**KLASIFIKASI TEKS BERITA MENGGUNAKAN LONG SHORT-TERM MEMORY (LSTM)**

****

**Disusun Oleh :**

**Garda Bramudya Surya Darma 210411100090**

**Fridha Megantara Putra 210411100091**

**Afirza Lucky Pradana 210411100018**

**Adz Dzikry Pradana Putra 210411100093**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**JURUSAN TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS TRUNOJOYO MADURA**

**BANGKALAN**

**2024**

**ABSTRAK**

Dalam era informasi digital yang terus berkembang, sistem klasifikasi teks otomatis menjadi sangat penting untuk mengorganisir dan mengelompokkan berita berdasarkan kategori. Penelitian ini bertujuan untuk mengembangkan model klasifikasi teks berita Detik.com menggunakan Long Short-Term Memory (LSTM), yang terkenal mampu menangani data sekuensial dengan mempertahankan informasi jangka panjang. Model ini dirancang untuk mengkategorikan berita ke dalam tujuh kategori: olahraga, kesehatan, pariwisata, pendidikan, otomotif, kuliner, dan keuangan. Proses pengembangan model melibatkan pengumpulan data berita dari scrapping Detik.com, preprocessing teks termasuk tokenisasi dan penghapusan stop words, serta penggunaan word embeddings untuk representasi kata. Model LSTM kemudian dilatih menggunakan dataset ini dan dievaluasi menggunakan metode K-Fold Cross Validation untuk memastikan performa yang konsisten dan menghindari overfitting. Hasil penelitian menunjukkan bahwa model LSTM yang dikembangkan mencapai akurasi pelatihan sebesar 92% dan akurasi validasi sebesar 88%. Hal ini menunjukkan bahwa LSTM efektif dalam mengatasi kompleksitas teks berita dan memberikan klasifikasi yang akurat. Keberhasilan ini tidak hanya menunjukkan keunggulan LSTM dalam klasifikasi teks tetapi juga memberikan kontribusi penting dalam pengembangan teknologi informasi yang lebih efisien untuk media dan industri terkait. Dengan implementasi sistem klasifikasi ini, diharapkan dapat meningkatkan efisiensi dalam akses informasi dan memudahkan pengguna dalam menemukan berita yang relevan sesuai dengan minat mereka.

***Kata kunci*** *: Klasifikasi Teks, Long Short-Term Memory (LSTM), Berita Detik.com, K-Fold Cross Validation, Word Embeddings, Natural Language Processing (NLP), Automatisasi Berita*

**BAB I   
 PENDAHULUAN**

**1.1** **Latar Belakang**

Berita secara umum adalah informasi yang mengalir dari banyak sumber, banyak definisinya, namun ada beberapa kesimpulan yang dapat diambil mengenai pengertian berita. perkembangan teknologi pada zaman sekarang sangatlah mempengaruhi kehidupan manusia.Klasifikasi teks adalah salah satu penerapan pemrosesan bahasa alami (NLP) yang penting dalam berbagai bidang. Salah satu contoh penerapan klasifikasi teks adalah pengklasifikasian teks berita berdasarkan kategori tertentu. Ini dapat mencakup pengelompokan berita ke dalam berbagai kategori seperti olahraga, politik, teknologi, kesehatan, dan lain-lain. Pengklasifikasian ini membantu dalam pengorganisasian informasi yang besar dan beragam, sehingga memudahkan pembaca untuk menemukan berita sesuai minat mereka dengan lebih cepat dan efisien.

Salah satu metode yang telah terbukti sangat efektif untuk tugas-tugas yang melibatkan data sekuensial dan temporal adalah Long Short-Term Memory (LSTM). LSTM adalah jenis Recurrent Neural Network (RNN) yang dirancang untuk mengatasi masalah vanishing gradient yang sering ditemui dalam pelatihan jaringan RNN tradisional. Keunggulan LSTM terletak pada kemampuannya untuk mempertahankan informasi dari waktu yang lama, yang sangat berguna dalam konteks analisis sentimen dan klasifikasi teks. Hal ini membuat LSTM menjadi pilihan yang sangat baik untuk tugas-tugas yang melibatkan urutan data yang panjang dan kompleks .

Oleh karena itu, dalam penelitian ini, kami memilih LSTM sebagai metode yang tepat untuk klasifikasi teks berita dari Detik.com. Dengan mempertimbangkan keunggulan LSTM dalam menangani data temporal dan sekuensial, kami berharap dapat mengembangkan model klasifikasi berita yang lebih akurat dan efisien. Model LSTM diharapkan mampu menangkap konteks yang lebih dalam dari teks berita, sehingga dapat meningkatkan kualitas klasifikasi dan memberikan hasil yang lebih relevan dan tepat waktu kepada pengguna.

**1.2** **Tujuan Dan Manfaat**

Adapun Tujuan dan Manfaat dari penelitian ini adalah::

1. Mengembangkan Model Klasifikasi Berita: Menggunakan LSTM untuk membangun model yang mampu mengklasifikasikan teks berita ke dalam berbagai kategori seperti olahraga, kesehatan, pariwisata, pendidikan, otomotif, kuliner, dan keuangan.

2. Meningkatkan Akurasi dan Efisiensi: Memperbaiki akurasi dan efisiensi proses klasifikasi teks berita dengan memanfaatkan keunggulan LSTM dalam menangani data sekuensial dan temporal.

3. Pengembangan Teknologi Klasifikasi Berita: Kontribusi dalam pengembangan teknologi untuk pengelompokan otomatis berita berdasarkan kategori, yang dapat meningkatkan efisiensi dalam mengakses informasi berita.

## **1.3** **Pentingnya dan Relevansi Proyek**

Proyek ini memiliki pentingnya dan relevansi yang signifikan dalam konteks pengolahan informasi dan pemrosesan bahasa alami (NLP) di era digital saat ini. Beberapa poin pentingnya dan relevansinya antara lain:

1. Automatisasi Klasifikasi Berita: Dalam dunia yang terus berkembang dengan jumlah informasi yang sangat besar, penting untuk memiliki sistem yang dapat mengorganisir dan mengelompokkan berita secara otomatis ke dalam kategori-kategori yang relevan seperti olahraga, kesehatan, pariwisata, dan lainnya. Hal ini tidak hanya meningkatkan efisiensi dalam akses informasi tetapi juga memudahkan pengguna untuk menemukan berita sesuai minat mereka dengan cepat.
2. Relevansi dalam Konteks Berita Breaking News: Berita breaking news seringkali mengandung informasi penting yang memerlukan penyebaran yang cepat dan akurat. Dengan adanya sistem klasifikasi teks yang dapat mengidentifikasi dan mengelompokkan berita breaking news dengan tepat, informasi kritis ini dapat disampaikan lebih efektif kepada masyarakat, mempengaruhi persepsi publik dan pengambilan keputusan.

