

# Prediksi Genre Film Dengan Klasifikasi Multi Kelas Sinopsis\_INTECOMS

*by Admin Dinamik*

---

**Submission date:** 14-Apr-2024 07:14PM (UTC+0700)

**Submission ID:** 2193122335

**File name:** Genre\_Film\_Dengan\_Klasifikasi\_Multi\_Kelas\_Sinopsis\_INTECOMS.pdf (273K)

**Word count:** 3912

**Character count:** 24687

## **PREDIKSI GENRE FILM DENGAN KLASIFIKASI MULTI KELAS SINOPSIS MENGUNAKAN JARINGAN LSTM**

### **FILM GENRE PREDICTION WITH MULTI-CLASS CLASSIFICATION SYNOPSIS USING LSTM NETWORK**

**Riska Amelia<sup>1</sup>, Dwi Budi Santoso<sup>2</sup>**

<sup>1,2</sup>Universitas Stikubank

riskaamelia@mhs.unisbank.ac.id

#### **ABSTRACT**

*A film synopsis provides important information about storylines and themes that can be used to predict a film's genre. However, predicting a film's genre manually by looking at the synopsis in advance can make it less effective because the synopsis is long and varies a lot. Therefore, an automatic approach is needed by using multi-class classification techniques with LSTM (Long Short-Term Memory) networks to predict film genres based on their synopsis in order to make a useful contribution to the film industry. This study uses the LSTM network for multi-class classification to predict film genres with a datasets taken from the Kaggle website which consist of synopsis of films of different genres. After that, the synopsis of the film is processed through a process of tokenization, vectorization, word embedding using NLP (Natural Language Processing). The results of testing the dataset using the multi-class classification method with the LSTM network are able to achieve an accuracy rate of 98% and a loss of 5% in predicting film genres based on synopsis.*

**Keywords:** Multi-Class Classification, LSTM, film synopsis, film genre

#### **ABSTRAK**

Sinopsis film menyajikan informasi penting mengenai alur cerita dan tema yang dapat digunakan untuk memprediksi genre film. Namun, memprediksi genre film secara manual dengan melihat sinopsis terlebih dahulu dapat menjadikannya kurang efektif karena sinopsis yang panjang dan sangat bervariasi. Oleh karena itu, diperlukannya pendekatan otomatis dengan menggunakan teknik klasifikasi multi kelas dengan jaringan LSTM (*Long Short-Term Memory*) untuk memprediksi genre film berdasarkan sinopsisnya guna memberikan kontribusi yang bermanfaat dalam industri perfilman. Penelitian ini menggunakan jaringan LSTM untuk klasifikasi multi kelas untuk memprediksi genre film dengan dataset yang diambil dari situs Kaggle yang terdiri dari berbagai sinopsis film dengan genre yang berbeda-beda. Setelah itu, sinopsis film tersebut diolah melalui proses tokenisasi, vektorisasi, *embedding* kata menggunakan NLP (*Natural Language Processing*). Hasil pengujian terhadap dataset dengan menggunakan metode klasifikasi multi kelas dengan jaringan LSTM mampu mencapai tingkat akurasi 98% dan loss 5% dalam memprediksi genre film berdasarkan sinopsis.

**Kata Kunci:** Kalsifikasi Multi Kelas, LSTM, Sinopsis Film, Genre Film

#### **PENDAHULUAN**

Perkembangan industri film yang pesat dan peningkatan minat masyarakat terhadap film-film dari berbagai genre telah mendorong pengembangan teknik-teknik yang dapat memprediksi genre film secara otomatis (Nurlaela, 2020). Dalam beberapa tahun terakhir, dengan adanya kemajuan besar dalam bidang pengolahan bahasa alami dan pembelajaran mesin, telah muncul berbagai metode yang memanfaatkan kecerdasan buatan untuk memprediksi genre film berdasarkan

sinopsisnya. Salah satu metode yang menarik adalah menggunakan jaringan LSTM (*Long Short-Term Memory*) untuk melakukan klasifikasi multi kelas pada sinopsis film (Gunawan & Kristian, 2020).

Jaringan LSTM adalah jenis jaringan saraf rekuren (*recurrent neural network*) yang memiliki kemampuan untuk mengingat informasi jangka panjang dan menangani masalah urutan data (Amrastian et al., 2022). Dalam prediksi genre film, sinopsis film dapat dianggap sebagai data yang memiliki urutan kata-kata yang

penting untuk memahami konsep film tersebut. Dengan menggunakan jaringan LSTM, kita dapat memanfaatkan kemampuan jaringan ini untuk mengekstraksi fitur-fitur penting dari sinopsis dan memprediksi genre film secara akurat (Ningrum et al., 2021).

