
PREDIKSI RATING FILM ANIMASI BERDASARKAN ELEMEN MISE EN SCENE MENGGUNAKAN NEURAL NETWORK

Odi Isya Winanda^{1*}, Selly Artaty Zega^{2**}, Raihan Hilmawan^{3**}

* Teknik Informatika, Politeknik Negeri Batam

** Teknik Multimedia Jaringan, Politeknik Negeri Batam

odi.isya17@gmail.com¹, selly@polibatam.ac.id², raihan.hilmawanbikers@gmail.com³

ABSTRACT

A production house can make adjustments in releasing the film if knows the possibility of success to the related film. The goal is to get maximum profit after the film is released. It can even use a prediction to know how market developments are. In this study, the authors used data from IMDb to predict the rating on a film. Because IMDb is the largest database consisting of relevant and comprehensive information related to films. It is necessary to consider the additional features of visual, which basically refers to everything that appears on the screen. Everything captured by the camera consisting of settings, lighting, the human figure, and composition, these four elements are part of the mise en scene. To determine the level of accuracy in film rating predictions, the author uses the science of soft computing, namely the neural network, focusing on the prediction part. The accuracy of the animated movie rating prediction is based on the mise en scene element which is equal to 44%, the accuracy is based on the average of each test option.

Keyword:

Film Rating, IMDb, Mise En Scene, Neural Network

I. PENDAHULUAN

Industri film adalah industri multi-miliar dolar, setiap tahunnya dapat menghasilkan pendapatan sekitar sepuluh miliar dolar. Bahkan tiap filmnya bisa menghasilkan lebih dari satu miliar dolar. Rumah produksi biasanya hanya mengandalkan beberapa film yang sangat mahal disetiap tahunnya, untuk memastikan tetap mendapatkan keuntungan. Diperkirakan sekitar 80% dari keuntungan industri film selama dekade terakhir, dihasilkan dari 6% film yang dirilis, dan 78% film telah kehilangan uang dari periode waktu yang sama (Darin Im and Minh Thao Nguyen, 2011).

Di Amerika Serikat, 1000 film dirilis setiap tahunnya. Sejak tahun 1920-an, industri film Amerika telah meraup lebih banyak uang per tahunnya dibandingkan dengan negara lain. Miliaran dolar yang dihabiskan hanya untuk iklan saja, kampanye iklan berkontribusi besar terhadap total anggaran film. Terkadang hasil investasi merugi cukup besar bagi produsen (Nithin Vr, 2014). Mencoba menemukan cara untuk memproduksi film yang lebih memungkinkan untuk berhasil. "Hollywood adalah lahan dugaan dan tebakan liar," jadi sulit untuk memprediksi apakah film dengan anggaran tinggi akan menghasilkan untung yang maksimal (Darin Im dan Minh Thao Nguyen, 2011).

Warner Brothers, salah satu rumah produksi terbesar mengalami penurunan pendapatan tahun lalu, Sebuah rumah

produksi bisa melakukan penyesuaian dalam merilis filmnya, jika mengetahui lebih dulu kemungkinan keberhasilan terkait film tersebut. Tujuannya untuk mendapatkan keuntungan maksimal setelah film dirilis. Bahkan bisa menggunakan prediksi untuk tahu perkembangan pasar. Ini menunjukkan adanya kebutuhan mendesak pada perangkat lunak terkait untuk dikembangkan. Sejauh ini belum ada penelitian yang berhasil membuat model yang cukup bagus untuk digunakan di industri (Nithin Vr, 2014). Dalam studi ini, penulis menggunakan data dari IMDb untuk memprediksi *rating* pada suatu film. *Rating* film dipengaruhi oleh banyak faktor, sehingga akurasi dalam memprediksi *rating* sebuah film baru dapat menjadi sebuah tantangan. Dalam beberapa tahun terakhir, berbagai teknik analisis semantik berhasil diterapkan untuk menganalisis ulasan pengguna, yang pada gilirannya digunakan untuk memprediksi peringkat film IMDb dikarenakan IMDb adalah *database* besar yang terdiri dari informasi yang relevan dan komprehensif tentang film.

Kunci untuk meningkatkan kinerja, memerlukan pertimbangan fitur tambahannya yaitu terkait visual yang pada dasarnya mengacu pada segala sesuatu yang muncul pada layar. Segalanya yang tertangkap oleh kamera yang terdiri dari *setting*, *lighting*, *the human figure*, dan *composition* (Maria Pramaggiore dan Tom Wallis, 2005). Elemen diatas ialah bagian dari elemen yang terdapat pada *mise en scene*.

Berdasarkan pengamatan penulis, penggunaan ilmu *soft computing* yaitu *neural network* dapat membantu dalam pengolahan data pada prediksi *rating* film. Penulis akan memfokuskan dibagian *prediction*. Penggunaan *neural network* pernah diterapkan pada prediksi kesuksesan film *box-office*, yang bertujuan untuk membantu industri film animasi dalam memprediksi kisaran pendapatan yang diharapkan sebelum film dirilis (Riwinoto *et al*,2015). Sehingga penelitian ini dapat dijadikan sebagai acuan dalam penggunaan *neural network* pada prediksi *rating* film animasi layar lebar. Pada penelitian ini, akan terlihat besarnya peran sebuah media IMDb dan juga elemen *mise en scene* untuk mempengaruhi keputusan penonton dalam menentukan pilihan serta menilai kualitas film (Irvando Dityawan Alpha Putra,2016). Diharapkan dapat menjadi acuan dan juga penyelesaian masalah terkait prediksi *rating* film.

II. DASAR TEORI

A. *Mise en scene*

Mise en scene ditentukan selama pra-produksi dan produksi yang melibatkan banyak orang. Perencanaan yang hati-hati dilakukan perancang produksi yang sangat berkontribusi dalam koherensi *mise en scene*. Sutradara dan desainer produksi membuat keputusan tentang bagaimana ceritanya akan terlihat baik sebelum *shooting* dimulai.

Mengendalikan elemen-elemen mana yang harus difokuskan dan mengatur unsur-unsurnya dalam *frame* sesuai unsur estetika. Setiap elemen dari *mise en scene* yaitu *setting*, *the human figure*, *lighting*, dan *composition* mempengaruhi pengalaman audiens tentang cerita, karakter, ruang, dan waktu. Pembuat film menggunakan detail secara sistematis, terintegrasi tidak hanya untuk menciptakan dunia dilayar, tetapi juga untuk menunjukkan pengembangan karakter, menyajikan motif, memperkuat tema, dan membangun suasana hati.

