

# Pengembangan Sistem *Deployment* Chatbot Dengan Teknologi LSTM untuk *Customer Service* Industri Batik di Karanganyar

Silvester Aditya Damaswara<sup>1</sup>, Abu Salam<sup>2</sup>

Jurusan Teknik Informatika, Universitas Dian Nuswantoro  
Jalan Imam Bonjol No. 207 & Jalan Nakula I No. 5 - 11, Pendrikan Kidul, Kec. Semarang Tengah, Kota Semarang,  
Jawa Tengah 50131, Indonesia

<sup>1</sup>111202012868@mhs.dinus.ac.id

<sup>2</sup>abu.salam@dsn.dinus.ac.id

---

## Abstrak

Peningkatan pemahaman masyarakat tentang batik sebagai warisan budaya menjadi salah satu upaya penting dalam melestarikan nilai-nilai budaya lokal. Penelitian ini bertujuan mengembangkan sebuah chatbot berbasis Long Short-Term Memory (LSTM) yang berfungsi sebagai sarana penyebaran informasi mengenai batik, khususnya di wilayah Karanganyar. Dataset chatbot yang digunakan disusun dalam format JSON dengan 720 baris data, mencakup 59 kategori atau kelas, dan diproses melalui tahapan preprocessing seperti case folding, lemmatisasi, tokenisasi, dan padding. Model dikembangkan menggunakan arsitektur Bidirectional LSTM dan GRU, diikuti dengan layer Dense dan regularisasi untuk meningkatkan performa. Hasil evaluasi menunjukkan model mencapai akurasi 96% pada data pelatihan dan 85% pada data validasi, dengan nilai F1 Score sebesar 84%. Meskipun terdapat indikasi overfitting, perbedaannya tidak signifikan sehingga model tetap mampu memberikan respons yang memadai. Model ini kemudian diintegrasikan ke dalam aplikasi web berbasis Flask dengan antarmuka yang interaktif untuk mempermudah akses masyarakat. Penelitian ini menunjukkan bahwa penggunaan LSTM efektif dalam mengatasi tantangan vanishing gradient dan menghasilkan chatbot yang mampu menjawab pertanyaan dengan akurasi tinggi, memberikan kontribusi positif terhadap pelestarian dan penyebaran informasi mengenai batik.

**Kata kunci:** Batik; chatbot; deep learning; flask; LSTM

## Abstract

*Enhancing public knowledge of batik as a cultural heritage is a crucial effort in preserving local cultural values. This study aims to develop a chatbot based on Long Short-Term Memory (LSTM) that serves as a medium for disseminating information about batik, particularly in the Karanganyar region. The chatbot dataset was structured in JSON format with 720 data rows, encompassing 59 categories or classes, and processed through preprocessing stages such as case folding, lemmatization, tokenization, and padding. The model was developed using a Bidirectional LSTM and GRU architecture, followed by Dense layers and regularization to enhance performance. Evaluation results indicate that the model achieved 96% accuracy on training data and 85% accuracy on validation data, with an F1 Score of 84%. Although there were indications of overfitting, the difference was not significant, allowing the model to still provide adequate responses. The model was then integrated into a Flask-based web application with an interactive interface to facilitate public access. This study demonstrates that the use of LSTM is effective in addressing vanishing gradient challenges and producing a chatbot capable of answering questions with high accuracy, contributing positively to the preservation and dissemination of information about batik.*

**Keywords:** Batik; chatbot; deep learning; flask; LSTM

---

## I. PENDAHULUAN

Batik merupakan warisan budaya Indonesia yang memiliki nilai seni, makna simbolis, dan identitas yang kuat bagi bangsa[1]. Batik telah berkembang

menjadi bagian penting dalam kehidupan masyarakat Indonesia, baik sebagai busana sehari-hari, acara adat, maupun sebagai elemen budaya yang diakui dunia[2]. Sebagai salah satu hasil karya seni yang kaya akan motif dan warna, batik

mencerminkan nilai-nilai tradisional, filosofi, dan sejarah dari berbagai daerah di Indonesia. Motif batik seringkali menggambarkan alam, kepercayaan, dan kehidupan sehari-hari, sehingga setiap pola memiliki makna yang mendalam dan unik[3]. Batik tidak hanya berfungsi sebagai pakaian, tetapi juga menjadi sarana ekspresi sosial dan budaya. Banyak penelitian menunjukkan bahwa batik dapat memperkuat identitas lokal serta meningkatkan kebanggaan nasional. Dengan pengakuan UNESCO sebagai Warisan Budaya Takbenda, batik telah memperkuat posisi Indonesia di kancah internasional[4]. Hal ini turut mendorong peningkatan industri batik yang berkelanjutan di Indonesia, baik dari segi produksi, pemasaran, maupun pemahaman masyarakat tentang nilai budayanya. Namun, perkembangan batik juga menghadapi tantangan, seperti perubahan selera pasar dan persaingan dengan produk tekstil lain[5]. Oleh karena itu, diperlukan inovasi dan dukungan dari berbagai pihak agar batik tetap relevan dan berkelanjutan di era modern. Pemerintah dan berbagai organisasi kebudayaan terus berupaya mempromosikan batik, baik melalui pameran budaya, pendidikan, hingga penerapan batik dalam berbagai produk modern[6]. Upaya ini bertujuan agar batik tidak hanya dikenal sebagai warisan leluhur, tetapi juga menjadi bagian dari gaya hidup masyarakat masa kini. Pendidikan tentang batik juga semakin digalakkan di sekolah-sekolah untuk meningkatkan apresiasi generasi muda terhadap budaya bangsa[7]. Industri batik mengalami peningkatan yang drastis terutama tentang cara pembuatan batik yang dulu menggunakan teknik tulis, lukis dan cap, ketiga metode pembuatan batik merupakan karya seni yang dibuat tangan (handmade), kemudian mengalami perubahan dengan munculnya metode batik cetak, perubahan ini juga didasari oleh permintaan pembeli yang selalu meningkat. Kendala yang dihadapi industri batik adalah pada hubungan pelanggan[8]. Oleh karena itu diperlukan sebuah metode cerdas untuk mengatasi permasalahan pada komunikasi hubungan dengan pelanggan. Industri Batik mempunyai potensi untuk memanfaatkan teknologi chatbot, dalam kondisi ini chatbot berfungsi untuk memberikan aksesibilitas yang lebih baik bagi masyarakat dalam memperoleh informasi dan layanan terkait informasi batik[9]. Mesin dan komputer juga mampu melakukan pekerjaan apapun layaknya manusia, salah satu yang menjadi perhatian merupakan kemampuan untuk dapat melakukan interaksi dan percakapan dengan manusia menggunakan metode Natural Language Processing (NLP)[10]. NLP merupakan bagian dari Artificial Intelligence yang berkaitan dengan percakapan maupun interaksi antara komputer