Penelitian ini bertujuan untuk mengembangkan dan menerapkan metode prediksi genre film dengan menggunakan klasifikasi multi kelas pada sinopsis menggunakan jaringan LSTM. Pendekatan ini diharapkan dapat memberikan kontribusi dalam memperbaiki kualitas sistem rekomendasi film, serta memberikan pemahaman yang lebih baik tentang struktur dan karakteristik film dari berbagai genre.

Metode yang digunakan dalam penelitian ini melibatkan tahap-tahap seperti pengumpulan dataset sinopsis film dari sumber terpercaya, pembersihan dan preprocessing data untuk menghilangkan noise dan mengubah data mentah menjadi format yang sesuai untuk pemrosesan oleh jaringan LSTM. Selanjutnya, dilakukan proses pelatihan (training) jaringan LSTM menggunakan dataset yang telah dipersiapkan. Setelah pelatihan selesai, dilakukan tahap pengujian (testing) untuk mengevaluasi performa model yang dikembangkan.

Melalui penelitian ini, diharapkan dapat ditemukan hasil yang menunjukkan kemampuan jaringan LSTM dalam memprediksi genre film secara akurat. Selain itu, diharapkan juga dapat mengidentifikasi faktor-faktor yang berkontribusi pada performa model, seperti ukuran dataset, panjang sinopsis, dan representasi teks yang digunakan. Hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan teknik-teknik prediksi genre film yang lebih canggih dan efektif di masa depan.

Penelitian sebelumnya yang berjudul “Analisis Sentimen Multi-Class pada Sosial Media menggunakan Metode *Long Short Term-Memory* (LSTM)” yang dilakukan

oleh Yuliya Astari, Afiyati, Saddam Wahid Rozaqi pada tahun 2021. Penelitian ini menggunakan dataset yang langsung diambil dan dikumpulkan dari tulisan pengguna pada media sosial. Hasil pengujian dengan melakukan uji coba sebanyak 5 kali menunjukkan nilai akurasi tertinggi 91,9% dan nilai rata-rata dari multi-class mendapatkan hasil 89,45% (Astari & Wahib Rozaqi, 2021).

Penelitian lainnya yang berjudul “Algoritma LSTM-CNN untuk Sentimen Klasifikasi dengan Word2vec pada Media Online” yang dilakukan oleh Dedi Tri Hermanto, Arief Setyanto, Emha Taufiq Luthfi pada tahun 2021. Penelitian ini menggunakan dataset judul artikel berbahasa Indonesia yang diambil dari situs Detik Finance. Berdasarkan hasil pengujian memperlihatkan bahwa metode LSTM, LSTM-CNN, CNN-LSTM memiliki hasil akurasi sebesar 62%, 65% dan 74% (Hermanto et al., 2021).

Kedua penelitian sebelumnya menggunakan metode LSTM yang dimana metode tersebut diterapkan juga dalam penelitian ini. Perbedaan penelitian sebelumnya dengan penelitian ini yaitu dataset yang digunakan. Pada penelitian ini menggunakan dataset 1005 data judul film. Dengan memanfaatkan keunggulan LSTM dalam memahami konteks dan dependensi jangka panjang (Arief et al., 2023), tujuan disusunnya penelitian ini untuk membangun model yang dapat mengklasifikasikan sinopsis film kedalam beberapa golongan genre secara efisien dan akurat.

Tujuan akhir dari penelitian ini yaitu untuk mengembangkan sistem yang dapat membantu industri film dalam mengklasifikasikan genre film. Sehingga mampu menyajikan rekomendasi film kepada penonton yang lebih relevan sesuai preferensi mereka dengan lebih akurat dan efisien. Selain itu, penelitian ini juga mampu memberikan wawasan dalam analisis teks dan pemrosesan bahasa alami

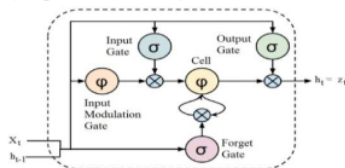
**METODE**

**Metode LSTM untuk Klasifikasi Data**

Metode LSTM (Long Short-Term Memory) adalah salah satu jenis jaringan saraf rekuren (recurrent neural network) yang memiliki kemampuan khusus untuk memahami dan mengolah data yang memiliki urutan atau urutan waktu, seperti teks atau data deret waktu. Dalam konteks prediksi genre film berdasarkan sinopsis, metode LSTM dapat digunakan untuk menganalisis urutan kata-kata dalam sinopsis film dan mempelajari pola-pola yang terkait dengan genre tertentu (Dwi et al., n.d.).

