



Analisis Sentimen Jogja Darurat Sampah di Twitter menggunakan Ekstraksi Fitur Model Word2Vec dan Convolutional Neural Network

Yoga Yusanto*, Mutaqin Akbar

Fakultas Teknologi Informasi, Prodi Informatika, Universitas Mercu Buana Yogyakarta, Yogyakarta, Indonesia

Email: ^{1,*}18111034@student.mercubuana-yogya.ac.id, ²mutaqin@mercubuana-yogya.ac.id

Email Penulis Korespondensi: 18111034@student.mercubuana-yogya.ac.id

Abstrak—Karena mengalami darurat sampah, Daerah Istimewa Yogyakarta menjadi sorotan publik dan ramai menjadi perbincangan. Banyak kelompok masyarakat yang melemparkan opininya melalui berbagai media sosial, khususnya *Twitter*. Tidak bisa dipungkiri, saat ini memang *Twitter* adalah salah satu tempat untuk bebas mengutarakan pendapat. Oleh sebab itu, analisis sentimen perlu berperan untuk upaya mengelompokkan opini masyarakat terkait sesuatu yang sedang *hype* atau viral dan membaginya ke dalam tiga kategori yaitu, positif, negatif, dan netral. Pada pelatihan ini, dataset yang diambil dengan menggunakan teknik *scrapping* dan *tools tweetscraper* dari *actor APIFY* yakni web.harvester/easy-twitter-search-scraper. Metode yang digunakan pada tinjauan ini adalah metode klasifikasi *Convolutional Neural Network* (CNN) menggunakan ekstraksi model *Word2Vec*. Tinjauan ini menggunakan 505 *tweet* dalam Bahasa Indonesia dengan tagar *JogjaDaruratSampah* (*#JogjaDaruratSampah*) dan *TPSTPiyungan* (*#TPSTPiyungan*) sebagai data, dan dari 505 *tweet*, 381 *tweet* digunakan sebagai data latih. Sedangkan 124 *tweet* sisanya digunakan sebagai data uji. Hasil pengujian data latih akurasi paling tinggi didapat pada *epoch* ke 19 dengan nilai 90%, dapat disimpulkan dengan adanya proses pengujian bahwa tinjauan ini dapat mengenali sentimen positif, negatif, dan netral dengan ketepatan 53%. Hasil analisis sentimen menunjukkan banyaknya cuitan negatif sebesar 49.7% dari total 505 *tweets*.

Kata Kunci: Analisis Sentimen; *JogjaDaruratSampah*; CNN; *Word2Vec*; *Twitter*

Abstract—Due to a waste emergency, the Special Region of Yogyakarta has garnered public attention and sparked discussions. Numerous community groups express their opinions through various social media platforms, especially *Twitter*. It's undeniable that *Twitter* is currently one of the places for freely expressing opinions. Therefore, sentiment analysis plays a crucial role in efforts to categorize public opinions on something trending or viral into three categories: positive, negative, and neutral. In this study, the dataset was obtained using scraping techniques and the tweetscraper tool from the *APIFY* actor web.harvester/easy-twitter-search-scraper. The method employed in this analysis is the *Convolutional Neural Network* (CNN) classification method using *Word2Vec* model extraction. The study involves 505 tweets in Bahasa Indonesia with the hashtags *#JogjaDaruratSampah* (*#JogjaDaruratSampah*) and *#TPSTPiyungan* as data. Out of these, 381 tweets are utilized as training data, and the remaining 124 tweets are used as test data. The highest accuracy in testing the training data was achieved in the 19th epoch with a 90% accuracy rate. It can be concluded from the testing process that this study can identify positive, negative, and neutral sentiments with an accuracy of 53%. The sentiment analysis results indicate a significant amount of negative tweets, accounting for 49.7% of the total 505 tweets.

Keywords: Sentiment Analysis; *JogjaDaruratSampah*; CNN; *Word2Vec*; *Twitter*

1. PENDAHULUAN

Sejak dulu sampah menjadi polemik permasalahan dari berbagai belahan dunia, baik di negara berkembang ataupun di negara maju. Dengan sifat yang susah terurainya, sampah akan terus menerus bertambah. Hingga pada akhirnya di era sekarang permasalahan sampah yang sangat krusial ini menjadi hal yang lumrah. Populasi manusia yang meningkat dengan cepat, mengakibatkan produksi sampah yang dihasilkan masyarakat juga meningkat. Sedangkan kepedulian dan kebiasaan masyarakat untuk mengelola sampah masih minim. Indonesia sendiri tak luput dari permasalahan sampah, bahkan Indonesia termasuk negara kedua penyumbang sampah terbesar setelah China (Ratnaningtyas, 2020). Banyaknya sampah yang tidak dikelola dengan benar dan dibuang secara sembarangan mengakibatkan banyak bencana yang terjadi seperti banjir, tanah tercemar, air sungai tercemar, dan udara yang tercemar. Kemudian sialnya D.I Yogyakarta adalah salah satu provinsi di Indonesia yang kerap kali terjadi darurat sampah. Puluhan ton sampah terlihat mangkrak di pinggir jalan, di tempat penampungan sampah sementara, dan juga di tempat sampah pribadi warga yang puncaknya pada tahun 2022 kemarin (Susilo et al., 2023).