1. *Setting*

Pengaturan mengacu pada tempat-tempat di mana aksi film ini terbentang. Tempat-tempat ini mungkin lokasi umum atau spesifik, tempat nyata atau imajiner. Membangun sebuah set memerlukan tingkat kontrol maksimum terkait lingkungan dan juga pengambilan gambar. Pada set didalam ruangan, sutradara dan sinematografer tidak perlu khawatir bersaing dengan buruknya cuaca, kebisingan, dan kondisi pencahayaan yang tidak dapat diandalkan.

2. *The Human Figure*

Kinerja para aktor berkontribusi banyak pada makna film. Kebanyakan film naratif menceritakan kisah tentang manusia makhluk hidup dan konflik yang dihadapi. *Casting* (pemilihan aktor), *acting style*, penempatan dan pergerakan tokoh mempengaruhi audiens dalam menanggapi karakter. Memilih aktor adalah salah satu keputusan terpenting yang harus dilakukan seorang sutradara. Pada waktu bersamaan, karena biaya aktor mewakili sebagian besar biaya produksi, dikarenakan penonton sering pergi ke bioskop untuk melihat bintang favorit, maka seorang eksekutif studio

lebih memilih untuk meminimalkan risiko dan tetap dengan "hal yang pasti" dalam hal casting.

3. *Lighting*

Cahaya adalah persyaratan penting pembuatan film. Tanpa cahaya masuk lensa kamera, tidak ada gambar yang akan direkam. *Lighting* adalah unsur *mise en scene* karena menerangi set dan aktor dan bertujuan untuk membuat suasana hati dan efek tertentu. Tetapi juga terkait dengan masalah sinematografi, karena sifat fotokimia dari stok film, penggunaan lensa dan *filter*, dan teknik pemrosesan lab semua mempengaruhi tampilan film. *Lighting* semakin menyempurnakan pemahaman penonton tentang karakter, menggarisbawahi tindakan tertentu, membangun tema, dan menetapkan suasana hati.

4. *Composition*

Elemen terakhir dari *mise en scene* ialah komposisi, yang didefinisikan sebagai pengaturan visual objek, aktor, dan ruang di dalam *frame*. Ruang *frame* dapat dianggap sebagai ruang dua dimensi, di mana terdapat prinsip-prinsip seni rupa. Satu prinsip penting adalah pastikan ada keseimbangan atau simetri di dalam bingkai. Bingkai dapat dipartisi secara horizontal, pada sumbu kiri-kanan, dan secara vertikal, dari atas ke bawah. Keseimbangan komposisi harus distribusikan merata dari daerah terang dan gelap, warna mencolok, benda dan atau *figure*.

B. *Neural Network*

Neural network adalah model matematika yang mencoba untuk mensimulasikan struktur dan fungsi jaringan saraf biologis. *Neural network* telah menunjukkan hasil yang bagus dan menjanjikan dalam sistem prediksi di berbagai *domain* aplikasi seperti pasar saham, medis, analisis media sosial dan juga prediksi *box office* untuk film. Namun, banyak model *neural network* yang berbeda, strategi desain dan variabel yang digunakan dalam proses analisis sangat mempengaruhi hasil dan akurasi dari prediksi.

Pada penelitian tersebut menggunakan *neural network three layer back-propagation* dan data yang diekstraksi dari 100 pendapatan terlaris dari film-film Amerika yang dirilis antara tahun 2010 dan 2015, untuk memprediksi dan mengklasifikasikan keberhasilan film berdasarkan pada pendapatan dan biayanya. Dalam memvalidasi pendekatan menggunakan *crossentropy validation* dan hasil akhir menunjukkan akurasi 91% dalam prediksi.

Penelitian sebelumnya dapat dijadikan sebagai acuan dalam pemilihan metode yang tepat dalam menggunakan *neural network*, sehingga dapat memudahkan dalam melakukan proses prediksi *rating* film animasi layar lebar pada penelitian ini.

C. *Multilayer Perceptron*

Multilayer Perceptron adalah topologi yang paling umum dari *neural network*, setiap *perceptron* terkoneksi membentuk beberapa lapisan (*layer*). *Multilayer perceptron* memiliki *input layer*, minimal satu *hidden layer* dan *output layer* (Negnevitsky, Michael, 2005).

TABEL I
ANALOGI JARINGAN SARAF BIOLOGIS DENGAN *NEURAL NETWORK*

| Jaringan saraf biologis | Jaringan saraf tiruan |
|-------------------------|--------------------------|
| <i>Soma</i> | Neuron |
| <i>Dendrite</i> | Masukan (<i>input</i>) |
| <i>Axon</i> | Luaran (<i>Output</i>) |
| <i>Synapse</i> | Bobot (<i>Weight</i>) |

Back-propagation ialah metode yang paling banyak digunakan dalam pembelajaran atau pelatihan *multilayer perceptron*. Terdiri dari empat langkah dari metode ini ialah inialisasi (*initialization*), aktivasi (*activation*), pelatihan bobot (*weight training*) dan iterasi (*iteration*). Pada tahapan inialisasi, nilai awal dari bobot dan ambang batas (*threshold*) ditentukan secara acak tetapi dengan batasan tertentu. Selanjutnya pada tahap aktivasi, diberikan *input* dan nilai *output* yang diharapkan (*desired output*). Lalu pada tahap pelatihan bobot, nilai *actual output* dibandingkan dengan *desired output* dan dilakukan penyesuaian bobot.

D. IMDb (Internet Movie Database)

IMDb ialah situs web yang menyediakan informasi mengenai film dari seluruh dunia, terdiri dari orang-orang yang terlibat didalamnya mulai dari aktor/aktris, sutradara, penulis sampai penata rias dan *soundtrack*. Situs web ini sekarang dimiliki oleh Amazon.com. Koleksi informasi film yang ditampilkan cukup lengkap, kita bisa melihat informasi film lama maupun film terbaru yang akan rilis di bioskop. IMDb diluncurkan pada tanggal 17 oktober 1990 dan pada tahun 1998 diakuisisi oleh Amazon.com. IMDb menyediakan data yang kaya tentang film, dan peringkat pengguna untuk film-film tersebut.

