

DETEKSI UJARAN ANCAMAN BERBASIS WEBSITE PADA POSTINGAN MEDIA SOSIAL TWITTER MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK

WEBSITE BASED DETECTION OF THREATS IN SOCIAL MEDIA TWITTER USING CONVOLUTIONAL NEURAL NETWORK

Azhar Eka Mulia Wiguna¹, Muhammad Nasrun S.Si, M.T.² Ratna Astuti Nugrahaeni S.T, M.T.³

^{1,2,3}Prodi S1 Teknik Komputer, Fakultas Teknik Elektro, Universitas Telkom

¹azharekamulia@telkomuniversity.ac.id, ²muhammadnasrun@telkomuniversity.ac.id,

³ratnaan@telkomuniversity.ac.id

Abstrak

Di era teknologi zaman sekarang media sosial sangat penting bagi kehidupan manusia. Twitter merupakan media sosial yang sering digunakan oleh masyarakat khususnya di Indonesia. Twitter memiliki fungsi untuk memposting kalimat yang diunggah oleh penggunanya. Ada banyak macam postingan yang disampaikan oleh komunitas maupun masyarakat, ada yang positif, dan ada juga yang negatif. Pihak berwenang kesulitan menangani ujaran ancaman yang ada karena banyak variasi dari ujaran ancaman tersebut. Oleh karena itu sistem yang dibuat yaitu mendeteksi ujaran ancaman pada postingan media sosial twitter dengan metode klasifikasi *Convolutional Neural Network*. Dengan adanya sistem ini, diharapkan dapat membantu pihak yang berwenang untuk menangani kasus ujaran ancaman khususnya pada postingan Twitter. Dari hasil penelitian tugas akhir ini dalam mengklasifikasikan ujaran ancaman pada postingan Twitter mendapatkan *precision* sebesar 81%, *recall* sebesar 78%, *f-1 score* sebesar 79%, dan *accuracy* sebesar 80.63%

Kata kunci : Media Sosial, *Convolutional Neural Network*, Ujaran Ancaman, Twitter.

Abstract

In today's technological era, social media is very important for human life. Twitter is a social media that is often used by people, especially in Indonesia. Twitter has a function to post sentences uploaded by its users. There are many kinds of posts submitted by the community and society, some are positive, and some are negative. The authorities have difficulty dealing with existing threat utterances because of the wide variety of threat phrases. Therefore the system created is to detect threat utterances in Twitter social media posts using the Convolutional Neural Network classification method. With this system, it is hoped that it can help the authorities to handle cases of threatening speech, especially on Twitter posts. From the results of this research, in classifying threat utterances in Twitter posts, get 81% precision, 78% recall, 79% f-1 score, and 80.63% accuracy.

Keywords: Social media, *Convolutional Neural Network*, Threat Speech, Twitter.

1. Pendahuluan [10 pts/Bold]

Di Dalam beberapa tahun belakang ini, media sosial menjadi bagian dalam kehidupan sehari-hari. Kebanyakan orang sangat tidak asing dengan media sosial twitter, facebook, ataupun media sosial lainnya. Sosial media banyak digunakan untuk berkomunikasi, mencari informasi, iklan, dan acara sosial. Banyak orang di media sosial seperti twitter memposting kata dengan bebas dan banyak juga yang berupa ujaran ancaman. Sosial media merupakan faktor penting kemajuan internet di indonesia sendiri, berdasarkan survei yang dilakukan global web index, Indonesia adalah Negara yang memiliki pengguna sosial media yang paling aktif di asia. Indonesia memiliki 79,7% user aktif di social media mengalahkan Filipina 78%, Malaysia 72%, Cina 67% pengguna mendapat banyak celah untuk melakukan ujaran ancaman tersebut.

