

Analisis Sentimen Pilkada di Tengah Pandemi Covid-19 Menggunakan *Convolution Neural Network (CNN)*

Sukma Nindi Listyarini^{*1}, Dimas Aryo Anggoro²

^{1,2} Informatika, Fakultas Komunikasi dan Informatika, Universitas Muhammadiyah Surakarta, Indonesia
Email: ¹sukmanindi30@gmail.com, ²Dimas.a.anggoro@ums.ac.id

Abstrak

Pemilihan kepala daerah 2020 menjadi kontroversi, sebab dilaksanakan ditengah pandemi covid-19. Komentar muncul di berbagai lini media sosial seperti twitter. Banyak masyarakat yang setuju pilkada dilanjutkan, namun banyak juga yang berpendapat untuk menunda pilkada sampai masa pandemi berakhir. Melihat perbedaan pendapat seperti ini, perlu dilakukan analisis sentimen, dengan tujuan untuk memperoleh persepsi atau gambaran umum masyarakat terhadap penyelenggaraan pilkada 2020 saat pandemi covid-19. Sebanyak 500 tweet diperoleh dengan cara *crawling* data dari twitter API menggunakan *library tweepy*, berdasarkan *keyword* yang telah ditentukan. Dataset yang didapat diberi label ke dalam dua kelas, negatif dan positif. Penelitian ini mengusulkan pendekatan *deep learning* dengan algoritma *Convolution Neural Network (CNN)* untuk klasifikasi, yang terbukti efektif untuk tugas *Natural Language Processing (NLP)* dan mampu mencapai kinerja yang baik dalam klasifikasi kalimat. Percobaan dilakukan dengan menerapkan 4-layer *convolutional* dan mengamati pengaruh jumlah *epoch* terhadap akurasi model. Variasi *epoch* yang digunakan adalah 50, 75, 100. Hasil dari penelitian menunjukkan bahwa, metode CNN dengan dataset pilkada ditengah pandemi mendapatkan akurasi tertinggi sebesar 90% dengan 4-layer *convolutional* dan 100 *epoch*. Didapatkan pula bahwa, semakin banyak *epoch* yang digunakan dalam model, akurasi cenderung meningkat.

Kata kunci: *analisis sentimen, covid-19, cnn, deep learning, NLP*

Sentiment Analysis Regional Elections During Pandemic Covid-19 Using Convolution Neural Network (CNN)

Abstract

The 2020 regional head election is controversial, because it was held during the Covid-19 pandemic which was endemic in Indonesia. Comments appear on various social media such as twitter. Many people agree that the elections will be continued, but there are also many who think that they should postpone the elections until the pandemic period ends. Seeing differences of opinion like this, it is necessary to carry out sentiment classification or sentiment analysis, with the aim of obtaining public perceptions or public opinions of the 2020 regional elections during the Covid-19 pandemic. Total 500 tweets were obtained by crawling data from the twitter API using the tweepy library, based on predetermined keywords. The data sets obtained were labeled into two classes, negative and positive. This study developed a deep learning approach with the Convolution Neural Network (CNN) algorithm for classification, proven to be effective for Natural language processing (NLP) tasks and able to achieve good performance in sentence classification. The experiment was implement by applying 4 convolutional layers and observe the effect of total epochs on the model accuracy. The epoch variations used are 50, 75, 100. The results showed that the CNN method with the local election dataset in the middle of the pandemic received the highest accuracy of 90% with 4-layers convolutional and 100 epoch. It was also found that, the more epochs used in the model, the accuracy tends to increase.

Keywords: *covid-19, cnn, deep learning, NLP, sentiment analysis*

1. PENDAHULUAN

Pelaksanaan pilkada atau pemilihan kepala daerah 2020 menjadi kontroversi, sebab dilaksanakan dimasa pandemi covid-19 yang melanda Indonesia. Dimana Kementerian Kesehatan telah mengeluarkan peraturan pemberlakuan PSBB (pembatasan sosial berskala besar) di seluruh Indonesia [1]. Meskipun demikian, KPU tetap menjadwalkan ulang pelaksanaan pilkada hingga akhir tahun 2020. KPU yakin dengan peraturan KPU

no.13 tahun 2020, pilkada 2020 akan tetap dilaksanakan dengan mengutamakan protokol kesehatan pada setiap tahapan pilkada.