Pemilihan *website* IMDb sebagai *database* untuk data film yang digunakan pada analisis ini dikarenakan telah dilakukannya perbandingan dengan *website* lain yaitu *the Open Movie Database* (OMDb API), dari kedua *website* tersebut IMDb memiliki informasi yang lebih lengkap dibanding OMDb API, sehingga memungkinkan penulis mengumpulkan data film animasi layar lebar dan memudahkan dalam pengumpulan atribut dari elemen *mise en scene*.

E. MATLAB

Nama MATLAB adalah singkatan dari MATrix LABoratory. MATLAB awalnya ditulis untuk menyediakan akses yang mudah ke perangkat lunak matriks yang dikembangkan oleh LINPACK (paket sistem linier) dan proyek EISPACK (paket sistem Eigen). MATLAB adalah bahasa berkinerja tinggi untuk komputasi. Mengintegrasikan komputasi, visualisasi, dan lingkungan pemrograman. Memiliki struktur data yang canggih, berisi alat pengeditan dan debugging bawaan, dan mendukung pemrograman berorientasi objek.

Matlab memiliki rutin *built-in* yang kuat yang memungkinkan berbagai macam perhitungan. Aplikasi

spesifik dikumpulkan dalam paket yang disebut sebagai kotak peralatan. Ada kotak peralatan untuk pemrosesan sinyal, perhitungan simbolis, teori kontrol, simulasi, optimasi, dan beberapa bidang ilmu terapan dan teknik lainnya.

Pada penelitian ini matlab digunakan dalam proses ekstraksi gambar, menghitung dan menampilkan nilai RGB (*red, green, blue*), proses ini dilakukan hanya pada atribut di salah satu elemen *mise en scene* yaitu *lighting*. Data tersebut akan dikumpulkan dengan data atribut dari elemen lainnya.

F. WEKA

Weka adalah kumpulan algoritma pembelajaran untuk memecahkan masalah data *mining*. Weka ditulis menggunakan *java* dan berjalan di hampir semua *platform*. Algoritma dapat diterapkan langsung ke kumpulan data atau dipanggil.

Weka adalah *front-end* TCL / TK untuk algoritma pemodelan (sebagian besar pihak ketiga) yang diimplementasikan dalam bahasa pemrograman lain, ditambah utilitas *preprocessing* data dalam C, dan sistem berbasis *makefile* untuk menjalankan eksperimen pembelajaran. Versi asli ini dirancang sebagai alat untuk menganalisis data dari domain pertanian, tetapi versi berbasis *java* sepenuhnya baru-baru ini (Weka 3), yang pembangunannya dimulai pada tahun 1997, sekarang digunakan di banyak bidang aplikasi yang berbeda, khususnya untuk pendidikan dan penelitian.

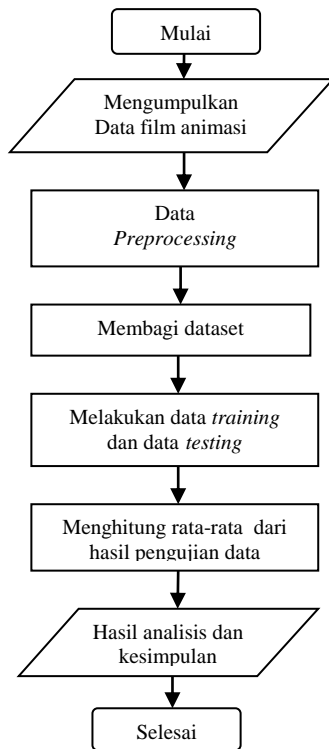
Keuntungan dari weka termasuk:

1. Ketersediaan gratis di bawah Lisensi Publik Umum GNU
2. Portabilitas, karena itu sepenuhnya diimplementasikan dalam bahasa pemrograman *java* dan dengan demikian berjalan hampir setiap *platform* komputasi *modern*.
3. Sebuah koleksi komprehensif teknik *preprocessing* dan pemodelan data
4. Mudah digunakan, dari segi antarmuka pengguna grafisnya.

Pada penelitian ini weka digunakan untuk membagi data menjadi dua bagian, yaitu data *training* dan data *testing*. Kedua data tersebut berisi data dari 125 film animasi layar lebar yang akan di analisis berdasarkan elemen *mise en scene*. Tujuan pembagian data ini ialah memudahkan penulis dalam proses analisis nantinya.

III. PERANCANGAN PENELITIAN

Dalam penelitian ini, terdapat beberapa kebutuhan dan parameter yang perlu diidentifikasi agar proses pembuatan berjalan dengan lancar dan sesuai dengan hasil yang diharapkan. Adapun perancangan yang digunakan sebagai tahapan dalam proses penelitian ini dapat lihat sebagai berikut :



Gambar 1. Flowchart Perancangan Penelitian

A. Mengumpulkan data film animasi

Pada tahap pertama yang perlu disiapkan ialah data film yang akan di analisis. Berisi film animasi layar lebar yang di rilis tahun 2010 hingga 2018, penentuan tahun rilis berdasarkan data yang didapat dari *website* IMDb yaitu *Top Animated Movies: 2000-2018* dan *The 150 Best Animated Movies of The 2010s*. Pada kedua data tersebut akan dilakukan proses modifikasi data, menghapus data, penggabungan data, generalisasi data dan normalisasi data. Dari proses tersebut sehingga menghasilkan data sebanyak 125 film animasi layar lebar yang terdiri dari kelas dan atribut yang siap untuk diolah pada tahap selanjutnya.

B. Data preprocessing

Berdasarkan data film yang sudah di dapat sebelumnya, penulis akan melakukan observasi langsung terhadap film untuk dikumpulkan atribut dari elemen *mise en scene*, maupun data yang didapat dari *website* IMDb. Tahapan dari data *preprocessing* dimulai dari data *cleaning*, data *integration*, data *transformation*.