dengan manusia menggunakan NLP [11]. Penerapan NLP kerap kali digunakan dalam fitur aplikasi chatbot. Selain itu, NLP juga dimanfaatkan sebagai penerjemah seperti Google Translate, personal asisten misalnya penerapan Siri pada produk Apple, pengecekan dan ketepatan penggunaan bahasa seperti Grammarly[12]. Perkembangan teknologi di bidang Artificial Intelligent semakin digemari dan kerap digunakan dalam kehidupan sehari-hari bahkan hingga saat ini[13]. Dimana pada saat ini Artificial Intelligent banyak sekali pemanfaatannya dalam berbagai bidang untuk menjawab pertanyaan mulai dari percakapan yang kompleks hingga sederhana. Salah satu bidang yang memanfaatkan teknologi AI adalah pelayanan bisnis. Pemanfaatan AI dalam bidang bisnis yang sering kita temui adalah Chatbot[14]. Penggunaan metode LSTM sering kali digunakan dalam pembuatan arsitektur chatbot dikarenakan peningkatan efisiensi yang diberikan. Salah satu teknologi yang menunjukkan potensi besar dalam informasi batik adalah penggunaan chatbot berbasis LSTM (Long Short-Term Memory)[15]. Chatbot, yang dikenal juga sebagai aplikasi cerdas percakapan, adalah software komputer yang dirancang agar berkomunikasi dengan user melalui interaksi teks maupun suara. Dalam konteks pengolahan limbah sampah, chatbot berbasis LSTM dapat digunakan untuk memfasilitasi interaksi antara pengguna memberikan informasi, dan saran terkait informasi batik di Karanganyar. Chatbot merupakan program yang menstimulasikan percakapan manusia melalui perintah suara, teks atau keduanya[16].

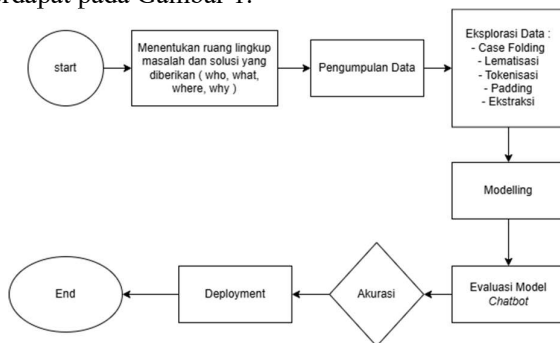
Chatbot sendiri adalah jenis dari *Artificial Intelligence* dan contoh nyata hubungan interaksi manusia dan komputer Dengan adanya chatbot, pengguna dapat dengan mudah mendapatkan panduan dan instruksi yang relevan, memperoleh informasi tentang jenis batik, bahan yang digunakan dalam batik, dan warna batik. Hal ini dapat membantu meningkatkan Masyarakat untuk mengenali batik lebih dalam. Metode yang digunakan dalam chatbot ini adalah Long Short-Term Memory[17], [18]. Dalam pembuatan chatbot dapat menggunakan berbagai macam metode, diantaranya seperti penggunaan Recurrent Neural Network dan Long Short Term Memory (RNN-LSTM)[19], dan Natural Language Processing. Untuk memudahkan komputer dalam memproses Natural Language Processing, user dapat menggunakan metode Long Short-Term Memory yang merupakan algoritma dalam proses pengolahan data berupa penerimaan input dan menghasilkan output data sequences yang merupakan pengembangan dari Recurrent Neural Network[20].

Penelitian ini bertujuan untuk mengembangkan chatbot berbasis LSTM yang dapat digunakan oleh industri batik dalam memberikan layanan informasi yang lebih baik kepada pelanggan. Selain itu, penelitian ini juga bertujuan untuk menganalisis efektivitas penerapan LSTM dalam meningkatkan kualitas interaksi chatbot di industri kreatif. Kontribusi dari penelitian ini diharapkan dapat memperkaya literatur terkait pengembangan chatbot untuk industri kreatif, khususnya dalam konteks pelestarian budaya melalui teknologi.

## II. METODE PENELITIAN

### 2.1 Tahapan Penelitian

Tahapan dalam pengembangan chatbot ini menggunakan AI Project Cycle yang memiliki langkah – langkah seperti penentuan lingkup masalah, pengumpulan data, eksplorasi data, modelling dan evaluasi [21]. Tahapan penelitian terdapat pada Gambar 1.



**Gambar 1.** Flowchart tahapan penelitian Chatbot

Pada tahap penentuan ruang lingkup masalah peneliti menetapkan rumusan masalah dan fungsi penelitian ini, kemudian mengumpulkan data berupa dataset percakapan sebanyak 42 tag *class* percakapan. Pada tahap eksplorasi data dilakukan preprocessing data untuk 42 tag class sebelum masuk ke model training. Modelling yang diimplementasikan dalam chatbot ini menggunakan LSTM agar membentuk model AI, lalu model tersebut di evaluasi dan mengeluarkan *output* akurasi model sehingga bisa dilakukan *deploy* pada web *localhost*.

#### 2.1.2 Ruang Lingkup Masalah

Dalam proses ini untuk mengetahui ukuran masalahnya. Langkahnya adalah mengidentifikasi dan memproses masalah secara komprehensif, setelah itu menentukan fungsi untuk menyelesaikan masalah secara efisien. Metode 4Ws digunakan untuk mejadi acuan dalam menentukan ruang

lingkup masalah dan meringankan dalam proses menyelesaikan masalah dengan menganalisa siapa, apa, dimana dan mengapa. Metode ini berkontribusi menjelaskan aspek-aspek yang penting. “Siapa” yaitu keteranga masalah telibat yang relevan, “Apa” menganalisa perilaku masalah, “kapan” yaitu mengarah pada waktu masalah itu muncul, dan “mengapa” muncul pada solusi yang diperlukan. Dengan menerangkan setiap aspek 4W secara detail, dapat memudahkan menganalisa masalah secara menyeluruh dan merumuskan solusi khusus untuk mengatasi masalah tersebut.