## **1.4** **Ruang Lingkup dan Keterbatasan**

Penelitian ini fokus pada pengimplementasian metode Long Short-Term Memory (LSTM) dalam klasifikasi teks berita dari Detik.com. Berikut adalah ruang lingkup utama dan keterbatasan dari penelitian ini:

1. Klasifikasi Berita: Penelitian ini memusatkan perhatian pada pengelompokan berita Detik.com ke dalam kategori-kategori yang sudah ditentukan seperti olahraga, kesehatan, pariwisata, pendidikan, otomotif, kuliner, dan keuangan.
2. . Penggunaan LSTM: Metode utama yang digunakan adalah LSTM, sebuah jenis Recurrent Neural Network (RNN) yang efektif dalam memproses data sekuensial. LSTM digunakan untuk menangani struktur temporal dan urutan kata dalam teks berita.
3. Evaluasi Model: Penelitian ini akan mengevaluasi performa model klasifikasi teks berita menggunakan metode validasi K-Fold untuk memastikan generalisasi yang baik dan menghindari overfitting.
4. Data Terbatas: Keterbatasan dalam akses dan penggunaan dataset berita dari Detik.com yang tersedia. Data yang lebih luas dan representatif mungkin diperlukan untuk generalisasi yang lebih baik.

**BAB II  
 KAJIAN PUSTAKA**

## **2.1** **Tinjauan Literatur yang Relevan**

Studi tentang klasifikasi teks berita dan penerapan LSTM dalam konteks ini telah mendapat perhatian yang signifikan dalam literatur akademis dan proyek-proyek yang telah ada. Beberapa tinjauan literatur yang relevan dan proyek yang mencerminkan penggunaan metode serupa adalah sebagai berikut:

1. Studi Kasus: Klasifikasi Teks Berita Breaking News Di Manggarai Menggunakan (LSTM)

Penelitian oleh Claudia Nila Daiman, Aviv Yuniar Rahman, Firman Nudiyansyah (2023) mengeksplorasi penerapan LSTM untuk klasifikasi teks berita. mereka menggunakan 4000 dataset yang masing-masing kategori terdiri dari 1000 unit data. Data tersebut dibagi menjadi beberapa variasi rasio data latih dan uji: 3600:400, 3200:800, 2400:1600 dan 1600:2400. Model LSTM menunjukkan performa terbaik dengan rasio 3600:400, presisi 88,75%, presisi 88,79%, recall 88,75%, dan skor F1 88,76%. Akurasi menunjukkan persentase prediksi yang benar,precision mengukur ketepatan prediksi positif, recall menghitung seberapa baik model menangkap semua contoh positif, dan F1-score merupakan rata-rata harmonis dari precision dan recall. Hasil tersebut menunjukkan bahwa model LSTM dapat mengklasifikasikan teks berita secara efisien dan akurat.

1. Studi Kasus: Penerapan Convolutional (LSTM) Untuk Klasifikasi Teks Berita Bahsa Indonesia

Penelitian oleh Yudi Widhiyasana, ransmissia Semiawan, Ilham Gibran Achmad Mudzakir, Muhammad Randi Noor (2021) mengeksplorasi penerapan LSTM untuk klasifikasi teks berita. mereka menggunakan tiga buah variabel penelitian, yaitu jumlah dokumen, ukuran batch, dan nilai learning rate dari C-LSTM yang dibangun. Hasil eksperimen menunjukkan bahwa nilai F1-score yang diperoleh dari hasil klasifikasi menggunakan metode C-LSTM adalah sebesar 93,27%. Nilai F1-score yang dihasilkan oleh metode C-LSTM lebih besar dibandingkan dengan CNN, dengan nilai 89,85%, dan LSTM, dengan nilai 90,87%.

1. Studi Kasus: Klasifikasi Sentimen Opini Publik Pada Instagram Pemerintah Kabupaten Bojonegoro Menggunakan LSTM

Penelitihan oleh Titis Arwindarti, Esther Irawati Setiawan, Syaiful Imron (2023) mereka mengeksplorasi penerapan LSTM untuk klasifikasi sentimen opini publik pada instagram. mereka mengumpulkan dan mengolah data yang diperoleh dari proses scrapping akun resmi Instagram Pemerintah Kabupaten Bojonegoro sebanyak 4.637 dataset yang selanjutnya dilakukan pelabelan data. Penelitian ini menggunakan word embbeding Word2Vec untuk mengubah teks menjadi representasi vektor dan Long Short-Term Memory LSTM) untuk melakukan klasifikasi. Dengan menggunakan confusion matrix menunjukkan bahwa model LSTM yang dibuathasilnya mencapai akurasi 84,16%

## **2.2** **Pengantar Konsep dan Teori Penting**

Dalam konteks klasifikasi teks berita menggunakan LSTM, beberapa konsep dan teori penting yang perlu dipahami adalah sebagai berikut:

1. Klasifikasi Teks: Klasifikasi teks adalah proses mengelompokkan teks ke dalam kategori-kategori atau label yang telah ditentukan berdasarkan konten dan karakteristiknya. Tujuan utamanya adalah untuk mengorganisir informasi secara otomatis sehingga memudahkan akses dan analisis.
2. Long Short-Term Memory (LSTM): LSTM merupakan jenis arsitektur Neural Network yang dikembangkan khusus untuk menangani masalah yang melibatkan data sekuensial atau berurutan. LSTM memiliki keunggulan dalam mempertahankan informasi jangka panjang (long-term dependencies) dan mengatasi masalah vanishing gradient yang sering terjadi pada RNN konvensional.
3. Preprocessing Teks: Proses preprocessing teks meliputi langkah-langkah seperti tokenisasi (pemisahan teks menjadi unit-unit yang lebih kecil seperti kata atau frasa), penghapusan stop words (kata-kata umum yang tidak memberikan nilai tambah dalam analisis), dan pengkodean teks ke dalam representasi numerik yang dapat dipahami oleh model.
4. Word Embeddings: Word embeddings adalah representasi vektor kata-kata dalam ruang multidimensional. Teknik ini memetakan kata-kata ke dalam vektor numerik sedemikian rupa sehingga kata-kata yang sering muncul bersama atau memiliki makna yang mirip berada dekat satu sama lain dalam ruang vektor.