Data preprocessing ialah proses yang dilakukan diawal yang bertujuan untuk membantu dalam membuat data mentah menjadi data yang berkualitas dalam arti mencegah kesalahan dalam *data mining*. Dalam proses membuat data tidak menutup kemungkinan akan terjadinya *incomplete*, *noisy* dan *inconsistent*. Maka dari itu pada penelitian ini penulis

menerapkan *data preprocessing* dalam membuat data, karena hasil yang berkualitas berdasarkan dari data yang berkualitas.

1. Data Cleaning

Penerapan *data cleaning* pada penelitian ini digunakan pada pengumpulan data film animasi layar lebar dan juga pengumpulan atribut *mise en scene*, dikarenakan pada dua tahap tersebut proses pengumpulan data berasal dari sumber yang berbeda. Mengacu pada identifikasi tidak lengkap, maupun tidak relevan, kemudian dilakukan proses modifikasi atau menghapus data. Proses *data cleaning* dilakukan pada tahap berikut:

1. Data film yang didapat dari *website* IMDb terdiri dari dua data yaitu *Top Animated Movies: 2000-2018* dan *The 150 Best Animated Movies of The 2010s*. Dari dua data tersebut penulis melakukan proses *data cleaning* pada tiap film berdasarkan *title type* dan tahun rilisnya, berikut tabel yang berisi beberapa data film yang harus dilakukan proses *data cleaning* pada *title type*:

TABEL II
DATA FILM ANIMASI LAYAR LEBAR

| Posisi | Judul | Title Type | Tahun |
|--------|---------------------------------|---------------------|-------|
| 19 | Ping Pong the Animation | <i>tvMiniSeries</i> | 2014 |
| 22 | Sanzoku no musume Rōnya | <i>tvSeries</i> | 2014 |
| 41 | Over the Garden Wall | <i>tvMiniSeries</i> | 2014 |
| 64 | Adventure Time with Finn & Jake | <i>tvSeries</i> | 2010 |
| 85 | Sidonia no Kishi | <i>tvSeries</i> | 2014 |
| 89 | The Animatrix | <i>video</i> | 2003 |
| 96 | Moonbeam City | <i>tvSeries</i> | 2015 |
| 103 | World of Tomorrow | <i>short</i> | 2015 |

Berdasarkan tabel diatas proses data cleaning yang dilakukan ialah pengelompokan data berdasarkan *title type* diluar dari *movie* atau film layar lebar, sehingga *title type* seperti *tvMiniSeries*, *tvSeries*, *video* dan *short* akan dihapus, sehingga film pada kategori tersebut tidak digunakan dalam proses pengolahan data nantinya.

2. Proses *data cleaning* selanjutnya ialah berdasarkan tahun rilis dari sebuah film, pada penelitian ini film yang digunakan dengan range waktu dari tahun 2010 hingga 2018. Berikut tabel yang berisi beberapa data film yang harus dilakukan proses *data cleaning* pada tahun rilisnya:

TABEL III
DATA FILM ANIMASI LAYAR LEBAR

| Posisi | Judul | Title Type | Tahun |
|--------|---------------------------------|------------|-------|
| 31 | Ratatouille | movie | 2007 |
| 32 | Millennium Actress | movie | 2001 |
| 34 | Shrek | movie | 2001 |
| 35 | The Girl Who Leapt Through Time | movie | 2006 |

Berdasarkan data pada tabel, tahun rilis dari tiap filmnya akan dikelompokkan berdasarkan *range* waktu yang telah ditetapkan, maka film dengan tahun rilis sebelum tahun 2010 tidak digunakan dalam proses pengolahan data nantinya.

Setelah dilakukan proses *data cleaning* maka akan didapat data film yang sesuai dengan kategori yang diinginkan, berikut beberapa data film animasi layar lebar yang digunakan untuk proses pengolahan data selanjutnya:

TABEL IIIV
DATA FILM ANIMASI LAYAR LEBAR

| Posisi | Judul | Title Type | Tahun |
|--------|---------------------------------|------------|-------|
| 1 | Coco | movie | 2017 |
| 2 | Your Name. | movie | 2016 |
| 3 | Toy Story 3 | movie | 2010 |
| 4 | Inside Out | movie | 2015 |
| 5 | Wolf Children | movie | 2012 |
| 6 | A Silent Voice | movie | 2016 |
| 7 | Incredibles 2 | movie | 2018 |
| 8 | Song of the Sea | movie | 2014 |
| 9 | How to Train Your Dragon | movie | 2010 |
| 10 | The Tale of the Princess Kaguya | movie | 2013 |

2. Data Integration

Proses penggabungan data dari sumber yang berbeda atau sumber data yang tidak hanya dari satu tempat disebut dengan *data integration*, proses tersebut juga diterapkan pada penelitian ini diantaranya dalam penggabungan data yang diperoleh dari IMDb dan observasi langsung. Diantaranya:

1. Atribut yang diperoleh dari *website* IMDb akan dikumpulkan pada *microsoft excel* seperti *human figure*, aspek rasio dan rating.
2. Data yang diperoleh dari observasi langsung juga akan digabungkan dengan atribut lainnya di *microsoft excel*, seperti *setting* dan *lighting*.

Data dari dua aplikasi yang berbeda juga akan digabungkan terhadap atribut lain, misalnya pada aplikasi matlab. Diantaranya:

1. Data hasil pengolahan warna gambar (RGB), menggunakan aplikasi matlab maka penulis harus memindahkan data hasil olahan pada *file excel* yang sama.

Penggabungan data dari IMDb dan observasi langsung menghasilkan tabel sebagai berikut:

TABEL V
ATRIBUT ELEMEN MISE EN SCENE DAN RATING

| No/ID | Judul | setting | Human Figure | Aspek Ratio | Lighting | Rating |
|-------|-------|---------|--------------|-------------|----------|--------|
| | | | | | | |

Tabel diatas digunakan untuk pengumpulan data atribut dari elemen *mise en scene*, sehingga membantu dalam pengolahan data menjadi data yang berkualitas dan data tersebut dapat digunakan untuk proses selanjutnya.

3. Data Transformation

Pada penelitian ini *data transformation* cukup berperan penting dalam proses membuat maupun menambahkan data. Karena dibutuhkannya beberapa tahap yang harus dilakukan yaitu *generalization* dan *normalization*. Penerapan tahapan tersebut digunakan pada saat menggabungkan nilai dari sebuah atribut dalam satu kategori, dan menormalisasikan semua nilai atribut pada data set menjadi numerik.

