

PEMANFAATAN MACHINE LEARNING SEBAGAI SENSOR BERBASIS MEDIA SOSIAL UNTUK PEMANTAUAN BENCANA BANJIR DI LAHAN BASAH

Mohammad Reza Faisal^{1,*}, Irwan Budiman¹, Friska Abadi¹

¹ Universitas Lambung Mangkurat, Jl Jendral A. Yani KM 36, Banjarbaru, Indonesia

*Penulis korespondensi: reza.faisal@ulm.ac.id

Abstrak. Pesan bencana alam di media sosial dapat digunakan sebagai peringatan dini dan pemantauan saat terjadi bencana alam. Machine learning dapat membantu menentukan pesan media sosial yang dikategorikan sebagai pesan bencana alam. Riset klasifikasi pesan bencana alam terdahulu menggunakan algoritma klasifikasi shallow learning dan teknik ekstraksi fitur data teks dengan vector space representation. Teknik ekstraksi fitur ini menghasilkan data berdimensi tinggi dan jumlahnya tidak bisa ditentukan oleh manusia namun tergantung oleh data teks yang digunakan. Algoritma deep learning 1D CNN umumnya digunakan untuk kasus sinyal. Pada riset ini 1D CNN untuk kasus yang baru yaitu data teks pesan bencana banjir. Penelitian ini melakukan perbandingan dengan teknik Long Short-Term Memory (LSTM). Hasilnya akurasi yang dihasilkan dengan menggunakan LSTM adalah 68,33%. Sedangkan metode yang diusulkan dapat bekerja lebih baik dengan akurasi 74%.

Kata kunci: 1D CNN, LSTM, bencana alam, media sosial, sensor

1. PENDAHULUAN

Pada tahun 2021 bencana banjir melanda 11 dari 13 kabupaten dan kota di Kalimantan Selatan. Penyebaran informasi dan berita saat banjir tersebut sangat cepat karena adanya media sosial seperti Twitter dan Instagram. Pada media sosial, saksi mata memberikan informasi berupa gambar atau video disertai caption teks yang menjelaskan apa yang terjadi. Pada pesan yang diunggah oleh saksi mata juga dapat diketahui lokasi dan waktu kejadian. Pengguna-pengguna lain pada media sosial kerap membagi ulang pesan yang diunggah oleh saksi mata tersebut sehingga informasi tersebut semakin luas menyebar dengan cepat. Pesan ini juga dikategorikan pesan bencana alam namun dibagikan oleh pengguna yang bukan saksi mata. Pengguna pada media sosial dapat menyaring pesan berdasarkan kata kunci, hal ini dapat membantu pengguna untuk mencari pesan bencana alam. Contoh kata kunci adalah “banjir”, “air pasang” dan lain-lain. Namun pesan hasil penyaringan bisa saja bukan pesan tentang bencana alam walaupun di dalam pesan tersebut terdapat kata “banjir” atau “air pasang”. Contoh pesan tersebut adalah sebagai berikut “hari ini ini aku banjir air mata”.

Dari penjelasan di atas maka pesan-pesan bencana alam di media sosial dapat dikategorikan menjadi tiga kategori yaitu pesan bencana dari saksi mata, pesan bencana bukan dari saksi mata, dan pesan bukan tentang bencana. Pembagian tiga kategori ini seperti telah dilakukan pada beberapa penelitian tentang pemanfaatan media sosial sebagai sensor bencana alam (Delimayanti *et al.*, 2021; Nooralifa *et al.*, 2021; Zahra *et al.*, 2020). Dimana manusia berperan sebagai sensor yang mendeksi dan melakukan dokumentasi ketiga bencana alam terjadi. Berdasarkan hasil riset tersebut maka media sosial dapat digunakan sebagai sistem peringatan dini dan pemantauan ketika terjadi bencana.