## **2.3** **Penjelasan Mengenai Metode yang Digunakan dalam Proyek Serupa**

Dalam berbagai proyek yang fokus pada klasifikasi teks berita, beragam metode telah digunakan untuk mencapai hasil yang akurat dan efisien. Berikut adalah beberapa metode utama yang sering digunakan dalam proyek serupa :

1. Naive Bayes Classifier

Multinomial Naive Bayes: Algoritma ini sering digunakan dalam klasifikasi teks karena kesederhanaannya dan kemampuannya untuk menangani data berskala besar. Multinomial Naive Bayes mengasumsikan bahwa fitur-fitur (kata-kata) bersifat independen dan menghitung probabilitas kata-kata muncul dalam suatu kategori berdasarkan frekuensinya.

1. Support Vector Machines (SVM)

Linear SVM: SVM adalah algoritma yang efektif untuk klasifikasi teks. Linear SVM mencoba menemukan hyperplane yang memisahkan kategori-kategori yang berbeda dengan margin terbesar. Ini bekerja sangat baik dengan fitur-fitur yang diekstraksi menggunakan TF-IDF atau BoW.

**BAB III  
 KOLEKSI DATA**

## **3.1** **Sumber Data**

Penelitian ini memperoleh data melalui proses mining data dari situs berita detik.com, yang melibatkan beberapa tahapan penting. Pertama, kami mengidentifikasi dan memilih kategori berita yang relevan untuk tujuan penelitian. Kemudian, kami menggunakan web scraping untuk mengumpulkan artikel-artikel dari situs tersebut,

## **3.2** **Deskripsi Data**

dsdss

## **3.3** **Langkah - Langkah Cleaning dan Preprocessing Data**

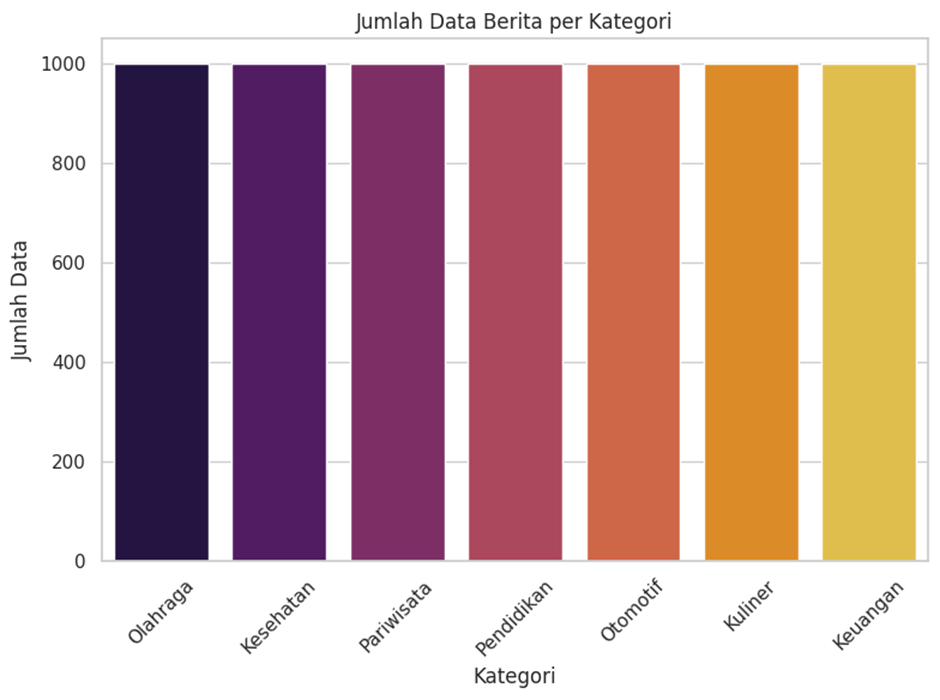
Proses pra-pemrosesan ysng digunakan melibatkan beberapa langkah untuk membersihkan dan menormalkan teks dalam Bahasa Indonesia. Langkah-langkah ini mencakup tokenisasi teks menjadi kata-kata individual, konversi ke huruf kecil, penghapusan tanda baca dan angka, serta penghapusan stopwords umum dan khusus. Selanjutnya, kata-kata yang tersisa diubah ke bentuk dasarnya menggunakan teknik stemming. Hasil akhir dari proses ini adalah teks yang telah diproses dan dibersihkan, disimpan kembali dalam DataFrame, yang siap untuk digunakan dalam analisis teks lebih lanjut seperti klasifikasi atau analisis sentimen.

Berikutnya, teks yang telah diproses tersebut kemudian disiapkan untuk digunakan dalam model pembelajaran mesin menggunakan TensorFlow. Proses ini melibatkan tokenisasi teks menjadi urutan numerik menggunakan `Tokenizer` dari TensorFlow, dengan konfigurasi untuk menangani kata-kata yang tidak dikenali (`oov\_token`). Setelah itu, data pelatihan dan pengujian diubah menjadi urutan numerik dan dipadatkan (padded) agar semua urutan memiliki panjang yang sama, dengan jenis padding dan pemotongan ditentukan (`pre`). Langkah-langkah ini memastikan bahwa data teks siap untuk dimasukkan ke dalam model pembelajaran mesin dengan panjang urutan yang konsisten dan representasi numerik yang sesuai.