#### 2.1.3 Pengumpulan Data

Pelatihan model prediktif adalah tentang belajar dari data yang diambil. Karena itu, memiliki data berkualitas tinggi merupakan prasyarat untuk setiap aktivitas pembagian mesin. Pada penelitian ini penulis mengumpulkan data question dan answering terkait informasi tentang industri batik menggunakan studi pustaka atau menggunakan jurnal jurnal topik yang penulis angkat pada penelitian ini. Dataset yang akan penulis gunakan untuk melatih model chatbot ini berformat JSON file dan dataset tersebut penulis petakan menggunakan konsep 4W. Pada dataset yang digunakan untuk pemodelan chatbot LSTM ini terdapat beberapa objek seperti :

- a) *Intens* : Berisi 3 objek yaitu Tags, Pattern, dan Responses
- b) *Tags* : Pada tags berisi kategori *question* dan *answer*
- c) *Pattern* : Pattern berisi tentang pertanyaan yang akan digunakan untuk melatih model chabot
- d) *Responses* : Untuk responses sendiri berisi tentang tanggapan atau jawaban yang akan dijadikan sebagai *output* dari hasil prediktif yang diberikan oleh pengguna.

#### 2.1.4 Eksplorasi

Tahap pemrosesan data adalah langkah awal yang sangat berpengaruh terhadap hasil akhir dari pemodelan chatbot. Pada penelitian ini, data preprocessing mencakup beberapa bagian meliputi:

- a) *Case folding*

*Case folding* adalah langkah dalam pemrosesan teks yang bertujuan untuk mengganti keseluruhan huruf kapital didalam suatu teks menjadi huruf kecil maupun besar secara konsisten. Proses ini bertujuan untuk memastikan keseragaman dalam penulisan, sehingga memudahkan saat melakukan

pencarian dan perbandingan teks tanpa perlu memikirkan perbedaan kapitalisasi. Dalam konteks analisis data dan pengolahan teks, case folding memainkan peran penting karena variasi kapitalisasi dapat menyebabkan hasil pencarian yang tidak konsisten, terutama ketika teks harus dibandingkan dengan pola tertentu atau dikelompokkan berdasarkan kata-kata yang serupa. Dengan mengonversi semua huruf menjadi bentuk yang seragam, case folding membantu memastikan bahwa setiap kata atau frasa dalam teks dapat dianalisis tanpa hambatan perbedaan huruf kapital, meningkatkan akurasi dalam analisis lebih lanjut, seperti dalam pengklasifikasian data atau ekstraksi informasi.

b) Lematisasi

Lematisasi adalah proses transformasi kata-kata dalam teks menjadi bentuk dasarnya atau bentuk akar, yang berfungsi untuk menyatukan berbagai variasi kata sehingga lebih mudah dianalisis. Dalam analisis teks dan pemrosesan bahasa alami, lematisasi sangat penting karena banyak kata memiliki variasi bentuk yang berbeda akibat konjugasi, perubahan tense, atau bentuk pluralitas. Dengan mengonversi setiap kata menjadi bentuk dasarnya, lematisasi membantu menyederhanakan data teks, memungkinkan perangkat lunak atau algoritma analisis untuk melihat setiap kata sebagai entitas yang konsisten. Hal ini sangat membantu dalam proses klasifikasi, pengelompokan, dan ekstraksi informasi, karena makna dasar dari kata tetap terjaga, sementara variasi bentuk yang berbeda dapat diabaikan.

c) Tokenisasi

Tokenisasi merupakan langkah dalam pemrosesan teks kata-kata dipecah menjadi bagian-bagian kecil yang disebut "token," yang bisa berbentuk kata, frasa, maupun simbol tertentu. Peran utama dari tokenisasi bertujuan dalam mempermudah analisis dan pengolahan teks dengan memisahkannya menjadi unit-unit yang lebih kecil dan lebih terstruktur. Dalam analisis data teks, tokenisasi sangat bermanfaat karena memungkinkan setiap kata atau frasa dianalisis secara individual, sehingga dapat mendeteksi pola-pola tertentu atau memahami struktur kalimat dengan lebih jelas. Proses ini menjadi dasar bagi berbagai teknik pengolahan bahasa alami, seperti analisis sentimen, klasifikasi teks, dan ekstraksi informasi, karena setiap token dapat diolah dan dianalisis sebagai satuan informasi tersendiri yang merepresentasikan makna tertentu dalam konteks

yang lebih besar. Dengan tokenisasi, teks mentah dapat diuraikan menjadi unit-unit yang lebih sederhana, yang kemudian siap untuk diproses lebih lanjut dalam model atau algoritma bahasa pemrograman.

d) Padding

Padding adalah tahapan pengembangan elemen, berupa angka nol atau atribut tertentu, ke dalam data untuk memastikan panjangnya seragam atau sesuai dengan yang dibutuhkan. Dalam konteks ilmu/pengetahuan terkait machine learning, padding digunakan secara luas untuk mengatur panjang input data, seperti teks atau deretan angka, sehingga model dapat memprosesnya dengan efektif walaupun data memiliki panjang yang berbeda-beda. Dengan demikian, padding membantu menciptakan konsistensi dalam panjang data yang diumpankan ke model, memungkinkan proses komputasi yang lebih efisien dan hasil yang lebih akurat.

e) Ekstraksi

Ekstraksi output encoding adalah tahapan mengubah data dari suatu format atau bentuk ke format maupun bentuk lain, misalnya memodifikasi teks menjadi bilangan biner atau mengganti atribut menjadi representasi angka. Dalam dunia komputasi, encoding dapat juga mengarah pada perubahan data ke bentuk format tertentu, seperti UTF8 untuk teks, yang mampu untuk merepresentasikan atribut dari berbagai bahasa.

