



Komparasi Algoritma Klasifikasi Dan Penerapan NER pada Analisis Sentimen Bencana Alam Banjir

Aprilisa Arum Sari^{1*}, Muhammad Risky², Ayu Aprillia³

¹ Universitas Duta Bangsa Surakarta, ²³Magister Ilmu Komputer, Universitas Budi Luhur
¹Jl. Bhayangkara no. 55-57 Tipes, Surakarta

²Direktorat Bina Umrah dan Haji Khusus, Kementerian Agama Republik Indonesia
Jalan Lapangan Banteng Barat No. 3 - 4, Jakarta Pusat 10710

³ Universitas Budi Luhur, Jl. Ciledug Raya, Petukangan Utara, Jakarta Selatan, 12260, DKI
Jakarta, Indonesia

*Email Penulis Koresponden: 2011600711@student.budiluhur.ac.id

Abstrak:

Bencana alam akhir-akhir ini terjadi di Indonesia. Layanan jejaring sosial twitter memberi ruang bagi masyarakat untuk menanggapi terkait bencana yang disebabkan oleh faktor alam, namun tanggapan yang diberikan oleh masyarakat belum diklasifikasikan. Pada penelitian ini, analisis sentimen dilakukan terhadap tweet yang mengklasifikasikannya ke dalam kategori katastrofik dan non-katastrofik menggunakan metode klasifikasi komparatif dan pemodelan lain yang disebut *Entity Recognition Modeling* (NER). Penelitian ini menggunakan 6 algoritma klasifikasi, yaitu multinominal naive bayes, random forest, linier regression, support vector machine, decision tree dan KNN. Untuk data training dan data testing diambil dengan metode random sampling dengan presentase data training 80% dan data testing 20%. Pemodelan NER dilakukan dengan Spacy untuk mendapatkan LOCATION, ORGANIZATION, PERSON, QUANTITY, TIME. Setelah dilakukan pemodelan NER dengan spacy, dilanjutkan pengukuran accuracy, precision, recall, f1-score. Support. Perhitungan precision, recall, dan f-measure mendapatkan nilai terbaik yaitu 80%, 100%, dan 89% untuk metode NER Sedangkan untuk hasil prediksi bencana dan non bencana didapatkan hasil yaitu 81,60%, 82% dan 82%.

Kata Kunci:

Naive bayes;
Random forest;
Linier regression;
Support vector machine;
Decision tree;
KNN;
Named Entity Recognition (NER);

Riwayat Artikel:

Diserahkan 5 September, 2023
Direvisi 14 Januari, 2024
Diterima 25 Juli, 2024

DOI:

10.22441/incomtech.v14i2.22962



1. PENDAHULUAN

Indonesia ialah negeri tropis terbanyak di dunia yang wilayahnya dilintasi garis khatulistiwa. Hawa tropis yang terdapat di Indonesia sudah menimbulkan banyak

bencana alam yang wajib dirasakan Indonesia. Bencana alam akhir-akhir ini terjadi di Indonesia. Indonesia berada di wilayah tropis sehingga menyebabkan rawan bencana alam, salah satunya adalah bencana banjir. Banjir ialah bencana alam yang sangat mengusik kegiatan warga. Banjir pula mengganggu infrastruktur dan menghambat aktivitas ekonomi [1]. Seperti yang terjadi pada Jumat, 15 Juli 2022 malam, Jakarta diguyur hujan deras yang mengakibatkan beberapa wilayah di Jakarta terdampak banjir yang meluas dari 71 RT menjadi 92 RT. Berbagai tanggapan dan cara masyarakat Indonesia terhadap peristiwa tersebut dilakukan dengan bantuan media sosial, salah satunya media sosial twitter.

