

Ekstraksi Emosi Majemuk Kalimat Bahasa Indonesia Menggunakan *Convolutional Neural Network* (*Compound Emotional Extraction of Indonesian Sentences Using Convolutional Neural Network*)

Aripin¹, Wisnu Agasty², Hanny Haryanto³

Abstract—Facial expressions can strengthen the information conveyed in interactive communication. In the field of developing virtual characters specifically for facial characters, facial expressions are needed to animate a facial virtual character to make it look natural like a human. One type of emotional expression is a compound emotional expression, which is a combination of two or more basic emotions. For example, the expression of disappointed emotions is a combination of anger and sadness. Facial expressions can appear due to emotional stimulation, one of which is the meaning of the sentence. This research aims to extract emotional data from Indonesian sentences using the multi-label classification process of the CNN model so as to produce compound facial expressions that are applied in virtual character animation. The basic emotion classes used in the classification process are anger, disgust, fear, happiness, sadness, and surprise. Based on the experimental results, the CNN model can produce an accuracy of 94.5% with the composition of training data and test data is 8: 2. The classification process result shows that each sentence can produce more than one basic emotion class that forms compound expressions. The results of the visualization of compound expressions for each sentence can represent compound expressions.

Intisari—Ekspresi wajah dapat memperkuat informasi yang disampaikan dalam sebuah komunikasi interaktif. Dalam bidang pengembangan karakter virtual khusus karakter wajah, ekspresi wajah diperlukan untuk menghidupkan karakter virtual agar terlihat natural dan alami seperti manusia. Salah satu jenis ekspresi emosi adalah ekspresi emosi majemuk (kompleks) yang merupakan gabungan dari dua atau lebih emosi dasar. Misalnya, ekspresi emosi wajah kecewa merupakan gabungan dari emosi marah dan sedih. Ekspresi wajah dapat muncul oleh adanya rangsangan emosi, salah satunya dari makna kalimat. Makalah ini bertujuan untuk mengekstraksi data emosi dari kalimat bahasa Indonesia dengan menggunakan proses klasifikasi multi-label model *Convolution Neural Network (CNN)* sehingga menghasilkan ekspresi wajah majemuk yang diterapkan dalam animasi karakter virtual. Adapun kelas-kelas emosi yang digunakan dalam proses klasifikasi teks adalah enam kelas emosi dasar yang meliputi marah, jijik, takut, senang, sedih, dan terkejut. Berdasarkan hasil eksperimen, model CNN dapat

menghasilkan akurasi sebesar 94,5% dengan komposisi data latihan dan data uji adalah 8:2. Hasil dari proses klasifikasi tersebut menunjukkan bahwa setiap kalimat dapat menghasilkan lebih dari satu kelas emosi dasar yang membentuk emosi majemuk. Tiap kelas emosi majemuk tersebut divisualisasikan menggunakan karakter wajah virtual untuk membentuk ekspresi wajah yang dapat merepresentasikan emosi majemuk.

Kata Kunci—Klasifikasi Teks, *Convolutional Neural Network*, Emosi Majemuk, Kalimat Bahasa Indonesia.

I. PENDAHULUAN

Komunikasi *non-verbal* yang meliputi ekspresi wajah, intonasi suara, dan gerakan ekspresif lainnya berperan penting dalam komunikasi interaktif manusia [1]. Dalam industri animasi, *interactive game*, dan pengembangan animasi karakter virtual lainnya, dibutuhkan ekspresi wajah dengan tingkat kealamian yang tinggi. Tingkat kealamian yang tinggi mampu menunjang suatu informasi yang disampaikan dalam komunikasi interaktif. Dalam pembentukan ekspresi wajah, dibutuhkan data emosi yang terdapat dalam teks, audio, visual, dan komponen informasi lainnya. Data emosi dari suatu teks yang diekstraksi dengan baik dapat menghasilkan ekspresi wajah yang mampu mengekspresikan emosi yang selaras dengan informasi dari teks tersebut. Emosi adalah suatu rangsangan yang mendorong suatu tindakan/reaksi terhadap suatu objek atau kejadian [2]. Emosi dalam kalimat bahasa Indonesia dapat dikenali melalui kata sifat [3].

Bahasa Indonesia merupakan bahasa resmi yang digunakan masyarakat Indonesia untuk saling berkomunikasi. Struktur kalimat bahasa Indonesia meliputi subjek, predikat, objek, dan keterangan [4]. Terdapat dua jenis kalimat bahasa Indonesia berdasarkan jumlah klausa atau gagasannya, yaitu kalimat tunggal dan majemuk [5]. Kalimat tunggal adalah kalimat yang hanya memiliki satu klausa, sedangkan kalimat majemuk adalah kalimat dengan lebih dari satu klausa yang dipisahkan dengan kata penghubung. Kalimat berstruktur majemuk berpotensi memiliki emosi majemuk (gabungan) yang mengacu pada struktur kalimat tersebut yang memiliki lebih dari satu gagasan.

Ekstraksi data emosi dari kalimat bahasa Indonesia dapat dihasilkan melalui proses *supervised-learning* seperti klasifikasi teks. Klasifikasi teks berdasarkan enam kelas emosi dapat menghasilkan kategori-kategori emosi dari sebuah kalimat. Terdapat berbagai macam model yang dapat digunakan dalam proses klasifikasi teks, seperti *Naive Bayes*, K-NN, SVM, dan *neural networks*. Ekstraksi data emosi yang

¹Jurusan Teknik Biomedis Fakultas Teknik Universitas Dian Nuswantoro, Jl. Nakula I No. 5 - 11 Semarang 50131 (telp: 024-3517261; fax: 024-3569684; HP: 081331866866, e-mail: arifin@dsn.dinus.ac.id)

^{2,3}Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Dian Nuswantoro, Jln. Imam Bonjol 207 Semarang 50131 INDONESIA (telp: 024-3517261; fax: 024-3569684; e-mail: wisnu.aga@gmail.com, hanny.haryanto@dsn.dinus.ac.id)

telah dilakukan menggunakan model *multinomial Naïve Bayes* dapat menghasilkan kategori-kategori emosi berdasarkan frekuensi kemunculan kata sifat dari suatu kalimat [3]. Namun, emosi dalam suatu kalimat tidak hanya dinilai dari kemunculan kata sifat saja. Ekspresi emosi dapat dihasilkan dari letak kata sifat atau pola urutan kata dalam suatu kalimat.

