

Mengoptimalkan Akurasi pada Klasifikasi Emosi Majemuk Berdasarkan Semantik Kalimat Menggunakan XLM-RoBERTa

Aripin¹, Steven Adi Santoso², Hanny Haryanto³

¹ Program Studi Teknik Biomedis Fakultas Teknik Universitas Dian Nuswantoro, Semarang, 50131 INDONESIA (tel.: 024-70793733; fax: 024-3569684; email: ¹arifin@dsn.dinus.ac.id)

^{2,3} Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Dian Nuswantoro, Semarang 50131 INDONESIA (tel.: 024-3517261; fax: 024-3560567; email:

²sstevenadi@gmail.com, ³hanny.haryanto@dsn.dinusl.ac.id)

[Diterima: 9 Desember 2022, Revisi: 2 Februari 2023]

Corresponding Author: Aripin

INTISARI — Emosi dasar dibagi menjadi enam, yaitu marah, sedih, senang, jijik, kaget, dan takut. Gabungan lebih dari satu emosi dasar dapat menciptakan sebuah emosi baru, yaitu emosi majemuk. Emosi majemuk dapat diimplementasikan untuk *chat-bot*, penerjemahan bahasa, *text summarization*, dan sebagainya. Penelitian mengenai klasifikasi emosi berdasarkan teks bahasa Indonesia telah banyak dilakukan dengan menggunakan beberapa model tradisional, seperti *multinomial naïve Bayes*, *SVM*, *k-nearest neighborhood*, dan *term frequency-inverse document frequency* (TF-IDF). Penelitian tersebut memiliki kelemahan, antara lain kinerja yang kurang optimal karena model hanya dapat mengklasifikasi dari data yang telah dipelajarinya, diperlukan pemrosesan teks terlebih dahulu, dan diperlukannya waktu yang lama dalam proses pelatihan dengan data berukuran besar. Penelitian ini bertujuan untuk mengatasi beberapa kelemahan penelitian sebelumnya dengan menggunakan model *cross-lingual language model-robustly optimized bidirectional encoder representations from transformers approach* (XLM-RoBERTa) untuk mengklasifikasi emosi majemuk berdasarkan semantik atau makna kalimat dan kata. XLM-RoBERTa merupakan sebuah model *transformer* yang dapat mengetahui sebuah makna kata dari *attention mechanism* pada kata tersebut dan merupakan sebuah vektor yang merepresentasikan sebuah konteks atau makna kata. *Attention mechanism* merupakan sebuah representasi kata berbentuk vektor untuk mengetahui penggunaan dan posisi kata pada suatu kalimat dan merupakan cara agar model dapat mengetahui makna dari sebuah kata. Dengan *attention mechanism*, model dapat melihat pola kalimat dari penggunaan kata dan mengklasifikasikan kalimat tersebut sesuai dengan pola dan urutan kata, sehingga semantik kalimat dapat diketahui. Hasil eksperimen menunjukkan bahwa model yang diusulkan mampu mengklasifikasi teks berbahasa Indonesia ke dalam kelas-kelas emosi dasar dan kombinasinya sebagai dasar pembentukan emosi majemuk dengan akurasi sebesar 95,56%. Nilai akurasi ini merupakan nilai akurasi yang lebih unggul dibandingkan dengan penelitian klasifikasi kelas emosi majemuk dengan menggunakan model tradisional.

KATA KUNCI — Klasifikasi Emosi Majemuk, Kalimat Berbahasa Indonesia, *Multi-Label*, Semantik Kalimat, XLM-RoBERTa.

I. PENDAHULUAN

Emosi merupakan sebuah perasaan yang dialami oleh seseorang atau sesuatu. Emosi juga merupakan reaksi seseorang terhadap suatu kejadian [1]. Reaksi tersebut dapat berupa perasaan senang, sedih, takut, dan sebagainya. Emosi juga merupakan sesuatu yang unik bagi makhluk hidup. Manusia berkomunikasi menggunakan emosi sebagai penunjuk perasaan seseorang. Emosi juga dapat memengaruhi perilaku seseorang, seperti ketika seseorang merasa senang, maka orang tersebut akan selalu tersenyum ataupun tertawa. Begitu juga sebaliknya, jika seseorang merasakan sedih, maka orang tersebut akan menjadi tidak mau bersosialisasi dengan orang lain atau bahkan mengurung diri.

Emosi tidak hanya dapat ditunjukkan melalui perilaku seseorang, tetapi dapat juga ditunjukkan melalui suara ataupun teks, seperti teks dalam cerita atau puisi. Emosi dapat diketahui dengan memperhatikan tanda yang menyertai suatu emosi, yaitu *cues to emotions* [2]. Dijelaskan juga bahwa tanda-tanda emosi itu meliputi tuturan vokal atau verbal (*verbal cues*) dan perilaku nonverbal atau ekspresi wajah (*facial cues*). Pertanda verbal dapat terlihat dari penuturan kata-kata yang digunakan dalam berkomunikasi. Penelitian tersebut membuktikan bahwa bahasa dapat digunakan untuk melakukan interaksi dengan nilai-nilai sosial serta budaya penuturnya. Salah satu contoh komunikasi verbal adalah teks.

Dari teks tersebut akan dilakukan proses klasifikasi untuk mendapatkan emosi yang terkandung di dalam teks. Proses

klasifikasi sendiri merupakan proses *supervised-learning*, yaitu membuat komputer belajar dari data yang sudah diberi label atau data yang menjadi sebuah fitur untuk menebak sebuah jawaban. Proses klasifikasi dilakukan dengan teks yang memiliki label berjumlah enam, sesuai dengan emosi utama [3]. Banyak sekali penelitian yang melakukan klasifikasi teks dengan menggunakan metode yang berbeda-beda, seperti *support vector machine* (SVM), *naïve Bayes*, *random forest*, dan *convolutional neural network* [4]–[7]. Model *cross-lingual language model-robustly optimized bidirectional encoder representations from transformers approach* (XLM-RoBERTa) dapat meningkatkan kinerja klasifikasi ujaran kebencian (*hate speech*) teks berbahasa Indonesia menjadi 89,52% dari penelitian sebelumnya, yaitu 77,36%, yang menggunakan *long short term memory* (LSTM) [1].

XLM-RoBERTa merupakan sebuah model *transformer* yang dibuat oleh para peneliti Facebook pada tahun 2019 [8]. XLM-RoBERTa merupakan sebuah model yang lebih unggul dari model sebelumnya, yaitu *multilingual bidirectional encoder representations from transformers* (mBERT) yang merupakan model *transformer* multilingual [9]. Model mBERT adalah evolusi dari model BERT awal, yaitu model *transformer* yang dapat mengidentifikasi *attention* dalam sebuah teks, sehingga dapat memprediksi kata-kata selanjutnya berdasarkan *attention* setiap kata [10].

