

Opinion Mining Terhadap Pemberitaan Corona di Instagram menggunakan Convolutional Neural Network

Ahmad Rizki Maulana¹, Naim Rochmawati²

^{1,2} Teknik Informatika, Universitas Negeri Surabaya
ahmadmaulana16051204042@mhs.unesa.ac.id
naimrochmawati@unesa.ac.id

Abstrak—Keadaan Indonesia di tengah pandemi yang sedang terjadi saat ini menjadi berita yang sangat dibutuhkan oleh masyarakat Indonesia. Penyebaran berita saat ini sudah banyak dilakukan melalui media sosial, karena kemudahan akses dan jangkauan pasar yang sangat luas menjadikan media sosial sebagai lahan untuk menyebarkan berita. Perusahaan penyedia berita nasional telah ikut bergabung di beberapa platform media sosial. Instagram merupakan salah satu media sosial dengan pengguna terbanyak di Indonesia. Pemberitaan mengenai corona di Indonesia melalui Instagram mengizinkan pembacanya untuk dapat mengetahui reaksi pembaca lainnya terhadap berita tersebut. *Opinion mining* adalah suatu analisis yang dilakukan terhadap opini untuk diklasifikasikan kedalam beberapa kelas. *Convolutional Neural Network* merupakan algoritma yang dapat menganalisis dan mengklasifikasikan opini kedalam kelas yang telah ditentukan. Penelitian ini melakukan proses *opinion mining* menggunakan data komentar di Instagram terhadap unggahan pemberitaan terkait corona pada tiga akun berita nasional, yaitu @kompascom, @cnnindonesia, dan @detikcom. Proses klasifikasi yang dilakukan menggunakan algoritma *Convolutional Neural Network* mengasilkan akurasi sebesar 88% pada data training, 82% pada data testing, nilai presisi sebesar 96% dan recall sebesar 68%.

Kata Kunci—Opinion Mining, CNN, Komentar Instagram, Corona

I. PENDAHULUAN

Sosial media sudah menjadi bagian penting bagi sebagian besar masyarakat Indonesia. Sosial media menyediakan keterbukaan dan kemudahan akses bagi semua kalangan. *We Are Social* [1] merilis data pengguna media sosial di Indonesia mengalami peningkatan sebanyak 12 Juta dari April 2019 hingga Januari 2020. Di Indonesia, terdapat beberapa platform sosial media yang akrab dengan masyarakat Indonesia. Menurut *We Are Social*, Instagram merupakan salah satu platform sosial media yang paling populer di Indonesia dengan presentase sebesar 79%.

Pergeseran tren ini dimanfaatkan oleh perusahaan penyedia berita nasional. Perusahaan penyedia berita yang dulunya aktif memberitakan melalui media cetak, kini juga harus bersaing melalui media sosial agar dapat menjangkau pasar yang lebih luas. Peredaran unggahan berita-berita di media sosial memiliki peluang yang lebih banyak untuk sampai ke konsumen. Hal ini dikarenakan proses penerbitan berita di media sosial jauh lebih cepat daripada media cetak yang memiliki jadwal tertentu untuk penerbitannya. Salah satu keunggulan penyampaian berita melalui media sosial adalah pemberitaan terkini dapat langsung diberitakan.

Topik yang banyak dicari akhir-akhir ini salah satunya adalah penyebaran virus *corona* di Indonesia. TRAS N CO Indonesia

telah melakukan survei digital terhadap tren pencarian tentang virus *corona*, hasilnya selama Januari 2020 ada 4.090.000 pencarian dengan kata kunci “Virus Corona” [2]. Akun-akun berita nasional turut memberitakan tentang penyebaran virus *corona* di Indonesia. Tersebar virus *corona* di Indonesia telah berdampak banyak bagi masyarakat dan pemerintahan di Indonesia.

Opinion Mining atau juga dikenal dengan analisis sentimen merupakan analisis terhadap suatu dokumen ataupun teks yang berisi opini tentang suatu objek dengan tujuan untuk mengekstraksi komponen objek yang telah dikomentari dari dokumen tersebut untuk menentukan apakah termasuk pada komentar positif, negatif atau netral [3]. *Opinion mining* dapat dimanfaatkan untuk mengetahui respon pengguna media sosial terhadap suatu isu, sehingga apabila diperlukan pengambilan kebijakan bagi banyak pihak, hasil dari *opinion mining* dapat dijadikan bahan pertimbangan.

Penelitian [4] telah melakukan analisis sentimen terhadap efek dari karantina karena penyebaran virus corona. Penelitian ini menganalisa *tweet* dari pengguna *twitter* untuk dapat menganalisa sentiment masyarakat Filipina khususnya di wilayah Luzon. Penggunaan *Natural Language Processing* untuk menentukan sentiment dari *tweet* masyarakat Luzon. Hasilnya, sebagian besar masyarakat Luzon memiliki sentiment negatif terhadap Covid-19. Sentimen negatif ini terkait permasalahan yang ikut timbul karena pemberlakuan karantina masyarakat yang ekstrem. Selain itu, penelitian ini juga menganalisis dampak lain dari Covid-19 terhadap gaya hidup pribadi. Pendekatan kualitatif dilakukan untuk mengetahui dampak karantina yang diberlakukan dalam menghadapi Covid-19.

