collecting Data

Proses *collecting data* pada penelitian ini, menggunakan metode *crawling data* dalam proses pengumpulan informasi atau data yang berkaitan dengan penelitian. Untuk sumber data berasal dari *platform Youtube* dan *platform Instagram*, dengan memanfaatkan komentar pada postingan dengan menggunakan *keyword* keputusan kenaikan UKT mahasiswa 2024. Jumlah data komentar yang dikumpulkan sebanyak 2583 data.

2.2 Preprocessing Data

Preprocessing teks adalah langkah yang dilakukan sebelum mengategorikan data agar lebih cocok untuk dianalisis atau diproses lebih lanjut [9]. Tujuan dari prosedur ini adalah untuk mengekstraksi data dari unsur-unsur yang tidak perlu untuk memperoleh informasi yang berkualitas tinggi dan

relevan. Ada beberapa tahapan dalam proses *preprocessing* ini antara lain *cleaning data*, *case folding*, *stopwords*, *tokenizing*, *stemming* seperti yang diilustrasikan pada Gambar 2.



Gambar 2. Alur Tahapan *Preprocessing*

a. *Cleaning Data*

Pada tahapan ini dilakukan proses pembersihan data seperti menghapus *hashtag*, *link*, gambar, *retweet*, *emoticon* serta simbol-simbol yang tidak bermakna pada komentar.

b. *Case Folding*

Pada tahap ini proses perubahan jenis tulisan dilakukan, dengan mengubah semua teks menjadi huruf kecil atau *lower* [10].

c. *Tokenizing*

Pada tahapan ini dilakukan pemberian token dengan cara membagi teks menjadi beberapa potongan kecil atau diubah ke unit-unit kecil yang disebut dengan token [11].

d. *Stopwords*

Setelah data menjadi token, kemudian dilanjutkan dengan proses pengurangan kata-kata yang tidak penting atau bermakna. Penghapusan kata *stopwords* membantu dalam *focus* pada kata-kata kunci yang lebih relevan dalam menentukan *sentiment* [12]. Kemudian data tersebut akan dimasukkan ke dalam daftar *stoplist* yang sudah dibuat. Setiap data yang masuk akan dihapus.

e. *Stemming*

Stemming adalah proses mengubah kata dari bahasa lisan ke bentuk akarnya, atau akar kata. Tujuan dari *stemming* adalah untuk mengurangi frekuensi kata atau turunannya sehingga kata-kata yang mirip dapat diidentifikasi sebagai satu kesatuan dalam analisis teks. *Stemming* membantu dalam penyederhanaan kata-kata dengan akar yang sama [13]. Tahapan ini adalah tahapan akhir dalam *preprocessing*.

2.3 Labelling Data

Pelabelan data adalah proses pemberian *tag* atau kategori pada data sehingga dapat diidentifikasi dan disusun berdasarkan kriteria tertentu. Pada proses ini dilakukan pelabelan secara manual. Jumlah data yang telah melalui *preprocessing* adalah 1000. Dari data tersebut ditetapkan tingkat penurunan dan kenaikan terhadap minat dari calon mahasiswa baru. Untuk penurunan berlabel negatif atau 0 dan kenaikan berlabel positif atau 1.

2.4 Word Embedding

Word Embedding merupakan proses menggabungkan kata-kata menjadi vektor panjang yang sudah ditetapkan sehingga tidak dibatasi dengan kosakata yang lebih banyak [14]. Setelah proses pembersihan data dan *labelling*. Selanjutnya masuk ke proses perubahan kata-kata yang memiliki hubungan *semantic* dengan angka dan dimasukkan ke dalam kata-kata. Dalam tahapan *word embedding* ada beberapa metode yang digunakan yaitu vektorisasi kata, *learning representations*, dan *using representation*.

