



9.25%

SIMILARITY OVERALL

SCANNED ON: 17 JUL 2024, 7:02 PM

Similarity report

Your text is highlighted according to the matched content in the results above.

● IDENTICAL
0.3%

● CHANGED TEXT
8.94%

Report #22060727

BAB I PENDAHULUAN Pendahuluan terdiri dari latar belakang dilakukannya penelitian ini, permasalahan yang diidentifikasi, tujuan dari penelitian, manfaat yang diharapkan, kebaruan serta kerangka penulisan yang akan diimplementasikan. 1.1 Latar Belakang Masalah Bidirectional Long Short-Term Memory atau disingkat BiLSTM merupakan salah satu jenis algoritma deep learning yang dimanfaatkan untuk menangkap hubungan temporal dalam data yang berurutan (sequence). Keunggulan BiLSTM terletak pada kemampuannya memproses data dari dua arah, yaitu dari awal ke akhir (forward) dan dari akhir ke awal (backward), sehingga mampu memahami konteks informasi secara lebih mendalam. BiLSTM telah digunakan untuk menyelesaikan berbagai tugas seperti klasifikasi teks, penerjemahan bahasa, dan analisis sentimen. Genre film merupakan salah satu faktor penentu utama bagi penonton dalam memilih film yang akan ditonton (Immanuel & Isa, 2021). Genre film dapat mencakup kategori seperti aksi, komedi, drama, horor, dan lainnya, memberikan gambaran bagi penonton untuk mengantisipasi jenis pengalaman yang akan didapatkan setelah menonton film tersebut. Dalam industri film, genre film berperan penting sebagai strategi pemasaran dalam menentukan film yang akan diproduksi serta menyesuaikan dengan genre yang sedang populer pada waktu tertentu (Saputra, et al., 2019). Mengingat pentingnya genre film sebagai salah satu faktor dalam memilih film dan menjadi komponen penting pada segi

pemasaran, penelitian ini mengusulkan penggunaan teknik klasifikasi teks pada sinopsis film untuk mengidentifikasi genre secara otomatis. Klasifikasi teks terhadap sinopsis ini bertujuan untuk membantu proses pengkategorian film. Dengan menggunakan algoritma BiLSTM, model klasifikasi genre dilatih untuk mengenali pola-pola spesifik tiap genre dalam sinopsis. Pelatihan dilakukan terhadap data sinopsis yang telah memiliki label genrenya masing-masing, sehingga nantinya model yang diintegrasikan dengan perangkat lunak pada penelitian ini dapat mengidentifikasi genre dari sebuah sinopsis. Diharapkan sistem klasifikasi genre berdasarkan sinopsis yang dibuat dalam penelitian ini dapat membantu pelaku industri film mengidentifikasi kata kunci setiap genre, sehingga dapat membuat sinopsis yang sesuai dengan genre yang ditentukan pada proses pembuatan film.

1.2 Identifikasi Masalah Penelitian ini menetapkan rumusan dan batasan masalah yang menjadi acuan selama proses penelitian berlangsung. Berikut adalah identifikasi masalah yang dirumuskan.

1.2.1 Rumusan Masalah Berdasarkan pendahuluan serta identifikasi masalah yang telah diuraikan di atas, masalah dapat dirumuskan sebagai berikut.

1. Bagaimana implementasi algoritma Bidirectional Long Short-Term Memory (BiLSTM) pada proses klasifikasi?
2. Bagaimana membangun model klasifikasi genre film berdasarkan sinopsis film?

1.2.2 Batasan Masalah Batasan masalah dirumuskan untuk menjaga fokus penelitian pada masalah yang ditetapkan dan memperoleh hasil yang optimal. Adapun batasan masalah yang ditetapkan pada penelitian ini dijabarkan sebagai berikut.

1. Penelitian ini menggunakan dataset yang diperoleh dari situs kaggle.com, yang terakhir diperbarui pada tahun 2020.
2. Penelitian ini akan menggunakan film dari berbagai negara, dengan sinopsis film berbahasa inggris.
3. Genre yang digunakan dalam penelitian ini terdiri dari action, adventure, animation, comedy, crime, drama, horror, romance, sci-fi, dan thriller.

1.3 Tujuan Penelitian Tujuan dari penelitian ini yaitu membuat sebuah perangkat lunak yang mampu melakukan klasifikasi genre film terhadap sinopsis film menggunakan algoritma BiLSTM.

1.4 Manfaat Penelitian Penelitian ini memiliki tiga manfaat yaitu untuk masyarakat, peneliti, dan ilmu pengetahuan.

Dengan berfokus pada tujuan yang ingin dicapai, diharapkan bahwa penelitian ini dapat membawa dampak positif bagi dunia pendidikan. Adapun manfaat tersebut dijabarkan sebagai berikut.

1.4.1 Manfaat bagi Masyarakat
Penelitian ini diharapkan memiliki manfaat bagi masyarakat yang ingin mengetahui genre dari sebuah film. Penerapan Natural Language Processing pada aplikasi yang dikembangkan, diharapkan dapat membantu masyarakat khususnya para pelaku di industri film dalam mengidentifikasi kata kunci pada genre film. Aplikasi ini akan membantu mengenali genre yang sesuai dengan plot atau sinopsis film berbahasa Inggris secara otomatis.

1.4.2 Manfaat bagi Peneliti
Manfaat penelitian ini bagi peneliti meliputi peningkatan kompetensi di bidang Informatika, khususnya dalam bidang Artificial Intelligence, dengan mengimplementasikan algoritma Bidirectional LSTM untuk mengklasifikasi genre film berdasarkan sinopsis. Selain itu, penelitian ini juga mengembangkan pengetahuan peneliti dalam pembuatan sistem tersebut.

1.4.3 Manfaat bagi Pengetahuan
Manfaat penelitian ini bagi ilmu pengetahuan, yaitu menghasilkan pengembangan sistem perangkat lunak berbasis Artificial Intelligence, khususnya pada lingkup Natural Language Processing dalam mengklasifikasi genre film berdasarkan sinopsis menggunakan algoritma Bidirectional LSTM.

1.5 Kebaruan
Kebaruan yang terletak pada penelitian ini adalah penerapan algoritma Bidirectional LSTM untuk mengklasifikasi genre film berdasarkan sinopsis dalam beberapa kategori. Model NLP ini akan diintegrasikan ke dalam sebuah perangkat lunak yang memungkinkan pengguna untuk mengklasifikasi genre film berdasarkan sinopsis yang dimasukkan. Aplikasi ini dilengkapi dengan fitur yang dapat mengklasifikasi genre film beberapa sinopsis sekaligus dan menyimpan hasil klasifikasi tersebut dalam riwayat, sehingga bisa dilihat kembali.

1.6 Kerangka Penulisan Penelitian berupa Tugas Akhir ini dirancang berdasarkan pedoman dari Fakultas Teknologi dan Desain Universitas Pembangunan Jaya yang terdiri dari enam bab, dirincikan sebagai berikut.

2 13 **BAB I**

PENDAHULUAN Bab ini menguraikan secara rinci mengenai latar belakang penelitian, mengidentifikasi masalah yang terbagi menjadi rumusan dan

batasan masalah, serta menetapkan tujuan dan manfaat penelitian. Selain itu, bab ini juga menunjukkan kebaruan yang dihasilkan dari penelitian yang dilakukan dan menjelaskan metode penyajiannya dalam dokumen ini. **15** **BAB II TINJAUAN PUSTAKA** Bab ini mengulas pencapaian yang telah dilakukan pada penelitian terdahulu dan teori-teori yang relevan dengan pelaksanaan penelitian ini dalam konteks tinjauan teoritis. **BAB III TAHAPAN PELAKSANAAN** Bab ini menguraikan langkah-langkah pelaksanaan dan pendekatan pengujian yang dijalankan dalam rangkaian penelitian ini. **BAB IV PERANCANGAN** Bab ini menjabarkan terkait analisis sistem pada penelitian sebelumnya, identifikasi kebutuhan sistem baru, dan proses perancangan sistem. **1 2 3 26** **BAB V HASIL DAN PEMBAHASAN** Bab ini secara rinci menyajikan temuan serta pembahasan penelitian yang telah terealisasi. Hasil serta pembahasan ini berlandaskan pada uraian tujuan dan rancangan penelitian yang dipaparkan pada bab – bab sebelumnya. **2 3**

18 **BAB VI PENUTUP** Bab ini memuat kesimpulan atau ringkasan penelitian yang dilakukan serta saran yang diberikan oleh peneliti untuk penelitian selanjutnya. Bagian ini juga menjelaskan inti dari penelitian tersebut, sehingga pembaca dapat memahami tujuan serta hasil yang telah dicapai oleh peneliti. **8** **BAB II TINJAUAN PUSTAKA** Pada bab ini disajikan tinjauan pustaka yang mencakup referensi penelitian sebelumnya serta tinjauan teoritis yang menjadi acuan bagi peneliti untuk memperkaya teori dalam melakukan perancangan aplikasi serta memperkuat landasan penelitian. **2.1 Pencapaian Terdahulu** Pada penelitian ini, referensi penelitian sebelumnya digunakan sebagai bahan rujukan dan landasan untuk penulis. Oleh karena itu, referensi yang diacu mencakup topik-topik mengenai penerapan metode komputasi yang sejenis. **Tabel 2. 1 Hasil Pencapaian Terdahulu** No . Nama Peneliti Publikasi Judul Hasil 1. **23** **Junita Amalia, Juanda Pakpahan, Melani Pakpahan, Yeni Panjaitan JATISI (Jurnal Teknik Informatika dan Sistem Informasi), 2022. Natural Language Processing (NLP) Model Klasifikasi Berita Palsu Menggunakan Bidirectional LSTM Dan Word2Vec Sebagai Vektorisasi Penelitian ini membuat model untuk mendeteksi berita palsu yang tersebar di media sosial dan platform sms di era digital. Model dilatih untuk mendeteksi berita palsu dari segi judul dan konten**

menggunakan metode Bidirectional LSTM dan embedding Word2Vec. Hasil pengujian menunjukkan bahwa akurasi tertinggi diperoleh untuk judul dan isi berita, yaitu 79,18% dan 92,80%. Penggunaan metode BiLSTM dapat secara efektif mendeteksi berita palsu dari segi judul, khususnya konten berita.