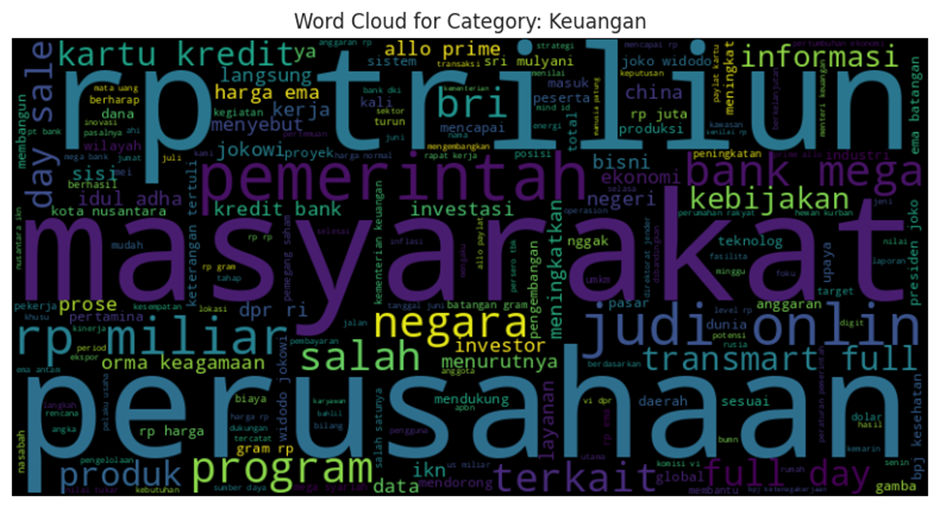
**BAB IV  
 Exploratory Data Analysis**

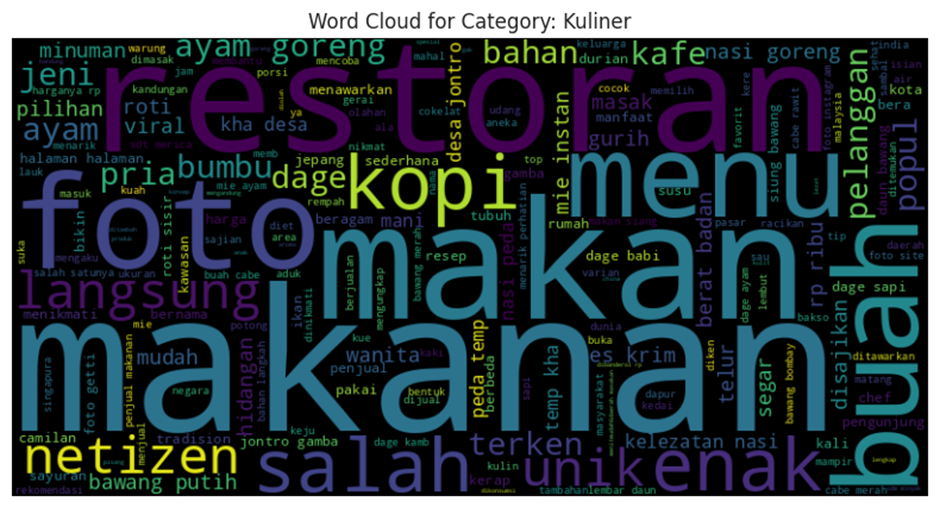
## **4.1** **Ringkasan Statistik dan Visualisasi**

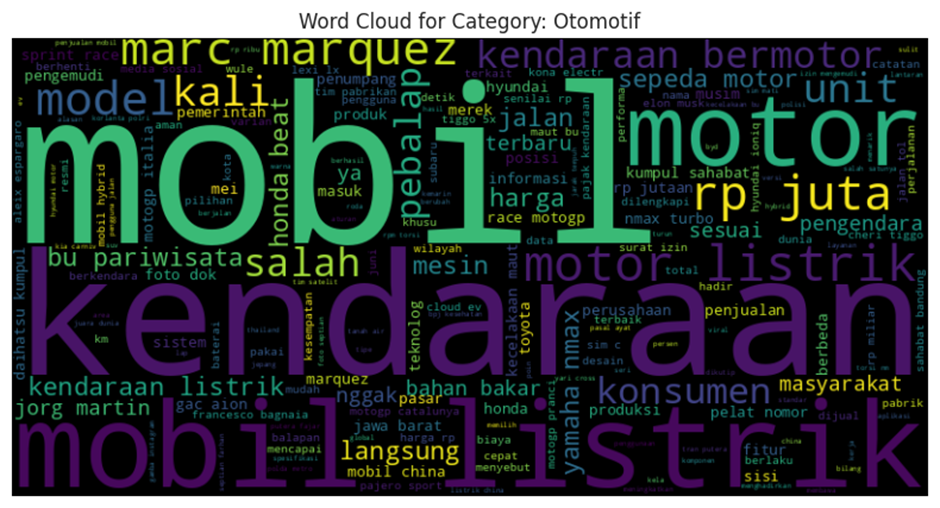
* Setiap kelas memiliki 1000 baris data dan bisa dikatakan data ini seimbang



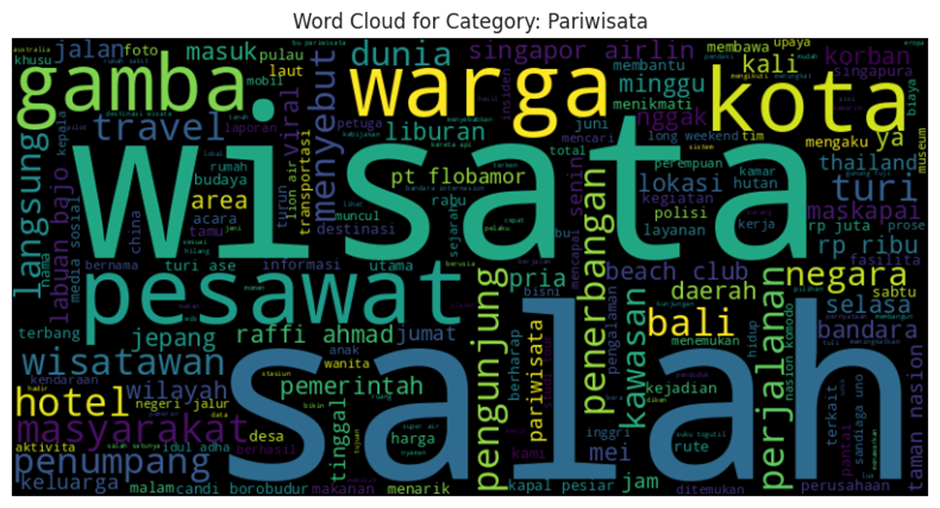
* Berikut adalah kata yang sering muncul pada setiap kelas