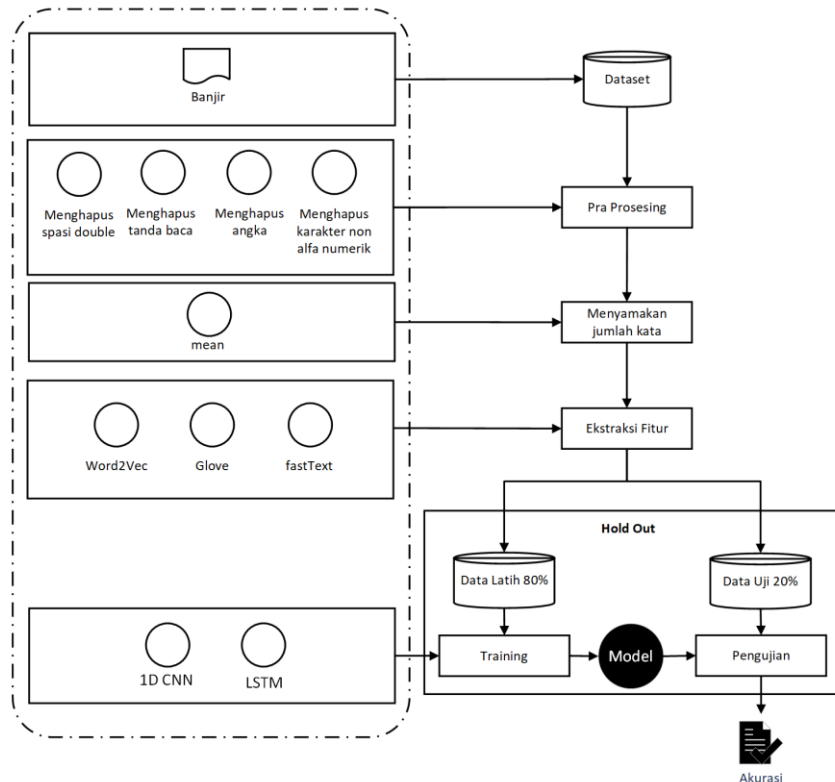
Manusia yang membaca pesan-pesan tersebut dapat dengan mudah membedakan pesan berdasarkan kategori tersebut. Namun manusia akan kesulitan jika harus memeriksa pesan-pesan pada media sosial yang jumlahnya sangat banyak. Pada beberapa penelitian telah dilakukan untuk melakukan klasifikasi secara otomatis pesan-pesan bencana alam menjadi tiga kategori dengan menggunakan komputer yang disematkan program berbasis kecerdasan buatan (Delimayanti *et al.*, 2020; Faisal *et al.*, 2021; Firlia *et al.*, 2021). Pada penelitian tersebut digunakan algoritma klasifikasi berbasis *shallow learning* seperti Random Forest, Support Vector Machine dan lain-lain. Seiring perkembangan pengetahuan ditemukan algoritma klasifikasi berbasis *deep learning* seperti Deep Neural Networks (DNN) (Samek *et al.*, 2021), Convolutional Neural Networks (CNN) (Rinaldi *et al.*, 2021), Long Short Term Memory Network (LSTM) (Jena *et al.*, 2021), atau Recurrent Neural Network (RNN) (Hernandez-Suarez *et al.*, 2019).

Untuk kasus klasifikasi teks, biasanya digunakan algoritma deep learning yang umum digunakan adalah LSTM atau RNN. Hal ini dikarenakan algoritma ini sensitif urutan data. Ini dikarenakan LSTM memiliki

kemampuan untuk mengingat informasi sebelumnya untuk mengenali pola sehingga akurasi klasifikasi dapat meningkat. Hal ini sesuai dengan makna kalimat yang sangat tergantung dengan urutan kata-kata didalamnya. Sedangkan CNN umumnya untuk melakukan klasifikasi gambar dengan 2D CNN (Nugroho *et al.*, 2020) atau video dengan 3D CNN (Kusumastuti *et al.*, 2021). Selain itu juga dikenal 1D CNN yang sering digunakan untuk kasus klasifikasi sinyal (Kiranyaz *et al.*, 2019). Pada (Rachman & Santoso, 2021), 2D CNN digunakan untuk kasus klasifikasi teks namun 1D CNN masih kurang dimanfaatkan untuk klasifikasi teks secara umum terlebih lagi untuk kasus klasifikasi pesan bencana alam. Pada penelitian ini dilakukan perbandingan klasifikasi pesan bencana alam dengan LSTM dan 1D CNN untuk mengetahui algoritma mana yang dapat bekerja lebih baik.