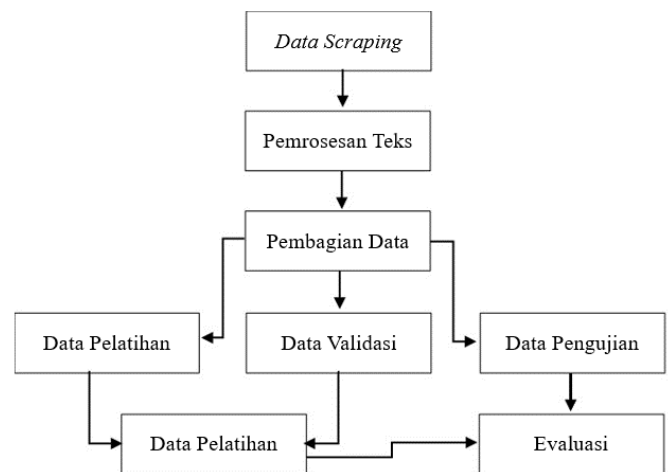
Beberapa penelitian yang dilakukan sebelumnya juga sudah berhasil meningkatkan kinerja model dalam melakukan proses

klasifikasi teks [11] dengan menggunakan *transformer*, seperti melakukan klasifikasi *offensive text*, ulasan Internet Movie Database (IMDb), dan analisis sentimen. Kinerja model *transformer* mengalahkan model-model tradisional pada umumnya. Pada penelitian tersebut, model dapat memberikan hasil yang baik terhadap proses klasifikasi teks, tetapi masih terdapat beberapa kekurangan. Model melakukan klasifikasi teks dengan menggunakan *single label*, yaitu data hanya memiliki dua label, yakni 1 dan 0.

Terdapat penelitian serupa dengan menggunakan model *naïve Bayes* untuk mengekstraksi data emosi dari sebuah data teks [12]. Model *naïve Bayes* dapat menghasilkan akurasi sebesar 75,47%. Teks diklasifikasi dengan menggunakan *multinomial naïve Bayes* sehingga menghasilkan nilai probabilitas dari enam kelas emosi. Langkah selanjutnya adalah menggunakan nilai ambang untuk menemukan kelas-kelas emosi dominan. Dalam penelitian tersebut, model dilatih menggunakan data teks sebanyak 2.187 data dengan rata-rata 364 data per label. Kekurangan dari model ini adalah model tidak dilatih dengan data yang besar dan terbatasnya kosakata dari data yang sedikit. Model memiliki kinerja yang baik terhadap data yang pernah dihadapinya. Namun, model memiliki kinerja yang kurang baik apabila data berbentuk struktur kalimat yang mengandung kata-kata yang berbeda, seperti ‘memakan’ dan ‘makannya’ yang memiliki arti berbeda, tetapi karena proses *stemming*, kedua kata tersebut menjadi hanya memiliki satu arti.

Penelitian yang diusulkan ini bertujuan untuk mengurangi kekurangan dari model penelitian sebelumnya [13], dengan menggunakan model *transformer* multibahasa yaitu XLM-RoBERTa. Dengan model ini, dilakukan proses klasifikasi terhadap teks berbahasa Indonesia dengan mempelajari makna dan posisi setiap kata dari suatu kalimat, sehingga akurasi klasifikasi dapat meningkat secara signifikan. Pada model tradisional, biasanya digunakan *term frequency-inverse document frequency* (TF-IDF) yang memiliki representasi berbeda-beda sesuai dengan frekuensi kemunculan kata dalam kalimat atau teks tersebut. Beberapa kelebihan model XLM-RoBERTa antara lain: 1) mampu melakukan proses klasifikasi dengan pola atau kata-kata yang berbeda dan tetap mendapatkan hasil prediksi yang benar; 2) memiliki pengetahuan kata berbahasa Indonesia dengan proses pelatihan yang dapat diparalelkan, sehingga mempercepat proses pelatihan; dan 3) dapat melakukan klasifikasi multilabel, yaitu sebuah kata dapat memiliki beberapa label, sehingga dapat memprediksi emosi majemuk. Kemampuan prediksi emosi majemuk dengan tingkat akurasi yang tinggi dari hasil penelitian ini dapat dimanfaatkan untuk berbagai aplikasi, antara lain aplikasi analisis ekspresi wajah, dan analisis situasi sosial.

Keseluruhan bagian yang ada dalam makalah ini telah disusun secara terstruktur. Tiap bagian memiliki fungsi masing-masing untuk mendeskripsikan hasil penelitian. Bagian pendahuluan berfungsi untuk mendeskripsikan latar belakang masalah, penelitian sebelumnya yang terkait, dan tujuan penelitian. Bagian kajian emosi majemuk mendeskripsikan kajian tentang emosi dasar dan emosi majemuk bahasa Indonesia. Bagian metodologi mendeskripsikan tahapan penelitian yang dilakukan. Bagian hasil dan pembahasan mendeskripsikan hasil eksperimen dan analisisnya, sedangkan bagian kesimpulan menyajikan rangkuman hasil eksperimen dan kesimpulan penelitian.



Gambar 1. Ringkasan proses klasifikasi emosi majemuk.

II. KAJIAN EMOSI MAJEMUK

Emosi dasar manusia dibagi menjadi enam, yaitu senang, sedih, marah, jijik, takut, dan terkejut [3]. Senang merupakan perasaan ketika manusia memiliki rasa puas akan sesuatu, sedangkan sedih merupakan perasaan ketika manusia tidak mendapatkan hal yang diinginkan dan penyesalan. Marah merupakan perasaan seseorang seperti frustrasi dengan sesuatu. Jijik merupakan perasaan penolakan terhadap sesuatu. Sementara itu, takut merupakan perasaan akan kekhawatiran terhadap sesuatu dan terkejut merupakan suatu perasaan ketika seseorang mengalami sesuatu secara tidak terduga.

Dari kumpulan emosi dasar tersebut, akan tercipta sebuah emosi majemuk, yaitu sebuah emosi baru dari gabungan emosi-emosi dasar [14]. Emosi majemuk ini dapat diidentifikasi dengan melihat titik-titik otot yang berada di wajah. Emosi majemuk juga dapat dihasilkan dari kombinasi kelas-kelas emosi dasar yang terbentuk dari proses klasifikasi teks berbahasa Indonesia. Salah satu jenis emosi majemuk adalah terharu, yang merupakan gabungan dari emosi bahagia dan sedih.

Emosi majemuk terdiri atas beberapa kategori, yaitu 1) terkejut senang, 2) senang jijik, 3) senang sedih, 4) sedih takut, 5) sedih marah, 6) terkejut sedih, 7) takut marah, 8) terkejut takut, 9) takut jijik, 10) marah kaget, 11) marah jijik, 12) terkejut jijik, 13) kaget, 14) kebencian, dan 15) kagum. Identifikasi emosi majemuk dapat diimplementasikan pada beberapa aplikasi, seperti *chat-bot*, penerjemahan bahasa, dan *text summarization*, sedangkan manfaat identifikasi emosi majemuk adalah untuk menganalisis situasi sosial.