2. Widi Afandi, Satria Nur Saputro, Andini Mulia Kusumaningrum, Hikari Ardiansyah, Muhammad Hilmi Kafabi, Sudianto Jurnal Informatika: Jurnal pengembangan IT (JPIT), 2022. Natural Language Processing (NLP) Klasifikasi Judul Berita Clickbait menggunakan RNN-LSTM Penelitian ini membuat model untuk mengklasifikasi berita clickbait pada media elektronik. Model dilatih untuk mendeteksi judul berita yang bersifat sensasional atau tidak sesuai dengan isi berita. Model klasifikasi dilatih menggunakan arsitektur RNN-LSTM. Berdasarkan hasil pengujian, diperoleh akurasi sebesar 77%, sehingga dapat digunakan untuk mengklasifikasi judul berita clickbait dan non- clickbait .

3. Dian Sukma Hani, Chanifah Indah Ratnasari Jurnal Media Informatika Budidarma, 2023. Natural Language Processing (NLP) Klasifikasi Masalah Pada Komunitas Marah-Marah di Twitter Menggunakan Long Short-Term Memory Penelitian ini membuat model untuk mengklasifikasi tweet dalam sebuah komunitas media sosial twitter yaitu komunitas marah-marah. Model dilatih untuk mengklasifikasi jenis permasalahan yang dibahas dalam tweet tersebut. Pengklasifikasian dilakukan menggunakan algoritma LSTM dengan enam kategori yaitu studi, percintaan, keluarga, karier/pekerjaan, personal, dan umpatan. Berdasarkan hasil pengujian, model memperoleh akurasi sebesar 91,94%, sehingga model LSTM tersebut efektif dalam mengklasifikasikan tweet komunitas marah-marah.

4. Muhammad Navi Nugraha, Muhammad Arrofiq Journal of Internet and Software Engineering (JISE), 2024. Natural Language Processing (NLP) Purwarupa Sistem Klasifikasi Legalitas Investasi Berbasis Algoritma Bidirectional Long Short Term Memory Penelitian ini membuat aplikasi sederhana berbasis website yang diintegrasikan dengan model machine learning untuk mengklasifikasi pesan iklan investasi. Model dilatih menggunakan dataset hasil scraping dari grup media sosial Telegram. Pengklasifikasian dilakukan menggunakan algoritma Bidirectional LSTM dengan

dua kategori yaitu legal dan ilegal. Berdasarkan hasil pengujian, didapatkan akurasi terbesar model yaitu 96%, sehingga model tersebut dapat digunakan untuk mengklasifikasikan pesan investasi legal dan ilegal. 5 5. Steven Dharmawan, Viny Christanti Mawardi, Novario Jaya Perdana Jurnal Ilmu Komputer dan Sistem Informasi (JIKSI), 2023. Natural Language Processing (NLP) Klasifikasi Ujaran Kebencian Menggunakan Metode FeedForward Neural Network (IndoBERT) Penelitian ini mengembangkan aplikasi sederhana berbasis website yang diintegrasikan dengan model machine learning untuk melakukan mengklasifikasikan sebuah komentar. Model dilatih menggunakan dataset hasil scraping dari twitter dan youtube. Klasifikasi dilakukan menggunakan metode FeedForward Neural Network dan IndoBERT dengan dua kategori yaitu normal dan ujaran kebencian. Berdasarkan hasil 2.2 Tinjauan Teoritis Tinjauan teoritis berfungsi sebagai fondasi dalam mendukung teori yang relevan dengan topik pada penelitian ini. Penjelasan rinci mengenai setiap tinjauan teoritis yang diaplikasikan pada penelitian ini, diuraikan sebagai berikut. 2.2.1 Artificial Intelligence Artificial Intelligence (AI) merupakan salah satu bidang dalam ilmu komputer yang didedikasikan untuk pengembangan mesin yang mampu meniru tugas-tugas yang biasanya memerlukan kemampuan kognitif manusia. Penciptaan sistem yang memanfaatkan AI bertujuan untuk menciptakan mesin yang dapat belajar, bernalar, dan melakukan tindakan dengan cara yang mirip dengan manusia. Dalam domain AI yang luas, terdapat berbagai macam teknik dan metodologi yang dapat digunakan, termasuk pemrosesan bahasa alami. Pemrosesan bahasa alami memungkinkan mesin untuk memahami seluk-beluk kontekstual bahasa dan memahami makna secara spesifik kata-kata dalam kalimat. 2.2.2 Deep Learning Deep learning adalah algoritma jaringan saraf tiruan yang memanfaatkan data sebagai masukan dan memprosesnya dengan memanfaatkan beberapa lapisan tersembunyi (Zakiyamani, Cahyani, Riana, & Hardianti, 2022). Deep learning memanfaatkan pendekatan pembelajaran yang terinspirasi oleh struktur otak manusia, sehingga memungkinkan komputer untuk belajar dari data yang sangat besar dan kompleks. Pendekatan pembelajaran tersebut terdiri dari berbagai macam

algoritma yang dirancang agar dapat mudah beradaptasi dan dapat diterapkan dalam berbagai skenario pembelajaran. Algoritma ini juga diterapkan pada tugas-tugas seperti pengenalan wajah, gambar, suara, klasifikasi teks, dan lain sebagainya.

2.2.3 Natural Language Processing

Natural Language Processing (NLP) adalah bidang studi yang berfokus pada penelitian dan pengembangan metode komputer untuk memahami dan memproses bahasa manusia, baik dalam bentuk teks atau ucapan (Jesse & Belinkov, 2019). Tahapan dalam NLP melibatkan langkah pra-pemrosesan untuk mengubah data teks atau ucapan menjadi informasi yang dapat dipahami oleh komputer. Langkah-langkah umum tersebut mencakup memecah teks menjadi kata, membersihkan data dari noise, menghilangkan kata stopwords, dan mengubahnya kata ke bentuk dasar. Setelah tahap pra-pemrosesan, data tersebut siap untuk dianalisis lebih lanjut menggunakan berbagai teknik NLP, seperti analisis sentimen, klasifikasi teks, dan text-to-speech.

2.2.4 Long Short-Term Memory (LSTM)

Menurut (Pratama, Wijaya, Santosa, & Tamba, 2023), Long Short-Term Memory (LSTM) merupakan jenis khusus dari jaringan saraf tiruan (neural network) yang dirancang untuk mengatasi masalah pemodelan dan prediksi pada data yang memiliki urutan atau pola tertentu. LSTM mampu mengingat informasi lampau yang muncul sebelumnya, sehingga memungkinkan prediksi yang lebih akurat. Kemampuan mengingat informasi dalam jangka panjang, membuat LSTM digunakan di berbagai tugas yang membutuhkan pemahaman konteks, seperti analisis sentimen dan klasifikasi teks. Arsitektur LSTM memungkinkan adanya proses penambahan atau penghapusan kumpulan informasi pada memori cell state atau kondisi sel dengan struktur yang disebut dengan gates atau gerbang. Gerbang berfungsi sebagai komponen yang menjaga informasi dan mengontrol aliran data di dalam jaringan.

24 Gerbang yang terdiri dari tiga komponen inti yaitu input gate, forget gate, dan output gate terdapat pada setiap unit. Adapun penjelasan mengenai keterkaitan setiap komponen pada gerbang dan kondisi sel, dijabarkan sebagai berikut:

Gambar 2.1 Arsitektur LSTM

a) Input Gate Gerbang pertama dalam proses LSTM adalah input gate (i_t) yang berperan untuk menerima informasi lampau dari sel pada waktu

sebelumnya dan informasi terbaru yang berasal dari masukkan pada keadaan saat ini atau waktu t . Informasi tersebut selanjutnya digabung, lalu diproses oleh lapisan sigmoid dan tanh. Lapisan sigmoid, yang juga dikenal dengan sebutan "input gate layer" berperan untuk memilah informasi mana yang akan diperbarui dengan mengubah nilai menjadi 0 hingga 1. Selanjutnya, lapisan tanh akan membuat vektor untuk menampung kandidat nilai baru atau \tilde{C}_t ke dalam kondisi sel yang hasilnya berupa nilai -1 hingga 1. Langkah selanjutnya yaitu menggabungkan hasil dari masing-masing lapisan untuk memperbarui kondisi sel. Adapun persamaan input gate, sebagai berikut:

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

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (2)$$

Dimana pada i_t atau input gate, W_i adalah bobot dari input gate pada waktu saat ini, h_{t-1} adalah luaran dari timestamp sebelumnya, x_t adalah masukkan dari timestamp saat ini, b_i adalah nilai bias pada input gate, σ adalah lapisan sigmoid, \tilde{C}_t adalah kandidat nilai baru dari timestamp saat ini, tanh adalah lapisan tanh, W_c adalah berat untuk kandidat \tilde{C}_t , dan b_c adalah nilai bias untuk kandidat \tilde{C}_t .

b) Forget Gate Gerbang kedua dalam proses LSTM adalah forget gate (f_t), atau gerbang lupa. Gerbang ini berfungsi untuk menentukan bobot yang diberikan pada informasi masa lalu dalam kondisi sel. Bobot ini yang akan menentukan informasi mana yang akan dipertahankan dan dihapus, sehingga LSTM dapat fokus pada informasi yang relevan untuk prediksi saat ini atau waktu t . Forget gate terdiri dari lapisan sigmoid yang berperan untuk menyaring informasi dengan memberikan luaran berupa 0 dan 1. Pada prosesnya, forget gate memerlukan informasi berupa hidden state yang berasal dari cell pada waktu sebelumnya dan informasi saat ini. Setelah itu menerapkan fungsi aktivasi sigmoid. Jika $f_t = 0$ maka informasi pada keadaan sel sebelumnya akan dibuang, sementara jika $f_t = 1$ maka informasi akan disimpan. Berikut adalah persamaan dari forget gate.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3)$$

Dimana p

ada f_t atau forget gate, W_f adalah bobot dari forget gate pada waktu saat ini, h_{t-1} adalah luaran berupa vektor dari timestep sebelumnya, x_t adalah masukan dari timestep saat ini, dan b_f adalah nilai bias pada forget gate. c) Cell State Komponen cell state pada jaringan LSTM merupakan komponen vital yang berfungsi sebagai penyimpanan informasi jangka panjang. Komponen ini berperan untuk memperbarui kondisi sel pada waktu sebelumnya atau C_{t-1} menjadi kondisi sel baru pada waktu saat ini yang ditandai dengan C_t .