Berdasarkan permasalahan yang dijabarkan diatas, penulis akan membuat sistem aplikasi website yang berfokus pada perancangan dengan metode Convolutional Neural Network untuk mendeteksi ujaran ancaman yang terdapat pada media sosial. Pada perancangan ini penulis akan menggunakan metode dari machine learning yaitu Convolutional Neural Network yang termasuk

juga dalam *Convolutional Neural Network*. Metode *Convolutional Neural Network* sendiri dipilih oleh penulis karena memberikan performansi dan kedalaman jaringan yang tinggi, sehingga banyak diaplikasikan pada data citra dengan menghasilkan hasil yang baik. Dengan pembuatan sistem ini diharapkan dapat mendeteksi ujaran ancaman pada postingan di media sosial Twitter dan membantu pihak yang berwenang dalam menangani kasus ujaran ancaman yang marak di media sosial.

2. Dasar Teori

2.1 Ujaran Ancaman

Dalam KBBI (Kamus Besar Bahasa Indonesia) Ujaran Ancaman (Threat Speech) adalah perkataan, perilaku, tulisan, ataupun pertunjukan yang dilarang karena dapat memicu terjadinya tindakan kekerasan dan sikap prasangka entah dari pihak pelaku pernyataan tersebut ataupun korban dari tindakan tersebut, dan memberi pertanda atau peringatan mengenai kemungkinan malapetaka yang bakal terjadi. Kebanyakan dari pelaku ujaran ancaman melakukan tindakannya melalui media sosial seperti twitter, dengan memanfaatkan fitur postingan ataupun video situs ini menggunakan Forum Internet dan Berita untuk mempertegas suatu sudut pandangan tertentu.

2.2 Twitter

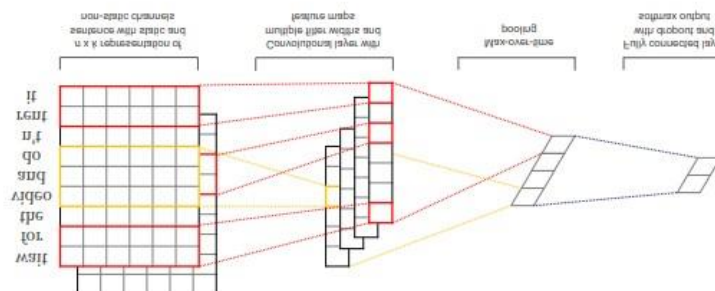
Twitter adalah salah satu media sosial yang sangat populer di Indonesia yang didirikan oleh Jack Dorsey yang mulai beroperasi pada juli tahun 2006. Melalui Twitter, seseorang dapat membagikan postingan berupa kalimat, foto, video, maupun menandakan kepada teman maupun pengguna twitter lainnya yang memiliki akun Twitter. Pada Twitter memiliki fitur retweet, favorite, dan reply sehingga pengguna dapat dengan mudah menyampaikan pendapatnya tentang postingan kalimat, foto, dan video yang dibagikan. Hal ini membuat Twitter menjadi salah satu media sosial dengan penyampaian ujaran ancaman terbanyak di Indonesia[1].

2.3 Keterangan Tabel dan Gambar

Convolutional Neural Network (CNN) merupakan pengembangan dari Multilayer Perceptron (MLP) yang dirancang untuk mengolah data dua dimensi. CNN termasuk dalam jenis deep neural network karena kedalaman jaringannya yang tinggi dan banyak digunakan pada data gambar maupun teks. CNN dirancang khusus untuk mengenali bentuk 2 dimensi, dengan tingkat inversi tinggi, terhadap translation, scaling dan distorsi lainnya. CNN termasuk dalam jenis deep neural network karena Jaringan memiliki kedalaman yang tinggi dan banyak digunakan pada data citra. Convolutional Neural Network biasanya digunakan untuk klasifikasi citra atau deteksi objek. Dalam Convolutional Neural Network terbagi menjadi dua bagian utama, yaitu fitur extraction layer dan Fully connected layer. fitur extraction terdiri dari convolutional layer dan pooling layer[4].

