

# NERO<sup>®</sup>

Networking Engineering Research Operation



## Jurnal Multimedia

P-ISSN : 2355-2190  
E-ISSN : 2615-6539

Volume 7, No 1, April 2022

Diterbitkan oleh Jurusan Teknik Informatika Fakultas Teknik Universitas Trunojoyo Madura  
SK Menristek/BRIN Akreditasi Jurnal Ilmiah No. 200/M/KPT/2020 berlaku 2020-2024



## DEWAN REDAKSI

### Editor in Chief

Ika Oktavia Suzanti, Universitas Trunojoyo Madura

### Editorial Board

Dahlan Abdullah, Universitas Malikussaleh

I Dewa Made Cipta Santosa, Politeknik Negeri Bali

Fika Hastarita Rachman, Universitas Trunojoyo Madura

Aeri Rachmad, Universitas Trunojoyo Madura

Lukman Heryawan, Universitas Gadjah Mada

### Bestari partners

Muhammad Jauharul Fuady, Universitas Negeri Malang

Husni, Universitas Trunojoyo Madura

Rosa Ariani Sukanto, Universitas Pendidikan Indonesia

Ika Candradewi, Universitas Gadjah Mada

Baginda Anggun Nan Cenka, Universitas Indonesia

Iis Hamsir Ayub Wahab, Universitas Khairun Ternate

Adityo Hidayat, PT. Gamatechno Indonesia

### Web Administrator

Doni Abdul Fatah, Universitas Trunojoyo Madura

### Sekretariat Jurnal

Jurusan Teknik Informatika, Fakultas Teknik,

Universitas Trunojoyo Madura

Jalan Raya Telang, PO BOX 2 Kamal, Bangkalan

Telp: (031)3011147 Fax: (031)3011147

E-mail: [nero@trunojoyo.ac.id](mailto:nero@trunojoyo.ac.id)

Website : [nero.trunojoyo.ac.id](http://nero.trunojoyo.ac.id)



# SERTIFIKAT AKREDITASI NERO

**SERTIFIKAT**  
Kementerian Riset dan Teknologi/  
Badan Riset dan Inovasi Nasional



  
RISTEK-BRIN

Petikan dari Keputusan Menteri Riset dan Teknologi/  
Kepala Badan Riset dan Inovasi Nasional  
Nomor 200/M/KPT/2020  
Peringkat Akreditasi Jurnal Ilmiah Periode III Tahun 2020  
Nama Jurnal Ilmiah  
**Network Engineering Research Operation [NERO]**  
E-ISSN: 26156539  
Penerbit: Universitas Trunojoyo Madura  
Ditetapkan sebagai Jurnal Ilmiah

**TERAKREDITASI PERINGKAT 3**

Akreditasi Berlaku selama 5 (lima) Tahun, yaitu  
Volume 5 Nomor 1 Tahun 2020 sampai Volume 9 Nomor 2 Tahun 2024  
Jakarta, 23 Desember 2020  
Menteri Riset dan Teknologi/  
Kepala Badan Riset dan Inovasi Nasional  
Republik Indonesia.  
  
Bambang P. S. Brodjonegoro





## DAFTAR ISI

PREDIKSI RATING FILM PADA WEBSITE IMDB MENGGUNAKAN METODE <i>NEURAL NETWORK</i> ..... <b>Vincentius Riandaru Prasetyo, Mirella Mercifia, Anasthasya Averina, Lauren Sunyoto, Budiarto</b>	1 – 8
ANALISIS KOMPARASI METODE <i>EDGE DETECTION ROBINSON</i> DAN <i>KIRSCH</i> PADA DETEKSI BATIK PARANG RUSAK ..... <b>Farida, Muhamad Zun Afif Amrulloh</b>	9 – 20
IMPLEMENTASI METODE <i>FUZZY TOPSIS</i> DALAM PEMILIHAN OBYEK WISATA TERBAIK DI PULAU BAWEAN ..... <b>Nour Haqiki, Weny Mistarika Rahmawati, Maftahatul Hakimah</b>	21 – 28
KLASIFIKASI CITRA OBJEK WISATA DI KECAMATAN SAWAHAN MENGGUNAKAN <i>DOMINANT COLOR STRUCTURE DESCRIPTOR (DCSD)</i> ..... <b>Nurul Mahpiroh, Julian Sahertian, Resty Wulanningrum, Doni Abdul Fatah</b>	29 – 42
IMPLEMENTASI <i>CHATBOT</i> LAYANAN INFORMASI PENDAFTARAN MAHASISWA BARU PROGRAM PASCASARJANA DEPARTEMEN TEKNIK INFORMATIKA ITS..... <b>Eva Mursidah, Lina Ambarwati, Faiz Ainun Karima</b>	43 – 52
FAKTOR PENENTU PENERIMAAN TEKNOLOGI SISTEM PEMBAYARAN TAGIHAN BULANAN MELALUI <i>E-MARKETPLACE</i> MENGGUNAKAN METODE <i>COMBINED- THEORY OF PLANNED BEHAVIOUR-TECHNOLOGY ACCEPTANCE MODEL (C-TPB-TAM)</i> ..... <b>Rian Fitriana, S. Thya Safitri, Citra Wiguna</b>	53 – 68
IDENTIFIKASI HAMA DAN PENYAKIT TANAMAN JAGUNG DENGAN MENGGUNAKAN METODE KLASIFIKASI <i>SUPPORT VECTOR MACHINE (SVM)</i> ..... <b>Bain Khusnul Khotimah, Eko Setiawan, Verdi Sasmeka, Aulya Fridayanti, Ikbar Maulana, Arwinda Mifta Zulfida</b>	69 – 74
ANALISIS DATA LALU LINTAS JARINGAN DI KANTOR <i>CANGEHGAR CYBER OPERATION CENTER</i> MENGGUNAKAN ALGORITMA <i>K-MEANS</i> ..... <b>Rubangiya, Tuti Hartati, Yudhistira Arie Wijaya</b>	75 – 84