Masyarakat Indonesia membagikan reaksi mereka terhadap bencana alam melalui media sosial twitter, tetapi reaksi ini tidak terklasifikasi [2]. Data mining adalah istilah untuk mencari informasi dalam database. Penambangan data terdiri dari proses tercantum memakai matematika, statistika, kecerdasan buatan, serta metode pendidikan mesin buat mengekstraksi serta mengenali informasi serta kumpulan informasi yang bermanfaat dari bermacam basis informasi [3]. Text mining merupakan pelaksanaan konsep pengetahuan serta metode penambangan buat menciptakan pola dalam proses menganalisis bacaan buat menciptakan data yang bermanfaat buat tujuan tertentu [4]. Klasifikasi merupakan metode yang terbuat buat memprediksi tiap peristiwa ataupun fitur informasi. Proses klasifikasi dicoba sehabis informasi lewat sesi pre- processing. Pada sesi preprocessing dicoba pengecekan duplikasi informasi, pemilihan fitur, confusion matrix [5]. Dalam konteks ini, perlu dilakukan pengklasifikasian respon masyarakat terhadap bencana alam banjir.

Penelitian ini, didapatkan data dari twitter sebanyak 1.500 data dengan hastag (#) banjir yang kemudian dilakukan pelabelan secara manual. Hasil pengolahan data tersebut dengan 2 kategori yaitu data yang termasuk dalam kategori bencana dan bukan bencana. Setelah didapatkan data tweet-tweet bencana banjir, selanjutnya akan dilakukan tahap bagian dari penggalian data informasi untuk pengklasifikasi suatu dokumen atau materi yang diklasifikasikan dalam beberapa kategori seperti orang (*person*), organisasi (*organisation*), lokasi (*location*), waktu (*time*) dan lainnya. Pemodelan ini biasanya disebut pemodelan Named Entity Recognition dan disingkat dengan NER. Pemodelan ini bisa menolong proses ekstraksi data dengan metode mengenali sesuatu entitas nama.

Penelitian sebelumnya oleh Adi Mashabbi Maksun (2021) [6] melakukan *sentiment analyse natural disaster* pada salah satu provinsi di pulau Kalimantan pada Twitter menggunakan metode algoritma Naive Bayes dengan dataset kecil yaitu 520 data uji dengan menggunakan 5 fold sedangkan Imam Fahrur Rozi (2020) [7] melakukan *sentiment analyze* di Twitter mengenai pasca bencana menggunakan metode algoritma klasifikasi Naive Bayes dengan menggunakan fitur dari N-gram yaitu unigram dan bigram, dengan jumlah data yang sedikit yaitu 300. Empat uji coba dilakukan dalam tahap pengujian penelitian ini. Namun kedua penelitian ini tidak menggunakan pendekatan pemodelan yang berbeda yaitu *Named Entity Recognition* (NER), dan tidak mempertimbangkan keberadaan entitas sebagai unsur utama data peristiwa.

Banyak penelitian perbandingan algoritma juga telah dilakukan, misalnya Vinita [8] yang meneliti tentang Komparasi beberapa Algoritma Klasifikasi *Machine Learning* serta *Feature Selection* untuk Analisis Sentimen buat *Review Film* dengan algoritma Naive Bayes, SVM dan ANN. Dari hasil perbandingan ketiga algoritma

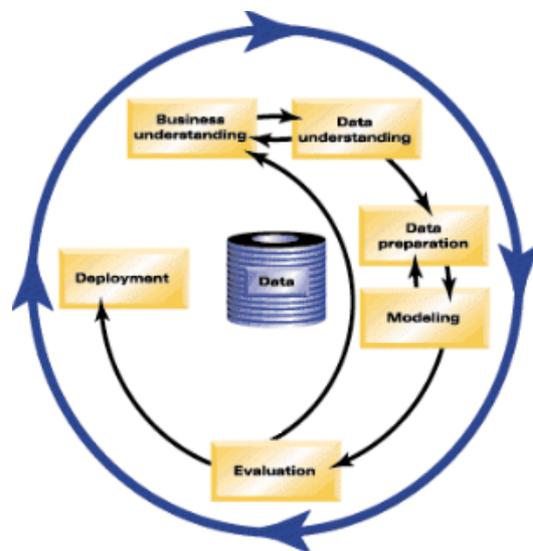
tersebut, algoritma pemodelan terbaik adalah algoritma SVM dengan hasil akurasi sebesar 81,10%. Mawaddah Harahap dkk [8] juga meneliti tentang komparasi 4 algoritma untuk analisis sentimen *review* penjualan produk UMKM. Dari hasil komparasi 4 algoritma didapatkan algoritma terbaik untuk pemodelannya yaitu algoritma SVM dengan nilai akurasi 98,8%.

