Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 873−883
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7398

Perbandingan Algoritma LSTM, BI-LSTM, dan CNN untuk Klasifikasi Komentar Masyarakat: Pembangkitan Serigala Direwolf pada Media X

Agung Novriyandi, Nirwana Hendrastuty*

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
Email: [1]Agung_noriyandi@teknokrat.ac.id, [2,*]nirwanahendrastuty@teknokrat.ac.id
Email Penulis Korespondensi: nirwanahendrastuty@teknokrat.ac.id
Submitted: **19/05/2025**; Accepted: **30/06/2025**; Published**: 30/06/2025**

Abstrak−Pembangkitan Serigala Direwolf oleh Colossal Biosciences, sebuah perusahaan bioteknologi yang berbasis di Dallas, Texas, AS, melalui teknologi kloning dan pengeditan gen telah memicu perdebatan dan diskusi luas di masyarakat. Kloning adalah proses pembuatan salinan genetik yang identik dari suatu organisme, dan dalam konteks ini, digunakan untuk menghidupkan kembali Serigala Direwolf, spesies yang telah punah sekitar 12.500 tahun lalu. Penelitian ini bertujuan untuk membandingkan kinerja algoritma Long Short-Term Memory (LSTM), Bidirectional LSTM (BI-LSTM), dan Convolutional Neural Network (CNN) dalam mengklasifikasikan komentar masyarakat terkait pembangkitan Serigala Direwolf pada Media X. Dengan menggunakan dataset yang terdiri dari 3400 komentar, setelah melalui tahap *cleaning* dan *preprocessing* untuk menghilangkan *noise* dan memperbaiki kualitas data, diperoleh 1424 komentar yang valid, yang terdiri dari 869 komentar berlabel negatif, 270 komentar berlabel positif, dan 285 komentar berlabel netral. Dalam Penelitian ini data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Penelitian ini akan mengevaluasi kinerja ketiga algoritma tersebut berdasarkan metrik seperti akurasi, *presisi,* dan *recall*. Hasil evaluasi menunjukkan bahwa model LSTM memiliki akurasi tertinggi sebesar 73%, diikuti oleh BI-LSTM sebesar 70%, dan CNN sebesar 66%. Berdasarkan hasil tersebut, pendekatan LSTM dapat dianggap sebagai pendekatan yang lebih baik dalam mengklasifikasikan komentar masyarakat terkait topik pembangkitan Serigala Direwolf. Hasil penelitian ini diharapkan dapat memberikan informasi yang berguna bagi pengembangan sistem analisis sentimen dan memahami opini masyarakat terkait teknologi kloning dan pengeditan gen.

Kata Kunci: Pembangkitan Serigala Direwolf; Analisis Sentimen; LSTM; BI-LSTM; CNN; Kloning dan Pengeditan Gen; Klasifikasi Komentar

Abstract−The resurrection of the Direwolf by Colossal Biosciences, a biotechnology company based in Dallas, Texas, USA, through cloning and gene editing technology has sparked widespread debate and discussion in society. Cloning is the process of creating a genetically identical copy of an organism, and in this context, it is used to bring back the Direwolf, a species that has been extinct for around 12,500 years. This study aims to compare the performance of Long Short-Term Memory (LSTM), Bidirectional LSTM (BI-LSTM), and Convolutional Neural Network (CNN) algorithms in classifying public comments related to the resurrection of the Direwolf on Media X. Using a dataset of 3400 comments, after undergoing cleaning and preprocessing to eliminate noise and improve data quality, 1424 valid comments were obtained, consisting of 869 negative, 270 positive, and 285 neutral comments. This study will evaluate the performance of the three algorithms based on metrics such as accuracy, precision, and recall. The evaluation results show that the LSTM model has the highest accuracy at 73%, followed by BI-LSTM at 70%, and CNN at 66%. Based on these results, the LSTM approach can be considered a better approach in classifying public comments related to the topic of Direwolf resurrection. The results of this study are expected to provide useful information for the development of sentiment analysis systems and understanding public opinion related to cloning and gene editing technology.

Keywords: Direwolf Resurrection; Sentiment Analysis; LSTM; BI-LSTM; CNN; Cloning and Gene Editing; Comment Classification

1. PENDAHULUAN

Perkembangan media sosial dan *platform* online telah menghasilkan volume data yang besar, termasuk opini dan sentimen masyarakat. Dalam konteks ini, X dipilih sebagai *platform* sosial yang diteliti. Menurut data We Are Social pada Oktober 2023, X memiliki 666,2 juta pengguna di seluruh dunia, menjadikannya salah satu platform media sosial terkemuka. Platform ini sering digunakan untuk analisis data terkait tren dan opini publik [1]. *Platform X, yang sebelumnya dikenal sebagai X, memungkinkan penggunanya untuk berbagi informasi, pengalaman, dan pandangan pribadi secara langsung dan interaktif* [2]. Serigala Dire (Aenocyon dirus) adalah spesies serigala besar yang telah punah dan hidup pada Zaman Es di Amerika Utara dan Selatan. Mereka juga dikenal sebagai "serigala mengerikan" karena ukuran dan kekuatan mereka [3]. Serigala Dire (Aenocyon dirus) spesies serigala besar yang telah punah, dengan ukuran tubuh yang mengesankan. Mereka memiliki panjang tubuh 1,5-1,8 meter dan berat hingga 100 kg, membuatnya lebih besar dari serigala abu-abu modern. Meskipun mirip secara visual, analisis genetik menunjukkan bahwa Serigala Dire adalah spesies yang berbeda dengan garis keturunan terpisah, yang hidup berdampingan dengan serigala abu-abu tanpa bukti perkawinan silang [4]. Ilmuwan di Colossal Biosciences mengklaim telah berhasil "menghidupkan kembali" serigala dire yang telah punah melalui penyuntingan gen canggih, dan telah melahirkan tiga anak serigala yang membawa DNA serigala dire setelah lebih dari 12.000 tahun [5]. Para peneliti berhasil mengumpulkan *genom Aenocyon dirus* yang berkualitas tinggi dari DNA purba yang diekstrak dari dua fosil dire wolf, memungkinkan mereka memiliki informasi genetik yang lengkap tentang spesies tersebut [6]. keberhasilan penelitian asal ilmuan AS telah memicu polemik dalam beberapa waktu terakhir. Hal ini memicu munculnya berbagai opini dan pendapat di media sosial, terutama di *X*

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 873−883
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7398

Analisis sentimen dapat digunakan untuk memahami reaksi masyarakat terhadap pembangkitan serigala dire dengan mengidentifikasi dan mengklasifikasikan opini publik yang tersebar di platform digital seperti media sosial, forum diskusi, dan artikel berita [7]. Analisis sentimen adalah teknik komputasi yang digunakan untuk mengidentifikasi dan mengklasifikasikan emosi atau sentimen dalam teks digital, dengan hasil yang dikategorikan menjadi tiga jenis utama: positif, netral, dan negatif [8]. Algoritma klasifikasi teks yang umum digunakan antara lain LSTM (Long Short-Term Memory), BI-LSTM (Bidirectional Long Short-Term Memory), dan CNN (Convolutional Neural Network). LSTM bekerja dengan memproses data sekuensial untuk mengenali pola jangka panjang[9], sementara BI-LSTM melakukan hal serupa namun dengan mempertimbangkan konteks dari kedua arah [10], CNN menggunakan filter konvolusi untuk mengenali pola lokal dalam data teks [11].