Sekber Kartamantul mencatat bahwa rata-rata 700 ton sampah masuk ke TPST Piyungan setiap harinya yang merupakan akumulasi dari tiga kabupaten yaitu Kabupaten Bantul, Sleman, dan Yogyakarta (Istriani, 2023). Saat ini krisis sampah lagi-lagi menggemparkan masyarakat Yogyakarta karena penutupan Tempat Pembuangan Sampah Terpadu Piyungan (TPST). Permasalahan sampah yang terjadi secara berturut-turut setiap tahun membuat warga sekitar TPA Piyungan geram dan menutup akses jalan masuk ke TPST Piyungan dengan harapan agar TPST Piyungan ditutup secara permanen, warga sekitar tidak henti-hentinya terus mendesak pemerintah setelah mengeluhkan jalan yang rusak sebab mobilitas truk sampah yang berlalu-lalang, juga tercemarnya kualitas air dan lingkungan sekitar (Susilo et al., 2023). Dampaknya banyak depo-depo dan juga tempat pembuangan sampah yang mengalami ledakan sampah.

Tempat pembuangan yang telah melebihi batas maksimal akan berakibat fatal, di antaranya adalah sampah yang berserakan di jalanan juga bau busuk yang menyengat sehingga mengganggu kenyamanan warga sekitar maupun pengguna jalan. Oleh karena itu, timbul berbagai respon akibat sampah di Yogyakarta yang menumpuk, banyak media

massa turut menyoroiti baik media lokal atau nasional. Berbagai surat kabar, cetak, elektronik, maupun digital ikut andil memberikan informasi mengenai fenomena tersebut.

Kejadian permasalahan sampah di Yogyakarta memunculkan respon viral di media sosial. Salah satu yang digunakan publik untuk mengungkapkan sikap mereka untuk permasalahan tersebut yaitu Twitter. Di Indonesia sendiri, pengguna sosial media menyentuh kisaran 150 juta pengguna, dan lebih dari 52 persennya adalah pengguna Twitter dari jumlah total pengguna sosial media (Rachman & Pramana, 2020). Hal ini menunjukkan peluang sumber data yang begitu besar untuk dimanfaatkan pada sebuah tinjauan. Twitter merupakan salah satu platform media sosial yang sangat populer dengan jutaan pengguna aktif. Kemudahan dan kesederhanaannya, serta kebebasan beropini menjadikan Twitter sangat menarik di mata masyarakat. Masyarakat atau pengguna dapat dengan bebas melontarkan opini atau pendapat dalam bentuk teks yang disebut *tweets*, dan *tweest* juga nantinya yang akan mejadi sumber data yang penting untuk analisis sentimen (Fikri et al., 2020).

Analisis sentimen merupakan bagian dari *Natural Language Processing* (NLP) yang memiliki tujuan untuk membantu proses memahami dan mengidentifikasi isi dataset perasaan, emosi, serta opini yang terkandung dalam teks terhadap suatu fenomena yang bersifat positif, negatif, atau netral (Fikri et al., 2020). Banyak cara untuk mengukur analisis sentimen ini, di antaranya adalah *Naïve Bayes* (NB), *K-Nearest Neighbour* (K-NN), *Support Vector Machine* (SVM), dan *Convolutional Neural Network* (CNN).

Pada penelitian sebelumnya oleh Darwis dkk pada tahun 2021, metode yang diterapkan untuk klasifikasi data tersebut adalah *Naïve Bayes Classifier*. Klasifikasi tersebut kemudian dibagi menjadi tiga kelas di antaranya adalah positif, negatif, dan netral. Hasil yang didapatkan yakni hasil uji akurasi klasifikasi sebesar 69,97% (Darwis et al., 2021). Penelitian lainnya oleh Laurensz dan Eko Sedyono pada tahun 2021 yang membandingkan dua klasifikasi, yaitu *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB). Tinjauan ini juga membagi dua klasifikasi sentimen menjadi dua kelas, yaitu positif dan negatif. Hasil menunjukkan bahwa *Naïve Bayes* (NB) memiliki akurasi lebih tinggi dibandingkan dengan SVM yakni 85,95% dibanding 84,41%. Oleh sebab itu, tinjauan ini menarik kesimpulan bahwa *Naïve Bayes* adalah metode klasifikasi terbaik untuk analisis sentimen terhadap tindakan vaksinasi dalam usaha mengatasi pandemi COVID-19 (Laurensz & Eko Sedyono, 2021).