# PREDIKSI RATING FILM PADA WEBSITE IMDB MENGUNAKAN METODE NEURAL NETWORK

## FILM RATING PREDICTION ON IMDB WEBSITE USING NEURAL NETWORK

**Vincentius Riandaru Prasetyo<sup>1)</sup>, Mirella Mercifia<sup>2)</sup>, Anasthasya Averina<sup>3)</sup>, Lauren Sunyoto<sup>4)</sup>, Budiarjo<sup>5)</sup>**

*1), 2), 3), 4), 5) Jurusan Teknik Informatika, Fakultas Teknik, Universitas Surabaya  
Jl. Raya Rungkut, Kali Rungkut, Surabaya, Jawa Timur 60293  
Email : [vincent@staff.ubaya.ac.id](mailto:vincent@staff.ubaya.ac.id)<sup>1)</sup>, [s160418026@student.ubaya.ac.id](mailto:s160418026@student.ubaya.ac.id)<sup>2)</sup>,  
[s160418083@student.ubaya.ac.id](mailto:s160418083@student.ubaya.ac.id)<sup>3)</sup>, [s160418034@student.ubaya.ac.id](mailto:s160418034@student.ubaya.ac.id)<sup>4)</sup>,  
[s160418051@student.ubaya.ac.id](mailto:s160418051@student.ubaya.ac.id)<sup>5)</sup>*

### Abstrak

*Film adalah sebuah rekaman gambar bergerak yang menceritakan sebuah cerita atau kejadian yang dapat ditonton pada bioskop atau televisi. Kualitas atau kesuksesan dari film dapat dinilai dengan menggunakan sistem angka yang disebut rating. Rating biasanya diasosiasikan dengan rentang angka 0 sampai 5 atau 0 sampai 10. Rating film dapat dipengaruhi oleh banyak faktor, seperti alur cerita, sinematografi, pemeran, musik dan lainnya. Faktor-faktor tersebut merupakan data kualitatif sehingga untuk menentukan rating film berdasarkan rentang angka merupakan hal yang sulit. Untuk mengatasi masalah tersebut, maka dibuatlah sistem untuk memprediksi rating film berdasarkan kriteria-kriteria yang dimiliki oleh suatu film. Label Encoding, Normalizer, MinMax Scaler, dan Standard Scaler akan dipakai untuk preprocessing pada penelitian ini. Proses prediksi rating akan menggunakan metode Artificial Neural Network (ANN) dan Convolutional Neural Network (CNN). Dari hasil pengujian dengan menggunakan CNN didapatkan nilai akurasi sebesar 76.45%, dan hasil pengujian menggunakan ANN didapatkan nilai akurasi yang lebih baik yaitu sebesar 85.27%.*

**Kata kunci:** ANN, CNN, Film, Prediksi, Rating

### Abstract

*Film is a motion picture recording that tells a story or event that can be watched on a cinema or television. The quality or success of the film is judged by the rating it gets. Ratings are usually associated with a number range from 0 to 5 or up to 10. Film ratings can be influenced by many factors, such as storyline, cinematography, cast, music and others. These factors are qualitative data, so it is difficult to determine a film's rating based on a range of numbers. To overcome this problem, a system is made to predict film ratings based on criteria for a film. Encoding, Normalizer, MinMax Scaler, and Standard Scaler labels will be used for preprocessing in this research. The rating prediction will use Artificial Neural Network (ANN) and Convolutional Neural Network (CNN) methods. From test results using CNN obtained an accuracy value of 76.45%, and ANN obtained a better accuracy value of 85.27%.*