Menanggapi dua kondisi yang sama pentingnya, komentar muncul diberbagai lini media sosial seperti twitter. Kekhawatiran netizen (*citizen of the net*) muncul apabila pilkada serentak 2020 tetap dilaksanakan, kemungkinan covid-19 semakin tidak dapat dikendalikan. Pilkada ini akan menjadi ajang penyebaran dan penularan covid-19 secara besar-besaran. Bagaimanapun juga, tidak dapat diperkirakan kapan berakhirnya pandemi covid-19, sedangkan pesta demokrasi pilkada juga penting dilaksanakan untuk merangsang pembangunan dan kesejahteraan bagi masyarakat daerah [2].

Melihat kondisi seperti ini, perlu dilakukan suatu analisis sentimen atau dikenal juga sebagai *opinion mining*[3]. Analisis sentiment dilakukan dengan tujuan untuk memperoleh persepsi atau gambaran umum masyarakat terhadap penyelenggaraan pilkada di masa pandemi covid. Apakah masyarakat setuju dengan pelaksanaan pilkada (sentimen positif) atau tidak setuju dilaksanakannya pilkada 2020 (sentimen negatif). Sehingga hasil analisis dapat dijadikan landasan bagi pemerintah untuk mengambil keputusan terkait pelaksanaan pikada.

Analisis sentimen merupakan proses mengekstraksi dan mengidentifikasi opini atau informasi subjektif pada sebuah teks. Analisis sentimen dapat menentukan sikap penulis terhadap suatu entitas, apakah positif, negatif atau netral [4]. Analisis sentimen melibatkan komputasi dalam memahami sentimen tersirat di dalam sebuah teks. *Natural language processing* (NLP) dikenal sebagai komputasi linguistik yang mampu menyelesaikan masalah praktis dalam memahami bahasa manusia [5]. NLP memungkinkan mesin memproses dan menerjemahkan bahasa alami manusia ke dalam format yang dapat dimengerti oleh mesin[6]. Penelitian oleh Cui dkk [7] membuktikan bahwa penerapan algoritma *Convolution Neural Network* (CNN) efektif untuk tugas NLP dan mampu mencapai kinerja yang baik dalam klasifikasi kalimat.

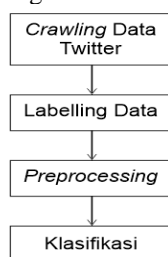
Banyak penelitian mengangkat tema analisis sentimen menggunakan NLP dan penerapan algoritma CNN sebagai klasifikasi. Penelitian yang dilakukan oleh Guan dkk [8] mengusulkan gabungan antara *Weakly-supervised Deep Embedding* (WDE) dan CNN untuk klasifikasi sentimen tingkat kalimat. Penelitian Poria dkk [9] membangun model CNN untuk mendeteksi sarkastik pada twitter. Penelitian oleh Rhanoui dkk [10] menjelaskan bahwa model CNN sangat kuat dalam ekstraksi fitur dan cocok untuk analisis opini artikel panjang. Penelitian lain oleh Rehman dkk [4] menjelaskan bahwa model CNN yang menggunakan *convolutional layer* dan *max pooling layer* secara efisien dapat mengekstrak fitur lebih tinggi.

Bedasarkan studi literatur yang telah dilakukan, masih sedikit yang meneliti pengaruh banyaknya *layer convolutional* dan banyaknya *epoch* yang diterapkan di dalam model CNN. *Epoch* merupakan *step* dalam proses *training* CNN, dimana seluruh dataset telah melewati *forward* dan *backward propagation* dalam satu kali putaran [11]. *Epoch* kecil dapat menghasilkan kinerja yang buruk karena penyesuaian parameter yang tidak lengkap, sementara *epoch* besar mungkin meningkatkan waktu komputasi serta risiko *overfitting* [12].