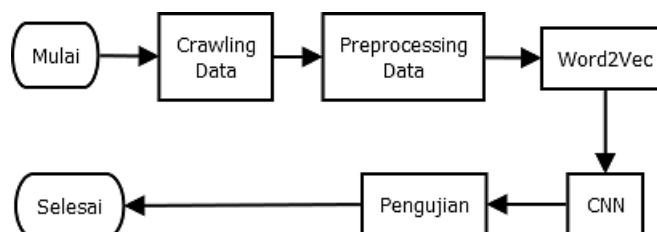
Selain pengukuran akurasi klasifikasi, waktu proses analisis juga dapat dibandingkan seperti penelitian yang dilakukan oleh Nasution dan Hayaty pada tahun 2019. Pada penelitian tersebut menguji dua algoritma klasifikasi, yaitu *K-Nearest Neighbor* (K-NN) dan *Support Vector Machine* (SVM). Tinjauan ini membangun model dari tiap-tiap algoritma menggunakan dataset sentimen. Tujuannya untuk mengetahui mana yang lebih baik dalam hal akurasi dan waktu proses. Hasil menunjukkan bahwa SVM memiliki akurasi lebih tinggi yaitu 89,70% tanpa *K-Fold Cross Validation* dan 88,76 dengan *K-Fold Cross Validation*. Akan tetapi, K-NN mempunyai waktu yang lebih cepat daripada SVM, yaitu 0.0160 detik tanpa *K-Fold Cross Validation*, dan 0.1505 detik dengan *K-Fold Cross Validation* (Nasution & Hayaty, 2019). Penelitian lainnya oleh Putra dan Syafira pada tahun 2023 dengan membandingkan kinerja kedua klasifikasi tersebut dalam menganalisis data sentimen Twitter yang membahas politik. Hasilnya menunjukkan bahwa *Naïve Bayes* mendapatkan akurasi sedikit lebih tinggi dibanding CNN, yaitu 0,69 dengan 0,68. Meski adanya perbedaan akurasi, selisihnya begitu kecil, dengan begitu menunjukkan bahwa kedua metode tersebut cukup efektif untuk menganalisis sentimen data Twitter dengan pembahasan politik (Putra A P & Syafira A F, 2023).

Tinjauan yang dilakukan menggunakan metode *word2vec* dilakukan oleh Pradana pada tahun 2022. Tinjauan ini memanfaatkan metode *deep learning*, khususnya LSTM atau *Long Short Term Memory* serta *Word2Vec* yang digunakan untuk menganalisis sentimen *tweets* masyarakat. Model LSTM yang dikembangkan dalam tinjauan ini menghasilkan *accuracy* sebesar 95%, *recall* sebesar 93%, *precision* sebesar 93%, serta *F1-Measure* sebesar 95%. Hasil tersebut menunjukkan bahwa metode ini efektif dalam menganalisis sentimen masyarakat terkait pemindahan ibu kota dan dapat memberikan pemahaman yang lebih baik mengenai pandangan publik terhadap perubahan tersebut (Pradana et al., 2022).

Berdasarkan latar belakang dan tinjauan pustaka tersebut, maka pada artikel ini diusulkan sebuah pendekatan analisis sentimen terhadap penanganan permasalahan sampah di provinsi Daerah Istimewa Yogyakarta menggunakan ekstraksi fitur model Word2Vec dan CNN.

2. METODE PENELITIAN

2.1 Tahapan Penelitian



Gambar 1. Tahap Penelitian



Pada penelitian ini, ada beberapa tahapan yang perlu dilewati supaya dapat memperoleh hasil yang maksimal. Secara garis besar tahapan penelitian ini adalah crawling data, preprocessing data, Word2Vec, klasifikasi dengan CNN, dan pengujian yang dapat dilihat pada Gambar 1.

2.1.1 Crawling Data

Crawling data atau pengambilan data menggunakan Teknik *web scrapping* dengan *tool tweetscrapper* dari *actor APIFY* yaitu *web.harvester/easy-twitter-search-scraper*. Pengambilan data dengan menggunakan *keyword* #JogjaDaruratSampah dan #TPSTPiyungan. Sebanyak 711 *tweets* data yang dikumpulkan dengan rentang pengambilan data mulai tanggal 01 Januari 2023 sampai 31 Desember 2023. Kemudian data tersebut diproses secara manual untuk menghilangkan *tweets* yang mengandung iklan dan *tweets* yang tidak relevan dengan *keyword* menjadi 505 data yang kemudian dilakukan pelabelan data secara manual menjadi sentimen positif, negatif, dan netral. setelah itu data dibagi menjadi data latih sebanyak 381 data dan data uji sebanyak 124 data.

2.1.2 Preprocessing Data

Mayoritas data teks yang diambil dari Twitter menggunakan kata yang tidak baku, misalnya penggunaan bahasa daerah, atau singkatan yang tidak ditemukan dalam Kamus Besar Bahasa Indonesia. Maka, diperlukan langkah *preprocessing* untuk mengolah data supaya lebih terstruktur. *Preprocessing data* adalah sebuah proses pembersihan dan persiapan teks untuk dilakukan proses klasifikasi. Seluruh tahapan *preprocessing* melibatkan beberapa langkah yaitu:

a. *Cleansing*

Cleansing adalah langkah membersihkan data yang akan digunakan dari fitur hingga kata-kata yang tidak diperlukan, dengan tujuan untuk mengurangi *noise* yang dapat memicu proses perhitungan dalam pengklasifikasian tidak maksimal (Sawicki et al., 2023).

