



## KLASIFIKASI SITUASI BENCANA ALAM BANJIR MENGUNAKAN SUPPORT VECTOR MACHINE BERDASARKAN DATA TWITTER

Ramses Caniago<sup>1</sup>, Muhammad Habibi<sup>2\*</sup>

<sup>1,2</sup> Program Studi Informatika, Universitas Jenderal Achmad Yani Yogyakarta

<sup>1,2</sup> Jl. Siliwangi, Ringroad Barat, Banyuraden, Gamping<sup>12</sup> Sleman 55293

e-mail : [ramusecaniago@gmail.com](mailto:ramusecaniago@gmail.com)<sup>1</sup>, [muhammadhabibi17@gmail.com](mailto:muhammadhabibi17@gmail.com)<sup>2</sup>

### ABSTRAK

*Bencana alam banjir sering terjadi di Indonesia karena banyaknya pulau dan iklim tropisnya. Setiap harinya tweet masyarakat mengenai banjir di Twitter bertambah banyak dan dapat mencapai ribuan hanya dalam beberapa hari saja. Tujuan dari penelitian ini adalah membuat suatu model klasifikasi untuk melakukan klasifikasi situasi bencana alam banjir berdasarkan data Twitter. Penelitian ini menggunakan algoritma Support Vector Machine (SVM) yang merupakan salah satu metode dari data mining untuk melakukan klasifikasi. Tahapan yang dilakukan dalam penelitian ini, diantaranya, pengumpulan data, pelabelan manual membagi data ke dalam tiga jenis situasi yaitu ringan, sedang, dan berat. Pembobotan menggunakan TF-IDF dan dilakukan proses training untuk menghasilkan sebuah model. Hasil pengujian model dengan confusion matrix dan K-fold cross validation menghasilkan nilai akurasi sebesar 90,61% dan nilai F1-score sebesar 90,64%. Hasil klasifikasi tweet terkait data banjir menunjukkan bahwa sebanyak 67,40% tweet masuk ke dalam kategori ringan, 19,79% tweet kategori sedang, dan 12,81% tweet kategori berat.*

**Kata kunci :** *Banjir, Data Mining, Klasifikasi, SVM, Twitter.*

### ABSTRACT

*Due to its numerous islands and warm environment, Indonesia frequently experiences flood natural disasters. Tweets about floods on Twitter grow every day and can reach thousands in a matter of days. This study's objective is to develop a classification model for categorizing flood-related natural catastrophe events using data from Twitter. The Support Vector Machine (SVM) algorithm, a data mining technique for categorizing, is used in this work. Data gathering, manual labeling, and segmenting the data into three categories—mild, moderate, and severe—were all steps taken in this study. A training process is carried out to create a model before weighting is applied. The accuracy value and F1-score obtained by evaluating the model using the confusion matrix and K-fold cross-validation are 90.61% and 90.64%, respectively. 67.40% of the tweets classified as having flood-related data fell into the light category, followed by 19.79% of tweets classified as medium tweets, and 12.81% of tweets classified as heavy tweets.*

**Keywords:** *Classification, Data Mining, Flood, SVM, Twitter*

### 1. PENDAHULUAN

Kemajuan teknologi internet sekarang ini dapat dimanfaatkan masyarakat dengan cara melakukan *posting* pada media sosial. Salah satu media sosial yang digunakan masyarakat adalah Twitter. Berdasarkan data dari Oberlo, pengguna aktif bulanan Twitter tahun 2020 sebanyak 192

juta pengguna aktif harian (Lin, 2021). Postingan pada Twitter disebut dengan *tweet*. Setiap harinya *tweet* masyarakat mengenai banjir pada Twitter bertambah banyak dan dapat mencapai ribuan hanya dalam beberapa hari saja. *Tweet* yang berjumlah banyak tersebut kemudian akan digunakan sebagai data analisis untuk mengetahui



seberapa parah situasi banjir di Indonesia. Penelitian ini bertujuan untuk melakukan klasifikasi *tweet* banjir berdasarkan dampak yang ditimbulkan saat banjir maupun setelah banjir.