Untuk itu, penelitian ini mengusulkan melakukan analisis sentimen pilkada ditengah pandemi covid-19 menggunakan teknik NLP untuk mengolah dan mengekstrak data dari twitter. Hasil *preprocessing* berupa token-token kata selanjutnya diklasifikasikan dengan pendekatan *deep learning* menggunakan algoritma *Convolution Neural Network* (CNN). Algoritma CNN dipilih karena memiliki kemampuan ekstraksi fitur yang tinggi. Hasil penelitian ini berupa perbandingan akurasi dari penerapan 4-layer konvolusi dan pengaruh banyaknya *epoch* yang digunakan dalam model, serta hasil label prediksi pada uji kemampuan terhadap input kalimat baru.

2. METODE PENELITIAN

Persoalan utama dalam klasifikasi sentimen adalah cara mengidentifikasi sentimen dalam sebuah teks, apakah teks memiliki sentimen positif atau negatif. Gambar 1 merupakan alur yang diterapkan pada penelitian ini [13]. Langkah pertama, pengambilan data atau *crawling* data dari twitter, selanjutnya seluruh data diberi label ke dalam dua kelas, positif dan negatif. Langkah selanjutnya *preprocessing* menggunakan teknik NLP. Hasil dari *preprocessing* berupa token-token akan menjadi input pada proses klasifikasi. Proses klasifikasi menggunakan pendekatan *deep learning* dengan algoritma *Convolution Neural Network* (CNN).



Gambar 1. Langkah Penelitian

2.1. Crawling Data

Data diambil dari twitter menggunakan bahasa pemrograman *Python*, dengan *library tweepy* untuk menghubungkan ke Twitter API. Sebanyak 500 tweet dipilih dengan kata kunci penelusuran diantaranya “pilkada”, “pemilu”, ”corona”, “pandemi” dan “covid-19”. Contoh hasil *crawling* data twitter ditunjukkan pada tabel 1, tweet tersebut dipilih secara acak, merujuk pada kondisi pilkada di masa pandemi covid-19.

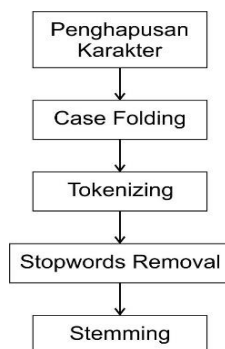
Tabel 1. Contoh Hasil *Crawling* Data Twitter

Tanggal	Tweet
2020-10-11	RT @masniamar: Udah lah gak usah paksakan Pilkada , ini tuh bahaya banget bisa” muncul cluster baru https://t.co/SbeiPhnLGJa..
2020-10-11	RT @Sadasuka: 18) Pandemi tak bisa dijadikan alasan untuk menunda hajatan lima tahunan yang sudah berjalan dan diagendakan lama. https://t...
2020-10-15	RT @victorastuti: Masyarakat antusias dan optimis Pilkada berjalan lancar. Sosialisasi melaksanakan disiplin protokol kesehatan ketat
2020-11-26	Patuhi protokol kesehatan saat pilkada stop penyebaran virus corona https://t.co/0uCZE6i5Oh
2020-10-11	Ntar pilkada jangan datang, biar aja mereka yang datang ke rumah. Alasannya covid aja kan namanya juga perwakilan\n#Pilkada2020

2.2. Labelling Data

Pemberian label pada dataset dilakukan secara manual dengan bantuan pakar. Data diberi label positif dan negatif. Dari hasil *labelling* didapatkan sebanyak 250 tweet sentimen negatif dan 250 tweet dengan sentimen positif.

2.3. Pre-processing



Gambar 2. Langkah *Preprocessing*

Data yang diperoleh dari twitter umumnya berupa kalimat yang tidak terstruktur contoh pada tabel 1, seperti terdapat karakter non alfanumerik, *link url*, tata bahasa tidak teratur, dll. Adanya *noise* dan ketidakteraturan data memengaruhi kinerja mesin [14]. *Preprocessing* NLP adalah langkah penghapusan *noise* yang tidak diinginkan dari dataset twitter [15], sehingga data yang dihasilkan menjadi lebih terstruktur. Penelitian oleh [16] mempelajari dampak *preprocessing* pada analisis sentimen, hasilnya terbukti bahwa proses *preprocessing* membantu dalam meningkatkan kinerja klasifikasi. Langkah *preprocessing* [17] ditunjukkan pada gambar 2.