a. Vektorisasi Kata

Setiap kata dipetakan ke vektor numerik unik. Pada tahapan ini, setiap kata dalam korpus (kelompok teks) diwakili oleh panah atau vektor jenis lain. Vektor ini kemudian dioptimalkan selama proses pembelajaran. Metode yang digunakan adalah *one-hot encoding*, dimana metode ini merepresentasikan vektor dengan panjang sama dengan ukuran korpus, dengan satu elemen bernilai 1 dan sisanya bernilai 0.

b. *Learning Representations*

Vektor-vektor kata dioptimalkan melalui proses pembelajaran yang berbasis jaringan saraf atau algoritme statistik untuk mencerminkan *semantic* dan kontekstual hubungan antara kata-kata. Kata-kata yang memiliki arti mirip akan memiliki vektor yang dekat di ruang vektor.

c. *Using Representation*.

Setelah diperoleh, representasi kata vektor digunakan dalam berbagai tugas NLP seperti klasifikasi teks, identifikasi entitas, pemahaman teks, dan sebagainya. Kata vektor yang dihasilkan dapat digunakan sebagai masukan untuk model pembelajaran mesin atau jaringan saraf yang lebih kompleks.

2.5 Modelling

Modelling merupakan langkah penting dalam membangun model LSTM yang bertujuan untuk menganalisis data secara lebih akurat. Proses ini terdiri dari beberapa tahapan utama, yaitu *splitting* data, vektorisasi teks, dan arsitektur jaringan LSTM. Proses *modeling* terbagi menjadi beberapa tahapan, yaitu *splitting* data, *vektorisasi* teks, dan arsitektur jaringan LSTM.

a. *Splitting* Data

Data dipisahkan menjadi set pelatihan dan pengujian. Analisis data dilakukan dengan pembagian rasio 80:20, di mana parameter *test_size = 0.2* menunjukkan bahwa sekitar 20% dari total *dataset* digunakan untuk pengujian. Dalam hal ini, sebanyak 800 data digunakan untuk proses pelatihan, sementara 200 data sisanya dialokasikan untuk pengujian. Selain itu, data diacak dengan *seed* sebesar 42 untuk memastikan konsistensi dalam proses pembagian.

b. Vektorisasi Teks

Pada langkah ini, teks divektorkan menggunakan modul Keras *tokenizer*. Vektorisasi teks menyediakan setiap bilangan bulat unik. Hasilnya, setiap kata akan diidentifikasi menggunakan algoritma bulat. Dengan jumlah kata teratas 20.000, jumlah teks ulasan 100, dan *train_df* menggunakan parameter label.

c. Arsitektur Jaringan

Dalam LSTM, pemrosesan memori yang lebih kompleks dapat dicapai melalui penggunaan *dynamic cell states* dan *gates*. Secara umum, *internal system state* untuk melupakan iterasi sebelumnya, *input gates* untuk menambahkan informasi atau input baru ke isi sel, dan *output gates* untuk mengontrol jumlah penyesuaian pada iterasi saat ini. sehingga, informasi lama dapat dibuang, sementara informasi baru dapat ditransfer ke memori dengan *lancer* [15]. Pemodelan mengacu pada proses pembuatan model LSTM [16]

$$\text{Forget Gate} \quad f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (2.1)$$

$$\text{Input Gate} \quad i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2.2)$$

$$C_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad (2.3)$$

$$\text{Output Gate} \quad o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (2.4)$$

$$\text{Hidden State} \quad h_t = o_t * \tanh(C_t) \quad (2.5)$$

Keterangan:

σ = Fungsi aktivasi *sigmoid*, mengubah input menjadi nilai antara 0 dan 1

tanh = Fungsi aktivasi *tanh*, mengubah input menjadi nilai antara -1 dan 1

W_f, W_i, W_C, W_o = Matriks bobot untuk *forget gate*, *input gate*, *candidate cell gate*, dan *output gate*.

b_f, b_i, b_C, b_o = Vektor bias untuk *forget gate*, *input gate*, *candidate cell gate*, dan *output gate*.