Penelitian sebelumnya telah menggunakan analisis sentimen untuk memahami opini masyarakat terkait pandemi COVID-19 di X. Perbandingan antara metode CNN dan Bi-LSTM dengan menggunakan Word2Vec menunjukkan bahwa Bi-LSTM memiliki performa lebih baik dengan akurasi 86%, sedangkan CNN memiliki akurasi 84%. Analisis sentimen dapat menjadi pendekatan yang efektif untuk memahami opini masyarakat terkait vaksinasi COVID-19, berdasarkan hasil penelitian sebelumnya [12]. Selain itu, Penelitian lain juga telah menggunakan algoritma LSTM untuk analisis sentimen kebijakan MBKM berdasarkan opini masyarakat di X, dengan hasil akurasi terbaik sebesar 80,42% setelah dilatih menggunakan dataset 658 tweet [9].

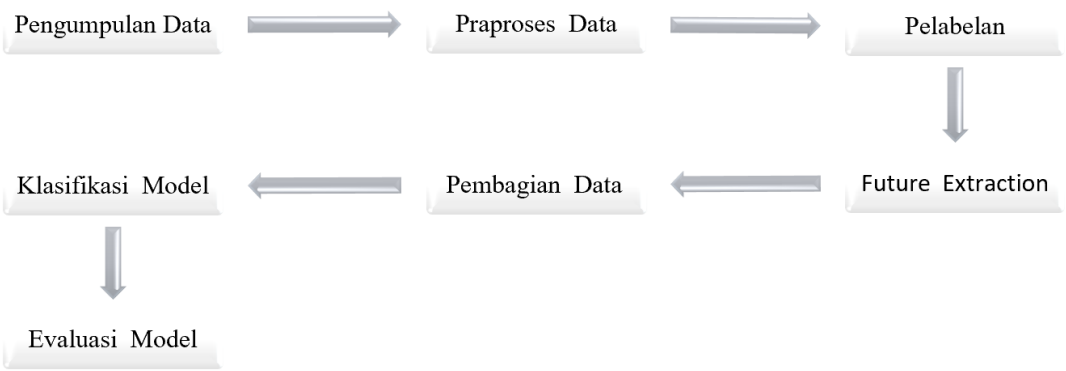
Penelitian terkait topic yang sama juga dilakukan dengan judul "Etika Pengobatan/Rekayasa Genetik dalam Islam sebagai Implikasi untuk Terapi Gen dan Teknologi DNA serta Tantangan Kontemporer" membahas topik kloning DNA dan mengidentifikasi tantangan seperti risiko manipulasi genetik, eugenika, dan konsekuensi etis. Penelitian ini menekankan pentingnya dialog terbuka antara ilmuwan, etika, dan komunitas agama untuk merumuskan panduan yang sesuai dengan nilai-nilai agama. Selain itu, penelitian lanjutan diperlukan untuk mengeksplorasi aplikasi praktis dan dampak jangka panjang terapi gen dalam konteks etika Islam serta mengembangkan kebijakan yang sejalan dengan prinsip-prinsip agama.

Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma LSTM, BI-LSTM, dan CNN dalam menganalisis sentimen terkait pembangkitan serigala dire wolf. Dengan menggunakan ketiga algoritma tersebut, penelitian ini juga berfokus pada mengidentifikasi tren sentimen masyarakat, termasuk sentimen positif, negatif, dan netral. Hasil penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik tentang efektivitas algoritma-algoritma tersebut dalam melakukan klasifikasi sentimen dan membandingkan akurasi ketiga algoritma untuk menentukan mana yang paling efektif dalam konteks pembangkitan serigala dire wolf.

2. METODOLOGI PENELITIAN

2.1 Tahapan Analisis Sentimen

Penelitian ini dilakukan melalui 7 tahap, meliputi pengumpulan data, *preprocessing, labeling* sentimen (positif, negatif, netral), ekstraksi fitur dengan *TF-ID*F, dan pembagian data menjadi set pelatihan dan pengujian. Tiga algoritma—*LSTM, BI-LSTM, dan CNN*—digunakan untuk klasifikasi sentimen, dengan *LSTM* mengenali pola jangka panjang, *BI-LSTM* memahami konteks dua arah, dan *CNN* mendeteksi pola lokal dalam teks. Proses dan tahapan penelitian secara keseluruhan ditunjukkan dalam Gambar 1.



Gambar 1. Tahapan Penelitian

Penelitian ini dilakukan dengan membandingkan tiga algoritma guna memperoleh pemahaman yang lebih komprehensif mengenai performa masing-masing dalam konteks pemrosesan data. Tujuan utama dari studi ini adalah memberikan kontribusi praktis yang dapat membantu para ilmuwan dalam merancang dan mengevaluasi penelitian di masa depan, khususnya dalam pengambilan keputusan yang berkaitan dengan teknik kloning *DNA* hewan purba, seperti serigala direwolf. Data yang digunakan berasal dari platform *X,* yang kemudian dianalisis menggunakan ketiga algoritma tersebut. Evaluasi kinerja model dilakukan dengan menggunakan sejumlah metrik pengukuran, yaitu *akurasi, presisi, recall, dan F1-Score.*

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 873−883
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7398

2.3 Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan mengakses API X menggunakan metode Tweet Harvest pada platform Google Colab dengan bahasa pemrograman Python. Proses crawling dilakukan untuk mengumpulkan tweet yang relevan dengan topik penelitian, yaitu serigala dire wolf dan kloning DNA, dengan menggunakan kata kunci seperti "Serigala dire wolf", "ilmuan AS", "kloning DNA", "Serigala purba", dan kata kunci terkait lainnya dalam periode waktu tertentu. Hasil dari proses crawling ini kemudian diolah dan disimpan dalam format file CSV, dengan total jumlah data yang terkumpul sebanyak 3400 data, yang selanjutnya akan diproses lebih lanjut dalam tahap preprocessing untuk memastikan kualitas data sebelum dilakukan analisis lanjutan.

Crawling data X adalah proses pengumpulan data yang menggunakan *API X* untuk mengambil informasi terkait pengguna atau *tweet* tertentu. Proses ini melibatkan pembuatan program yang dirancang untuk mencari dan mengumpulkan data berdasarkan kata kunci yang telah ditentukan sebelumnya [13]. *Tweet-Harvest* adalah sebuah *library open-source* yang dirancang untuk mengumpulkan *tweet* berdasarkan kata kunci tertentu dan batasan jumlah tweet yang diinginkan, sehingga memungkinkan pengguna untuk memperoleh data *tweet* yang relevan dengan lebih mudah dan efisien [14].