2. METODE

Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur penelitian

Dari gambar di atas dapat dilihat tahapan-tahapan penelitian yaitu: koleksi dataset, pre prosesing, menyamakan jumlah kata, ekstraksi fitur, pembagian data latih dan uji, training dengan menggunakan data latih dengan menggunakan algoritma 1D CNN dan LSTM, kemudian dilanjutkan dengan pengujian dan perhitungan kinerja klasifikasi.

2.1. Dataset

Sumber data adalah dari media sosial Twitter. Data diambil dengan menggunakan kata kunci terkait dengan bencana banjir seperti: banjir, banjir bandang, air pasang, air naik, air bah dan lain-lain. Dataset yang digunakan berasal dari penelitian terdahulu berikut ini (Delimayanti *et al.*, 2021; Firlia *et al.*, 2021). Detail dari dataset ini dapat dilihat pada Tabel 1.

Tabel 1. Dataset

Kategori	Jumlah
Saksi mata	1000
Non saksi mata	1000

Tidak tahu	1000
Jumlah	3000

Contoh pesan dari masing-masing kategori dapat dilihat pada Tabel 2.

Tabel 2. Contoh pesan bencana alam berdasarkan kategori

Kategori	Pesan
Saksi mata	hujan lebat gais banjir disini huhu
Saksi mata	ujannya baru berapa jam udah banjir aja cepet amat banjirnya
Non saksi mata	assalamualaikum wsalam menjawab salam selamat pagi salam sejahtera semoga warga pantai timuq tabah menghadapi bencana banjir
Non saksi mata	pkpu sumbar berikan trauma healing pada anak korban bencana solsel melawan pos keadilan peduli umat pkpu human initiative sumatra barat hadir di tengah tengah pengungsi bencana banjir
Tidak tahu	jadi gini rasanya banjir konten
Tidak tahu	jangan nangis ntar banjir yen

2.2. Pre Prosesing Teks dan Menyamakan Jumlah Kata

Teks pada dataset dinormalisasi dengan cara menghapus spasi ganda, menghapus tanda baca, menghapus angka, dan menghapus karakter non alfa numerik (Firlia *et al.*, 2021; Nooralifa *et al.*, 2021). Tujuannya agar teks bersih dan hanya berisi kata-kata bahasa alami manusia.

Setiap pesan teks pada dataset memiliki jumlah kata yang berbeda-beda. Hal ini menjadi permasalahan tersendiri karena pada proses ekstraksi fitur hanya dapat dilakukan jika setiap pesan memiliki jumlah kata yang sama. Teknik menyamakan jumlah kata ini dikenal dengan istilah *text padding*. Jumlah kata yang digunakan pada untuk padding berdasarkan perhitungan statistik mean atau rata-rata jumlah kata pada setiap pesan pada dataset.

2.3. Ekstraksi Fitur

Ekstraksi fitur data teks bertujuan mengubah teks sebagai data tidak terstruktur menjadi data terstruktur (Rusli *et al.*, 2019). Data terstruktur adalah data berbentuk tabel yang terdiri atas baris dengan jumlah kolom yang sama. Pada penelitian digunakan tiga model word embedding yaitu word2vec (Rusli *et al.*, 2020), Glove (Pennington *et al.*, 2014), dan fastText (Khomsah *et al.*, 2022). Ketiga model word embedding ini menjadi kamus untuk melakukan ekstraksi fitur. Setiap kata pada model word embedding ini direpresentasikan sebagai sebuah vektor dengan 100 nilai numerik.

Sebuah pesan adalah kalimat yang terdiri atas beberapa kata. Jika sebuah kata terdiri atas 100 nilai, maka kalimat yang terdiri atas lima kata akan memiliki jumlah 500 nilai. Ilustrasi penjelasan ini dapat dilihat pada.