III. METODOLOGI

Secara garis besar, tahapan penelitian ini terdiri atas *data scraping*; pemrosesan teks data; pembagian data latih, data validasi, dan data uji; pembuatan model klasifikasi; dan evaluasi model. Tahapan penelitian secara keseluruhan dapat dilihat pada Gambar 1. *Data scraping* merupakan kegiatan pengumpulan data. Tahap selanjutnya adalah pemrosesan kata agar dapat digunakan oleh model. Data kemudian dibagi menjadi tiga bagian, yaitu data latih, data validasi yang akan digunakan dalam proses pelatihan, dan data uji yang digunakan untuk proses evaluasi. Dalam proses evaluasi ini, data uji digunakan sebagai data yang akan diuji terhadap data latih.

A. DATA SCRAPING

Penelitian ini menggunakan *dataset* kalimat berbahasa Indonesia, yang terdiri atas teks yang memiliki emosi dan tidak



Gambar 2. Proses tokenisasi dengan model XLM-R.

memiliki emosi. *Dataset* diambil dari *data scraping* dengan menggunakan *application programming interface* (API) Twitter dan bantuan *library* Tweepy di Python [15]–[17]. Twitter merupakan sebuah platform media sosial yang memberikan layanan kepada para penggunanya untuk dapat menyampaikan sebuah ‘kicauan’ atau *tweet* yang merupakan sebuah teks pendek dengan panjang maksimal 140 kata.

Twitter memberikan akses API secara *open-source* untuk keperluan riset ataupun eksperimen dengan cara memberikan akses mengenai data-data seperti *tweet*, *username*, dan status, tetapi tetap menjaga data-data privasi pengguna. Data ini biasanya juga digunakan oleh para peneliti untuk klasifikasi sentimen [18]. Dalam penelitian ini, *dataset* yang diperoleh dari Twitter diberi label emosi oleh tim ahli bahasa berdasarkan enam emosi dasar manusia, yaitu marah, sedih, gembira, jijik, kaget, dan takut. Jumlah data yang diperoleh dari Twitter adalah 40.000 teks berbahasa Indonesia.

Dari 40.000 teks tersebut, keseimbangan jumlah label untuk tiap kelas emosi dianalisis. Hasil analisis menunjukkan bahwa *dataset* memiliki distribusi label setiap kelas yang kurang seimbang. Oleh karena itu, proses *scraping* dilakukan lagi sehingga didapatkan 25.100 teks bahasa Indonesia. Selanjutnya, seluruh *dataset* yang berjumlah 65.100 teks bahasa Indonesia diseleksi dengan memperhitungkan keseimbangan jumlah label tiap kelas emosi. Dari proses seleksi, diperoleh *dataset* berupa teks bahasa Indonesia sebanyak 29.171 teks.

Multilabel merupakan data yang memiliki label lebih dari satu, sedangkan klasifikasi teks multilabel merupakan tugas mengelompokkan teks ke dalam satu atau beberapa kelas. *Deep learning* merupakan salah satu teknik yang dapat digunakan untuk mengklasifikasi teks multilabel. Terdapat beberapa cara lain untuk melakukan proses klasifikasi multilabel, salah satunya dengan proses klasifikasi biner (*binary-classification*) pada setiap label yang ada dan menggabungkannya kembali. Pada *dataset* teks berbahasa Indonesia, sebuah teks dapat memiliki lebih dari satu kelas emosi yang berbeda. Oleh karena itu, setiap teks dapat memiliki label lebih dari satu. Hal inilah yang disebut dengan multilabel.

B. PEMROSESAN TEKS

Pemrosesan teks ini dilakukan agar model dapat lebih mudah melakukan klasifikasi [19]. Pada penelitian tersebut dijelaskan bahwa pemrosesan teks dapat meningkatkan kinerja klasifikasi, sehingga nilai akurasi lebih baik. Pemrosesan teks dalam penelitian ini meliputi *case folding* dan menghapus beberapa karakter ataupun tanda, seperti *emoji*, entitas, *hashtag*, tanda baca, dan spasi yang berlebihan.

Case folding merupakan proses mengubah karakter kapital menjadi karakter kecil. Hal ini bertujuan untuk memudahkan proses tokenisasi agar tidak diperlukan pembuatan token baru. Prapemrosesan teks yang lain adalah penapisan (*filtering*) teks, seperti menghapus *emoji* 😊; menghapus entitas dalam teks,

TABEL I
CONTOH PROSES TOKENISASI

Teks	Token Teks	Token
Aku senang sekali bisa hidup di dunia ini tanpa ada nya halangan keuangan	['Aku', 'senang', 'sekali', 'bisa', 'hidup', 'di', 'dunia', 'ini', 'tanpa', 'adanya', 'halangan', 'keuangan']	[1, 101, 2, 18, 9, 10, 38, 49, 12, 34, 8, 103, 5]

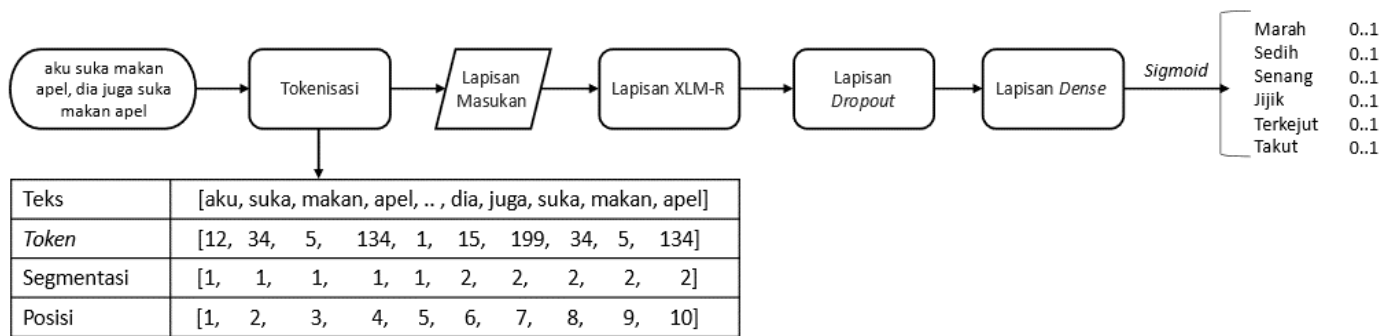
yakni tautan (*link*), karakter non-ASCII, kalimat tanda baca, *mention* dalam Twitter, *hashtag*, dan *retweet* [20]; serta menghapus spasi yang berlebihan.

Prapemrosesan teks selanjutnya adalah membuang data yang memiliki jumlah kata sedikit atau memiliki jumlah karakter kurang dari 40 karakter. Dari tahap pemrosesan teks ini, jumlah data yang dapat digunakan dalam proses klasifikasi adalah 29.171 data dari 40.000 data.

C. TOKENISASI

Proses tokenisasi merupakan proses umum pada bidang *natural language processing*. Tokenisasi merupakan proses pembagian kalimat menjadi beberapa kata dan mengubah kata-kata tersebut menjadi sebuah urutan angka yang nantinya akan digunakan untuk melakukan proses lebih lanjut [21]. Contoh proses tokenisasi ditunjukkan pada Tabel I. Pada setiap teks yang ada perlu dilakukan tokenisasi agar suatu kata lebih mudah dipahami maknanya tanpa adanya pembatas-pembatas kata [21].