**Keywords :** ANN, CNN, Film, Prediction, Rating

## 1. PENDAHULUAN

Film adalah suatu karya seni visual yang digunakan untuk mengekspresikan ide, cerita, perasaan, dan lain-lain lewat media gambar yang bergerak. Dalam konteks kegiatan produksi film, sebuah film dapat memiliki banyak atribut yang mempengaruhi kesuksesan atau jumlah keuntungan ketika proses perilisasi nanti [1]. Kualitas film sangat bergantung pada pemilihan proyek film yang tepat dan melibatkan manajemen serta tim kreatif yang tepat (pemeran, penyutradaraan, desain visual, skor, fotografi, kostum, desain set, pengeditan, dan banyak spesialisasi tambahan). Salah satu contoh atribut yang mungkin berpengaruh terhadap kualitas film adalah sutradara dan aktor-aktris. Beberapa sutradara dan aktor-aktris biasanya hanya akan memilih proyek film dengan *screenplay* berkualitas. Dengan demikian, sutradara dan aktor-aktris tersebut cenderung memiliki reputasi tinggi karena film yang disutradarai atau dibintangi selalu memiliki *rating* tinggi (*critically-acclaimed*) [2].

Terdapat berbagai *platform online* yang memiliki catatan mengenai informasi detail dari sebuah film, salah satu contohnya adalah *website* *imdb.com* (*Internet Movie Database*). Berdiri sejak tahun 1990, IMDb tercatat telah memiliki jutaan data film dalam *database*-nya. Untuk memberikan *review*, pengguna diwajibkan untuk melakukan registrasi / *login* terlebih dahulu. Sistem *movie review* yang diadopsi oleh IMDb adalah pemberian komentar dan skala *rating* 1-10. Semakin tinggi *rating*, semakin bagus sentimen yang diberikan terhadap suatu film [3].

Penelitian ini bertujuan untuk membangun sistem yang dapat memprediksi *rating* suatu film berdasarkan atribut-atribut yang ada. Pemilihan kombinasi atribut film yang sesuai sebelum memulai produksi dapat meningkatkan kualitas film nantinya. Pihak produsen film dapat memakai sistem ini untuk mendapatkan rekomendasi dalam pemilihan atribut-atribut film berdasarkan prediksi *output* kinerja berbagai atribut tersebut. Karena *dataset* yang digunakan berasal dari *imdb.com*, hasil prediksi *output* dikemas ke dalam bentuk klasifikasi *rating* 1-10 sesuai dengan cara kerja *review* film pada *website* *imdb.com*.

Metode yang digunakan untuk membangun sistem ini adalah salah satu jenis *machine learning* yaitu algoritma *neural network*. *Neural Network* (NN) adalah model komputasi yang terinspirasi dari cara kerja sistem saraf makhluk hidup, yaitu syaraf otak. Beberapa ciri-ciri mendasar dari aplikasi NN adalah kemampuan sistem untuk belajar dari pengalaman. Dengan demikian, NN memiliki potensi untuk diaplikasikan sebagai penyelesaian beberapa masalah dalam ilmu di bidang sains, salah satunya adalah *pattern recognition/classification* [4].

Penelitian mengenai pembuatan sistem prediksi *rating* serupa salah satunya dilakukan oleh Abarja dan Wibowo [5] yang meneliti beberapa arsitektur model CNN dan membuat prediksi berdasarkan beberapa jenis *dataset*. Metode evaluasi yang dilakukan adalah perhitungan *mean squared error* dan *mean absolute error*. Penelitian tersebut menyimpulkan bahwa arsitektur *one dimensional CNN* (1DCNN) dapat memberikan hasil prediksi yang optimal.

Yumin, et.al. [6] melakukan perbandingan berbagai metode *machine learning* yang dapat digunakan untuk klasifikasi seperti *K-Nearest Neighbors*, *Decision Tree*, *Support Vector Machine*, dan *Naive*. Pada penelitian tersebut ditekankan bahwa tahap *pre-processing* sangat menunjang hasil prediksi dengan akurasi tinggi. Hasil evaluasi menunjukkan bahwa model prediksi dapat menghasilkan hingga 89.8% untuk prediksi yang dilakukan dengan metode *K-Nearest Neighbors*.

Berbeda lagi dengan penelitian yang dilakukan Asad, et.al. [7], atribut-atribut *dataset* penelitian berupa atribut-atribut film pada umumnya, namun *output* yang dihasilkan berupa prediksi tingkat popularitas film. Dalam penelitiannya, Asad, et.al. [7] menyimpulkan bahwa terdapat korelasi antara *budget* produksi film dengan tingkat kepopulerannya, dimana semakin besar *budget*, film tersebut akan memiliki tingkat kepopuleran yang tinggi.