2.3.3 Konsep CNN

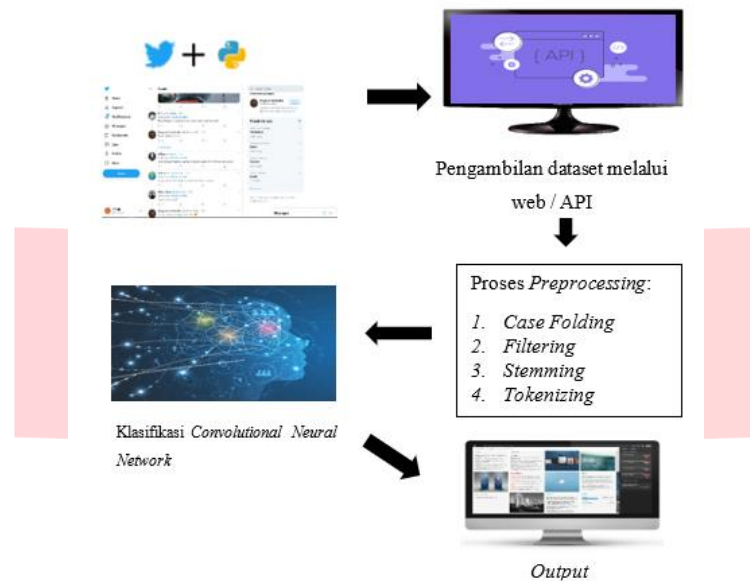
Dalam Convolutional Neural Network terbagi menjadi dua bagian utama, yaitu fitur fully connected layer dan fitur extraction layer. Fitur extraction layer terdiri dari convolutional layer dan pooling layer. Arsitektur Convolutional Neural Network adalah sebagai berikut.



Gambar 2. 1 Arsitektur Convolutional Neural Network 1D

3. Pembahasan

3.1. Gambaran Umum Sistem



Gambar 3. 1 Gambaran Umum Sistem

Pada Gambar 3.1 menunjukkan proses deteksi ujaran ancaman. Studi kasus yang akan dibahas di penelitian ini adalah mendeteksi ujaran ancaman pada kolom postingan Twitter yang berbahasa Indonesia. Input yang diberikan adalah postingan dalam bahasa Indonesia, dan output dari sistem adalah muncul tidaknya postingan dalam bentuk ujaran ancaman.. *Dataset* diperoleh dari kolom postingan Twitter. *Dataset* yang telah diambil langsung masuk pada UI yang telah di rancang. Setelah itu, data akan diolah dan diindikasikan apakah merupakan pernyataan ancaman atau bukan.

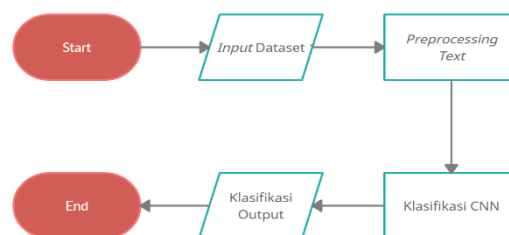
3.2 Text Preprocessing

Pada penelitian yang dilakukan, penulis menggunakan dataset berupa postingan Twitter berbahasa Indonesia yang dikumpulkan melalui beberapa akun pada aplikasi Twitter. Postingan yang diambil akan disimpan dalam bentuk (.csv). Terdapat 4 tahap dalam text pre-processing yaitu:

1. *Case Folding* merupakan tahap untuk proses mengubah huruf capital menjadi huruf kecil.
2. *Filtering* adalah proses pembuangan kata-kata yang tidak relevan untuk di proses selanjutnya, berupa tanda hubung, kata-kata yang tidak baku, dan lainnya. Contohnya imbuhan meng-, me-, kan-, di- dll.
3. Tahap selanjutnya yaitu tahap *stemming*. Tahap ini yaitu mengubah kata-kata yang ada menjadi kata dasar.
4. *Tokenizing* adalah proses untuk memotong kalimat atau string menjadi beberapa bagian yang biasa disebut dengan token. Pada saat yang bersamaan biasa membuang karakter tertentu, seperti tanda baca