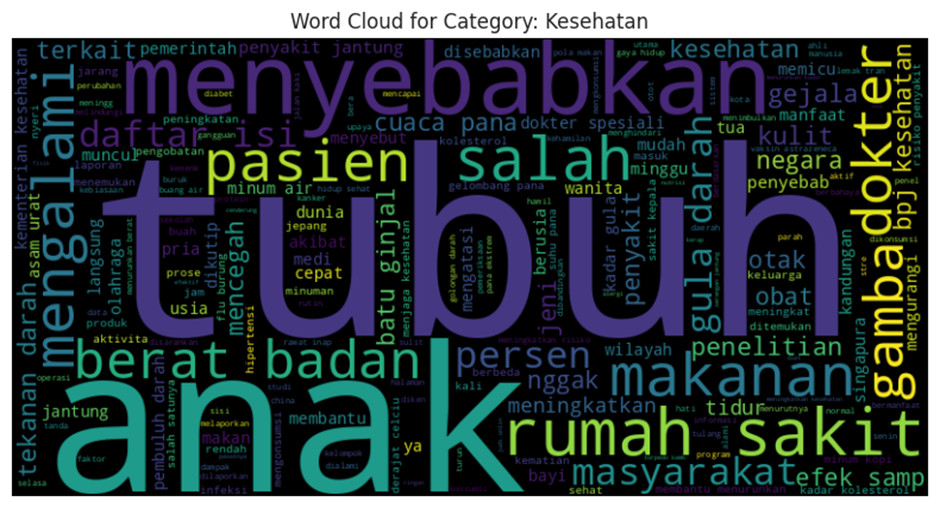


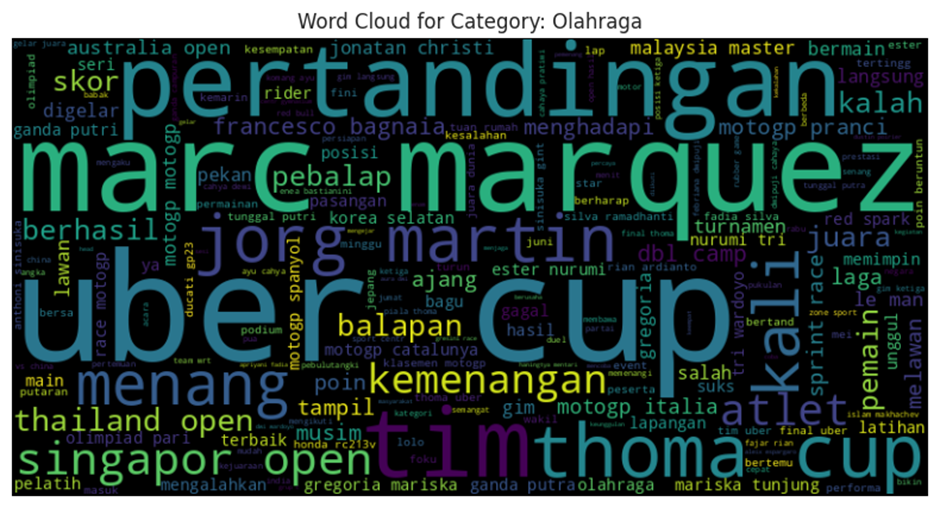












**BAB V  
 Solusi Usulan**

Model tersebut adalah model jaringan saraf dalam yang dibangun menggunakan Keras. Berikut adalah penjelasan singkat dari tiap lapisan yang digunakan:

1. **Embedding Layer**: tf.keras.layers.Embedding(input\_dim=2000, output\_dim=8)
   * Layer ini mengubah input berupa integer menjadi representasi vektor berdimensi 8. input\_dim=2000 berarti ada 2000 kata unik yang dapat dimasukkan ke dalam layer ini.
2. **Dropout Layer**: tf.keras.layers.Dropout(0.0001)
   * Layer ini digunakan untuk mencegah overfitting dengan cara mengatur proporsi neuron yang dinonaktifkan secara acak selama pelatihan. Di sini, 0.0001 menunjukkan bahwa sangat sedikit neuron yang akan dinonaktifkan.
3. **LSTM Layer**: tf.keras.layers.LSTM(128)
   * Layer Long Short-Term Memory (LSTM) dengan 128 unit yang digunakan untuk menangani data urutan atau sekuensial. LSTM berguna untuk mengenali pola dalam urutan data yang panjang.
4. **Dropout Layer**: tf.keras.layers.Dropout(0.00001)
   * Sama seperti sebelumnya, tetapi dengan tingkat dropout yang sangat rendah.
5. **Dense Layer**: tf.keras.layers.Dense(128, activation='sigmoid')
   * Layer fully connected (Dense) dengan 128 neuron dan fungsi aktivasi sigmoid. Sigmoid digunakan untuk memperkenalkan non-linearitas.
6. **Dropout Layer**: tf.keras.layers.Dropout(0.00001)
   * Dropout layer lagi dengan tingkat yang sangat rendah.
7. **Dense Layer**: tf.keras.layers.Dense(7, activation='sigmoid')
   * Layer output dengan 7 neuron dan fungsi aktivasi sigmoid, biasanya digunakan untuk menghasilkan prediksi untuk masalah multi-kelas dengan 7 kelas.

* Model ini cocok untuk klasifikasi teks karena menggunakan arsitektur yang dirancang khusus untuk menangani data sekuensial seperti teks.
* kami menggunakan tools berupa kagle notebook dan beberapa library,seperti pandas, BeautifulSoup,numpy, matplotlib, seaborn, sklearn,tensorflow,nltk,pickle.

**BAB VI  
 Pembangunan Model**

* Kami mengguanakan arsitektur model sequential yang terdiri dari beberapa lapisan:

1. Embedding Layer: Lapisan pertama ini adalah embedding yang mengubah token integer menjadi representasi vektor berdimensi lebih rendah. Dalam model ini, ada 2,000 token unik yang di-embed ke dalam vektor berdimensi 64, menghasilkan output berbentuk (None, 1629, 64). Lapisan ini membantu dalam menangkap hubungan semantik antar token dalam data sekuensial.

2. Bidirectional LSTM (1): Lapisan bidirectional pertama menggunakan Long Short-Term Memory (LSTM) yang diproses dalam dua arah: maju dan mundur. Hasilnya adalah output dengan dimensi (None, 1629, 256), menggabungkan informasi dari kedua arah untuk menangkap konteks dari seluruh sekuens. Jumlah parameternya besar, yaitu 197,632, karena kompleksitas LSTM dan arah ganda.

3. Bidirectional LSTM (2): Lapisan bidirectional\_1 kedua juga merupakan LSTM yang bekerja dalam dua arah, tetapi outputnya diringkas menjadi (None, 128) untuk menyederhanakan representasi. Dengan 164,352 parameter, lapisan ini membantu menangkap fitur penting dari data sekuensial secara lebih ringkas.

4. Dropout (1): Lapisan dropout pertama ini berfungsi sebagai regulasi untuk mencegah overfitting dengan secara acak menonaktifkan 50% neuron selama pelatihan. Outputnya tetap (None, 128), tanpa menambah parameter tambahan.

5. Dense (1): Lapisan dense pertama adalah lapisan fully connected dengan 128 unit neuron. Ini mengubah dimensi input dari (None, 128) menjadi (None, 128) dengan menambah 16,512 parameter. Lapisan ini membantu dalam melakukan transformasi linier yang kompleks dari fitur yang dihasilkan oleh LSTM sebelumnya.