V1	...	V100	V1	...	V100	V1	...	V100	V1	...	V100	V1	...	V100
Vektor kata1			Vektor kata2			Vektor kata3			Vektor kata4			Vektor kata5		

Gambar 2. Pembentukan vektor kalimat

Pada penelitian ini juga dilakukan penggabungan hasil ekstraksi fitur dari ketiga teknik word embedding. Artinya jika vektor sebuah kalimat memiliki 500 nilai, maka hasil penggabungan tiga teknik word embedding ini menghasilkan 1500 nilai. Ilustrasinya dapat dilihat pada Gambar 3.

V1	V2	...	V500	V1	V2	...	V500	V1	V2	...	V500
Word2vec				Glove				fastText			

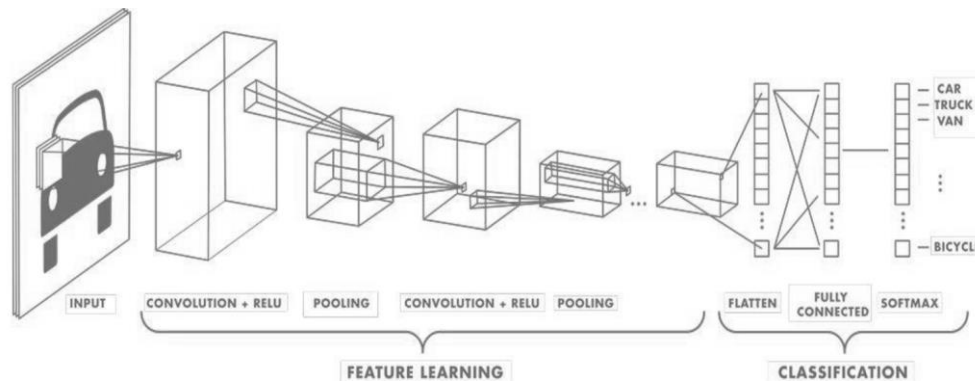
Gambar 3. Pembentukan vektor kata dari tiga teknik word embedding

2.4. 1D Convolutional Neural Network

Convolutional Neural Network (CNN) adalah salah satu algoritma *deep learning*. CNN sudah dikenal luas dalam bidang pengolahan gambar. CNN meniru cara sel-sel syaraf berkomunikasi dengan neuron yang saling

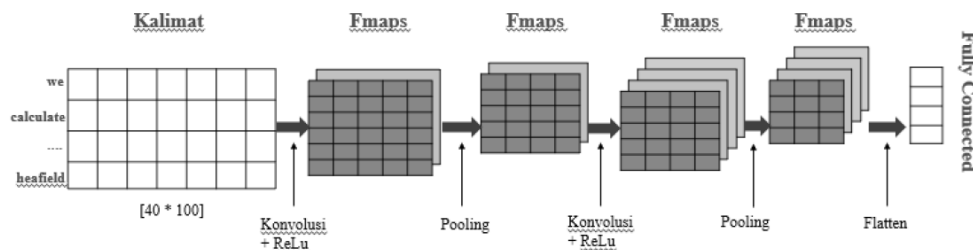


berhubungan. Arsitektur CNN memiliki keunikan dibandingkan jaringan syaraf tiruan yang lain yaitu adanya operasi *convolutional*. Operasi ini melakukan filter pada setiap bagian input untuk mengekstraksi pola dan *features maps* (Kholik, 2021). Arsitektur CNN dapat dilihat pada Gambar 4. Arsitektur ini merupakan arsitektur 2D CNN.



Gambar 4. Arsitektur 2D CNN untuk klasifikasi gambar (Kholik, 2021)

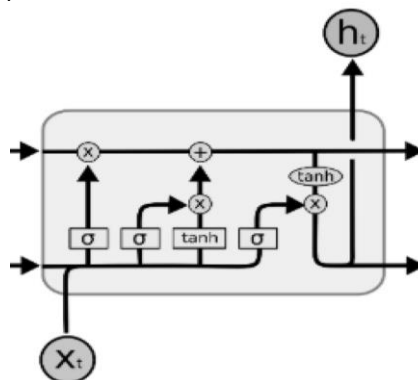
Sedangkan untuk kasus klasifikasi data teks dengan menggunakan 1D CNN telah dilakukan oleh (Mahendra *et al.*, 2020) dengan arsitektur seperti pada Gambar 5.



Gambar 5. Arsitektur 1D CNN untuk klasifikasi teks (Mahendra *et al.*, 2020).