Luaran yang dihasilkan oleh jaringan LSTM sebelumnya atau h_{t-1} akan dikalikan dengan f_t . Tujuannya adalah agar cell state hanya mengingat informasi yang dianggap penting dan melupakan informasi lainnya. Setelah itu, hasilnya akan dijumlahkan dengan hasil perkalian antara i_t dengan \tilde{C}_t . Adapun persamaan untuk proses pembaruan kondisi sel sebagai berikut. $C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$ (4) Dimana

na pada C_t atau cell state, f_t adalah nilai forget gate pada waktu saat ini, C_{t-1} adalah kondisi sel pada timestamp sebelumnya atau $t-1$, i_t adalah nilai input gate pada timestamp saat ini atau t , dan \tilde{C}_t adalah kandidat cell state baru. d) Output

Gate Pada gerbang output gate (o_t), terjadi proses penentuan luaran berupa informasi yang akan dibagikan. Langkah pertama, informasi akan diproses oleh lapisan sigmoid untuk menentukan informasi apa yang akan kita jadikan sebagai luaran. Selanjutnya informasi tersebut diproses oleh lapisan tanh yang akan mengubah informasi tersebut menjadi rentang antara -1 dan 1. Hasil tersebut akan menentukan seberapa penting informasi pada waktu saat ini untuk diteruskan ke sel LSTM berikutnya. **4** Berikut

adalah persamaan output gate. $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$ (5) $h_t = o_t * \tanh(C_t)$ (6) Dimana

o_t atau output gate, σ adalah fungsi aktivasi lapisan sigmoid, W_o adalah bobot pada output gate, h_{t-1} adalah keadaan tersembunyi pada timestamp sebelumnya atau $t-1$, x_t adalah masukan dari stempel waktu saat ini atau t , b_o adalah nilai bias dari output gate.

Sementara h_t atau keadaan tersembunyi pada stempel waktu saat ini, \tanh adalah fungsi aktivasi hyperbolic tangen, dan C_t adalah kondisi sel yang baru pada stempel waktu saat ini atau t . 2.2.5

Bidirectional Long Short-Term Memory (BiLSTM) BiLSTM (Bidirectional Long Short-Term Memory) dikembangkan dari arsitektur Long Short-Term Memory (LSTM) yang dirancang untuk meningkatkan kemampuan dalam memahami konteks kata dalam sebuah kalimat.

21 Berbeda dengan LSTM yang hanya memproses informasi secara backward, BiLSTM memproses informasi secara dua arah, yaitu forward dan backward. Pemrosesan secara dua arah ini memungkinkan

model untuk menangkap konteks dari keadaan sebelumnya dan yang akan datang dalam urutan, sehingga memberikan pemahaman yang lebih menyeluruh terhadap teks (Airlangga, 2024). Dengan kemampuannya, BiLSTM telah banyak

digunakan dalam berbagai aplikasi pemrosesan bahasa alami seperti klasifikasi teks, ringkasan teks, terjemahan mesin, dan lainnya. Gambar 2.2 merupakan ilustrasi arsitektur BiLSTM yang di dalamnya terdapat tiga unit LSTM dengan timestep yang terdiri dari waktu lampau atau ($t-1$), saat ini (t), dan waktu berikutnya ($t+1$).

Lapisan forward pada BiLSTM memproses informasi dari awal kalimat ke akhir dengan memasukkan dari timestep sebelumnya dan timestep saat ini. Lapisan backward memproses informasi dari akhir kalimat ke awal dengan memasukkan yang diperoleh dari timestep saat ini dan timestep berikutnya. Pada dasarnya luaran yang dihasilkan BiLSTM adalah penjumlahan antara nilai probabilitas output pada timestep sebelumnya atau $t-1$ dengan nilai probabilitas output pada timestep setelahnya atau $t+1$.

Adapun persamaan BiLSTM adalah sebagai berikut: $y_t = \hat{h}_t + h_t$ (7) Dimana y_t

adalah output BiLSTM, \hat{h}_t adalah output yang dihasilkan dari lapisan forward, dan h_t adalah output yang dihasilkan dari lapisan

backward. 2.2.6 Text Classification Text classification merupakan salah satu teknik yang digunakan dalam penerapan pemrosesan bahasa alami.

Menurut (Miranda, Gabriella, Wahyudi, & Chai, 2023), text classification adalah proses menandai teks dengan label kategori tertentu dari sekumpulan

data yang telah ditetapkan sebelumnya. Teknik ini memungkinkan pengelompokan teks ke dalam kategori-kategori yang relevan, sehingga informasi yang memiliki makna tertentu dapat diekstraksi dari kumpulan teks yang besar. Dalam penerapannya, teknik ini dapat digunakan untuk mem filter email spam , mengklasifikasi sentimen ulasan produk menjadi positif atau negatif, dan membuat bot yang dapat menjawab pertanyaan sesuai dengan keyword tertentu. Adapun tools yang digunakan dalam merancang sistem pada penelitian ini, yaitu: a) Python Python merupakan bahasa pemrograman canggih yang dikenal karena kemudahan mempelajarinya. Diciptakan oleh Guido van Rossum pada akhir tahun 1980-an, Python berhasil menjadi pilihan utama pada pengembangan perangkat lunak, termasuk pengembangan web, kecerdasan buatan, dan otomasi sistem. Python dirancang dengan tujuan khusus agar kode dan sintaks dapat dibaca secara sederhana, membuatnya ideal untuk pemula maupun programmer berpengalaman. b) Google Colaboratory Google Colaboratory merupakan platform pengembangan berbasis cloud yang diciptakan Google. Platform cloud ini dirancang khusus untuk memenuhi kebutuhan pemrograman, analisis data, dan pembelajaran mesin. Keunggulan utama platform ini adalah dapat menjalankan kode Python secara interaktif. Pengguna dapat memanfaatkan platform ini untuk menjalankan kode Python pada environment Jupyter Notebook yang dapat diakses secara lokal melalui browser. c) NLTK Natural Language Toolkit (NLTK) adalah salah satu pustaka yang dirancang untuk memfasilitasi pengguna dalam mengolah data berupa teks. NLTK menyediakan berbagai macam fungsi yang memfasilitasi pengguna untuk melakukan beragam tugas pemrosesan bahasa alami. Fungsi-fungsi tersebut mencakup tokenisasi kalimat, mengubah kata menjadi bentuk dasarnya, dan mengubah teks menjadi format yang dapat memudahkan pemrosesan. NLTK dapat menjadi pilihan yang tepat bagi para peneliti, pengembang, dan praktisi yang ingin menjelajahi dunia pemrosesan bahasa alami. d) Keras Gambar 2. 2 Arsitektur Bidirectional LSTM Keras adalah salah satu Application Programming Interface (API) yang dibuat untuk menyederhanakan proses pengembangan model deep learning menggunakan

Python. Keras berperan sebagai jembatan antara pengguna dan library deep learning sehingga menjadi lebih mudah dan sederhana. Library yang disediakan oleh Keras antara lain seperti Tensorflow, Theano, dan CNTK.

e) SKLearn Scikit-Learn atau yang biasa disingkat SKLearn adalah library pada bahasa pemrograman Python yang dimanfaatkan untuk pengembangan machine learning . Library ini menyediakan berbagai fitur untuk analisis dan pemodelan data, termasuk algoritma untuk klasifikasi, regresi, clustering , dan algoritma serupa. Fitur SKLearn yang digunakan oleh peneliti yaitu fungsi untuk memisahkan data latih dan data uji pada dataset. 2.2.7

Black Box Metode pengujian black box merupakan pengujian kualitas perangkat lunak yang difokuskan pada aspek fungsionalitas dari perangkat lunak tersebut (Wijaya & Astuti, 2021). Pengujian black box dilakukan dengan maksud untuk menemukan fungsi yang belum berjalan sesuai dengan tujuan, kesalahan pada interface , masalah dalam struktur data, kendala kinerja, kesalahan pada proses awal hingga akhir. Pada dasarnya, pengujian ini memastikan bahwa semua fitur pada perangkat lunak bekerja sesuai dengan tujuan dan ketentuan awal. 2.2.8 White Box Metode pengujian menggunakan white-box testing berfokus pada analisis struktur internal kode program. Berbeda dengan pengujian black box , metode ini berfokus pada pengujian jalur eksekusi, percabangan kondisional, dan logika dalam kode untuk mengidentifikasi kesalahan atau kegagalan yang dapat terjadi. Penguji yang menerapkan white box sebagai metode pengujian perangkat lunak harus mengetahui dan memahami kode sumber perangkat lunak tersebut secara mendalam (Sie, Musdar, & Bahri, 2022). BAB III TAHAPAN PELAKSANAAN Pada

sub bab ini, tahapan pelaksanaan yang terdiri dari serangkaian langkah dijelaskan oleh penulis secara rinci. Setiap langkah telah diuraikan untuk memberikan wawasan mengenai proses implementasi. 3.1 Langkah-langkah Pelaksanaan Langkah pelaksanaan yang terdapat pada penelitian ini memaparkan secara rinci bagaimana tugas akhir diselesaikan. Dalam penelitian ini, diagram alir digunakan sebagai panduan untuk menyelesaikan langkah – langkah pelaksanaan yang tersegmentasi menjadi beberapa tahapan. Gambar 3. 1