Proses tokenisasi dengan menggunakan model XLM-RoBERTa berbeda dengan proses tokenisasi secara umum. Proses tokenisasi dengan model XLM-RoBERTa merupakan model *transformer* dengan segmen tambahan dalam proses tokenisasi, yaitu *token embedding*, *segment embedding*, dan *position embedding*, seperti terlihat pada Gambar 2. *Token embedding* merupakan proses mengubah teks menjadi urutan angka tertentu sehingga teks dapat diproses secara matematis. Proses ini dimulai dari pembagian teks menjadi kata-kata yang dinotasikan dengan *E* (merepresentasikan *embedding*) sehingga menjadi E_{aku} , E_{suka} , dan seterusnya. Notasi $[CLS]$ dan $[SEP]$ merupakan token tambahan pada sebuah kalimat. Notasi $[CLS]$ berfungsi sebagai token awal dalam sebuah kalimat dan notasi $[SEP]$ berfungsi sebagai token pemisah atau akhir kalimat. Selanjutnya, *segment embedding* bertugas mengubah teks menjadi beberapa bagian sesuai dengan *delimiter* dalam sebuah teks. Notasi E_A menunjukkan bagian kalimat pertama, E_B menunjukkan bagian kalimat kedua, dan seterusnya. Proses ini bertujuan untuk membagi sebuah teks panjang menjadi beberapa kalimat. Dengan demikian, korelasi antar kalimat dapat diperoleh dari proses pelatihan *next sentence prediction* pada *transformer*. Tahap *position embedding* bertugas memberikan angka untuk kata yang sesuai dengan urutan kata



Gambar 3. Arsitektur model XLM-RoBERTa.

dalam sebuah teks, sehingga model dapat mempelajari konteks (makna) dalam sebuah kalimat. Setiap kata dinotasikan dengan huruf E dan diikuti posisi kata dalam sebuah teks. Penomoran posisi kata dimulai dari indeks 0.

Prapemrosesan teks selanjutnya adalah *padding* dan *truncating*. *Padding* merupakan proses membuat matriks token dan apabila memiliki panjang yang kurang dari yang ditentukan, akan diberikan data tambahan, yaitu 0, agar memiliki bentuk matriks yang sama. Sementara itu, *truncation* merupakan proses memangkas teks dengan panjang berlebihan agar memiliki panjang yang sesuai dengan yang ditentukan.

Tabel I memperlihatkan proses tokenisasi yang memisah kata dengan spasi dan menjadikan sebuah daftar kata yang nantinya diubah menjadi angka sesuai dengan *word corpus* yang ada. Sebagai contoh, pada Tabel I kata yang ada dipisah dan setiap kata diubah menjadi angka, seperti “aku” menjadi 1, “senang” menjadi 101, dan seterusnya. Pemberian angka dari tiap kata harus selalu konsisten (selalu sama). Ini berarti bahwa apabila terdapat beberapa teks yang memiliki kata “aku”, pemberian angka untuk kata “aku” selalu angka 1. Demikian juga dengan kata-kata yang lain.

D. DATASET

Pembagian *dataset* merupakan proses membagi *dataset* untuk proses pelatihan, data untuk validasi, dan data pengujian. Pembagian data ini bertujuan untuk melihat hasil kinerja model yang digunakan. Data yang didapatkan berasal dari hasil *scraping* dari Twitter. Distribusi pembagian *dataset* terdiri atas 90% data pelatihan, 5% data validasi, dan 5% data pengujian. Oleh karena itu, jumlah data untuk masing-masing bagian data adalah 26.254 data pelatihan, 1.458 data validasi, dan 1.459 data pengujian.

E. XLM-ROBERTA

XLM-RoBERTa lebih akurat dari pendahulunya, yaitu XLM [8], dengan peningkatan akurasi sebesar 14% pada *dataset cross-lingual natural language inference* (XNLI) dan 2,4% *F1-score* pada *named entity recognition*. Model ini dilatih menggunakan seratus bahasa dengan ukuran *file* yang sangat besar [8] untuk meningkatkan kinerja klasifikasi. Model *transformer* XLM-RoBERTa dapat mengetahui suatu makna kata dengan *self attention mechanism*, yang merupakan sebuah metode *embedding* untuk kata yang menggunakan posisi kata dan keterkaitan kata terhadap kata-kata di sekitarnya untuk menghasilkan bobot yang berbeda-beda, tergantung pada penggunaan kata tersebut. Rumus dari *self attention mechanism* ditunjukkan pada (1).

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

dengan variabel Q merupakan vektor *embedding* pada sebuah kata dalam kalimat, K merupakan vektor *embedding* pada kata-kata yang berada dalam kalimat, V merupakan sebuah vektor yang berisi perhitungan *dot product* dari *embedding* pada kata dengan parameter tertentu, dan d adalah dimensi vektor Q , K , dan V . Lalu, hasil dari $\left(\frac{QK^T}{\sqrt{d_k}}\right)$ dihitung menggunakan fungsi *softmax*, seperti ditunjukkan pada (2).

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (2)$$

dengan notasi \vec{z} adalah vektor masukan, e^{z_i} adalah eksponen dari vektor masukan, K adalah jumlah total kelas, dan e^{z_j} adalah eksponen dari vektor keluaran. Dari kedua fungsi tersebut, akan didapatkan sebuah bobot dinamis pada sebuah kata yang memiliki *embedding* berbeda-beda untuk setiap kata dalam kalimat.

Untuk dapat menggunakan model XLM-RoBERTa pada proses klasifikasi, diperlukan *fine tuning* dengan menambahkan beberapa lapisan (*layer*) pada arsitektur model. Lapisan-lapisan tersebut ialah lapisan masukan sebagai lapisan pertama untuk memasukkan data; lapisan model yang merupakan lapisan utama dalam model yang berisikan model *transformer* dari XLM-RoBERTa; lapisan *dropout* yang berfungsi untuk menonaktifkan beberapa neuron di dalam lapisan tersembunyi; dan lapisan *dense* yang berisi neuron yang saling terhubung.

Pada penelitian ini, arsitektur model XLM-RoBERTa disajikan pada Gambar 3. Pada sebuah teks dilakukan proses tokenisasi agar menjadi masukan bagi model utama melalui lapisan masukan. Proses klasifikasi dilakukan pada lapisan XLM-RoBERTa. Hasilnya dilewatkan pada lapisan *dropout* untuk mematikan beberapa neuron agar hasil pelatihan tidak menjadi *overfit*, yaitu model memiliki kinerja baik pada data latih tetapi tidak pada data validasi. Selanjutnya, data menuju lapisan *dense* yang merupakan neuron untuk mengeluarkan hasil keluaran dari lapisan-lapisan sebelumnya dengan melakukan perhitungan lagi menggunakan fungsi aktivasi *sigmoid* agar keluaran menjadi angka dengan rentang dari 0 hingga 1. Dari fungsi tersebut, terdapat enam hasil angka yang masing-masing mewakili emosi dalam sebuah teks, yakni marah, sedih, senang, terkejut, jijik, dan takut. Hasil prediksi adalah nilai probabilitas kelas-kelas emosi pada masing-masing label. Gabungan beberapa kelas emosi digunakan sebagai dasar pembentukan emosi majemuk.

