

# Klasifikasi Sentimen Ulasan Film Indonesia dengan Konversi Speech-to-Text (STT) Menggunakan Metode Convolutional Neural Network (CNN)

Nadhifa Ayu Shafirra dan Irhamah

Departemen Statistika, Institut Teknologi Sepuluh Nopember (ITS)

*e-mail*: irhamah@statistika.its.ac.id

**Abstrak**—Ulasan film adalah sebuah opini yang bersifat subjektif. Ulasan film memiliki media yang beragam, seperti tulisan, audio, dan video. Ulasan film dapat diolah dengan menggunakan klasifikasi sentimen, agar ucapan seseorang terkait film dapat ditentukan sebagai sentimen tertentu. Di masa sekarang, data memiliki berbagai bentuk, pemilihan jenis data yang lebih baik juga dapat mempengaruhi klasifikasi sentimen. Data video dapat di-konversi menjadi data teks dengan bantuan Speech-to-Text (STT). Data teks digunakan karena kata atau kalimat dapat dibedakan secara negatif atau positif. Data ulasan dikelompokkan berdasarkan aspek penilaian film dan klasifikasi sentimen dilakukan pada keseluruhan potongan ulasan serta di tiap aspek yang ada. Dengan menggunakan metode Convolutional Neural Network, didapatkan bahwa model klasifikasi sentimen tiap aspek memiliki nilai AUC lebih baik dibandingkan model klasifikasi sentimen dengan keseluruhan data.

**Kata Kunci**—Aspek Penilaian, Convolutional Neural Network, Speech-to-Text, Ulasan Film

## I. PENDAHULUAN

FILM adalah alternatif hiburan bagi masyarakat di seluruh dunia. Sebagian besar masyarakat menikmati film-film yang beredar di setiap waktu. Dengan pilihan film yang beragam, masyarakat membaca ulasan film untuk menentukan film yang ingin mereka tonton. Ulasan film telah beredar di berbagai media, seperti tulisan ataupun video. Ulasan film dalam bentuk video juga sudah sangat banyak di Youtube, Instagram, dan di media lainnya.

Beberapa *channel* Youtube Indonesia yang khusus berisi konten ulasan film adalah Cine Crib, #sumatranbigfoot, Kawan Review, dan lain sebagainya. Ulasan film adalah sebuah opini yang bersifat subjektif dan tentu saja tiap orang pasti akan berbeda-beda dalam memberikan opininya terkait film yang telah ditonton. Seringkali antar dua ulasan memiliki poin yang berbeda dalam menilai suatu film. Dibutuhkan suatu metode yang dapat menentukan apakah ucapan dari *reviewer* bersifat positif atau negatif pada suatu film tersebut. Alternatif untuk menyelesaikan permasalahan ini adalah analisis sentimen. Analisis sentimen atau bisa juga disebut *opinion mining* adalah suatu bidang yang mempelajari cara menganalisis opini, sentimen, evaluasi, pujian, sikap, dan emosi seseorang terhadap suatu entitas seperti produk, jasa, organisasi, individu, isu, acara, topik, dan lain sebagainya [1].

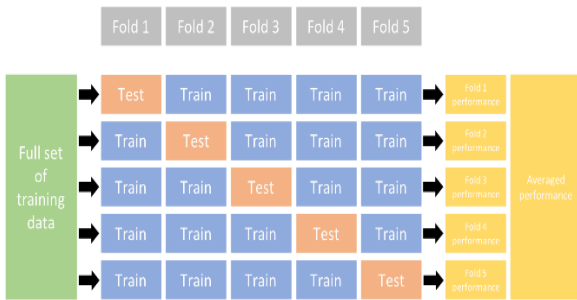
Data di masa sekarang tidak hanya dalam bentuk angka saja, mulai dari tulisan, audio, gambar, ataupun video sekarang dapat menjadi sumber data yang dapat diolah. Analisis sentimen juga dapat diterapkan dalam data audio. Namun,

data audio tidak bisa langsung digunakan sebagai input untuk klasifikasi sentimen. Data audio dapat diubah menjadi data teks yang selanjutnya digunakan untuk analisis sentimen. Proses perubahan data suara menjadi data teks seringkali disebut *Speech-to-Text* [2]. Banyak API yang menyediakan fasilitas *Speech-to-Text* secara gratis, salah satu contohnya adalah *Google Cloud Speech*.

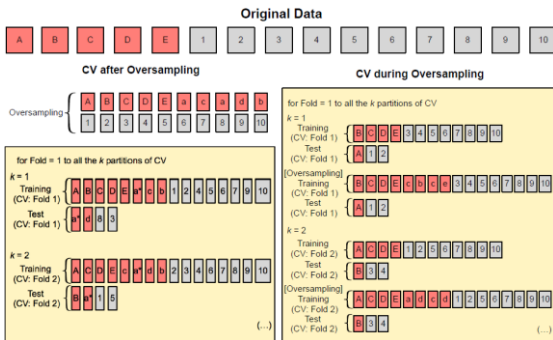
Data yang telah diekstraksi selanjutnya dianalisis menggunakan metode klasifikasi. Data teks digunakan karena kata atau kalimat dalam suatu teks dapat dibedakan secara negatif atau positif berdasarkan makna dari kumpulan kata tersebut. Karena ulasan film memiliki durasi dan panjang teks yang berbeda-beda akan kurang efektif jika melakukan klasifikasi sentimen dengan variabel dependen berupa ulasan. Sehingga untuk membuat klasifikasi sentimen lebih efektif, ulasan yang ada dapat dibagi berdasarkan aspek-aspek penilaian seorang *reviewer* dalam menilai suatu film. Aspek penilaian yang digunakan adalah film secara keseluruhan, pemeran, direktor (sutradara), jalan cerita, adegan, dan musik [3]. Sehingga, tiap ulasan akan memiliki enam variabel dependen berdasarkan jenis aspeknnya.