*Dataset* dari IMDb juga digunakan untuk proses prediksi *rating* yang dilakukan oleh Hsu, et.al. [8]. Penelitian ini mengembangkan model prediksi *rating* yang menggunakan teknik klasifikasi (*linear combination*, *multiple linear regression*, *neural networks*). Total *dataset* penelitian meliputi 32968 film, 31506 film data latih, dan lainnya data uji. Dua temuan penelitian yang utama adalah : pertama, perhitungan nilai *absolute error* berada  $< 0,82$ ; kedua, penggunaan model *neural networks* lebih akurat daripada yang lain.

## 2. DASAR TEORI

### 2.1. *Crawling*

*Crawling* adalah teknik pengumpulan data dari internet yang biasanya didapatkan dari struktur HTML suatu *website*. *Crawling* berfungsi untuk mendapatkan informasi secara cepat dan legal yang ditampilkan pada sebuah halaman *website* [9]. Salah satu *library* yang dapat digunakan untuk proses *crawling* adalah *BeautifulSoup*. *BeautifulSoup* merupakan sebuah *library* Python yang digunakan untuk web *scraping*, *library* ini bertujuan untuk mengambil data atau informasi dari file HTML dan XML, *library* ini membuat sebuah *parse tree* dari *source code* halaman HTML dalam bentuk hierarki yang lebih mudah dibaca [10].

## 2.2. Preprocessing

*Preprocessing* merupakan teknik untuk mengubah data yang telah dikumpulkan dari berbagai sumber yang akan digunakan untuk pengolahan selanjutnya. *Preprocessing* digunakan untuk mengatasi data yang memiliki *missing value*, data *redundant*, *outliers*, ataupun format data yang tidak sesuai sistem di mana jika data memiliki data tersebut dapat mengganggu hasil *output* [11]. Ada beberapa *preprocessing* yang akan digunakan dalam penelitian ini antara lain: *Normalizer*, *MinMax Scaler*, dan *Standard Scaler*, di mana metode-metode tersebut akan digunakan secara terpisah untuk dilakukan perbandingan antar jenis *preprocessing*.

*Preprocessing* yang pertama adalah *normalizer*. *Normalizer* merupakan metode *preprocessing* di mana dilakukan proses normalisasi kepada setiap sampel data. Metode *normalizer* digunakan pada dataset yang memiliki banyak nilai 0 dengan atribut dari skala yang berbeda-beda. *Normalizer* digunakan untuk mengatasi data *outlier*. *Normalizer* yang akan digunakan adalah *Normalizer L2* yang akan bekerja dengan memastikan data untuk memiliki jumlah kuadrat bernilai 1 [12].

Metode *preprocessing* sebagai pembanding yang kedua adalah *MinMax Scaler*. *MinMax Scaler* adalah metode *preprocessing* dimana dilakukan transformasi fitur dengan cara menskala kan setiap fitur secara individual dengan rentang tertentu. *Preprocessing MinMax Scaler* dilakukan agar rentang setiap sampel pada suatu fitur tidak terlalu besar. *Preprocessing* ini dilakukan dengan melakukan pengurangan pada sampel dengan nilai sampel terkecil pada fitur dan akan dilakukan pembagian dengan nilai sample terbesar pada fitur yang telah dikurangi dengan nilai sample terkecil pada fitur [13]. Rumus dari *MinMaxScaler* ditunjukkan pada persamaan (1), di mana  $X$  adalah nilai sampel.

$$X_{sc} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

*Preprocessing* yang terakhir yang digunakan sebagai pembanding adalah *Standard Scaler*. *Standard Scaler* merupakan metode *preprocessing* di mana metode tersebut akan melakukan standarisasi fitur dengan menghapus rata-rata dan menskalakan unit varian. Proses ini akan dilakukan pada setiap *fitur* pada *sampel*. *Preprocessing* ini dilakukan untuk mencegah adanya data yang memiliki nilai terlalu besar dibanding dengan nilai yang lain yang akan dapat mengakibatkan proses *training* tidak sesuai keinginan [14]. Rumus dari *Standard Scaler* ditunjukkan pada Persamaan (2), di mana  $\bar{X}$  adalah rata-rata nilai sampel dan  $\sigma$  adalah standar deviasi.