Neural network merupakan arsitektur model yang dikembangkan dalam subbidang *deep learning* yang menggunakan *multi-layer perceptron* dalam penentuan keputusan. *Neural network* bekerja dengan memberikan bobot dari setiap data berulang kali hingga mendapatkan bobot yang sesuai. Dalam model *neural network* terdapat berbagai jenis arsitektur, seperti *Deep Neural Network (DNN)*, *Convolutional Neural Network (CNN)*, dan *Recurrent Neural Network (RNN)*. *CNN* adalah arsitektur model *neural network* yang sering dimanfaatkan dalam kasus *computer vision*. Secara umum, model tersebut terdiri atas *convolutional layer* dan *max pooling layer* yang diikuti dengan *layer-layer* tambahan. Proses pemilihan fitur terjadi dalam *convolutional layer* yang dilanjutkan pada *max pooling layer*. Namun, telah ada penelitian yang memanfaatkan *CNN* model dalam kasus klasifikasi teks [6].

Berkaitan dengan masalah dalam makalah ini, telah ada penelitian-penelitian tentang klasifikasi teks menggunakan kelas emosi, seperti penelitian tentang klasifikasi teks menggunakan model *multinomial Naïve Bayes* yang menghasilkan emosi majemuk berdasarkan enam kelas emosi [7]. Dalam penelitian tersebut dilakukan proses klasifikasi teks yang menghasilkan nilai probabilitas. Nilai probabilitas kemudian diseleksi menggunakan *threshold* yang dihasilkan dari metode persamaan batas dominan. Kelas emosi yang memiliki nilai probabilitas di atas nilai *threshold* dipilih sebagai kelas yang dominan, sehingga dalam setiap kalimat dapat terkandung satu emosi tunggal atau emosi majemuk yang dihasilkan dari gabungan emosi dominan. Namun, penelitian tersebut hanya menentukan emosi dari frekuensi kemunculan kata sifat.

Salah satu penelitian terkait lainnya adalah penelitian mengenai klasifikasi sentimen ulasan film Indonesia dengan menggunakan metode *CNN* [8]. *Dataset* yang digunakan adalah data teks berbahasa Indonesia yang merupakan hasil konversi dari data video. Data tersebut dikelompokkan berdasarkan aspek penilaian film dan diklasifikasi dengan metode *CNN*. Proses klasifikasi menggunakan enam *convolution layer* dengan satu *pooling layer* antara *layer* kedua dan ketiga. Rata-rata nilai akurasi dari proses klasifikasi menggunakan metode *CNN* adalah 83,7%.

Berdasarkan uraian di atas, makalah ini mengusulkan model *CNN* untuk proses ekstraksi emosi majemuk dari kalimat bahasa Indonesia. Model *CNN* dipilih karena memiliki kinerja yang lebih baik dalam menangani permasalahan klasifikasi teks dibandingkan model *multinomial Naïve Bayes*. Dalam makalah ini, emosi majemuk yang dihasilkan dari proses klasifikasi teks *multi-label* dapat menghasilkan gabungan emosi dari setiap kalimat yang mengandung emosi ganda. Klasifikasi *multi-label* memungkinkan setiap kelas emosi saling berkombinasi dengan kelas lain [9], [10]. Adapun kelas yang digunakan dalam proses



Gbr. 1 Kelas-kelas emosi dasar.

klasifikasi teks adalah enam emosi dasar yang meliputi marah, jijik, takut, senang, sedih, dan terkejut [2]. Pembentukan ekspresi wajah majemuk pada karakter virtual didasarkan pada kelas emosi majemuk yang dihasilkan.

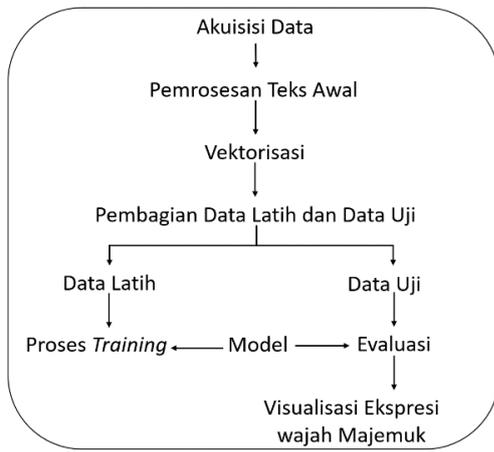
Penggunaan model *CNN* yang diusulkan diharapkan mampu meningkatkan akurasi dibandingkan penelitian sejenis [7], [8]. Model *CNN* dengan klasifikasi *multi-label* juga diharapkan mampu menggantikan dan menyederhanakan penggabungan model *multinomial Naïve Bayes* dan persamaan batas dominan [7] dengan hasil kinerja yang lebih baik.

II. KAJIAN EMOSI

Emosi digambarkan sebagai kondisi perasaan yang memengaruhi perilaku individu, perubahan fisiologis, dan pikiran. Emosi dari setiap individu dapat ditularkan pada individu lain, seperti emosi kemarahan dan ketakutan yang terangsang saat melihat dua orang sedang berkelahi atau terlibat emosi kemarahan. Demikian juga dengan rangsangan emosi kebahagiaan yang muncul saat melihat orang tersenyum. Menurut Paul Ekman, terdapat enam emosi dasar yang ada dalam diri manusia, seperti yang disajikan pada Gbr. 1, meliputi: 1) senang, yaitu emosi yang digambarkan sebagai rasa kepuasan terhadap sesuatu; 2) sedih, yaitu emosi yang menggambarkan rasa penyesalan terhadap suatu kegagalan; 3) marah, yaitu emosi yang menggambarkan rasa frustrasi atau ganjalan terhadap suatu objek; 4) takut, yaitu emosi yang menggambarkan rasa kekhawatiran terhadap sesuatu yang sedang atau akan terjadi, 5) jijik, yaitu emosi yang menggambarkan rasa penolakan terhadap objek yang menimbulkan rasa tidak nyaman; dan 6) terkejut, yaitu emosi yang menggambarkan reaksi yang terjadi terhadap sesuatu yang mendadak.