Metode klasifikasi yang akan digunakan adalah *Convolutional Neural Network*. Hasil analisis sentimen data *tweet* menggunakan CNN lebih baik dibandingkan metode SVM untuk keseluruhan kelas [4]. Antara metode *convolutional neural network* (CNN), LSTM, dan *distributed CNN*, dalam melakukan klasifikasi emosi pada data hasil *speech recognition* didapat nilai akurasi paling baik adalah *distributed CNN* sebesar 86,65%, CNN sebesar 86,06%, dan terakhir LSTM sebesar 78,31% [5]. Berdasarkan penelitian terdahulu untuk klasifikasi sentimen pada data audio tentang klasifikasi suara paru-paru dari 7 kelas yang ada nilai presisi metode CNN lebih baik di 4 kelas dibandingkan metode SVM [[6]. Berdasarkan beberapa penelitian terdahulu didapatkan bahwa metode CNN cukup baik dalam menangani permasalahan klasifikasi sentimen.

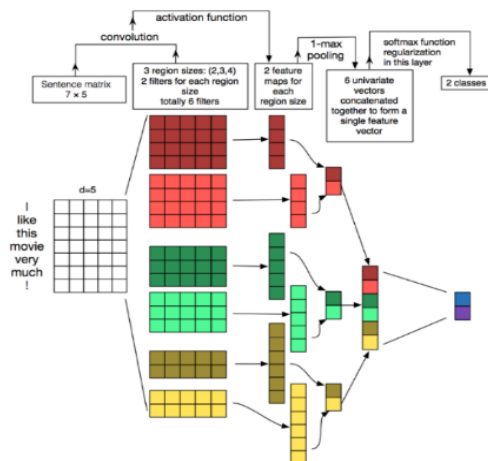
*Convolutional Neural Network* (ConvNet/CNN) adalah algoritma *deep learning* yang dapat menggunakan gambar sebagai input, menetapkan kepentingan (bobot dan bias yang dapat dipelajari) untuk berbagai aspek dan objek dalam gambar agar dapat membedakan satu dengan yang lain [7]. Dalam penelitian ini akan didapatkan proses konversi data audio menggunakan STT. Selain itu, klasifikasi sentimen akan dilakukan terhadap semua potongan ulasan serta ulasan yang telah dikelompokkan berdasarkan aspeknnya. Setelah mendapatkan hasil klasifikasi menggunakan metode CNN dari



Gambar 1. Ilustrasi K-Fold Cross Validation dengan k=5



Gambar 2. Ilustrasi SMOTE dalam K-Fold Cross Validation



Gambar 3. Ilustrasi Text Mining Menggunakan CNN

masing-masing data ekstraksi, hasil klasifikasi tersebut dihitung menggunakan evaluasi kinerja klasifikasi untuk menentukan hasil klasifikasi terbaik antara kedua data yang telah digunakan. Grafik prosentase diperoleh untuk menampilkan visualisasi klasifikasi sentimen suatu film. Diharapkan hasil penelitian ini dapat dimanfaatkan untuk penelitian selanjutnya terkait klasifikasi sentimen dengan metode CNN.

## II. TINJAUAN PUSTAKA

### A. Speech Recognition

Speech recognition atau pengenalan suara adalah teknologi dengan suara sebagai input ke suatu sistem [8]. Komponen pertama dari pengenalan suara adalah ucapan. Suara ini harus dikonversikan dari suara fisik ke sinyal listrik dengan mikrofon, kemudian ke data digital dengan konverter. Setelah didigitalkan, beberapa model dapat digunakan untuk menyalin audio ke teks. Hal ini sering disebut juga Speech-to-Text (STT)[2].

### B. Text Mining

Text mining adalah serangkaian teknik dan metode yang digunakan untuk pemrosesan otomatis untuk data teks bahasa (natural language text) [9]. Seperti data mining, text mining sebagian berasal dari tanggapan terhadap data teks berukuran besar yang dibuat dan disebar di masyarakat (seperti hukum, perintah, peraturan, atau kontrak), dan sebagian untuk tujuan generalisasi input dan penyimpanan data di sistem komputer [9]. Ilustrasi text minning menggunakan CNN dapat dilihat pada Gambar 3.

Text preprocessing mengambil input raw text dan mengembalikan output berupa token yang telah dibersihkan. Setiap tahap dari text preprocessing akan menghapus informasi yang tidak diperlukan dari teks aslinya [10]. Tahapan text preprocessing terdiri dari case folding, stemming, stopwords removal, dan tokenizing.

Case folding adalah tahapan yang berfungsi untuk mengubah font, serta mengubah semua huruf menjadi huruf lowercase [11]. Stopwords removal adalah tahapan text preprocessing yang akan menghilangkan stopwords dalam suatu teks [12]. Contoh stopwords dalam bahasa Indonesia adalah “yang”, “dan”, “di”, dan lain sebagainya. Stemming adalah proses untuk menggati kata-kata menjadi kata dasarnya [12]. Tokenizing adalah proses untuk memisahkan kata per kata dalam suatu teks [1].

Proses selanjutnya adalah melakukan word embedding di mana word embedding adalah pendekatan untuk vektor dari representasi kata-kata berdasarkan makna kata tersebut [13]. Salah satu metode word embedding yang biasa digunakan adalah Word2Vec. Word2Vec adalah word embedding di mana kata-kata dipetakan dalam ruang vektor tetap, sehingga kemiripan cosinus antara dua kata adalah estimator yang baik dari kesamaan semantic. Manfaat utama Word2Vec adalah bahwa metode training tidak perlu membangun dan memvalidasi model yang kompleks secara manual, tetapi model Word2Vec belajar dengan mengekstraksi makna kata dengan mempertimbangkan kata-kata yang berdekatan dalam satu set kalimat[14].