Data transformation diterapkan pada proses generalisasi data dan pembentukan atribut atau fitur berdasarkan data dari IMDb dan observasi langsung. Dikarenakan data yang diperoleh pada saat observasi langsung terhadap elemen *mise en scene* cenderung tidak relevan, maka dilakukannya proses normalisasi data.

Normalisasi data dilakukan terhadap atribut dari elemen *mise en scene*, diantaranya:

1. Setting

Pada elemen ini penulis membutuhkan data jumlah *setting* pada tiap film, penulis harus melakukan proses *screenshots* terlebih dahulu. Proses *screenshots* dilakukan setiap adanya pergantian *setting* pada film.

Setelah semua hasil *screenshots* dari tiap *setting* diperoleh selanjutnya dihitung jumlahnya dan data tersebut dipindahkan pada *microsoft excel*.

TABEL VI
ATRIBUT ELEMEN SETTING

| NO | Judul | Setting |
|----|-------------|---------|
| 1 | Zootopia | 27 |
| 2 | Isle of Dog | 22 |

Tahap selanjutnya ialah pembagian kelas interval pada atribut *setting*, proses ini diperlukan karena pada atribut *setting* memiliki cukup banyak data yang berbeda-beda tiap filmnya, maka perlu untuk mengelompokkan atribut tersebut menjadi beberapa kelas, sehingga menghindari hasil yang *error* atau bias pada aplikasi *weka* nantinya. Dalam menentukan kelas interval penulis menggunakan rumus :

$$k = 1 + 3,3 \log n$$

Rumus menentukan kelas interval

Keterangan :

k : Banyaknya kelas interval

n : Banyaknya data

Untuk mendapatkan nilai jangkauan menggunakan rumus sebagai berikut :

$$\text{Datum terbesar} - \text{Datum terkecil}$$

Keterangan :

Datum Terbesar : Nilai terbesar

Datum terkecil : Nilai terkecil

Selanjutnya menentukan panjang interval kelas menggunakan rumus sebagai berikut:

$$c = J/k$$

Keterangan :

c : Panjang interval kelas

J : Jangkauan

k : Banyak kelas interval

Selanjutnya dari hasil penentuan kelas interval maka akan di dapat tabel distribusi frekuensi sebagai berikut:

TABEL VII
DISTRIBUSI FREKUENSI ELEMEN SETTING

| Kelas Interval | Nilai Statistik |
|----------------|-----------------|
| 1 | 1 - 4 |
| 2 | 5 - 8 |
| 3 | 9 - 12 |
| 4 | 13 - 16 |
| 5 | 17 - 20 |
| 6 | 21 - 24 |
| 7 | 25 - 28 |
| 8 | 29 - 33 |
| 9 | 34 - 37 |

Setelah didapatkan kelas interval, maka selanjutnya dilakukan proses pengelompokan pada atribut *setting* di microsoft excel, Selain mendapatkan jumlah *setting*, gambar dari hasil *screenshots* juga digunakan dalam proses ekstraksi gambar untuk mendapatkan nilai RGB.

2. Lighting

Pada penelitian ini penulis menggunakan film animasi layar lebar sebagai bahan penelitian, untuk mengetahui informasi terkait lighting pada sebuah film penulis akan melakukan ekstraksi gambar dari setiap pergantian setting pada film. Adapun penelitian sebelumnya yang digunakan sebagai acuan dalam menentukan lighting melalui warna pada setiap scene, diantaranya penelitian yang dilakukan oleh Navenka Dimitrova dkk, pada penelitiannya menganalisis transformasi warna menjadi sebuah suasana (mood), dengan warna dominan yang ditentukan menggunakan RDC (*representative dominant Color*) dan untuk keseluruhan film mereka menggunakan MPH (*movie palette histogram*) dengan mengumpulkan statistik warna palet. Dengan menggunakan *color palette* dan *mood dynamic* mereka

melakukan pendekatan klasifikasi dan menghasilkan akurasi sekitar 80% dan juga pada film dengan genre yang sama memiliki MPH (*movie palette histogram*) serupa.

Penelitian selanjutnya mengenai warna sebagai *look* dan *mood* pada videografi film televisi pancer, pada penelitian ini menyimpulkan melalui warna penonton dapat menangkap gambar yang memiliki *look* (nuansa) dan merasakan melalui *mood* (suasana) yang dibangun dari setiap adegannya maupun untuk efek tertentu (mandella majid prachihara). Oleh karena itu penulis melakukan ekstraksi gambar dan pengolahan warna gambar untuk mencari nilai *red*, *green*, *blue* (RGB) dari tiap filmnya. Gambar yang akan diolah ialah gambar yang didapat dari hasil *screenshots* beberapa *scene* dari atribut elemen *setting*. Aplikasi yang digunakan untuk mengekstraksi gambar ialah matlab.



Gambar 2. GUI dari aplikasi pengolahan warna gambar

Setelah nilai RGB didapatkan, tahap selanjutnya dilakukan proses input data pada *microsoft excel*, kemudian dihitung nilai rata-rata dari keseluruhan *scene* pada film tersebut sehingga memudahkan saat proses penggabungan atribut dengan elemen yang lain.

TABEL VIII
ATRIBUT ELEMEN LIGHTING

| NO | Judul | Lighting | x1 | x2 | x3 | x4 | x-n | Rata- Rata |
|----|-------|----------|----|----|----|----|-----|------------|
| | | R | | | | | | |
| | | G | | | | | | |
| | | B | | | | | | |

Karena nilai yang akan diperoleh cenderung unik dan bias, pada elemen ini penulis juga menentukan kelas interval dan tabel distribusi frekuensi, dengan tabel sebagai berikut:

TABEL IX
DISTRIBUSI FREKUENSI ELEMEN LIGHTING

| Kelas Interval | Nilai Statistik |
|----------------|-----------------|
| 1 | 35 - 51 |
| 2 | 52 - 68 |
| 3 | 69 - 85 |
| 4 | 86 - 102 |
| 5 | 103 - 119 |
| 6 | 120 - 136 |
| 7 | 137 - 153 |
| 8 | 154 - 170 |
| 9 | 171 - 187 |

3. The Human figure

Pada elemen ini penulis memerlukan data terkait tingkat terkenalnya seorang aktor, karena peran aktor sangat mempengaruhi audiens dalam menonton sebuah film. Untuk memperoleh data tersebut penulis menggunakan *website* IMDb, karena IMDb menyediakan informasi mengenai film dari seluruh dunia, maupun orang-orang yang terlibat didalamnya termasuk aktor/aktris.