$$Z = \frac{X_i - \bar{X}}{\sigma} \quad (2)$$

## 2.3. Neural Network

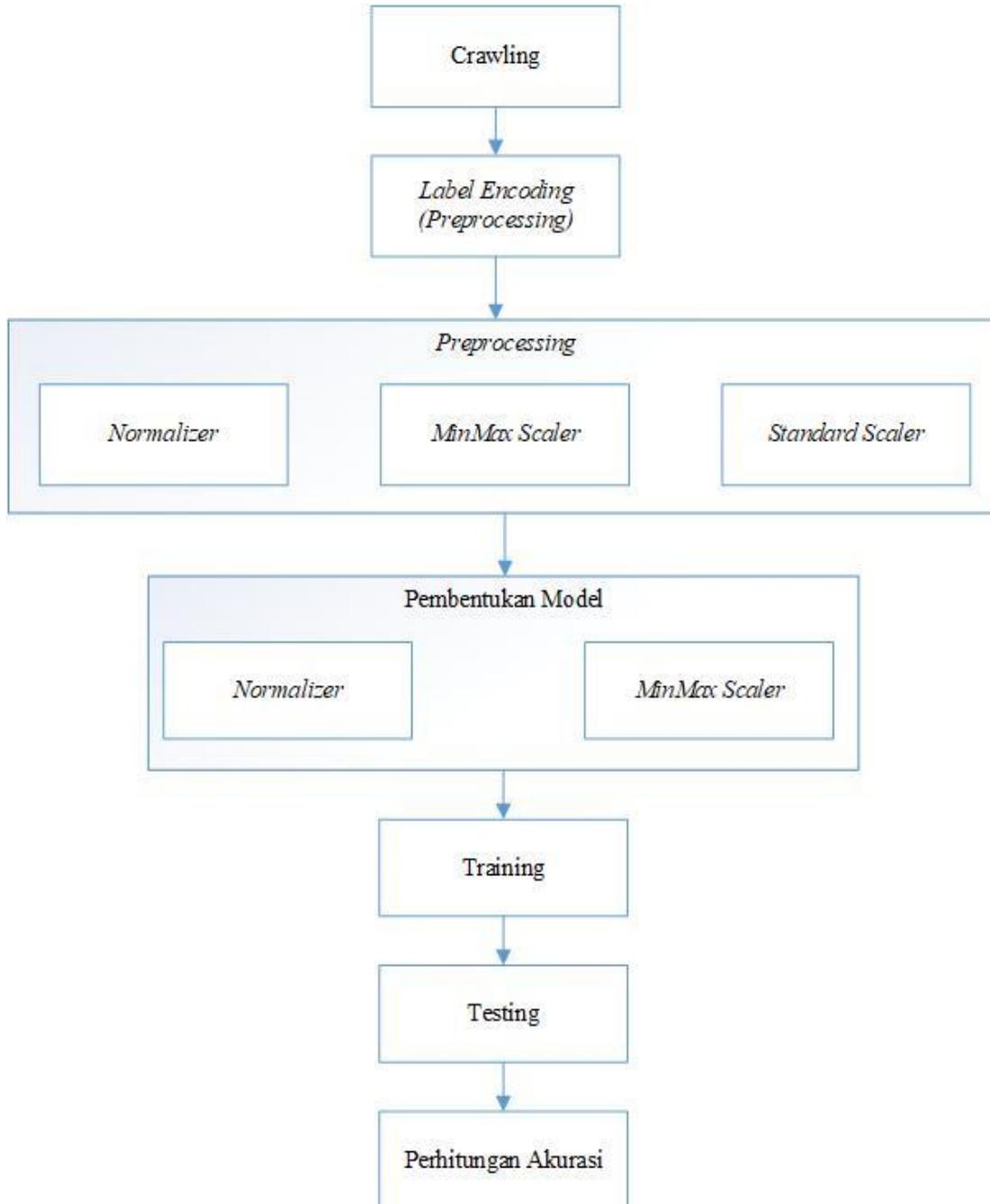
*Neural Network* merupakan cabang ilmu dari kecerdasan buatan, dimana menerapkan kemampuan otak manusia untuk menerima informasi dan informasi tersebut diproses untuk mengeluarkan sebuah *output*. *Neural Network* dapat digunakan untuk pengklasifikasian pola dan prediksi [15]. Terdapat dua model *neural network* yang akan dipakai dalam penelitian ini yaitu *Artificial Neural Network* (ANN) dan *Convolutional Neural Network* (CNN). Kedua metode tersebut akan dibandingkan tingkat keakuratannya.

*Artificial Neural Network* (ANN) merupakan metode pengolahan informasi yang meniru cara kerja sistem saraf biologis, khususnya manusia. *Neural Network* bekerja dengan cara memecah suatu masalah tertentu dan setelah itu membentuk beberapa instruksi kecil yang terstruktur. Instruksi tersebut akan dikonversi menjadi program komputer dan kemudian ke dalam kode mesin yang akan dapat digunakan untuk mendeteksi tren yang lebih kompleks [16].

*Convolutional Neural Network* (CNN) adalah teknik *machine learning* yang dikembangkan dari *Multi Layer Perceptron* (MLP) yang dirancang untuk memproses data dua dimensi. Cara kerja CNN mirip dengan MLP, tetapi pada CNN setiap *neuron* diwakili dalam bentuk dua dimensi, sedangkan MLP setiap neuron hanya memiliki memiliki satu dimensi [17].