2.3.1. Penghapusan Karakter

Dataset yang telah diberi label selanjutnya dibersihkan dari karakter yang tidak relevan seperti link url, *hashtag*, *username*, tanda baca dan karakter non alfanumerik (kecuali spasi) seperti ! @ # \$ & % dll. Karakter tersebut sebagian besar dapat memengaruhi akurasi proses analisis, karena bukan termasuk teks yang dapat ditemukan di kamus [18]. Penghapusan karakter menggunakan *library regular expression* milik *python*.

2.3.2. Case Folding

Proses menyamaratakan seluruh kata, dengan mengkonversi huruf kapital di dataset menjadi huruf kecil [19].

2.3.3. Tokenization

Suatu proses pemotongan atau penguraian teks menjadi per kata. Hasil dari proses ini disebut token [20]. Sebagai contoh proses tokenisasi kalimat terdapat pada tabel 2.

Tabel 2. Proses Tokenisasi

Teks Asli	ayo pilkada jangan takut corona
Setelah Tokenisasi	<ol style="list-style-type: none"> 1. 'ayo', 2. 'pilkada', 3. 'jangan', 4. 'takut', 5. 'corona'

2.3.4. Stopwords Removal

Penerapan *stopword* bertujuan untuk menghapus kata-kata yang dianggap tidak mempunyai arti penting, yang kemungkinan berpengaruh terhadap kecepatan dan kinerja proses analisis [21]. Proses *stopwords* dalam penelitian ini menggunakan *library python* sastrawi. Beberapa kata pada *stoplist python* sastrawi adalah 'yang', 'untuk', 'pada', 'dan', 'ke', 'dari', dll.

2.3.5. Stemming

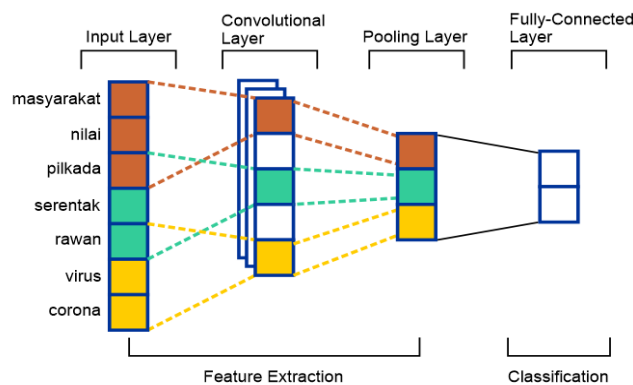
Merupakan langkah menghilangkan imbuhan berupa awalan, akhiran maupun campuran (awalan-akhiran) menjadi bentuk kata dasar sesuai kaidah bahasa Indonesia [22]. Proses *stemming* ini menggunakan *library python* sastrawi, sebuah *library* sederhana yang didesain untuk memperoleh kata dasar bahasa Indonesia dengan mudah. Pada tabel 3 terdapat contoh proses *stemming* pada kata berimbuhan.

Tabel 3. Contoh Proses *Stemming*

Sebelum di <i>Stemming</i>	Sesudah di <i>Stemming</i>
<ol style="list-style-type: none"> 1. Mengatur 2. Diatur 3. Teratur 	Atur
<ol style="list-style-type: none"> 1. Bayangkan 2. Membayangkan 3. Bayangan 	Bayang
<ol style="list-style-type: none"> 1. Berkerumun 2. Kerumunan 	Kerumun

2.4. Klasifikasi

Setelah melewati tahap *Preprocessing* dari NLP, selanjutnya data diklasifikasikan dengan pendekatan *deep learning* menggunakan algoritma *Convolutional Neural Networks*. CNN merupakan bagian dari *Artificial Neural Network* (ANN) yang mampu mendeteksi informasi dengan akurasi tinggi [10]. Model CNN telah memecahkan permasalahan dalam pemrosesan gambar dan saat ini peneliti-peneliti telah mengembangkan CNN untuk NLP seperti analisis sentimen, klasifikasi polaritas emosional, *text summary*, dll [23].