Dengan demikian, untuk memudahkan masyarakat dan pihak terkait dalam memperoleh informasi bencana alam banjir yang terjadi, dilakukan penelitian ini melakukan pemodelan lebih lanjut dengan menerapkan pemodelan *Named Entity Recognition* (NER) untuk mendeteksi entitas banjir dari suatu teks, yang berguna saat mengekstraksi data banjir dengan menggunakan 5 entitas yaitu orang (person), organisasi (organisation), lokasi (location), waktu (time) dan jumlah (quantity).

Pada penelitian ini diusulkan dilakukan pengklasifikasian dengan beberapa algoritma untuk menggali pengetahuan terkait banjir yang terjadi di Indonesia serta penerapan NER untuk mengetahui entitas location, person, time, quantity dan organization pada data twitter yang akan bermanfaat bagi pihak-pihak yang terlibat dalam mengatasi permasalahan banjir di Indonesia.

2. METODE

Dalam riset ini, metodologi informasi data mining CRISP- DM digunakan selaku pemecah permasalahan universal dalam bisnis serta riset. Tata cara ini terdiri dari enam tahapan ialah *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, serta *Deployment* [9]. Proses tahapan untuk penelitian ini terdiri dari enam tahapan yang dapat dijelaskan sebagai berikut.



1. Business Understanding (Pemahaman Bisnis)

Berbagai perhal dicoba dalam fase ini, semacam menguasai kebutuhan serta tujuan perspektif bisnis, menginterpretasikan informasi dalam perhal mendefinisikan permasalahan informasi mining, serta menetapkan rencana serta strategi buat menggapai tujuan informasi mining. [10].

2. Data Understanding (Pemahaman Data)

Pada tahap ini dilakukan dengan cara pengumpulan data dari twitter menggunakan tools orange, deskripsi serta evaluasi mutu data [11].

3. Data Preparation (Persiapan Data)

Pada tahap dilakukan persiapan dan kumpulan informasi final dari informasi mentah. Berbagai perihal dicoba antara lain dengan pembersihan atau *cleaning* data, kemudian dilanjutkan dengan pemilihan data (data Selection) yang akan digunakan, selanjutnya dilakukan *record* data serta atribut- atribut, serta pula melaksanakan transformasi terhadap data (data Transformation) buat dijadikan masukan dalam sesi pemodelan.

4. Modelling (Pemodelan)

Berbagai hal yang harus dilakukan dalam tahap pemodelan antara lain, termasuk pemilihan teknik pemodelan, pemodelan, evaluasi model [11]. Pada tahapan ini dilakukan secara langsung untuk mencakup machine learning buat penentuan metode informasi mining, perlengkapan bantu informasi mining dan algoritma informasi mining. Enam algoritma klasifikasi digunakan dalam penelitian ini kemudian dilakukan penerapan NER untuk mengetahui entitas seperti location, person, time, quantity dan organization.

5. Evaluation (Pengujian)

Pada tahap ini dicoba dengan cara melihat tingkatan dari standar pola yang dihasilkan oleh suatu algoritma. Sehingga parameter yang akan digunakan dibuat untuk mengevaluasi suatu algoritma perbandingan merupakan Confusion Matrix serta dilihat ketentuan nilai akurasi, presisi serta recall.

6. Deployment (Penyebaran)

Pada tahapan ini dilakukan pembuatan laporan berupa hasil yang didapatkan dan artikel jurnal yang menggunakan suatu model algoritma yang dihasilkan.

3. HASIL DAN PEMBAHASAN

A. Business Understanding

Penerapan data mining untuk penelitian ini berhubungan secara langsung dengan data terkait bencana alam yaitu banjir untuk menggali pengetahuan tentang banjir yang terjadi di Indonesia. Serta penerapan NER untuk mengetahui entitas location, person, time, quantity dan organization pada data twitter.

B. Data Understanding

Data yang diperoleh pada penelitian ini yaitu melalui twitter yang ditambah dengan tools orange dengan keyword #banjir. Data tersebut merupakan data twitter bencana banjir pada periode Juli 2022.