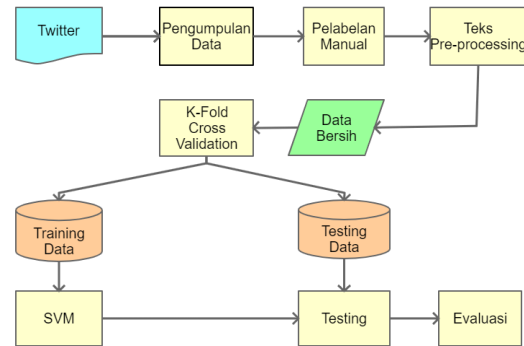
Metode yang digunakan adalah *Support Vector Machine* (SVM), keunggulan SVM dibandingkan dengan metode lain adalah kemampuan SVM untuk menentukan jarak antara kelas, sehingga dapat mempercepat proses komputasi. Penelitian yang dilakukan oleh (Akbar et al., 2020) melakukan klasifikasi sentimen publik terhadap isu perpindahan ibu kota di Indonesia. Penelitian lain yaitu melakukan analisis sentiment mengenai pelayanan Badan Penyelenggara Jaminan Sosial (BPJS) Kesehatan di Indonesia menggunakan SVM (Rahmawati & Habibi, 2020). Pada penelitian ini melakukan analisis sentimen *tweet* tentang pelayanan BPJS ke dalam 4 klasifikasi sentimen, yaitu puas, kecewa, senang dan sedih.

*Hoax* merupakan berita palsu atau isu yang sering beredar di masyarakat. Sebagai upaya pencegahan *hoax*, dilakukan penelitian untuk mengidentifikasi dan menganalisis *hoax* (Mubarq & Setiawan, 2020). Penelitian ini menggunakan metode SVM. Pada penelitian yang dilakukan oleh (Widiastuti, 2017) dijelaskan bahwa pemerintah DKI Jakarta pernah menggunakan media sosial untuk mengetahui perkembangan pasca bencana banjir di Jakarta.

Berikut beberapa penelitian yang menggunakan data banjir, diantaranya terkait prediksi curah hujan yang berpotensi banjir (Syamsurizal et al., 2022), prediksi banjir (Hasanah et al., 2021), dan klasifikasi mitigasi bencana banjir (Febriansyah et al., 2020). Penelitian yang dilakukan memiliki perbedaan dengan penelitian terdahulu, yaitu klasifikasi yang dilakukan menggunakan dukungan beberapa faktor seperti curah hujan, frekuensi (rentang waktu kejadian). Ketinggian air, fasilitas umum rusak, rumah rusak, korban terdampak, dan korban meninggal. Pemilihan faktor tersebut berdasarkan atribut yang digunakan oleh Badan Nasional Penanggulangan Bencana (BNPB) dan penelitian sebelumnya dalam pendataan banjir (Mendrofa, 2019). Tujuan dari penelitian ini adalah untuk memberikan informasi situasi banjir di Twitter, sehingga dapat dimengerti bahaya banjir dan dapat diketahui daerah dampak banjir yang membutuhkan prioritas bantuan.

**2. METODOLOGI PENELITIAN**

Penelitian ini dilakukan melalui beberapa tahapan. Tahapan penelitian dapat dilihat pada Gambar 1.



**Gambar 1.** Tahapan Penelitian

**a. Pengumpulan Data**

Pengumpulan data dilakukan dengan menggunakan *library* Tweepy pada Python. Tweepy digunakan untuk mengakses *Application Programming Interface* (API) Twitter sehingga dapat dilakukan pengumpulan data. Tahapan awal dari pengumpulan data adalah membuat token yaitu *consumer key*, *consumer secret*, *access token*, dan *access secret*. Keempat token tersebut dapat dibuat dengan cara membuat akun *twitter developer*. Setelah pengambilan token, token akan digunakan untuk autentikasi. Data Twitter dapat dikumpulkan menggunakan beberapa kriteria tertentu seperti kata kunci, tanggal, lokasi, dll. Pada penelitian ini kata kunci yang digunakan adalah banjir, banjir digunakan untuk mendapatkan *tweet* dalam Bahasa Indonesia. *List* yang telah berisi data *tweet* kemudian dapat dimasukan ke *file* berbentuk csv atau xlsx. Tabel 1 menunjukkan hasil dari pengumpulan data *sampel* yang digunakan sebagai contoh