Salah satu keunggulan utama LSTM adalah kemampuannya untuk mengatasi masalah yang dikenal sebagai vanishing gradient problem, di mana gradien yang digunakan untuk melatih jaringan rekuren dapat menghilang atau meledak seiring dengan meningkatnya kedalaman jaringan (Setyo Nugroho et al., 2021). Dalam LSTM, ini dicapai melalui penggunaan unit memori khusus yang disebut sel memori (memory cell) yang dapat menyimpan informasi jangka panjang, serta gerbang-gembok (gate) yang mengatur aliran informasi di dalam sel memori.

LSTM terdiri dari beberapa gerbang yang bertanggung jawab untuk mengatur aliran informasi dalam dan keluar dari sel memori. Gerbang pertama adalah gerbang lupa (forget gate), yang mengontrol sejauh mana informasi lama akan dilupakan dalam sel memori. Gerbang kedua adalah gerbang input (input gate), yang menentukan sejauh mana informasi baru akan ditambahkan ke sel memori. Gerbang ketiga adalah gerbang keluaran (output gate), yang mengatur sejauh mana informasi dalam sel memori akan digunakan untuk menghasilkan output (Uly et al., 2023).



**Gambar 1. Gerbang dalam LSTM**

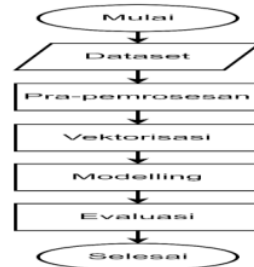
Selain itu, LSTM juga dapat memiliki jalur terhubung langsung yang disebut jalur sel (cell state), yang mengizinkan aliran informasi tanpa ada perubahan oleh gerbang. Jalur sel memungkinkan LSTM untuk mempelajari dan mengingat informasi jangka panjang (Rolangon et al., 2023) yang relevan untuk prediksi genre film.

Dalam prediksi genre film, jaringan LSTM akan dilatih menggunakan dataset sinopsis film yang memiliki label genre yang sesuai. Selama proses pelatihan, jaringan akan mempelajari pola dan hubungan antara urutan kata-kata dalam sinopsis dengan genre film yang sesuai. Kemudian, setelah pelatihan selesai, jaringan akan dapat menerima sinopsis film baru dan menghasilkan prediksi genre berdasarkan pola yang telah dipelajari selama pelatihan.

Dengan memanfaatkan kemampuan LSTM dalam mengingat informasi jangka panjang dan memahami urutan data, metode ini dapat membantu meningkatkan akurasi dalam prediksi genre film. LSTM dapat menangkap hubungan kompleks antara kata-kata dalam sinopsis dan genre film yang relevan, sehingga memungkinkan model untuk menghasilkan prediksi genre yang lebih akurat.

**Tahapan Penelitian**

Tahapan klasifikasi data yang dilakukan pada penelitian ini terdiri dari pengumpulan *dataset* berupa sinopsis film, pra-pemrosesan, pembentukan representasi vektor data, *modelling* menggunakan metode LSTM, dan evaluasi.



**Gambar 2. Tahapan Penelitian**

## 1. Dataset

Dalam penelitian ini, digunakan dataset yang diunduh dari platform Kaggle dengan judul "IMDB Synopsis Indonesian Movies" yang dapat diakses melalui tautan <https://www.kaggle.com/datasets/antoniuscs/imdb-synopsis-indonesian-movies>.

Dataset ini berisi ringkasan film dari berbagai genre dalam bahasa Indonesia, dengan jumlah 1005 data judul film dan sinopsisnya. Contoh sampel data dapat dilihat pada tabel 1.