### 2.1.5 Pemodelan Data

Tahap pemodelan data dilakukan jika dataset sudah melalui tahap preporocessing. Setelah dataset sudah siap Langkah berikutnya adalah sebagai berikut :

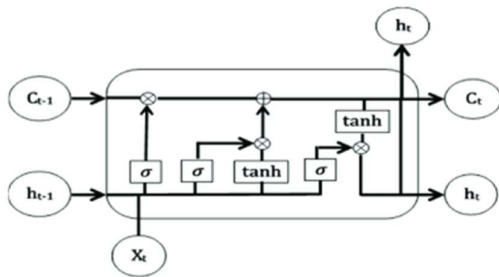
- a) Dataset dibagi menjadi dua bagian, data training dan data validation guna mendapatkan hasil pemodelan yang baik. Pada pembagian data ini atau bisa disebut split data, rasio pembagiannya adalah data training sebesar 80% dan data validation sebesar 20%.
- b) Untuk data training sendiri akan digunakan sebagai bahan pelatihan model dan untuk data validation akan digunakan sebagai bahan uji pengujian model yang telah dibuat.
- c) Langkah selanjutnya adalah pembuatan layer model, layer model menentukan seberapa prediktif model yang dibuat. Disini penulis menggunakan beberapa layer seperti layer embedding, layer LSTM, layer bidirectional, layer flatten, layer dense, layer dropout, dan

layer regulasi. Dari beberapa layer tersebut menggunakan aktivasi layer relu dan softmax.

- d) Setelah layer model dibuat, selanjutnya melakukan proses training menggunakan dataset yang telah di split sebelumnya. Pada tahapan ini penulis menggunakan optimizer Adam dengan Learning rate 0.001. Tahapan training model ini mempunyai epochs 400 atau melakukan training secara berulang sebanyak 400 kali dan akan berhenti jika accuracy mencapai 0.9 serta validation accuracy nya sudah menyentuh 0.8 agar hasil prediktif yang dihasilkan sesuai dengan apa yang diharapkan.

**2.1.6 Modelling**

Dalam penelitian ini, peneliti memilih model LSTM karena metode ini sangat sesuai untuk memproses data berurutan. LSTM memiliki sel yang berfungsi menyimpan informasi yang dikumpulkan oleh algoritma LSTM, dan terdapat komponen yang disebut gate untuk mengatur modifikasi memori. Terdapat tiga jenis gate dalam algoritma LSTM, yaitu *Forget Gate*, *Input Gate*, dan *Output Gate*.



**Gambar 2.** Arsitektur LSTM

- a) *Forget Gate*  
Berperan dalam menghilangkan informasi yang kurang relevan dari sel. Prosesnya melibatkan penilaian output biner berdasarkan dua input, yaitu  $x(t)$  dan output sel sebelumnya, yang kemudian melalui fungsi aktivasi untuk menghasilkan output biner.
- b) *Input Gate*  
Input gate mengubah kondisi sel dengan menambahkan informasi terkait. Pendekatannya serupa dengan forget gate, menggunakan input  $h(t-1)$ , dan dimulai dengan mengatur informasi menggunakan fungsi sigmoid untuk menyaring nilai yang akan disimpan
- c) *Output Gate*

Tugas output gate adalah mengambil informasi yang berarti dari cell state saat ini dan mengirimkannya sebagai nilai keluaran. Fungsi tanh pertama kali

digunakan untuk membuat vector dari sel. Data tersebut kemudian diatur menggunakan fungsi sigmoid, yang mem-filter input menjadi nilai yang disimpan.

**III. HASIL DAN PEMBAHASAN**

**3.1 Ruang Lingkup Masalah**

Dalam penjabaran cakupan masalah, peneliti menggunakan konsep unsur 4Ws yang meliputi Who, What, Where, dan Why. Berikut cakupan dalam pengembangan aplikasi ini :

**Tabel 1.** Konsep 4Ws

| Konsep 4Ws | Penjelasan  |
|------------|---|
| Who        | Pada penelitian ini, sasaran utama dalam pengembangan aplikasi adalah masyarakat. Aplikasi ini dirancang dengan tujuan untuk melayani kebutuhan dan kepentingan masyarakat luas sebagai pengguna utamanya.  |
| What       | Bagian ini bertujuan untuk memverifikasi adanya masalah yang menjadi fokus pembahasan penelitian, agar dapat memberikan landasan pengetahuan yang kuat untuk melanjutkan penelitian terkait. Masalah utama yang menjadi fokus adalah meningkatkan pemahaman dan wawasan masyarakat mengenai batik, agar mereka memiliki pengetahuan yang lebih luas tentang warisan budaya tersebut.  |
| Where      | Akar permasalahan yang mendorong dilakukannya penelitian ini terletak pada kurangnya informasi yang tersedia mengenai kain batik di Karanganyar. Situasi ini menjadi pendorong utama untuk melakukan penelitian, dengan tujuan mengatasi kekurangan informasi tersebut dan meningkatkan kesadaran serta pengetahuan masyarakat tentang batik di daerah tersebut   |
| Why        | Pada bagian ini, dijelaskan alasan-alasan yang mendukung pentingnya melanjutkan penelitian ini. Penelitian ini bertujuan untuk memberikan informasi yang komprehensif dan menawarkan solusi melalui website yang berfokus pada informasi mengenai batik. Tujuan utamanya adalah memanfaatkan website sebagai platform untuk menyebarluaskan pengetahuan tentang batik, sehingga masyarakat dapat lebih mudah mengakses informasi yang relevan dan akurat terkait dengan kain batik. |

### 3.2 Perolehan Data

Data yang digunakan untuk melatih model algoritma chatbot dalam penelitian ini berasal dari kumpulan data manual yang disimpan dalam format file JSON. Data ini mencakup beberapa elemen penting yang bersama-sama membentuk dasar dari kemampuan chatbot untuk memahami dan merespons input pengguna. Elemen pertama adalah intents, yang mewakili tujuan atau maksud percakapan yang ingin dikenali oleh chatbot. Setiap intent ini memiliki label tag yang digunakan untuk mengelompokkan berbagai intents ke dalam kategori tertentu, memudahkan proses klasifikasi saat chatbot beroperasi. Selain itu, elemen pattern berisi pola atau frasa yang mungkin digunakan oleh pengguna, yang terkait dengan intent tertentu. Misalnya, pola-pola ini dapat berupa pertanyaan atau pernyataan yang diharapkan oleh sistem untuk mengenali maksud dari pengguna. Terakhir, elemen response adalah tanggapan yang telah disiapkan oleh chatbot berdasarkan intent yang telah diidentifikasi dari input pengguna. Dengan memanfaatkan data yang terstruktur seperti ini, chatbot dapat mengenali maksud pengguna secara akurat dan memberikan respons yang relevan, menciptakan pengalaman interaktif yang lebih alami.

```
{
  "tag": "sejarah_batik",
  "patterns": [
    "Bagaimana sejarah batik di Karanganyar berkembang?",
    "Apa yang menjadi asal mula perkembangan batik di Karanganyar?",
    "Ceritakan tentang sejarah dan perkembangan batik di Karanganyar.",
    "Bagaimana asal mula batik bisa berkembang di daerah Karanganyar?",
    "Bagaimana latar belakang sejarah batik di Karanganyar?",
    "Apa yang mempengaruhi perkembangan batik di Karanganyar?",
    "Bagaimana kisah sejarah batik di Karanganyar bisa menjadi seperti sekarang?",
    "Bagaimana proses sejarah batik mulai dan berkembang di Karanganyar?",
    "Apa saja yang menjadi bagian dari sejarah batik di Karanganyar?",
    "Bagaimana sejarah batik dimulai dan berkembang di Karanganyar?"
  ],
  "responses": [
    "Batik di Karanganyar memiliki sejarah panjang yang berakar dari pengaruh Keraton Surakarta pada abad ke-19"
  ]
}
```