**Tabel 1.** Contoh data *tweet*

<i>text</i>	<i>date</i>	<i>user</i>
@caceciii Udah diundang di mana mana chaa. Apalagi kalo ramadhan banjir jobb gak ada cb pas ramadhan	06/08/2021 09:37:05	pinkiemun
Tiku banjir doun	06/08/2021 09:36:38	kangensungjin



hari ini banjir konten lagi aku nya	06/08/2021 09:34:48	Lullabeehw
---	------------------------	------------

**b. Pelabelan Manual**

Pada tahapan ini data yang telah terkumpul diberikan sebuah label. Tujuan dari pelabelan tersebut adalah mengelompokan fitur kedalam kelas atau kategori yang sesuai (Wibawa et al., 2018). Label yang digunakan adalah angka yaitu satu, dua, dan tiga. Angka satu memiliki arti keadaan banjir yang ringan, dua banjir sedang, dan tiga banjir berat. Proses ini merupakan proses yang sangat penting karena dapat mempengaruhi hasil evaluasi SVM. Terdapat faktor pendukung yang digunakan dalam pelabelan manual yaitu curah hujan, frekuensi, ketinggian air, fasilitas umum rusak, rumah rusak, korban terdampak, dan korban meninggal. Tabel 2 merupakan contoh dari pelabelan manual.

**Tabel 2.** Contoh Pelabelan manual

text	label
Diguyur hujan deras, belasan rumah di Bojonggede terendam banjir <a href="https://t.co/t04rbmywif">https://t.co/t04rbmywif</a>	1
pemkab pesisir selatan mulai perbaiki jembatan ambruk akibat banjir bandang di Nagari Binjai Tapan	2
rt @beritarienews: banjir bandang ntt, data terbaru bnpb: 174 orang meninggal dunia, 48 orang hilang #ntt #prayfortnt	3

**c. Teks Preprocessing**

*Preprocessing* merupakan tahapan awal dimana dokumen akan ditransformasi terlebih dahulu kedalam bentuk yang berbeda. Transformasi digunakan untuk menganalisis frasa dan kata yang akan digunakan ataupun dihilangkan (Hickman et al., 2020) (Habibi & Winarko, 2017). Terdapat enam tahapan *preprocessing* yang digunakan pada penelitian ini yaitu:

- 1) *Cleaning data*, merupakan proses untuk menghilangkan *noise*. *Noise* akan dihilangkan dengan cara membersihkan teks dari karakter selain huruf seperti *tag* dan tanda baca.
- 2) *Case folding*, merupakan proses dimana semua karakter huruf pada dokumen akan diubah menjadi huruf kecil atau sebaliknya.

Tujuan dari tahap ini adalah mempermudah pencarian dokumen.

- 3) *Tokenizing*, merupakan proses dimana kalimat dipotong menjadi beberapa bagian berdasarkan kata perkata, kemudian potongan kata akan disebut dengan token. Tujuan dari pemotongan ini untuk mempermudah pemrosesan tiap unit (frasa dan kata).
- 4) *Stopword Removal*, merupakan proses membuang kata yang dianggap tidak berpengaruh pada makna kata (Nasron & Habibi, 2020). Proses ini dilakukan dengan daftar *stoplist*. *Stoplist* merupakan kumpulan kata yang tidak digunakan. Kumpulan kata tersebut akan digunakan sebagai kata kunci dan kemudian kata tersebut akan dihilangkan.
- 5) *Normalization*, merupakan proses dimana setiap unit akan diubah dari kata *slang* menjadi kata baku. Proses ini memerlukan sebuah dokumen yang memiliki kata kunci dan kata baku.
- 6) *Stemming*, merupakan proses merubah kata yang tidak baku menjadi baku. proses ini melakukan pengecekan tiap unit sehingga memerlukan waktu eksekusi yang lebih lama daripada proses lainnya.