### 3. METODOLOGI PENELITIAN

Metodologi penelitian yang digunakan terdiri dari beberapa tahapan, antara lain *crawling*, *preprocessing*, pembentukan model, *training*, *testing*, dan perhitungan akurasi. Gambar 1 menunjukkan alur dari penelitian yang dilakukan.



**Gambar 1.** Tahapan Penelitian

Tahapan pertama adalah *crawling*, di mana pada penelitian ini URL yang menjadi target *crawling* adalah [https://www.imdb.com/feature/genre/?ref\\_=nv\\_ch\\_gr](https://www.imdb.com/feature/genre/?ref_=nv_ch_gr). Proses *crawling* pada penelitian ini menggunakan library *BeautifulSoup*. Selain *BeautifulSoup* juga dibutuhkan modul *requests*, yang bertujuan untuk mengirim HTTP *request* menggunakan Python, yang akan mengembalikan hasil berupa *response object* yang berisi konten HTML, *encoding*, status dan lainnya. Informasi-informasi yang akan diambil dari proses *crawling* ini adalah judul, *url* film,

tahun rilis, durasi, *certificate*, sutradara, penulis, *vote*, negara, bahasa, perusahaan produksi dan *rating*.

Tahapan selanjutnya adalah *preprocessing*, di mana proses ini terpisah menjadi 3 bagian dengan metode yang berbeda-beda, yang nantinya akan dibandingkan nilai akurasi. Sebelum dilakukan *preprocessing*, setiap tabel yang memiliki data bertipe *string/text* akan diubah menjadi bentuk numerikal melalui proses *label encoding*, sehingga masing-masing dari *string/text* akan menjadi deretan angka yang akan mewakili setiap variasi teks. Data yang telah diubah melalui *label encoding*, kemudian akan dilakukan *preprocessing* dengan bantuan *library scikit-learn* yang tersedia secara gratis. Hasil *preprocessing* ini akan digunakan pada proses *training*.

Dalam penelitian ini, proses penentuan *rating* film akan menggunakan metode ANN dan CNN sebagai metode *training*. Metode ANN dan CNN akan menggunakan *library* dari *Keras* untuk membantu dalam pembentukan modelnya. Model yang dipakai adalah *sequential model*, di mana untuk ANN terdapat 5 *layer* dalam model ini. Untuk metode CNN, terdapat 6 *layer* yang digunakan. Pada penelitian ini, *dataset* dibagi menjadi 80% sebagai data *training* dan 20% sebagai data *testing*. Setelah itu akan dilakukan *epoch* sebesar 100 pada proses *training*. Proses *training* dilakukan untuk melatih dan membentuk model berdasarkan data *training* yang telah disiapkan. Selanjutnya, pada tahap *testing* akan dilakukan pengujian pada model yang telah terbentuk pada proses *training*.

Hasil dari proses *training* dan *testing* sebelumnya akan digunakan untuk melakukan proses validasi untuk menghitung tingkat akurasi dari tiap model yang dihasilkan, sehingga akan ditemukan model mana yang dapat melakukan prediksi *rating* film dengan baik. Persamaan (3) menunjukkan rumus dari perhitungan akurasi, di mana TP dan TN adalah jumlah data *testing* yang diprediksi dengan benar, sedangkan FP dan FN adalah jumlah data *testing* yang diprediksi dengan salah [18]. *Library scikit-learn* masih digunakan untuk membantu proses perhitungan akurasi ini.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

## 4. PENGUJIAN DAN PEMBAHASAN

### 4.1. Implementasi Sistem

Sebelum sistem mampu melakukan prediksi, maka dibutuhkan *dataset* yang diperoleh melalui proses *crawling* untuk pembentukan model prediksi oleh metode ANN dan CNN. *Dataset* yang didapatkan berasal dari hasil *crawling* yang dilakukan pada *website* IMDb dengan *url* [https://www.imdb.com/feature/genre/?ref\\_=nv\\_ch\\_gr](https://www.imdb.com/feature/genre/?ref_=nv_ch_gr). Pada proses *crawling*, sistem akan mengambil daftar *genre* film yang ada pada URL tersebut, kemudian pada setiap *genre* yang didapatkan akan mengambil daftar film yang ada. Setelah itu, pada *url* tiap film akan dilakukan proses *crawling* untuk mengambil informasi-informasi pada film tersebut seperti judul, *url* film, tahun rilis, durasi, *certificate*, sutradara, penulis, *vote*, negara, bahasa, perusahaan produksi dan *rating*. Data *rating* nantinya akan digunakan untuk sebagai pembanding antara prediksi yang dihasilkan sistem, sehingga dapat diukur nilai akurasi. Sedangkan data lainnya akan digunakan sebagai input parameter. Hasil *crawling* yang didapat ini akan disimpan pada *file* *csv* untuk dilakukan proses selanjutnya. Total data yang didapatkan dari proses *crawling* adalah 28,274 data film dari berbagai negara dan tahun.