Tabel 1. Contoh *Cleansing*

Teks <i>Tweet</i>	Hasil <i>Cleansing</i>
Kondisi ini diperparah bila Yogyakarta diguyur hujan.. sampah yang menumpuk menjadi basah, lembek dan mudah busuk. #JogjaDaruratSampah	Kondisi ini diperparah bila Yogyakarta diguyur hujan sampah yang menumpuk menjadi basah lembek dan mudah busuk
Tempat sampah di rumahku udah ga muat...bau lagi..bingung mau buang sampah dimana #jogjadaruratsampah	Tempat sampah di rumahku udah ga muat bau lagi bingung mau buang sampah dimana

b. *Tokenization*

Tokenization adalah sebuah proses untuk memisah satuan kata di dalam kalimat atau paragraf menjadi *token* atau potongan kata tunggal (*tremmed word*). Selain itu *tokenization* juga menghilangkan atau menghapus kata tertentu yang dianggap sebagai tanda baca (Toraman et al., 2023).

Tabel 2. Contoh *Tokenization*

Teks <i>Tweet</i>	Hasil <i>Tokenization</i>
Kondisi ini diperparah bila Yogyakarta diguyur hujan sampah yang menumpuk menjadi basah lembek dan mudah busuk	“Kondisi”, “ini”, “diperparah”, “bila”, “Yogyakarta”, “diguyur”, “hujan”, “sampah”, “yang”, “menumpuk”, “menjadi”, “basah”, “lembek”, “dan”, “mudah”, “busuk”
Tempat sampah di rumahku udah ga muat bau lagi bingung mau buang sampah dimana	“Tempat”, “sampah”, “dirumahku”, “udah”, “ga”, “muat”, “bau”, “lagi”, “bingung”, “mau”, “buang”, “sampah”, “dimana”

c. *Case-folding*

Case-folding adalah langkah menyamakan *case* di dalam dokumen. Proses ini dimanfaatkan untuk memudahkan pencairan dalam dokumen (Yuliska & Syaliman, 2020).

Tabel 3. Contoh *Case-folding*

Teks <i>Tweet</i>	Hasil <i>Case-folding</i>
“Kondisi”, “ini”, “diperparah”, “bila”, “Yogyakarta”, “diguyur”, “hujan”, “sampah”, “yang”, “menumpuk”, “menjadi”, “basah”, “lembek”, “dan”, “mudah”, “busuk”	“kondisi”, “ini”, “diperparah”, “bila”, “yogyakarta”, “diguyur”, “hujan”, “sampah”, “yang”, “menumpuk”, “menjadi”, “basah”, “lembek”, “dan”, “mudah”, “busuk”
“Tempat”, “sampah”, “dirumahku”, “udah”, “ga”, “muat”, “bau”, “lagi”, “bingung”, “mau”, “buang”, “sampah”, “dimana”	“tempat”, “sampah”, “dirumahku”, “udah”, “ga”, “muat”, “bau”, “lagi”, “bingung”, “mau”, “buang”, “sampah”, “dimana”

d. Penghilangan *Stopwords*

Stopwords diartikan sebagai *term* yang tidak memiliki relasi dengan subjek utama dari basis data meskipun kata tersebut kerap kali muncul dalam sebuah dokumen.

Tabel 4. Contoh *Stopwords*

Teks <i>Tweet</i>	Hasil <i>Stopwords</i>
“kondisi”, “ini”, “diperparah”, “bila”, “yogyakarta”, “diguyur”, “hujan”, “sampah”, “yang”, “menumpuk”, “menjadi”, “basah”, “lembek”, “dan”, “mudah”, “busuk”	“kondisi”, “diperparah”, “Yogyakarta”, “diguyur”, “hujan”, “sampah”, “menumpuk”, “basah”, “lembek”, “mudah”, “busuk”
“tempat”, “sampah”, “dirumahku”, “udah”, “ga”, “muat”, “bau”, “lagi”, “bingung”, “mau”, “buang”, “sampah”, “dimana”	“tempat”, “sampah”, “rumahku”, “udah”, “ga”, “muat”, “bau”, “lagi”, “bingung”, “buang”, “sampah”, “dimana”

e. *Stemming*

Stemming merupakan langkah ketika kata yang tidak mengandung *stopwords* direduksi ke bentuk *stemmed word (term)* yang sesuai. Kata tersebut mengalami proses *stemming* untuk memperoleh bentuk awal atau akar dari kata tersebut dengan cara menghilangkan imbuhan. Dengan menggunakan cara ini, akan diperoleh kelompok kata yang memiliki makna yang sama tapi dengan bentuk berbeda secara sintaksis (Rifai & Winarko, 2019).