Penerapan *Opinion Mining* harus didukung dengan metode klasifikasi. Seperti pada [5] menggunakan SVM untuk mengklasifikasikan komentar Instagram yang termasuk dalam penindasan melalui dunia maya. Sebanyak 42% remaja berusia 12-20 tahun telah menjadi korban penindasan melalui dunia maya. Bahaya kejahatan siber ini juga datang dari media sosial Instagram. Analisis sentimen dirasa perlu dilakukan untuk dapat mengetahui komentar pada akun Instagram. Data yang digunakan sebanyak 400 data, dengan presentase data *training* sebesar 70% dan data uji sebesar 30%. Penelitian ini menggunakan 1799 fitur dan hasil akurasi tertinggi mencapai 90%. Hasil akurasi tertinggi didapat dengan nilai *degree polynomial* 2, nilai *learning rate* 0,0001 dan jumlah iterasi maksimum yang digunakan adalah 200 kali.

Metode *Naïve Bayes* yang digunakan pada [6] diterapkan untuk klasifikasi untuk mendeteksi komentar spam di akun Instagram. Deteksi spam ini dilakukan karena banyaknya

komentar yang ditujukan untuk mendapatkan perhatian. Dataset diambil dari komentar yang ditujukan pada akun-akun figur publik di Indonesia yang memiliki lebih dari satu juta pengikut. Penelitian ini menerapkan *preprocessing* (*tokenization*, *stop words removal*, dan *stemming*), TF-IDF, dan *naive bayes* untuk mendeteksi komentar spam dan menghasilkan rata-rata tingkat akurasi sebesar 77,25% pada *dataset* yang telah diproses dan 74,31% pada *dataset* yang belum diproses.

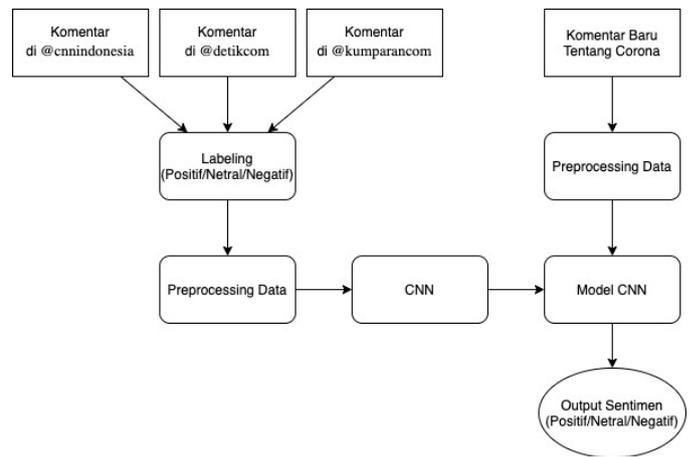
Pada [7] algoritma *Naive Bayes* digunakan untuk menganalisis opini terkait kasus *Rohingya* dari media sosial *twitter*. Konflik antara populasi Rakhine dan tentara Myanmar menimbulkan banyak pro dan kontra dalam mengatasi masalah ini. Penelitian ini menganalisis opini masyarakat dunia yang didapat dari *twitter*. *Twitter* dipilih karena menjadi salah satu media sosial yang populer dan paling sering dikunjungi, sehingga dirasa banyak data yang dapat ditemukan dari *twitter* untuk diproses dan dianalisis. Opini dikelompokkan menjadi positif, netral, dan negatif. Berdasarkan penelitian tersebut, hasilnya adalah opini negatif seimbang dengan opini positif, sedangkan kelompok netral jumlahnya lebih sedikit. Kendala pada penelitian ini banyaknya *tweet* yang tidak masuk kategori dikarenakan keterbatasan kamus dan hal lainnya.

Penerapan CNN digunakan untuk analisis sentiment pada [8]. Analisis sentiment ini memanfaatkan media *twitter* untuk memahami situasi dunia. Metode ini dapat dimanfaatkan untuk memprediksi kepuasan pengguna terhadap suatu produk, seberapa parah situasi setelah terjadi bencana dan juga kebahagiaan pada lingkungan tertentu. CNN dianggap lebih efektif karena CNN memiliki *convolutional layer* untuk mengekstrak informasi dengan teks yang lebih besar. Penelitian ini menggunakan dua data set, STS Gold dan MR. Hasil penelitian ini menunjukkan bahwa CNN memiliki akurasi yang lebih tinggi dalam klasifikasi sentiment *twitter* dibandingkan dengan SVM dan *Naive bayes*. CNN memiliki akurasi pada dataset *training* sebesar 74.5% pada dataset MR dan 68% pada dataset STS Gold.

Penelitian ini akan berfokus untuk menerapkan *Opinion Mining* terkait pemberitaan mengenai penyebaran virus *corona* di Indonesia melalui komentar pada beberapa akun *Instagram* penyedia berita nasional menggunakan metode CNN untuk mengklasifikasikan komentar kedalam kelompok komentar positif, komentar negatif dan komentar netral.