2.5. Long Short-Term Memory

LSTM merupakan pengembangan dari metode *deep learning* RNN yang memiliki kelebihan mampu memproses data yang relatif panjang (*long-term dependency*) (Widayat, 2021). Metode ini pertama kali diperkenalkan pada tahun 1997 oleh Schmidhuber. LSTM sendiri dapat menemukan lapisan tersembunyi dari setiap sel dan dirancang untuk menyimpan informasi sel sebelumnya. Metode LSTM digunakan dengan cara mengklasifikasi data secara jangka panjang dengan menyimpan pada sel memori. LSTM sendiri memiliki empat komponen utama yaitu: input gate, koneksi berulang, forget gate dan output gate (yuli Astari *et al.*, 2021). Arsitektur LSTM dapat dilihat pada Gambar 6.



Gambar 6. Arsitektur LSTM (Widayat, 2021)

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Berikut ini adalah hasil-hasil dari setiap tahapan-tahapan penelitian. Hasil pada tahap pre prosesi teks dan menyamakan jumlah kata adalah data teks yang sudah bersih dengan jumlah setiap pesan memiliki jumlah kata sebanyak delapan kata.

Pada tahap berikutnya adalah ekstraksi fitur berbasis word embedding. Hasilnya adalah beberapa data terstruktur dengan detail seperti pada Tabel 3.

Tabel 3. Hasil ekstraksi fitur data teks banjir

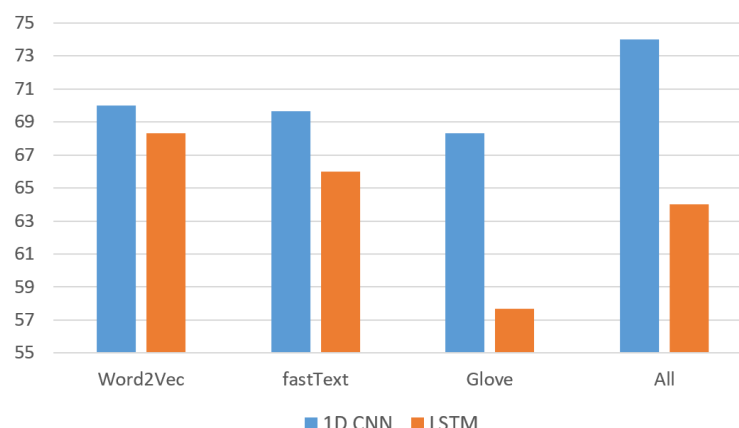
Input Data Teks	Teknik Word Embedding	Output Data Terstruktur	Jumlah Fitur
Banjir	Word2vec	B_W	800
	fastText	B_F	800
	Glove	B_G	800
	Word2vec, fastText, Glove	B_All	2400

Dataset pesan bencana banjir yang diproses oleh ekstraksi fitur berbasis word2vec akan menghasilkan data terstruktur dengan kode B_W dan dengan jumlah fitur 800. Selanjutnya hasil ekstraksi fitur berbasis fastText dan Glove menghasilkan data terstruktur dengan kode B_F dan B_G dengan jumlah fitur yang sama yaitu 800. Sedangkan data terstruktur dengan menggunakan ekstraksi fitur penggabungan tiga teknik word embedding diberi kode B_All dengan jumlah fitur adalah 2400.

Dari hasil tersebut maka dilakukan keempat dataset hasil tahapan tersebut diproses oleh teknik 1D CNN dan LSTM. Hasil dari tahapan ini adalah kinerja klasifikasi dari masing-masing teknik. Kinerja klasifikasi yang digunakan adalah akurasi yang dapat dilihat pada Tabel 4.

Tabel 4. Kinerja klasifikasi 1D CNN dan LSTM

Output Data Terstruktur	Akurasi (%)	
	1D CNN	LSTM
B_W	70	68.33
B_F	69.67	66
B_G	68.33	57.67
B_All	74	64



Gambar 7. Perbandingan akurasi

3.2. Pembahasan

Dari hasil yang ditampilkan pada Tabel 4 dapat diketahui beberapa informasi. Yang pertama adalah ekstraksi fitur berbasis word2vec dapat membantu untuk memberikan kinerja klasifikasi terbaik baik dengan

menggunakan teknik klasifikasi 1D CNN atau LSTM. Hal tersebut dapat dilihat pada Gambar 7. Sedangkan ekstraksi fitur berbasis Glove memberikan kinerja terendah dibanding teknik word embedding yang lain.