**d. K-fold Cross Validation**

*Cross validation* merupakan metode statistik yang digunakan untuk menilai kemampuan suatu *machine learning*. Salah satu dari metode tersebut adalah *K-fold cross validation* (Roy & Saha, 2022). Pada metode tersebut terdapat sebuah parameter K yang merujuk kepada jumlah grup yang akan dipisahkan dari data sampel. Langkah awal untuk melakukan metode tersebut adalah mengacak *dataset*. Setelah data diacak, *dataset* dipisahkan menggunakan perbandingan tertentu menjadi K grup (Yan et al., 2022). Setiap grup memiliki sebagian data *test* dan data *train*. Langkah selanjutnya adalah menaruh model ke dalam *training set* dan mengevaluasi model pada data *testing* (Habibi & Kusumaningtyas, 2023). Selanjutnya akan dihasilkan skor evaluasi model. Skor evaluasi model tersebut dapat digunakan untuk mengukur kemampuan suatu model (Pisner & Schnyer, 2020).

**e. Training Data**

*Training data* akan dilakukan terhadap data yang telah diberikan label secara manual dan telah dibersihkan. Sebelum melakukan proses *training* perlu dilakukan *feature extraction* (Zhou et al., 2021). Pada *feature extraction* akan digunakan



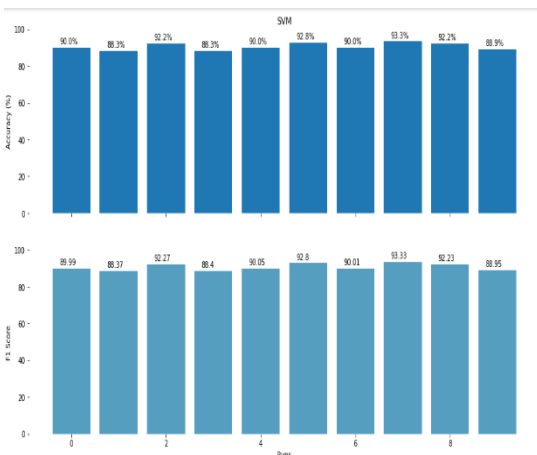
*Term Frequency- Inverse Document Frequency* (TF-IDF) untuk memberikan bobot pada tiap kalimat (Chen et al., 2023).

**f. Testing Data**

*Testing* data akan dilakukan dengan cara mengecek tingkat *accuracy*, *precision*, dan *recall*. Pengecekan dapat dilakukan dengan menggunakan *confusion matrix* (Maarif et al., 2023). *Confusion matrix* merupakan metode untuk menganalisis sebuah model klasifikasi. Proses analisis tersebut akan menghasilkan nilai *accuracy*, *precision*, serta *recall*. *Confusion matrix* berbentuk tabel yang kemudian didalamnya akan menyatakan jumlah data uji yang benar dan salah pada klasifikasi (Rahman et al., 2017).

**g. SVM**

Metode SVM merupakan metode yang digunakan pada salah satu algoritma *machine learning*. Sebagai salah satu algoritma *machine learning*, SVM memiliki dua tujuan utama yaitu mengoptimalkan akurasi dan memastikan klasifikasi dapat digunakan ke data baru (mengoptimalkan reproduktifitasnya)



**Gambar 2.** Grafik *K-fold cross validation* SVM

Berdasarkan Gambar 2. dapat dilihat bahwa data diiterasi hingga 10 kali. *Setiap* iterasi dapat memiliki nilai yang berbeda, hal ini dikarenakan perbedaan letak partisi data *test* saat proses iterasi. *Fold* ketujuh memiliki hasil *accuracy* dan *F1-score* tertinggi berturut-turut 93,3%, 93,33%. Data baru (*data testing*) juga dilakukan *K-fold cross validation*. Perbedaannya terdapat pada jumlah data yang digunakan, data baru menggunakan 10.000 data. Data baru dibagi menjadi 10 bagian, dengan perbandingan 90:10.