II. METODOLOGI

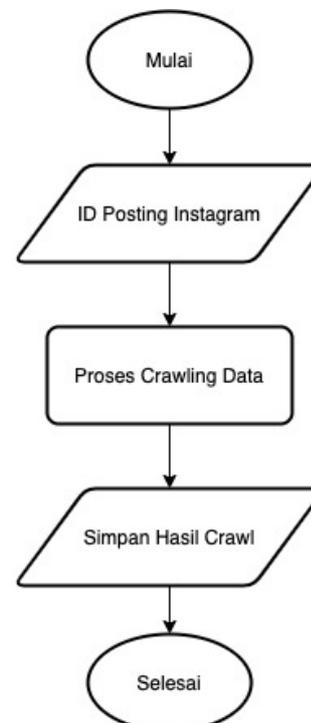
Pada penelitian ini, dataset komentar terhadap corona yang berasal dari Instagram akan dilatih menggunakan algoritma CNN [8]. Model yang dihasilkan oleh data latih kemudian akan digunakan untuk memprediksi sentiment dari komentar baru terkait corona. Rancangan penelitian dapat dilihat pada Gbr. 1



Gbr. 1 Rancangan Penelitian

A. Crawling Data

Pada penelitian ini, dataset yang digunakan untuk melatih model merupakan dataset komentar yang berasal dari komentar di Instagram terkait corona. Komentar tersebut berasal dari 3 akun berita di Indonesia yang terpercaya dan memiliki lebih dari 1 juta pengikut. Akun tersebut adalah kumparan (@kumparancom), cnn (@cnnindonesia), dan detik (@detikcom). Di Instagram terdapat dua jenis tipe komentar. Komentar terhadap unggahan dan balasan dari komentar yang telah dituliskan. Komentar yang diambil pada penelitian ini adalah komentar terhadap unggahan dan bukan merupakan balasan terhadap komentar yang telah dituliskan. Dataset komentar yang dipilih pada penelitian ini diambil dari 16 unggahan dari ketiga akun diatas dalam rentang waktu 1 bulan.



Gbr. 2 Proses Crawling

Data komentar diambil dengan melakukan proses *crawling* seperti pada Gbr 2. Pertama, peneliti mengumpulkan *list id* unggahan terkait corona dari 3 akun berita. Kemudian dari id unggahan tersebut akan dilakukan *crawling* data komentar menggunakan *Graph API Instagram*. *Graph API Instagram* mengirimkan detail id unggahan termasuk komentarnya dalam bentuk data JSON. Selanjutnya, data JSON yang berisi komentar akan disimpan kedalam sebuah *file* dengan format *tsv*. Data komentar yang akan diambil berjumlah 1000 komentar dari ketiga akun tersebut.

B. Labelling Data

Data komentar yang telah didapatkan dari proses *crawling* selanjutnya akan diberi label positif, netral, atau negatif terhadap unggahan pada akun berita tersebut yang berkaitan dengan corona. Pelabelan dilakukan secara manual untuk setiap komentar. Contoh komentar yang akan dilabeli seperti pada tabel 1.

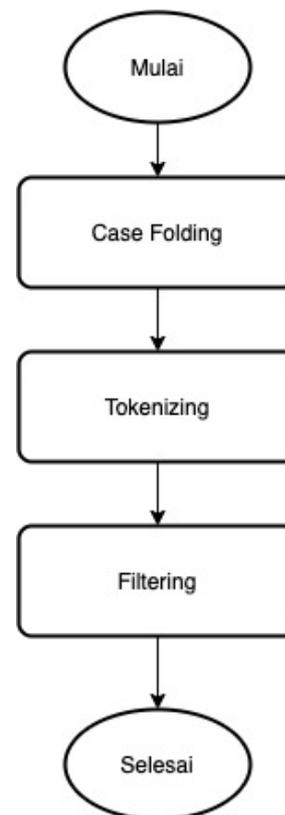
TABEL I
KALIMAT DAN PELABELAN

Komentar	Label
PSBB ga ada efek nya apa gimana sih pemerintah ini, ga becus ngurusi negara	Negatif
Alhamdulillah, akhirnya banyak yg sembuh dari pd yg meninggal, kerja bguus tim mediss	Positif
Bodo amat lahh gue didesa tenang	Netral

Setelah semua komentar telah diberikan label, data akan dipisah menjadi 2 jenis data yakni data *training* dan data *testing*. Data *training* digunakan untuk melatih model *cnn* dan data *testing* digunakan untuk melakukan *testing* terhadap model *cnn* yang dihasilkan pada tahap *training*.

C. Preprocessing

Komentar di sosial media terutama Instagram masih banyak memuat karakter, huruf, simbol yang dianggap tidak penting dalam proses klasifikasi sentimen pada komentar tersebut. Karakter, huruf, dan simbol yang tidak penting akan mengurangi performa dari model yang dibuat. Selain itu karakter, huruf, dan simbol tersebut akan membuat teks menjadi tidak terstruktur. Untuk mengatasi permasalahan ini akan dilakukan *preprocessing* pada teks komentar agar model yang dihasilkan memiliki performa yang bagus. Alur *preprocessing* dapat dilihat pada Gbr. 3



Gbr. 3 Alur Preprocessing

1) Case Folding

Case Folding adalah proses untuk mengubah semua teks menjadi bentuk standar dan memiliki kesetaraan bentuk setiap kata didalam teks yang diolah. Pada penelitian ini, jenis *case folding* yang diterapkan adalah mengubah semua teks menjadi huruf kecil. Jenis *case folding* yang diterapkan ini merupakan *case folding* yang sering dilakukan untuk menstandarisasi teks yang akan diolah. Komentar di sosial media banyak mengandung inkonsistensi penulisan huruf besar dan huruf kecil. Ketidak teraturan penulisan ini menyebabkan suatu algoritma akan terpengaruh beda makna. Contoh kata “Positif” dengan kata “positif” atau dengan kata “PosiTif” akan dianggap berbeda oleh algoritma. Algoritma akan mempelajari terlalu berlebihan pada kata yang seharusnya dianggap sama. Sehingga dengan melakukan *case folding* dengan mengubah semua teks menjadi huruf kecil akan meningkatkan performa algoritma terhadap teks.