### C. Klasifikasi Sentimen

K-Fold Cross Validation menjadi langkah pertama dalam klasifikasi sentimen karena berfungsi untuk meminimalkan bias yang terkait dengan pengambilan sampel acak dari data training dan data sampel dalam membandingkan akurasi prediksi dua metode atau lebih. Kumpulan data lengkap dibagi secara acak menjadi k subset dengan ukuran yang sama. Model klasifikasi yang ada diuji sebanyak k kali. Setiap fold di-training kecuali satu fold dan diuji untuk satu fold yang tersisa [15].

Jika hasil K-Fold Cross Validation tidak seimbang, maka pembagian data training dan testing akan disertai dengan resampling. Salah satu teknik resampling adalah Synthetic Minority Oversampling Technique (SMOTE) yang dirancang oleh V. Chawla et. all (2002). Cara menentukan data synthetic dirumuskan dalam persamaan berikut.

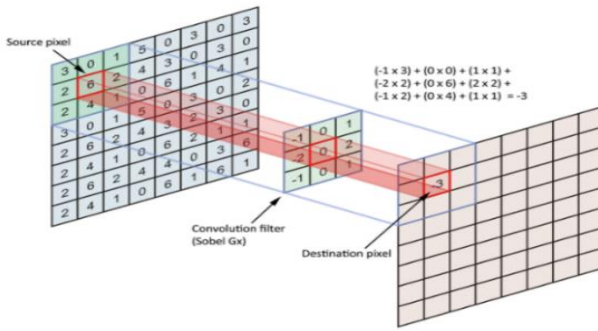
$$x_{syn} = x_i + \delta(x_{knn} - x_i), \tag{1}$$

keterangan:

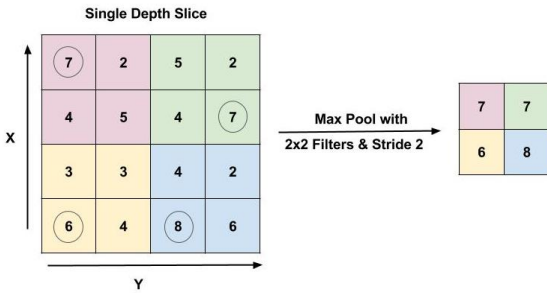
$x_{syn}$  = nilai bobot synthetic

$x_i$  = nilai bobot data ke-i di kelas minoritas

$x_{knn}$  = nilai bobot dari data di kelas minoritas yang



Gambar 4. Ilustrasi Convolutional Layer



Gambar 5. Ilustrasi Pooling Layer

memiliki jarak terdekat dengan  $x_i$   
 $\delta$  = bilangan acak antara 0 dan 1

SMOTE akan diterapkan di prosedur *K-Fold Cross Validation* pada masing-masing data *training* di setiap *fold* Dapat dilihat pada Gambar 1. Ilustrasi penerapan SMOTE pada *K-Fold CV* ditunjukkan pada Gambar 2.

Metode yang akan digunakan dalam klasifikasi sentimen ini adalah *Convolutional Neural Network (ConvNet/CNN)*. CNN termasuk dalam jaringan saraf tiruan *deep feed-forward* (di mana koneksi antar *node* tidak membentuk siklus. CNN umumnya digunakan dalam *computer vision*, namun CNN juga mulai banyak diterapkan pada berbagai tugas NLP (*Natural Language Processing*) [16]. Penggunaan CNN pada pada Arsitektur CNN *Unigram* dalam Klasifikasi Data Teks Gambar 8.

Pola bisa berupa ekspresi dan CNN dapat mengidentifikasinya dalam kalimat tertentu terlepas posisi kata-kata tersebut [16]. Selain itu CNN juga dapat diterapkan menggunakan data yang dihasilkan oleh ekstraksi menggunakan MFCC. *Layer* yang digunakan dalam CNN terdiri dari 3 jenis yaitu *convolutional layer*, *pooling layer*, dan *fully-connected layer*. *Layer* pertama adalah *convolutional layer* berisi serangkaian *filter* dengan ukuran yang tetap untuk digunakan dalam melakukan konvolusi pada data. *Output* dari *layer* ini adalah matriks baru yang disebut *feature maps* [17]. Persamaan operasi *convolutional* ditunjukkan pada Persamaan 12.

$$FM_{a,b} = bias + \sum_c^C \sum_d^D Z_{c,d} \times X_{a+c-1,b+d-1}, \quad (2)$$

keterangan:

- $FM_{a,b}$  = *feature map* ke- $a,b$ ,
- $bias$  = bias pada *feature map*,
- $Z_{c,d}$  = bobot pada *convolution kernel* ke- $c,d$ ,
- $X$  = *input*,
- $a$  = 1,2, ...,  $A$ .  $A$  adalah panjang *pixel feature map*,
- $b$  = 1,2, ...,  $B$ .  $B$  adalah lebar *pixel feature map*,
- $c$  = 1,2, ...,  $C$ .  $C$  ukuran panjang *convolution kernel*,
- $d$  = 1,2, ...,  $D$ .  $D$  ukuran lebar *convolution kernel*.

Berdasarkan perhitungan tersebut berikut ditunjukkan ilustrasi proses pada *convolutional layer* pada Gambar 4. *Pooling layer* ini memastikan bahwa jaringan hanya berfokus pada pola yang paling penting. Pada *pooling layer*, data dirangkum dengan menggeser jendela melintasi *feature maps*, lalu menerapkan beberapa operasi linear atau non-linear pada data yang ada di dalam jendela. *Pooling layer* berfungsi untuk mengurangi dimensi dari *feature maps* yang akan digunakan di *layer* selanjutnya [17]. Ilustrasi *pooling layer* dapat dilihat pada Gambar 5.