Tahapan Pelaksanaan Adapun penjelasan detail mengenai setiap tahapan pelaksanaan yang terdapat pada gambar di atas, diuraikan sebagai berikut:

a) Identifikasi Masalah Langkah pertama pada penelitian ini yaitu mengidentifikasi masalah yang kemudian dirumuskan untuk memudahkan pencarian solusi. b) Tinjauan Pustaka Tujuan tahapan tinjauan pustaka adalah untuk mengumpulkan dan meninjau penelitian terkait dengan topik yang menjadi fokus. Dengan meninjau penelitian sebelumnya, peneliti dapat menemukan konsep dan teori fundamental yang relevan dengan topik penelitian. c) Analisis Kebutuhan Analisis kebutuhan adalah tahap di mana peneliti mengidentifikasi, mendokumentasikan, dan menganalisis kebutuhan yang terkait dengan penelitian. Langkah ini diawali dengan mengumpulkan informasi, data, dan spesifikasi yang diperlukan sesuai dengan metodologi dan tujuan penelitian. Tujuannya agar peneliti dapat memastikan bahwa penelitian dirancang dan dilaksanakan dengan cara yang optimal, sehingga menghasilkan luaran yang selaras dengan tujuan penelitian. d) Pengolahan Data Pengolahan data adalah tahap yang penting dalam berbagai bidang, khususnya pada penelitian. Pada tahap ini, data mentah yang dikumpulkan diubah menjadi informasi yang dapat dipahami, dianalisis, dan dimanfaatkan untuk berbagai keperluan. Tahap ini mencakup pembersihan data untuk menghilangkan inkonsistensi dan duplikasi, serta perubahan format data agar dapat diolah sesuai dengan algoritma atau metodologi yang digunakan. e) Perancangan dan Pengujian Model NLP Perancangan dan pengujian model NLP adalah bagian utama dalam penelitian ini. Model NLP yang akan dirancang akan berperan penting terhadap masalah yang diangkat, yaitu mengklasifikasi genre film berdasarkan sinopsis. Perancangan model menggunakan arsitektur Bidirectional LSTM dan memanfaatkan framework serta library yang tersedia sebagai alat, dan akan disimpan dengan format keras. Sementara pengujian model akan menggunakan metrik evaluasi seperti nilai akurasi. f) Implementasi Model ke dalam Sistem Model yang telah dilatih dan diuji kemudian diintegrasikan ke dalam sistem yang telah dirancang agar dapat digunakan oleh pengguna. Sistem yang akan dibuat dapat berupa aplikasi

berbasis desktop atau website. Tahap ini melibatkan beberapa pengembangan yaitu pada sisi antarmuka dan sisi logika pengolahan data. g) Pengujian Sistem Sistem yang telah dirancang dan diintegrasikan akan diuji untuk meminimalkan kesalahan, baik dari sisi antarmuka pengguna maupun bug pada program.

6 27

Pengujian akan dilakukan dengan memanfaatkan dua metode yang umumnya digunakan yaitu black box dan white box . Metode black box berfungsi untuk

mengevaluasi seberapa baik aplikasi bekerja sesuai dengan fungsinya.

Sementara itu, metode white box memastikan bahwa kode yang ditulis

berjalan dengan baik. h) Penulisan Laporan Penulisan laporan merupakan

tahap mendokumentasikan setiap rangkaian yang telah dilakukan secara

tertulis. Tahap ini dilakukan setelah seluruh rangkaian kegiatan selesai,

dengan tujuan untuk mendokumentasikan seluruh proses penelitian, temuan, dan

kesimpulan yang diperoleh. 3.2 Metode Pengujian Sistem klasifikasi genre

film yang telah dirancang dan dikembangkan akan diuji menggunakan dua

metode, yaitu black box dan white box . Penjelasan rinci mengenai

setiap metode pengujian tersebut, diuraikan sebagai berikut. 3.2.1 Black

Box Metode Black Box merupakan rancangan pengujian yang lebih berfokus

untuk mengevaluasi fungsionalitas sistem agar sesuai dengan kebutuhan

pengguna. Pada penelitian ini, metode tersebut diterapkan untuk menguji

luaran pada setiap fungsi dalam mengklasifikasikan genre film yang telah

diintegrasikan dengan model klasifikasi menggunakan algoritma Bidirectional

Long Short-Term Memory . 3.2.2 White Box Metode White Box merupakan

rancangan pengujian yang berperan untuk menilai apakah proses pada sistem

telah sesuai dengan rancangan yang telah ditetapkan sebelumnya. Metode ini

berfokus pada pemeriksaan internal sistem, dengan cara mengamati kode

program yang telah dirancang untuk setiap logika sistem klasifikasi genre

film. Tujuannya adalah untuk memverifikasi bahwa sistem berfungsi sesuai

dengan spesifikasi yang telah ditetapkan sebelumnya. 10 BAB IV PERANCANGAN Pada

bab ini akan disajikan penelitian yang telah dilaksanakan oleh peneliti

sebelumnya, spesifikasi untuk kebutuhan sistem baru, perancangan sistem dan

antarmuka, serta langkah-langkah perancangan sistem secara rinci. 14 4.1 Analisis

Sistem Terdahulu Analisis mendetail terhadap sistem terdahulu dilakukan untuk mengidentifikasi kekurangan, kelemahan, serta menghindari terjadinya duplikasi dengan penelitian sebelumnya. Peneliti secara aktif mengobservasi sistem yang telah ada untuk memahami kebutuhan dalam mengembangkan sistem baru. Analisis yang dilakukan terhadap sistem sebelumnya dapat menjadi bahan pertimbangan dalam menambahkan fitur-fitur tertentu yang dapat menjadi peningkatan pada sistem yang akan dibuat. Pada penelitian yang dilakukan oleh Nugraha dan Arrofiq, aplikasi website sederhana yang dibuat berfungsi untuk melakukan klasifikasi teks berupa analisis sentimen terhadap pesan investasi. Metode pengumpulan data yang digunakan sebagai data pelatihan diperoleh menggunakan teknik scrapping pada grup Telegram. 20 Tahapan yang dilakukan dalam membangun model analisis sentimen tersebut meliputi case folding , filtering , tokenisasi, stemming , dan stopwords removal . Aplikasi klasifikasi teks yang dibangun hanya memiliki form input dan hasil klasifikasi yang terbagi menjadi dua kelas, yaitu legal dan ilegal.

4.2 Spesifikasi

Kebutuhan Sistem Baru Perumusan spesifikasi kebutuhan sistem baru dilakukan setelah tahapan analisis mendalam terhadap sistem terdahulu agar dapat menemukan spesifikasi pada sistem lama. 12 Diperlukan suatu sistem baru untuk merancang dan membuat sistem klasifikasi genre film, kebutuhan tersebut dibagi menjadi dua, yaitu spesifikasi perangkat lunak (software) dan perangkat keras (hardware). Berikut adalah penjelasan masing-masing bagian tersebut.

4.2.1 Spesifikasi Perangkat Lunak (Software)

Pada penelitian ini, penggunaan perangkat lunak dibutuhkan sebagai media untuk membantu proses perancangan desain sistem hingga penulisan kode program. Perangkat lunak yang digunakan berupa sistem operasi dan development environment . Adapun rincian spesifikasi perangkat lunak tersebut, sebagai berikut.

No.	Perangkat Lunak	Keterangan
1.	Windows 10 (32 bit)	Batas minimal sistem operasi yang harus dipenuhi oleh pengguna
2.	Visual Studio Code	Code editor yang digunakan untuk mengembangkan aplikasi utama dalam penelitian ini
3.	Google Colaboratory Platform	berbasis cloud yang digunakan untuk pengembangan dan pelatihan model
4.	Kaggle	

Platform yang digunakan untuk memperoleh dataset Tabel 4.1 menunjukkan daftar perangkat lunak yang digunakan oleh peneliti untuk merancang dan mengembangkan sistem klasifikasi genre film. Berikut adalah penjelasan mengenai fungsi setiap perangkat lunak secara rinci sebagai berikut. a) Windows 10 Windows 10 merupakan sistem operasi yang digunakan sebagai wadah yang memastikan kompatibilitas dan kelancaran penggunaan setiap perangkat lunak. Sistem operasi memungkinkan perangkat lunak seperti Kaggle , Google Colaboratory , Jupyter Notebook , Visual Studio Code , dan alat-alat lainnya dapat dijalankan untuk pengembangan dan pembuatan model klasifikasi genre film. b) Visual Studio Code Visual Studio Code atau disingkat VSCode adalah editor kode sumber yang ringan serta dilengkapi berbagai macam fitur yang dapat memudahkan proses pengembangan dan pembuatan perangkat lunak. Peneliti menggunakan kode editor ini untuk mengintegrasikan model klasifikasi genre film dengan perangkat lunak yang terdiri dari logika program untuk dapat memproses masukan dari pengguna, sehingga mengeluarkan luaran yang sesuai. c) Google Colaboratory Google Colaboratory digunakan sebagai platform untuk menulis dan menjalankan kode Python, khususnya dalam pengembangan model Natural Language Processing untuk sistem klasifikasi genre film pada penelitian ini. d) Kaggle Kaggle merupakan platform online yang dirancang khusus untuk mengembangkan model machine learning . Salah satu keunggulan platform ini adalah menyediakan berbagai macam dataset yang mencakup beragam domain machine learning , salah satunya yaitu dataset film yang digunakan pada penelitian ini untuk melatih model.

4.2.2 Spesifikasi Perangkat Keras (Hardware)

Perangkat keras berperan sebagai basis fisik untuk menjalankan perangkat lunak dalam proses perancangan sistem. Dalam penelitian ini, perangkat keras yang menjadi faktor pendukung meliputi processor , storage , dan memory . Tabel berikut merinci spesifikasi perangkat keras yang digunakan oleh peneliti, termasuk spesifikasi minimum yang diperlukan. Tabel 4. 2 Spesifikasi Perangkat Keras Minimum

No.	Perangkat Keras	Keterangan
1.	Processor	1 Ghz atau lebih
2.	Storage	128 GB Hard Disk/SSD
3.		

Memory 4 GB RAM Tabel berikut menampilkan spesifikasi perangkat keras yang digunakan peneliti dalam merancang dan membangun sistem klasifikasi genre film. Tabel 4. 3 Spesifikasi Perangkat Keras Peneliti No.

Perangkat Keras Keterangan

1. Processor 2.6 Ghz
2. Storage 512 GB SSD
3. Memory 8 GB RAM

a) Processor Processor berperan sebagai inti atau pusat pengendali dalam sebuah sistem dan terletak pada perangkat keras. Ini berfungsi untuk dapat melakukan berbagai operasi perhitungan dan pengolahan data yang diperlukan, termasuk mengeskusi instruksi-instruksi yang terdapat dalam perangkat lunak dan telah terintegrasi dengan model yang menggunakan algoritma BiLSTM. b) Storage Storage atau penyimpanan digunakan sebagai tempat untuk menyimpan data yang diperlukan oleh sistem klasifikasi genre film, seperti model BiLSTM, dataset pelatihan, serta kode sumber perangkat lunak yang telah dibuat sebelumnya. c) Memory Memory memiliki peran untuk menyimpan data dan program sementara yang sedang berjalan saat ini. Perangkat keras ini juga digunakan sebagai penyimpanan model BiLSTM selama perangkat lunak sedang digunakan secara realtime .