**Tabel 1. Sampel Data**

| Sinopsis  | Genre    |
|---|----------|
| Raden Mas Said putra sulung Tumenggung Wilarikta di bawah Kerajaan Majapahit yang berkuasa di wilayah Tuban melihat sekeluarga miskin yang menderita busung lapar. Ia merasa sangat prihatin dan hati nuraninya tergugah untuk menolong. Kemudian ia mencoba secara diam-diam mengambil makanan dari lumbung orang tuanya. Perbuatan itu tidak disetujui orang tuanya bahkan ia dihukum sekap di gudang makanan itu. Sejak kejadian itu RM Said yang tumbuh dewasa tidak betah tinggal di rumah. Ia berkelana dari daerah satu ke daerah lainnya. | Drama    |
| Steve dan beberapa sahabatnya merayakan pesta bujang dan bermain "Jangka Setan" permainan memanggil arwah. Keesokan harinya mereka mulai dihantui berbagai macam kejadian aneh.   | Horror   |
| Suatu pagi yang tenang di bulan Januari 2016 mendadak menjadi genting dan mencekam setelah terjadi serangan bom di Jalan MH Thamrin Jakarta. Seorang polisi bernama Ardi (Ario Bayu) dengan satuan kepolisian segera memburu pelaku pengeboman.   | Laga     |
| Johny WW Lengkong, seorang pelaut yang telah meninggalkan keluarganya selama hampir 20 tahun, berniat kembali ke Manado dengan harapan keluarganya mau memaafkannya. Namun  | Romantis |

ternyata istrinya telah meninggal dunia dan meninggalkan seorang anak bernama Pingkan. Johny pun berjanji akan merawat Pingkan, namun ternyata semua tidak berjalan dengan mudah.

Dono (Abimana Aryasatya) Kasino (Vino G. Bastian) dan Indro (Tora Sudiro) kembali beraksi di tengah hiruk pikuk kota Jakarta. Sekali lagi mereka memainkan peran sekelompok perwira yang dipekerjakan oleh lembaga swasta CHIPS (A Great Way untuk Mengatasi Masalah-masalah Sosial). Meskipun mereka bersemangat melayani masyarakat mereka juga terus menimbulkan masalah karena tindakan konyol dan lucu mereka.

## 2. Pra-pemrosesan

Pra-pemrosesan dalam penelitian ini untuk mempersiapkan data sebelum digunakan dalam proses pemodelan. Dalam penelitian ini, sebelum sinopsis film dimasukkan ke dalam model klasifikasi multi kelas dengan jaringan LSTM perlu dilakukannya pra-pemrosesan data yang dimana termasuk pada serangkaian langkah. Tujuan pra-pemrosesan data yaitu untuk membersihkan data teks, mengubah format, atau mengubah data agar sesuai dengan kebutuhan model. Adapun tahapan-tahapan pra-pemrosesan yang dilakukan:

a) **Penghapusan karakter khusus**, dalam pra-pemrosesan data, penghapusan karakter khusus merupakan tahap membersihkan teks dari karakter yang tidak relevan atau mengganggu (Aditya Quantano Surbakti et al., 2021), serta tidak memberikan kontribusi dalam pengklasifikasian genre film. Penelitian ini menggunakan library NLTK (Natural Language Toolkit) untuk melakukan penghapusan karakter khusus pada sinopsis film sebelum proses pemrosesan lebih lanjut. Metode penghapusan karakter khusus dengan menggunakan library NLTK dilakukan dengan memanfaatkan berbagai alat pemrosesan bahasa alami



yang telah tersedia, memungkinkan untuk mengidentifikasi dan menghapus karakter khusus seperti tanda baca, simbol, angka, dan karakter non-alfanumerik lainnya, dengan hasil seperti pada tabel 2, dengan contoh 3 data teratas.

**Tabel 2. Hasil Menghapus Karakter Khusus**

| Sinopsis   |
|--|
| Raden Mas Said putra sulung Tumenggung Wilarikta di bawah Kerajaan Majapahit yang berkuasa di wilayah Tuban melihat sekeluarga miskin yang menderita busung lapar Ia merasa sangat prihatin dan hati nuraninya tergugah untuk menolong Kemudian ia mencoba secara diamdiam mengambil makanan dari lumbung orang tuanya Perbuatan itu tidak disetujui orang tuanya bahkan ia dihukum sekap di gudang makanan itu Sejak kejadian itu RM Said yang tumbuh dewasa tidak betah tinggal di rumah Ia berkelana dari daerah satu ke daerah lainnya Dari sanalah ia tahu betapa banyak penyelewengan dan kesewenangwenangan para lurah yang munafik |
| Soe Hok Gie adalah seorang aktivis yang hidup di tahun enam puluhan Terletak di era paling gelap sejarah modern Indonesia GIE adalah interpretasi dari apa yang terjadi berdasarkan jurnalnya  |
| Guru Bangsa Tjokroaminoto menceritakan tentang setelah lepas dari era tanam paksa di akhir tahun 1800 Hindia Belanda memasuki babak baru yang berpengaruh dalam kehidupan masyarakatnya yaitu dengan gerakan politik etis yang dilakukan oleh pemerintah Belanda Tetapi kemiskinan masih banyak terjadi Rakyat banyak yang belum mengenyam pendidikan dan kesenjangan sosial antaretnis dan kasta masih terlihat jelas   |