**3. HASIL DAN PEMBAHASAN**

Penelitian ini menggunakan data Twitter dengan kata kunci “banjir”. Data tersebut diambil pada kurun waktu antara tanggal 14 April sampai dengan 2 Juni 2021. *Data training* pada penelitian ini menggunakan 1800 data *tweet*. *Data training* tersebut dibagi kedalam tiga label yaitu ringan, sedang, dan berat, dengan jumlah data yang sama rata 600 data. Dalam pembuatan model klasifikasi menggunakan SVM kemudian digunakan sebagai model prediksi label pada data *tweet* yang baru. Data baru tersebut memiliki jumlah 10.000 *data tweet* dan tidak memiliki label didalamnya.

**a. Hasil Evaluasi K-Fold Cross Validation**

*K-fold cross validation* merupakan tahapan SVM dimana data akan dilipatkan menjadi *K-fold*. *K-fold cross validation* dilakukan untuk mengetahui tingkat akurasi dari tiap *fold*. Selain akurasi terdapat juga *F1-score* yang berguna untuk menyeimbangkan data. Pada penelitian ini data dilipat menjadi 10 kali dan kemudian ditampilkan menggunakan Matplotlib. Hasil dari *K-Fold cross validation* dapat dilihat pada Gambar 2.

**b. Hasil Evaluasi Confusion Matrix**

*Confusion matrix* adalah sebuah model yang dapat digunakan untuk mengukur *accuracy* dan *F1-score*. *Confusion matrix* merupakan hasil dari rata-rata *K-fold cross validation*. Hasil rata-rata tersebut berupa angka dalam bentuk desimal. *Confusion matrix* memiliki struktur tabel seperti pada Tabel 3.

**Tabel 3.** *Confusion matrix* model SVM

Klasifikasi yang benar	Diprediksi sebagai		
	Ringan (1)	Sedang (2)	Berat (3)
Ringan (1)	52,3	5,2	1,2
Sedang (2)	4	56,3	1,7
Berat (3)	1,7	3,1	54,5

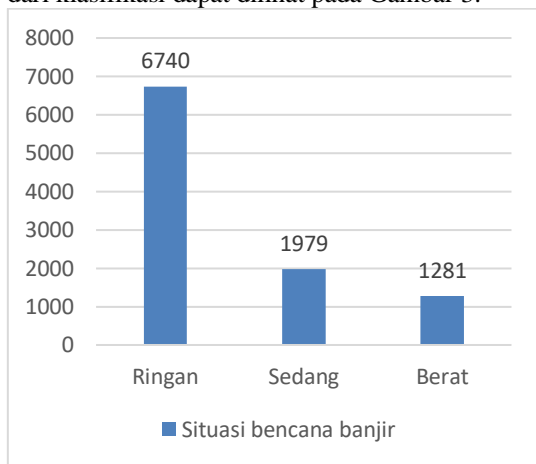
Data berjumlah 1800 yang telah dilakukan *K-fold cross validation* memiliki *confusion matrix* masing-masing disetiap iterasinya. *Confusion matrix* tersebut kemudian dijumlahkan dan dibagi sejumlah 10. Dhasilkan *confusion matrix* rata-rata seperti pada Tabel 3 di atas. Sebagai contoh, data yang diprediksi ringan, ternyata benar ringan berjumlah 52,3 (TP), dapat diartikan sebagai, rata-rata 523 data berhasil dan benar diprediksi ringan.



**c. Hasil Analisis**

.Pada penelitian ini diambil data terkait banjir dengan jumlah 11800 *data tweet*. *Data tweet* tersebut dipisahkan menjadi dua bagian yaitu *data training* dan data yang akan diklasifikasikan (data baru). *Data training* yang digunakan berjumlah 1800 *data tweet*, dan 10.000 *data tweet* digunakan untuk data baru.

*Data training* akan diberikan label terlebih dahulu. Jumlah label pada *data training* disamaratakan menjadi 600 *data tweet* tiap label. *Data training* kemudian digunakan sebagai dasar pembuatan model SVM. Data baru akan digunakan pada model tersebut (klasifikasi). Hasil dari klasifikasi dapat dilihat pada Gambar 3.