Gambar 3. Arsitektur CNN

CNN merupakan jaringan multilayer atau biasa disebut *fully-connected layer*, oleh karena itu *output* dari satu layer akan menjadi input dari layer berikutnya. Umumnya terdiri dari *layer input*, *hidden layer* dan *layer output*[10]. Penelitian ini menggunakan CNN sederhana seperti gambar 3 yang dijelaskan oleh Pambudi dan Suprpto[24] pada penelitiannya, yang terdiri dari *convolutional layer*, *max pooling layer*, *dropout layer* dan *fully connected output layer*.

Pengaturan parameter dalam penelitian ini diterapkan sebanyak 300 *vector size*, 8 unit *convolutional* tiap layer dengan filter 16, 32, 64. *Optimizer* yang dipilih Adam [25] yang dianggap sebagai optimasi terbaik [26] dan aktivasi *Rectified Linear Unit* (ReLU) sebagai aktivasi non linier yang memiliki keuntungan kompleksitas rendah [17].

3. HASIL DAN PEMBAHASAN

Sebanyak 500 dataset twitter digunakan dalam penelitian ini. Data tersebut dikelompokkan dalam dua kategori sentimen, 0 dengan sentimen negatif dan 1 dengan sentimen positif. Pada tahapan *preprocessing*, dataset dibersihkan dari kebisingan, selanjutnya melalui proses tokenisasi dari 500 tweet dipecah menjadi 1.806 kata atau token. Pada proses *stopwords*, kata-kata yang dianggap tidak memiliki arti penting, yang terdapat pada *stoplist* Sastrawi dihapus. Jumlah kata berkurang menjadi 1.352 kata. Tabel 4 menunjukkan contoh dataset sebelum dan sesudah *preprocessing*.

Tabel 4. Contoh Proses *Preprocessing*

Sentimen	Sebelum <i>Preprocessing</i>	Sesudah <i>Preprocessing</i>
1	Patuhi protokol kesehatan saat pilkada stop penyebaran virus corona https://t.co/0uCZE6i5Oh }	patuh protokol sehat pilkada stop sebar virus corona
0	Corona itu bahaya, lebih baik pilkada ditunda saja!	corona bahaya lebih baik pilkada tunda
1	Mari menyukseskan Pilkada 2020!! #madepmantebmanetep #pilkadalancar #pilkada2020	sukses pilkada 2020

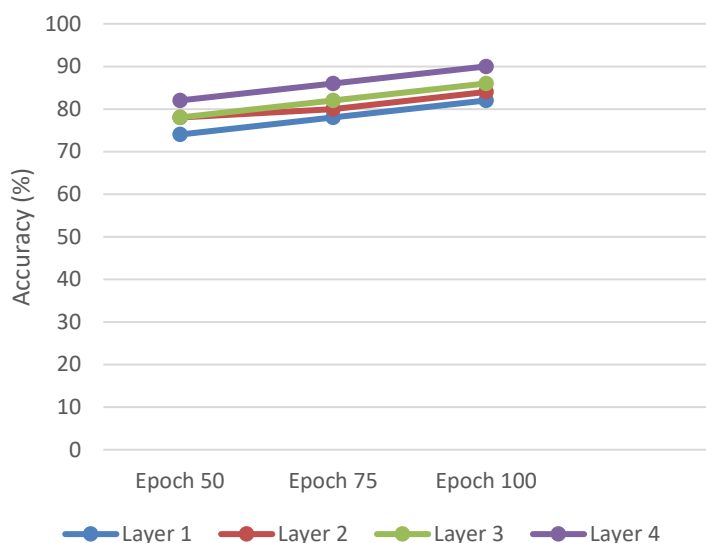
Kumpulan kata/token hasil *preprocessing* selanjutnya dijadikan input untuk proses klasifikasi. Klasifikasi pada penelitian ini mengusulkan algoritma CNN untuk membangun model dengan akurasi yang tinggi, dengan mengamati pengaruh kedalaman layer konvolusi dan pengaruh banyaknya *epoch* yang digunakan. Dataset penelitian dibagi menjadi tiga bagian, 70% data *training*, 20% data *validation* dan 10% data *testing* [27]. Pengamatan dilakukan pada 4-layer konvolusi dengan variasi *epoch* 50, 75, 100.