III. METODOLOGI

Tahapan penelitian untuk menghasilkan emosi majemuk secara garis besar terdiri atas proses akuisisi data, prapemrosesan teks, pembagian data latih dan data uji, klasifikasi teks, evaluasi, dan visualisasi ekspresi wajah



Gbr. 2 Tahapan penelitian ekstraksi data emosi majemuk.

majemuk pada karakter virtual. Urutan tahapan tersebut disajikan di Gbr. 2.

A. Akuisisi Data

Makalah ini menggunakan *dataset* kalimat bahasa Indonesia dengan struktur kalimat tunggal dan majemuk. *Dataset* diambil dari beberapa sumber seperti media sosial, meliputi unggahan dan komentar dalam Twitter, Facebook, Instagram, LinkedIn, dan media digital lainnya. Kemudian *dataset* diberi label secara manual oleh tim *linguistic analyst* berdasarkan kategori enam emosi dasar. Seluruh data yang didapatkan berjumlah 1.306 kalimat. Sampel kalimat dalam *dataset* disajikan pada Tabel I.

Untuk proses klasifikasi *multi-label*, maka setiap kalimat dapat memiliki label emosi lebih dari satu. Kelas emosi yang menjadi label ditandai dengan nilai angka 1, seperti yang disajikan dalam Tabel II. Proses pelabelan dilakukan dengan bantuan analisis dari ahli bahasa, sedangkan hasil proses pelabelan divalidasi oleh psikolog.

B. Pemrosesan Teks Awal

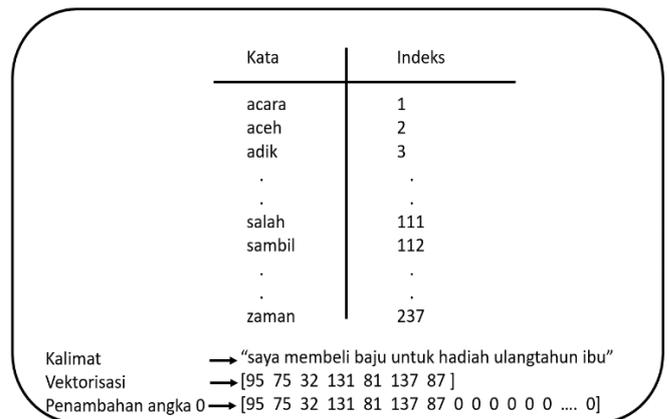
Dalam kasus *Natural Language Processing* (NLP), seperti klasifikasi teks, pemrosesan teks awal merupakan tahap penting untuk mengolah data teks menjadi data yang lebih terstruktur [11]. Proses ini dapat menentukan baik buruknya fitur yang dihasilkan dari suatu data. Adapun tahap pemrosesan teks awal dalam makalah ini meliputi *case-folding*, *filtering*, *stemming*, dan *unstemming*. *Case-folding* adalah tahap mengubah seluruh huruf dalam *dataset* menjadi huruf kecil (*lowercase*). Tahap *filtering* meliputi *filtering stopwords* dan *punctuation*. *Filtering stopwords* adalah mengambil kata-kata penting dan membuang kata-kata yang tidak memiliki makna, misalnya “dan”, “dari”, “di”, sedangkan *filtering punctuation* adalah proses menghapus tanda baca dan karakter khusus. *Stemming* adalah proses untuk mengubah setiap kata menjadi kata dasar dengan menghilangkan imbuhan. Proses *unstemming* merupakan proses kebalikan dari *stemming*, yaitu mempertahankan setiap kata dalam bentuk aslinya dengan tidak mengubahnya menjadi kata dasar. Dalam kalimat bahasa Indonesia, terdapat beberapa kata yang dapat berbeda arti dengan kata dasarnya, seperti kata “bertaruh” dengan kata “taruh” dan kata “cobaan” dengan kata “coba”.

TABEL I
SAMPEL KALIMAT

No.	Kalimat
#1	sungguh bagus sekali ulahmu tadi pagi
#2	betapa terkejutnya aku mendengar berita bodoh itu
#3	senang sekali dia tiba-tiba datang ke acara ulang tahunku
#4	aku takut melihat kecoa di dinding, apalagi kalau dia terbang
#5	teganya kalian meninggalkan aku sendirian di tempat sepi

TABEL II
LABEL KALIMAT

No.	Marah	Jijik	Takut	Senang	Sedih	Terkejut
#1	1	0	0	0	0	1
#2	1	0	0	0	0	1
#3	0	0	0	1	0	1
#4	0	1	1	0	0	0
#5	0	0	1	0	1	0



Gbr. 3 Vektorisasi.