Dari Gambar 7 juga dapat dilihat bahwa kinerja teknik klasifikasi 1D CNN dapat bekerja lebih baik jika dibandingkan dengan kinerja LSTM. Hal ini artinya 1D CNN juga dapat dimanfaatkan untuk kasus klasifikasi data teks. Selain itu walaupun LSTM dilaporkan dapat bekerja baik untuk data sensitif dengan urutan, namun terlihat data terstruktur berbasis word embedding tidak bekerja baik pada teknik LSTM. Sedangkan 1D CNN terlihat cocok untuk bekerja dengan data terstruktur berbasis word embedding terkhusus teknik word2vec.

Pada penelitian ini juga dapat dilihat bahwa data terstruktur hasil dari penggabungan ketiga teknik word embedding meningkatkan kinerja klasifikasi jika menggunakan teknik 1D CNN yaitu didapat hasil akurasi 74%. Nilai akurasi ini tertinggi dibandingkan hasil-hasil lainnya. Namun data hasil penggabungan tiga teknik word embedding tersebut malah menurunkan kinerja klasifikasi saat digunakan pada LSTM yaitu nilai tertingginya adalah 68,33% dengan menggunakan word2vec kemudian menjadi turun 64%.

4. SIMPULAN

Teknik klasifikasi 1D CNN mempunyai potensi untuk digunakan pada klasifikasi teks dengan cara menggabungkan beberapa teknik word embedding. Sebagaimana diketahui bahwa CNN juga dapat memproses data 2D dan 3D, berdasarkan hal tersebut maka pada penelitian selanjutnya akan dilakukan penggabungan teknik word embedding untuk diproses oleh teknik klasifikasi 2D CNN dan 3D CNN.

5. UCAPAN TERIMA KASIH

Komputer yang digunakan pada penelitian ini disediakan oleh Laboratorium Data Science, Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Penelitian ini merupakan Program Dosen Wajib Meneliti (PDWM) dari hibah PNPB Universitas Lambung Mangkurat