3.3 Text Preprocessing

Berikut merupakan *flowchart* klasifikasi dengan metode convolutional neural network:



Gambar 3. 2 Flowchart Sistem dengan Metode Convolutional Neural Network

Pada gambar 3.10 dapat dilihat proses sistem yang telah dibuat. Data masukan merupakan postingan Twitter yang selanjutnya di kondisikan dalam kasus ini yaitu di *preprocessing* dan telah diberi label. Setelah itu data yang telah di *preprocessing* dibagi menjadi dua yaitu data latih dan data uji. Kemudian melewati proses klasifikasi *convolutional neural network*. Setelah itu data latih yang telah di proses menjadi model *training* sedangkan data uji akan diuji berdasarkan model *training* dan menjadi *output*. Lalu dilanjutkan dengan perhitungan CNN dengan *word index*

3.4 Perhitungan Recall, Precision, dan Accuracy

Untuk mengetahui kualitas sistem dalam data mining, maka eksperimen diperlukan untuk mengukur seberapa baik dalam penggunaan system tersebut. Percobaan meliputi perhitungan precision, recall, dan accuracy yang dapat dihitung menggunakan bantuan tabel confusion matriks[10].

Precision adalah adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh system. Menghitung nilai precision:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Menghitung nilai recall :

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

Accuracy diartikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai actual. Menghitung nilai accuracy:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (3)$$

F1-Score merupakan evaluasi yang terdiri dari gabungan antar precision dan recall. Menghitung F1-Score

$$F1 \text{ score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

Keterangan:

TP adalah Data yang diklasifikasikan sebagai klasifikasi ujaran ancaman.
TN adalah Data yang bukan ujaran ancaman tetapi system mengklasifikasikan sebagai ujaran ancaman.

FP adalah Data yang merupakan ujaran ancaman tetapi sistem mengklasifikasikan bukan ujaran ancaman.

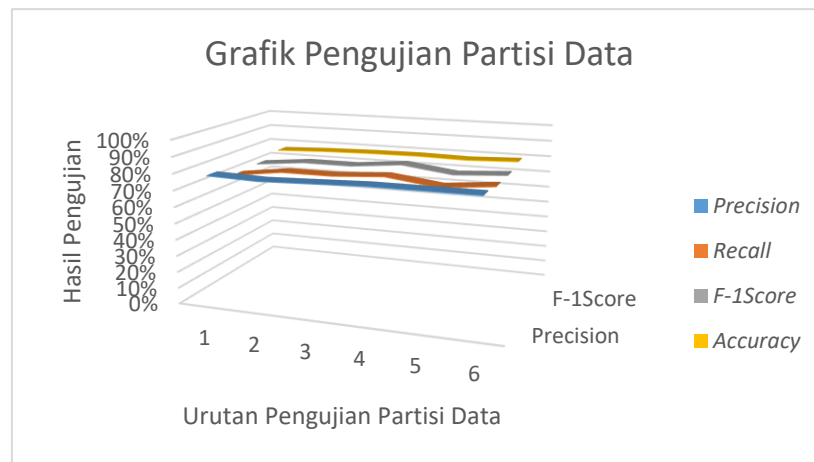
FN adalah Data yang diklasifikasikan bukan ujaran ancaman.

4. Implementasi dan Pengujian Sistem

Pengujian ini berfungsi untuk mengetahui performansi dari penggunaan metode *convolutional neural network* yang telah dibuat. Kemudian dilakukan pengujian dengan parameter yang ada. Parameter tersebut antara lain partisi data uji dan data latih, nilai *learning rate*, dan nilai *epoch*.