- b) **Tokenisasi**, tokenisasi adalah proses membagi teks yang dapat berupa kalimat, paragraf, atau dokumen menjadi unit yang lebih kecil seperti ata atau frasa yang disebut token (Jamaluddin et al., 2020). Misalnya tokenisasi sinopsis “Kisah cinta anak remaja yang penuh drama” menghasilkan tujuh token, yakni: “Kisah”, “cinta”, “anak”, “remaja”, “yang”, “penuh”, “drama”.

**Tabel 3. Hasil Tokenisasi**

| Sinopsis   |
|--|
| Raden, Mas, Said, putra, sulung, Tumenggung, Wilarikta, di, bawah, |

Kerajaan, Majapahit, yang, berkuasa, di, wilayah, Tuban, melihat, sekeluarga, miskin, yang, menderita, busung, lapar, Ia, merasa, sangat, prihatin, dan, hati, nuraninya, tergugah, untuk, menolong, Kemudian, ia, mencoba, secara, diamdiam, mengambil, makanan, dari, lumbung, orang, tuanya, Perbuatan, itu, tidak, disetujui, orang, tuanya, bahkan, ia, dihukum, sekap, di, gudang, makanan, itu, Sejak, kejadian, itu, RM, Said, yang, tumbuh, dewasa, tidak, betah, tinggal, di, rumah, Ia, berkelana, dari, daerah, satu, ke, daerah, lainnya, Dari, sanalah, ia, tahu, betapa, banyak, penyelewengan, dan, kesewenangwenangan, para, lurah, yang, munafik

Soe, Hok, Gie, adalah, seorang, aktivis, yang, hidup, di, tahun, enam, puluhan, Terletak, di, era, paling, gelap, sejarah, modern, Indonesia, GIE, adalah, interpretasi, dari, apa, yang, terjadi, berdasarkan, jurnalnya

Guru, Bangsa, Tjokroaminoto, menceritakan, tentang, setelah, lepas, dari, era, tanam, paksa, di, akhir, tahun, 1800, Hindia, Belanda, memasuki, babak, baru, yang, berpengaruh, dalam, kehidupan, masyarakatnya, yaitu, dengan, gerakan, politik, etis, yang, dilakukan, oleh, pemerintah, Belanda, Tetapi, kemiskinan, masih, banyak, terjadi, Rakyat, banyak, yang, belum, mengenyam, pendidikan, dan, kesenjangan, sosial, antaretnis, dan, kasta, masih, terlihat, jelas

- c) **Penghapusan Stopword**, stopword merupakan kata-kata yang muncul namun tidak memiliki makna yang signifikan dalam analisis. Misalnya dalam bahasa Indonesia adalah “dan”, “di”, “dari”, “yang”. Penghapusan stopword adalah salah satu langkah yang umum dilakukan dalam pra-pemrosesan teks untuk menghilangkan kata-kata umum yang tidak memberikan banyak informasi penting dalam analisis teks. Proses penghapusan stopword bertujuan untuk membersihkan teks dari kata-kata yang umum dan tidak relevan, sehingga meningkatkan akurasi dan keberlanjutan analisis (Khomsah & Agus Sasmito Aribowo, 2020). Penghapusan stopword dapat membantu mengurangi kebisingan dalam teks dan memfokuskan perhatian pada kata-kata yang lebih informatif untuk ekstraksi fitur atau analisis berbasis kata.