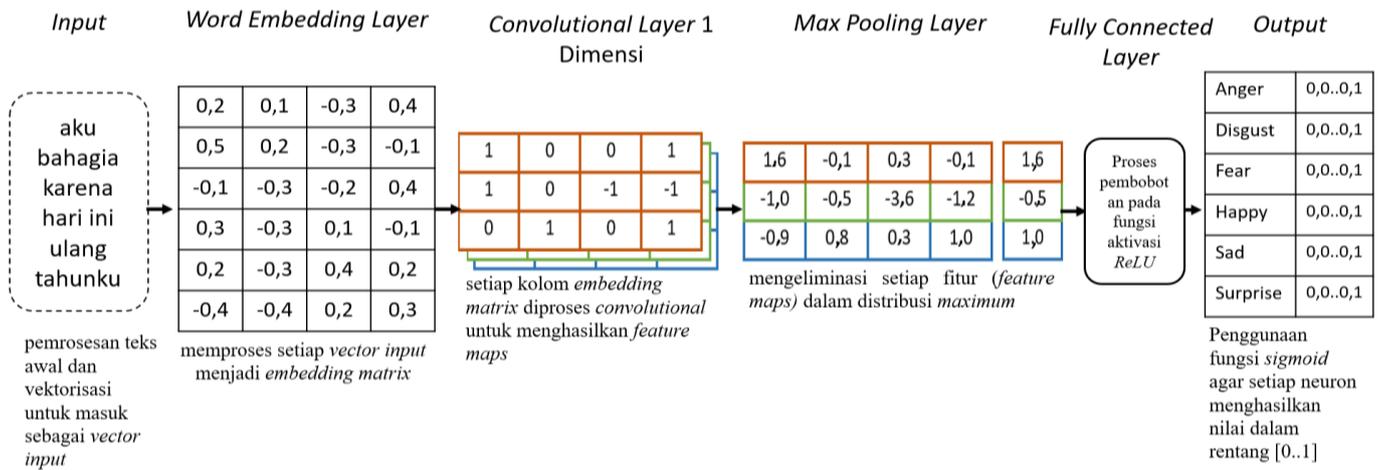
C. Vektorisasi

Vektorisasi adalah tahap konversi setiap kalimat dalam *dataset* menjadi sebuah vektor yang merepresentasikan kalimat tersebut. Dalam arsitektur *neural network*, *input* yang diterima hanya berupa angka, sehingga setiap kata harus dikonversi menjadi sebuah angka. Proses ini diilustrasikan dalam Gbr. 3.

Proses ini merepresentasikan setiap kata unik dalam *dataset* sebagai nilai integer yang disimpan dalam *vocab dictionary* [7]. Kemudian, setiap kalimat diubah menjadi vektor satu-dimensi berdasarkan *vocab dictionary*, sehingga dari setiap kalimat dihasilkan satu vektor yang merepresentasikan kalimat tersebut. Namun, setiap kalimat dalam *dataset* memiliki panjang yang berbeda-beda, sedangkan *sequential model* membutuhkan *input* dengan panjang yang tetap (*fixed length*). Oleh karena itu, setiap *vector input* ditambahkan dengan nilai 0 dalam proses *padding*. Nilai 0 tidak merepresentasikan kata apapun dalam *vocab dictionary*, sehingga tidak akan mengubah informasi dari kalimat tersebut.

D. Membagi Data Latih dan Data Uji

Hasil dari tahap vektorisasi adalah data vektor satu-dimensi yang berjumlah 1.306 vektor. Dalam proses pelatihan (*training*) dan evaluasi model, data dikelompokkan menjadi dua bagian, yaitu data latih dan data uji, secara acak. Komposisi



Gbr. 4 Arsitektur model Convolution Neural Network (CNN).

TABEL III
PEMBAGIAN DATA

Data	Porsi Data	Jumlah Data
Data Latih	80%	1.044
Data Uji	20%	262

pembagian data latih dan data uji adalah 80% data latih dan 20% data uji, sehingga dihasilkan data latih berjumlah 1.044 dan data uji berjumlah 262, seperti yang disajikan dalam Tabel III.

Data latih digunakan untuk proses pelatihan model, sedangkan data uji digunakan untuk proses evaluasi model dalam memprediksi *unseen* data atau data yang belum pernah dilihat oleh model.

E. Model CNN

CNN adalah salah satu model sekuensial yang umumnya digunakan untuk kasus *computer vision*. Namun, model ini juga mampu digunakan dalam kasus NLP [12], [13]. Arsitektur model CNN dalam makalah ini meliputi *word embedding layer*, *one-dimensional convolution layer*, *max-pooling layer*, *fully connected layer (dense layer)*, dan *output layer* [14], yang diilustrasikan dalam Gbr. 4.

Dalam model CNN, setiap kalimat yang telah diproses dalam tahap-tahap sebelumnya meliputi pemrosesan teks awal dan vektorisasi, yang selanjutnya masuk sebagai *vector input* dalam *embedding layer*, layer ini menghasilkan *embedding matrix* dari setiap *vector input*. Kemudian, dari setiap kolom dalam *embedding matrix* dilakukan proses *convolutional* dalam *convolutional layer* yang menghasilkan *feature maps*. Seluruh *feature maps* yang dihasilkan diteruskan dalam *max-pooling layer* untuk mengeliminasi setiap fitur dalam distribusi *maximum*. Setelah dihasilkan *feature maps* yang dieliminasi dalam *max-pooling layer*, kemudian setiap *feature maps* diteruskan dalam setiap neuron *fully connected layer* yang menggunakan fungsi aktivasi *Rectified Linear Unit (ReLU)*. Dalam *fully connected layer* dilakukan proses pembobotan, sedangkan fungsi aktivasi *ReLU* berfungsi untuk mengubah nilai negatif menjadi nilai 0. Hasil dari *layer* tersebut kemudian diteruskan dalam *output layer* yang terdiri atas enam neuron

dengan fungsi aktivasi *sigmoid*. Penggunaan fungsi *sigmoid* bertujuan agar setiap neuron dapat menghasilkan nilai dalam rentang [0..1]. Setiap neuron mewakili label masing-masing emosi dasar.

F. Embedding Layer

Setiap data *input* diproses dalam *word embedding layer*. Layer ini berfungsi untuk mengubah setiap *vector input* menjadi *embedding matrix*. Proses ini memetakan setiap kata dalam dimensi vektor yang lebih tinggi. Setiap kata ke-*i* dalam setiap kalimat dilambangkan sebagai $x_i \in R^k$, dengan *k* adalah dimensi *embedding matrix*. *Embedding matrix* dilambangkan dalam (1).

$$x = [x_1, x_2, \dots, x_l] \tag{1}$$

dengan $x \in R^{l \times k}$ dan *l* adalah panjang dari kalimat [6]. *Embedding matrix* dapat diperbarui seiring dengan proses pelatihan pada model.