Table 4. 1 Rangkuman Pengujian Partisi Data

Pengujian ke-	Data Latih (%)	Data Uji (%)	Precision (%)	Recall (%)	F-1 Score (%)	Accuracy (%)
1	50	50	78	72	73	76.56
2	55	45	77	76	76	77.75
3	60	40	78	76	76	78.45
4	70	30	79	78	79	78.74
5	80	20	79	74	75	78.15
6	90	10	79	77	77	79.34

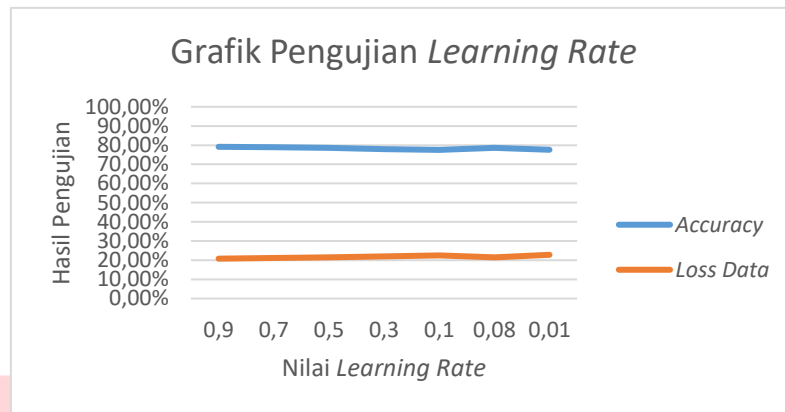


Gambar 4. 1 Grafik Partisi Data

Pada gambar 4.1 dapat dilihat bahwa pengujian keenam memperoleh tingkat accuracy tertinggi yaitu 79.34% yang dimana nantinya partisi tersebut akan dipakai untuk pengujian selanjutnya. Karena semakin banyak data latih maka semakin bagus akurasi sistem yang diperoleh. Tetapi pada beberapa proses pelatihan contohnya pada pengujian kelima mengalami penurunan accuracy, lalu naik kembali pada pengujian keenam. Lalu pada beberapa proses pelatihan sistem mengalami overfitting. Ini disebabkan dengan penggunaan learning rate yang cukup besar yang mengakibatkan proses pembelajaran terjadi dengan cepat, tetapi juga dapat menjadi tidak stabil atau bahkan tidak belajar sama sekali [14][15].

Table 4. 2 Rangkuman Pengujian Learning Rate

No.	Pengujian	Learning Rate	Epoch	Accuracy (%)	Loss Data
1.	Pertama	0.9	10	79.14	0.208
2.	Kedua	0.7		78.94	0.210
3.	Ketiga	0.5		78.55	0.214
4.	Keempat	0.3		77.95	0.220
5.	Kelima	0.1		77.45	0.225
6.	Keenam	0.08		78.55	0.214
7.	Ketujuh	0.01		77.55	0.228