**Tabel 4. Hasil Penghapusan Stopword****Sinopsis**

Raden Mas Said putra sulung Tumenggung Wilarikta bawah Kerajaan Majapahit berkuasa wilayah Tuban melihat sekeluarga miskin menderita busung lapar merasa sangat prihatin hati nuraninya tergugah untuk menolong Kemudian mencoba secara diam-diam mengambil makanan lumbung orang tuanya Perbuatan itu tidak disetujui orang tuanya bahkan dihukum sekap gudang makanan itu Sejak kejadian itu RM Said tumbuh dewasa tidak betah tinggal rumah Ia berkelana daerah satu ke daerah lainnya Dari sanalah tahu betapa banyak penyelewengan kesewenang-wenangan para lurah munafik

Soe Hok Gie seorang aktivis hidup tahun enam puluhan Terletak era paling gelap sejarah modern Indonesia GIE adalah interpretasi terjadi berdasarkan jurnalnya

Guru Bangsa Tjokroaminoto menceritakan tentang setelah lepas era tanam paksa akhir tahun 1800 Hindia Belanda memasuki babak baru berpengaruh dalam kehidupan masyarakatnya dengan gerakan politik etis dilakukan pemerintah Belanda Tetapi kemiskinan masih banyak terjadi Rakyat banyak belum mengenyam pendidikan kesenjangan sosial antaretnis kasta masih terlihat jelas

- d) **Stemming**, merupakan proses dalam pra-pemrosesan teks yang bertujuan untuk mengubah kata-kata menjadi bentuk dasarnya yang disebut "kata dasar" atau "akar kata" (Guterres et al., 2019). Tujuan dari stemming adalah untuk mengurangi variasi bentuk kata dalam teks sehingga kata-kata dengan akar yang sama dianggap sebagai entitas yang sama, meskipun ada variasi bentuk morfologis. Penelitian ini, menggunakan library Sastrawi untuk melakukan stemming pada sinopsis film. Sastrawi merupakan library pemrosesan bahasa Indonesia yang dirancang khusus untuk memenuhi kebutuhan pemrosesan teks dalam bahasa Indonesia, termasuk stemming (Rahmawati & Santoso, 2023). Stemming menggunakan aturan dan algoritma linguistik untuk menghapus imbuhan (affixes) pada kata-kata dan menghasilkan akar kata yang umum. Proses stemming dapat melibatkan

pemangkasan (trimming) imbuhan kata-kata, penggantian imbuhan dengan bentuk dasar yang tepat, atau kombinasi keduanya.

**Tabel 5. Hasil Stemming****Sinopsis**

raden, mas, said, putra, sulung, tumenggung, wilarikta, raja, majapahit, kuasa, wilayah, tuban, keluarga, miskin, derita, busung, lapar, prihatin, hati, nurani, gugah, tolong, kemudian, coba, diam-diam, ambil, makan, lumbung, orang, tua, buat, tuju, orang, tua, hukum, sekap, gudang, makan, sejak, jadi, rm, said, tumbuh, dewasa, betah, tinggal, rumah, ia, kelana, daerah, daerah, dari, sana, betapa, seleweng, kesewenang-wenangan, lurah, munafik

soe, hok, gie, aktivis, hidup, enam, puluh, letak, era, gelap, sejarah, modern, indonesia, gie, interpretasi, dasar, jurnal

guru, bangsa, tjokroaminoto, cerita, lepas, era, tanam, paksa, 1800, hindia, belanda, pasuk, babak, pengaruh, hidup, masyarakat, gera, politik, etis, perintah, belanda, tetapi, miskin, rakyat, kenyam, didik, senjang, sosial, antaretnis, kasta

### 3. Modelling

Pada tahapan ini, dimodelkan sebuah model jaringan saraf rekuren (RNN) dengan arsitektur STM (Long Short-Term Memory) untuk memprediksi genre film berdasarkan sinopsisnya. Model ini dirancang dengan tujuan untuk menggabungkan kecerdasan buatan dan pemrosesan bahasa alami dalam upaya meningkatkan kualitas sistem rekomendasi film, dengan menggunakan library TensorFlow untuk memodelkan jaringan saraf rekuren. Model ini terdiri dari beberapa layer yang berfungsi untuk memproses dan mempelajari pola dari data sinopsis film.

Pertama, terdapat layer embedding yang bertugas mengonversi kata-kata dalam sinopsis menjadi representasi vektor numerik. Kemudian, layer LSTM digunakan untuk memahami urutan kata-kata dan mempertahankan informasi jangka panjang. Proses ini memungkinkan model untuk memahami konteks dan relasi antara kata-kata dalam sinopsis film.

Selanjutnya, terdapat layer dense dengan fungsi aktivasi ReLU yang

bertindak sebagai layer tersembunyi dalam model. Layer ini membantu dalam mengenali pola-pola kompleks dan memperkuat fitur yang diekstraksi dari layer LSTM sebelumnya.