G. Convolutional Layer

Convolutional layer berfungsi untuk mengekstrak fitur dari *embedding layer* dalam operasi konvolusi satu-dimensi. Proses konvolusi satu dimensi menggunakan *kernel filter* untuk mendeteksi fitur dari setiap data dengan menghasilkan *feature map*. Operasi konvolusi dirumuskan dalam (2).

$$m_j = f(w \cdot x_{i:i+s-1} + b) \tag{2}$$

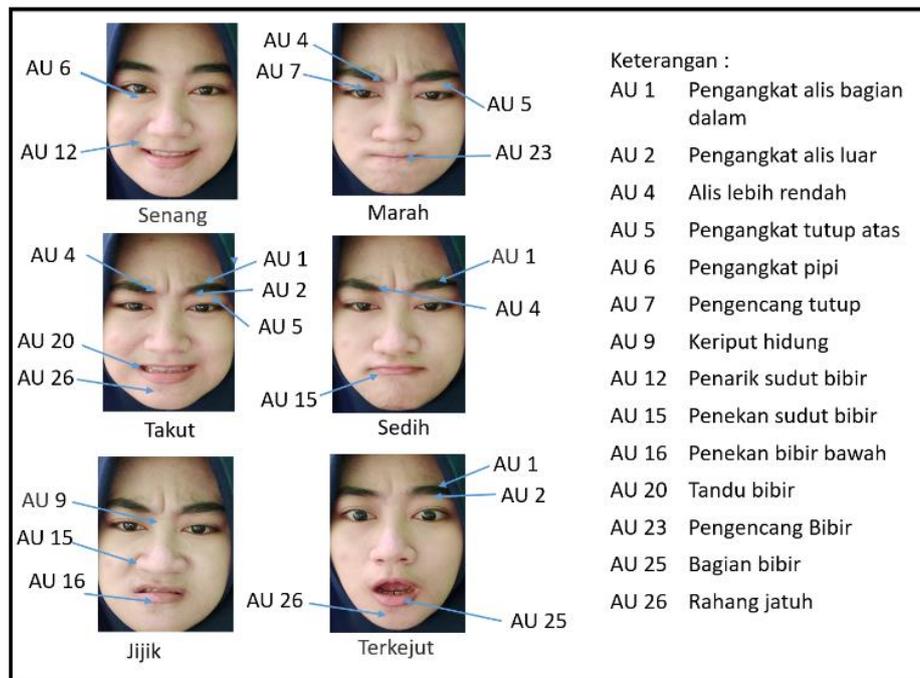
dengan $w \in R^{s \times k}$, *s* adalah ukuran filter, notasi \cdot adalah operasi *dot matrix*, *b* adalah nilai *bias*, *f* adalah fungsi nonlinear, dan $x_{i:i+s-1}$ dirumuskan dalam (3).

$$x_{i:i+s-1} = x_i + x_{i+1} + \dots + x_{i+s-1} \tag{3}$$

dengan notasi $+$ adalah *concatenation of vector lines*. Selanjutnya, untuk menghasilkan *feature map* digunakan rumusan dalam (4).

$$m = [m_1, m_2, \dots, m_{l-s+1}] \in R^{l-s+1} \tag{4}$$

Dalam *convolutional layer* satu-dimensi, terdapat *n-filter* yang menghasilkan *n-feature map*, sehingga seluruh proses dalam *convolution layer* dapat dinotasikan dalam (5).



Gbr. 5 Action Unit (AU) tiap ekspresi emosi dasar.

$$M = [m_1; m_2; \dots; m_n] \in R^{n \times (l-s+1)} \quad (5)$$

dengan M adalah matriks *output* dari *convolutional layer* dan tanda *semicolons* mewakili *line vector*.

H. Max Pooling Layer

Feature maps yang dihasilkan dari *convolutional layer* satu-dimensi diteruskan dalam *max-pooling layer*. Dalam *max-pooling layer*, setiap *feature map* diproses dalam distribusi *maximum* yang menghasilkan nilai tertinggi dari setiap *feature map*. Proses ini membantu mempercepat proses pelatihan tanpa mengorbankan informasi penting dari setiap data.

I. Fungsi Aktivasi

Dalam arsitektur *neural network* terdapat beberapa fungsi aktivasi yang digunakan untuk mengubah nilai dalam setiap neuron [15]. Adapun fungsi aktivasi yang digunakan dalam *layer-layer* berikutnya meliputi *ReLU* dan *sigmoid*. *ReLU* merupakan fungsi aktivasi nonlinear yang dapat mengubah nilai negatif menjadi nilai 0, yang dinotasikan dalam (6).

$$ReLU(z) = \max(0, z) \quad (6)$$

dengan z adalah nilai *input* dan $\max()$ adalah operasi *maximum*. Dalam fungsi aktivasi *ReLU*, nilai *output* memiliki rentang antara 0 hingga positif tak terhingga. Fungsi aktivasi *ReLU* umumnya dimanfaatkan untuk menghindari nilai negatif.

Sigmoid adalah fungsi aktivasi nonlinear yang mengubah nilai *input* menjadi nilai dengan rentang 0 hingga 1, yang dinotasikan dalam (7).

$$sigmoid(z) = \frac{1}{1+e^{-z}} \quad (7)$$

dengan z adalah nilai *input* dan e adalah nilai eksponensial. Fungsi aktivasi *sigmoid* digunakan pada *output layer*.

J. Output Layer

Output layer dalam arsitektur model makalah ini memiliki enam neuron sesuai dengan jumlah kelas emosi yang digunakan. Fungsi aktivasi *sigmoid* digunakan dalam *output layer* untuk mendapatkan nilai independen dari setiap neuron, sehingga setiap neuron yang mewakili masing-masing emosi dapat memiliki nilai dalam rentang [0..1]. Nilai independen dari setiap nilai memungkinkan terbentuknya emosi gabungan dari beberapa emosi yang dihasilkan.