Setelah *dataset* dikumpulkan, maka proses selanjutnya adalah *preprocessing* untuk mempersiapkan data sebelum diolah untuk pembentukan model prediksi. *Preprocessing* pertama adalah *Label Encoding*, di mana setiap tabel yang memiliki data bertipe *string/text* akan diubah menjadi bentuk numerikal. Hasil dari proses *label encoding*, akan masuk ke *preprocessing* yang kedua dengan menggunakan salah satu dari 3 metode yaitu: *Normalizer*, *Standard Scaler*, atau *MinMax Scaler*. Ketiga metode tersebut digunakan sebagai pembanding dari nilai akurasi model yang dihasilkan pada proses *training*.

Proses selanjutnya adalah *training*, di mana *output* dari proses ini adalah model prediksi yang menggunakan metode CNN atau ANN. Kedua metode tersebut digunakan sebagai

pembandingan untuk menentukan metode mana yang paling akurat dalam memprediksi *rating* dari suatu film. Untuk melakukan sebuah prediksi *rating* dari suatu film, *user* harus menginputkan parameter-parameter yang ditentukan seperti yang ditunjukkan pada Gambar 2. Setelah parameter-parameter diinputkan, user diharuskan untuk menekan tombol “Submit” untuk memulai proses prediksi.

**Gambar 2.** Halaman Utama Sistem

Dengan menggunakan model yang ada, sistem akan melakukan prediksi terhadap *rating* film dan hasil dari prediksi akan ditampilkan kembali pada halaman pengguna. Hasil dari prediksi akan ditampilkan dalam rentang nilai 1-10, sesuai dengan rentang *rating* yang digunakan pada IMDb. Contoh tampilan hasil prediksi dapat dilihat pada Gambar 3.

Rating Result

8/10

Reset

**Gambar 3.** Halaman Prediksi Sistem

**4.2. Evaluasi Sistem**

Setelah melakukan implementasi metode ke sistem, maka dilakukan perbandingan akurasi untuk menentukan metode *preprocessing* dan metode *training* yang terbaik dalam melakukan prediksi *rating* film. Beberapa percobaan dilakukan dengan mengkombinasikan masing-masing dari metode *preprocessing* yang ditentukan dengan metode ANN dan CNN. Percobaan dilakukan dengan menggunakan *epoch* sebesar 100, 20% data *testing*, dan *random state* adalah 0. Tabel 1 memperlihatkan perbandingan hasil akurasi dari masing-masing *preprocessing* dan metode *training* yang digunakan.

**Tabel 1.** Perbandingan Akurasi Tiap Preprocessing dan Metode Training

<b>Preprocessing</b>	<b>Metode Training</b>	<b>Akurasi (%)</b>
<i>Standard Scaler</i>	ANN	85.09
<i>Normalizer</i>		79.72
<i>MinMax Scaler</i>		81.02
<i>Standard Scaler</i>	CNN	75.33
<i>Normalizer</i>		75.33
<i>MinMax Scaler</i>		76.45

Dari hasil evaluasi yang ditunjukkan pada Tabel 1, dapat disimpulkan bahwa *preprocessing* dengan *Standard Scaler* dan metode *training* ANN memiliki hasil akurasi tertinggi yaitu sebesar 85.09%. Dikarenakan *standard scaler* dengan metode *training* ANN memiliki hasil akurasi yang terbaik, maka dilakukan percobaan selanjutnya dengan jumlah *hidden layer* yang berbeda-beda. Perbandingan akurasi dengan jumlah *hidden layer* yang berbeda dapat dilihat pada Tabel 2. Setelah melakukan perbandingan akurasi antara jumlah *hidden layer* yang berbeda, maka dapat dilihat bahwa hasil akurasi dengan *hidden layer* berjumlah 3 menghasilkan akurasi tertinggi yaitu 85.27%, di mana akurasi ini tidak terlalu jauh berbeda dengan jumlah *hidden layer* lainnya.

**Tabel 2.** Perbandingan Akurasi Antar Hidden Layer ANN

<b>Jumlah Hidden Layer</b>	<b>Akurasi (%)</b>
3	85.27
4	85.23
5	85.09

**5. KESIMPULAN**

Dari percobaan yang telah dilakukan sebelumnya, maka dapat disimpulkan bahwa sistem yang dibangun mampu memprediksi *rating* film dengan baik, berdasarkan data yang ada di situs *imdb.com*. Hal ini dapat dilihat dengan nilai akurasi terbaik adalah 85.27%, dengan menggunakan metode *Standard Scaler* sebagai *preprocessing* dan ANN sebagai metode *training*, dengan jumlah *hidden layer* adalah 3.