**Gambar 4.** Dataset dalam bentuk JSON

Gambar 4 menunjukkan dataset dalam format JSON yang digunakan sebagai basis data untuk chatbot. Dalam dataset ini, elemen "tag" bertindak sebagai kategori atau label yang mengklasifikasikan pernyataan atau pertanyaan dari *user*, mewakili topik atau konsep tertentu. Sebagai contoh, tag "greeting" digunakan untuk menyapa *user* saat awal interaksi, sementara tag "terima kasih" digunakan untuk mengelompokkan ungkapan terima kasih untuk mengakhiri percakapan. Elemen "patterns" mencakup pola-pola atau susunan kalimat yang berhubungan dengan setiap tag, yaitu contoh-contoh pertanyaan dan pernyataan yang sesuai dengan konsep dari tag terkait. Sebagai contoh, untuk tag "greeting," pola dapat mencakup variasi sapaan seperti "Hai" atau "Helo."

Elemen "responses" berisi daftar jawaban yang telah ditetapkan untuk setiap tag, yang akan dipilih oleh chatbot sebagai respons yang sesuai terhadap input pengguna. Sebagai contoh, untuk tag "greeting," respons yang mungkin diberikan adalah "Hai, saya adalah Batik+ ada yang bisa saya bantu?" atau variasi sapaan serupa. Dengan susunan JSON yang mencakup elemen-elemen seperti tag, patterns, dan responses, chatbot dapat mempelajari dan mengenali konteks dari input pengguna serta memberikan respons yang relevan. Struktur ini memungkinkan chatbot untuk merespons sesuai dengan konsep atau topik yang telah diatur dalam dataset, meningkatkan akurasi dan kesesuaian respons terhadap kebutuhan pengguna.

### 3.3 Eksplorasi Data

Tahapan berikutnya dalam pengembangan model chatbot mencakup proses upload dataset yang telah disatukan dalam format JSON dan kemudian mengonversinya ke dalam bentuk dataframe. Dalam dataframe ini, terdapat dua kolom utama, yaitu pola (patterns) dan kelas (tags), yang berfungsi untuk mengorganisasi data secara lebih terstruktur dan mudah diakses dalam proses pelatihan model. Tujuan utama dari langkah ini adalah menyusun data dalam format yang sistematis sehingga setiap pola teks dapat diklasifikasikan sesuai dengan label yang mewakili kategori atau konteks tertentu. Elemen "responses" dalam dataset ini dimanfaatkan untuk memberikan respons atau jawaban yang sesuai dengan label (tags) yang diprediksi oleh model, sehingga chatbot mampu merespons sesuai dengan maksud atau tujuan pengguna. Dataset ini berisi total 720 baris data yang terbagi dalam 59 kategori atau kelas yang berbeda, memberikan keragaman pola interaksi yang dapat dikenali oleh chatbot.

Setelah mempelajari karakteristik dataset, proses preprocessing data dimulai dengan tahapan case folding. Case folding adalah proses normalisasi teks dengan menyatukan penulisan huruf dalam dokumen agar konsisten, sehingga pencarian dan analisis teks menjadi lebih mudah. Mengingat tidak semua dokumen mengikuti aturan yang konsisten dalam penggunaan huruf kapital, case folding memungkinkan seluruh teks dalam dokumen diubah menjadi huruf kecil. Hal ini bertujuan untuk menciptakan format teks standar, di mana setiap kata atau frasa dapat dibandingkan tanpa perlu memperhatikan perbedaan kapitalisasi yang dapat mengganggu konsistensi data.

Langkah selanjutnya dalam proses preprocessing adalah menerapkan lemmatisasi sebagai upaya untuk melakukan normalisasi data. Pos tagged yang sudah

dibuat sebelumnya kemudian dikonfigurasi kembali ke dalam fungsi lemmatisasi guna mengembalikan setiap kata ke bentuk dasarnya yang sesuai dengan Kamus Bahasa Indonesia. Melalui proses ini, diperoleh 241 kata unik yang mewakili keseluruhan dataset, serta total 720 dokumen yang terdiri dari elemen patterns dan intents.

Setelah lemmatisasi, fitur-fitur dalam data diekstraksi melalui tokenisasi. Tokenisasi adalah proses pemisahan teks dalam dokumen menjadi unit-unit kecil yang disebut token. Setiap kalimat dalam dokumen dipecah menjadi token-token individual dan dibatasi hingga 2000 kata yang paling sering muncul. Kata-kata yang berada di luar batas ini dianggap sebagai kata tidak dikenal dan diabaikan dalam proses selanjutnya.

Setelah tokenisasi selesai, setiap urutan token diindeks dan dikonversi menjadi bentuk urutan angka yang merepresentasikan kata-kata dalam data patterns. Untuk memastikan bahwa semua urutan memiliki panjang yang seragam, padding diterapkan dengan cara menambahkan angka nol di awal atau akhir urutan. Hal ini dilakukan agar seluruh input memiliki panjang yang sama, mengingat lapisan embedding dalam model hanya dapat menerima input dengan panjang yang konsisten. Pada kasus ini, panjang maksimal urutan yang diinginkan adalah 14 kata per dokumen. Setelah proses padding, langkah berikutnya adalah menerapkan label encoding pada variabel target, yaitu kolom tags dalam data. Label encoding mengonversi data kategorikal pada kolom tags menjadi representasi numerik dalam bentuk vektor biner, sehingga model dapat mengenali kategori tersebut dengan lebih baik. Proses preprocessing diakhiri dengan menyimpan hasil label encoding dan tokenisasi ke dalam file pickle, yang akan memudahkan dalam pemanggilan data pada tahap pelatihan model berikutnya.