Penulis menggunakan informasi terkait aktor dengan menampilkan tiga aktor teratas. Penulis menggunakan data tersebut pada setiap filmnya. Data yang digunakan ialah jumlah penghargaan yang dimiliki setiap aktor.

TABEL X
SKOR AKTOR

| Skor | Keterangan |
|------|-------------------------|
| 5 | Tidak masuk nominasi |
| 10 | Masuk nominasi |
| 15 | Memenangkan penghargaan |

Setelah skor didapat pada setiap aktor, data tersebut akan dikumpulkan dalam satu *microsoft excel* yang nantinya skor tersebut akan dijumlahkan perfilmnya dengan tabel sebagai berikut:

TABEL XI
ATRIBUT ELEMEN THE HUMAN FIGURE

| NO | Judul | | Aktor | | Jumlah |
|----|-------|---|-------|-----|--------|
| | | I | II | III | |
| | | | | | |

4. Composition

Pada elemen ini memerlukan data terkait *aspect ratio* pada film. Data tersebut diperoleh dari *website* IMDb. Berdasarkan data film yang sudah didapat, penulis melakukan pengelompokan terhadap aspek rasio pada setiap filmnya.

TABEL XII
KATEGORI ASPEK RASIO

| KATEGORI | ASPEK RASIO |
|----------|-------------|
| 1 | 1.33 : 1 |
| 2 | 1.66 : 1 |
| 3 | 1.78 : 1 |
| 4 | 1.85 : 1 |
| 5 | 2.24 : 1 |
| 6 | 2.35 : 1 |
| 7 | 2.39 : 1 |

5. Rating

Proses *data transformation* yang terakhir yaitu data tentang *rating* dari sebuah film. Untuk mendapatkan data tersebut penulis menggunakan *website* IMDb sebagai sumbernya. sehingga penulis dapat menambahkan langsung pada file yang diinginkan.

Pada atribut *rating* juga perlu menentukan kelas interval, karena nilai yang diperoleh cenderung unik dan bias, pada elemen ini penulis juga menentukan kelas interval dan tabel distribusi frekuensi, dengan tabel sebagai berikut:

TABEL XIII
DISTRIBUSI FREKUENSI RATING

| Kelas Interval | Nilai Statistik |
|----------------|-----------------|
| H | 4.6 - 5.1 |
| G | 5.2 - 5.7 |
| F | 5.8 - 6.3 |
| E | 6.4 - 6.9 |
| D | 7.0 - 7.5 |
| C | 7.6 - 8.1 |
| B | 8.2 - 8.7 |
| A | 8.8 - 9.3 |

C. Hasil Akhir

Berdasarkan hasil data *preprocessing*, adapun hasil akhir yang didapat berupa gabungan data dari *website* IMDb dan observasi langsung, Adapun beberapa data dari 125 film terdapat pada tabel XIV. Kedua data tersebut ialah hasil akhir dari atribut elemen *mise en scene*.

TABEL XIV
 ATRIBUT ELEMEN *MISE EN SCENE*

| SETTING | HUMAN FIGURE | COMPOSITION | | LIGHTING | | RATING |
|---------|--------------|-------------|---|----------|---|--------|
| | | | R | G | B | |
| 2 | 45 | 2.35:1 | 4 | 3 | 3 | B |
| 3 | 35 | 1.78:1 | 6 | 6 | 6 | B |
| 3 | 45 | 1.66:1 | 4 | 3 | 2 | B |
| 4 | 45 | 1.85:1 | 6 | 5 | 5 | B |
| 5 | 45 | 1.85:1 | 5 | 5 | 5 | B |

IV. HASIL DAN PEMBAHASAN

A. Hasil

Pada tahap ini akan dilakukan proses *data training* menggunakan dataset berisi data 125 film animasi layar lebar yang telah dikelompokkan berdasarkan elemen *mise en scene* dan telah di *preprocessing* sebelumnya. Selanjutnya dataset tersebut akan dibagi menjadi empat *test option* dengan persentase yang berbeda-beda, yang mana proses tersebut akan dilakukan pada aplikasi weka.

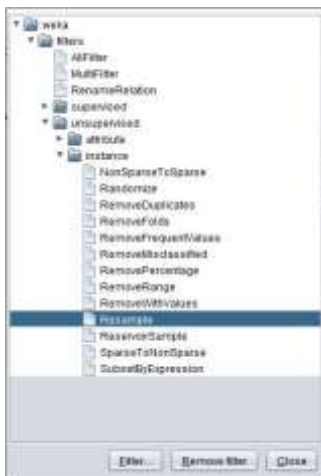
1. Teknik Membagi Dataset

Pada tahap ini dataset yang digunakan ialah *file* dataset *full.arff* yang sudah disimpan pada proses sebelumnya, untuk membagi data penulis akan menggunakan *unsupervised filter* yaitu *resample* yang telah tersedia pada aplikasi weka.

Penulis akan membagi file tersebut menjadi empat persentase yang berbeda diantaranya ialah:

1. 70% data *training* dan 30% data *testing*
2. 30% data *training* dan 70% data *testing*
3. 50% data *training* dan 50% data *testing*
4. 30%, 50% dan 70% dari dataset full

Pada tahap membagi data penulis menerapkan *unsupervised filter* pada aplikasi weka yaitu *resample* dalam membagi data sesuai dengan persentase yang diinginkan.



Gambar 3. *Unsupervised filter* pada aplikasi weka

2. Melakukan Data Training dan Data Testing

Pada tahap ini dilakukan proses pengolahan data menggunakan aplikasi weka, data yang diolah ialah dataset yang telah dibagi berdasarkan persentase yang telah ditetapkan. Data akan diolah menggunakan *classifier* dan akan diuji menggunakan *test option* yang tersedia pada aplikasi weka. Hasil olahan data berupa tingkat akurasi dan juga *neural network*.