*Layer* terakhir yang digunakan adalah *fully-connected layer* untuk memahami pola yang dihasilkan oleh *layer* sebelumnya. Neuron di *layer* ini memiliki koneksi penuh ke semua aktivasi di *layer* sebelumnya [17]. Metode CNN juga menggunakan fungsi aktivasi yang dilakukan ketika berada di antara *convolutional layer* dan *pooling layer*. Aktivasi di antara kedua *layer* tersebut menggunakan fungsi aktivasi ReLU. Sedangkan untuk fungsi aktivasi *output* menggunakan *sigmoid*. Persamaan fungsi aktivasi ReLU adalah sebagai berikut.

$$f_h(0, FM_{a,b}) = \max(0, FM_{a,b}) = \begin{cases} FM_{a,b}, & \text{jika } FM_{a,b} \geq 0, \\ 0, & \text{jika } FM_{a,b} < 0, \end{cases} \quad (3)$$

keterangan:

- $FM_{a,b}$  = *feature map* ke- $a,b$ ,
- $a$  = 1,2, ...,  $A$ .  $A$  panjang *pixel* pada *feature map*,
- $b$  = 1,2, ...,  $B$ .  $B$  panjang *pixel* pada *feature map*,

Fungsi aktivasi pada *output* menggunakan fungsi *sigmoid* yang dikhususkan pada klasifikasi dengan dua kelas karena memiliki *range* nilai 0 hingga 1 [18]. Persamaan fungsi aktivasi *sigmoid* adalah sebagai berikut.

$$f(x) = \frac{1}{1+e^{-x}} \quad (4)$$

Fungsi terakhir adalah *loss function* untuk menghitung *loss* (nilai *error*) dengan menggunakan *binary cross-entropy*. Persamaan dari *loss function* adalah sebagai berikut.

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N y_i \times \log(p(y_i)) + (1 - y_i) \times \log(1 - p(y_i)) \quad (5)$$

keterangan:

- $N$  = jumlah banyak data,
- $y_i$  = kelas pada klasifikasi (0 atau 1),
- $p(y_i)$  = nilai peluang  $y_i$ .

Setelah mencapai *layer* terakhir dan menghitung fungsi aktivasi *output*, tahapan selanjutnya adalah melakukan optimasi parameter untuk meminimumkan nilai *loss* menggunakan *Adam optimizer*. *Adam optimizer* adalah metode optimisasi stokastik yang diterapkan pada model CNN untuk memperbaiki parameter pada jaringan sehingga fungsi tujuan dapat dioptimalkan [19]. *Adam* menggunakan *gradient*, estimasi momen pertama dan kedua, serta melakukan koreksi melalui *bias correction*. Langkah-langkah algoritma dari optimasi Adam adalah sebagai berikut.

1. Menentukan nilai  $\beta_1$  dan  $\beta_2$  secara *default* berturut-turut sebesar 0,9 dan 0,999.
2. Menentukan fungsi  $f(x)$  yang merupakan *stochastic objective function*.
3. Menginisiasi parameter *moment* pertama dan kedua ( $m_0, v_0$ ).
4. Menghitung gradien dari fungsi (15).
5. Melakukan *update* pada *moment* pertama dan kedua dengan formula berikut.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{13}$$

Tabel 2.

Kata Kunci dan Aspek yang Dinilai dalam Ulasan Film	
Aspek	Kata Kunci
Pemeran	akting, aktris, aktor, peran, karakter, penjahat, kinerja, performansi, dimainkan, pemeran, pemain, <i>chemistry</i> , pesona
Direktor	arah, arahan, penyutradaraan, sutradara, difilmkan, pembuatan film, pembuat film, sinematik, edisi, sinematografi, direktor
Jalan Cerita	jalan cerita, cerita, kisah, roman, dialog, skrip, bercerita, berakhir, <i>ending</i> , menceritakan, balas dendam, pengkhianatan, plot, penulisan, <i>twist</i> , drama, misteri, <i>thriller</i> , horor
Adegan	pemandangan, animasi, kekerasan, skenario, aksi, efek khusus, <i>special effect</i> , <i>stunt</i> , visual, alat peraga, properti, kamera, grafik, efek, kamera
Musik	lirik, suara, musik, audio, <i>musical</i> , trek, efek suara, <i>sound effect</i> , <i>soundtrack</i>
Keseluruhan	film, <i>remake</i> , adaptasi, <i>sequel</i> , hiburan

Tabel 3.

Jumlah Potongan Ulasan Tiap Aspek		
Aspek	Film	Jumlah Potongan Ulasan
Aspek 1 (Pemeran)	Bumi Manusia	9
	Dilan 1991	7
	Dua Garis Biru	10
	Gundala	9
	Makmum	9
Aspek 2 (Direktor)	Bumi Manusia	9
	Dilan 1991	6
	Dua Garis Biru	9
	Gundala	7
	Makmum	9
Aspek 6 (Keseluruhan)	Bumi Manusia	10
	Dilan 1991	10
	Dua Garis Biru	10
	Gundala	10
	Makmum	10

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1)g_t \tag{6}$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2)g_t \odot g_t \tag{7}$$

6. Menghitung *bias correction* dengan formula berikut.

$$\hat{m}_t = \frac{m_t}{(1 - \beta_1^t)} \tag{8}$$

$$\hat{v}_t = \frac{v_t}{(1 - \beta_2^t)} \tag{9}$$

7. Melakukan *update parameter* jika konvergen maka iterasi berhenti, jika tidak maka mengulang langkah ke-4. Persamaan untuk menghitung tahap terakhir ini ditunjukkan pada persamaan (20).

$$\theta_t = \theta_{t-1} - \frac{\eta \cdot \hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \tag{10}$$

$\theta$  merupakan parameter yang terdiri dari *bias* pada *input layer* ( $a_j$ ), bobot antara *input layer* dan *hidden layer* ( $w_{i,j}$ ), *bias* pada *hidden layer* ( $b_k$ ), dan bobot antara *input layer* dan *output layer* ( $v_{j,k}$ ).

Evaluasi kinerja klasifikasi dihitung untuk membandingkan hasil klasifikasi sentimen dan dibuat sesuai dengan rumus berikut [20].