F. FINE TUNING

Fine tuning merupakan proses mengubah arsitektur model dengan mengubah lapisan masukan dan keluaran sebuah model, sehingga dapat digunakan untuk menyelesaikan berbagai

macam masalah seperti klasifikasi, *chatbot*, dan *named entity recognition* (NER). *Fine tuning* meliputi penambahan beberapa lapisan pada model utama dan melakukan *hypertuning* pada beberapa lapisan. Oleh karena itu, penelitian yang diusulkan ini menggunakan model XLM-Roberta dengan *fine-tuning*. Tidak banyak perubahan yang dilakukan terhadap model karena proses *fine-tuning* hanya menambah lapisan masukan serta keluaran yang disesuaikan dengan kebutuhan.

Penambahan lapisan ini berfungsi sebagai pintu masuk serta keluar dari sebuah model. Model menggunakan *layer mask* (LM) dan *next sentence prediction* untuk menghasilkan keluaran, yaitu sebuah teks yang memiliki sebuah *token [MASK]* serta teks dengan label kalimat selanjutnya untuk dilakukan proses pelatihan pada model awal. Model juga menggunakan lapisan masukan yang berbeda untuk proses *question answering* model dengan memasukkan teks yang berupa *start/end span*, yaitu teks dengan pembatasan kalimat yang menentukan bagian pertanyaan serta jawabannya dan lapisan keluaran yang menghasilkan keluaran teks dengan *question* sebagai fitur dan *answer* sebagai label dari *question*.

Lapisan masukan berfungsi sebagai lapisan yang menerima data berupa angka untuk dilakukan proses perhitungan pada model di bawahnya. Pada data yang diterima harus dilakukan prapemrosesan terlebih dahulu agar data tersebut dapat digunakan untuk lapisan lainnya. Lapisan ini sangat penting karena tanpa adanya penerima data, lapisan di bawahnya tidak dapat digunakan. Lapisan XLM-RoBERTa merupakan lapisan dari model utama, yaitu XLM-RoBERTa. Lapisan ini berisi lapisan-lapisan lain yang terdapat pada model utama. Lapisan ini memiliki sebanyak 559.896.432 lapisan lainnya.

Lapisan *dropout* merupakan lapisan yang berfungsi untuk mencegah sebuah model mempelajari hal yang sama secara berulang-ulang sehingga model menjadi *overfitting* [22]–[24]. *Dropout* dilakukan dengan cara menonaktifkan beberapa neuron yang ada dalam sebuah lapisan tersembunyi. Parameter pada lapisan ini digunakan untuk menentukan persentase banyaknya neuron yang akan dinonaktifkan. Lapisan *dense* merupakan lapisan yang digunakan sebagai lapisan keluaran dalam arsitektur ini. Lapisan ini berguna untuk memberikan hasil dari kalkulasi lapisan-lapisan sebelumnya. Hasil keluaran merupakan sebuah larik berisi enam angka yang menandakan kategori teks. Setiap angka memiliki jarak dari 0 hingga 1. Makin tinggi nilai angka, makin yakin sebuah model dalam memberikan kategori tersebut.

G. FUNGSI AKTIVASI

Fungsi aktivasi merupakan fungsi yang digunakan untuk menghitung total masukan dan bobot sebuah neuron dan menentukan sebuah *neuron* perlu dinonaktifkan atau diaktifkan [25], [26]. Salah satu fungsi aktivasi yang digunakan dalam penelitian adalah *sigmoid*. Fungsi aktivasi ini merupakan sebuah fungsi aktivasi nonlinear yang dapat mengubah masukan menjadi angka dalam rentang 0 hingga 1. Fungsi aktivasi *sigmoid* dirumuskan dengan menggunakan (3).

$$S(x) = \frac{1}{1+e^{-x}} \quad (3)$$

dengan notasi $S(x)$ merupakan fungsi *sigmoid*, x merupakan masukan data, dan e merupakan eksponen. Fungsi *sigmoid* digunakan untuk membuat hasil keluaran dari sebuah model dengan berbagai macam hasilnya selalu menjadi angka dalam rentang 0 hingga 1. Fungsi aktivasi ini sangat cocok digunakan dalam kasus klasifikasi biner.

TABEL II
NILAI AKURASI TIAP ITERASI

Iterasi	Akurasi Pelatihan	Akurasi Validasi	Nilai Loss Pelatihan	Nilai Loss Validasi
1	91,06%	95,48%	30,02%	16,27%
2	94,96%	95,48%	18,68%	15,40%
3	94,97%	95,50%	17,42%	14,25%
4	94,95%	95,50%	16,12%	13,46%
5	95,02%	95,56%	15,04%	13,34%

Dengan fungsi aktivasi *sigmoid*, keluaran model berisi enam angka dengan rentang dari 0 hingga 1, yang menentukan kelas-kelas emosi dari sebuah teks. Hasil proses klasifikasi dengan probabilitas tinggi untuk memiliki lebih dari satu label menunjukkan bahwa teks tersebut merupakan teks dengan emosi majemuk. Sebaliknya, jika teks tidak menghasilkan nilai probabilitas untuk label apapun, maka teks tersebut merupakan teks yang netral.

IV. HASIL DAN PEMBAHASAN

Di bagian ini dipaparkan hasil pelatihan dengan pembahasan terhadap kinerja model yang telah dilatih. Dalam eksperimen ini, model XLM-RoBERTa menghasilkan akurasi yang lebih baik, yaitu menjadi 95,56%. Model ini juga memiliki kinerja tinggi untuk data yang belum terlihat. Hal ini terjadi karena model *transformer* teks diprediksi berdasarkan dengan semantik (makna) dari sebuah kata, bukan dari frekuensi kemunculan sebuah kata.

Hasil eksperimen menunjukkan bahwa model yang diusulkan dapat memprediksi kelas-kelas emosi dari masing-masing teks berbahasa Indonesia. Hasil prediksi menunjukkan bahwa setiap teks berbahasa Indonesia dapat memiliki satu atau lebih kelas-kelas emosi dasar yang nantinya digabungkan untuk membentuk emosi majemuk. Pada bagian ini, dijelaskan lebih detail mengenai eksperimen pada proses pelatihan dan prediksi.

A. PELATIHAN

Proses pelatihan model hanya dilakukan sebanyak lima iterasi dengan menggunakan *loss function binary cross entropy* (BCE) [27] dengan menggunakan (4).

$$BCE = \frac{1}{n} \sum_{i=1}^n y_i \cdot \log \hat{y}_i + (1 - y_i) \cdot \log(1 - \hat{y}_i) \quad (4)$$

dengan variabel \hat{y}_i adalah label diprediksi, y_i adalah label aktual, dan n adalah jumlah label. *Loss function* berfungsi untuk menghitung jarak perbedaan antara label diprediksi terhadap label aktual. *Loss function* ini digunakan dalam proses klasifikasi biner, yaitu sebuah data memiliki label 1 dan 0.