Pada aplikasi weka terdapat empat *classifier* yang dapat digunakan untuk mengevaluasi sesuai dengan opsi yang ditetapkan dengan mengklik di kotak *test option*, empat metode uji tersebut ialah:

1. *Use training set*, pengklasifikasi yang mengevaluasi pada seberapa baik ia memprediksi kelas *instance* yang dilatihnya.
2. *Supplied test set*, pengklasifikasi yang mengevaluasi pada seberapa baik ia memprediksi kelas dari *set instance* yang dimuat dari file. Mengklik tombol *set* menampilkan dialog yang memungkinkan pengguna memilih file yang akan diuji.
3. *Cross-validation*, pengklasifikasi yang mengevaluasi dengan validasi silang, menggunakan jumlah lipatan yang dimasukkan dalam bidang teks lipatan.
4. *Percentage split*, pengklasifikasi yang mengevaluasi pada seberapa baik ia memprediksi persentase tertentu dari data yang digunakan untuk pengujian. Jumlah data yang disimpan tergantung pada nilai yang dimasukkan di kolom persentase.



Gambar 4. *Test option* pada aplikasi weka

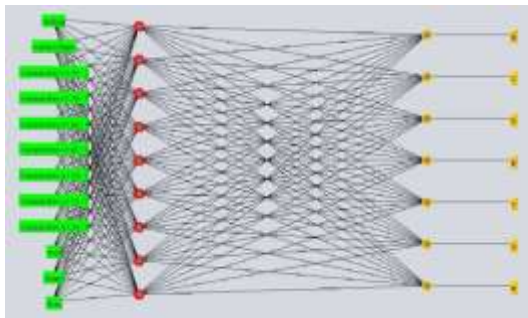
Tahap selanjutnya ialah memilih *classifier* yang akan digunakan untuk pengujian dataset, karena pada penelitian ini akan menggunakan *neural network* sebagai hasil akhirnya, maka penulis menggunakan *classifier* yang bernama *multilayer perceptron*.



Gambar 5. Menu *classifier* pada aplikasi weka

B. Penerapan Neural Network

Pada pengujian ini tidak hanya mendapatkan hasil akurasi tetapi juga menampilkan hasil dari jaringan saraf tiruan (*neural network*) yang didapat pada setiap pengujian. *Perceptron* terhubung membentuk beberapa *layer* (lapisan). Sebuah *multilayer perceptron* mempunyai *input layer* (lapisan masukan), minimal satu *hidden layer* (lapisan tersembunyi), dan *output layer* (lapisan luaran). berikut tampilan *neural network* berdasarkan hasil pengujian dataset:



Gambar 4.49 Hasil *neural network* pada aplikasi weka

Pada gambar posisi *input layer* berada pada sisi paling kiri, yaitu *node* yang berwarna hijau, *input layer* berisi atribut dari dataset. Sedangkan *hidden layer* ialah *node* yang berwarna merah, berada diantara *input layer* dan *output layer*. Posisi *output layer* berada pada sisi paling kanan gambar yaitu *node* yang berwarna kuning, pada *output layer* berisi kelas yang diprediksi. Berdasarkan gambar *input layer* memiliki 12 *neuron*, pada *hidden layer* memiliki 9 *neuron* dan pada *output layer* memiliki 7 *neuron*.

Selain itu informasi yang dapat diambil berdasarkan gambar ialah pada tiap *layer* memiliki bobot (*weight*), terdapat 9 *weight* diantara *input layer* dan *hidden layer*, dan diantara *hidden layer* dan *output layer* terdapat 7 *weight*.

C. Hasil Uji Validasi

Pada tabel dibawah terdapat hasil uji validasi data yang telah dikelompokkan berdasarkan *test option*.

TABEL XV
HASIL RATA-RATA MEAN ABSOLUTE ERROR

| Jenis Error | UTS | STS | 10 CROSS | PS |
|---------------------|--------|--------|----------|--------|
| Mean Absolute Error | 0.0989 | 0.1956 | 0.1997 | 0.2111 |

Penggunaan nilai *mean absolute error* dalam uji validasi data bertujuan sebagai ukuran kualitas prediksi, nilai MAE yang kecil menunjukkan akurasi prediksi model yang diteliti semakin akurat. Pada penelitian ini nilai MAE didapat secara otomatis berdasarkan proses pengolahan data pada aplikasi weka.

Berdasarkan tabel diatas nilai MAE terendah didapat pada *test option* "use training set" yaitu sebesar 0,0989 dan nilai MAE terbesar berjumlah 0.2111 didapat pada *test option* "percentage split".

D. Hasil Uji Correct/Incorrect Classified Instances

Performansi yang diukur adalah persentase akurasi dari hasil klasifikasi. Berdasarkan hasil akurasi yang didapat setiap persentase dari *test option*, penulis akan menghitung rata-ratanya untuk mendapatkan hasil akhir dari tingkat akurasi.

Tahap pertama penulis akan menghitung rata-rata dari tiap *test option*, berikut hasil yang didapat:

1. Use Training Set

TABEL XVI
HASIL RATA-RATA PADA USE TRAINING SET

| | FULL | 70% | 50% | 30% | Rata-rata |
|-----------|------|-----|-----|-----|-----------|
| Correct | 72% | 74% | 85% | 86% | 79.25% |
| Incorrect | 28% | 26% | 15% | 14% | 20.75% |

2. Supplied Test Set

TABEL XVII
HASIL RATA-RATA PADA SUPPLIED TEST SET

| | 70% - 30% | 30%-70% | 50%-50% | Rata-rata |
|-----------|-----------|---------|---------|-----------|
| Correct | 47% | 32% | 35% | 38% |
| Incorrect | 53% | 68% | 65% | 62% |

3. 10 Cross Validation

TABEL XVIII
HASIL RATA-RATA PADA 10 CROSS VALIDATION

| | Full | 70% | 50% | 30% | Rata-rata |
|-----------|------|-----|-----|-----|-----------|
| Correct | 39% | 32% | 39% | 24% | 33.5% |
| Incorrect | 61% | 68% | 61% | 76% | 66.5% |

4. Percentage Split

TABEL XIX
HASIL RATA-RATA PADA PERCENTAGE SPLIT

| | 70% | 66% | 50% | 30% | Rata-rata |
|-----------|-----|-----|-----|-----|-----------|
| Correct | 22% | 38% | 24% | 24% | 27% |
| Incorrect | 78% | 62% | 76% | 76% | 73% |