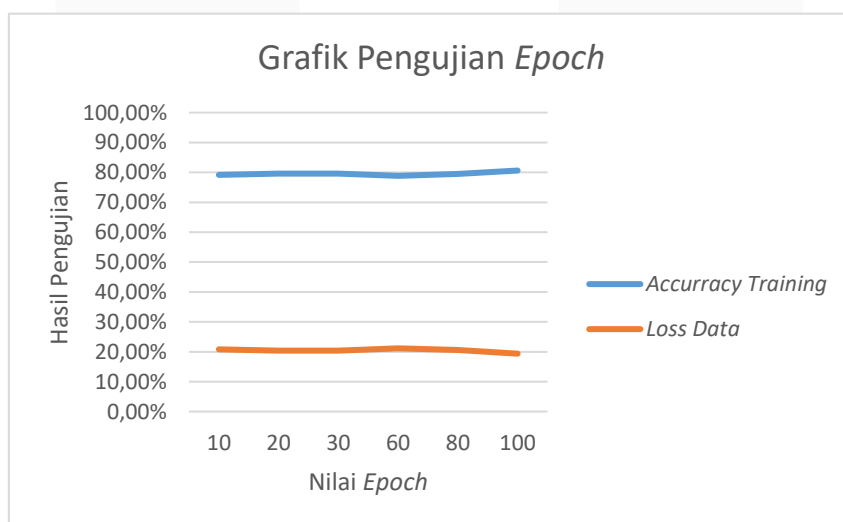


Gambar 4. 2 Grafik Pengujian Learning Rate

Pada gambar 4.4 dapat dilihat grafik hasil pengujian pada parameter *learning rate*. Dari hasil tersebut dapat diambil kesimpulan bahwa semakin kecil nilai *learning rate* maka semakin baik *accuracy* yang diperoleh[14][15]. Hal ini disebabkan karena penggunaan nilai *learning rate* untuk menentukan seberapa cepat sistem belajar dari data latih yang diperoleh. Tetapi jika terlalu kecil maka akan menyebabkan pelatihan menjadi lambat, terbukti pada gambar 4.4 *learning rate* 0.01 menghasilkan *loss data* yang lebih besar dibandingkan *learning rate* 0.9. Maka dari itu, *learning rate* 0.9 akan dipakai untuk pengujian selanjutnya.

Table 4. 3 Rangkuman Pengujian Epoch

No.	Pengujian	Epoch	Accuracy (%)	Loss Data
1.	Pertama	10	79.14	0.2085
2.	Kedua	20	79.64	0.2035
3.	Ketiga	30	79.64	0.2035
4.	Keempat	60	78.84	0.2115
5.	Kelima	80	79.44	0.2055
6.	Keenam	100	80.63	0.1936



Gambar 4.3 Grafik Pengujian Epoch

Pada gambar 4.5 dapat dilihat grafik hasil pengujian pada parameter *epoch*. Dari hasil tersebut dapat diambil kesimpulan bahwa semakin besar nilai *epoch* maka semakin mempengaruhi *accuracy* dan *loss data* yang dihasilkan. Maka nilai *epoch* 100 merupakan nilai terbaik pada pengujian ini yaitu menghasilkan *accuracy* sebesar 80,63% dan *loss data* sebesar 0.1936.

Dari hasil pengujian yang dilakukan dengan metode Convolutional Neural Network dengan 1007 dataset yang sudah divalidasi oleh Balai Bahasa Jawa Barat yang mengandung 60% data bukan ancaman dan 40% data ancaman bisa disimpulkan dengan hasil yang didapat dari tujuh kali pengujian yaitu pengujian partisi data keenam dengan 90% data *train* dan 10% data test memperoleh tingkat *accuracy* tertinggi yaitu 79.34% dan tidak terjadi overfitting karena tidak terjadi *high loss* dan *low accuracy*. Karena semakin

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pengujian dan analisa yang telah dilakukan pada tugas akhir ini, maka dapat ditarik kesimpulan bahwa:

1. Sistem mendeteksi ujaran ancaman berbasis website pada postingan media sosial twitter menggunakan metode Convolutional Neural Network berhasil mengklasifikasikan postingan Twitter berupa ujaran ancaman atau bukan ujaran ancaman.
2. Berdasarkan pengujian partisi data, semakin banyak data latih maka *accuracy* klasifikasi yang dihasilkan semakin bagus. Dalam hal ini partisi data yang terbaik pada proses pengujian yaitu 90% data latih sebanyak 906 data sedangkan 10% data uji yaitu sebanyak 101 data.
3. Nilai *learning rate* mempengaruhi kinerja sistem dan *accuracy* yang dihasilkan. Dalam kasus ini nilai *learning rate* terbaik pada proses pengujian yaitu sebesar 0.9.
4. Nilai *epoch* mempengaruhi kinerja sistem pada loss data. Dalam kasus ini nilai *epoch* terbaik pada proses pengujian yaitu sebesar 100 *epoch*, dimana memperoleh akurasi sebesar 80,63% dan loss data yang dihasilkan sebesar 0.1936.
5. Pada proses pengujian sistem diperoleh nilai rata-rata parameter *precision* sebesar 81%, *recall* sebesar 78%, *f-1 score* sebesar 79%, dan *Accuracy* sebesar 80.63%.

5.2 Saran

Berdasarkan hasil penelitian, pengujian dan analisa yang telah dilakukan pada tugas akhir ini, maka saran yang dapat diusulkan untuk penelitian lebih lanjut yaitu:

1. Implementasi sistem ujaran ancaman di media sosial lain seperti Instagram ataupun facebook.
2. Implementasi dataset yang digunakan harus balance.