2) Tokenizing

Tokenizing merupakan proses untuk memisah kata-kata penyusun dari sebuah kalimat. *Tokenizing* berguna untuk memudahkan algoritma dalam memahami makna pada tingkat kata. *Tokenizing* tidak akan mengubah urutan kata yang dipisah, ia akan tetap berurutan sesuai kalimat sebelum dipisahkan. Sehingga tidak akan mempengaruhi makna dari kalimat tersebut. Kata-kata yang dihasilkan dari proses *tokenizing* disebut *token*. Pada penelitian ini *token* diperlukan untuk proses *cnn*. Setiap *token* (kata) pada

kalimat akan dijadikan sebuah *word vector* yang kemudian dimasukkan kedalam *layer input* pada *cnn*.

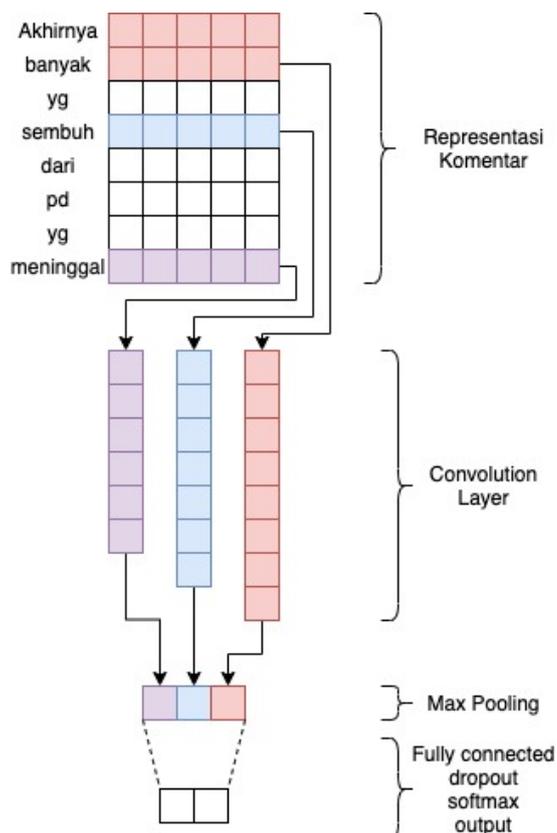
3) Filtering

Filtering merupakan proses untuk mengambil kata-kata yang dianggap penting untuk mendukung sebuah pemrosesan teks. *Filtering* akan memisah *token-token* yang dihasilkan dari proses *tokenizing* yang dianggap tidak penting. Contoh *token* yang tidak dianggap penting adalah tanda baca atau kata-kata yang tidak memiliki makna. Pada penelitian ini *filtering* dilakukan hanya untuk membuang tanda baca pada daftar *token* yang dihasilkan proses sebelumnya. *Filtering* juga akan membersihkan *emoticon*, *link url* dan kata-kata yang tidak memiliki pengaruh pada proses pengklasifikasian. Untuk memfilter kata-kata yang tidak memiliki makna akan sangat sulit karena pengguna sosial media sering sekali menggunakan bahasa yang tidak baku, biasanya menggunakan bahasa gaul. Jika dilakukan *filter* yang berlebihan akan memungkinkan perubahan makna dan struktur kalimat akan habis tanpa menyisakan *token*.

D. Convolutional Neural Network

Convolutional Neural Network (CNN) merupakan pengembangan dari *multilayer perceptron* (MLP) yang didesain untuk mengolah data dua dimensi [9]. *Convolutional Neural Network* termasuk golongan dari *Deep Neural Network* karena memiliki kedalaman jaringan yang tinggi. Dikarenakan memiliki Layer Konvolusi atau *Convolution Layer* [9]. CNN merupakan jenis *neural network* yang mengolah data dengan dimensi tinggi dan mengandung *neuron* yang saling terhubung satu dengan yang lainnya. *Convolutional Neural Network* pada umumnya terdiri dari beberapa lapisan penyusun, lapisan-lapisan ini saling terhubung untuk memproses data yang di *input*-kan, lapisan-lapisan tersebut adalah lapisan *convolution*, *pooling*, dan *fully connected*. Lapisan *convolution* dan *pooling* merupakan lapisan yang bertugas untuk mempelajari fitur pada data yang dimasukkan pada CNN, lapisan ini akan mengolah data dengan rumus matematika, sedangkan *fully connected layer* akan bertugas untuk melakukan klasifikasi terhadap data input yang diberikan.