2.4 *Pre-processing*

Tujuan dari preprocessing data adalah untuk meningkatkan kualitas data dengan membersihkan dan menstrukturkan data sehingga dapat meningkatkan akurasi model prediksi. Pada data X, preprocessing sangat penting karena bahasa yang digunakan sering kali mengandung karakteristik yang tidak penting dan perlu dihilangkan untuk mengoptimalkan analisis sentimen [15]. Dalam penelitian ini, beberapa metode preprocessing yang digunakan yaitu

2.4.1 *Cleaning*

Data cleaning adalah proses pembersihan data dari elemen-elemen yang tidak relevan atau tidak penting, seperti *URL*, nama pengguna, *retweet, kode HTML, dan hashtag*, untuk meningkatkan kualitas dan akurasi data yang digunakan dalam analisis [16].

2.4.2 *Case Folding*

Case folding adalah proses mengubah semua teks menjadi huruf kecil untuk menyeragamkan format dan memudahkan pengolahan data, sehingga analisis teks dapat dilakukan dengan lebih efektif dan akurat [17].

2.4.3 *Normalisasi*

Normalisasi dalam analisis sentimen adalah proses untuk mengubah data teks ke dalam format yang standar dan konsisten, sehingga memudahkan proses analisis selanjutnya. Tujuannya adalah untuk mengurangi variasi yang tidak perlu dalam data, seperti perbedaan huruf besar-kecil, tanda baca, atau format teks, sehingga model analisis dapat fokus pada makna dan sentimen yang terkandung dalam teks [18]**.**

2.4.3 *Tokenizing*

Tokenizing adalah proses memecah kalimat menjadi kata-kata individu, sehingga setiap kata dapat diolah secara terpisah. Setelah melalui proses cleansing dan case folding, kalimat tweet dipecah menjadi kata-kata yang terpisah untuk analisis lebih lanjut [19].

2.4.4 *Filtering / Stopword removal*

Stopword removal adalah teknik preprocessing yang digunakan untuk menghilangkan kata-kata umum yang tidak memiliki makna penting dalam analisis teks, seperti "ke", "dari", "dan", atau "atau". Tujuannya adalah untuk meningkatkan efisiensi dan akurasi analisis teks dengan menghilangkan kata-kata yang tidak berkontribusi pada makna yang signifikan.[20]

2.4.5 *Stemming*

Stemming adalah proses mengubah kata-kata berimbuhan menjadi bentuk dasarnya dengan menghilangkan awalan dan akhiran, sehingga diperoleh kata dasar yang murni. Tujuannya adalah untuk memudahkan analisis dan pengolahan teks dengan menyeragamkan kata-kata yang memiliki makna serupa [18].

2.5 Pelabelan Data

Pelabelan adalah proses mengkategorikan data ke dalam kelas atau kategori tertentu berdasarkan sentimen atau karakteristik yang relevan, sehingga data dapat diolah dan dianalisis lebih lanjut dengan lebih terstruktur.[21] Pelabelan dalam penelitian ini digunakan untuk mengklasifikasikan data menjadi tiga kelas sentimen, yaitu positif, negatif, dan netral menggunakan Lexicon Based, Lexicon Based adalah metode analisis sentimen yang memanfaatkan kamus kata-kata sentimen untuk menentukan sentimen teks. Metode ini sederhana, efektif, dan praktis untuk menganalisis data dari media sosial seperti X dan Facebook. yang menghasilkan pelabelan data sebanyak 1.424 komentar, terdiri dari 869 komentar negative, 285 netral dan 270 komentar positif.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 873−883
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7398

2.6 Ekstraksi Fitur

Ekstraksi fitur (feature extraction) dalam analisis teks bertujuan mengubah teks menjadi bentuk yang dapat dipahami oleh mesin pembelajaran. Metode Term Frequency (TF) dan Inverse Document Frequency (IDF) digunakan untuk menghitung bobot kata dalam teks berdasarkan frekuensinya. TF mengukur frekuensi kemunculan kata dalam dokumen, sedangkan IDF mengukur kepentingan kata di seluruh koleksi dokumen. Kombinasi keduanya, TF-IDF, membantu menentukan signifikansi kata dalam konteks yang lebih luas.[22]
a. Term Frequency (IDF) :

$$TF(d,t) = \frac{jumlah\ kemunculan\ term\ t\ dalam\ dokumen\ d}{total\ jumlah\ term\ dalam\ dokumen\ d} \quad (1)$$

TF mengukur seberapa sering sebuah kata muncul dalam teks dibandingkan dengan jumlah total kata.
b. Inverse Document *Frequency* (IDF)

$$IDF(t,d) = \log\left(\frac{jumlah\ total\ dokumen\ d}{jumlah\ dokumen\ dalam\ D\ yang\ mengandung\ term\ t+1}\right. \quad (2)$$

c. TF-IDF

$$\text{TF-IDF (t,d,D)} - \text{TF(td) x IDF (t,D)} \quad (3)$$

TF-IDF dihasilkan dari kombinasi nilai TF dan IDF, memberikan skor yang lebih akurat dengan mempertimbangkan frekuensi istilah dalam dokumen (TF) dan signifikansinya di seluruh koleksi dokumen (IDF). Ini membantu menentukan kepentingan istilah dalam konteks yang lebih luas.

2.7 Pembagian Data

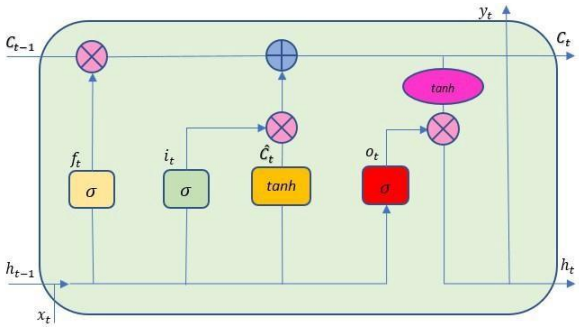
Pembagian data menjadi data pelatihan dan data uji sangat penting dalam pembuatan model. Data pelatihan digunakan untuk melatih model mengenali pola, sedangkan data uji digunakan untuk mengevaluasi kinerjanya. Dalam penelitian ini, data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian, untuk mengoptimalkan hasil model.