Tabel 5. Komparasi Akurasi 4-Layer Konvolusi Pada *Epoch* yang Berbeda

		<i>Epoch</i> 50	<i>Epoch</i> 75	<i>Epoch</i> 100
Layer 1	<i>Accuracy</i>	74%	78%	82%
	<i>Loss</i>	1.0542	0.7871	0.6157
Layer 2	<i>Accuracy</i>	78%	80%	84%
	<i>Loss</i>	0.6790	0.8444	1.1142
Layer 3	<i>Accuracy</i>	78%	82%	86%
	<i>Loss</i>	1.0848	1.0974	0.7902
Layer 4	<i>Accuracy</i>	82%	86%	90%
	<i>Loss</i>	1.3596	1.2395	0.7744

Tabel 5 menunjukkan hasil pengujian, akurasi tertinggi sebesar 90%, *loss* 0.7744 pada 4-layer konvolusi dengan 100 *epoch*. Berdasarkan tabel 5, seiring dengan peningkatan jumlah layer, akurasi yang didapatkan mengalami kenaikan. Peningkatan akurasi berkisar 2-4% dari akurasi layer sebelumnya. Meskipun, pada 2-layer dan 3-layer dengan *epoch* 50 memiliki akurasi yang sama, namun *loss validation* 2-layer lebih kecil daripada 3-layer.

Selanjutnya, penelitian ini mengamati pengaruh jumlah *epoch* yang diterapkan pada model. Gambar 4 menunjukkan grafik hasil akurasi dari variasi *epoch* yang diterapkan. Pada 1-layer konvolusi hasil akurasi sebesar 74%, 78%, 82%, pada 2-layer konvolusi hasil akurasi 78%, 80%, 84%, sampai 4-layer konvolusi hasil akurasi 82%, 86%, 90%.



Gambar 4. Pengaruh Epoch Pada Grafik Akurasi

Bedasarkan gambar 4, terbukti bahwa akurasi empat layer meningkat seiring bertambahnya epoch. Dari epoch 50, 75 sampai 100 mengalami kenaikan akurasi berkisar antara 2-4 % dari variasi epoch sebelumnya. Penerapan epoch yang berbeda pada proses training akan menghasilkan kemampuan klasifikasi yang berbeda. Pada epoch kecil, proses penyesuaian parameter masih belum selesai karena terbatas waktu iterasi, sehingga menghasilkan akurasi yang relatif rendah. Dengan peningkatan jumlah epoch, parameter dioptimalkan lebih lanjut, sehingga kinerja klasifikasi mengalami peningkatan secara bertahap [12].

Pengamatan selanjutnya yaitu uji klasifikasi pada kalimat baru yang tidak ada di dataset, untuk melihat apakah label prediksi sesuai dengan label aslinya. Hasil pengamatan tertera pada tabel 6.

Tabel 6. Perbandingan Label Prediksi Dan Label Asli

Twitter	Label Prediksi	Label Asli
Males banget mau berangkat nyoblos ke tps rawan corona	Negatif	Negatif
Dukung seluruh proses pilkada, tetap jaga jarak terapkan 3M	Positif	Positif
Mari kita sukseskan pilkada 2020	Positif	Positif

Pada tabel 6 model diuji dengan kalimat baru, hasil label prediksi pada kalimat baru sesuai dengan label aslinya. Namun, beberapa kalimat baru berlabel positif justru diklasifikasikan ke dalam sentimen negatif. Kesalahan klasifikasi tersebut dapat disebabkan oleh kalimat yang mengandung makna ambigu atau makna tersirat, penggunaan bahasa informal, ejaan tidak konsisten dan terdapat kemiripan kata dalam data training dan data testing [17].

4. KESIMPULAN

Hasil yang diperoleh dari penelitian analisis sentimen pilkada di tengah pandemi mendapatkan akurasi tertinggi sebesar 90% dengan 4-layer convolutional dan 100 epoch. Akurasi dari keempat layer meningkat seiring bertambahnya jumlah epoch yang diterapkan. Dari epoch 50, 75 sampai 100 mengalami kenaikan akurasi berkisar antara 2-4% dari variasi epoch sebelumnya. Hasil uji klasifikasi kalimat baru berhasil memberikan label prediksi sesuai dengan label aslinya. Namun terdapat beberapa kesalahan klasifikasi yang disebabkan oleh beberapa faktor, diantaranya terdapat kata dengan makna ambigu atau makna tersirat, penggunaan bahasa informal, ejaan tidak konsisten serta terdapat kemiripan kata dalam data training dan data testing. Penelitian kedepan disarankan menambahkan dataset dan menggabungkan beberapa metode lain seperti LSTM untuk meningkatkan akurasi.