Pada penelitian ini CNN yang akan digunakan tampak seperti pada Gambar 4. Model dibuat khusus untuk menerima input berupa teks. Pertama teks komentar akan diubah menjadi matriks. Matriks terdiri dari token-token yang dihasilkan pada proses *preprocessing*. CNN hanya dapat mengolah data *numeric*, sehingga token-token yang telah didapat akan diubah menjadi data *numeric* [10]. Input token merupakan representasi *word vector* setiap kata penyusunnya. Dimensi dari *word vector* adalah *d*. Jika panjang sebuah komentar dinyatakan dalam *s*, dimensi matriks dari komentar tersebut adalah $s \times d$. Pada penelitian ini, matriks kalimat sama dengan matriks gambar, jadi pada penelitian ini menerapkan *convolution* pada matriks dengan filter linear.



Gbr. 4 CNN berbasis teks

1) Convolution Layer

CNN memiliki lapisan yang sangat dikenal dan menjadi ciri utama dari algoritma ini, lapisan tersebut adalah lapisan konvolusi. Lapisan konvolusi akan melakukan konvolusi terhadap data masukan atau data output dari lapisan sebelumnya dengan menggeser sebuah filter agar dapat menghasilkan sebuah output yang disebut *feature map* atau *activation map*. Satu *feature map* atau *activation map* saja tidak cukup untuk mengetahui makna dari teks yang dimasukkan. Sehingga pada penelitian ini akan diberikan banyak filter pada lapisan konvolusi di CNN.

Pada proses pelatihan atau *training* pada data yang dilatih, bobot pada setiap lapisan konvolusi akan mengalami proses *update* untuk memperbaiki hasil klasifikasi. Pemilihan fungsi aktivasi yang sesuai pada lapisan konvolusi juga menjadi faktor penting keberhasilan pelatihan dengan algoritma CNN. Pada setiap lapisan konvolusi akan dilakukan regulasi L2 atau disebut juga *Ridge Regression* agar mengurangi *Overfitting* yang terjadi di setiap Epoch saat proses pelatihan data.

2) Pooling Layer

Feature map yang dihasilkan pada lapisan konvolusi masih cukup besar. Hal ini akan mempengaruhi kecepatan komputasi proses *training data*. Untuk mengurangi *feature map* yang masih cukup besar akan dilakukan pengurangan dengan lapisan *pooling*. Lapisan *pooling* adalah salah satu cara untuk meringkas data masukan atau *feature map* menjadi data dengan ukuran yang lebih kecil. Terdapat dua

jenis layer *pooling* yang bisa dilakukan, yakni *max pooling* dan *average pooling*.

Max pooling akan meringkas data masukan atau *feature map* dengan mencari nilai atau value terbesar dari *feature map* berdasarkan penggeseran *window pooling*. Sedangkan *average pooling*, akan menghitung nilai rata-rata dari setiap *feature map* pada setiap kali penggeseran *window pooling*. Pada penelitian ini lapisan *pooling* yang akan dimasukkan diantara lapisan konvolusi secara berturut-turut. Penerapan ini dapat secara progresif untuk mengurangi mengurangi dimensi *Feature Map*, sehingga mengurangi jumlah parameter dan perhitungan pada lapisan konvolusi, untuk mengendalikan *Overfitting* yang sering terjadi pada CNN. *Overfitting* adalah suatu kondisi dimana tingkat akurasi saat proses *training* lebih besar daripada saat proses *testing*. Pada penelitian ini bagian *pooling layer* akan menggunakan *max pooling* untuk mengekstrak dari *feature map*.

3) Fully Connected

Fitur fitur yang dihasilkan pada proses sebelumnya kemudian akan dilakukan proses klasifikasi. Proses klasifikasi ini akan dilakukan pada lapisan *fully connected*. Pada lapisan *fully connected* terdapat beberapa lapisan lagi yakni lapisan *input*, lapisan tersembunyi, dan lapisan *output*, yang mirip seperti pada *Multi Layer Perceptron* (MLP). Setiap lapisan yang terdapat pada lapisan *fully connected* memiliki *neuron* yang saling terhubung pada setiap lapisan penyusunnya. *Neuron-neuron* ini akan memiliki *weight* atau bobot yang akan membantu pada proses pemetaan fitur ke dalam kategori masing-masing. Perbedaan antara lapisan *Fully-Connected* dan lapisan konvolusi adalah *neuron* yang terdapat pada lapisan konvolusi akan terhubung hanya ke daerah tertentu pada *input*, sedangkan lapisan *Fully-Connected* memiliki *neuron* yang terhubung saling terhubung secara keseluruhan. Kemudian untuk mendapatkan hasil akhir klasifikasi, pada penelitian ini menggunakan fungsi *softmax*. Pada *layer softmax* akan diimplementasikan *Dropout Regulation*. *Dropout* merupakan sebuah teknik regulasi jaringan syaraf dengan tujuan memilih beberapa *neuron* secara acak dan tidak akan dipakai selama proses pelatihan, dengan kata lain *neuron-neuron* tersebut dibuang secara acak.