Tabel 5. Contoh *Stemming*

Teks <i>Tweet</i>	Hasil <i>Stemming</i>
“kondisi”, “diperparah”, “Yogyakarta”, “diguyur”, “hujan”, “sampah”, “menumpuk”, “basah”, “lembek”, “mudah”, “busuk”	“kondisi”, “parah”, “yogyakarta”, “guyur”, “hujan”, “sampah”, “tumpuk”, “basah”, “lembek”, “mudah”, “busuk”
“tempat”, “sampah”, “rumahku”, “udah”, “ga”, “muat”, “bau”, “lagi”, “bingung”, “buang”, “sampah”, “dimana”	“tempat”, “sampah”, “rumah”, “udah”, “ga”, “muat”, “bau”, “lagi”, “bingung”, “buang”, “sampah”, “mana”

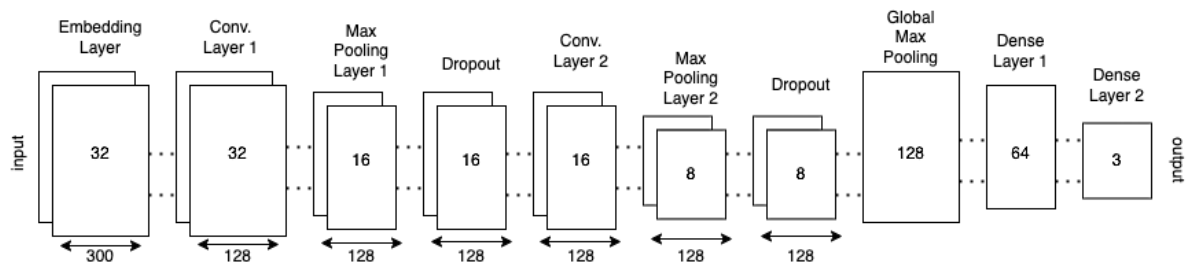
2.1.3 Word2Vec

Word2Vec adalah salah satu algoritma *word embedding* yang memetakan setiap kata ke dalam vektor. *Word2Vec* merepresentasikan kata menjadi vektor yang memuat makna semantik dari kata tersebut. Metode *word embedding* ini termasuk ke dalam penerapan *unsupervised learning* menggunakan *deep neural network*. *Word2Vec* memiliki dua arsitektur representasi konteks yaitu *Continuous Bag Of Word (CBOW)* dan *Skip-gram*. *Continuous Bag of Words (CBOV)* adalah model arsitektur yang memperkirakan kata target atau kata tengah berdasar pada konteks kata yang berdekatan dari beberapa kata sebelum dan sesudah kata target (Al-Saqqa & Awajan, 2019). Sedangkan *Skip-gram*, model arsitektur yang memperkirakan kata target dalam rentang sebelum dan sesudah kata target (Nurdin et al., 2020).

Tabel 6. Contoh *Word2Vec*

<i>Tweet</i>	<i>Word2vec</i>
kondisi ini diperparah bila yogyakarta diguyur hujan sampah yang menumpuk menjadi basah lembek dan mudah busuk	[0.00315225 0.00315225 0.00547597 0.00741556 0.00912817 0.01068435 0.01212381 0.01347162 0.01474487 0.0159558]

2.1.4 Convolutional Neural Network (CNN)



Gambar 2. Arsitektur CNN

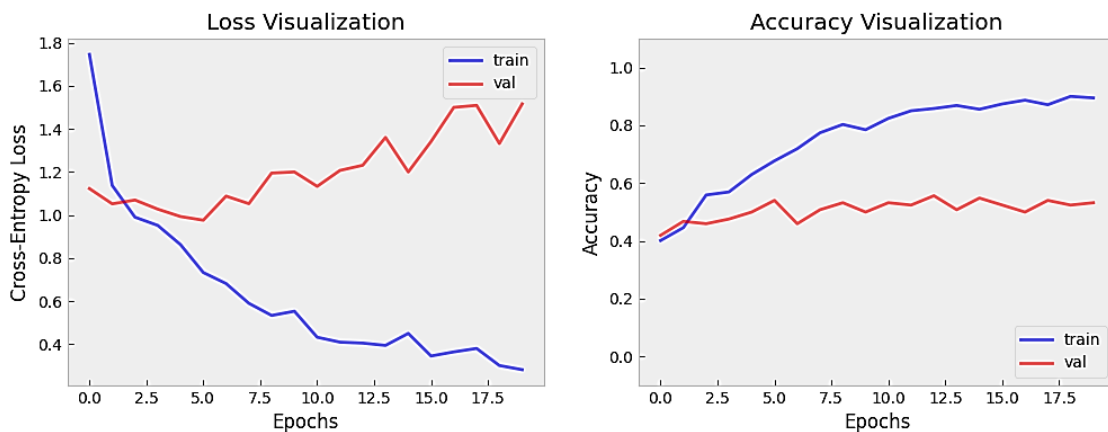
Convolutional Neural Network (CNN) adalah pengembangan dari *Multi Layer Perceptron (MLP)* yang dirancang khusus untuk mengolah data dua dimensi, seperti gambar dan teks. CNN menggunakan teknik konvolusi mengekstrak karakter penting dari data (Akbar, 2021). Konvolusi merupakan sebuah *sliding window* yang biasa disebut *filter* yang berfungsi untuk membagi matriks representasi teks menjadi beberapa *window* dan setelah itu menjumlahkannya (Akbar et al., 2022). Hasil penjumlahan disebut *feature maps*, teknik yang lain *max-pooling* yaitu pemilihan nomor paling besar dari setiap *feature map* (Yuliska et al., 2021). Arsitektur CNN yang digunakan pada tinjauan ini yaitu 2 *convolutional layer*, 2 *max-pooling layer*, 2 *dropout*, *global max pooling* dan *dense layer*. Model *Convolutional Neural Network* yang digunakan dapat dilihat pada Gambar 2.