Dalam proses pelatihan, model diberi data validasi untuk melihat kinerjanya dalam mengenali data yang belum pernah terlihat sebelumnya pada proses pelatihan. Data validasi juga memiliki manfaat untuk membantu mengatur *hyperparameter* dengan melakukan prediksi terhadap data pengujian dan memberikan evaluasinya untuk dijadikan acuan pengaturan *hyperparameter*, seperti penambahan lapisan tersembunyi pada penggunaan fungsi aktivasi lain atau perubahan arsitektur model. Data validasi dapat diibaratkan sebagai data pengujian untuk proses pelatihan.

Hasil eksperimen yang disajikan di Tabel II menunjukkan bahwa model yang diusulkan memiliki akurasi yang sangat tinggi. Data validasi serta data latih memiliki akurasi lebih dari 90% pada iterasi pertama dan terus meningkat pada setiap iterasi. Tabel II juga menyajikan peningkatan akurasi yang stabil dari setiap iterasi untuk data pelatihan dan data validasi.

TABEL III
HASIL KLASIFIKASI KELAS-KELAS EMOSI MAJEMUK

No.	Teks berbahasa Indonesia	Marah	Sedih	Senang	Jijik	Takut	Kaget	Label Sebenarnya	Hasil Prediksi	Emosi Majemuk
1	waktu smk pernah tuh jumpa sama cewe namanya dea lokasi tu di kolam renang yang <i>family friendly</i> lah harga masuknya awal nya saling lempar senyum aja agak salting sih pas si dea terus ngelitin terus pas akunya nunjukin gaya renang pake ngomongin lagi sama temannya	0%	100%	84%	1%	0%	0%	Senang, Sedih	Senang, Sedih	Terharu
2	ini pak lieus lagi cari perhatian pak jokowi kita ketawain aja maaf pak bukan saya yang edit	0%	0%	98%	0%	11%	0%	Senang, Takut	Senang, Takut	Greget
3	saat konstantinopel takhluk ayasofya tetap dijaga bahkan saat perang pun rumah ibadah tak boleh diganggu perusakan rumah ibadah di kab sintang kalimantan barat sangat disayangkan saya turut mengecam dan berduka cita semoga pelakunya dapat diproses hukum dengan adil	95%	0%	98%	0%	0%	0%	Marah, Sedih	Marah, Sedih	Kecewa
4	dia nih yang bikin <i>umbrella academy</i> tebal coba aja dia mau nurutin apa kata bapaknya buat gausah aneh aneh pasti <i>umbrella academy</i> tipis banget wkwk	72%	0%	81%	0%	0%	2%	Marah, Senang	Marah, Terkejut	Geram
5	firasat alam jatuh dari jabatan dalam waktu dekat terperosok dalam lumpur yang menjijikan insya allah wallahu alam bishowab	2%	100%	0%	11%	2%	0%	Sedih, Jijik	Sedih, Jijik	Tertekan

Sementara itu, nilai *loss* untuk data pelatihan dan data validasi mengalami penurunan cukup stabil dari iterasi ke-1 sampai dengan iterasi ke-5. Makin kecil nilai *loss*, makin baik kinerja model.

Analisis terhadap hasil eksperimen menunjukkan bahwa model memiliki nilai *loss* yang cukup kecil pada iterasi pertama, yaitu sebesar 30,02% untuk data pelatihan dan 16,27% untuk data validasi. Hal ini membuktikan bahwa model yang diusulkan dapat melakukan proses klasifikasi tanpa perlunya proses pelatihan yang banyak. Nilai *loss* untuk data pelatihan dan data validasi dari setiap eksperimen tersaji di Tabel II.

Deskripsi lebih detail hasil eksperimen untuk lima kali perulangan adalah sebagai berikut. Pada iterasi pertama, akurasi pelatihan mencapai 91,06% dan akurasi validasi mencapai 95,48%. Pada iterasi kedua, akurasi bertambah sebesar 3,90% sehingga menjadi 94,96%, sedangkan akurasi validasi tetap. Selama proses pelatihan berlangsung, akurasi selalu bertambah sedikit demi sedikit dan mencapai puncaknya pada iterasi terakhir, yaitu iterasi ke-5, dengan akurasi pelatihan mencapai 95,02% dan akurasi validasi sebesar 95,56%.

Hasil eksperimen yang mendeskripsikan perkembangan nilai *loss* pelatihan dan nilai *loss* validasi adalah sebagai berikut. Nilai ini mengalami penurunan dari iterasi ke-1 sampai dengan iterasi ke-5. Pada iterasi pertama, nilai *loss* pelatihan mencapai nilai paling tinggi, yaitu sebesar 30,02% dan nilai *loss* validasi sebesar 16,27%. Pada setiap iterasi, nilai *loss* pelatihan dan nilai *loss* validasi makin menurun hingga mencapai puncaknya pada iterasi ke-5, yaitu sebesar 15,04% untuk nilai *loss* pelatihan dan 13,34% untuk nilai *loss* validasi.

Hasil eksperimen yang mendeskripsikan kelas-kelas emosi dasar dari hasil proses klasifikasi terhadap kalimat-kalimat berbahasa Indonesia disajikan di Tabel III. Setiap kalimat menghasilkan kelas-kelas emosi dasar yang dominan sebagai dasar untuk membentuk kelas emosi majemuk. Sebagai contoh, hasil klasifikasi pada teks nomor 1 menghasilkan nilai probabilitas kelas emosi dasar 'sedih' sebesar 100% dan kelas emosi dasar 'senang' sebesar 84%. Nilai probabilitas kedua kelas emosi dasar tersebut merupakan nilai yang dominan dibandingkan dengan kelas-kelas emosi dasar lainnya. Kedua kelas emosi dasar yang dominan ini digunakan sebagai dasar untuk membentuk kelas emosi majemuk, yaitu terharu. Demikian juga dengan hasil klasifikasi terhadap teks berbahasa Indonesia lainnya, seperti ditunjukkan pada Tabel III.

Berdasarkan hasil klasifikasi di Tabel III, ekspresi wajah majemuk dibentuk berdasarkan nilai probabilitas dari dua kelas emosi dasar yang memiliki nilai probabilitas paling tinggi. Ekspresi wajah majemuk diimplementasikan pada animasi wajah 3D dengan menggunakan sistem *facial action coding system* (FACS). FACS merupakan sistem pengodean aktivitas otot wajah yang memberikan informasi ekspresi wajah. FACS dari setiap kelas emosi dasar terdiri atas beberapa *action unit* (AU) yang menunjukkan gerakan otot wajah. Kombinasi gerakan dari gabungan otot-otot wajah yang saling terkait dapat membentuk ekspresi emosi dasar tertentu.

Dalam mengimplementasikan ekspresi wajah 3D, nilai setiap AU ditentukan melalui suatu nilai dengan rentang antara 0 sampai dengan 1. Dalam penelitian ini, setiap AU dari tiap kelas emosi dasar diimplementasikan dengan menggunakan

nilai probabilitas hasil proses klasifikasi yang berkisar antara 0 sampai dengan 1, seperti yang ditunjukkan dalam Tabel III.