K. Visualisasi Ekspresi Wajah Majemuk

Visualisasi ekspresi wajah majemuk pada karakter virtual tiga-dimensi dibangun berdasarkan formasi *Action Unit* (AU) di wajah. AU merupakan gerakan dasar otot atau kelompok otot di wajah. Formasi AU di wajah mengacu pada *Facial Action Coding System* (FACS), yaitu kodifikasi yang dikembangkan oleh Eckman dan Friesen [16], yang menyatakan bahwa setiap ekspresi wajah memicu aktivitas otot (AU) tertentu. Dengan demikian, dapat diketahui jumlah dan posisi AU wajah yang berfungsi sebagai kontrol untuk mengendalikan gerakan pada wajah sehingga karakter virtual wajah menjadi lebih hidup. Setiap ekspresi emosi dasar didefinisikan sebagai satu kesatuan gerakan dari kelompok AU tertentu seperti disajikan dalam Gbr. 5. Untuk menghasilkan ekspresi wajah majemuk, dilakukan penggabungan AU yang didefinisikan oleh FACS [17].

Penggabungan kelompok AU dari masing-masing kelas emosi dasar menjadi kelompok AU pada ekspresi wajah majemuk mengikuti pola seperti yang tersaji di Tabel IV [17]. Setiap AU diimplementasikan ke dalam animasi *blend shape* untuk mengontrol gerakan otot wajah. Masing-masing AU memiliki kisaran nilai antara 0 sampai dengan 1 ([0..1]). Nilai

TABEL IV
CONTOH BEBERAPA POLA PENGGABUNGAN AU UNTUK MEMBENTUK
EKSPRESI WAJAH MAJEMUK

Kelompok AU Emosi Dasar #1	Kelompok AU Emosi Dasar #2	Kelompok AU Ekspresi Wajah Majemuk
Senang: AU 6, AU 12	Terkejut: AU 1, AU 2, AU 5, AU 25, AU 26	Terkejut Senang: AU 1, AU 2, AU 5, AU 12, AU 25, AU 26
Terkejut: AU 1, AU 2, AU 5, AU 25, AU 26	Takut: AU 1, AU 2, AU 4, AU 5, AU 7, AU 20, AU 26	Terkejut sekaligus Takut: AU 1, AU 2, AU 5, AU 20, AU 25, AU 26
Terkejut: AU 1, AU 2, AU 5, AU 25, AU 26	Jijik: AU 9, AU 15, AU 16	Terkejut sekaligus Jijik: AU 1, AU 2, AU 5, AU 9, AU 15, AU 16
Jijik: AU 9, AU 15, AU 16	Takut: AU 1, AU 2, AU 4, AU 5, AU 7, AU 20, AU 26	Jijik Menakutkan: AU 1, AU 2, AU 4, AU 5, AU 9, AU 15, AU 20, AU 26
Marah: AU 4, AU 5, AU 7, AU 23	Sedih: AU 1, AU 4, AU 15	Marah sekaligus Sedih: AU 1, AU 4, AU 5, AU 7, AU 15, AU 23

probabilitas tiap kelas emosi yang telah diperoleh dari hasil proses klasifikasi digunakan sebagai dasar untuk nilai tiap AU tersebut.

Implementasi nilai AU untuk visualisasi ekspresi wajah majemuk mengikuti aturan: 1) untuk AU yang hanya terdapat di tiap kelas emosi dasar, nilai AU tersebut menggunakan hasil nilai probabilitas dari kelas emosi tersebut; 2) untuk AU yang terdapat di kedua kelas emosi yang digabungkan, nilai AU tersebut menggunakan nilai probabilitas terbesar dari dua kelas emosi yang digabungkan.

IV. HASIL DAN PEMBAHASAN

Dalam bagian ini dipaparkan proses pelatihan dan evaluasi model, serta visualisasi hasil ekstraksi emosi dengan menggunakan karakter virtual.

A. Pelatihan dan Evaluasi Model

Dalam proses pelatihan model, data latih diproses dalam arsitektur model untuk menentukan bobot yang sesuai dengan fitur data. Proses ini menggunakan *Binary Cross Entropy* (BCE) sebagai *loss function*, yang dinotasikan dalam (8).

$$BCE = \frac{1}{n} \sum_{i=1}^n y_i \cdot \log \hat{y}_i + (1 - y) \cdot \log(1 - \hat{y}_i) \quad (8)$$

dengan \hat{y}_i adalah *predicted label*, y_i adalah *actual label*, dan n adalah jumlah label. *Loss function* berfungsi untuk menghitung jarak perbedaan antara *predicted label* terhadap *actual label* [18]. Setelah bobot yang sesuai dengan fitur yang terdapat dalam data latih didapatkan, selanjutnya dilakukan evaluasi menggunakan data latih.

Setelah melalui tahap pelatihan, dilanjutkan dengan evaluasi model. Dalam tahap ini, data latih diprediksi label emosinya menggunakan model CNN. Label prediksi kemudian

TABEL V
CONFUSION MATRIX

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP (True Positive)	FP (False Positive)
	Negative (0)	FN (False Negative)	TN (True Negative)

TABEL VI
HASIL EVALUASI MODEL DENGAN MENGHITUNG *PRECISION*, *RECALL*, *F1-SCORE*, DAN *ACCURACY* UNTUK TIAP KELAS EMOSI

Kelas	Precision	Recall	F1-Score	Accuracy
Marah	86,0%	76,0%	81,0%	91,0%
Jijik	94,0%	100,0%	97,0%	99,0%
Takut	100,0%	79,0%	88,0%	96,0%
Senang	88,0%	90,0%	89,0%	93,0%
Sedih	98,0%	83,0%	90,0%	92,0%
Terkejut	94,0%	71,0%	81,0%	97,0%
Rata-rata	93,3%	83,1%	87,6%	94,5%

dibandingkan dengan label aktual dalam tabel *confusion matrix* seperti disajikan di Tabel V [19].