Nilai *True Positive*:

$$TP Rate = \frac{TP}{TP+FN} \tag{11}$$

Nilai *False Positive*.

$$FP Rate = \frac{FP}{FP+TN} \tag{12}$$

Akurasi:

Tabel 1.

Variabel Penelitian Data Teks		
Variabel	Keterangan	Skala Data
Y	Kelas sentimen 0 = Positif 1 = Negatif	Nominal
X	<i>Word Embedding</i>	Rasio

AUC:

$$AUC = \frac{TP Rate + (1 - FP Rate)}{2} \tag{14}$$

#### D. Ulasan Film

Ulasan film adalah analisis dari suatu film yang dibuat oleh satu orang atau secara berkelompok lalu mengungkapkan pendapatnya tentang film tersebut [21]. Media dari sebuah ulasan beragam, mulai dari tulisan, suara, ataupun melalui video dapat di lihat pada Tabel 1. Ulasan ditujukan untuk menunjukkan pan-dangan atau penilaian terhadap suatu film serta membantu para pembaca dan penonton untuk mengetahui isi dari suatu film [22]. Para *reviewer* membuat ulasan berdasarkan aspek-aspek tertentu, sehingga klasifikasi sentimen nantinya tidak hanya untuk keseluruhan ulasan tapi berdasarkan aspek yang dinilai dalam ulasan tersebut [23].

### III. METODOLOGI PENELITIAN

#### A. Sumber Data

Data sekunder diambil dari Youtube. Data audio berupa suara orang yang mengulas film dikonversi menjadi data teks. Data yang digunakan terdiri dari 10 video *review* tiap film. Film yang dipilih adalah 5 film Indonesia yang tayang di tahun 2019. Data yang ada diolah melalui *preprocessing* dan selanjutnya akan didapatkan 242 potongan ulasan. Potongan ulasan tersebut telah dikelompokkan berdasarkan aspeknya dapat dilihat pada Tabel 2.

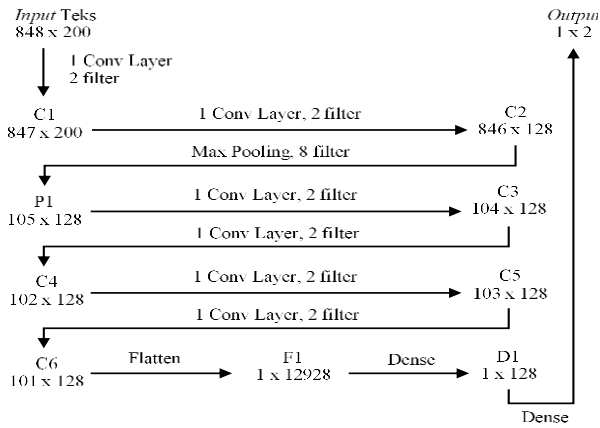
#### B. Variabel Penelitian

Variabel penelitian yang akan digunakan untuk klasifikasi sentimen menggunakan data teks dapat dilihat pada Tabel 3.

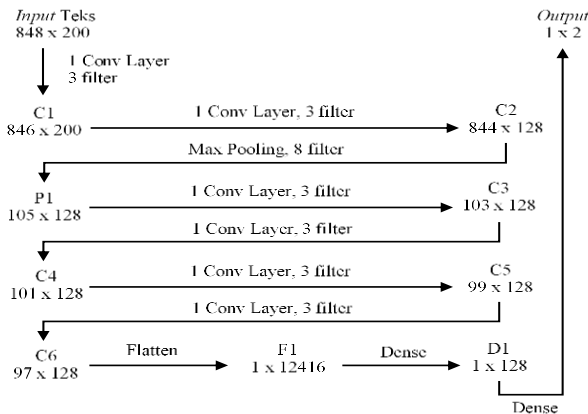
#### C. Langkah Analisis

Langkah analisis yang akan dilakukan pada penelitian ini adalah sebagai berikut.

1. Mengumpulkan audio dari video ulasan film Indonesia yang ada di Youtube.
2. Memotong audio menjadi potongan audio berdasarkan keheningan audio dengan ketentuan durasi keheningan (*silence*) minimal 350 ms dan audio tersebut dikatakan hening jika amplitudo suara lebih kecil dari -25 hingga -30 dBFS.
3. Melakukan proses *Speech-to-Text* pada data audio dengan bantuan API *Google Speech Cloud* untuk mengambil data teks.
4. Melakukan *text preprocessing* dengan tahap-tahap sebagai berikut.
  - a. *Case folding*,
  - b. *Stemming*,
  - c. *Stopword Removal*,
  - d. *Tokenizing*.



Gambar 6. Arsitektur CNN Bigram dalam Klasifikasi Data Teks



Gambar 7. Arsitektur CNN Trigram dalam Klasifikasi Data Teks

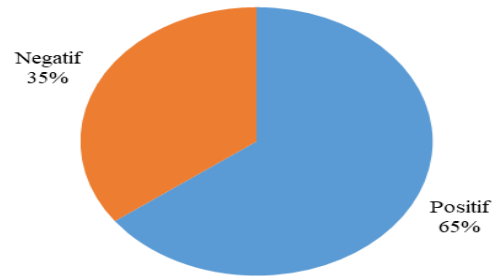
5. Mengelompokkan potongan teks tersebut berdasarkan aspek ulasan.
6. Pelabelan sentimen data ulasan film secara manual.
7. Melakukan pembobotan menggunakan *word embedding* dengan algoritma *Word2Vec*.
8. Membagi data menjadi data *training* dan data *testing* menggunakan *K-Fold Cross Validation*.
9. Melakukan teknik SMOTE jika pembagian data *training* dan data *testing* tidak seimbang (*imbalanced*).
10. Mengklasifikasi sentimen ulasan film Indonesia di Youtube menggunakan metode CNN untuk hasil *word embedding*.
11. Memilih model terbaik berdasarkan nilai AUC dan akurasi.
12. Interpretasi, mengambil kesimpulan dan saran.