3. HASIL DAN PEMBAHASAN

Model arsitektur *Convolutional Neural Network* (CNN) menghasilkan sebuah model dengan menggunakan beberapa layer yang dimiliki. Pembentukan model mempunyai suatu *output* informasi model yang dapat diproyeksikan menggunakan *summary()*. Layer yang terbentuk dapat dilihat pada Gambar 2. *Input layer* adalah *layer input data* yang pertama kali. *Embedding layer* dengan nilai keluaran (*none, 32, 300*) yang memiliki nilai guna membuat suatu proyeksi vektor dengan nilai dimensi sebesar 300 atau selaras dengan *inputan embedding dim*. *Conv1d layer* merupakan sebuah *layer* yang menerima *input* dari *output embedding layer*. Layer dengan nilai keluaran (*none, 32, 128*) memiliki fungsi untuk membuat sebuah *kernel* yang menjalankan *filter* dari *output embedding layer*. Setelah itu, *global_max_pooling_id layer* mengambil nilai maksimal dari perhitungan konvolusi oleh *kernel* yang ditetapkan menjadi nilai elemen kata terkuat. *Global_max_pooling_id* mengeluarkan *output* yang diterima *dropout layer* dan melakukan *drop* terhadap nilai *signal* yang kurang memuaskan, yaitu 0.5 pada nilai paling bawah. *Dense layer* bertindak sebagai *fully connected layer* mempunyai *neuron* yang terhubung dengan *dropout layer* tersebut.

Tabel 7. Output Epoch

Epoch	Loss	Accuracy	Validation loss	Validation accuracy
1	1.7462	0.4016	1.1230	0.4194
2	1.1377	0.4462	1.0521	0.4677
3	0.9902	0.5591	1.0698	0.4597
4	0.9529	0.5696	0.9931	0.4758
5	0.8627	0.6299	0.9931	0.5000
6	0.7331	0.6772	0.9762	0.5403
7	0.6818	0.7192	1.0880	0.4597
8	0.5901	0.7743	1.0529	0.5081
9	0.5337	0.8031	1.1950	0.5323
10	0.5533	0.7848	1.2003	0.5000
11	0.4325	0.8241	1.1333	0.5323
12	0.4096	0.8504	1.2073	0.5424
13	0.4052	0.8583	1.2313	0.5565
14	0.3946	0.8688	1.3606	0.5081
15	0.4503	0.8556	1.2003	0.5484
16	0.3454	0.8740	1.3417	0.5242
17	0.3644	0.8871	1.5004	0.5000
18	0.3804	0.8714	1.5096	0.5403
19	0.3011	0.9003	1.3325	0.5242
20	0.2817	0.8950	1.5161	0.5323

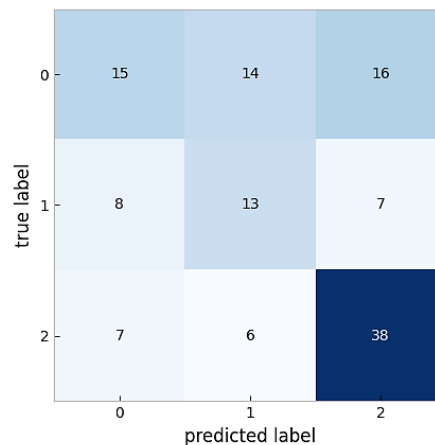


Gambar 3. Loss-Accuracy Visualization

Berdasarkan Tabel 7 pelatihan data menggunakan *epoch* 20, satu *epoch* berarti melatih *neural network* menggunakan semua data pelatihan dalam satu siklus, dan telah menghitung *accuracy* and *loss* dari data latih, juga menghitung *validation loss* dan *validation accuracy* dari data uji. Pada tabel 7 menunjukkan *accuracy* tertinggi ada pada *epoch* ke-19 bernilai 90%. Pada gambar 3 menunjukkan model yang dibuat memberikan akurasi sebesar 90% untuk data pelatihan. Peningkatan nilai *accuracy* dan nilai *loss* mengindikasikan model telah dilatih dengan baik. Kemudian pada *epoch* ke-13 memberikan nilai *validation accuracy* sebesar 55% untuk data pelatihan.

Pelatihan dilakukan menggunakan data yang sudah dipersiapkan sejumlah 381 data latih. Sebelum menjalankan pelatihan model data latih perlu di lakukan konversi menjadi urutan *integer* unik. Gambaran kinerja model dapat dilihat melalui pengujian menggunakan 124 data uji. Penghitungan *accuracy*, *precision*, *recall*, dan *f1-score* model terhadap