III. HASIL DAN PEMBAHASAN

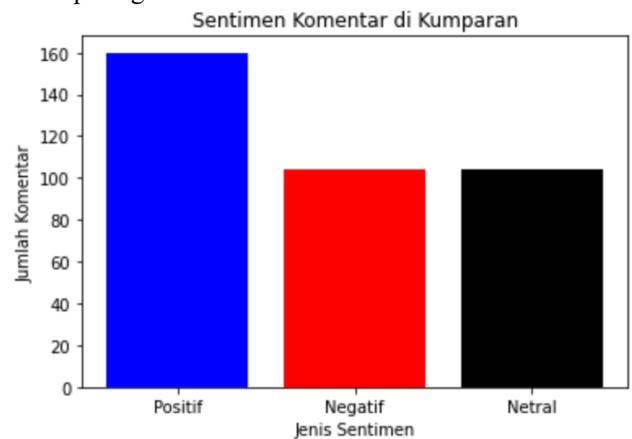
Pada penelitian ini klasifikasi sentimen teks komentar kedalam tiga kelas, yaitu positif, negatif dan netral dilakukan menggunakan algoritma *Convolutional Neural Network* (CNN). Proses utama dalam pembuatan model untuk melakukan klasifikasi sentimen teks komentar diawali dengan proses *crawling data*. *Crawling data* akan menghasilkan dataset yang akan digunakan pada penelitian ini dilakukan pada tiga akun berita, @kumparancom, @cnnindonesia, dan @detikcom. Dari dataset yang telah didapat, data akan dibagi menjadi dua kelompok, yaitu 10% data *testing* dan 90% data *training*.

Proses *training* dilakukan menggunakan *packages Keras*. *Keras* merupakan salah satu modul yang dibuat oleh tim dari *Google* untuk mempermudah pengguna dalam melakukan penelitian mengenai *neural network*. Proses *training data* akan

menghasilkan sebuah model yang kemudian digunakan untuk melakukan pengujian pada *data testing*. Parameter penilaian untuk mengukur tingkat keberhasilan model CNN adalah nilai akurasi. Nilai akurasi dari model yang dihasilkan dapat ditentukan melalui pengujian pada *data testing*.

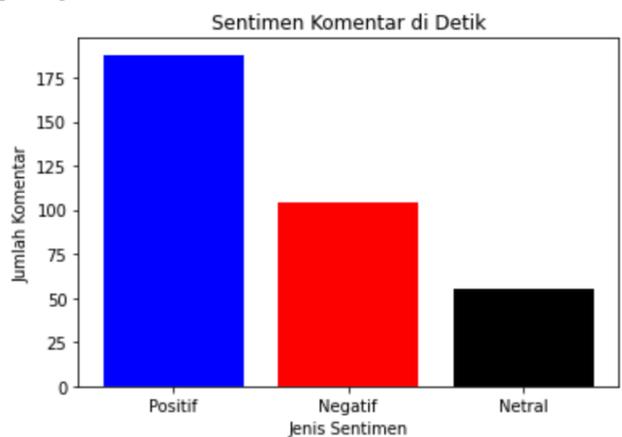
1) Hasil Crawling

Proses *crawling* data pada akun *Instagram* kumparan (@kumparancom) menghasilkan 368 komentar. *Dataset* yang telah dihasilkan dari proses *crawling* kemudian diberi label untuk mengidentifikasi kelas komentar. Proses *labeling* pada *dataset* @kumparancom menghasilkan komentar positif berjumlah 160 komentar, komentar negatif berjumlah 104 komentar, dan komentar netral sebanyak 104 komentar. Perbandingan jumlah komentar klasifikasi sentimen teks komentar pada akun @kumparancom dapat dilihat pada gbr. 5



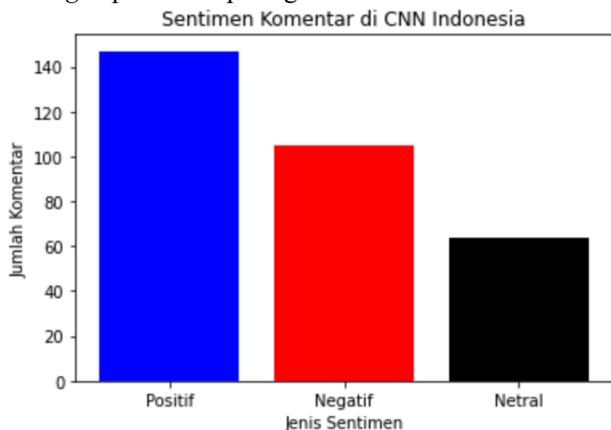
Gbr. 5 Hasil Klasifikasi sentimen teks komentar pada akun @kumparancom

Crawling Data pada akun *instagram* @detikcom menghasilkan 347 data komentar. Proses *crawling* ini menghasilkan 188 komentar berlabel positif, 104 komentar berlabel negatif 55 komentar netral. Perbandingan jumlah komentar yang telah melalui proses *labeling* dapat dilihat pada gbr.6