Reference:

- [1] D. Dwiki Adriadi Nur and K. Kunci, "PENGEMBANGAN APLIKASI SENTIMENT ANALYSIS MENGGUNAKAN METODE NAÏVE BAYES (Studi Kasus Sentiment Analysis dari media Twitter)," *Semin. Nas. Sist. Inf. Indones.*, no. November, pp. 2–3, 2015.
- [2] M. Allahyari, E. D. Trippe, and J. B. Gutierrez, "A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques," 2017.
- [3] K. Kowsari, D. E. Brown, M. Heidarysafa, K. J. Meimandi, M. S. Gerber, and L. E. Barnes, "HDLTex: Hierarchical Deep Learning for Text Classification," pp. 364–371, 2017.
- [4] E. N. Arrofiqoh and H. Harintaka, "Implementasi Metode Convolutional Neural Network Untuk Klasifikasi Tanaman Pada Citra Resolusi Tinggi," *Geomatika*, vol. 24, no. 2, p. 61, 2018.
- [5] H. Yanagisawa, T. Yamashita, and H. Watanabe, "A study on object detection method from manga images using CNN," 2018 *Int. Work. Adv. Image Technol. IWAIT 2018*, pp. 1–4, 2018.
- [6] Y. Wang, J. Zhang, Y. Cao, and Z. Wang, "A deep CNN method for underwater image enhancement," *Proc. - Int. Conf. Image Process. ICIP*, vol. 2017-Sept, pp. 1382–1386, 2018.
- [7] W. S. Eka Putra, "Klasifikasi Citra Menggunakan Convolutional Neural Network (CNN) pada Caltech 101," *J. Tek. ITS*, vol. 5, no. 1, 2016.
- [8] S. Ilahiyah and A. Nilogiri, "Implementasi Deep Learning Pada Identifikasi Jenis Tumbuhan Berdasarkan Citra Daun Menggunakan Convolutional Neural Network," *JUSTINDO (Jurnal Sist. dan Teknol. Inf. Indones.)*, vol. 3, no. 2, pp. 49–56, 2018.
- [9] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, no. 4, pp. 611–629, 2018.

- [10] H. Juwiantho, E. I. Setiawan, J. Santoso, and M. H. Purnomo, "Sentiment Analysis Twitter Bahasa Indonesia Berbasis Word2Vec Menggunakan Deep Convolutional Neural Network," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 1, pp. 181–188, 2020.
- [11] H. Zhang, Y. Du, S. Ning, Y. Zhang, S. Yang, and C. Du, "Pedestrian Detection Method Based on Faster R-CNN," *Proc. - 13th Int. Conf. Comput. Intell. Secur. CIS 2017*, vol. 2018-Janua, pp. 427–430, 2018.
- [12] A. Mishra and S. Vishwakarma, "Analysis of TF-IDF Model and its Variant for Document Retrieval," *Proc. - 2015 Int. Conf. Comput. Intell. Commun. Networks, CICN 2015*, pp. 772–776, 2016.
- [13] M. S. Wibawa and U. G. Mada, "Pengaruh Fungsi Aktivasi , Optimisasi dan Jumlah Epoch Terhadap Performa Pengaruh Fungsi Aktivasi , Optimisasi dan Jumlah Epoch Terhadap Performa Jaringan Saraf Tiruan," no. January 2017, 2018.
- [14] N. Triano, *Implementasi Deep Learning Untuk Image Classification Menggunakan Algoritma Convolutional Neural Network (CNN) Pada Citra Wayang Golek*. Universitas Islam Indonesia, Yogyakarta, 2018.
- [15] T. Brian, "ANALISIS LEARNING RATES PADA ALGORITMA BACKPROPAGATION UNTUK KLASIFIKASI PENYAKIT DIABETES," vol. 3, no. 1, pp. 21–27, 2016.