dataset uji dilakukan untuk mengetahui performa model tersebut. Evaluasi model diawali dengan melakukan pemanggilan *file* model CNN yang telah disimpan sebelumnya, yaitu ‘*model_cnn.h5*’ dengan sintaksis *load_model()* dari keras kemudian menginisiasikan ke variable *new_model*. *Data_test* baru yang sudah dilakukan dengan model tersebut, yaitu data pengujian. Untuk menerapkan model ke dalam *data_test* uji menggunakan sintaks *model.predict()*. Hasil *predict* yang didapat akan dimasukkan ke dalam variable *y_pred* kemudian dilakukan penerjemahan di dalam label sentimen yang ada di *y_test*. Kemudian adanya penghitungan *accuracy*, *precision*, *recall*, dan *f1-score* model terhadap dataset uji dilakukan untuk mengetahui performa model tersebut. pembentukan konvusi matriks yang mempunyai nilai *true label* dan *predicted label* seperti pada Gambar 4 dilakukan dengan perhitungan terhadap data tersebut. Konvusi matriks dibentuk menggunakan sintaks *confusion_matrix()*. Kemudian penggunaan *classification_report()* untuk menghitung *accuracy*, *precision*, *recall*, dan *f1-score* diperlakukan setelah membua konvusi matriks. Tinjauan ini menggunakan label “positif”, “negatif”, dan “netral” yang merupakan sebuah larik yang memiliki nilai kata atau kalimat sentiment.

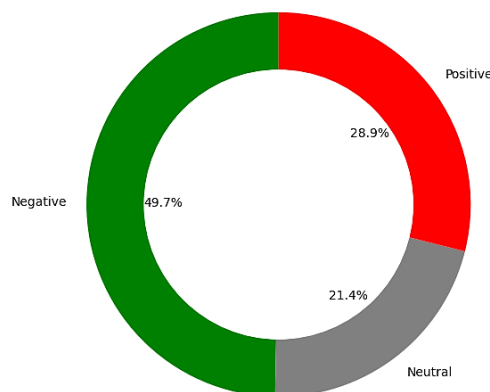


Gambar 4. Confusion Matrix

Berdasarkan Tabel 8, pelatihan menggunakan model *epoch 20* mendapatkan hasil *accuracy* sebesar 53%. Namun *precision* untuk *tweet* netral sebesar 50% menunjukkan ketepatan label aktual negatif dengan *output* label perkiraan netral yang dikeluarkan sistem sangat tinggi. Nilai *recall tweets positive* memperlihatkan keefektifan *system* guna memperoleh kembali label bernilai positif juga cukup rendah. Sedangkan di sisi lain *f1-score tweets* negatif dengan nilai 40% menunjukkan komparasi rata-rata *precision* dengan *recall tweet* negatif cukup baik.

Tabel 8. Hasil Evaluasi

Sentimen	Precision	recall	F1-score	support
Neutral	0.50	0.33	0.40	45
Positive	0.39	0.46	0.43	28
Negative	0.62	0.75	0.68	51
Accuracy			0.53	124



Gambar 5. Distribusi Sentimen

Pada masing-masing label seperti pada gambar 5 dapat dihitung persentase dari hasil analisis sentimen, yaitu sentimen positif, sentimen negatif, dan sentimen netral. Akumulasi keseluruhan terdapat 505 data *tweets* yang terkumpul, 146 data (28.9%) terdapat *tweets* positif, selanjutnya 108 data (21.4%) *tweets* netral, dan terakhir 251 data (49.7%) data *tweets* negatif. Berdasarkan data hasil analisis tersebut, permasalahan darurat sampah di Jogja ini banyak warga yang membuat cuitan negatif.



4. KESIMPULAN

Penggunaan Convolutional Neural Network (CNN) dan Word2Vec ini berhasil membangun program analisis sentimen terhadap tagar JogjaDaruratSampah (#JogjaDaruratSampah) dan tagar TPSTPiyungan (#TPSTPiyungan) di Twitter. Program ini ditujukan untuk mengelompokkan *tweets* yang dibagi menjadi 3 kelas sentimen, yaitu sentimen positif, negatif, dan netral dari pengguna aktif di Twitter. Pada tinjauan ini bahasa pemrograman yang dipakai untuk membentuk program yaitu *Python*, dengan keseluruhan data sebanyak 505 *tweets*. Data tersebut dibagi menjadi dua bagian yaitu, data latih dan data uji dengan jumlah masing-masing 381 data latih dan 124 data uji. Sebelum pengujian dilakukan, data harus melalui proses labeling dan *data preprocessing*. Setelah itu lanjut ke tahap *word embedding* menggunakan *Word2Vec* dan melakukan uji model CNN. Hasil pengujian model menghasilkan ketepatan tertinggi dengan menggunakan *epoch* 20 pada iterasi *epoch* yang ke-19 dengan ketepatan 90%. Kemudian hasil sentimen analisis menunjukkan bahwa banyak pengguna aktif Twitter membuat cuitan negatif terhadap polemik sampah di Jogja sebanyak 49.7% dari total 505 data *tweets*.