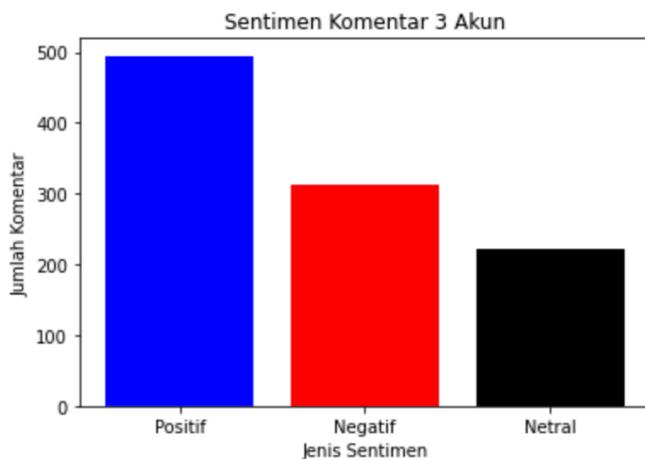
Gbr. 6 Perbandingan jumlah komentar pada akun @detikcom

Proses *crawling* pada akun @cnnindonesia menghasilkan total 316 data komentar, dengan rincian 147 komentar positif, 105 komentar negatif, dan 64 komentar netral. Perbandingan jumlah komentar yang didapatkan dari proses *labeling* dapat dilihat pada gbr.7



Gbr. 7 Perbandingan jumlah komentar pada akun @CNNIndonesia

Proses *crawling* pada ketiga akun *Instagram* telah menghasilkan total 1031 data keseluruhan. Total komentar yang dilabeli positif sebanyak 495, 313 data diberi label negatif, dan 223 data dilabeli netral. Perbandingan jumlah akumulasi dari hasil *crawling* ketiga akun dapat dilihat pada gbr.8



Gbr. 8 Perbandingan jumlah komentar dari ketiga akun *instagram*

2) WordCloud

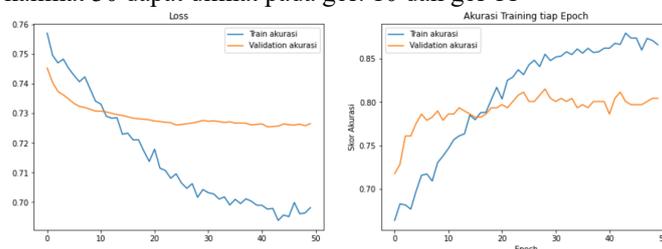
WordCloud dapat menampilkan kata dengan menyoroti kata yang sering muncul dari data. Hasil *crawling data* pada ketiga akun *Instagram* tersebut menghasilkan *wordcloud* yang dapat dilihat pada gbr.9



Gbr. 9 *WordCloud* dari ketiga akun *Instagram*

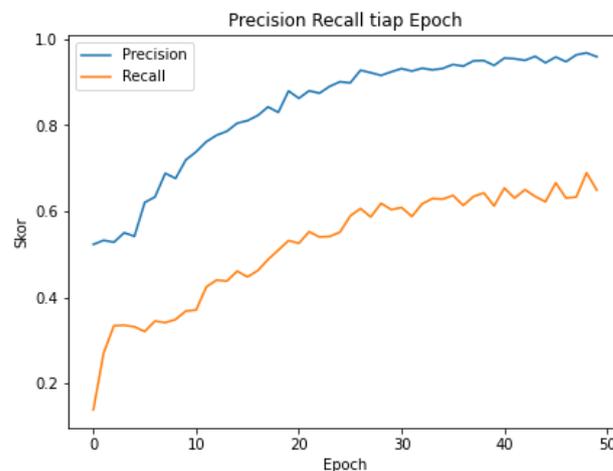
3) Presisi, Akurasi dan Recall

Presisi adalah tingkat ketepatan informasi yang diminta dengan jawaban yang dihasilkan oleh sistem. Akurasi adalah tingkat kedekatan antara jawaban yang dihasilkan oleh sistem dengan jawaban aktual. *Recall* adalah tingkat keberhasilan sistem dalam menemukan ulang sebuah informasi. Presisi, akurasi dan *recall* sangat dipengaruhi oleh *tunning hyperparameter*. *Tunning Hyperparameter* adalah melakukan perubahan nilai atau *value* dari *hyperparameter* pada model untuk menentukan model yang terbaik dari pilihan hasil yang ada berdasarkan parameter tersebut. *Hyperparameter* yang dilakukan *tunning* adalah jumlah *epoch*, ukuran *input* kalimat, skenario data, jumlah *filter*, ukuran *kernel*, dan *learning rate*. Dari proses *tunning hyperparameter* pada tiap parameter, telah didapatkan parameter yang menghasilkan nilai akurasi paling optimal. Kemudian nilai dari tiap parameter tersebut digunakan pada proses *training*. Proses *training* menggunakan 50 epoch, dengan nilai *learning rate* 0.001, ukuran *kernel* 2, *filter* 50, dan skenario 10:90 pada data *training* dengan ukuran input kalimat 30 dapat dilihat pada gbr. 10 dan gbr 11



Gbr. 10 Nilai akurasi pada data *training*

Berdasarkan gbr. 10 akurasi dari data *training* mencapai 88%, sedangkan nilai akurasi dari data *validation* mencapai 82%.



Gbr. 11 Nilai presisi dan recall pada data *training*

Pada gbr. 11 Penggunaan algoritma CNN mampu mencapai nilai presisi hingga 96% dan nilai recall hingga 68%. Maka dari itu perlu adanya optimasi pada arsitektur CNN sehingga dapat mencapai hasil yang lebih baik.