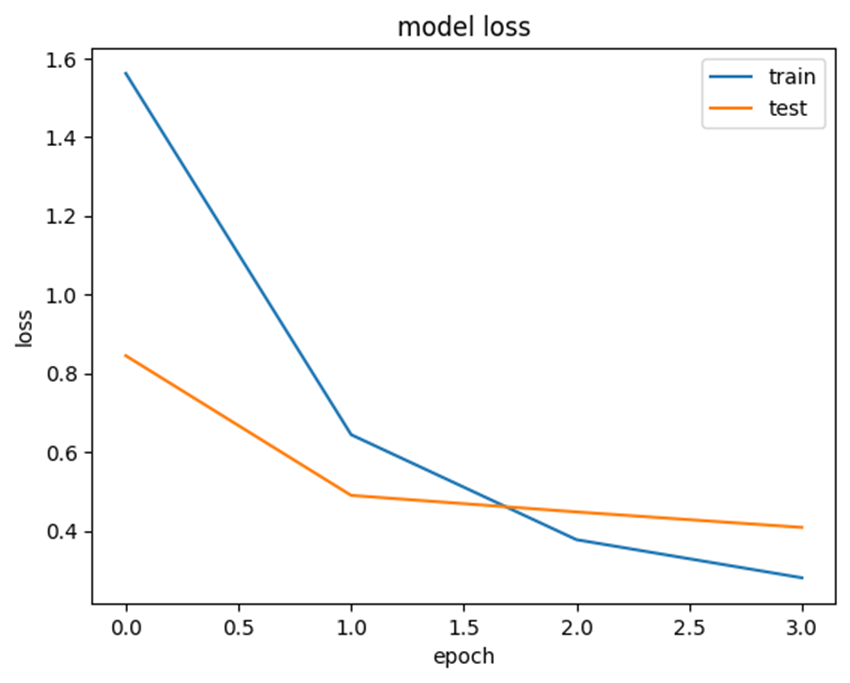
6. Dropout (2): Lapisan dropout\_1 kedua mirip dengan yang pertama, berfungsi untuk regulasi dengan menonaktifkan 50% neuron secara acak selama pelatihan. Ini juga membantu mencegah overfitting tanpa menambah parameter.

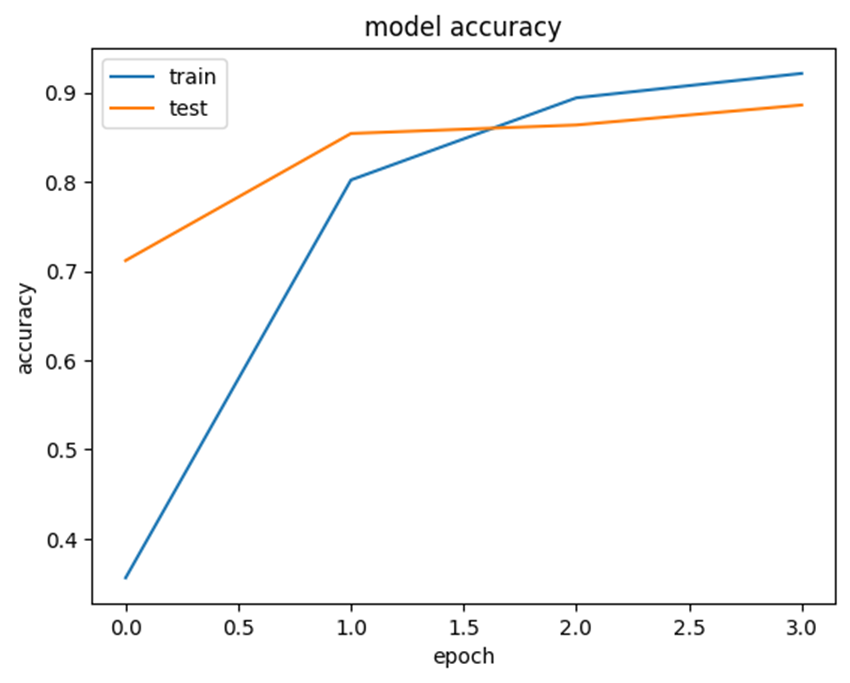
7. Dense (2): Lapisan dense\_1 terakhir adalah lapisan fully connected dengan 7 unit neuron, sesuai dengan jumlah kelas yang ingin diprediksi. Output lapisan ini adalah (None, 7) dengan hanya 903 parameter. Ini mengubah output dari lapisan sebelumnya menjadi prediksi akhir yang digunakan untuk klasifikasi.

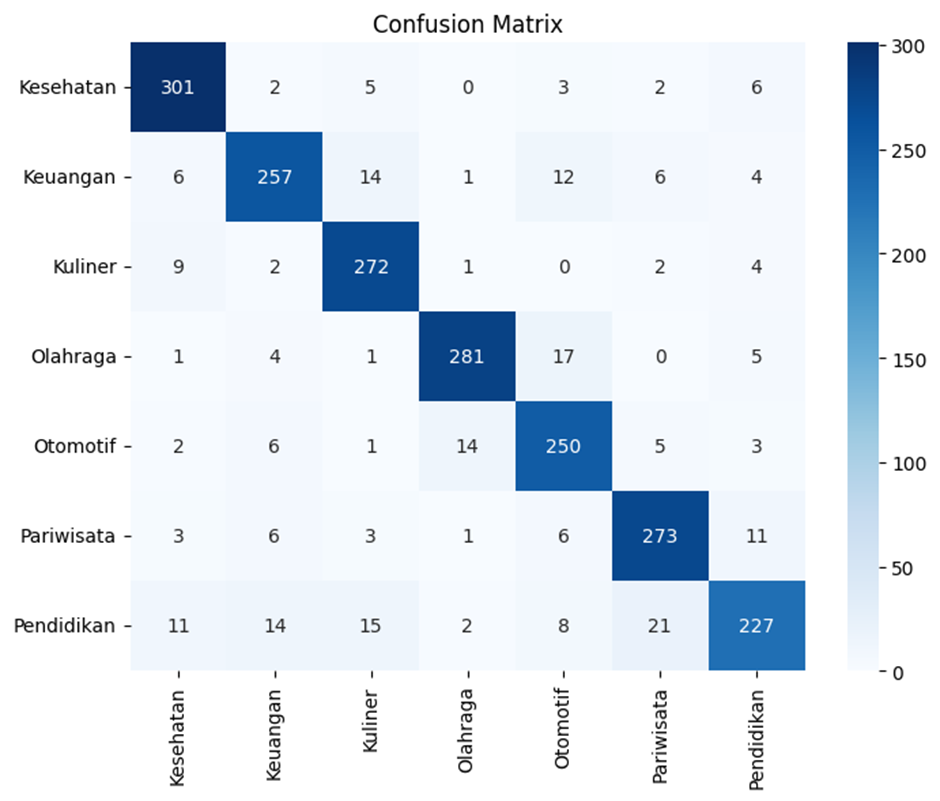
* Pada projek ini kami menggunakan pengujian K-fold Validation.