### 3.4 Modelling

Pertama, digunakan layer Embedding dengan ukuran embedding sebesar 100 untuk mengonversi kata-kata dari data sekuens menjadi vektor numerik yang merepresentasikan hubungan semantik antar kata. Kemudian, model dilanjutkan dengan Bidirectional LSTM yang memiliki 64 unit dan `'return_sequences=True'` untuk menangkap hubungan temporal baik dari arah maju maupun mundur, sambil mempertahankan urutan keluaran untuk setiap langkah waktu. Selanjutnya, dilakukan Dropout dengan tingkat 0.5 untuk mengurangi risiko overfitting dan `**Batch Normalization` untuk menstabilkan distribusi keluaran dari lapisan sebelumnya.

Proses ini diikuti oleh Bidirectional GRU dengan 128 unit, yang juga memiliki `'return_sequences=True'`, untuk menangkap pola sekuensial tambahan secara efisien. Keluaran dari lapisan ini kemudian diratakan menggunakan Flatten layer, mengubah data multi-dimensi menjadi vektor satu dimensi yang siap untuk diproses lebih lanjut. Setelah itu, Dense layer dengan 100 unit dan fungsi aktivasi ReLU diterapkan, dilengkapi dengan regularisasi L2 untuk mencegah overfitting. Dropout tambahan sebesar 0.5 dan Batch Normalization diterapkan kembali untuk menjaga stabilitas dan generalisasi model. Pada bagian akhir model, Output layer menggunakan Dense layer dengan fungsi aktivasi softmax, yang bertugas mengklasifikasikan data ke dalam kategori sesuai dengan jumlah kelas. Dalam bagian compiling the model, digunakan optimizer Adam dengan nilai learning rate default untuk mengoptimalkan proses pelatihan model. Fungsi kerugian yang digunakan adalah categorical cross-entropy, mengingat tugas ini adalah klasifikasi multi-kelas dengan keluaran berupa probabilitas untuk setiap kelas. Selain itu, metrik akurasi juga ditentukan untuk mengevaluasi kinerja model. Model yang telah dirancang kemudian dirangkum menggunakan fungsi `'model.summary()'` untuk memeriksa parameter dan arsitektur keseluruhan.

### 3.5 Evaluasi

Proses training algoritma model dilakukan dengan penuh kehati-hatian dan diulang berkali-kali untuk mengonfirmasi bahwa model dapat mencapai tingkat akurasi yang tinggi. Langkah ini juga bertujuan untuk mengurangi nilai kesalahan atau loss selama pelatihan, sehingga model dapat bekerja lebih efisien. Untuk mencapai hasil yang optimal, model dilatih melalui beberapa iterasi, di mana parameter-parameter model secara bertahap disesuaikan. Penyesuaian ini dilakukan untuk meningkatkan kemampuan model dalam memprediksi output yang benar, memastikan bahwa model dapat mempelajari pola dengan lebih baik dan memberikan hasil yang lebih akurat dalam prosesnya. Evaluasi kinerja model LSTM dilakukan dengan menggunakan beberapa metrik, salah satunya adalah akurasi. Akurasi ini mengukur seberapa sering model menghasilkan prediksi yang benar dibandingkan dengan total prediksi yang dilakukan. Selain itu, fungsi loss yang digunakan adalah `categorical_crossentropy`, yang merupakan fungsi standar untuk mengukur kesalahan dalam model klasifikasi multi-kelas. Fungsi ini dioptimalkan menggunakan metode adam, yaitu algoritma optimasi yang populer karena efisiensinya dalam memperbarui bobot model selama proses pelatihan. Dengan kombinasi metrik akurasi dan optimasi `*categorical_crossentropy`

menggunakan metode adam, model dapat mencapai performa yang lebih baik dalam memprediksi output yang sesuai dengan data input.

**Tabel 2.** Akurasi Loss Pelatihan Model LSTM Berdasarkan jumlah epochs

| Epochs | Accuracy (%) | Val_accuracy (%) | Loss (%) | Val_loss (%) |
|--------|--------------|------------------|----------|--------------|
| 100    | 87           | 81               | 1.75     | 2.07         |
| 200    | 93           | 81               | 1.85     | 2.87         |
| 300    | 95           | 84               | 1.55     | 2.85         |
| 400    | 96           | 85               | 1.30     | 3.02         |

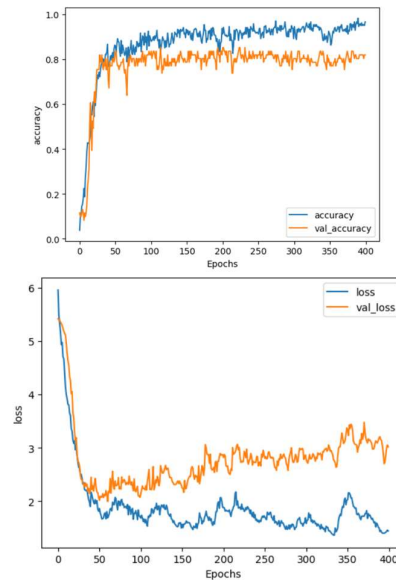
Setelah melalui 400 iterasi pelatihan yang dijelaskan dalam Tabel 2, model LSTM berhasil mencapai tingkat akurasi yang tinggi, yaitu 96%, dengan akurasi validasi sebesar 85 %. Selain itu, model menunjukkan nilai loss yang rendah, yaitu sekitar 1.3%, serta validasi loss sebesar 2.5%. Dari training model peneliti mendapat precision sebesar 87% dan recall sebesar 81%, akurasi dapat dihitung dari precision dan recall dengan cara sebagai berikut :

$$F1\ score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

$$F1\ score = 2 \cdot \frac{0.874 \cdot 0.819}{0.874 + 0.819}$$

$$= 0,844\ (84\%)$$

Untuk memberikan gambaran yang lebih jelas, visualisasi grafik yang menunjukkan analisis akurasi dan loss selama proses pelatihan model LSTM dapat dilihat pada Gambar 5. Grafik ini membantu dalam memahami bagaimana model berkembang seiring dengan bertambahnya iterasi.