Terakhir, terdapat layer dense output dengan fungsi aktivasi softmax untuk menghasilkan prediksi genre film. Layer ini memberikan probabilitas untuk setiap genre film berdasarkan sinopsis yang diberikan, sehingga kita dapat mengidentifikasi genre yang paling mungkin sesuai dengan sinopsis film.

**Tabel 6. Model LSTM**

| Layer     | Nilai                            |
|-----------|----------------------------------|
| Embedding | input_dim=5000,<br>output_dim=16 |
| LSTM      | 64                               |
| Dense     | 128, activation='relu'           |
| Dense     | 5,<br>activation='softmax'       |

#### 4. Evaluasi

Pertama, didefinisikan fungsi loss (loss function) yang akan digunakan dalam pelatihan model menggunakan `categorical_crossentropy` karena jenis klasifikasi yang digunakan multi kelas, di mana label target diubah menjadi representasi one-hot encoding.

Selanjutnya, menentukan metrik yang akan digunakan untuk mengukur kinerja model selama pelatihan dan pengujian. Dalam penelitian ini, menggunakan metrik `accuracy` (akurasi) yang mengukur persentase keberhasilan dalam klasifikasi dengan benar.

#### HASIL DAN PEMBAHASAN

Hasil uji pada klasifikasi dari model yang digunakan, telah diukur nilai akurasi dan loss dari metode LSTM. Tabel 7 memperlihatkan hasil klasifikasi menggunakan metode LSTM.

**Tabel 7. Hasil Uji Loss dan Accuracy**

| Epoch | Loss   | Accuracy |
|-------|--------|----------|
| 1     | 1.6101 | 0.1903   |
| 2     | 1.6063 | 0.2376   |
| 3     | 1.6499 | 0.3557   |
| 4     | 1.4392 | 0.4117   |
| 5     | 1.1555 | 0.4801   |

|    |        |        |
|----|--------|--------|
| 6  | 0.9392 | 0.4975 |
| 7  | 0.8180 | 0.5771 |
| 8  | 0.7089 | 0.6468 |
| 9  | 0.5358 | 0.7649 |
| 10 | 0.3745 | 0.8122 |
| 11 | 0.2639 | 0.9167 |
| 12 | 0.2294 | 0.9403 |
| 13 | 0.1379 | 0.9689 |
| 14 | 0.0752 | 0.9851 |
| 15 | 0.0506 | 0.9876 |

Dalam tabel tersebut, terlihat bahwa loss dan akurasi mengalami perubahan pada setiap *epoch*. Pada awal pelatihan yaitu pada *epoch* 1-3, loss dan akurasi masih rendah. Namun, seiring berjalannya pelatihan, loss semakin berkurang dan akurasi semakin meningkat secara signifikan.

Pada *epoch* 11, akurasi telah mencapai sebesar 0.9167. Dan pada *epoch* 15, akurasi terus meningkat menjadi 0.9876. Hal ini menunjukkan bahwa model telah mengalami peningkatan secara signifikan selama dilakukannya proses pelatihan.

Setelah melakukan uji coba dengan mengubah *epoch* yang berbeda-beda, terlihat hasil akurasi yang paling baik 98% dengan melakukan sebanyak 15 kali *epoch*

#### SIMPULAN

Dalam penelitian ini dilakukan percobaan mengenai penggunaan metode LSTM untuk klasifikasi multi kelas, didapatkan kesimpulan sebagai berikut:

1. Metode LSTM data diterapkan pada model untuk proses klasifikasi pada sinopsis film menggunakan data dari situs Kaggle sebanyak 1005 data judul film dan sinopsisnya.
2. Hasil uji coba pada metode LSTM memperoleh hasil akurasi sebesar 98% dan loss sebesar 5%.

#### DAFTAR PUSTAKA

Aditya Quantano Surbakti, Regiolina Hayami, & Januar Al Amien. (2021). Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter. *Jurnal CoSciTech (Computer Science and*