H_{t-1} = *Hidden state* dari langkah waktu sebelumnya.

X_t = Input pada waktu t

$[H_{t-1}, x_t]$ = Konkatenasi dari hidden state sebelumnya dan input saat ini.

2.6 Evaluasi Model

Evaluasi Model adalah proses menilai kinerja sebuah *machine learning* dengan menggunakan berbagai simetrik untuk menentukan seberapa baik model tersebut. Pada tahapan ini, dilakukan kinerja model yang melibatkan penggunaan matriks konfusi untuk menentukan analisis benar atau salah. Dalam konteks isu penurunan minat bagi mahasiswa baru terhadap keputusan kenaikan UKT, kita dapat menggunakan estimasi ketidaksepakatan untuk

mengevaluasi kinerja model yang bertujuan untuk mengetahui bagaimana pengaruhnya yaitu “Positif (Kenaikan)”, dan “Negatif (Penurunan)”. Ini adalah salah satu pendekatan validasi model yang paling sering digunakan untuk menentukan generalisasi model ke data baru.

a. *Precision* dalam konteks klasifikasi mengacu pada kemampuan model untuk mengidentifikasi positif yang sebenarnya secara akurat. Artinya, *precision* mengukur proporsi prediksi positif yang benar-benar positif. Jika mendekati 1, maka model sangat tinggi atau akurat. Jika bernilai mendekati 0,5 maka model tergolong moderat atau agak akurat. Sedangkan jika mendekati 0 maka, model tergolong rendah atau sering salah.

$$\text{Precision} = TP / (TP + FP) \quad (2.6)$$

b. *Recall* merupakan kemampuan model untuk mengidentifikasi semua sampel positif yang sebenarnya, tidak peduli model klasifikasi sampel lain dengan benar atau tidak. Artinya, *recall* mengukur proporsi sampel positif yang diklasifikasikan sebagai positif.

$$\text{Recall} = TP / (TP + FN) \quad (2.7)$$

c. *F1-score* merupakan model yang menggabungkan *precision* dan *recall* untuk memberikan gambaran yang lebih seimbang tentang kinerja model.

$$F1\text{-score} = \frac{2 * (\text{precision} * \text{recall})}{(\text{precision} + \text{recall})} \quad (2.8)$$

d. Akurasi adalah model evaluasi performa model dalam pembelajaran mesin yang mengukur seberapa akurat model dalam memprediksi label (kelas) dari data baru. Akurasi dihitung sebagai persentase jumlah prediksi yang benar dibagi dengan jumlah total prediksi yang dibuat.

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.9)$$

Keterangan:

TP(*True Positive*) = Jumlah sampel yang benar-benar positif dan diklasifikasikan sebagai positif.

TN(*True Negative*) = Jumlah sampel negatif yang diklasifikasikan dengan benar sebagai negatif.

FP(*False Positive*) = Jumlah sampel yang sebenarnya negatif tetapi diklasifikasikan sebagai positif (positif palsu)

FN(*False Negative*) = Jumlah sampel yang sebenarnya positif tetapi diklasifikasikan sebagai negatif(negatif palsu)

3. HASIL DAN PEMBAHASAN

Pada penelitian ini, dibuat sebuah sistem untuk mengklasifikasikan penurunan minat kuliah calon mahasiswa akibat kenaikan Uang Kuliah Tunggal (UKT). Algoritma yang digunakan pada penelitian ini adalah algoritma *Long Short-Term Memory* (LSTM) dan juga bahasa pemrograman python untuk menganalisis data yang telah dikumpulkan. Dalam penelitian ini, menggunakan data dari *platform* sosial media yaitu *Youtube* dan *Instagram* terkait kenaikan UKT dan minat kuliah calon mahasiswa untuk pembuatan sistem analisis. Dari hasil dari analisis ditetapkan dua kelas, yaitu positif berarti kenaikan dan negatif berarti penurunan.