Ekspresi emosi majemuk wajah 3D diimplementasikan dengan mengombinasikan AU dari dua kelas emosi dasar. Kombinasi nilai AU dari masing-masing kelas emosi dasar tersebut akan membentuk gerakan otot wajah yang dapat menampilkan ekspresi wajah majemuk tertentu. Nilai probabilitas dari tiap kelas emosi dasar digunakan sebagai nilai tiap AU. Oleh karena itu, apabila nilai probabilitas tiap kelas emosi tidak seimbang (cenderung lebih besar pada kelas emosi tertentu), ekspresi wajah majemuk yang terbentuk juga akan mengikuti kecenderungan untuk kelas emosi tertentu. Misalnya, pada kalimat nomor 2 di Tabel III, nilai probabilitas senang adalah 98%, sedangkan nilai probabilitas takut sebesar 11%, sehingga ekspresi wajah majemuk yang ditampilkan akan cenderung pada ekspresi emosi senang.

Dari hasil eksperimen juga dapat diketahui bahwa kelas-kelas emosi dasar dominan yang dihasilkan dari proses klasifikasi sama dengan kelas-kelas emosi dasar dominan yang sebenarnya. Ini menunjukkan bahwa model yang diusulkan dalam penelitian ini telah mampu menghasilkan kelas emosi majemuk yang merupakan gabungan dari kelas-kelas emosi dasar yang dominan dengan tingkat akurasi yang sangat baik.

B. PREDIKSI

Hasil prediksi untuk keseluruhan *dataset* sebanyak 29.172 mencapai akurasi yang sangat baik dengan iterasi pelatihan sebanyak lima kali. Model XLM-RoBERTa memiliki kinerja yang sangat baik terhadap data yang belum pernah dilihat sebelumnya karena model telah memiliki pengetahuan umum untuk sebuah bahasa. Kinerja model dapat meningkat jauh lebih baik lagi apabila data pelatihan yang digunakan dalam proses klasifikasi jumlahnya lebih banyak dan kualitasnya lebih baik.

Eksperimen selanjutnya digunakan untuk mengevaluasi data pelatihan dengan melakukan distribusi data pelatihan, data validasi, dan data pengujian dengan komposisi yang berbeda-beda. Hal ini bertujuan untuk mengetahui komposisi data yang dapat menghasilkan kinerja paling optimal. Komposisi data pertama yaitu data pelatihan sebesar 70%, data validasi sebesar 15%, dan data pengujian sebesar 15%. Dengan komposisi data ini, hasil akurasi yang diperoleh adalah 94,92% dan tingkat *loss* sebesar 14,59%. Dari beberapa komposisi data tersebut, diperoleh hasil bahwa makin besar data pelatihan, makin tinggi akurasi. Tingkat akurasi yang paling optimal diperoleh pada komposisi data pelatihan sebesar 90%, data validasi sebesar 5%, dan data pengujian sebesar 5%, yang menghasilkan akurasi sebesar 95,56% dan tingkat *loss* sebesar 12,90%. Hasil yang didapatkan telah melampaui hasil dari penelitian sebelumnya yang menggunakan *naive Bayes* dan TF-IDF, yaitu sebesar 75,47% [13].

Model *naive Bayes* dan TF-IDF masih menggunakan proses *embedding* TF-IDF yang menghitung kemunculan kata dalam sebuah kalimat atau teks. Hal ini sangat berbeda dengan penggunaan model XLM-RoBERTa yang diusulkan dalam penelitian ini. Setiap kata direpresentasikan berdasarkan posisi dan korelasi antara kata-kata dalam sebuah kalimat atau teks. Dengan demikian, sebuah kata memiliki representasi yang berbeda-beda, sesuai dengan konteks kalimat tersebut, sehingga sebuah kalimat menjadi lebih mudah diklasifikasi.

Model XLM-RoBERTa memiliki kemampuan untuk memprediksi sebagian besar teks dengan data yang kurang baik serta tidak dilakukan pemrosesan yang begitu detail. Hal inilah

yang membuat model XLM-RoBERTa lebih unggul dibandingkan dengan model tradisional lainnya.

V. KESIMPULAN

Berdasarkan hasil beberapa eksperimen yang telah dilakukan, dapat disimpulkan bahwa model XLM-RoBERTa dapat melakukan klasifikasi multilabel dengan hasil yang optimal hanya dengan proses pelatihan sebanyak lima iterasi. Tingkat akurasi optimal yang dapat dicapai adalah sebesar 95,56% dan nilai *loss* sebesar 12,90%. Model juga berhasil memprediksi emosi dominan dalam sebuah teks yang memiliki kelas emosi lebih dari satu kelas.

Untuk penelitian selanjutnya, kinerja klasifikasi dapat ditingkatkan lagi dengan menambahkan jumlah *dataset* dan distribusi *dataset* yang seimbang untuk tiap kelas. Peningkatan kinerja model *transformer* juga dapat dilakukan dengan mengganti atau memperbaiki arsitektur lapisan *transformer*-nya.

KONFLIK KEPENTINGAN

Dengan ini penulis menyatakan bahwa data-data maupun hasil analisis yang dipublikasikan pada makalah ini tidak memiliki konflik kepentingan dengan pihak mana pun. Apabila di kemudian hari ditemukan adanya hal-hal yang tidak sesuai, maka tanggung jawab sepenuhnya berada di pihak penulis.

KONTRIBUSI PENULIS

Konseptualisasi, Aripin; metodologi, Aripin; perangkat lunak, Steven Adi Santoso; validasi, Aripin; analisis formal, Aripin, Hanny Haryanto; narasumber, Aripin, Steven Adi Santoso; investigasi, Aripin, Hanny Haryanto; kurasi data, Aripin, Steven Adi Santoso; penulisan—penyusunan draf asli, Aripin, Steven Adi Santoso; penulisan—peninjauan dan penyuntingan, Aripin, Hanny Haryanto; visualisasi, Aripin, Steven Adi Santoso; administrasi penelitian, Aripin; akuisisi pendanaan, Aripin.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada Direktorat Riset dan Pengabdian Masyarakat Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi Republik Indonesia yang telah mendukung pendanaan penelitian ini melalui Program Hibah Kompetitif dalam Skema Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) berdasarkan Kontrak Nomor 10/061031/PB/SP2H/AK.04/2022.

Ucapan terima kasih juga disampaikan kepada seluruh pimpinan Universitas Dian Nuswantoro yang telah memberikan fasilitas penggunaan Laboratorium Sistem Cerdas sehingga kegiatan penelitian ini dapat diselesaikan dengan baik.