4.3 Identifikasi Keyword

Tahap identifikasi keyword atau kata kunci dilakukan untuk mendukung model dalam melakukan klasifikasi genre terhadap sinopsis. Kata kunci diidentifikasi pada setiap sinopsis di setiap genre. Pengidentifikasian kata kunci dilakukan dengan cara manual dengan menemukan kata yang unik dan terkait dengan suatu genre. Kata kunci tersebut tidak boleh terkandung dalam lebih dari satu genre, agar menghindari adanya kekeliruan dalam proses klasifikasi.

4.4 Pembuatan Model

Proses perancangan sistem klasifikasi genre film, yang bertujuan untuk menentukan kategori film berdasarkan sinopsisnya, melibatkan beberapa tahapan yang diilustrasikan pada gambar di bawah ini. Gambar 4.1 merupakan tahapan proses pembuatan model pada penelitian ini untuk membuat sistem klasifikasi genre berdasarkan sinopsis menggunakan Algoritma BiLSTM. Tahapan tersebut terdiri dari mengumpulkan dataset, melakukan training dan testing , hingga sistem dapat mengklasifikasi genre film. Adapun penjelasan mengenai setiap tahapan di atas, sebagai berikut.

4.4.1 Dataset Data

yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari situs kaggle.com dengan judul dataset yaitu “ Genre Classification Dataset IMDB . Dataset ini memuat informasi seperti judul, genre, dan sinopsis setiap film yang terdaftar di Internet Movie Database (IMDB) hingga tahun 2020. Dataset ini memiliki jumlah data sebanyak 54.214 data film dan memiliki empat atribut yaitu id, title , genre, serta description . Data dalam dataset tersebut memiliki format comma separated values (CSV) yang memungkinkan untuk diolah dan dianalisis menggunakan machine learning . Pada tahap awal, dilakukan proses penyaringan dataset untuk menyeleksi data film yang memiliki genre action , adventure, animation , comedy , crime , drama , horror , romance , science-fiction , dan thriller . Dari total 54.214 data film, proses penyaringan menghasilkan 29.267 data film yang memenuhi kriteria. Analisis genre film yang tersisa menunjukkan distribusi yang tidak seimbang, dengan genre mendominasi yaitu drama (46.5%), diikuti comedy (25.4%), horror (7.5%), thriller (5.4%), action (4.5%), adventure (2.6%), romance (2.3%), science-fiction (2.2%), crime (1.7%), dan animation (1.7%). Ketidakseimbangan data ini dikhawatirkan dapat berakibat pada hasil klasifikasi yang tidak optimal. Maka dari itu, dilakukan teknik resampling dengan menggunakan library imblearn di Python. Metode resampling yang digunakan adalah upsampling dengan memanfaatkan teknik RandomOverSampler dari library imblearn. Teknik ini bekerja dengan memilih data film secara acak dari dataset dan menggandakannya hingga jumlahnya sama dengan frekuensi data terbanyak, yaitu genre drama. Hal ini bertujuan untuk menyeimbangkan distribusi data dan meningkatkan performa model klasifikasi dalam menangani semua genre film dengan lebih baik.

4.4.2 Pre-processing

Preprocessing merupakan langkah-langkah yang digunakan untuk menyeragamkan format dan membersihkan data dari segala nilai yang tidak memberikan pengaruh besar terhadap proses klasifikasi.

3 Dalam klasifikasi teks, tahap ini memiliki peran penting agar model dapat mengolah data teks yang bersifat sekuensial atau berurutan. Tahapan preprocessing terdiri dari serangkaian proses yang memiliki tugasnya

masing-masing, tahapannya mencakup case folding , cleaning , tokenisasi, menghilangkan stopwords , dan stemming . Upaya ini bertujuan untuk mengurangi noise dan meningkatkan kualitas model saat tahap pelatihan. **31** Berikut adalah diagram alir yang menunjukkan tahapan preprocessing . **31** Gambar 4. **6** **11** **2**

Flowchart Preprocessing Data 1) Case Folding Case folding adalah tahap awal yang dilakukan pada preprocessing yang mentransformasi seluruh huruf besar pada kalimat menjadi huruf kecil atau lowercase . Tahap ini dilakukan dengan tujuan agar data memiliki format yang konsisten. Dengan begitu, data teks yang digunakan sebagai masukan dapat diolah lebih lanjut melewati proses selanjutnya, sehingga menghasilkan analisis yang akurat dan objektif. Proses ini juga membantu dalam mengurangi variasi yang tidak perlu dalam data teks. 2) Cleaning Cleaning adalah tahap membersihkan data dari elemen-elemen yang tidak diperlukan atau dapat berdampak negatif terhadap pelatihan model. Elemen-elemen tersebut mencakup hyperlinks atau url , hashtag , angka, simbol retweet, emoji, dan tanda baca. Tujuan dari tahap ini adalah untuk mengurangi noise pada data, yaitu elemen-elemen yang tidak relevan yang dapat menyamarkan pola dan makna penting dalam teks. Gambar 4. 3 Flowchart proses Cleaning Gambar 4.3 menunjukkan tahapan proses cleaning atau pembersihan data teks. Proses ini diawali dengan melakukan perubahan pada setiap kata teks lowercase atau menjadi huruf kecil. Selanjutnya, kata- kata tersebut digabungkan menjadi kalimat. Dalam kalimat tersebut, segala elemen yang menyebabkan noise seperti simbol atau non-karakter akan dihilangkan. Selain itu, mention , hyperlink atau url , dan hashtag juga dihilangkan, serta spasi yang berlebihan disesuaikan menjadi satu spasi. Langkah terakhir adalah menghilangkan tanda baca yang tidak diperlukan. 3) Tokenisasi Tokenisasi adalah tahap mengubah teks yang terdiri dari kumpulan kata menjadi token yang terdiri dari satu kata. Kata-kata yang telah dijadikan token memiliki indeks yang berbeda-beda sehingga dapat digunakan sebagai kamus bahasa atau kumpulan kosakata. 4) Menghilangkan Stopword Stopword adalah kata-kata umum yang sering ditemukan dalam teks, tetapi kurang memberikan informasi penting

mengenai konteks teks tersebut. Contohnya adalah kata-kata yang termasuk dalam daftar stopwords, mencakup konjungsi atau kata penghubung, pronouns atau kata ganti, dan preposisi atau kata depan serta kata-kata yang sering ditemukan dan dianggap tidak relevan.

5) Stemming Stemming adalah tahap dimana setiap kata dicari root atau kata dasarnya sehingga menghasilkan luaran berupa kata dasar tanpa imbuhan. Tujuan proses stemming adalah untuk mengurangi kata-kata yang memiliki bentuk berbeda tetapi dengan makna yang sama. Dengan begitu, kompleksitas dan ukuran data tidak terlalu besar, sehingga berdampak pada penghematan waktu dan memori. Pada penelitian ini, proses stemming diimplementasikan menggunakan bantuan salah satu library Python yaitu PorterStemmer dari NLTK.

4.4.3 Split Data

Split data adalah tahap pemisahan dataset yang telah melewati proses preprocessing beberapa bagian atau subset. Jumlah data setelah dilakukan proses resampling terkumpul sebanyak 136.130 data. Data tersebut terbagi menjadi 13.613 data karena terdapat 10 genre yang akan menjadi kategori klasifikasi. Secara umum, pembagian data diterapkan sehingga menghasilkan dua set data, yaitu data pelatihan (training set) dan data pengujian (testing set). Penjelasan kedua bagian tersebut dijelaskan sebagai berikut:

a. Data Training / Training Set Data training akan digunakan oleh model sebagai bahan pelatihan agar dapat mengenali pola untuk dapat mengklasifikasi genre film berdasarkan sinopsisnya. Dalam penelitian ini, 80% dari total populasi data, yaitu sebanyak 108.904 data, digunakan sebagai data pelatihan. Data tersebut terbagi secara merata ke dalam 10 genre, dengan masing-masing genre memiliki 10.890 data.

b. Data Testing / Testing Set Data testing akan digunakan untuk menguji performa model yang telah melewati proses pelatihan, dengan data yang belum pernah dilihat sebelumnya. Dalam penelitian ini, 20% dari total populasi data, yaitu sebanyak 27.226 data, digunakan sebagai data pengujian. Data tersebut terbagi secara merata ke dalam 10 genre, dengan masing-masing genre memiliki 2.722 data.