2.8 Klasifikasi Model

Model ini dilatih pada data pelatihan untuk mengenali pola tertentu dan kemudian menerapkan pengetahuan tersebut untuk memprediksi kategori atau label dari data baru yang belum pernah dilihat. Keakuratan model sangat bergantung pada kualitas data pelatihan, algoritma yang efektif, dan kemampuan generalisasi model terhadap pola yang dipelajari.:

2.9.1 *Bidirectional LSTM*

Algoritma Deep Learning seperti LSTM menawarkan performa lebih baik dan waktu komputasi yang lebih cepat dibandingkan algoritma machine learning tradisional. LSTM dirancang untuk mengatasi masalah vanishing gradient dan dapat menyimpan informasi dalam jangka panjang dengan menggunakan tiga gerbang utama: input gate, forget gate, dan output gate, sehingga efektif untuk tugas klasifikasi kata dan analisis data sekuensial[23]



Gambar 3. Arsitektur LSTM

Forget gate pada LSTM memilah informasi pada cell state dengan menentukan apakah informasi tersebut dibuang atau disimpan. Jika forget gate bernilai 0, informasi akan dibuang, sedangkan jika bernilai 1, informasi akan disimpan dalam cell state. Proses ini memungkinkan LSTM untuk mempertahankan atau menghapus informasi yang relevan atau tidak relevan dalam jangka panjang.

$$\text{ft} = \sigma(\text{Wf}.[\text{ht}-1,\text{Xt}] + \text{bf}) \quad (4)$$

Input gate pada LSTM memproses informasi melalui dua lapisan: lapisan sigmoid yang menentukan informasi mana yang akan diperbarui sesuai persamaan 2, dan lapisan tanh yang menghasilkan kandidat nilai baru sesuai persamaan 3. Cell state kemudian dihasilkan dari penggabungan nilai output kedua lapisan tersebut.

$$\text{it} = \sigma(\text{Wi}.[\text{ht}-1,\text{Xt}] + \text{bi}) \quad (5)$$

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 873−883
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7398

$$Ct = \tanh(Wc . [ht-1, Xt] + bc) \quad (6)$$

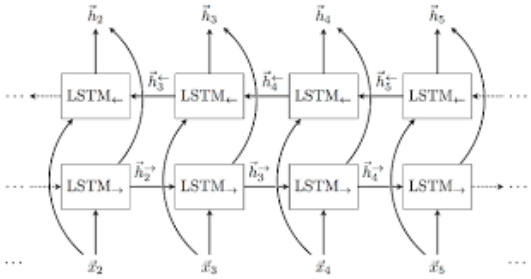
Setelah cell state diperbarui menggunakan persamaan 4, proses dilanjutkan ke output gate. Di sini, lapisan sigmoid menentukan bagian mana dari cell state yang akan menjadi output berdasarkan persamaan 5. Selanjutnya, output tersebut diproses melalui lapisan tanh sesuai persamaan 6 untuk menghasilkan output akhir LSTM.

$$Ct = ft . Ct-1 + it . Ct \quad (7)$$

$$Ot = \sigma(Wo . [ht-1, Xt] + bo \quad (8)$$

$$ht = Ot . \tanh(Ct) \quad (9)$$

LSTM memiliki keterbatasan karena hanya memproses kata dari satu arah (awal hingga akhir). Untuk mengatasi ini, penelitian ini menggunakan Bidirectional LSTM yang dapat memproses kata dari dua arah: awal hingga akhir dan akhir hingga awal. Dengan dua lapisan LSTM terpisah untuk arah maju dan mundur, Bidirectional LSTM dapat menangkap konteks kata secara lebih komprehensif sesuai persamaan 7.



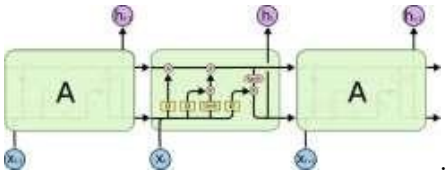
Gambar 4. Arsitektur Bidirectional LSTM

ht BiLSTM = ht forward ⊕ ht backward

Output Bidirectional LSTM dihasilkan dari penggabungan output lapisan LSTM maju ($ht\ forward$) dan lapisan LSTM mundur ($ht\ backward$) melalui operasi concatenation (⊕), sehingga model dapat menangkap konteks kata dari kedua arah secara komprehensif

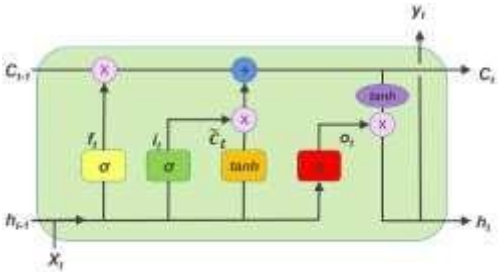
2.9.2 *LSTM Classification*

STM, yang diperkenalkan oleh Hochreiter & Schmidhuber (1997), adalah variasi RNN yang dirancang untuk mengatasi masalah penyimpanan informasi jangka panjang dalam proses pembelajaran. Dengan menggantikan node RNN tradisional di hidden layer dengan LSTM cells, model ini mampu menyimpan atau melupakan informasi sebelumnya secara selektif, sehingga lebih efektif dalam menangani data sekuensial yang kompleks.[24]



Gambar 2. RNN dengan LSTM Cells

LSTM memiliki tiga gerbang utama: Input Gate, Forget Gate, dan Output Gate. Ketiga gerbang ini berfungsi untuk mengatur aliran informasi dengan menentukan informasi mana yang disimpan, dilupakan, atau dikeluarkan. Berikut adalah persamaan-persamaan yang digunakan dalam proses tersebut.



Gambar 3. LSTM Cell

$$ft = \sigma(wf\,[ht-1, xt] + bf) \quad (9)$$

$$ot = \sigma(wo\,[ht-1, xt] + bo) \quad (10)$$

$$ct = ft * ct-1 + it * \tilde{c}t \quad (11)$$

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 873−883
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7398

$$it = \sigma(wi\,[ht{-}1, xt\,] + bi) \tag{12}$$

$$\tilde{c}t\ =\ \tanh(\mathrm{wc}\,[\mathrm{ht}-1,\mathrm{xt}\,]+\ \mathrm{bc}) \tag{13}$$

$$\mathrm{ht}\ =\ \mathrm{ot}\ *\ \tanh(\mathrm{c\,t}\,) \tag{14}$$

LSTM mengelola informasi melalui tiga gerbang: Forget Gate, Input Gate, dan Output Gate. Forget Gate menentukan informasi yang dilupakan dari cell state sebelumnya, Input Gate menentukan informasi baru yang ditambahkan, dan Output Gate menentukan output berdasarkan cell state. Proses ini diatur oleh beberapa persamaan matematika yang memungkinkan LSTM menyimpan, melupakan, dan mengeluarkan informasi secara selektif untuk tugas-tugas yang melibatkan data sekuensial.