Tabel 1. Data Twitter

RT @infobmgkya: Peringatan Dini Potensi Banjir Pesisir (ROB)
di Wilayah Pesisir Selatan Jawa Barat, Jawa Tengah, dan Yogyakarta.
Berlaks, @AchmadZ83168521
RT @AryBandung: Perubahan iklim timbukan dilema bg rantai pasokan. Haruskah kita bertindak skrg atau tunggu apa yg mungkin terjadi? Kita t&e
RT @infobmgkya: Peringatan Dini Potensi Banjir Pesisir (ROB)
di Wilayah Pesisir Selatan Jawa Barat, Jawa Tengah, dan Yogyakarta.
Berlaks, @SitorusSalmah
@Hermans93659231 @DedenLG @tatakujiyati jakarta masih banjir dan macet gak bang @setyobu83879420
Lokasi penanaman merupakan lahan kritis yang berpotensi banjir di bagian bawah bukit.A
#SahabatGanjar #GanjarPranowo #Ganjar,@4_gp2024
baru pulang, sehanan urusan berkas+nerobos banjir+ngejar deadline ktm yang gagal&y' https://t.co/o8xUuC3wpm.@gyubiinn
@Yuwii_Maaciw yaa smg kita banjir bigwin di akun baru,@Selamat031201

C. Data Preparation

Pada Tahap ini yaitu *Data Cleaning*, dataset yang terdiri dari 1500 dengan 733 dengan label bencana dan 754 dengan label bukan bencana. Setelah di *preprocessing* didapatkan data sebanyak 870 data yang akan digunakan dalam penelitian.

Tabel 2. Data Bersih

pas dodo masuk udah berita emyu banjir tropi
berkah banjir moment banget gak kuat
munyak banjir trophy mah prediksi musim wkwk
pusingg aiih nonton banjir bgt menit sampe huua
ngirim tanam banjir guys bisabisa air abis ngirim tanam aja lupa siram ya tanam semangatsalingbantu
mimpi banjir banjir rezeki si kaya
yah lester city aja juara epltrus knpa totenham gk nepati peringkat musim kemarin aja posisi lolos ucl dan man united prediksi
ya allah bantal sdh tdk bsa tamping banjir
seret tibatiba banjir momen

Pada Sesi Data Transformation kumpulan informasi diganti jadi format informasi yang cocok dalam proses informasi mining. Data tersebut dicoba pergantian dari format xls jadi format csv, Dataset bencana diberi label angka 0(nol) sebaliknya non bencana diberi label angka 1(satu). Data tersebut setelah itu ditaruh serta diimport ke tools informasi mining memakai python.

Tabel 3. Data Transformasi

saman raja negeri rana suara lantang aktiviti bala kawal amp	0
kalo rumah lu banjir gue kasih tau aja air bah tangis gue	0
banjir nder	0
tinjau progres bangun bendung kering dry dam ciawi bendung irigasi air b	0
jkt tmbh kumuh banjir	1
met bobooo beso bangun notif banjir congrats amin	0
trio serang favorit	0

D. Modelling

Pada fase ini, proses dimana model memakai metode klasifikasi terpilih menciptakan sesuatu pola data yang bisa mempermudah pihak- pihak yang berkepentingan. Model klasifikasi yang dihasilkan metode data mining ini digunakan buat memprediksi twitter bencana serta non bencana. Ada pula *tools* ataupun perlengkapan bantu buat pemrosesan pengolahan data mining merupakan *Visual Code* dengan bawah *Python 3. 9*. Sesi ini menunjukkan dan membagikan data menimpa kinerja algoritma dalam tata cara klasifikasi.

Pada penelitian ini menggunakan 6 algoritma klasifikasi, yaitu *multinomial naive bayes*, *random forest*, *linier regression*, *support vector machine*, *decission tree* dan *knn*. Untuk *data training* dan *data testing* dikumpulkan dengan metode *random sampling* dengan *presentase data training* 80% dan *data testing* 20%.