4.4.4 Rancangan Lapisan Model

Pembuatan model dilakukan dengan mendefinisikan setiap layer atau

lapisan yang digunakan untuk dapat membangun model sistem klasifikasi genre film. Pembuatan model pada penelitian ini akan menggunakan lapisan BiLSTM sebagai salah satu lapisan utama yang akan menjadi fondasi untuk model dalam melakukan pembelajaran. Lapisan BiLSTM membantu model dalam menangkap informasi secara dua arah dari sinopsis film, sehingga dapat memahami konteks dengan lebih baik untuk klasifikasi genre. Gambar 4.4 menunjukkan lapisan yang digunakan peneliti dalam membuat model sistem klasifikasi genre film. Adapun penjelasan mengenai setiap lapisan tersebut, dijabarkan sebagai berikut:

a) Embedded Layer Embedding layer adalah lapisan yang mengubah input data yang telah melewati proses tokenisasi menjadi representasi vektor dengan dimensi yang lebih rendah. Lapisan ini berfungsi untuk memetakan setiap kata (atau token) dalam teks menjadi vektor numerik yang memiliki dimensi bersifat statis atau tetap. Tujuannya adalah untuk menangkap hubungan antar kata dan mengurangi dimensi input sehingga memudahkan pemrosesan data teks oleh model.

b) BiLSTM Layer Lapisan BiLSTM akan mengklasifikasi genre film berdasarkan input yang diterima dari lapisan embedded . **22** Lapisan ini tersusun dari dua unit LSTM, yang masing-masing memiliki tiga jenis gerbang yaitu forget gate , input gate , dan output gate . Unit-unit ini dibagi menjadi dua bagian, bagian pertama disebut keadaan maju, yang memproses informasi dalam urutan kronologis dari awal hingga akhir, sementara bagian kedua disebut keadaan mundur, yang memproses informasi dalam urutan terbalik dari akhir hingga awal. Sebanyak 32 unit digunakan sebagai lapisan BiLSTM. Unit-unit tersebut berfungsi sebagai ukuran atau jumlah hidden state LSTM pada lapisan tersembunyi dalam mengolah informasi.

c) GlobalMaxPooling1D Lapisan Global Max Pooling 1D berperan untuk mengambil nilai maksimum dari setiap fitur dalam satu dimensi. Lapisan ini berfungsi untuk mengurangi dimensi data dengan mengambil nilai maksimum dari fitur tertentu, sehingga menghasilkan vektor dengan dimensi yang lebih rendah. Ini akan membantu model dalam mengurangi kompleksitas dan mencegah terjadinya overfitting . Tujuannya adalah untuk mengekstrak fitur yang paling menonjol dalam data

dan mengurangi ukurannya tanpa kehilangan informasi penting. d) Dense Layer Dense Layer atau biasa disebut dengan fully connected layer yang membuat setiap unit (neuron) dalam lapisan tersebut terhubung dengan setiap unit di lapisan sebelumnya. Lapisan ini berfungsi untuk melakukan kombinasi linier dari input dan menerapkan fungsi aktivasi softmax untuk menghasilkan output yang digunakan pada tugas klasifikasi. Tujuannya adalah untuk melakukan transformasi non-linear pada data yang diproses serta untuk menggabungkan informasi yang telah dipelajari oleh lapisan-lapisan sebelumnya dalam model untuk menghasilkan prediksi akhir, dalam hal ini adalah genre.

4.5 Perancangan Sistem

Perancangan sistem merupakan tahap dimana peneliti menyusun struktur dan komponen utama sistem klasifikasi genre film menggunakan pemodelan Unified Modelling Language (UML). Pemodelan tersebut mencakup flowchart , use case diagram , dan activity diagram

4.5.1 Flowchart Proses Bisnis Aplikasi

Gambar 4. 5 Flowchart Penggunaan Aplikasi Gambar 4.5 merupakan alur sistem klasifikasi genre berdasarkan synopsis yang dibuat oleh peneliti. Tahap yang dilakukan user setelah mengakses halaman awal adalah menentukan banyaknya form input sebagai tempat untuk memasukkan sinopsis film yang ingin diklasifikasi. Setelah itu, sistem akan menampilkan form input sesuai dengan jumlah yang telah ditentukan, lalu user dapat memasukkan sinopsis film pada setiap form input . Tahap selanjutnya adalah sistem melakukan klasifikasi hingga memunculkan hasil klasifikasi sinopsis film tersebut dan menyimpan hasil klasifikasi ke dalam session untuk digunakan sebagai riwayat.

4.5.2 Use Case Diagram

Gambar 4. 6 Use Case Diagram Sistem Klasifikasi Genre Film Gambar 4.6 menunjukkan rancangan use case diagram untuk sistem klasifikasi genre film yang dikembangkan oleh peneliti. Pada diagram tersebut user berperan sebagai aktor yang dapat memasukkan teks sinopsis, melihat hasil klasifikasi, dan melihat riwayat klasifikasi. Adapun penjelasan detail mengenai setiap use case dijabarkan pada tabel scenario di bawah ini.

Use Case	Deskripsi User
4.4 Skenario Use Case Memasukkan Sinopsis Nama	User memasukkan teks
Use Case Memasukkan Sinopsis Aktor	User memasukkan teks

sinopsis ke aplikasi. Tahapan 1. User masuk ke tampilan awal. 2. User menentukan jumlah form input. 3. User memasukkan sejumlah teks sinopsis sesuai jumlah form input. Tabel 4.4 menunjukkan tahapan use case ketika user memasukkan sinopsis ke dalam aplikasi. User dapat menentukan jumlah input berdasarkan jumlah sinopsis yang ingin diklasifikasi. Tabel 4.5 Skenario Use Case Melihat Hasil Klasifikasi Nama Use Case Melihat Hasil Klasifikasi Aktor User Deskripsi User dapat melihat hasil yang disajikan oleh sistem. Tahapan 1. User memasukkan teks sinopsis ke dalam form input. 2. User menekan tombol submit. 3. Sistem melakukan klasifikasi pada sinopsis. 4. Sistem menampilkan daftar sinopsis yang telah diklasifikasikan. 5. User melihat hasil klasifikasi sinopsis. Tabel 4.5 menunjukkan scenario use case ketika user melihat hasil klasifikasi yang dihasilkan oleh aplikasi. User dapat melihat hasil klasifikasi berdasarkan banyaknya jumlah sinopsis yang dimasukkan ke dalam form input. Tabel 4.6 Skenario Use Case Melihat Riwayat Hasil Klasifikasi Nama Use Case Melihat Riwayat atau History Hasil Klasifikasi Aktor User Deskripsi User dapat melihat hasil yang disajikan oleh sistem. Tahapan 1. User memasukkan teks sinopsis ke dalam form input. 2. User menekan tombol submit. 3. Sistem melakukan klasifikasi pada sinopsis. 4. Sistem menampilkan daftar sinopsis yang telah diklasifikasikan. 5. Sistem memasukkan hasil klasifikasi ke dalam riwayat. 6. User melihat riwayat hasil klasifikasi sinopsis sebelumnya. Tabel 4.6 menunjukkan scenario use case ketika user melihat riwayat dari hasil klasifikasi yang dihasilkan oleh aplikasi. User dapat melihat seluruh riwayat hasil klasifikasi yang pernah dilakukan. 4.5.3 Activity Diagram Gambar 4.7 Activity Diagram Sistem Klasifikasi Gambar 4.7 memperlihatkan activity diagram yang telah dibuat untuk mengetahui alur kerja sistem klasifikasi genre yang terdiri dari tiga entitas atau unit yaitu user, sistem, dan model. User perlu menentukan jumlah form input yang berfungsi sebagai tempat untuk teks sinopsis sebelum dikirimkan ke sistem. Setelah mengirimkan teks sinopsis, sistem akan melakukan praproses data dan tokenisasi untuk dimuat

ke dalam model sebagai masukan. Selanjutnya, model akan melakukan klasifikasi terhadap teks sinopsis tersebut, menyimpannya di dalam session untuk riwayat dan akan digunakan sebagai sistem untuk luaran. 4.6

Perancangan Antarmuka Tahap ini adalah tahap persiapan dan perancangan tampilan serta fungsi yang akan berinteraksi langsung dengan pengguna. Tahapan ini bertujuan untuk memastikan interaksi pengguna dengan sistem dapat dilakukan dengan mudah dan efektif. Adapun rancangan antarmuka, disajikan sebagai berikut. Gambar 4. 8 Rancangan Tampilan Awal Aplikasi

Gambar 4.8 menunjukkan hasil rancangan tampilan awal aplikasi atau sering disebut dengan desain homepage . Pada sisi kiri terdapat tabel yang menampung sampel data untuk setiap genre yang akan diidentifikasi. Tabel tersebut memiliki dua kolom yang terdiri dari judul dan genre film. Sementara itu, pada sisi kanan terdapat bagian yang berisi form input dimana pengguna perlu menentukan banyaknya sinopsis yang akan dimasukkan dengan menekan tombol plus atau tambah. Gambar 4. 9 Rancangan Tampilan Pengguna Menentukan Jumlah Input Gambar 4.9 menunjukkan rancangan tampilan ketika pengguna menentukan jumlah input, aplikasi secara otomatis akan menampilkan form input sesuai jumlah yang telah ditentukan. Gambar 4. 10 Rancangan Tampilan Hasil Klasifikasi Gambar 4.10 menunjukkan rancangan tampilan ketika pengguna menekan tombol submit dan aplikasi akan secara otomatis melakukan klasifikasi terhadap sinopsis yang telah dimasukkan. Selanjutnya, aplikasi akan menampilkan tabel berisi sinopsis yang telah dimasukkan pengguna dan genre yang merupakan hasil klasifikasi. Pada saat yang sama, aplikasi juga akan menyimpan hasil klasifikasi ke dalam dropdown history , sehingga pengguna dapat melihat riwayat sinopsis yang telah diklasifikasi sebelumnya. Gambar 4. 11 Rancangan Tampilan Pengguna Menekan History Gambar 4.11 menunjukkan rancangan tampilan ketika pengguna menekan tombol dropdown history . Pada bagian history , pengguna dapat melihat riwayat hasil klasifikasi yang telah disimpan di dalam session pada website. 4.7 Skenario Pengujian Pada penelitian ini, setelah model klasifikasi genre berhasil dirancang, dibuat, dan diintegrasikan sehingga

menjadi sistem yang utuh, dilakukan dua metode pengujian yang telah dijelaskan pada bab tiga. Adapun rincian mengenai setiap skenario dijelaskan sebagai berikut.

4.7.1 Skenario Pengujian Black Box

Pengujian ini bertujuan untuk mengevaluasi kinerja sistem dan meminimalisir bug pada tampilan saat waktu eksekusi berlangsung. Berikut adalah rancangan pengujian yang dibuat pada penelitian ini.

25 Tabel 4. **25** 7 Skenario Pengujian Black Box No. Skenario Pengujian Hasil yang diharapkan

1. User mendefinisikan jumlah sinopsis yang ingin dimasukkan Sistem menampilkan form input sesuai dengan jumlah yang dimasukkan
2. User mengisi form input dengan sinopsis film
3. Sistem menampilkan isi form yang dimasukkan oleh user
4. User menekan tombol submit untuk memulai proses klasifikasi genre film
5. Sistem menampilkan tabel yang berisi sinopsis dan hasil klasifikasi
6. User menekan tombol history
7. Sistem menampilkan riwayat sinopsis yang telah dimasukkan sebelumnya

4.7.2 Skenario Pengujian White Box

Pengujian pada skenario ini dirancang untuk melihat kesesuaian logika pemrograman berdasarkan masukan dan hasil yang diharapkan berupa keluaran. Adapun skenario pengujian white box ditunjukkan pada tabel di bawah ini.