Berdasarkan hasil *confusion matrix*, dapat dihitung nilai *precision*, *recall*, *f1-score*, dan *accuracy*. *Precision* adalah rasio prediksi positif benar (*true positive*) dibandingkan dengan keseluruhan hasil yang diprediksi positif. *Recall* adalah rasio prediksi positif benar dibandingkan dengan keseluruhan data positif benar. *F1-score* merupakan perbandingan rata-rata *precision* dan *recall* yang dibobotkan, sedangkan nilai akurasi dapat diartikan sebagai tingkat kedekatan atau ketepatan antara nilai yang diprediksi dengan nilai aktual. Semakin tinggi nilai akurasi dari suatu proses klasifikasi menunjukkan bahwa kinerja proses klasifikasi semakin baik. Nilai akurasi dapat dihitung dengan menggunakan (11).

$$Precision = \frac{TP}{(TP+FP)} \quad (9)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (10)$$

$$Accuracy = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (11)$$

$$F1\ score = 2 * \frac{Recall * Precision}{Recall + Precision} \quad (12)$$

dengan TP adalah *true positive*, FP adalah *false positive*, TN adalah *true negative*, dan FN adalah *false negative*. Hasil evaluasi model terhadap data uji dengan menghitung *precision*, *recall*, *f1-score*, dan *accuracy* disajikan pada Tabel VI.

Dari Tabel VI terlihat bahwa nilai *precision* dari setiap kelas menghasilkan rata-rata 93,3%. Nilai *recall* menghasilkan rata-rata 83,1%, nilai *f1-score* menghasilkan rata-rata sebesar 87,6%, dan akurasi yang dihasilkan sebesar 94,5%.

Berdasarkan hasil evaluasi model, makalah ini berhasil meningkatkan kinerja proses klasifikasi dibandingkan dengan penelitian sejenis yang telah ada [7], [8]. Hal ini dapat dilihat dari perbandingan nilai akurasi seperti yang disajikan pada

TABEL VII
PERBANDINGAN KINERJA PROSES KLASIFIKASI DARI BEBERAPA PENELITIAN

Materi Penelitian	Model yang Digunakan	Nilai Akurasi
Klasifikasi emosi [7]	<i>Multinomial Naïve Bayes</i>	75,5%
Klasifikasi sentimen [8]	<i>Convolution Neural Network</i>	83,7%
Ekstraksi emosi majemuk kalimat bahasa Indonesia	<i>Convolution Neural Network</i> untuk klasifikasi multilabel	94,5%

TABEL VIII
BEBERAPA KALIMAT YANG DIGUNAKAN DALAM SAMPEL PENGUJIAN

No.	Kalimat
#1	Budi merasa sangat murka karena uangnya dicopet dan dia masih harus memikirkan untuk mengembalikan uang tersebut.
#2	aku mendapat kado dari ayah, namun betapa terkejutnya aku ternyata ini adalah sepatu yang selama ini aku inginkan.
#3	sejak kecil dia sudah menjadi anak yatim piatu sehingga diasuh pamannya, namun dia sekarang hidup bahagia karena menjadi orang kaya dan bergelimang harta.

TABEL IX
NILAI PROBABILITAS DOMINAN SETIAP KELAS EMOSI

No.	Marah	Jijik	Takut	Senang	Sedih	Terkejut
#1	0,42	0	0	0	0,58	0
#2	0	0	0	0,26	0	0,74
#3	0	0	0	0,46	0,54	0

Tabel VII. Kinerja klasifikasi yang dibandingkan dari beberapa penelitian memiliki jenis data yang sama, yaitu teks berbahasa Indonesia.

B. Eksperimen Visual

Pada tahap ini dilakukan pengujian terhadap beberapa kalimat sampel seperti yang terlihat di Tabel VIII. Hasil pengujian berupa prediksi label emosi menggunakan model CNN disajikan pada Tabel IX.

Tabel IX menunjukkan bahwa setiap kalimat yang diujikan menghasilkan nilai probabilitas dominan pada lebih dari satu kelas emosi. Nilai probabilitas tersebut merepresentasikan tingkat emosi dari tiap kalimat. Kalimat yang memiliki nilai probabilitas dominan lebih dari satu kelas emosi menunjukkan bahwa kalimat tersebut mengandung emosi majemuk. Selanjutnya, kalimat yang mengandung emosi majemuk tersebut divisualisasikan ke dalam karakter virtual tiga-dimensi sehingga terbentuk ekspresi wajah majemuk seperti yang tersaji pada Tabel X.

Visualisasi ekspresi wajah majemuk dilakukan dengan mengimplementasikan proses transformasi nilai probabilitas tiap kelas emosi dominan ke dalam nilai AU yang terkait dari masing-masing kelas emosi. Kelompok AU dari tiap kelas emosi digabungkan dengan kelompok AU dari kelas emosi yang lain dengan mengikuti aturan pola penggabungan AU. Kemudian, implementasi ekspresi majemuk ke dalam karakter virtual tiga-dimensi dilakukan berdasarkan animasi *blend*

TABEL X
KELAS-KELAS EMOSI DOMINAN DAN VISUALISASI EKSPRESI WAJAH MAJEMUK

No.	Kelas Emosi Dominan	Visualisasi
#1	Marah (0,42) Sedih (0,58)	
#2	Senang (0,26) Terkejut (0,74)	
#3	Senang (0,46) Sedih (0,54)	

shape dari masing-masing nilai AU sebagai pengontrol gerakan otot wajah yang membentuk ekspresi wajah majemuk.

V. KESIMPULAN

Berdasarkan hasil eksperimen dan evaluasi model, dapat disimpulkan bahwa model CNN yang digunakan untuk klasifikasi multilabel terhadap kalimat-kalimat berbahasa Indonesia dapat menghasilkan kelas emosi majemuk dengan nilai akurasi sebesar 94,5%. Klasifikasi multilabel menggunakan model CNN yang diusulkan terbukti mampu meningkatkan kinerja klasifikasi dibandingkan dengan penelitian sejenis yang telah ada. Klasifikasi multilabel model CNN juga terbukti mampu menyederhanakan penggabungan model *multinomial Naïve Bayes* dan persamaan batas dominan untuk memperoleh kelas majemuk seperti yang dilakukan di penelitian sebelumnya dengan kinerja klasifikasi yang lebih baik.