#### IV. ANALISIS DAN PEMBAHASAN

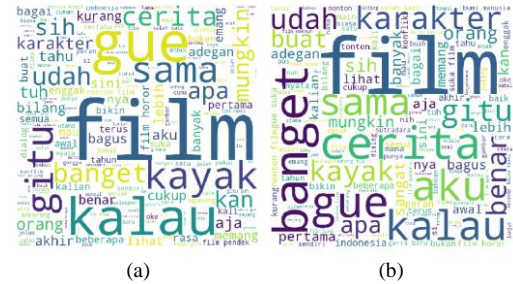
##### A. Karakteristik Data

Jumlah data yang akan digunakan sebanyak 242 sentimen ulasan film dengan prosentase antara sentimen positif dan negatif ditunjukkan pada Gambar 6.

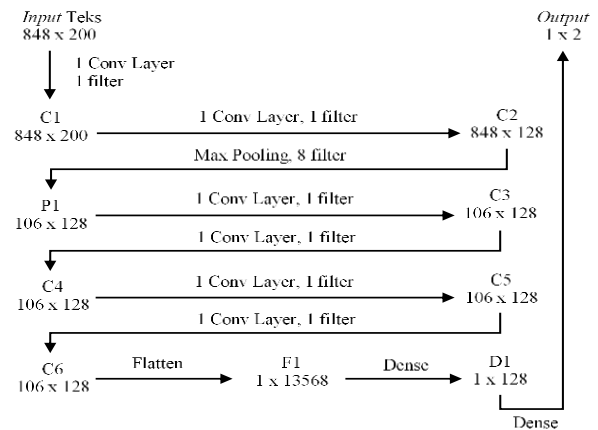
Data audio yang telah dikonversi menjadi data teks dan telah dikelompokkan berdasarkan aspek penilaian film, ditandai sentimen positif dan negatif dari masing-masing potongan ulasan dan *word cloud* untuk tiap sentimen ditunjukkan pada Gambar 7. Berdasarkan Gambar 7, teks ulasan film yang memiliki sentimen negatif dan positif tidak terlalu memiliki perbedaan. Sehingga, metode klasifikasi dapat membantu menentukan sentimen dari ulasan film yang ada. Arsitektur CNN Unigram dapat dilihat pada Gambar 8.



Gambar 8. Pie Chart Prosentase Sentimen Negatif dan Positif



Gambar 9. Word Cloud Ulasan Film Bersentimen (a) Negatif (b) Positif



Gambar 10. Arsitektur CNN Unigram dalam Klasifikasi Data Teks

##### B. Klasifikasi Data Teks Menggunakan CNN

Klasifikasi sentimen pada data teks dibagi menjadi dua yaitu dengan menggunakan semua ulasan yang ada dan pemodelan klasifikasi menggunakan data pada tiap-tiap aspek yang ada. Arsitektur CNN untuk masing-masing N-gram berbeda-beda.

Arsitektur CNN unigram yang ditunjukkan pada Gambar 9 terdiri dari 6 *convolution layer* dengan ukuran *filter* sebesar 1 sehingga *convolution window* bergeser tiap 1 kata dalam teks ulasan tersebut. Sehingga ukuran *output* dari *convolution layer* tidak berubah dari sebelumnya. Berdasarkan Gambar 10, arsitektur CNN *bigram* memiliki 6 *convolution layer* dengan ukuran *filter* 2, di mana *convolution window* bergeser tiap dua kata. Arsitektur CNN untuk *trigram* ditunjukkan pada Gambar 10 di mana tiap *convolution layer* memiliki ukuran *filter* sebesar 3. *Convolution window* akan bergeser setiap tiga kata dan ukuran *output* dari tiap *convolution layer* akan memiliki jumlah *input* dikurangi dengan dua. Contoh perhitungan dari *convolution layer* pertama dalam model *trigram* adalah sebagai berikut.

$$Y_{1,1} = -0,0018 + 0,0314X_{1,1} + (-0,0227)X_{2,1} + 0,0203X_{3,1}$$

Tabel 4.  
Nilai AUC Klasifikasi Keseluruhan Data Teks

K	Unigram		Bigram		Trigram	
	Training	Testing	Training	Testing	Training	Testing
	2	0,941	0,545	0,936	0,518	0,863
3	0,952	0,569	0,917	0,602	0,947	0,628
4	0,944	0,588	0,941	0,593	0,947	0,570
5	0,944	0,569	0,949	0,557	0,957	0,563
6	0,944	0,624	0,939	0,559	0,948	0,603
7	0,952	0,572	0,953	0,528	0,957	0,606
8	0,949	0,573	0,945	0,586	0,957	0,583
9	0,952	0,525	0,947	0,535	0,932	0,577
10	0,949	0,513	0,951	0,583	0,957	0,587

Tabel 5.  
Nilai Kinerja Klasifikasi Data Teks Tiap Aspek

Aspek	K	Model	AUC		Akurasi	
			Training	Testing	Training	Testing
			Pemeran	8	Unigram	0,996
Direktor	9	Trigram	0,997	0,599	0,997	0,483
Jalan Cerita	2	Bigram	1,000	0,590	1,000	0,646
Adegan	9	Unigram	1,000	0,608	1,000	0,550
Musik	3	Trigram	1,000	0,717	1,000	0,619
Keseluruhan	3	Trigram	1,000	0,655	1,000	0,680
Rata-Rata			0,999	0,648	0,999	0,617

Tabel 6.  
Evaluasi Kinerja Klasifikasi Model Tiap Data

Jenis Data		AUC		Akurasi	
		Training	Testing	Training	Testing
		Teks	Semua	0,947	0,628
	Aspek	0,999	0,648	0,999	0,617

$$Y_{1,2} = -0,0050 + (-0,0279)X_{1,2} + 0,0567X_{2,2} + (-0,0520)X_{3,2}$$

$$Y_{1,200} = -0,0079 + 0,0085X_{1,200} + (-0,0392)X_{2,200} + (-0,0133)X_{3,200}$$

$$Y_{846,200} = -0,0079 + 0,0145X_{846,200} + (-0,0036)X_{847,200} + (-0,0780)X_{848,200}$$

Klasifikasi sentimen menggunakan data teks dibagi menjadi tiga bagian di mana kalimat ulasan dipotong menjadi *unigram* (1 kata), *bigram* (2 kata), dan *trigram* (3 kata). Pada Tabel 4 hasil nilai akurasi pada model unigram, bigram, dan trigram dengan *k-fold* 2 hingga 10.