25 8 Tabel 4. **25** 8 Skenario Pengujian White Box No. **30** Skenario Pengujian Hasil yang diharapkan

1. Sistem mampu menyajikan form input sesuai dengan jumlah yang dimasukkan
2. Sistem mampu menyajikan hasil klasifikasi
3. **1 2 3** Sistem mampu menyajikan riwayat atau history hasil klasifikasi

BAB V HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil pengembangan sistem yang telah dilakukan peneliti dan membahas pengembangan sistem dalam dua sub bab, yaitu hasil dan pembahasan.

5.1 Hasil

Bagian ini menyajikan hasil penelitian yang telah dicapai, meliputi perancangan model, proses klasifikasi, serta implementasi sistem yang telah dikembangkan.

5.1.1 Pra-proses Data dan Pembuatan Model

a) Case Folding Gambar 5. **1** 1 Kode Penggalan Case Folding Gambar 5.1 menunjukkan penggalan kode untuk tahap pertama pra-pemrosesan yaitu case folding. Nama fungsi yang didefinisikan adalah `case_folding` dan memiliki satu parameter dengan nama value yang digunakan untuk menampung argumen atau sinopsis yang akan diterapkan case folding.

Argumen tersebut selanjutnya dimasukkan ke dalam variabel dengan nama result . Output yang dihasilkan dari fungsi ini adalah sinopsis yang telah diubah seluruh huruf kapitalnya menjadi huruf kecil sekaligus menghilangkan spasi pada awal dan akhir kalimat. Tabel 5. 1 Hasil Case Folding Input (Sebelum) Output (Sesudah) A gang of unemployed itinerant musicians play in the south of Stockholm. Then they get the chance to be an orchestra in a dance restaurant. It goes a gang of unemployed itinerant musicians play in the south of stockholm. then they get the chance to be an orchestra in a dance restaurant. it goes well until the female owner falls in love with one of them. well until the female owner falls in love with one of them. Tabel 5.1 menunjukkan perbandingan sinopsis sebelum dan sesudah penerapan case folding

. Perbedaan yang jelas terlihat adalah seluruh kalimat berubah dari huruf besar menjadi huruf kecil. b) Cleaning Gambar 5.

1 2 Kode Penggalan Cleaning

Gambar 5.2 memperlihatkan penggalan kode untuk tahap kedua pra-pemrosesan yaitu cleaning .

Nama fungsi yang didefinisikan adalah cleaning_synopsis dan memiliki satu parameter yang digunakan untuk menampung hasil pada tahap sebelumnya. Fungsi tersebut memiliki beberapa tahapan seperti memastikan bahwa setiap sinopsis hanya dipisahkan oleh satu spasi serta menghilangkan hyperlinks atau url , hashtag , angka, simbol retweet, emoji, dan tanda baca dengan menggunakan regex. Tahap ini memastikan bahwa data yang akan dikirimkan pada tahap selanjutnya hanya memiliki unsur elemen yang dibutuhkan saja. Tabel 5. 2 Hasil Cleaning Input (Sebelum) Output (Sesudah) a gang of unemployed itinerant musicians play in the south of stockholm. then they get the chance to be an orchestra in a dance restaurant. it goes well until the female owner falls in love with one of them. a gang of unemployed itinerant musicians play in the south of stockholm then they get the chance to be an orchestra in a dance restaurant it goes well until the female owner falls in love with one of them Tabel 5.2 menunjukkan perbandingan sinopsis sebelum dan sesudah penerapan cleaning . Perbedaan yang dapat dilihat

REPORT #22060727

adalah hilangnya tanda baca pada sinopsis. c) Tokenisasi Gambar 5.3

Kode Penggalan Tokenisasi Gambar 5.3 memperlihatkan penggalan kode tahap ketiga pra-pemrosesan, yaitu tokenisasi. Nama fungsi pada tahap ini adalah `tokenize_synopsis` yang memiliki satu parameter untuk menampung hasil tahap sebelumnya. Tokenisasi diterapkan dengan memanfaatkan salah satu fungsi pada Pustaka NLTK yaitu `word_tokenize`. Fungsi tersebut akan memecah sinopsis menjadi kata-kata individual dan mengubahnya dari tipe data string menjadi list. Tabel 5.3 Hasil Tokenisasi Input (Sebelum) Output (Sesudah) a

gang of unemployed itinerant musicians play in the south of stockholm

then they get the chance to be an orchestra in a dance restaurant

it goes well until the female owner falls in love with one of them

['a', 'gang', 'of', 'unemployed', 'itinerant', 'musicians', 'play', 'in',

'the', 'south', 'of', 'stockholm', 'then', 'they', 'get', 'the', 'chance',

'to', 'be', 'an', 'orchestra', 'in', 'a', 'dance', 'restaurant', 'it',

'goes', 'well', 'until', 'the', 'female', 'owner', 'falls', 'in', 'love',

'with', 'one', 'of', 'them'] Tabel 5.3 menunjukkan perbandingan sinopsis

sebelum dan sesudah penerapan tokenisasi. Kalimat yang berupa rangkaian

kata dengan tipe data string diubah menjadi list atau daftar dengan

jumlah token sesuai dengan jumlah kata yang ada pada sinopsis. d)

Stopword Removal Gambar 5.4 Kode Penggalan Stopword Removal Gambar 5.4

memperlihatkan penggalan kode tahap keempat pra-pemrosesan yang diketahui sebagai `stopword removal`. Nama fungsi pada tahap ini adalah `filtering_`

`stopwords` dan memiliki satu parameter untuk menampung argumen atau hasil

pada tahap sebelumnya. Daftar kata `stopword` bahasa inggris yang digunakan

untuk menyaring sinopsis berasal dari library NLTK. Fungsi tersebut akan

melakukan pengecekan terhadap sinopsis, apakah terdapat kata yang termasuk

ke dalam daftar kata `stopword`. Jika tidak terdaftar, kata tersebut

dimasukkan ke dalam list kata yang akan menjadi luaran. Tabel 5.4

Hasil Stopword Removal Input (Sebelum) Output (Sesudah) ['a', 'gang',

'of', 'unemployed', 'itinerant', 'musicians', 'play', 'in', 'the', 'south',

'of', 'stockholm', 'then', 'they', 'get', 'the', 'chance', 'to', 'be',

REPORT #22060727

'an', 'orchestra', 'in', 'a', 'dance', 'restaurant', 'it', 'goes', 'well', 'until', 'the', 'female', 'owner', 'falls', 'in', 'love', 'with', 'one', 'of', 'them'] ['gang', 'unemployed', 'itinerant', 'musicians', 'play', 'south', 'stockholm', 'get', 'chance', 'orchestra', 'dance', 'restaurant', 'goes', 'well', 'female', 'owner', 'falls', 'love', 'one']

Tabel 5.4 menunjukkan perbandingan sinopsis sebelum dan sesudah penerapan fungsi `stopword_removal`. Kata yang termasuk dalam daftar kata `stopword` akan dihilangkan agar mengurangi dimensi data. Adapun contoh kata yang termasuk ke dalam `stopword` yaitu `in`, `to`, `and`, `or`, `this`, dan kata lainnya yang tidak relevan.

e) Stemming

Gambar 5.5 Kode Penggalan Stemming

Gambar 5.5 menunjukkan penggalan kode tahap pra-pemrosesan ke-5 yaitu `stemming`. Nama fungsi pada tahap ini adalah `stemming_synopsis` dan memiliki satu parameter untuk menampung argumen atau hasil pada tahap sebelumnya. Pada tahap ini, setiap kata dalam daftar tersebut akan diubah menjadi bentuk dasarnya (stem) dengan menggunakan salah satu fungsi pada library NLTK yaitu `PorterStemmer`.

Tabel 5.5 Hasil Stemming

Input (Sebelum)	Output (Sesudah)
['gang', 'unemployed', 'itinerant', 'musicians', 'play', 'south', 'stockholm', 'get', 'chance', 'orchestra', 'dance', 'restaurant', 'goes', 'well', 'female', 'owner', 'falls', 'love', 'one']	['gang', 'unemploy', 'itiner', 'musician', 'play', 'south', 'stockholm', 'get', 'chanc', 'orchestra', 'danc', 'restaur', 'goe', 'well', 'femal', 'owner', 'fall', 'love', 'one']

Tabel 5.5 menunjukkan perbandingan hasil sinopsis yang telah melalui tahap pra-pemrosesan akhir yaitu `stemming`. Kumpulan kata yang ada pada list ditransformasi menjadi kata dasar atau root dengan menghilangkan seluruh imbuhan seperti “-ed”, “-ing”, “-s” dan sebagainya. Perbedaan dapat dilihat dari perubahan kata `unemployed` menjadi `unemploy`, `musicians` menjadi `musician`, `falls` menjadi `fall`.

f) Pembuatan Model

Gambar 5.6 Kode Penggalan Pendefinisian Model

Gambar 5.6 menunjukkan penggalan kode untuk mendefinisikan model yang menggunakan arsitektur BiLSTM untuk mengklasifikasi genre film. Model tersebut dari lapisan `embedding` yang berperan untuk

mengubah input teks menjadi representasi vektor numerik dan menyesuaikan ukuran dimensinya untuk dimasukkan ke layer BiLSTM. Pada penelitian ini, sebanyak 32 unit layer BiLSTM digunakan pada pendefinisian model. Layer GlobalMaxPooling1D berfungsi untuk mengambil nilai maksimum sepanjang dimensi dengan tujuan untuk mereduksi dimensi input data. 1 17

Layer dense pada tahap ini menggunakan sebanyak 10 unit dan akan menjadi output pengolahan jaringan dengan menggunakan fungsi aktivasi yang disebut dengan softmax . 5.2

Pembahasan Sub bab ini akan focus membahas evaluasi terhadap aplikasi yang telah dibuat. Pengujian akan dilakukan menggunakan dua metode yang sudah ditetapkan sebelumnya guna mengetahui kinerja dari aplikasi

klasifikasi genre film berdasarkan sinopsis. 5.2.1 Pengujian Black Box

Pengujian black box akan memusatkan perhatian untuk mengevaluasi daya kerja fitur yang telah dibangun. Tujuannya adalah untuk melakukan verifikasi pada setiap fitur berfungsi agar sesuai dengan yang diharapkan.