DAFTAR PUSTAKA

[1] M. Rudiyanto and R. Kasanova, "Dampak Covid-19 Terhadap Implementasi Pendidikan Karakter Impact of Covid-19 on The Implementation of Character Education," *Jurnal Pendidikan dan Teknologi*

- Indonesia (JPTI)*, vol. 1, no. 5, pp. 209–213, 2021, doi: <https://doi.org/10.52436/1.jpti.43>.
- [2] S. C. Rizki and Y. A. Hilman, “Menakar Perbedaan Opini Dalam Agenda Pelaksanaan Kontestasi Pilkada Serentak Di Tengah Covid-19,” *J. Ilm. Muqoddimah*, vol. 4, no. 2, pp. 143–155, 2020, doi: <http://dx.doi.org/10.31604/jim.v4i2.2020.143-155>.
- [3] L. Zhang, S. Wang, and B. Liu, “Deep learning for sentiment analysis: A survey,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 2018, vol. 8, no. 4, pp. 1–25, doi: 10.1002/widm.1253.
- [4] A. U. Rehman, A. K. Malik, B. Raza, and W. Ali, “A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis,” *Multimed. Tools Appl.*, vol. 78, no. 18, pp. 26597–26613, 2019, doi: 10.1007/s11042-019-07788-7.
- [5] D. W. Otter, J. R. Medina, and J. K. Kalita, “A Survey of the Usages of Deep Learning for Natural Language Processing,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 2, pp. 604–624, 2021, doi: 10.1109/TNNLS.2020.2979670.
- [6] A. Rajput, “Natural Language Processing , Sentiment Analysis and Clinical Analytics,” 2019, https://www.researchgate.net/publication/330871275_Natural_Language_Processing_Sentiment_Analysis_and_Clinical_Analytics/citations, (accessed Apr. 10, 2021)
- [7] H. Cui, Y. Lin, and T. Utsuro, “Sentiment Analysis of Tweets by CNN utilizing Tweets with Emoji as Training Data,” *Proceedings of KDD (Wisdom'18)*, 2018, no. 4, pp. 1–8.
- [8] Z. Guan, L. Chen, W. Zhao, Y. Zheng, S. Tan, and D. Cai, “Weakly-supervised deep learning for customer review sentiment classification,” *IJCAI Int. Jt. Conf. Artif. Intell.* 2016, pp. 3719–3725.
- [9] S. Poria, E. Cambria, D. Hazarika, and P. Vij, “A deeper look into sarcastic tweets using deep convolutional neural networks,” *COLING 2016 - 26th Int. Conf. Comput. Linguist. Proc. COLING 2016 Tech. Pap.*, 2016, pp. 1601–1612.
- [10] M. Rhanoui, M. Mikram, S. Yousfi, and S. Barzali, “A CNN-BiLSTM Model for Document-Level Sentiment Analysis,” *Mach. Learn. Knowl. Extr.*, vol. 1, no. 3, pp. 832–847, 2019, doi: 10.3390/make1030048.
- [11] S. Sharma, “Epoch vs Batch Size vs Iterations,” 2017. <https://towardsdatascience.com/epoch-vs-iterations-vs-batch-size-4dfb9c7ce9c9> (accessed Apr. 15, 2021)
- [12] D. Wang *et al.*, “Modulation Format Recognition and OSNR Estimation Using CNN-Based Deep Learning,” *IEEE Photonics Technol. Lett.*, vol. 29, no. 19, pp. 1667–1670, 2017, doi: 10.1109/LPT.2017.2742553.
- [13] M. Khader, A. Awajan, and G. Al-Naymat, “The impact of natural language preprocessing on big data sentiment analysis,” *Int. Arab J. Inf. Technol.*, vol. 16, no. 3ASpecial Issue, pp. 506–513, 2019.
- [14] Z. Jianqiang and G. Xiaolin, “Deep Convolution Neural Networks for Twitter Sentiment Analysis,” *IEEE Access*, vol. 3536, no. c, 2017, doi: 10.1109/ACCESS.2017.2776930.