Berdasarkan Tabel 4, model terbaik dengan nilai AUC tertinggi adalah model trigram dengan *fold* sebesar 3. Akurasi dari data *training* model ini sebesar 1 dan untuk akurasi data *testing* memiliki nilai sebesar 0,579. Model *trigram* dengan *fold* 3 ini akan dibandingkan dengan hasil klasifikasi MFCC.

Klasifikasi selanjutnya menggunakan data teks yang dikelompokkan pada tiap aspeknya. Aspek penilaian yang digunakan sebanyak 6 aspek yaitu pemeran, direktor, jalan cerita, adegan, musik, dan secara keseluruhan. Berdasarkan

Tabel 7.  
Evaluasi Kinerja Klasifikasi Terbaik Model Tiap Data

Aspek	K	Model	AUC		Akurasi	
			Train- ing	Test- ing	Train- ing	Test- ing
			Pemeran	8	Unigram	1,000
Direktor	9	Trigram	1,000	0,750	1,000	0,750
Jalan Cerita	2	Bigram	1,000	0,668	1,000	0,708
Adegan	9	Unigram	1,000	1,000	1,000	1,000
Musik	3	Trigram	1,000	0,916	1,000	0,857
Keseluruhan	3	Trigram	1,000	0,733	1,000	0,706
Rata-Rata			1,000	0,845	1,000	0,837

Tabel 8.  
Prosentase Sentimen Positif dan Negatif Tiap Film

Nama Film	Prosentase Sentimen Positif	Prosentase Sentimen Negatif	Jumlah Sentimen
Bumi Manusia	74%	26%	47
Dilan 1991	50%	50%	44
Dua Garis Biru	71%	29%	52
Gundala	67%	33%	48
Makmum	55%	45%	51

AUC dari data *testing* tertinggi, model terbaik dari masing-masing aspek dipilih dan model terpilih di tiap aspek dengan nilai AUC dan akurasi ditunjukkan pada Tabel 4.

Berdasarkan Tabel 5, rata-rata AUC data *testing* dari masing-masing model tiap aspek bernilai sebesar 0,648. Sedangkan untuk rata-rata akurasi data *testing* sebesar 0,617. Kumpulan model klasifikasi dari tiap aspek ini akan dibandingkan dengan model yang lain.

### C. Perbandingan Evaluasi Kinerja Klasifikasi

Model-model yang terpilih dari masing-masing jenis data dibandingkan untuk menentukan model terbaik berdasarkan nilai AUC dan akurasi dari data *testing*. Berdasarkan Tabel 6, dari keempat model yang ada, model dengan nilai AUC dari data *testing* tertinggi adalah model tiap aspek dengan menggunakan data teks. Meskipun nilai akurasi tidak tertinggi, namun nilai tersebut berada di atas 0,6. Klasifikasi menggunakan data MFCC memiliki akurasi yang lebih tinggi namun tidak dengan nilai AUC-nya, sehingga data teks lebih baik digunakan ketika data sentimen yang ada bersifat *imbalanced*.

Selanjutnya, dipilih salah satu hasil pada *fold* model dengan menggunakan data teks di tiap aspek berdasarkan nilai AUC tertinggi. Hasil pemilihan *fold* dengan nilai AUC tertinggi ditunjukkan pada Tabel 7. Rata-rata akurasi dengan model data teks di tiap aspek berdasarkan Tabel 7 mencapai 83,7%. Dengan rata-rata nilai AUC data *testing* yang diperoleh sebesar 0,845.

### D. Prosentase Klasifikasi Sentimen Ulasan Film

Setelah model dipilih, prosentase sentimen positif dan negatif berdasarkan hasil klasifikasi dihitung untuk menentukan nilai ulasan tiap film. Tabel 8 merupakan hasil prosentase sentimen positif dan negatif dari keseluruhan ulasan untuk tiap film. Hasil prosentase pada Tabel 8 menunjukkan dari kelima film, film Bumi Manusia memiliki prosentase

sentimen posi-tif tertinggi. Sedangkan untuk film dengan prosentase senti-men positif terendah adalah film Dilan 1991.

## V. KESIMPULAN/RINGKASAN

Berdasarkan penelitian yang telah dikerjakan dapat disimpulkan bahwa data audio yang telah dikelompokkan menghasilkan 242 potongan audio dengan jumlah sentimen positif sebesar 65% dan sentimen negatif sebesar 35%. Arsitektur CNN yang digunakan untuk klasifikasi sentimen adalah 6 *convolution layer* dengan 1 *pooling layer* antara *layer* kedua dan ketiga. Arsitektur CNN untuk data teks dibagi menjadi tiga yaitu *unigram*, *bigram*, dan *trigram*. Model terbaik untuk klasifikasi sentimen dengan data teks adalah model yang berdasarkan aspek dari penilaian film. Dari keempat model yang ada dibandingkan dan terpilih model dari data teks di tiap aspek. Setelah memilih model terbaik di tiap *fold* didapatkan rata-rata nilai AUC dan akurasi dari data *testing* berturut-turut sebesar 0,845 dan 0,837. Film dengan prosen-tase sentimen positif tertinggi adalah Bumi Manusia sebesar 74%. Sedangkan film Dilan 1991 merupakan film dengan prosentase senti-men positif terendah yaitu sebesar 50%. Di penelitian selanjutnya diharapkan dapat memilih model arsitektur CNN yang dapat mengklasifikasi sentimen dengan lebih baik karena jenis dan jumlah *layer* yang digunakan dapat mempengaruhi kebaikan model. Selain itu, penambahan jumlah data dapat dilakukan agar bisa menangani *imbalance* serta agar data tidak *overfitting*.