2.9.3 Algoritma *CNN*

Convolutional Neural Network (CNN) adalah jenis jaringan saraf tiruan deep feed-forward yang dapat mengidentifikasi pola dalam data, seperti ekspresi dalam kalimat, tanpa terpengaruh oleh posisi kata-kata. CNN terdiri dari tiga jenis layer utama: Convolutional Layer, Pooling Layer, dan Fully-Connected Layer. Pada Convolutional Layer, filter digunakan untuk mengkonvolusikan data dan menghasilkan feature maps yang merepresentasikan pola-pola penting dalam input. Proses ini diatur oleh operasi konvolusi yang dinyatakan dalam persamaan matematika tertentu.[25]

$$\mathrm{FMa,b} = \mathrm{bias} + \sum_c^C \sum_d^D \mathrm{Zc,d} + \mathrm{Xa} + \mathrm{c} - 1, \mathrm{b} + \mathrm{d} - 1 \tag{15}$$

Pooling layer memastikan bahwa jaringan hanya fokus pada pola yang paling penting serta data dirangkum dengan menggeser jendela melintasi feature maps, kemudian menerapkan beberapa operasi linear atau non linear pada data yang ada pada jendela. Poolinglayer memiliki fungsi untuk mengurangi dimensi dari feature maps yang akan digunakan pada layer selanjutnya.

$$(f(FMa,b)\ =\ \{\,FMa,b,jika\,FMa,b\ \geq\ 0 \tag{16}$$

$$\{\,0,jika\,FMa,b\ <\ 0$$

Dimana Rumus ReLU (Rectified Linear Unit) untuk feature map FM_{a,b} adalah f(FM_{a,b}) yang mengubah nilai menjadi 0 jika FM_{a,b} < 0 dan membiarkan nilai tetap sama jika FM_{a,b} ≥ 0, sehingga f(FM_{a,b}) = FM_{a,b} untuk nilai positif dan f(FM_{a,b}) = 0 untuk nilai negatif, yang secara efektif menghilangkan nilai negatif dan mempertahankan nilai positif dalam proses pembelajaran model neural network. Layer terakhir yang digunakan adalah fully-connectedlayer. layer ini digunakan untuk memahami pola yang dihasilkan dari layer sebelumnya. Neuron pada layer ini memiliki koneksi penuh ke semua aktivasi pada layer sebelumnya. Metode CNN juga menggunakan fungsi aktivasi yang dilakukan ketika berada di antara convolutionallayer dan poolinglayer. Aktivasi di antara kedua layer tersebut menggunakan fungsi aktivasi ReLU. Sedangkan untuk fungsi aktivasi output menggunakan softmax. Persamaan fungsi aktivasi ReLU terdapat pada Persamaan diatas.

Fungsi aktivasi softmax mempunyai tujuan untuk mendapatkan hasil klasifikasi serta menghasilkan nilai yang diinterpretasi sebagai probabilitas yang belum dinormalisasi untuk tiap kela Nilai kelas yang dihitung dengan menggunakan fungsi softmax ditunjukan oleh Persamaan dibawah

$$yijk = \frac{exijk}{\sum_{t=1}^{D} exijt} \tag{17}$$

Fungsi terakhir adalah loss function untuk menghitung loss (nilai error)dengan menggunakan categorical cross-entropy. Persamaan 5 adalah loss function yang dimaksud.

$$\mathrm{Llog(Y,Ypred)} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{C=1}^{C} 1yiecc\log P\ model[yi\varepsilon Cc \tag{18}$$

Rumus Cross-Entropy Loss digunakan untuk mengukur perbedaan antara distribusi probabilitas sebenarnya dan distribusi probabilitas prediksi model dalam masalah klasifikasi multi-kelas, dengan menghitung nilai loss berdasarkan label sebenarnya dan label prediksi, sehingga model dapat dioptimalkan untuk meningkatkan kemampuan klasifikasinya dengan meminimalkan nilai loss tersebut.

2.10 Evaluasi Model

Pengujian model BI-LSTM, LSTM dan CNN dilakukan untuk mengevaluasi kinerja model dalam memprediksi data dengan akurat, dengan menggunakan Confusion Matrix untuk menampilkan hasil prediksi dan menghitung metrik evaluasi seperti presisi, recall, skor F1, dan akurasi pada setiap data uji.

$$Accuracy\ =\ \frac{TP+TN}{TP+TN+FP+FN}x100\% \tag{19}$$

$$Precision = \frac{TP}{TF+FP} \tag{20}$$

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 873−883
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7398

$$Recall = \frac{TF}{TF+FN} \quad (21)$$

$$F1\ Score = 2x\frac{Precession\ x\ Recall}{Precession + Recall} \quad (22)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan mengakses API X menggunakan metode Tweet Harvest pada platform Google Colab dengan bahasa pemrograman Python. Proses crawling dilakukan untuk mengumpulkan tweet yang relevan dengan topik penelitian, yaitu serigala dire wolf dan kloning DNA, dengan menggunakan kata kunci seperti "Serigala dire wolf", "ilmuan AS", "kloning DNA", "Serigala purba", dan kata kunci terkait lainnya dalam periode waktu tertentu. Hasil dari proses crawling ini kemudian diolah dan disimpan dalam format file CSV untuk dianalisis sentiment pada Tabel 1

Tabel 1. Hasil Pengumpulan Data

Username	*Tweet*
Cymatics13	@oldyzach Apakah efek suara Anda bermasalah?
Nuhsmith655	@EmergeAmerica @KamalaHarris Terima kasih telah mengundangnya. Kami menantikannya

3.2 *Preprocessing*

Proses preprocessing data meliputi beberapa tahap penting, yaitu: pembersihan data, konversi huruf menjadi huruf kecil, pemecahan teks menjadi kata-kata individual, penghapusan kata-kata yang tidak signifikan, dan reduksi kata-kata ke bentuk dasarnya untuk meningkatkan kualitas data sebelum analisis lebih lanjut.

Tabel 2. Hasil Preprocessing

Tahapan	hasil
Dataset	@EmergeAmerica @KamalaHarris Terima kasih telah mengundangnya. Kami menantikannya
Cleaning	EmergeAmerica KamalaHarris Terima kasih telah mengundangnya Kami menantikannya
Casefolding	emergeAmerica kamalaHarris terima kasih telah mengundangnya kami menantikannya
Tokenizing	["emergeamerica", "kamalaharris", "terima", "kasih", "telah", "mengundangnya", "kami", "menantikannya"]
Filtering/Stopword Removal	["emergeamerica", "kamalaharris", "terima", "kasih", "mengundangnya", "menantikannya"]
Stemming	["emergeamerica", "kamalaharris", "terim", "kasih", "undang", "nantik"]

Setelah melakukan beberapa tahap preprocessing data, dapat dilihat pada Tabel 2. yaitu perubahan dari setiap tahapan dan jumlah data dalam dataset berkurang secara signifikan. Seperti yang terlihat pada Tabel 2, dari total 3400 data awal, hanya 1424 komentar yang valid, yang terdiri dari 869 komentar berlabel negatif, 270 komentar berlabel positif, dan 285 komentar berlabel netral data yang berhasil lolos proses pembersihan data. Dengan demikian, Data bersih yang diperoleh ini kemudian akan digunakan untuk pengujian selanjutnya

3.3 Pelabelan Data

Proses pelabelan data menggunakan Lexicon Based, Lexicon Based adalah metode analisis sentimen yang memanfaatkan kamus kata-kata sentimen untuk menentukan sentimen teks. Metode ini sederhana, efektif, dan praktis untuk menganalisis data dari media sosial seperti X dan Facebook. yang menghasilkan pelabelan data sebanyak 1.424 komentar, terdiri dari 869 komentar negative, 285 netral dan 270 komentar positif, dapat dilihat pada Gambar 4.