Setelah *dataset* dikumpulkan sebanyak 1000 data komentar dari *platform youtube* dan *instagram* dilanjutkan dengan tahapan *preprocessing* untuk memanipulasi *dataset* agar menghasilkan data yang berkualitas dan efisien. Data ini sudah melalui lima tahapan *preprocessing* seperti *cleaning data*, *case folding*, *stopwords*, *tokenizing*, *stemming*. Dengan menerapkan proses *preprocessing* yang tepat, model LSTM yang digunakan dalam penelitian ini dapat bekerja lebih optimal dalam mengklasifikasikan data sentimen. Hal ini memungkinkan sistem memberikan hasil analisis yang lebih akurat dan dapat diandalkan dalam menilai dampak kenaikan UKT terhadap minat kuliah calon mahasiswa. Berikut Hasil tahapan *preprocessing* dapat dilihat pada tabel 1.

Tabel 1. Hasil Preprocessing Data

Hasil Tahapan Preprocessing
lulus ptn senang ukt naikkan sedih orang tua bilang kerja saja
Niat lulus langsung kuliah milih kerja
Kuliah ugm data orang tua biaya sana hasil kuliah kampus kota
Mahasiswa baru pun tdk semua mampu om. Mereka ingin memperbaiki keluarga, dr tidak sekolah ortunya
anaknya sekolah hingga perguruan tinggi. Dari hanya pedagang kecil anaknya lulus s 1 bisa memperbaiki ekonomi keluarga. Sekarang lulusan sma/smu itu peluang kerjanya apa si?? Maksimal masa kerjanya berpa lama si??
Ponakan saya, ibunya honorer ayah tidak bekerja masih tinggal dgn nenek TDK punya rumah sendiri tapi masih kena UKT mahal, tolak ukurnya apa sih sebenarnya

3.1 Pembagian Data Testing dan Data Training

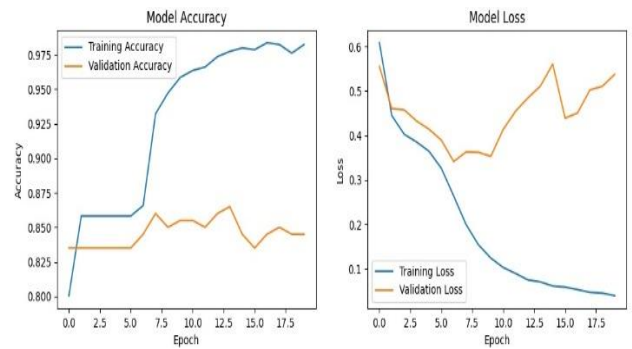
Dalam penelitian ini menggunakan sebanyak 1000 data komentar. Kemudian dilakukan pembagian data dengan 2 kategori yaitu data *training* dan data *testing*, dimana dalam pembagian ini menggunakan bahasa pemrograman *Python*. Data dibagi menjadi 80% data *training* dan 20% data *testing*. Dari hasil di atas didapatkan data *Training* sebesar 800 data komentar dan untuk data *Testing* sebesar 200 data komentar terkait isu kenaikan UKT, sehingga hasil inilah yang digunakan dalam penelitian ini.

3.2 Arsitektur LSTM

Lapisan pertama dalam arsitektur jaringan adalah *embedding layer*, yang menggunakan vektor dengan dimensi sepanjang 32 dan jumlah kata sebanyak 200. Selanjutnya, lapisan kedua adalah LSTM layer dengan 100 neuron. Lapisan terakhir merupakan *Dense layer* dengan 1 neuron serta fungsi aktivasi sigmoid. Karena penelitian ini bertujuan untuk klasifikasi biner, yaitu menentukan sentimen positif atau negatif, digunakan *loss function binary_crossentropy* dari Keras serta algoritma optimasi Adam. Parameter lainnya meliputi *batch size* sebesar 64, jumlah *epoch* sebanyak 20, serta *validation split* sebesar 0.2 atau 20%. Berdasarkan parameter yang telah dijelaskan, berikut adalah arsitektur jaringan yang terbentuk dalam model ini seperti yang ditunjukkan pada Tabel 2.