- Information Technology*), 2(2), 91–97.  
<https://doi.org/10.37859/coscitech.v2i2.2851>
- Amrustian, M. A., Widayat, W., & Wirawan, A. M. (2022). Analisis Sentimen Evaluasi Terhadap Pengajaran Dosen di Perguruan Tinggi Menggunakan Metode LSTM. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(1), 535. <https://doi.org/10.30865/mib.v6i1.3527>
- Arief, B., Kholifatullah, H., & Prihanto, A. (2023). Penerapan Metode Long Short Term Memory Untuk Klasifikasi Pada Hate Speech. *Journal of Informatics and Computer Science*, 04.
- Astari, Y., & Wahib Rozaqi, S. (2021). Analisis Sentimen Multi-Class pada Sosial Media menggunakan metode Long Short-Term Memory (LSTM). *Jurnal Linguistik Komputasional*, 4(1).
- Dwi, R., Santosa, W., Arif Bijaksana, M., & Romadhony, A. (n.d.). Implementasi Algoritma Long Short-Term Memory (LSTM) untuk Mendeteksi Penggunaan Kalimat Abusive Pada Teks Bahasa Indonesia. In *Jurnal Tugas Akhir Fakultas Informatika* (Vol. 12, Issue 2020).
- Gunawan, R., & Kristian, Y. (2020). Automatic Parental Guide Scene Classification Menggunakan Metode Deep Convolutional Neural Network Dan LSTM. *Journal of Intelligent System and Computation*, 2(2), 86–90. <https://doi.org/10.52985/insyst.v2i2.124>
- Guterres, A., Gunawan, & Santoso, J. (2019). Stemming Bahasa Tetun Menggunakan Pendekatan Rule Based. *Teknika*, 8(2), 142–147. <https://doi.org/10.34148/teknika.v8i2.224>
- Hermanto, D. T., Setyanto, A., & Luthfi, E. T. (2021). Algoritma LSTM-CNN untuk Binary Klasifikasi dengan Word2vec pada Media Online. *Creative Information Technology Journal*, 8(1), 64. <https://doi.org/10.24076/citec.2021v8i1.264>
- Jamaluddin, T., Arif Bijaksana, M., & Asror, I. (2020). Perbandingan Algoritma Sentencepiece BPE dan Unigram Pada Tokenisasi Artikel Bahasa Indonesia. *E-Proceeding of Engineering*, 8323–8331.
- Khomsah, S., & Agus Sasmito Aribowo. (2020). Text-Preprocessing Model Youtube Comments in Indonesian. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(4), 648–654. <https://doi.org/10.29207/resti.v4i4.2035>
- Ningrum, A. A., Syarif, I., Gunawan, A. I., Satriyanto, E., & Muchtar, R. (2021). Algoritma Deep Learning-LSTM untuk Memprediksi Umur Transformator. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(3), 539. <https://doi.org/10.25126/jtiik.2021834587>
- Nurlaela, D. (2020). Penerapan Adaboost Untuk Meningkatkan Akurasi Naive Bayes Pada Prediksi Pendapatan Penjualan Film. *INTI Nusa Mandiri*, 14(2), 181–188. <https://doi.org/10.33480/inti.v14i2.1220>
- Rahmawati, L., & Santoso, D. B. (2023). Implementasi Metode Naive Bayes Untuk Klasifikasi Ulasan Aplikasi E-Commerce Tokopedia Implementation Of Naive Bayes Method For Classification Of Tokopedia E-Commerce Application Review. *Journal of Information Technology and Computer Science (INTECOMS)*, 6(1).
- Rolangon, A., Weku, A., & Sandag, G. A. (2023). Perbandingan Algoritma LSTM Untuk Analisis Sentimen

Pengguna Twitter Terhadap Layanan Rumah Sakit Saat Pandemi Covid-19. *TeIKa*, 13(01), 31–40. <https://doi.org/10.36342/teika.v13i01.3063>

- Setyo Nugroho, K., Akbar, I., & Nizar Suksmawati, A. (2021). Deteksi Depresi Dan Kecemasan Pengguna Twitter Menggunakan Bidirectional Lstm. *The 4th Conference on Innovation and Application of Science and Technology (CIASTECH 2021)*, 287–296.
- Uly, N., Hendry, H., & Iriani, A. (2023). CNN-RNN Hybrid Model for Diagnosis of COVID-19 on X-Ray Imagery. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, 14(1), 57–67. <https://doi.org/10.31849/digitalzone.v14i1.13668>

# Prediksi Genre Film Dengan Klasifikasi Multi Kelas

## Sinopsis\_INTECOMS

---

ORIGINALITY REPORT

---

**14%**  
SIMILARITY INDEX

**13%**  
INTERNET SOURCES

**3%**  
PUBLICATIONS

**2%**  
STUDENT PAPERS

---

MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

---

1%  
★ **j-innovative.org**  
Internet Source

---

Exclude quotes      On  
Exclude bibliography      On

Exclude matches      Off