**Gambar 5.** Grafik Analisis Akurasi, Val\_accuracy, Loss dan Val\_loss Pelatihan Model LSTM

Gambar 4 menunjukkan bahwa model chatbot yang dilatih menggunakan algoritma LSTM menunjukkan kinerja yang cukup baik, namun terdapat indikasi terdapatnya over fitting pada tingkatan yang tergolong rendah. Hal tersebut terlihat dari perbedaan antara loss pada data pelatihan dan data validasi, di mana loss pada data pelatihan lebih rendah dibandingkan dengan loss pada data validasi. Meskipun overfitting terjadi, dampaknya terhadap kualitas jawaban chatbot tidak terlalu signifikan, karena model masih mampu memberikan respons yang memadai. Overfitting ini mungkin disebabkan oleh model yang terlalu beradaptasi dengan rincian-rincian kecil dalam data pelatihan, sehingga kemampuannya untuk menggeneralisasi pada situasi nyata yang berbeda dari data pelatihan bisa terganggu. Meskipun demikian, hal ini perlu diperhatikan dan diatasi untuk meningkatkan kemampuan model dalam menghadapi berbagai situasi dan pertanyaan yang mungkin dihadapi di dunia nyata.

### 3.6 Deployment

Penelitian ini berfokus dalam upaya menyederhanakan penggunaan model chatbot yang telah dikembangkan dengan cara melalui proses deployment. Tahap deployment ini bertujuan mengintegrasikan model chatbot ke dalam sebuah aplikasi web menggunakan framework Flask, di mana Javascript, HTML, dan CSS digunakan untuk membangun antarmuka pengguna yang intuitif dan menarik. Dengan demikian, aplikasi chatbot memiliki halaman utama yang mencakup komponen

navigasi serta area khusus untuk interaksi chatbot, yang dirancang untuk memberikan pengalaman pengguna yang nyaman dan responsif.

Fungsionalitas chatbot ini ditingkatkan melalui penerapan empat fungsi JavaScript. Fungsi pertama menangani proses pengiriman dan tampilan pesan dari pengguna ke chatbot, sementara fungsi kedua bertugas memanggil API melalui AJAX untuk mengirimkan pesan dan menerima respons dari model chatbot. Fungsi ketiga berfokus pada menampilkan hasil respons yang diterima, sedangkan fungsi keempat mencakup pemanggilan waktu yang akan tampil bersamaan dengan percakapan, memberikan konteks waktu yang lebih nyata dalam interaksi.

Selanjutnya, tahap perancangan routing dilakukan untuk mengaitkan alamat URL tertentu dengan fungsi-fungsi spesifik dalam aplikasi. Misalnya, halaman utama chatbot diakses melalui rute "/" sementara respons API chatbot diakses melalui rute "/get". Setelah semua rute ditentukan, model chatbot dan elemen pendukung lainnya dimuat ke dalam aplikasi, yang kemudian dijalankan menggunakan railway sebagai server. Dengan railway, aplikasi dapat diakses secara online maupun melalui jaringan lokal, memudahkan pengguna untuk mencoba fitur chatbot di mana saja.

Gambar 6 menunjukkan tampilan halaman chatbot yang siap digunakan, menampilkan antarmuka yang telah dirancang dengan elemen navigasi, area interaksi, serta respons yang muncul secara real-time, sehingga memberikan pengalaman pengguna yang lengkap dan interaktif.



**Gambar 6.** Tampilan Chatbot pada Website

Gambar 6 menunjukkan tampilan antarmuka aplikasi berbasis web yang berfungsi sebagai chatbot. Dalam antarmuka ini, pengguna dapat berkomunikasi dengan chatbot melalui kolom pesan yang tersedia. Ketika pengguna mengetik pesan dan menekan tombol kirim, chatbot akan memproses input tersebut dan memberikan jawaban yang relevan sesuai dengan pertanyaan atau pernyataan yang diajukan. Jika pengguna memiliki pertanyaan

tambahan, mereka cukup mengetik pesan baru di kolom yang sama dan mengirimkannya kembali. Fitur ini memungkinkan percakapan yang berkelanjutan antara pengguna dan chatbot dalam satu sesi interaksi.

### 3.7 Pembahasan

Berdasarkan pengujian sistem diatas chatbot ii memiliki akurasi 84% melalui 400 perulangan dari 720 bari data hal ini menunjukkan bahwa sistem mampu menjawab dengan baik. LSTM adalah salah satu metode pembelajaran yang sering digunakan dalam pemrosesan teks. Keunggulannya dibandingkan dengan RNN terletak pada kemampuannya dalam menangani urutan data yang panjang. RNN memiliki kendala berupa vanishing gradient, yaitu hilangnya efektivitas gradien selama proses pelatihan. Namun, masalah tersebut dapat diatasi dengan menggunakan LSTM[18].

## IV. KESIMPULAN

Penelitian ini berfokus pada pengembangan aplikasi chatbot untuk meningkatkan pemahaman masyarakat tentang batik, terutama di wilayah Karanganyar. Dengan menerapkan konsep 4Ws (Who, What, Where, Why), penelitian ini mengidentifikasi bahwa masyarakat luas adalah sasaran utama dari aplikasi ini. Tujuan dari aplikasi ini adalah untuk mengatasi kekurangan informasi yang ada mengenai batik dan menyediakan solusi dengan menyebarkan pengetahuan yang lebih luas dan mudah diakses. Model chatbot yang dikembangkan dilatih menggunakan algoritma Long Short-Term Memory (LSTM), yang dikenal sangat efektif dalam memproses urutan data teks. Dataset yang digunakan dalam pelatihan diambil dari berbagai pola percakapan yang disusun dalam format JSON. Proses pelatihan model dilakukan dengan teliti dan berulang kali untuk mencapai akurasi yang tinggi. Hasil pelatihan menunjukkan akurasi model sebesar 96% setelah 400 iterasi, meskipun ada sedikit overfitting yang terlihat dari perbedaan loss antara data pelatihan dan data validasi. Namun, meski ada overfitting, model tetap mampu memberikan jawaban yang memadai dan relevan. Setelah pelatihan, model chatbot diintegrasikan ke dalam aplikasi web menggunakan Flask, sebuah framework yang memungkinkan integrasi mudah dengan teknologi web lainnya seperti HTML, CSS, dan JavaScript. Pengguna dapat berinteraksi dengan chatbot melalui antarmuka yang dirancang secara intuitif, di mana mereka dapat mengirim pertanyaan dan menerima jawaban secara

real-time. Aplikasi ini tidak hanya mempermudah akses informasi tentang batik, tetapi juga menciptakan pengalaman interaktif yang dapat membantu meningkatkan kesadaran dan pengetahuan masyarakat tentang warisan budaya ini.