- [15] R. Rahutomo, A. Budiarto, K. Purwandari, A. S. Perbangsa, T. W. Cenggoro, and B. Pardamean, “Ten-year compilation of #savekpk twitter dataset,” *Proc. 2020 Int. Conf. Inf. Manag. Technol. ICIMTech 2020*, pp. 185–190, doi: 10.1109/ICIMTech50083.2020.9211246.
- [16] S. Alam and N. Yao, “The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis,” *Comput. Math. Organ. Theory*, vol. 25, no. 3, pp. 319–335, 2019, doi: 10.1007/s10588-018-9266-8.
- [17] D. A. N. Taradhita and I. K. G. D. Putra, “Hate speech classification in Indonesian language tweets by using convolutional neural network,” *J. ICT Res. Appl.*, vol. 14, no. 3, pp. 225–239, 2021, doi: 10.5614/itbj.ict.res.appl.2021.14.3.2.
- [18] W. Y. Chong, B. Selvaretnam, and L. K. Soon, “Natural Language Processing for Sentiment Analysis: An Exploratory Analysis on Tweets,” *Proc. - 2014 4th Int. Conf. Artif. Intell. with Appl. Eng. Technol. ICAIET 2014*, pp. 212–217, 2014, doi: 10.1109/ICAIET.2014.43.
- [19] M. K. Sohrabi and F. Hemmatian, “An efficient preprocessing method for supervised sentiment analysis by converting sentences to numerical vectors: a twitter case study,” *Multimed. Tools Appl.*, 2019, doi: 10.1007/s11042-019-7586-4.
- [20] A. Bayhaqy, S. Sfenrianto, K. Nainggolan, and E. R. Kaburuan, “Sentiment Analysis about E-Commerce from Tweets Using Decision Tree, K-Nearest Neighbor, and Naïve Bayes,” *2018 Int. Conf. Orange Technol. ICOT, 2018*, pp. 1–6, doi: 10.1109/ICOT.2018.8705796.
- [21] S. Anastasia and I. Budi, “Twitter sentiment analysis of online transportation service providers,” *2016 Int. Conference Adv. Comput. Sci. Inf. Syst. ICACSIS, 2016*, pp. 359–365.
- [22] T. B. Rohman, D. D. Purwanto, and J. Santoso, “Sentiment Analysis Terhadap Review Rumah Makan di Surabaya Memanfaatkan Algoritma Random Forest,” *Prosiding Seminar Nasional Teknologi Informasi & Aplikasinya*, 2018, pp. 7–11.
- [23] L. xia Luo, “Network text sentiment analysis method combining LDA text representation and GRU-

-
- CNN,” *Pers. Ubiquitous Comput.*, vol. 23, no. 3–4, pp. 405–412, 2019, doi: 10.1007/s00779-018-1183-9.
- [24] A. Pambudi and S. Suprpto, “Effect of Sentence Length in Sentiment Analysis Using Support Vector Machine and Convolutional Neural Network Method,” *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 15, no. 1, p. 21, 2021, doi: 10.22146/ijccs.61627.
- [25] J. Chen, S. Yan, and K. C. Wong, “Verbal aggression detection on Twitter comments: convolutional neural network for short-text sentiment analysis,” *Neural Comput. Appl.*, vol. 32, no. 15, pp. 10809–10818, 2020, doi: 10.1007/s00521-018-3442-0.
- [26] N. Purnama, P. K. Negara, and S. Primakara, “Deteksi Masker Pencegahan Covid19 Menggunakan Convolutional Neural,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 1, no. 10, pp. 576–583, 2021.
- [27] W. Supriyanti and D. A. Anggoro, “Classification of Pandavas Figure in Shadow Puppet Images using Convolutional Neural Networks,” *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 7, no. 1, 2021, doi: <https://doi.org/10.23917/khif.v7i1.12484>.