Tabel 5. 6 Pengujian Black Box No. 30 Skenario Pengujian Hasil yang diharapkan 1.

User mendefinisikan jumlah sinopsis yang ingin dimasukkan Sistem menampilkan form input sesuai dengan jumlah yang dimasukkan Hasil:

Kesimpulan: Sistem berhasil menyajikan form input sesuai jumlah yang

dimasukkan 2. User mengisi form input dengan sinopsis film Sistem

menampilkan isi form yang dimasukkan oleh user Hasil: Kesimpulan: Sistem

berhasil menyajikan isi dari form yang dimasukkan oleh user 3. User

menekan tombol submit untuk memulai proses klasifikasi genre film Sistem

menampilkan tabel yang berisi sinopsis dan hasil klasifikasi Kesimpulan:

Sistem berhasil menyajikan tabel yang berisi sinopsis dan hasil

klasifikasi 4. User menekan tombol history Sistem menampilkan riwayat

sinopsis yang telah dimasukkan sebelumnya Hasil: Kesimpulan: Sistem berhasil

menampilkan riwayat sinopsis yang telah dimasukkan sebelumnya 5.2 1 2 Pengujian

White Box Pengujian white box akan memusatkan perhatian pada evaluasi

struktur kode yang dikembangkan. 29 Pengujian ini dilakukan untuk memverifikasi

apakah kode program berjalan sesuai dengan yang diharapkan. Adapun hasil dari

pengujian white box , ditunjukkan pada tabel di bawah ini. Tabel 5.

7 Pengujian White Box No. Hasil yang diharapkan Kode Program 1. Sistem mampu menyajikan form input sesuai dengan jumlah yang dimasukkan Hasil: Kesimpulan: Sistem berhasil menyajikan form input sejumlah yang dimasukkan oleh user 2. Sistem mampu menyajikan hasil klasifikasi Hasil: Kesimpulan: Sistem berhasil menyajikan hasil klasifikasi 3. 9 Sistem mampu menyajikan riwayat atau history hasil klasifikasi Hasil: Kesimpulan: Sistem berhasil menyajikan riwayat hasil klasifikasi BAB VI PENUTUP Bab ini memaparkan kesimpulan serta saran yang diperoleh dari penelitian yang telah dilakukan. Adapun penjelasan perihal kesimpulan dan saran dijabarkan sebagai berikut. 6.1 Kesimpulan Pengembangan aplikasi klasifikasi genre film berdasarkan sinopsis menggunakan algoritma Bidirectional Long Short-Term Memory (BiLSTM) telah dilakukan. Adapun kesimpulan yang diambil dari penelitian ini, dijabarkan sebagai berikut. 1) Melalui penelitian ini berhasil dikembangkan perangkat lunak berbasis Natural Language Processing berupa sistem klasifikasi genre film berdasarkan sinopsis. Sistem ini dibuat dengan menggunakan dataset yang diperoleh dari kaggle.com dan telah melewati proses resampling sehingga penggunaan algoritma BiLSTM dapat dimaksimalkan. 2) Pengembangan sistem klasifikasi genre film berdasarkan sinopsis dengan algoritma BiLSTM mampu melakukan klasifikasi terhadap sinopsis sehingga menghasilkan luaran berupa salah satu dari sepuluh genre film. 7 Sepuluh genre film tersebut mencakup action, adventure, animation, comedy, crime, drama, horror, romance, sci-fi, dan thriller. 3) Sistem klasifikasi yang dibangun menghasilkan rata-rata akurasi sebesar 41%. Akurasi ini perlu ditingkatkan pada penelitian selanjutnya. 6.2 Saran Berdasarkan hasil serta pembahasan yang telah diuraikan pada bab sebelumnya, beberapa saran untuk penelitian selanjutnya dapat diuraikan sebagai berikut. (1) Menggunakan teknik word embedding seperti Word2Vec , GloVe , FastText , BERT atau teknik lainnya agar model dapat menangkap konteks kata kunci pada setiap genre secara rinci. (2) Memiliki dataset yang jumlahnya seimbang untuk setiap genre film yang ingin diidentifikasi agar dapat memaksimalkan potensi Bidirectional LSTM dalam melakukan klasifikasi tanpa melakukan proses resampling . (3) Mengidentifikasi lebih

REPORT #22060727

lanjut penyebab mengapa algoritma BiLSTM belum dapat mencapai performa maksimal dalam melakukan klasifikasi.



REPORT #22060727

Results

Sources that matched your submitted document.

● IDENTICAL ● CHANGED TEXT

INTERNET SOURCE		
1.	2% eprints.upj.ac.id https://eprints.upj.ac.id/id/eprint/6052/12/12.%20BAB%20V.pdf	● ●
INTERNET SOURCE		
2.	1.09% repository.uinjkt.ac.id https://repository.uinjkt.ac.id/dspace/bitstream/123456789/75077/1/SEVIA%20F..	●
INTERNET SOURCE		
3.	0.98% repository.uinjkt.ac.id https://repository.uinjkt.ac.id/dspace/bitstream/123456789/77128/1/ANNISA%2...	●
INTERNET SOURCE		
4.	0.63% mti.binus.ac.id https://mti.binus.ac.id/2019/12/02/long-short-term-memory-lstm/	●
INTERNET SOURCE		
5.	0.62% typeset.io https://typeset.io/journals/e-jurnal-ilmu-komputer-dan-sistem-informasi-1nnaw..	●
INTERNET SOURCE		
6.	0.46% elibrary.unikom.ac.id https://elibrary.unikom.ac.id/id/eprint/4767/8/UNIKOM_Dio%20Ilham%20Djatia...	●
INTERNET SOURCE		
7.	0.41% ejurnal.tunasbangsa.ac.id https://ejurnal.tunasbangsa.ac.id/index.php/jurasik/article/download/643/616	●
INTERNET SOURCE		
8.	0.38% eprints.upj.ac.id https://eprints.upj.ac.id/id/eprint/6056/9/BAB%20II.pdf	●
INTERNET SOURCE		
9.	0.36% e-journal.uajy.ac.id http://e-journal.uajy.ac.id/2333/5/4KOM02589.pdf	●



REPORT #22060727

INTERNET SOURCE		
10.	0.34% eprints.upj.ac.id	●
	https://eprints.upj.ac.id/id/eprint/7763/11/11.%20BAB%20IV.pdf	
INTERNET SOURCE		
11.	0.3% bisa.ai	●
	https://bisa.ai/portofolio/detail/MzA5OA	
INTERNET SOURCE		
12.	0.3% repota.jti.polinema.ac.id	●
	http://repota.jti.polinema.ac.id/608/5/Bab%20IV.pdf	
INTERNET SOURCE		
13.	0.3% repository.uinjkt.ac.id	●
	https://repository.uinjkt.ac.id/dspace/bitstream/123456789/65577/1/RAHMAWA...	
INTERNET SOURCE		
14.	0.29% eprints.upj.ac.id	●
	https://eprints.upj.ac.id/id/eprint/7599/11/BAB%20IV.pdf	
INTERNET SOURCE		
15.	0.28% eprints.upj.ac.id	●
	https://eprints.upj.ac.id/id/eprint/4083/9/BAB%20II.pdf	
INTERNET SOURCE		
16.	0.26% ettheses.uin-malang.ac.id	●
	http://ettheses.uin-malang.ac.id/59980/7/19650125.pdf	
INTERNET SOURCE		
17.	0.24% journal.ipm2kpe.or.id	●
	https://journal.ipm2kpe.or.id/index.php/INTECOM/article/download/6961/4332/	
INTERNET SOURCE		
18.	0.23% repository.iainkudus.ac.id	●
	http://repository.iainkudus.ac.id/10812/8/08.%20BAB%20V.pdf	
INTERNET SOURCE		
19.	0.23% repositori.buddhidharma.ac.id	●
	https://repositori.buddhidharma.ac.id/1973/3/20191000079_Ananda_Skripsi%20...	
INTERNET SOURCE		
20.	0.23% conference.upnvj.ac.id	●
	https://conference.upnvj.ac.id/index.php/senamika/article/download/2177/1665	



REPORT #22060727

INTERNET SOURCE		
21. 0.21%	kc.umn.ac.id https://kc.umn.ac.id/id/eprint/19945/2/BAB_II.pdf	●
INTERNET SOURCE		
22. 0.2%	jurnal.peneliti.net https://jurnal.peneliti.net/index.php/JIWP/article/download/2226/1706/	●
INTERNET SOURCE		
23. 0.18%	www.semanticscholar.org https://www.semanticscholar.org/paper/Emotion-Classification-Using-Machine-...	●
INTERNET SOURCE		
24. 0.18%	ejournal.undip.ac.id https://ejournal.undip.ac.id/index.php/jsinbis/article/download/56815/pdf	●
INTERNET SOURCE		
25. 0.18%	elibrary.unikom.ac.id https://elibrary.unikom.ac.id/id/eprint/1489/10/UNIKOM_Fathia%20Rahma_Bab..	● ●
INTERNET SOURCE		
26. 0.17%	dspace.uui.ac.id https://dspace.uui.ac.id/bitstream/handle/123456789/48009/19523122.pdf?sequ...	●
INTERNET SOURCE		
27. 0.16%	elibrary.unikom.ac.id https://elibrary.unikom.ac.id/id/eprint/956/10/UNIKOM_RIFKI%20VIANDA_BAB%..	●
INTERNET SOURCE		
28. 0.15%	dqlab.id https://dqlab.id/visual-studio-code-code-editor-populer-di-python	●
INTERNET SOURCE		
29. 0.15%	ejournal.unisbablitar.ac.id https://ejournal.unisbablitar.ac.id/index.php/antivirus/article/download/2501/1...	●
INTERNET SOURCE		
30. 0.12%	repository.uinjkt.ac.id https://repository.uinjkt.ac.id/dspace/bitstream/123456789/56343/1/NADIA%20...	●
INTERNET SOURCE		
31. 0.11%	jurnal.radenfatah.ac.id https://jurnal.radenfatah.ac.id/index.php/jusifo/article/view/13885/5621	● ●