REFERENSI

- [1] I.F. Putra dan A. Purwarianti, "Improving Indonesian Text Classification Using Multilingual Language Model," *2020 7th Int. Conf. Adv. Inform.: Concepts, Theory, Appl. (ICAICTA)* 2020, hal. 1-5, doi: 10.1109/ICAICTA49861.2020.9429038.
- [2] S. Du, Y. Tao, dan A.M. Martinez, "Compound Facial Expressions of Emotion," *PNAS*, Vol. 111, No. 15, hal. E1454-E1462, Mar. 2014, doi: 10.1073/pnas.1322355111.
- [3] P. Ekman, "An Argument for Basic Emotions," *Cogn., Emot.*, Vol. 6, No. 3-4, hal. 169-200, Jan. 2008, doi: 10.1080/02699939208411068.
- [4] V. Dogra dkk., "A Complete Process of Text Classification System Using State-of-the-Art NLP Models," *Comput. Intell., Neurosci.*, Vol. 2022, hal. 1-26, Jun. 2022, doi: 10.1155/2022/1883698.
- [5] T.H. Saputro dan A. Hermawan, "The Accuracy Improvement of Text Mining Classification on Hospital Review Through the Alteration in the Preprocessing Stage," *Int. J. Comput., Inf. Technol. (IJCIT)*, Vol. 10, No. 4, hal. 140-146, Jul. 2021, doi: 10.24203/ijcit.v10i4.138.

- [6] W.-H. Khong, L.-K. Soon, dan H.-N. Goh, "A Comparative Study of Statistical and Natural Language Processing Techniques for Sentiment Analysis," *J. Teknol.*, Vol. 77, No. 18, hal. 155–161, Nov. 2015, doi: 10.11113/jt.v77.6502.
- [7] Aripin, H. Haryanto, dan W. Agastya, "Synthesis of Compound Facial Expressions Based on Indonesian Sentences Using Multinomial Naïve Bayes Model and Dominance Threshold Equations," *Eng. Lett.*, Vol. 30, No. 1, hal. 1-10, Mar. 2022.
- [8] A. Conneau dkk., "Unsupervised Cross-lingual Representation Learning at Scale," *Proc. 58th Annu. Meeting Assoc. Comput. Linguist.*, 2020, hal. 8440-8451, doi: 10.18653/v1/2020.acl-main.747.
- [9] H. Gonen, S. Ravfogel, Y. Elazar, dan Y. Goldberg, "It's not Greek to mBERT: Inducing Word-Level Translations from Multilingual BERT," *Proc. Third BlackboxNLP Workshop Anal., Interpreting Neural Netw. NLP*, 2020, hal. 45-56, doi: 10.18653/v1/2020.blackboxnlp-1.5.
- [10] J. Devlin, M.-W. Chang, K. Lee, dan K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proc. 2019 Conf. North Amer. Chapter Assoc. Comput. Linguist.: Human Lang. Technol.*, 2019, hal. 4171-4186, doi: 10.18653/v1/n19-1423.
- [11] K. Taneja dan J. Vashishtha, "Comparison of Transfer Learning and Traditional Machine Learning Approach for Text Classification," *2022 9th Int. Conf. Comput. Sustain. Glob. Development (INDIACom)*, 2022, hal. 195-200, doi: 10.23919/INDIACom54597.2022.9763279.
- [12] Aripin, W. Agastya, dan H. Haryanto, "Ekstraksi Emosi Majemuk Kalimat Bahasa Indonesia Menggunakan Convolution Neural Network," *J. Nas. Tek. Elekt., Teknol. Inf. (JNTETI)*, Vol. 10, No. 2, hal. 148-155, Mei 2021, doi:10.22146/jnteti.v10i2.1051.
- [13] W. Agastya dan Aripin, "Pemetaan Emosi Dominan pada Kalimat Majemuk Bahasa Indonesia Menggunakan Multinomial Naïve Bayes," *J. Nas. Tek. Elekt., Teknol. Inf. (JNTETI)*, Vol. 9, No. 2, hal. 171-179, Mei 2020, doi: 10.22146/jnteti.v9i2.157.
- [14] S. Du dan A.M. Martinez, "Compound Facial Expressions of Emotion: From Basic Research to Clinical Applications," *Dialogues Clin. Neurosci.*, Vol. 17, No. 4, hal. 443-455, 2015, doi: 10.31887/DCNS.2015.17.4/sdu.
- [15] A. Wibowo dan E. Winarko, "Paper Review: Data Mining Twitter," *Konf. Nas. Sist., Inform. (KNS&I)*, 2014, hal. 1-6.
- [16] N. Azam, Jahiruddin, M. Abulaish, dan N.A.H. Haldar, "Twitter Data Mining for Events Classification and Analysis," *2015 Second Int. Conf. Soft Comput., Mach. Intell. (ISCMI)*, 2015, hal. 79–83 doi: 10.1109/ISCMI.2015.33.
- [17] R. Batool, A. Khattak, J. Hashmi, dan S. Lee, "Precise Tweet Classification and Sentiment Analysis," *2013 IEEE/ACIS 12th Int. Conf. Comput., Inf. Sci. (ICIS)*, 2013, hal. 461–466. doi: 10.1109/ICIS.2013.6607883.
- [18] S. Vosoughi, H. Zhou, dan D. Roy, "Enhanced Twitter Sentiment Classification Using Contextual Information," *Proc. 6th Workshop Comput. Approaches Subjectivity Sentiment, Social Media Anal.*, 2016, hal. 16-24, doi: 10.18653/v1/W15-2904.
- [19] M.D. Samad, N.D. Khounviengxay, dan M.A. Witherow, "Effect of Text Processing Steps on Twitter Sentiment Classification using Word Embedding," 2020, *arXiv:2007.13027*.
- [20] A.I. Kadhim, "An Evaluation of Preprocessing Techniques for Text Classification," *Int. J. Comput. Sci., Inf. Secur.*, Vol. 16, No. 6, hal. 22-32, Jun. 2018.
- [21] J.J. Webster dan C. Kit, "Tokenization as the Initial Phase in NLP," *Proc. 14th Conf. Comput. Linguist.*, 1992, Vol. 4, hal. 1106–1110, doi:10.3115/992424.992434.
- [22] N. Srivastava dkk., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res.*, Vol. 15, No. 56, hal. 1929-1958, Jun. 2014.
- [23] A. Labach, H. Salehinejad, dan S. Valaee, "Survey of Dropout Methods for Deep Neural Networks," 2019, *arXiv:1904.13310*, doi: 10.48550/arXiv.1904.13310.
- [24] S. Cai dkk., "Effective and Efficient Dropout for Deep Convolutional Neural Networks," 2020, *arXiv:1904.03392*, doi: 10.48550/arXiv.1904.03392.
- [25] S.R. Dubey, S.K. Singh, dan B.B. Chaudhuri, "Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark," 2022, *arXiv:2109.14545*, doi: 10.48550/arXiv.2109.14545.
- [26] C. Nwankpa, W. Ijomah, A. Gachagan, dan S. Marshall, "Activation Functions: Comparison of Trends in Practice and Research for Deep Learning," 2018, *arXiv:1811.03378*, doi: 10.48550/arXiv.1811.03378.
- [27] A.U. Ruby, P. Theerthagiri, I.J. Jacob, dan Y. Vamsidhar, "Binary Cross Entropy with Deep Learning Technique for Image Classification," *Int. J. Adv. Trends Comput. Sci., Eng.*, Vol. 9, No. 4, hal. 5393–5397, Agu. 2020, doi: 10.30534/ijatcse/2020/175942020.