Berdasarkan kedua grafik pada Gbr. 10 dan Gbr. 11 presisi, akurasi *training* dan *recall* akan terus meningkat

seiring bertambahnya *epoch*, namun akurasi dari data *validation* tidak mengalami peningkatan. Hal ini disebabkan adanya *noise* di dalam data. Sehingga perlu dilakukan *early stop* atau memberikan *epoch* yg paling optimal agar model tidak terlalu mempelajari *noise* sehingga model menjadi *overfitting*.

IV. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan, kesimpulan yang diperoleh dari penggunaan CNN untuk *Opinion Mining* terhadap pemberitaan terkait *corona* pada akun media sosial berita nasional adalah:

1. Sistem yang dirancang telah berhasil melakukan proses *Opinion Mining* pada *dataset* yang bersumber dari komentar pada unggahan terkait pemberitaan corona di akun berita nasional yaitu kumparan (@kumparancom), cnn (@cnnindonesia), dan detik (@detikcom) di media sosial *Instagram*. *Dataset* yang didapatkan dari proses crawling terhadap tiga akun tersebut mencapai 1031 data. Total komentar yang dilabeli positif sebanyak 495, 313 data diberi label negatif, dan 223 data dilabeli netral.
2. Sistem mampu mengklasifikasikan komentar di *instagram* kedalam kategori netral, positif dan negatif. Penggunaan CNN dalam mengklasifikasikan komentar menggunakan 50 *epoch*, dengan nilai *learning rate* 0.001, ukuran kernel 2, filter 50, dan skenario 10:90 dengan ukuran *input* kalimat 30 menghasilkan presentase presisi mencapai 96%, akurasi mencapai 88% pada data *training* dan 82% pada data *testing*, dan recall mencapai 68%.

V. SARAN

Pengambilan data komentar dari media sosial *Instagram* memiliki kelemahan pada tata bahasa yang digunakan, akun yang berkomentar kebanyakan menggunakan bahasa yang tidak baku. Untuk penelitian selanjutnya, disarankan menambahkan tahapan *preprocessing* yang lebih baik dan terinci untuk dapat mengolah data komentar tidak beraturan dari media sosial sehingga hasil dari analisis sentimen yang dilakukan dapat lebih baik.

UCAPAN TERIMA KASIH

Terima kasih kepada orang tua dan keluarga yang selalu memberikan dukungan dan kesabaran dalam menemani saya menuntut ilmu. Juga kepada seluruh civitas akademika jurusan Teknik Informatika Universitas Negeri Surabaya, khususnya teman-teman S1 Teknik Informatika angkatan 2016, untuk seluruh bantuan dan dukungan sehingga penelitian ini dapat dilakukan dengan tepat dan benar.

REFERENSI

- [1] W. A. Social, "DIGITAL 2020: INDONESIA," Hootsuite, 2020. [Online]. Available: <https://datareportal.com/reports/digital-2020-indonesia>. [Diakses 25 April 2020].
- [2] Info Brand Group, "Survei TRAS N CO Indonesia: Jawa Barat Paling Banyak Cari Informasi Virus Corona di Internet," 2020. [Online]. Available: <https://infobrand.id/survei-tras-n-co-indonesia-jawa-barat-paling-banyak-cari-informasi-virus-corona-di-internet.phtml>. [Diakses 25 April 2020].
- [3] B. Liu, "Opinion Mining," tanpa tahun. [Online]. Available: <https://www.cs.uic.edu/~liub/FBS/opinion-mining.pdf>. [Diakses 25 April 2020].
- [4] C. K. L. Pastor, "Sentiment Analysis of Filipinos and Effects of Extreme Community Quarantine Due to Coronavirus (COVID-19) Pandemic," 13 April 2020. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3574385. [Diakses 30 April 2020].
- [5] W. A. Luqyana, I. Cholissodin dan R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer Vol.2(11)*, pp. 4704-4713, 2018.
- [6] A. R. C dan Y. Lukito, "Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes," *ULTIMATICS, Vol. IX(1)*, pp. 50-58, 2017.
- [7] N. Rochmawati dan S. C. Wibawa, "Opinion Analysis on Rohingya using Twitter Data," dalam *IOP Conference Series: Materials Science and Engineering*, 2018.
- [8] S. Liao, J. Wang, R. Yu, o. Sato dan Z. Cheng, "CNN for situations understanding based on sentiment analysis of twitter data," dalam *8th International Conference on Advances in Information Technology*, Macau, 2016.
- [9] R. Rokhana, J. Priambodo, T. Karlita, I. M. G. Sunarya, E. M. Yuniarno, I. K. E. Purnama dan M. H. Purnomo, "Convolutional Neural Network untuk Pendeteksian Patah Tulang Femur pada Citra Ultrasonik B-Mode," *JNTETI*, vol. 8, no. 1, pp. 59-67, 2019.
- [10] X. Ouyang, P. Zhou, C. H. Li dan L. Liu, "Sentiment Analysis Using Convolutional Neural Network," dalam *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, Liverpool, 2015.
- [11] B. M. H. Hidayat dan R. E. Putra, "Penerapan CNN dengan Filter Gabor sebagai feature extractor untuk Content-Based Image Retrieval," *Journal of Informatics and Computer Science*, vol. 1, no. 1, pp. 16-25, 2019.