	Model	% Accuracy
0	Naive Bayes	0.816092
1	Random Forest	0.701149
2	Linier Regression	0.764368
3	SVM	0.712644
4	Decission Tree	0.678161
5	KNN	0.591954

Gambar 1. Hasil Komparasi 6 Algoritma

Pada Gambar 1. merupakan hasil komparasi dari 6 algoritma klasifikasi. Dari hasil komparasi tersebut didapatkan hasil menunjukan bahwa algoritma naive bayes adalah algoritma terbaik dengan nilai akurasi 81,6092% dibandingkan dengan algoritma klasifikasi yang lainnya.

Tweet yang merupakan bencana banjir dilakukan *Name Entity Recognition (NER)* dengan pemodelan *Spacy*. Pemodelan *Spacy*, telah memiliki pemodelan beberapa bahasa, diantaranya Bahasa: *Catalan, Chinese, Croatian, Danish, English, French, Italian, Korean* dan sebagainya. Sedangkan untuk Bahasa Indonesia, *Spacy* belum memiliki pemodelan Bahasa Indonesia [12].

Dikarenakan *Spacy* belum memiliki pemodelan untuk Bahasa Indonesia, maka pada penelitian ini digunakan *data training* sebanyak 3.561 data (<https://spacy.io/models>). *Data training* ini terdiri dari teks dan juga anotasi [13]. Dilakukan *preprocessing data*, yaitu *selection attribute*. Atribut yang digunakan adalah *tweet* yang merupakan bencana banjir. Kemudian memberikan huruf kapital di awal kata [14], seperti untuk untuk nama, nama lokasi, nama *organization*.

Pemodelan NER dilakukan dengan *Spacy* untuk mendapatkan LOCATION, ORGANIZATION, PERSON, QUANTITY, TIME. Setelah dilakukan pemodelan NER dengan *spacy* [15], dilanjutkan pengukuran *accuracy, precision, recall, fl-score* [16] *Support*.

Tabel 4. Hasil NER dengan data testing 733 tweet

Entitas	LOCATION	ORGANIZATION	PERSON	QUANTITY	TIME
LOCATION	21				
ORGANIZATION		7			
PERSON			7		
QUANTITY				9	
TIME					11

Pada hasil NER disajikan dalam tabel confusion matrix seperti pada Tabel 4 Pada Tabel 5 terlihat bahwa jumlah kejadian yang terdeteksi lebih sedikit dari dataset hal ini dikarenakan banyaknya *tweet* yang tidak menyebutkan LOCATION, ORGANIZATION, PERSON, QUANTITY, ataupun TIME.

Tabel 5. True-False NER

Label	TRUE	FALSE	Total
LOCATION	21	10	31
ORGANIZATION	7	1	8
PERSON	7	4	11
QUANTITY	9	2	11
TIME	11	10	11
Total	55	27	72

Pada Tabel 5. Ditampilkan performa NER, yaitu dari 31 LOCATION yang terdeteksi, terdapat 21 LOCATION true, 10 yang false. Sedangkan ORGANIZATION, true 7 dan false 1. Untuk PERSON 7 untuk true dan 4 untuk false. QUANTITY terdapat 9 true dan 2 false. Dan pada TIME dapat dilihat, ada 11 true dan 10 false dari total 11.

Sedangkan akurasi tertinggi, diperoleh dari algoritma Naïve Bayes 80,5%, selanjutnya disusul *Random Forest*, *Linier Regression*, *SVM*, *KNN* dengan nilai yang sama yaitu 77,8%. Terakhir adalah *Decision tree* dengan akurasi 75%.

E. Evaluation

Setelah model komparasi algoritma klasifikasi dirasa telah baik hingga berikutnya merupakan sesi evaluation. Pada sesi ini merupakan sesi penilaian komparasi algoritma dengan parameter yang digunakan yakni Confusion Matrix yang pada dasarnya merupakan buat membagikan data perbandingan hasil yang sudah dicoba oleh model dengan hasil klasifikasi sesungguhnya dengan memandang nilai presisi serta recall. Berikut hasil penilaian pada model yang dihasilkan oleh algoritma ini.