Kelas emosi majemuk yang dihasilkan dari makalah ini dapat digunakan sebagai dasar untuk membentuk ekspresi wajah majemuk. Implementasi ekspresi wajah majemuk pada karakter virtual dapat dihasilkan dengan mentransformasikan

nilai probabilitas tiap kelas dari hasil proses klasifikasi ke dalam nilai AU. Visualisasi ekspresi wajah majemuk pada karakter virtual terbentuk dari animasi *blend shape* dari kelompok AU yang terkait.

Hasil visualisasi ekspresi wajah pada karakter virtual dalam makalah ini terlihat masih kurang natural, terutama pada gerakan otot wajah yang membentuk kerutan di sekitar hidung dan alis. Oleh karena itu, diperlukan formula untuk menstransformasikan nilai probabilitas ke dalam nilai AU, sehingga nilai AU yang digunakan di animasi *blend shape* sesuai dengan tingkat emosi.

UCAPAN TERIMA KASIH

Diucapkan banyak terima kasih kepada Kementerian Riset, Teknologi dan Pendidikan Tinggi melalui Direktorat Riset dan Pengabdian Kepada Masyarakat yang telah mendukung pendanaan penelitian ini melalui program hibah bersaing dengan skema Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT). Ucapan terima kasih juga disampaikan kepada seluruh pimpinan Universitas Dian Nuswantoro yang telah memberikan dorongan, dukungan, spirit, dan fasilitas sarana prasarana sehingga penelitian ini dapat diselesaikan.

REFERENSI

- [1] T.I. Kusumawati, "Komunikasi Verbal dan Nonverbal," *Jurnal Pendidik dan Konseling*, Vol. 6, No. 2, hal. 83–98, 2016.
- [2] P. Ekman dan D. Cordaro, "What is Meant by Calling Emotions Basic," *Emot. Rev.*, Vol. 3, No. 4, hal. 364–370, Okt. 2011.
- [3] J.E. Prawitasari, "Mengenal Emosi Melalui Komunikasi Nonverbal," *Bul. Psikol.*, Vol. 3, No. 1, hal. 27–43, 2016.
- [4] B. Denafri, "Struktur Informasi Kalimat Bahasa Indonesia," *J. Ilmiah Kebahasaan dan Kesastraan*, Vol. 6, No. 1, hal. 43–49, Jun. 2018.
- [5] S.S.T.W. Sasangka, *Seri Penyuluhan Bahasa Indonesia: Kalimat*, Jakarta, Indonesia: Pusat Pembinaan dan Pemasyarakatan; Badan Pengembangan dan Pembinaan Bahasa; Kementerian Pendidikan dan Kebudayaan, 2015.
- [6] Y. Kim, "Convolutional Neural Networks for Sentence Classification," *Proc. the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, hal. 1746–1751.
- [7] W. Agastya dan Aripin, "Pemetaan Emosi Dominan pada Kalimat Majemuk Bahasa Indonesia Menggunakan Multinomial Naïve Bayes," *J. Nas. Teknik Elektro dan Teknologi Informasi (JNTETI)*, Vol. 9, No. 2, hal. 171–179, 2020.
- [8] N.A. Shafira dan Irhamah, "Klasifikasi Sentimen Ulasan Film Indonesia dengan Konversi Speech-to-Text (STT) Menggunakan Metode Convolution Neural Network (CNN)," *Jurnal Sains dan Seni ITS*, Vol. 9, No. 1, hal. 95–101, 2020.
- [9] I.C. Irsan dan M.L. Khodra, "Hierarchical Multi-Label News Article Classification with Distributed Semantic Model Based Features," *Int. Journal of Advances in Intelligent Informatics*, Vol. 5, No. 1, hal. 40–47, 2019.
- [10] S. Sirattanajakarin dan P. Thusaranon, "Movie Genre in Multi-Label Classification Using Semantic Extraction from Only Movie Poster," *ICCCM 2019: Proc. the 2019 7th International Conference on Computer and Communications Management*, 2019, hal. 23–27.
- [11] A. Jain, G. Kulkarni, dan V. Shah, "Natural Language Processing," *Int. Journal of Computer Sciences and Engineering (IJCSSE)*, Vol. 6, No. 1, hal. 161–167, 2018.
- [12] N.I. Widiastuti, "Convolution Neural Network for Text Mining and Natural Language Processing," *IOP Conference Series: Materials Science and Engineering*, Vol. 662, No. 5, hal. 1–6, 2019.
- [13] P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, dan B. Xu, "Text Classification Improved by Integrating Bidirectional LSTM with Two-Dimensional Max Pooling," *Proc. COLING 2016 - 26th International Conference on Computational Linguistics*, 2016, hal. 3485–3495.
- [14] Satyabrata, S. Chakraborty, dan H.-C. Kim, "Convolution Neural Network-Based Model for Web-Based Text Classification," *Int. Journal of Electrical and Computer Engineering (IJECE)*, Vol. 9, No. 6, hal. 5185–5191, 2019.
- [15] G. Onwujekwe dan V. Yoon, "Activation Functions and their Impact on the Training and Performance of Convolution Neural Network Models," *Proc. the Americas Conference on Information Systems (AMCIS)*, 2020, hal. 1–5.
- [16] P. Ekman dan W.V. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Palto Alto, USA: Consulting Psychologies Press, 1978.
- [17] S. Du, Y. Tao, dan A.M. Martinez, "Compound Facial Expressions of Emotion," *Proc. of the National Academy of Sciences*, 2014, hal. 1454–1462.
- [18] F. Nie, Z. Hu, dan X. Li, "An Investigation for Loss Functions Widely Used in Machine Learning," *Communication in Information and System*, Vol. 18, No. 1, hal. 37–52, 2018.
- [19] I. Dutsch dan G. Gediga, "Confusion Matrices and Rough Set Data Analysis," *Journal of Physics: Conf. Series*, Vol. 1229, hal. 1–6, 2019.