REFERENCES

- Akbar, M. (2021). Traffic sign recognition using convolutional neural networks. *Jurnal Teknologi Dan Sistem Komputer*, 9(2), 120–125. <https://doi.org/10.14710/jtsiskom.2021.13959>
- Akbar, M., Purnomo, A. S., & Supatman, S. (2022). Multi-Scale Convolutional Networks untuk Pengenalan Rambu Lalu Lintas di Indonesia. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 11(3), 310–315. <https://doi.org/10.32736/sisfokom.v11i3.1452>
- Al-Saqqa, S., & Awajan, A. (2019). The Use of Word2vec Model in Sentiment Analysis: A Survey. *Proceedings of the 2019 International Conference on Artificial Intelligence, Robotics and Control*, 39–43. <https://doi.org/10.1145/3388218.3388229>
- Darwis, D., Siskawati, N., & Abidin, Z. (2021). PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN REVIEW DATA TWITTER BMKG NASIONAL. *Jurnal Tekno Kompak*, 15(1), 131. <https://doi.org/10.33365/jtk.v15i1.744>
- Fikri, M. I., Sabrila, T. S., & Azhar, Y. (2020). Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA JURNAL*, 10(02), 71–76. <https://doi.org/10.32664/smatika.v10i02.455>
- Istriani, E. (2023). PKM Pendampingan Manajemen Sampah di RW 2 Pakualaman Yogyakarta. *Prosiding Seminar Nasional Pengabdian Kepada Masyarakat*, 27–33. <http://sendimas2023.ukrida.ac.id/pubs/6/pkm-pendampingan-manajemen-sampah-di-rw-2-pakualaman-yogyakarta>
- Laurensz, B. & Eko Sedyono. (2021). Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 10(2), 118–123. <https://doi.org/10.22146/jnteti.v10i2.1421>
- Nasution, M. R. A., & Hayaty, M. (2019). Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter. *Jurnal Informatika*, 6(2), 226–235. <https://doi.org/10.31311/ji.v6i2.5129>
- Nurdin, A., Anggo Seno Aji, B., Bustamin, A., & Abidin, Z. (2020). PERBANDINGAN KINERJA WORD EMBEDDING WORD2VEC, GLOVE, DAN FASTTEXT PADA KLASIFIKASI TEKS. *Jurnal Tekno Kompak*, 14(2), 74. <https://doi.org/10.33365/jtk.v14i2.732>
- Pradana, A. I., Rustad, S., Shidik, G. F., & Agus Santoso, H. (2022). Indonesian Traffic Signs Recognition Using Convolutional Neural Network. *2022 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 426–430. <https://doi.org/10.1109/iSemantic55962.2022.9920448>
- Putra A P & Syafira A F. (2023). Analisis Sentimen Data Twitter Topik Politik Dengan Metode Naive Bayes Dan Convolutional Neural Networks (Cnn). <https://doi.org/10.5281/ZENODO.8396579>
- Rachman, F. F., & Pramana, S. (2020). Analisis Sentimen Pro dan Kontra Masyarakat Indonesia tentang Vaksin COVID-19 pada Media Sosial Twitter. *Indonesian of Health Information Management Journal*, 8(2), 100–109.
- Ratnaningtyas, Rr. P. (2020). Sampah Dalam Kacamata Media Online. *Jurnal Komunikasi*, 12(1), 16. <https://doi.org/10.24912/jk.v12i1.5287>
- Rifai, W., & Winarko, E. (2019). Modification of Stemming Algorithm Using A Non Deterministic Approach To Indonesian Text. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(4), 379. <https://doi.org/10.22146/ijccs.49072>
- Sawicki, J., Ganzha, M., & Paprzycki, M. (2023). The State of the Art of Natural Language Processing—A Systematic Automated Review of NLP Literature Using NLP Techniques. *Data Intelligence*, 5(3), 707–749. https://doi.org/10.1162/dint_a_00213
- Susilo, M. E., Prayudi, P., & Florestyanto, M. Y. (2023, Oktober). PENGELOLAAN SAMPAH RUMAH TANGGA UNTUK MEMBANTU MENGATASI KRISIS SAMPAH DI YOGYAKARTA. *Prosiding Seminar Nasional LPPM UPN Veteran Yogyakarta*. Seminar Nasional Pengabdian Masyarakat “Pemberdayaan Masyarakat Berkelanjutan di Era Society 5.0 sebagai Implementasi Bela Negara,” Yogyakarta.
- Toraman, C., Yilmaz, E. H., Şahinuç, F., & Özcelik, O. (2023). Impact of Tokenization on Language Models: An Analysis for Turkish. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(4), 1–21. <https://doi.org/10.1145/3578707>



TIN: Terapan Informatika Nusantara

Vol 4, No 10, March 2024, page 679-686

ISSN 2722-7987 (Media Online)

Website <https://ejournal.seminar-id.com/index.php/tin>

DOI 10.47065/tin.v4i10.4952

- Yuliska, Y., Qudsi, D. H., Lubis, J. H., Syaliman, K. U., & Najwa, N. F. (2021). Analisis Sentimen pada Data Saran Mahasiswa Terhadap Kinerja Departemen di Perguruan Tinggi Menggunakan Convolutional Neural Network. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 8(5), 1067. <https://doi.org/10.25126/jtiik.2021854842>
- Yuliska, Y., & Syaliman, K. U. (2020). Literatur Review Terhadap Metode, Aplikasi dan Dataset Peringkasan Dokumen Teks Otomatis untuk Teks Berbahasa Indonesia. *IT Journal Research and Development*, 5(1), 19–31. [https://doi.org/10.25299/itjrd.2020.vol5\(1\).4688](https://doi.org/10.25299/itjrd.2020.vol5(1).4688)