Tabel 3. Hasil Pelabelan Data

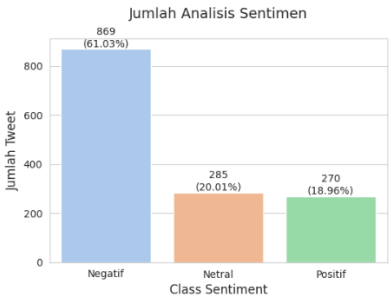
Tweet	Label
IanJaeger29 tidak terjadi	Negatif
oldyzach Apakah efek suara Anda bermasalah?	Netral
EmergeAmerica KamalaHarris Terima kasih telah mengundangnya Kami menantikannya	Positif

3.4 Akurasi

Setelah melalui proses pre-processing data, dari 3400 data awal, tersisa 1424 data yang kemudian dianalisis sentimennya menggunakan metode Lexicon Based, yang memanfaatkan kamus kata-kata sentimen. Hasil analisis

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 873−883
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7398

menunjukkan bahwa dari 1424 komentar, terdapat 869 komentar negatif, 285 komentar netral, dan 270 komentar positif. Dengan demikian, dapat disimpulkan bahwa sentimen negatif mendominasi dengan proporsi sekitar 61% dari total komentar yang dianalisis, menunjukkan adanya ketidakpuasan atau kritik yang signifikan dari pengguna. Dapat dilihat pada Gambar 4



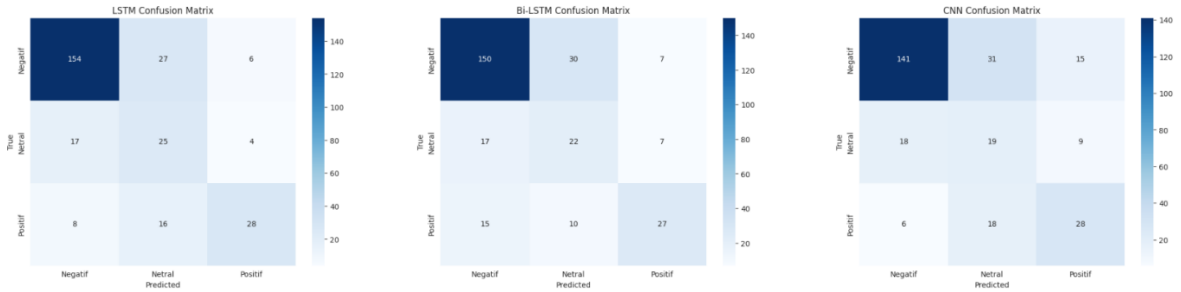
Gambar 4. Jumlah sentimen.

Berdasarkan Gambar 4 hasil analisis sentimen terhadap 1.424 komentar, terlihat bahwa sentimen negatif mendominasi dengan jumlah 869 komentar (sekitar 61%). Hal ini menunjukkan bahwa mayoritas pengguna menyampaikan ketidakpuasan, keluhan, atau kritik terhadap isu atau layanan yang menjadi objek penelitian. Dominasi komentar negatif ini mengindikasikan adanya permasalahan yang cukup signifikan yang perlu mendapatkan perhatian khusus Sementara itu, komentar netral berjumlah 285 (20%) menunjukkan bahwa sebagian pengguna menyampaikan pendapat tanpa kecenderungan emosi yang kuat, baik positif maupun negatif. Ini dapat terjadi karena komentar bersifat informatif atau kurang ekspresif dalam menyampaikan emosi. Adapun komentar positif hanya berjumlah 270 (19%), yang menunjukkan bahwa apresiasi atau persepsi baik dari pengguna masih tergolong rendah dibandingkan dengan komentar bernada negatif. Langkah selanjutnya adalah melatih dan menguji model algoritma. Dalam penelitian ini Hasil evaluasi menunjukkan bahwa model LSTM memiliki akurasi tertinggi sebesar 73%, diikuti oleh BI-LSTM sebesar 70%, dan CNN sebesar 66%. Berdasarkan hasil tersebut, pendekatan LSTM dapat dianggap sebagai pendekatan yang lebih baik dalam mengklasifikasikan komentar masyarakat terkait topik pembangkitan Serigala Direwolf

Tabel 4. Hasil Akurasi

Model	*Class*	*Accuracy*	*Precesion*	*Recalll*	*F1-Score*
LSTM	Negatif	73%	86%	82%	84%
	Netral		37%	54%	44%
	Positif		74%	54%	62%
BI-LSTM	Negatif	70%	82%	80%	81%
	Netral		35%	48%	61%
	Positif		66%	52%	58%
CNN	Negatif	66%	85%	75%	80%
	Netral		28%	41%	33%
	Positif		54%	54%	54%

Berdasarkan Tabel 4 hasil evaluasi model, dapat dilihat bahwa model LSTM memiliki performa yang baik dalam mengklasifikasikan sentimen negatif dengan accuracy 73%, precision 86%, recall 82%, dan F1-score 84%. Sementara itu, model BI-LSTM CNN memiliki performa yang lebih baik dalam mengklasifikasikan sentimen positif dengan accuracy 80%, precision 75%, recall 54%, dan F1-score 61%.Setelah menganalisis performa model menggunakan metrik accuracy, precision, recall, dan F1-score, langkah selanjutnya adalah menganalisis Confusion Matrix untuk memahami lebih detail tentang kinerja model dalam mengklasifikasikan sentimen negatif, netral, dan positif. Dengan Confusion Matrix, kita dapat mengetahui jumlah true positive, false positive, true negative, dan false negative untuk setiap kelas sentimen, sehingga dapat mengidentifikasi area perbaikan model lebih spesifik tabel confusion matrix dari setiap model dapat dilihat pada Gambar 5



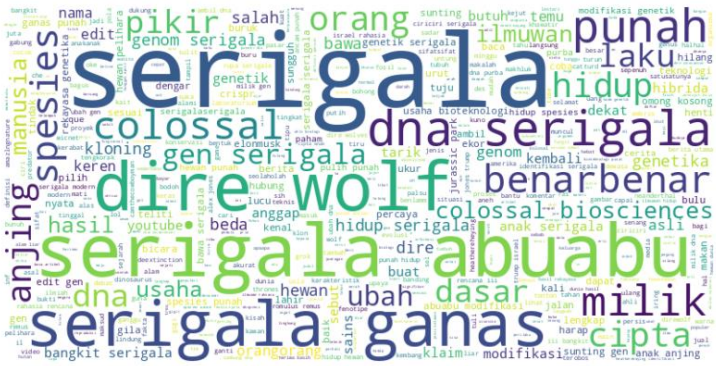
Gambar 5. Confusion Matrix masing-masing Model