6. DAFTAR PUSTAKA

- Delimayanti, M. K., Sari, R., Laya, M., Faisal, M. R., Pahrul, & Naryanto, R. F. (2020). The effect of pre-processing on the classification of twitter's flood disaster messages using support vector machine algorithm. *Proceedings of ICAE 2020 - 3rd International Conference on Applied Engineering, February 2021*. <https://doi.org/10.1109/ICAE50557.2020.9350387>
- Delimayanti, M. K., Sari, R., Laya, M., Faisal, M. R., & Pahrul, P. (2021). Pemanfaatan Metode Multiclass-SVM pada Model Klasifikasi Pesan Bencana Banjir di Twitter. *Edu Komputika Journal*, 8(1), 39–47.
- Faisal, M. R., Nugroho, R. A., Ramadhani, R., Abadi, F., Herteno, R., & Saragih, T. H. (2021). Natural Disaster on Twitter: Role of Feature Extraction Method of Word2Vec and Lexicon Based for Determining Direct Eyewitness. *Trends in Sciences*, 18(23), 680.
- Firlia, K. Y., Faisal, M. R., Kartini, D., Nugroho, R. A., & Abadi, F. (2021). Analysis of New Features on the Performance of the Support Vector Machine Algorithm in Classification of Natural Disaster Messages. *2021 4th International Conference of Computer and Informatics Engineering (IC2IE)*, 317–322.
- Hernandez-Suarez, A., Sanchez-Perez, G., Toscano-Medina, K., Perez-Meana, H., Portillo-Portillo, J., And Luis, V. S., & Javier Garcia Villalba, L. (2019). Using Twitter Data to Monitor Natural Disaster Social Dynamics: A Recurrent Neural Network Approach with Word Embeddings and Kernel Density Estimation. *Sensors (Basel, Switzerland)*, 19(7). <https://doi.org/10.3390/s19071746>
- Jena, R., Naik, S. P., Pradhan, B., Beydoun, G., Park, H.-J., & Alamri, A. (2021). Earthquake vulnerability assessment for the Indian subcontinent using the Long Short-Term Memory model (LSTM). *International Journal of Disaster Risk Reduction*, 66, 102642.
- Kholik, A. (2021). Klasifikasi Menggunakan Convolutional Neural Network (CNN) Pada Tangkapan Layar Halaman Instagram. *Jurnal Data Mining Dan Sistem Informasi*, 2(2), 10–20.
- Khomsah, S., Ramadhani, R. D., Wijaya, S., & others. (2022). The Accuracy Comparison Between Word2Vec and FastText On Sentiment Analysis of Hotel Reviews. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(3), 352–358.
- Kiranyaz, S., Ince, T., Abdeljaber, O., Avci, O., & Gabbouj, M. (2019). 1-D convolutional neural networks for signal processing applications. *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 8360–8364.
- Kusumastuti, A. R., Kristian, Y., & Setyati, E. (2021). Klasifikasi Ketertarikan Belajar Anak PAUD Melalui Video Ekspresi Wajah Dan Gestur Menggunakan Convolutional Neural Network. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 10(2), 182–188.
- Mahendra, Y. E., Ilyas, R., & Kasyidi, F. (2020). Klasifikasi Kalimat Ilmiah Menggunakan 1D Convolutional Neural Networks. *Prosiding Industrial Research Workshop and National Seminar*, 11(1), 503–509.
- Nooralifa, S. M., Faisal, M. R., Abadi, F., Nugroho, R. A., & Aziz, M. (2021). IDENTIFIKASI OTOMATIS PESAN SAKSI MATA PADA MEDIA SOSIAL SAAT BENCANA GEMPA. *KLIK-KUMPULAN JURNAL ILMU KOMPUTER*, 8(2), 129–138.
- Nugroho, P. A., Fenriana, I., & Arijanto, R. (2020). Implementasi Deep Learning Menggunakan Convolutional Neural Network (Cnn) Pada Ekspresi Manusia. *Algor*, 2(1), 12–20.
- Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.
- Rachman, F. P., & Santoso, H. (2021). Perbandingan Model Deep Learning untuk Klasifikasi Sentiment Analysis dengan Teknik Natural

- Language Processing. *Jurnal Teknologi Dan Manajemen Informatika*, 7(2), 103–112.
- Rinaldi, Faisal, M. R., Mazdadi, M. I., Nugroho, R. A., Abadi, F., & Others. (2021). Eye Witness Message Identification on Forest Fires Disaster Using Convolutional Neural Network. *Journal of Data Science and Software Engineering*, 2(02), 100–108.
- Rusli, M., Faisal, M. R., & Budiman, I. (2019). Ekstraksi Fitur Menggunakan Model Word2vec untuk Analisis Sentimen pada Komentar Facebook. *Soliter*, 2, 104–109.
- Rusli, M., Faisal, M. R., Budiman, I., Nugroho, R. A., & Farmadi, A. (2020). Ekstraksi Fitur Menggunakan Model Word2Vec Pada Sentiment Analysis Kolom Komentar Kuisisioner Evaluasi Dosen Oleh Mahasiswa. *Klik-Kumpul. J. Ilmu Komput*, 7(1), 35.
- Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K.-R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. *Proceedings of the IEEE*, 109(3), 247–278.
- Widayat, W. (2021). Analisis Sentimen Movie Review menggunakan Word2Vec dan metode LSTM Deep Learning. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(3), 1018–1026.
- yuli Astari, Y., Afiyati, A., & Rozaqi, S. W. (2021). Analisis Sentimen Multi-Class pada Sosial Media menggunakan metode Long Short-Term Memory (LSTM). *Jurnal Linguistik Komputasional*, 4(1), 8–12.
- Zahra, K., Imran, M., & Ostermann, F. O. (2020). Automatic identification of eyewitness messages on twitter during disasters. *Information Processing and Management*, 57(1), 102107. <https://doi.org/10.1016/j.ipm.2019.102107>