## DAFTAR PUSTAKA

- [1] B. Liu, "Sentiment Analysis and Opinion Mining," *Synth. Lect. Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012, doi: doi.org/10.2200/S00416ED1V01Y201204HLT016).
- [2] D. Amos, "The Ultimate Guide To Speech Recognition With Python," *Real Python*. <https://realpython.com/python-speech-recognition/> (accessed Jun. 17, 2020).
- [3] T. T. Thet, J. C. Na, and C. S. G. Khoo, "Aspect-based sentiment analysis of movie reviews on discussion boards," *J. Inf. Sci.*, vol. 36, no. 6, pp. 823–848, 2010, doi: 10.1177/0165551510388123.
- [4] H. Cui, Y. Lin, and T. Utsuro, "Sentiment Analysis of Tweets by CNN utilizing Tweets with Emoji as Training Data," *WISDOM*, no. August, pp. 1–8, 2018, [Online]. Available: [http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki\\_vector/](http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector/).
- [5] W. Lim, D. Jang, and T. Lee, "Speech emotion recognition using convolutional and Recurrent Neural Networks," *2016 Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. APSIPA 2016*, 2017, doi: 10.1109/APSIPA.2016.7820699.
- [6] D. Bardou, K. Zhang, and S. M. Ahmad, "Lung sounds classification using convolutional neural networks," *Artif. Intell. Med.*, vol. 88, pp. 58–69, 2018, doi: 10.1016/j.artmed.2018.04.008.
- [7] S. Saha, "A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way," *toward data science*, Dec. 16, 2018. <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53> (accessed Jun. 17, 2020).
- [8] N. Unuth, "What Is Speech Recognition?," 2019. [Online]. Available: <https://www.lifewire.com/>.
- [9] Tufféry Stéphane, *Data Mining and Statistics for Decision Making*. A John Wiley & Sons, Ltd. Publication, 2011.
- [10] M. Anandarajan, C. Hill, and T. Nolan, *Practical Text Analytics Maximizing the Value of Text Data*, 2nd ed. Cham, Switzerland: Springer, 2019.
- [11] A. A. Lutfi, A. E. Permanasari, and S. Fauziati, "Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 4, no. 1, pp. 57–64, 2018, doi: 10.20473/jisebi.4.2.169.
- [12] Y. Ledeneva, *Effect of Preprocessing on Extractive Summarization with Maximal Frequent Sequences*, vol. LNAI 5317. Av. Juan de Dios Batiz s/n, D.F., Mexico: Springer-Verlag Berlin Heidelberg, 2008.
- [13] J. Brownlee, *Deep Learning for Natural Language Processing: Develop Deep Learning Models ... - Jason Brownlee - Google Books*, V1.4. Jason Brownlee, 2019.
- [14] D. S. Charles, J. Carmona, V. M. Mulero, and M. Sole, *Reducing Even Variability in Logs by Clustering of Word Embeddings*, vol. 132, no. March. Barcelona: Springer International Publishing, 2018.
- [15] D. L. Olson and D. Delen, *Advanced data mining techniques*. Berlin: Springer, 2008.
- [16] A. Maheswari, "Report on Text Classification using CNN, RNN & HAN," *Medium*, Jul. 17, 2018. <https://medium.com/jatana/report-on-text-classification-using-cnn-rnn-han-f0e887214d5f> (accessed Jun. 17, 2020).
- [17] A. Ferreira and G. Giraldo, "Convolutional Neural Network approaches to granite tiles classification," *Expert Syst. Appl.*, vol. 84, pp. 1–11, 2017, doi: 10.1016/j.eswa.2017.04.053.
- [18] C. Gautam and V. Ravi, "Auto associative extreme learning machine based hybrids for data imputation," *Handb. Res. Intell. Tech. Model. Appl. Mark. Anal.*, vol. 5, pp. 75–99, 2016, doi: 10.4018/978-1-5225-0997-4.ch005.
- [19] E.-Y. Huan *et al.*, "Deep Convolutional Neural Networks for Classifying Body Constitution Based on Face Image," *Hindawi Comput. Math. Methods Med.*, vol. 9887 LNCS, pp. 128–135, 2017, doi: 10.1007/978-3-319-44781-0\_16.
- [20] Y. Al Amrani, M. Lazaar, and K. E. El Kadirp, "Random forest and support vector machine based hybrid approach to sentiment analysis," *Procedia Comput. Sci.*, vol. 127, pp. 511–520, 2018, doi: 10.1016/j.procs.2018.01.150.
- [21] bestessays, "Glossary - Movie Review," *bestessays*, 2019. [https://www.bestessays.com/glossary/movie\\_review.php](https://www.bestessays.com/glossary/movie_review.php) (accessed Jun. 17, 2020).
- [22] B. Syahid, "Pengertian Teks Ulasan, Contoh, Ciri, Tujuan, Struktur, dan Kaidahnya," *guru pendidikan*, Jan. 13, 2020. <https://www.gurupendidikan.co.id/teks-ulasan/> (accessed Jun. 17, 2020).
- [23] D. Anand and D. Naorem, "Semi-supervised Aspect Based Sentiment Analysis for Movies Using Review Filtering," *Procedia Comput. Sci.*, vol. 84, no. September, pp. 86–93, 2016, doi: 10.1016/j.procs.2016.04.070.