Sebagai saran untuk penelitian berikutnya, dapat dicoba untuk melakukan *feature selection* atau *dimensionality reduction* terhadap atribut *dataset* yang digunakan. Harapannya dengan menambahkan hal tersebut dapat meningkatkan nilai akurasi dari sistem. Selain itu, tipe *encoding* yang digunakan pada data kategori dapat dicoba untuk menggunakan tipe lain seperti *One Hot Encoding* sebagai pembanding.

**DAFTAR PUSTAKA**

- [1] A. K. Pratama, Analisis Konten Film Anak Jalanan Di RCTI, Palembang: UIN Raden Patah, 2018.
- [2] A. D. Kesuma, Faktor-Faktor Yang Mempengaruhi Kesuksesan Film, Medan: Universitas Sumatera Utara, 2018.
- [3] I. D. A P, Pengaruh Rating Dalam Situs IMDb Terhadap Keputusan Menonton Di Kota Bandung, Bandung: Telkom University, 2016.

- [4] I. N. da Silva, D. H. Spatti, R. A. Flauzino, L. H. B. Liboni and S. F. d. R. Alves, *Artificial Neural Network Architectures and Training Processes*, Springer, 2016.
- [5] R. A. Abarja and A. Wibowo, "Movie Rating Prediction using Convolutional Neural Network based on Historical Values," *International Journal of Emerging Trends in Engineering Research*, vol. 8, no. 5, pp. 2156-2164, 2020.
- [6] S. YuMin, Z. Yuan and Y. JinYao, "Neural Network Based Movie Rating Prediction," in *2018 International Conference on Big Data and Computing*, New York, 2018.
- [7] K. I. Asad, T. Ahmed and M. S. Rahman, "Movie popularity classification based on inherent movie attributes using C4.5, PART and correlation coefficient," in *2012 International Conference on Informatics, Electronics & Vision (ICIEV)*, Dhaka, 2012.
- [8] P. Y. Hsu, Y. H. Shen and X. A. Xie, "Predicting Movies User Ratings with Imdb Attributes," in *International Conference on Rough Sets and Knowledge Technology*, Shanghai, 2014.
- [9] V. R. Prasetyo, "Searching Cheapest Product on Three Different E-Commerce Using K-Means Algorithm," in *2018 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, Bali, 2018.
- [10] L. Richardson, "beautifulsoup4 4.10.0," [Online]. Available: <https://pypi.org/project/beautifulsoup4/>. [Accessed 30 August 2021].
- [11] V. R. Prasetyo, B. Hartanto and A. A. Mulyono, "Penentuan Pembimbing Tugas Akhir Mahasiswa Jurusan Teknik Informatika Universitas Surabaya Dengan Metode Dice Coefficient," *TEKNIKA*, vol. 8, no. 1, pp. 44-51, 2019.
- [12] S. K. Patro and K. K. Sahu, "Normalization: A Preprocessing Stage," *International Advanced Research Journal in Science, Engineering and Technology*, vol. 2, no. 3, pp. 20-22, 2015.
- [13] D. A. Nasution, H. H. Khotimah and N. Chamidah, "Perbandingan Normalisasi Data Untuk Klasifikasi Wine Menggunakan Algoritma K-NN," *CESS (Journal of Computer Engineering System and Science)*, vol. 4, no. 1, pp. 78-82, 2019.
- [14] A. Ambarwati, Q. J. Adrian and Y. Herdiyeni, "Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman," *Jurnal RESTI*, vol. 4, no. 1, pp. 117-122, 2020.
- [15] A. Chamsudin, "Implementasi Neural Network Untuk Memprediksi Jumlah Penderita Tuberculosis," *Emitor: Jurnal Teknik Elektro*, vol. 16, no. 1, pp. 43-50, 2016.
- [16] A. R. M. Izati and H. B. Notobroto, "Penerapan Metode Artificial Neural Network Dalam Peramalan Jumlah Kunjungan Ibu Hamil (K4)," *Jurnal Biometrika dan Kependudukan*, vol. 8, no. 1, pp. 11-20, 2019.
- [17] A. Kurniadi, "Implementasi Convolutional Neural Network Untuk Klasifikasi Varietas Pada Citra Daun Sawi Menggunakan Keras," *Journal of Computer and Information Technology*, vol. 4, no. 1, pp. 25-33, 2020.
- [18] T. M. S. Mulyana, D. Widyaningrum and Herlina, "OCR Huruf Jawa Dengan Fitur Kode Rantai Dan Levenshtein Distance," *Jurnal Ilmiah NERO*, vol. 6, no. 1, pp. 64-73, 2021.