Tabel 2. Arsitektur LSTM

Model: "sequential"		
Layer (type)	Output Shape	Param #
<i>embedding (Embedding)</i>	(None, 200, 32)	320032
<i>lstm (LSTM)</i>	(None, 100)	53200
<i>dense (Dense)</i>	(None, 1)	101
Total Params: 373333 (1.42 MB)		
Trainable Params: 373333 (1.42)		
Non-trainable Params: 0 (0.00 byte)		



Gambar 3. Grafik Pelatihan LSTM

Berdasarkan Gambar 3 Hasil pelatihan dari model LSTM terhadap kasus isu kenaikan UKT dan dampaknya terhadap penurunan minat calon mahasiswa baru. Grafik-grafik tersebut menunjukkan bahwa nilai akurasi pada data pelatihan mencapai sekitar 97.5% pada *epoch* ke-20. Namun, nilai akurasi pada data validasi cenderung stabil, berada di sekitar 85%. Grafik Model Loss pada gambar di atas menggambarkan perubahan tingkat kerugian (*loss*) model LSTM selama proses pelatihan. Kerugian pelatihan seiring berjalannya waktu, model meningkatkan kemampuannya untuk belajar dari data. *Validation loss* mulai menurun, namun kemudian stabil dan fluktuatif,

menyebabkan potensi *overfitting*, dimana model tidak dapat digeneralisasikan dengan baik pada data baru.

Dengan menggunakan model pelatihan terbaik, dilakukan pengujian terhadap berbagai kalimat yang diinputkan ke dalam sistem. Model LSTM akan mengklasifikasikan setiap kalimat ke dalam kategori sentimen positif atau negatif berdasarkan pola yang telah dipelajari selama pelatihan. Semakin banyak data yang digunakan dalam proses pelatihan, semakin baik performa algoritma dalam mengenali pola sentimen, sehingga meningkatkan akurasi prediksi.

Selain jumlah data, faktor lain yang memengaruhi kinerja model adalah tahapan *preprocessing* yang dilakukan sebelum data dimasukkan ke dalam model. Proses ini mencakup pembersihan data, normalisasi teks, serta teknik lanjutan seperti *POS-Tagging (Part-of-Speech Tagging)* untuk memahami struktur kata dalam kalimat dan *Negation Handling* guna menangani kata-kata negatif yang dapat mengubah makna suatu kalimat. Dengan penerapan *preprocessing* yang optimal, model LSTM dapat mengekstraksi sentimen dengan lebih akurat, sehingga meningkatkan kualitas klasifikasi yang dihasilkan.

3.3 Evaluasi Model

Hasil yang ditunjukkan dalam tabel 3 di bawah ini menunjukkan bahwa metode LSTM yang disarankan untuk sentimen analisis dapat menghasilkan akurasi yang tinggi. Ini karena kernel LSTM sangat cocok untuk masalah klasifikasi data teks. LSTM adalah metode yang terbukti dapat menghasilkan hasil akurasi yang tinggi dalam mengklasifikasikan data teks.

Tabel 3. Hasil Evaluasi Model

	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Support</i>
Negatif	0.88	0.94	0.91	167
Positif	0.52	0.33	0.41	33
<i>Accuracy</i>			0.84	200
<i>Macro Avg</i>	0.70	0.64	0.66	200
<i>Weighted Avg</i>	0.82	0.84	0.82	200

Berdasarkan hasil pengujian yang dilakukan, implementasi metode yang digunakan berhasil mengklasifikasikan sentimen data dengan nilai akurasi 0,84. Pada tahap *preprocessing*, 800 data digunakan untuk pelatihan dan 200 data untuk pengujian. Model dapat mengklasifikasikan analisis sentimen menjadi dua kelas yaitu sentimen negatif dan sentimen positif, dengan proporsi 83,5% sentimen negatif dan 16,5% sentimen positif dari total 200 data uji.