Naive Bayes					Random Forest				
confusion matrix [[65 15] [17 77]]					confusion matrix [[69 11] [41 53]]				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.79	0.81	0.80	80	0	0.63	0.86	0.73	80
1	0.84	0.82	0.83	94	1	0.83	0.56	0.67	94
accuracy			0.82	174	accuracy			0.70	174
macro avg	0.81	0.82	0.82	174	macro avg	0.73	0.71	0.70	174
weighted avg	0.82	0.82	0.82	174	weighted avg	0.74	0.70	0.70	174
Linier Regression					SVM				
confusion matrix [[70 10] [31 63]]					confusion matrix [[67 13] [37 57]]				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.69	0.88	0.77	80	0	0.64	0.84	0.73	80
1	0.86	0.67	0.75	94	1	0.81	0.61	0.70	94
accuracy			0.76	174	accuracy			0.71	174
macro avg	0.78	0.77	0.76	174	macro avg	0.73	0.72	0.71	174
weighted avg	0.78	0.76	0.76	174	weighted avg	0.74	0.71	0.71	174
Decision Tree					KNN				
confusion matrix [[63 17] [39 55]]					confusion matrix [[56 24] [47 47]]				
	precision	recall	f1-score	support		precision	recall	f1-score	support
0	0.62	0.79	0.69	80	0	0.54	0.70	0.61	80
1	0.76	0.59	0.66	94	1	0.66	0.50	0.57	94
accuracy			0.68	174	accuracy			0.59	174
macro avg	0.69	0.69	0.68	174	macro avg	0.60	0.60	0.59	174
weighted avg	0.70	0.68	0.68	174	weighted avg	0.61	0.59	0.59	174

Gambar 2. Confusion Matrix, Presisi dan Recall Komparasi Algoritma.

Perhitungan akurasi, *precision*, *Recall*, *f1-score*, dan *support* menggunakan 6 algoritma yaitu, *Naïve Bayes*, *Random Forest*, *Logistic Regression*, *SVC*, *Decision Tree*, dan *KNN*.

Tabel 6. Performa Algoritme Naïve Bayes

	Precision	Recall	F1-score	Support
1.	0.8	1.00	0.89	28
2.	1.00	0.12	0.22	8

Tabel 7. Performa Algoritme *Random Forest*

	Precision	Recall	F1-score	Support
1.	0.78	1.00	0.88	28
2.	0.00	0.00	0.00	8

Tabel 8. Performa Algoritme *Logistic Regression*

	Precision	Recall	F1-score	Support
1.	0.78	1.00	0.88	28
2.	0.00	0.00	0.00	8

Tabel 9. Performa Algoritme *SVC*

	Precision	Recall	F1-score	Support
1.	0.8	1.00	0.89	28
2.	1.00	0.12	0.22	8

Tabel 10. Performa Algoritme *Decision Tree*

	Precision	Recall	F1-score	Support
--	-----------	--------	----------	---------

1.	0.8	1.00	0.89	28
2.	1.00	0.12	0.22	8

Tabel 11. Performa Algoritme *Decision Tree*

	Precision	Recall	F1-score	Support
1.	0.8	1.00	0.89	28
2.	1.00	0.12	0.22	8

Tabel 12. Performa Algoritme *KNN*

	Precision	Recall	F1-score	Support
1.	0.8	1.00	0.89	28
2.	1.00	0.12	0.22	8

F. Deployment

Setelah melakukan tahap evaluasi, dimana dilakukan evaluasi keseluruhan atau detail terhadap model algoritma, kemudian seluruh model yang dibangun diimplementasikan. Setelah itu, perubahan pada model algoritme klasifikasi untuk memungkinkannya agar dapat memberikan hasil yang konsisten dengan tujuan awal fase CRISP-DM.

4. KESIMPULAN

Metode NER dengan Spacy yang digunakan mampu melakukan pengenalan Named Entity Recognition (NER) yaitu LOCATION, ORGANIZATION, PERSON, QUANTITY, TIME. dengan hasil yang baik. Dari hasil komparasi 6 algoritma klasifikasi didapatkan bahwa algoritma naive bayes adalah algoritma terbaik untuk keduanya. Perhitungan precission, recall, dan f-measure mencapai nilai terbaik yaitu 80%, 100%, dan 89% untuk metode NER sedangkan hasil prediksi bencana dan non bencana sebesar 81,60%, 82% dan 82%.