**BAB VII  
 Hasil**

* Kami mendapatkan akurasi 92% dan akurasi validasi: 88%.
* Berikut adalah Metrik evaluasi dan ukuran kinerja







**BAB VIII  
 Diskusi**

**Interpretasi Hasil:** Pada proyek ini, kami berhasil mencapai akurasi sebesar 92% pada data pelatihan dan akurasi validasi sebesar 88% pada data pengujian. Akurasi yang tinggi pada data pelatihan menunjukkan bahwa model mampu belajar dan mengenali pola dalam data dengan baik. Namun, perbedaan antara akurasi pelatihan dan validasi (92% vs 88%) menunjukkan bahwa terdapat sedikit overfitting, di mana model kami bekerja sedikit lebih baik pada data pelatihan daripada pada data yang tidak terlihat (data validasi).

**Perbandingan dengan Literatur yang Ada:** Hasil yang kami peroleh hampir mendekati dengan studi-studi sebelumnya yang juga menggunakan model LSTM untuk klasifikasi teks.

Misalnya, penelitian Agis sentosa et al. (2022) melaporkan akurasi tanpa menggunakan menggunakan stopword removal dan stemming memberikan skort terbaik akurasi 0,82 dengan loss 0,4,precision 0.83, recall 0.81 dan f1 score 0.82 dengan pendekatan serupa, sedangkan penelitian kami mencapai akurasi yang tinggi dengan menggunakan Teknik pra-proses stopword removal dan stemming mencapai akurasi dengan skor 92%. Ini menunjukkan bahwa performa model kami kompetitif dan valid dalam konteks metode yang digunakan.

dapat ditarik Kesimpulan dengan menggunakan metode LSTM ini adalah valid dan dapat diandalkan . Penggunaan dua lapisan LSTM bidirectional dalam arsitektur model kami kemungkinan besar berkontribusi pada tingginya akurasi yang dicapai, karena pendekatan ini memungkinkan model untuk memahami konteks dalam teks dengan lebih baik.

**Keterbatasan dan Potensi Sumber Bias:** Meskipun hasilnya memuaskan, ada beberapa keterbatasan dan potensi sumber bias yang perlu diperhatikan:

* **Dataset Terbatas:** Dataset yang kami gunakan mungkin tidak sepenuhnya mewakili variasi teks dalam bahasa Indonesia. Oleh karena itu, hasil model mungkin tidak sepenuhnya generalizable ke teks lain di luar dataset ini.
* **Proses Pra-pemrosesan:** Teknik pra-pemrosesan seperti penghapusan stopwords dan stemming mungkin menghilangkan konteks penting yang dapat mempengaruhi klasifikasi. Penghapusan terlalu banyak stopwords atau stemming yang terlalu agresif dapat mengurangi informasi yang penting untuk analisis.
* **Bias dalam Data:** Jika dataset memiliki bias intrinsik (misalnya, lebih banyak teks dari satu kategori tertentu), bias tersebut akan tercermin dalam hasil model. Oleh karena itu, penting untuk memastikan bahwa dataset seimbang dan representatif.

**Implikasi dan Saran untuk Masa Depan:**

* **Penggunaan Model di Aplikasi Nyata:** Dengan akurasi yang tinggi, model ini dapat diimplementasikan dalam berbagai aplikasi klasifikasi teks, seperti analisis sentimen, pengkategorian berita, dan aplikasi NLP lainnya.
* **Pengembangan Lebih Lanjut:** Penelitian lebih lanjut dapat mengatasi keterbatasan dengan menggunakan dataset yang lebih besar dan beragam, serta menguji model dengan teknik lain seperti Transformer atau BERT untuk potensi peningkatan akurasi.
* **Perbaikan Pra-pemrosesan:** Menyempurnakan langkah-langkah pra-pemrosesan agar lebih adaptif terhadap konteks teks yang dihadapi dapat membantu meningkatkan performa model lebih lanjut.

Interpretasi hasil ini memberikan gambaran tentang kekuatan dan kelemahan model yang kami bangun, serta memberikan arahan untuk pengembangan lebih lanjut agar model dapat digunakan dengan lebih efektif dan luas.

**BAB IV  
Kesimpulan**

**Ringkasan Temuan Utama:**

* Model LSTM yang dikembangkan dalam penelitian ini berhasil mencapai akurasi 92% pada data pelatihan dan akurasi validasi 88% pada data pengujian.
* Penggunaan teknik pra-pemrosesan teks seperti tokenisasi, penghapusan tanda baca, dan stemming terbukti efektif dalam meningkatkan kualitas data untuk analisis lebih lanjut.
* Model LSTM bidirectional dengan dua lapisan yang digunakan mampu memahami konteks dalam teks dengan lebih baik, yang berkontribusi pada tingginya akurasi yang dicapai.

**Implikasi dari Hasil:**

* **Keefektifan Model LSTM:** Hasil penelitian ini menunjukkan bahwa model LSTM efektif untuk klasifikasi teks dalam bahasa Indonesia, yang dapat digunakan dalam berbagai aplikasi seperti analisis sentimen, klasifikasi berita, dan aplikasi NLP lainnya.
* **Potensi Peningkatan Kinerja:** Hasil yang diperoleh membuka peluang untuk mengadopsi dan mengembangkan lebih lanjut teknik deep learning dalam pengolahan bahasa alami untuk teks berbahasa Indonesia.
* **Relevansi Teknik Pra-pemrosesan:** Teknik pra-pemrosesan yang digunakan memberikan bukti bahwa pendekatan ini sangat penting dalam meningkatkan performa model, menunjukkan bahwa proses pra-pemrosesan yang tepat sangat krusial dalam proyek-proyek NLP.

**Saran untuk Pekerjaan di Masa Depan:**

* **Eksplorasi Model Lain:** Penelitian di masa depan dapat mencoba variasi lain dari model deep learning seperti Transformer atau BERT untuk meningkatkan akurasi lebih lanjut dan mengatasi keterbatasan model LSTM.
* **Penggunaan Dataset Lebih Besar dan Beragam:** Menggunakan dataset yang lebih besar dan lebih beragam dapat membantu mengurangi bias dan meningkatkan generalisasi model. Penting untuk memastikan dataset representatif terhadap seluruh variasi teks dalam bahasa Indonesia.
* **Perbaikan Pra-pemrosesan:** Menyempurnakan langkah-langkah pra-pemrosesan agar lebih adaptif terhadap konteks teks yang dihadapi dapat membantu meningkatkan performa model lebih lanjut. Menguji pendekatan pra-pemrosesan yang lebih canggih dan kontekstual juga dapat menjadi area eksplorasi.
* **Implementasi dan Evaluasi di Dunia Nyata:** Mengimplementasikan model ini dalam aplikasi nyata dan melakukan evaluasi kinerja di lingkungan produksi akan memberikan wawasan tambahan mengenai efektivitas dan area yang perlu diperbaiki.