Dari hasil akurasi yang diperoleh nilai *correct* dan *incorrect* dari tiap *test option* penulis melakukan perhitungan rata-ratanya, adapun hasil yang didapat ialah sebagai berikut:

TABEL XX
HASIL RATA-RATA PADA KESELURUHAN

| | UTS | STS | 10 CROSS | PS | Rata-rata |
|-----------|--------|-----|----------|-----|-----------|
| Correct | 79.25% | 38% | 33.5% | 27% | 44% |
| Incorrect | 20.75% | 62% | 66.5% | 73% | 56% |

Hasil perbandingan uji *correctly/incorrectly classified instances* diperoleh persentase 79.25% untuk *instances* menggunakan *test option "use training set"* hasil tersebut merupakan yang tertinggi dibandingkan dengan *test option "percentage split"* hanya sebesar 27%. Berdasarkan hasil rata-rata diatas maka dapat disimpulkan persentase *correct* dan *incorrect* dari pegujian dataset menggunakan *multilayer perceptron* menghasilkan 44% akurasi.

V. KESIMPULAN

Berdasarkan tugas akhir yang telah diselesaikan penulis dengan judul "Prediksi *Rating* Film Animasi Berdasarkan Elemen *Mise En Scene* Menggunakan *Neural Network*" terdapat beberapa kesimpulan yaitu:

1. Proses pengumpulan data 125 film animasi layar lebar dari tahun 2010 hingga 2018 sangat terbantu dengan adanya *website* IMDb, karena dari *website* tersebut ada beberapa elemen *mise en scene* yang langsung terpenuhi seperti *the human figure* dan *composition*. Untuk elemen *setting* dibutuhkan proses *screenshot* pada tiap filmnya, sedangkan elemen *lighting* perlu melakukan ekstraksi gambar menggunakan MATLAB.
2. Visualisasi dari *neural network* yang ditampilkan pada antar muka di aplikasi *weka* cukup jelas dalam menampilkan *input layer*, *hidden layer* dan *output layer*, hanya saja tidak bisa diperbesar untuk melihat jumlah *weight* yang terhubung pada tiap *node* jika

input layer atau atribut yang terdapat pada dataset cukup banyak.

3. Dari hasil uji validasi terhadap *test option* yang digunakan nilai *error* terendah ialah pada *test option "use training set"* yaitu sebesar 0.0989. Sedangkan hasil perbandingan *correctly/incorrectly classified instances* menggunakan *classifier "multilayer perceptron"* diperoleh persentase 79.25% sebagai akurasi tertinggi didapat pada *test option "use training set"*. Adapun rata-rata yang didapat dari hasil akurasi tiap *test option* sebesar 44%.

DAFTAR PUSTAKA

- [1] Nithin Vr., Sarath Babu Pb., (2014). Predicting Movie Success Based on IMDB Data.
- [2] Darin Im., Minh Thao Nguyen., (2011). Predicting Box-Office Success of Movies in the U.S. Market.
- [3] Deniz Demir., Olga Kapralova., Hongze Lai., (15 December, 2012). Predicting IMDB movie ratings using Google Trends.
- [4] Maria Pramaggiore., Tom Wallis., (2005). Film: A Critical Introduction ISBN 1-85669-442-9 Laurence King Publishing Ltd.
- [5] Riwinoto., Selly Artaty Zega., Gia Irlanda., (2015). Predicting Animated Film of Box-Office Success with Neural Networks.
- [6] Jayakrishnan Sreekumar., Amrita Vishwa Vidyapeetham., (September 2015). Creating Meaning through Interpretations: A Mise-En-Scene Analysis of the Film 'The Song of Sparrows' Online Journal of Communication and Media Technologies.
- [7] Muhammad Dahria., (2 Agustus 2008). Kecerdasan Buatan (artificial Intelligence) Jurnal Saintikom.
- [8] David Houcque., (Agustus 2005). Introduction to Matlab for Engineering Student version 1.2
- [9] Dr. Sudhir B. Jagtap., (2013). Census Data Mining and Data Analysis using WEKA.
- [10] Andrej Krenker., Janez Bešter., Andrej Kos., (2011). Introduction to the Artificial Neural Networks, Artificial Neural Networks - Methodological Advances and Biomedical Applications, Prof. Kenji Suzuki (Ed.), ISBN: 978-953-307-243-2, InTech.
- [11] 15th IEEE International Conference on Machine Learning and Applications (2016). Predicting Movie Box Office Profitability A Neural Network Approach.
- [12] Jiawei Han., Micheline Kamber., (2000). Data Mining: Concepts and Techniques.
- [13] Hartarto Junaedi., Herman Budianto., Indra Maryati., Yuliana Melani., (2011) Data Transformation Pada Data Mining.
- [14] Negnevitsky, Michael. (2005). Artificial Intelligence: A Guide to Intelligent
- [15] Systems (2nd ed.). Pearson Education.
- [16] Samuel Sena. (2017). Pengenalan Deep Learning Part 3 : Back-Propagation Algorithm

-
- [17] Mandella Majid Prachihara. (2016). Warna Sebagai Look dan Mood Pada Videografi Film Televisi "Pancer".
- [18] Nevenka Dimitrova.,Cheng-Yu Wei., Shih-Fu Chang.(2004).Color-Mood Analysis of Films Based on Syntactic and Psychological Models.
- [19] Richard Kirkby.,Eibe Frank.(2004).Weka Explorer User Guide for Version 3-4-3. The University of Wikato.
- [20] Remco R. Bouckaert.,Eibe Frank.,Mark Hall.,Richard Kirkby.,Peter .,Reutemann.,Alex Seewald.,David Scuse.(2012).Weka Manual for Version 3-6-8. The University of Wikato.
- [21] Ian H. Witten.(2013).More Data Mining with Weka Class 2 - Lesson 1. The University of Waikato.
- [22] Ian H. Witten.(2014).More Data Mining with Weka Class 5 - Lesson 1. The University of Waikato.
- [23] Nunik Purwaningsih.(2016).Penerapan *Multilayer Perceptron* Untuk Klasifikasi Jenis Kulit Sapi Tersamak.ISSN: 2338-2724.
-