4. KESIMPULAN

Dalam penelitian ini, penggunaan algoritma *Long Short-Term Memory (LSTM)* berhasil mengklasifikasikan

sentimen terkait isu kenaikan UKT dengan akurasi 0,84 atau 84%. Model dapat mengidentifikasi dua jenis sentimen, yaitu negatif dan positif, dengan lebih banyak data yang diklasifikasikan sebagai negatif. Hasil evaluasi menunjukkan bahwa kinerja model kurang optimal pada kategori sentimen positif, dengan nilai *Precision*, *Recall*, dan *F1-score* yang lebih rendah dibandingkan kategori negatif. Untuk meningkatkan hasil ini, disarankan untuk menyeimbangkan jumlah data positif dan negatif agar model lebih akurat dalam mengklasifikasikan kedua jenis sentimen.

DAFTAR PUSTAKA

- [1] N. Setyawati, "Proses Mofofenemik Prefiks {Meng-} Dalam Pemberitaan Seputar Kenaikan Ukt Di Jawapos.Com," *Journal Of Language And Literature Education (Jolale)*, Vol. 1, No. 2, Pp. 3046–7373, 2024.
- [2] S. Rokhman, F. Rozi, And R. A. Asmara, "Pengembangan Sistem Penunjang Keputusan Penentuan Ukt Mahasiswa Dengan Menggunakan Metode Moora Studi Kasus Politeknik Negeri Malang," *Jurnal Informatika Polinema*, Vol. 3, No. 4, Pp. 36–43, 2017.
- [3] Fatmah, "Maraknya Aksi Mahasiswa Menentang Kenaikan Ukt Dan Uang Pengembangan," *Jurnal Pembelajaran Dan Pengembangan Diri*, Vol. 4, No. 2, Pp. 493–502, 2024, Doi: 10.47353/Bj.V4i2.339.
- [4] A. Hanafiah, Y. Arta, H. O. Nasution, And Y. D. Lestari, "Penerapan Metode Recurrent Neural Network Dengan Pendekatan Long Short-Term Memory (Lstm) Untuk Prediksi Harga Saham," *Bulletin Of Computer Science Research*, Vol. 4, No. 1, Pp. 27–33, Dec. 2023, Doi: 10.47065/Bulletincsr.V4i1.321.
- [5] Y. Ardian Pradana, I. Cholissodin, And D. Kurnianingtyas, "Analisis Sentimen Pemindahan Ibu Kota Indonesia Pada Media Sosial Twitter Menggunakan Metode Lstm Dan Word2vec," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, Vol. 7, No. 5, Pp. 2389–2397, May 2023, [Online]. Available: [Http://J-Ptiik.Ub.Ac.Id](http://J-Ptiik.Ub.Ac.Id)
- [6] S. J. Pipin And H. Kurniawan, "Analisis Sentimen Kebijakan Mbkm Berdasarkan Opini Masyarakat Di Twitter Menggunakan Lstm," *Jurnal Sifo Mikroskil*, Vol. 23, No. 2, Pp. 197–208, 2022.
- [7] N. M. Sedana Krisna, I. N. Wijaya Saputra, And I. K. Artana Resika, "Analisis Sentimen Berbahasa Inggris Dengan Metode Lstm Studi Kasus Berita Online Pariwisata Bali English Sentiment Analysis Using The Lstm Method Case Study Of Bali Tourism Online News," *Jurnal Teknologi Informasi*