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 873−883
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7398

Dari hasil analisis Confusion Matrix, dapat dilihat bahwa model LSTM cenderung memiliki kesulitan dalam mengklasifikasikan sentimen netral dengan benar, seperti yang terlihat pada rendahnya nilai true positive untuk kelas netral. Sementara itu, model BI-LSTM CNN menunjukkan hasil yang lebih baik dalam mengklasifikasikan sentimen positif dan negatif, dengan jumlah true positive yang lebih tinggi dibandingkan model LSTM. Namun, model BI-LSTM CNN juga menunjukkan kelemahan dalam mengklasifikasikan sentimen netral, dengan banyaknya false positive dan false negative pada kelas tersebut

3.5 Visualisasi Wordcloud

Teknik wordcloud digunakan untuk memvisualisasikan kata-kata yang paling sering muncul dalam dataset. Dengan menggunakan pustaka matplotlib Python, wordcloud dapat membantu mengidentifikasi kata-kata kunci yang dominan dalam teks. Hasil visualisasi wordcloud ditampilkan pada Gambar 6, menunjukkan kata-kata yang paling sering digunakan dalam dataset penelitian ini.

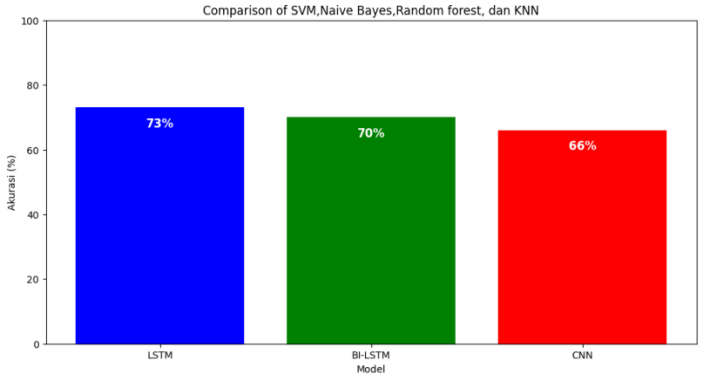


Gambar 6. Worldcloud

Gambar 6 menampilkan wordcloud yang memadukan sentimen negatif, positif, dan netral terkait pembangkitan serigala dire. Kata-kata yang paling menonjol seperti "Serigala", "Dorewolf", "Serigala Ganas", dan istilah lainnya mencerminkan reaksi dan asumsi masyarakat terhadap fenomena ini. Wordcloud ini menggambarkan isu-isu yang paling banyak dibicarakan dan sentimen yang muncul di kalangan masyarakat terkait pembangkitan serigala purba yang telah punah 10.000 tahun yang lalu.

3.5 Visualisasi Perbandingan Model

Visualisasi model dalam penelitian ini memperlihatkan perbandingan akurasi antara tiga model klasifikasi, yaitu LSTM, BI-LSTM, dan CNN. Hasil evaluasi menunjukkan bahwa model LSTM memiliki akurasi tertinggi sebesar 73%, diikuti oleh BI-LSTM sebesar 70%, dan CNN sebesar 66%. Perbandingan akurasi model dapat dilihat pada Gambar 7, yang menggambarkan perbedaan kinerja antara ketiga model tersebu dalam mengklasifikasikan sentimen



Gambar 7. Visualisasi model

Berdasarkan penelitian ini, dapat disimpulkan bahwa model LSTM menunjukkan kinerja terbaik dengan akurasi sebesar 73% dalam menganalisis sentimen terkait pembangkitan serigala dire wolf. Hasil ini menunjukkan bahwa LSTM dapat menangkap pola dan nuansa bahasa yang terkait dengan topik ini dengan lebih baik dibandingkan dengan BI-LSTM dan CNN. Sehingga, LSTM dapat dianggap sebagai pendekatan yang efektif untuk analisis sentimen pada topik pembangkitan serigala dire wolf, dan dapat menjadi referensi untuk penelitian lebih lanjut dalam bidang analisis sentimen.SVM dapat digunakan untuk memahami opini masyarakat terkait Pembangkitan serigala direwolf . Hasil ini juga dapat digunakan sebagai acuan untuk pendekatan pengambilan keputusan strategis dalam konteks yang lebih luas

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 873−883
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7398

4. KESIMPULAN

Penelitian ini berhasil mengumpulkan 3.400 data tweet melalui proses crawling dan menghasilkan 1.424 data bersih setelah proses pengolahan. Hasil analisis sentimen menunjukkan bahwa 869 tweet memiliki sentimen negatif, 285 tweet netral, dan 270 tweet memiliki sentimen positif. Berdasarkan analisis menggunakan model LSTM, BI-LSTM, dan CNN dengan pendekatan lexicon-based dan machine learning, penelitian ini menyimpulkan bahwa model LSTM unggul dalam menganalisis sentimen terkait pembangkitan serigala dire wolf dengan akurasi 73%. Hasil ini menunjukkan bahwa LSTM dapat menangkap nuansa bahasa dan pola sentimen dengan baik, sehingga menjadikannya pendekatan efektif untuk topik ini. Penelitian ini menggunakan pendekatan lexicon-based untuk mengidentifikasi sentimen dalam tweet, yaitu dengan menggunakan kamus kata-kata yang memiliki konotasi positif, negatif, atau netral. Selain itu, penelitian ini juga menggunakan teknik deep learning seperti LSTM, BI-LSTM, dan CNN untuk meningkatkan akurasi analisis sentimen. Penelitian ini memberikan kontribusi pada pengembangan sistem analisis sentimen dan pemahaman opini masyarakat terkait teknologi kloning dan pengeditan gen. Penelitian ini juga menunjukkan bahwa teknik deep learning dapat meningkatkan akurasi analisis sentimen dan dapat menjadi landasan bagi pengembangan sistem analisis sentimen yang lebih canggih dan efektif. Temuan ini dapat menjadi acuan bagi peneliti lain untuk mengembangkan model analisis sentimen yang lebih baik dan memberikan informasi berguna bagi masyarakat dan pengambil kebijakan..

REFERENCES

[1] A. Prasatya and N. Hendrastuty, “Analisis Sentimen: Perbandingan Performa Algoritma Naive Bayes, Support Vector Machine, Random Forest, dan K-Nearest Neighbor Dalam Pemecatan Shin Tae Yong pada Media X”, *bits*, vol. 6, no. 4, pp. 2612-2623, Mar. 2025, doi 10.47065/bits.v6i4.6987

[2] I. Zukhrufillah, “Gejala Media Sosial X Sebagai Media Sosial Alternatif,” *Al-I’lam J. Komun. dan Penyiaran Islam*, vol. 1, no. 2, p. 102, 2018, doi: 10.31764/jail.v1i2.235.