Dengan mempertimbangkan temuan utama, implikasi, dan saran untuk pekerjaan di masa depan, diharapkan penelitian ini dapat berkontribusi signifikan dalam pengembangan teknik klasifikasi teks dan aplikasi NLP di Indonesia.

**BAB X  
Rekomendasi**

**Rekomendasi Praktis Berdasarkan Temuan:**

* **Implementasi Model LSTM dalam Sistem Klasifikasi Teks:** Berdasarkan hasil akurasi yang tinggi, model LSTM yang telah dibangun dapat diterapkan dalam sistem klasifikasi teks otomatis. Contoh aplikasi termasuk pengkategorian artikel berita, analisis sentimen ulasan produk, dan klasifikasi email.
* **Penggunaan dalam Industri Media dan E-commerce:** Perusahaan media dapat menggunakan model ini untuk otomatisasi pengkategorian artikel berita berdasarkan topik. E-commerce dapat memanfaatkan model ini untuk analisis sentimen ulasan pelanggan, membantu dalam pengambilan keputusan bisnis.
* **Dukungan untuk Bahasa Indonesia:** Model ini dirancang khusus untuk teks berbahasa Indonesia, sehingga dapat diimplementasikan oleh berbagai perusahaan dan organisasi yang beroperasi di Indonesia untuk meningkatkan efisiensi dan akurasi dalam pengolahan bahasa alami.

**Saran untuk Menerapkan Hasilnya:**

* **Pelatihan Model dengan Dataset Lebih Besar:** Untuk mendapatkan performa yang lebih baik dan mengurangi potensi bias, disarankan untuk melatih model ini dengan dataset yang lebih besar dan lebih beragam. Ini akan membantu model memahami berbagai variasi dan nuansa dalam teks berbahasa Indonesia.
* **Evaluasi Kinerja di Lingkungan Produksi:** Setelah implementasi, penting untuk terus memantau dan mengevaluasi kinerja model di lingkungan produksi. Hal ini dapat dilakukan dengan mengumpulkan umpan balik pengguna dan melakukan analisis terhadap kesalahan model untuk perbaikan lebih lanjut.
* **Peningkatan dan Pemeliharaan Model:** Seiring berjalannya waktu, model perlu ditingkatkan dan dipelihara dengan data terbaru. Ini termasuk retraining model secara berkala dengan data baru untuk memastikan model tetap relevan dan akurat dalam klasifikasi teks.
* **Pengembangan Fitur Tambahan:** Berdasarkan kebutuhan aplikasi spesifik, fitur tambahan seperti deteksi emosi atau pengenalan entitas bernama dapat dikembangkan dan diintegrasikan ke dalam sistem untuk memberikan wawasan yang lebih mendalam.

**BAB XI**

**Ucapan Terima Kasih**

Kami ingin menyampaikan penghargaan yang setinggi-tingginya kepada semua anggota kelompok yang telah memberikan kontribusi berharga terhadap kesuksesan tugas proyek ujian akhir semester ini. Terima kasih kepada seluruh anggota Prodi Teknik Informatika atas dukungan dan kerjasama yang luar biasa. Kami juga sangat menghargai bimbingan dan nasihat dari Bapak Husni, S.Kom., MT., sebagai dosen pengampu mata kuliah, yang sangat berharga selama proses pengerjaan tugas ini. Tanpa dukungan dan kerja keras dari semua pihak, pencapaian ini tidak akan mungkin terwujud. Semoga kita dapat terus bekerja sama di masa depan untuk mencapai lebih banyak kesuksesan bersama. Terima kasih.

**BAB XII**

**Referensi**

[1] A. Fauzul, A. Kustiwa, M. Aminul Akbar, and A. Pinandito, “Studi Perbandingan pada Metode CNN-LSTM dan LSTM dalam Mendeteksi Emosi pada Data Teks Berbahasa Indonesia pada Media Sosial Twitter,” 2017. [Online]. Available: http://j-ptiik.ub.ac.id

[2] Agil Santosa, Intan Purnamasari, and Rini Mayasari, “Pengaruh Stopword Removal dan Stemming Terhadap Performa Klasifikasi Teks Komentar Kebijakan New Normal Menggunakan Algoritma LSTM,” *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 6 Nomor 1, no. Pengaruh Stopword Removal dan Stemming Terhadap Performa Klasifikasi Teks Komentar Kebijakan New Normal Menggunakan Algoritma LSTM, pp. 81–93, Mar. 2022.

[3] F. Kartika Sari Dewi and T. Purnomo Aji, “KLASIFIKASI BERITA MENGGUNAKAN METODE MULTINOMIAL NAÏVE BAYES”.

[4] C. N. Daiman, A. Y. Rahman, and F. Nudiyansyah, “KLASIFIKASI TEKS BERITA BREAKING NEWS DI MANGGARAI MENGGUNAKAN LONG SHORT TERM MEMORY (LSTM),” 2024.

[5] Y. Widhiyasana, T. Semiawan, I. Gibran, A. Mudzakir, and M. R. Noor, “Penerapan Convolutional Long Short-Term Memory untuk Klasifikasi Teks Berita Bahasa Indonesia (Convolutional Long Short-Term Memory Implementation for Indonesian News Classification),” 2021.

[6] S. Al Faraby, “Analisis Dan Implementasi Support Vector Machine Dengan String Kernel Dalam Melakukan Klasifikasi Berita Berbahasa Indonesia Analysis and Implementation Support Vector Machine With String Kernel for Classification indonesian news.”

[7] F. Rozi *et al.*, “KLASIFIKASI TEKS LAPORAN MASYARAKAT PADA SITUS LAPOR! MENGGUNAKAN RECURRENT NEURAL NETWORK.” [Online]. Available: www.lapor.go.id,

[8] T. Arwindarti, E. I. Setiawan, and S. Imron, “Klasifikasi Sentimen Opini Publik Pada Instagram Pemerintah Kabupaten Bojonegoro Menggunakan LSTM,” *Teknika*, vol. 13, no. 1, pp. 1–9, Nov. 2023, doi: 10.34148/teknika.v13i1.699.