## REFERENSI

- [1] T. M. Tanjung and R. Hidayat, "IMPLEMENTATION OF TWO LANGUAGE CHATBOT WEB TO FIND INFORMATION ON BATIK USING THE ARTIFICIAL NEURAL NETWORK MODEL," *TEKNOKOM*, vol. 6, no. 2, pp. 50–57, Jul. 2023, doi: 10.31943/teknokom.v6i2.125.
- [2] S. I. Syed Shaharuddin *et al.*, "A Review on the Malaysian and Indonesian Batik Production, Challenges, and Innovations in the 21st Century," *Sage Open*, vol. 11, no. 3, Jul. 2021, doi: 10.1177/21582440211040128.
- [3] Adi Kusrianto, *MENELUSURI ASAL USUL BATIK Benang Merah antara Sejarah, Dongeng Panji hingga Hasil Riset Modern*. Penerbit Andi, 2021.
- [4] I. S. Eko Nursanty, *BATIK HERITAGE Peningkatan Kemampuan Bersaing Kearifan Budaya Unggul Indonesia*. BUTTERFLY MAMOLI PRESS, 2021.
- [5] A. P. Siregar *et al.*, "Upaya Pengembangan Industri Batik di Indonesia," *Dinamika Kerajinan dan Batik: Majalah Ilmiah*, vol. 37, no. 1, Jun. 2020, doi: 10.22322/dkb.v37i1.5945.
- [6] Ari Wulandari, *Batik Nusantara: Makna Filosofis, Cara Pembuatan, dan Industri Batik*. Publisher Andi, 2022.
- [7] D. Nurcahyanti, A. Sachari, A. H. Destiarmand, and Y. Y. Sunarya, "Teaching Virtuous Values and the Implementation of New Strategies for the Sustainability of Batik in Girilayu, Karanganyar Regency, Indonesia." [Online]. Available: <https://www.researchgate.net/publication/355840598>
- [8] A. P. Sani Alhusain, D. Bidang Ekonomi dan Kebijakan Publik Gedung Nusantara, and S. R. DPR Jl Jend Gatot Subroto, "Achmad Sani Alhusain, Kendala dan Upaya Pengembangan Industri Batik di Surakarta Menuju Standardisasi | 199 KENDALA DAN UPAYA PENGEMBANGAN INDUSTRI BATIK DI SURAKARTA MENUJU STANDARDISASI (Efforts and Obstacles in the Development of Batik Industry in Surakarta towards Standardization)."
- [9] S. Patil, V. M. Mudaliar, P. Kamat, and S. Gite, "LSTM based Ensemble Network to enhance the learning of long-term dependencies in chatbot," *International Journal for Simulation and Multidisciplinary Design Optimization*, vol. 11, p. 25, Dec. 2020, doi: 10.1051/smdo/2020019.
- [10] S. K. Maher, R. M. Gaikwad, and S. S. Nimbhore, "Recurrent Neural Networks for Chatbot Excellence: Examining the Power of LSTM Architecture," 2024, pp. 513–522. doi: 10.1007/978-981-97-6678-9\_45.
- [11] A. S. Oktriwina, "NLP: Kecerdasan Buatan yang Bantu Komputer Pahami Bahasa Manusia," <https://glints.com/id/lowongan/natural-language-processing-adalah/#.YhTsm4hBxPY>. Accessed: Jan. 12, 2024. [Online]. Available: <https://glints.com/id/lowongan/natural-language-processing-adalah/#.YhTsm4hBxPY>
- [12] I. Anas and S. Zakir, "Artificial Intelligence: Solusi Pembelajaran Era Digital 5.0," 2024.
- [13] Y. Yuniati, F. A. Gurning, and H. Artikel, "Pengembangan Chatbot Batik Menggunakan Metode Long Short-Term Memory," vol. 4, no. 2, 2024, doi: 10.47709/digitech.v4i2.4391.
- [14] F. Fidiyanti, A. R. Subagja, R. P. Wachyu, and H. Madiistriyatno, "ANALISIS STRATEGI PENGEMBANGAN BISNIS MENGGUNAKAN TEKNOLOGI ARTIFICIAL INTELEGENGE," 2023.
- [15] M. I. Tri Khaqiqi and N. H. Harani, "Peningkatan Kinerja Chatbot NLP Asisten: Tinjauan Literatur tentang Metode dan Akurasi dalam Aplikasi Berbasis Percakapan," *Jurnal Informatika dan Teknologi Komputer (J-ICOM)*, vol. 5, no. 1, pp. 50–59, Apr. 2024, doi: 10.55377/jicom.v5i1.8242.
- [16] G. A. Santos, G. G. de Andrade, G. R. S. Silva, F. C. M. Duarte, J. P. J. Da Costa, and R. T. de Sousa, "A Conversation-Driven Approach for Chatbot Management," *IEEE Access*, vol. 10, pp. 8474–8486, 2022, doi: 10.1109/ACCESS.2022.3143323.
- [17] T. Lattifia, P. Wira Buana, N. Kadek, and D. Rusjyanthi, "Model Prediksi Cuaca Menggunakan Metode LSTM," 2022.
- [18] F. Pradana Rachman, H. Santoso, and A. History, "Jurnal Teknologi dan Manajemen Informatika Perbandingan Model Deep Learning untuk Klasifikasi Sentiment Analysis dengan Teknik Natural Language Processing Article Info ABSTRACT," vol. 7, no. 2, pp. 103–112, 2021, [Online]. Available: <http://http://jurnal.unmer.ac.id/index.php/jtmi>
- [19] P. Anki, A. Bustamam, H. S. Al-Ash, and D. Sarwinda, "Intelligent Chatbot Adapted from Question and Answer System Using RNN-LSTM Model," *J Phys Conf Ser*, vol. 1844, no. 1, p. 012001, Mar. 2021, doi: 10.1088/1742-6596/1844/1/012001.
- [20] D. Goutham, "AN INTUITIVE APPRAOCH ON LSTM-HYBRID MODEL FOR MULTI-DISEASE PREDICTION AND CHATBOT RECOMMENDATION SYSTEM."
- [21] A. Prasetyo and Taufik Ridwan, "ANALISIS SENTIMEN TERHADAP PEMBERHENTIAN TV ANALOG PADA TWITTER MENGGUNAKAN ALGORITMA NAIVE BAYES," *Jurnal Teknika*, vol. 15, no. 2, pp. 67–74, Sep. 2023, doi: 10.30736/jt.v15i2.991.