[3] Silmi Nurul UtamI, “Mengenal Direwolf, Serigala Raksasa Zaman Es yang Kembali Dibicarakan,” kOMPAS.COM. Accessed: May 18, 2025. [Online]. Available: https://www.kompas.com/skola/read/2025/04/14/100000669/mengenal-direwolf-serigala-raksasa-zaman-es-yang-kembali-dibicarakan

[4] Woro Anjar Verianty, “Dire Wolf: Mengenal Serigala Purba yang Kembali Hidup,” LIPUTAN 6. Accessed: May 18, 2025. [Online]. Available: https://www.liputan6.com/hot/read/5989002/dire-wolf-mengenal-serigala-purba-yang-kembali-hidup

[5] Christopel paino, “Bangkitnya Serigala Buas ‘Dire Wolf’ dari Kepunahan,” MONGABAY, Accessed: May 18, 2025. [Online]. Available: https://www.mongabay.co.id/2025/04/16/bangkitnya-serigala-buas-dire-wolf-dari-kepunahan/

[6] Redaksi 15, “Serigala Dire Wolf Punah Ribuan Tahun Lalu Muncul KembaliNo Title,” SeputarSumut.com, Accessed: May 18, 2025. [Online]. Available: https://www.seputarsumut.com/serigala-dire-wolf-punah-ribuan-tahun-lalu-muncul-kembali/

[7] P. Arsi and R. Waluyo, “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 1, p. 147, 2021, doi: 10.25126/jtiik.0813944.

[8] *S J Pipin*, “Analisis Sentimen Kebijakan MBKM Berdasarkan Opini Masyarakat di Twitter Menggunakan LSTM,” vol. 23, no. 2, pp. 28–36, 2022, doi 10.55601/jsm.v23i2.900

[9] S. Jurnalis Pipin and H. Kurniawan, “Analisis Sentimen Kebijakan MBKM Berdasarkan Opini Masyarakat di X Menggunakan LSTM,” *J. SIFO Mikroskil*, vol. 23, no. 2, pp. 197–208, 2022, doi: 10.55601/jsm.v23i2.900.

[10] S. M. P. Tyas, R. Sarno, and B. S. Rintyarna, “Analisis Perbandingan Metode Klasifikasi Sentimen Berita Saham: Pendekatan Machine Learning, Deep Learning, Transfer Learning, dan Graf,” *J. Penelit. IPTEKS*, vol. 9, no. 1, pp. 58–64, 2024, doi: 10.32528/penelitianipteks.v9i1.1479.

[11] F. A. Irawan and D. A. Rochmah, “Penerapan Algoritma CNN Untuk Mengetahui Sentimen Masyarakat Terhadap Kebijakan Vaksin Covid-19,” *J. Inform.*, vol. 9, no. 2, pp. 148–158, 2022, doi: 10.31294/inf.v9i2.13257.

[12] M. R. F. Kamarula and N. Rochmawati, “Perbandingan CNN dan Bi-LSTM pada Analisis Sentimen dan Emosi Masyarakat Indonesia Di Media Sosial X Selama Pandemik Covid-19 yang Menggunakan Metode Word2vec,” *J. Informatics Comput. Sci.*, vol. 04, pp. 219–228, 2022, doi: 10.26740/jinacs.v4n02.p219-228.

[13] J. Eka Sembodo, E. Budi Setiawan, and Z. Abdurahman Baizal, “Data Crawling Otomatis pada X,” September 2018, pp. 11–16, 2016, doi: 10.21108/indosc.2016.111.

[14] M. Rafli Ghufron *et al.*, “Analisis Sentimen Pengguna X Terhadap Pemilu 2024 Berbasis Model XLM-T,” J-INTECH (Journal of Information and Technology), vol. 11 no 2, no. 204, pp. 307–315, 2023, doi: https://doi.org/10.32664/j-intech.v11i2.1013.

[15] D B Setyohadi, “Perbaikan Performansi Klasifikasi Dengan Preprocessing Iterative Partitioning Filter Algorithm” Telematika, vol. 14, no. 1, 2017

[16] H. Sulastri and A. I. Gufroni, “Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 299–305, 2017, doi: 10.25077/teknosi.v3i2.2017.299-305.

[17] A. Filcha and M. Hayaty, “Implementasi Algoritma Rabin-Karp untuk Pendeteksi Plagiarisme pada Dokumen Tugas Mahasiswa,” *JUITA J. Inform.*, vol. 7, no. 1, p. 25, 2019, doi: 10.30595/juita.v7i1.4063.

[18] S. Wardani, “Analisis Sentimen Data Presiden Jokowi Dengan Preprocessing Normalisasi Dan Stemming Menggunakan Metode Naive Bayes Dan Svm,” *J. Din. Inform.*, vol. 5, no. November, pp. 1–13, 2015.

[19] A. Salam, J. Zeniarja, and R. S. U. Khasanah, “Analisis Sentimen Data Komentar Sosial Media Facebook Dengan K-Nearest Neighbor (Studi Kasus Pada Akun Jasa Ekspedisi Barang J&T Ekpress Indonesia),” *Pros. SINTAK*, pp. 480–486, 2018.

[20] K. Kevin, M. Enjeli, and A. Wijaya, “Analisis Sentimen Pengunaaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes,” *J. Ilm. Comput. Sci.*, vol. 2, no. 2, pp. 89–98, 2024, doi: 10.58602/jics.v2i2.24.

[21] K. S. Putri, I. R. Setiawan, and A. Pambudi, “Analisis Sentimen Terhadap Brand Skincare Lokal Menggunakan Naïve Bayes Classifier,” *Technol. J. Ilm.*, vol. 14, no. 3, p. 227, 2023, doi: 10.31602/tji.v14i3.11259.

Building of Informatics, Technology and Science (BITS)
Volume 7, No 1, June 2025 Page: 873−883
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v7i1.7398

[22] M. H. Mahendra, D. T. Murdiansyah, and K. M. Lhaksmana, "Analisis Sentimen Tweet COVID-19 menggunakan K-Nearest Neighbors dengan TF-IDF dan Ekstraksi Fitur CountVectorizer," *DIKE J. Ilmu Multidisiplin*, vol. 1, no. 2, pp. 37–43, 2023, doi: 10.69688/dike.v1i2.35.

[23] D. R. Alghifari, M. Edi, and L. Firmansyah, "Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia," *J. Manaj. Inform.*, vol. 12, no. 2, pp. 89–99, 2022, doi: 10.34010/jamika.v12i2.7764.

[24] M. Z. Rahman, Y. A. Sari, and N. Yudistira, "Analisis Sentimen Tweet COVID-19 menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 11, pp. 5120–5127, 2021

[25] E. Y. Hidayat and D. Handayani, "Penerapan 1D-CNN untuk Analisis Sentimen Ulasan Produk Kosmetik Berdasar Female Daily Review," *J. Nas. Teknol. dan Sist. Inf.*, vol. 8, no. 3, pp. 153–163, 2023, doi: 10.25077/teknosi.v8i3.2022.153-163.