Untuk pengembangan sistem kedepan, disarankan dapat mendeteksi kerugian dari bencana banjir yang dialami, sehingga menjadi bahan masukan untuk pemerintah pusat maupun daerah dalam penanggulangan bencana banjir.

REFERENSI

- [1] I. Fitriyaningsih dan Y. Basani, “Prediksi Kejadian Banjir dengan Ensemble Machine Learning Menggunakan BP-NN dan SVM,” *J. Teknol. dan Sist. Komput.*, vol. 7, no. 3, 2019.
- [2] F. Nonggala dan C. Faticah, “Klasifikasi jenis kejadian menggunakan kombinasi NeuroNER dan Recurrent Convolutional Neural Network pada data Twitter,” vol. 4, no. 2, hal. 81–90, 2018.
- [3] M. F. Rifai, H. Jatnika, dan B. Valentino, “Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS),” *PETIR*, vol. 12, no. 2, 2019, doi: 10.33322/petir.v12i2.471.
- [4] R. M. Yanti, I. Santoso, dan L. H. Suadaa, “Application of Named Entity Recognition via Twitter on SpaCy in Indonesian (Case Study : Power Failure in the Special Region of Yogyakarta),” *Indones. J. Inf. Syst.*, 2021, doi: 10.24002/ijis.v4i1.4677.
- [5] Y. Apridiansyah, N. D. M. Veronika, dan E. D. Putra, “Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu Menggunakan Metode Naive Bayes,” *JSAI (Journal Sci. Appl. Informatics)*, vol. 4, no. 2, 2021, doi: 10.36085/jsai.v4i2.1701.
- [6] A. M. Maksun, Y. A. Sari, dan B. Rahayudi, “Analisis Sentimen pada Twitter Bencana Alam di Kalimantan Selatan menggunakan Metode Naïve Bayes,” vol. 5, no. 12, hal. 5614–5621, 2021.
- [7] I. Fahrur Rozi, A. Taufika Firdausi, dan K. Islamiyah, “Analisis Sentimen Pada Twitter Mengenai Pasca Bencana Menggunakan Metode Naïve Bayes Dengan Fitur N-Gram,” *J. Inform. Polinema*, vol. 6, no. 2, hal. 33–39, 2020, doi: 10.33795/jip.v6i2.316.
- [8] V. Chandani dan R. S. Wahono, “Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film,” *J. Intell. Syst.*, vol. 1, no. 1, 2015.
- [9] J. Han, M. Kamber, dan J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- [10] D. A. Lestari dan D. Mahdiana, “Penerapan Algoritma K-Nearest Neighbor pada Twitter untuk Analisis Sentimen Masyarakat Terhadap Larangan Mudik 2021,” *Inform. J. Ilmu Komput.*, vol. 17, no. 2, 2021, doi: 10.52958/iftk.v17i2.3629.
- [11] R. Jayadi, H. M. Firmantyo, M. T. J. Dzaka, M. F. Suaidy, dan A. M. Putra, “Employee performance prediction using naïve bayes,” *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 6, 2019, doi: 10.30534/ijatcse/2019/59862019.
- [12] R. M. Yanti, I. Santoso, dan L. H. Suadaa, “Application of Named Entity Recognition via Twitter on SpaCy in Indonesian (Case Study : Power Failure in the Special Region of Yogyakarta),” vol. 4, no. 1, hal. 76–86, 2021.
- [13] R. R. Rizqiyah, L. Muflikhah, dan M. A. Fauzi, “Pengenalan Entitas Bernama untuk Identifikasi Transaksi Akuntansi Pengenalan Entitas Bernama untuk Identifikasi Transaksi Akuntansi Menggunakan Hidden Markov Model,” no. July, 2018.
- [14] P. Attewell dan D. Monaghan, “14. Association Rules,” *Data Min. Soc. Sci.*, vol. 8, no. 5, hal. 227–234, 2019, doi: 10.1525/9780520960596-015.
- [15] W. F. Oyewusi, O. Adekanmbi, I. Okoh, dan V. Onuigwe, “NaijaNER : Comprehensive Named Entity Recognition for 5 Nigerian Languages.,” 2021.
- [16] W. Liao dan S. Veeramachaneni, “A Simple Semi-supervised Algorithm For Named Entity Recognition,” no. June, hal. 58–65, 2009.