- Dan Ilmu Komputer (Jtiik)*, Pp. 1325–1334, 2024, Doi: 10.25126/Jtiik.2024118792.
- [8] L. Yosia Wibowo, N. Annisa, P. Ananda Khairunnisa Viktor Handrianus Pranatawijaya, And R. Priskila, “Implementasi Long Short-Term Memory Dalam Analisis Sentimen Pengguna Aplikasi Twitter Yang Mengandung Ujaran Kebencian,” *Jurnal Mahasiswa Teknik Informatika*, Vol. 8, No. 3, Pp. 3170–3174, Jun. 2024.
- [9] D. J. M. Pasaribu, K. Kusriani, And S. Sudarmawan, “Peningkatan Akurasi Klasifikasi Sentimen Ulasan Makanan Amazon Dengan Bidirectional Lstm Dan Bert Embedding,” *Inspiration: Jurnal Teknologi Informasi Dan Komunikasi*, Vol. 10, No. 1, Jun. 2020, Doi: 10.35585/Inspir.V10i1.2568.
- [10] D. Sintia Amelia And A. Ari Aldino, “Teks Dan Analisis Sentimen Pada Chat Grup Whatsapp Menggunakan Long Short Term Memory (Lstm),” *Jurnal Teknologi Dan Sistem Informasi (Jtsi)*, Vol. 2, No. 4, Pp. 56–61, 2021, [Online]. Available: [Http://Jim.Teknokrat.Ac.Id/Index.Php/Jtsi](http://jim.teknokrat.ac.id/index.php/jtsi)
- [11] M. Kamal Wisyaldin, G. Maya Luciana, H. Pariaman, And P. Pembangkitan Jawa Bali, “Pendekatan Long Short-Term Memory Untuk Memprediksi Kondisi Motor 10 Kv Pada Pltu Batubara,” *Kilat*, Vol. 9, No. 2, Pp. 311–318, 2020, Doi: 10.33322/Kilat.V9i2.997.
- [12] M. Apriliyani, M. Izzal Musyaffaq, S. Nur, M. Rini Handayani, And K. Umam, “Implementasi Analisis Sentimen Pada Ulasan Aplikasi Duolingo Di Google Playstore Menggunakan Algoritma Naïve Bayes,” *Aiti: Jurnal Teknologi Informasi*, Vol. 21, No. 2, Pp. 298–311, 2024.
- [13] A. Pranata And N. Azmi Verdikha, “Klasifikasi Teks Quick Count Pemilihan Presiden 2024 Pada Twitter Menggunakan Metode Tf-Idf Dan Naive Bayes,” *Jurnal Informatika Terpadu*, Vol. 10, No. 2, Pp. 93–100, 2024, [Online]. Available: [Https://Journal.Nurulfikri.Ac.Id/Index.Php/Jit](https://journal.nurulfikri.ac.id/index.php/jit)
- [14] M. David Rhman, A. Djunaidy, And F. Manahanto, “Penerapan Weighted Word Embedding Pada Pengklasifikasian Teks Berbasis Recurrent Neural Network Untuk Layanan Pengaduan Perusahaan Transportasi,” *Jurnal Sains Dan Seni Its*, Pp. 2337–3520, 2021.
- [15] S. Fachri And P. J. Ramdan, “Pemodelan Machine Learning : Analisis Sentimen Masyarakat Terhadap Kebijakan Ppkm Menggunakan Data Twitter,” *Jurnal Sistem Cerdas*, Vol. 5, No. 1, Pp. 12–20, 2022, [Online]. Available: [Https://T.Co/lenucgfuuuj](https://t.co/lenucgfuuuj)
- [16] D. R. Alghifari, M. Edi, And L. Firmansyah, “Implementasi Bidirectional Lstm Untuk Analisis Sentimen Terhadap Layanan Grab Indonesia,” *Jurnal Manajemen Informatika (Jamika)*, Vol. 12, No. 2, Pp. 89–99, Sep. 2022, Doi: 10.34010/Jamika.V12i2.7764.