**Gambar 3.** Grafik hasil klasifikasi

Label ringan memiliki jumlah terbanyak dengan 6740 *data tweet* (67,40%). Label sedang memiliki jumlah 1979 *data tweet* (19,79%). Label berat memiliki jumlah paling sedikit dengan 1281 *data tweet* (12,81%). Dari jumlah data tersebut dapat disimpulkan bahwa situasi bencana banjir ringan memiliki frekuensi yang tinggi dan sering terjadi. Banjir ringan dapat terjadi meskipun memasuki musim peralihan pada bulan April hingga Juni. Banjir ringan dapat dikategorikan sebagai banjir yang selalu terjadi setiap tahunnya saat peralihan musim dan musim kemarau. Banjir sedang memiliki frekuensi yang rendah, sama seperti banjir berat. Banjir sedang dapat dikategorikan sebagai banjir yang jarang terjadi dan tidak muncul terus-menerus. Banjir berat dapat dikategorikan sebagai banjir yang hanya terjadi apabila terdapat kejanggalan contoh seperti anomali iklim yang terjadi pada Nusa Tenggara Timur (NTT) di bulan April (BNPB, 2021).

**4. KESIMPULAN**

Dalam penelitian ini, kami mengklasifikasikan data twitter terkait bencana alam banjir menggunakan algoritma SVM. Klasifikasi data banjir diklasifikasikan ke dalam tiga kelompok, yaitu ringan, sedang dan berat. Hasil Pengujian dengan *confusion matrix* dan, K-fold *cross validation* menghasilkan nilai akurasi sebesar 90,61% dan nilai dari F1-score sebesar 90,64%. Hasil presentase klasifikasi *tweet* masyarakat terkait data banjir pada Twitter menunjukkan bahwa dari 10.000 *data tweet*, sebanyak 67,40% *tweet* masuk ke dalam kategori ringan, 19,79% *tweet* kategori sedang, dan 12,81% *tweet* kategori berat. Berdasarkan hasil klasifikasi, dapat dilihat bahwa bencana alam banjir pada awal tahun 2021 di Indonesia didominasi bencana banjir dengan kriteria ringan.

**5. REFERENSI**

Abdelkader, H. E., Gad, A. G., Abohany, A. A., & Sorour, S. E. (2022). An Efficient Data Mining Technique for Assessing Satisfaction Level With Online Learning for Higher Education Students during the COVID-19. *IEEE Access*, *10*, 6286–6303. <https://doi.org/10.1109/ACCESS.2022.3143035>

Akbar, M. R., Slamet, I., & Handajani, S. S. (2020). Sentiment analysis using tweets data from Twitter of Indonesian’s Capital City changes using classification method support vector machine. *International Conference on Science and Applied Science (ICSAS2020)*, *2296*, 020041. <https://doi.org/10.1063/5.0030357>

BNPB. (2021). *Info Bencana: Data dan Informasi Kebencanaan Bulanan Teraktual, Edisi April 2021* (April 2021). Badan Nasional Penanggulangan Bencana.

Chen, M., Yao, C., Li, X., & Shen, L. (2023). A Text Classification Model Based on Gaussian Multi-Head Self Attention Mechanism for Chinese Medical Data. *2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms, EEBDA 2023*, 229–232. <https://doi.org/10.1109/EEBDA56825.2023.10090693>

Febriansyah, A., Ramadhan, A., Gustiawan, M., R, M. R., Maulana, R., Y, R. J., G.E, R., & Firmansyah, R. (2020). Penerapan Machine Learning Dalam Mitigasi Banjir Menggunakan Data Mining. *Jurnal*



- Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 3(3), 215–218.  
<https://doi.org/10.32672/JNKTI.V3I3.2427>
- Habibi, M., & Kusumaningtyas, K. (2023). Customer Experience Analysis Skincare Products Through Social Media Data Using Topic Modeling and Sentiment Analysis. *JOURNAL OF SCIENCE AND APPLIED ENGINEERING*, 6(1), 1–9.  
<https://doi.org/10.31328/JSAE.V6I1.4169>
- Habibi, M., & Winarko, E. (2017). Analisis Sentimen dan Klasifikasi Komentar Mahasiswa pada Sistem Evaluasi Pembelajaran Menggunakan Kombinasi KNN Berbasis Cosine Similarity dan Supervised Model. In *Departemen Ilmu Komputer dan Elektronika, Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Gadjah Mada*.
- Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, 5(2), 103–108.  
<https://doi.org/10.30871/JAIC.V5I2.3200>
- Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2020). Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations. *Organizational Research Methods*, October, 1–33.  
<https://doi.org/10.1177/1094428120971683>
- Lin, Y. (2021, January 25). *10 Twitter Statistics Every Marketer Should Know in 2021 [Infographic]*. Oberlo.  
<https://www.oberlo.com/blog/twitter-statistics>
- Maarif, M. R., Saleh, A. R., Habibi, M., Fitriyani, N. L., & Syafrudin, M. (2023). Energy Usage Forecasting Model Based on Long Short-Term Memory (LSTM) and eXplainable Artificial Intelligence (XAI). *Information 2023*, Vol. 14, Page 265, 14(5), 265.  
<https://doi.org/10.3390/INFO14050265>
- Mendrofa, Y. (2019). Implementasi Algoritma C4.5 Untuk Memprediksi Tingkat Kerusakan Akibat Banjir (Studi Kasus: BPBD Prov.Sumut). *Jurnal Pelita Informatika*, 7(4), 584–592.
- Mubaroq, I. M., & Setiawan, E. B. (2020). *The Effect of Information Gain Feature Selection for Hoax Identification in Twitter Using Classification Method Support Vector*. 5(2), 107–118.  
<https://doi.org/10.21108/indojc.2020.5.2.499>
- Nasron, U. A., & Habibi, M. (2020). Analysis of Marketplace Conversation Trends on Twitter Platform Using K-Means. *Compiler*, 9(1), 51–61.  
<https://doi.org/10.28989/compiler.v9i1.579>
- Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In A. Mechelli & S. Vieira (Eds.), *MACHINE LEARNING Methods and Applications to Brain Disorders* (pp. 101–121). Elsevier.  
<https://doi.org/10.1016/B978-0-12-815739-8.00006-7>
- Rahman, M. F., Alamsah, D., Darmawidjaja, M. I., & Nurma, I. (2017). Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN). *Jurnal Informatika*, 11(1), 36.  
<https://doi.org/10.26555/jifo.v11i1.a5452>
- Rahmawati, S., & Habibi, M. (2020). Public Sentiments Analysis about Indonesian Social Insurance Administration Organization on Twitter. *IJID (International Journal on Informatics for Development)*, 9(2), 87–93.  
<https://doi.org/10.14421/IJID.2020.09205>
- Roy, J., & Saha, S. (2022). Ensemble hybrid machine learning methods for gully erosion susceptibility mapping: K-fold cross validation approach. *Artificial Intelligence in Geosciences*, 3, 28–45.  
<https://doi.org/10.1016/J.AIG.2022.07.001>
- Schnyer, D. M., Clasen, P. C., Gonzalez, C., & Beevers, C. G. (2017). Evaluating the diagnostic utility of applying a machine learning algorithm to diffusion tensor MRI measures in individuals with major depressive disorder. *Psychiatry Research: Neuroimaging*, 264, 1–9.  
<https://doi.org/10.1016/j.psychresns.2017.03.003>
- Syamsurizal, Cumel, Zamri, D., & Rahmaddeni. (2022). Perbandingan Metode Data Mining untuk Prediksi Banjir Dengan Algoritma Naïve Bayes dan KNN. *SENTIMAS: Seminar Nasional Penelitian Dan Pengabdian Masyarakat*, 40–48.  
<https://journal.irpi.or.id/index.php/sentimas/article/view/353>



- Wibawa, A. P., Guntur, M. P. A., Akbar, M. F., & Dwiyanto, F. A. (2018). *Metode-metode Klasifikasi*. 3(1), 134–138.
- Widiastuti, W. (2017). Penggunaan Twitter untuk Mendeteksi Banjir Melalui Pendekatan Text Mining dan Evaluasinya. *SEMINAR NASIONAL STATISTIKA FMIPA UNPAD 2017 (SNS VI)*, 47–54.
- Yan, T., Shen, S. L., Zhou, A., & Chen, X. (2022). Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm. *Journal of Rock Mechanics and Geotechnical Engineering*, 14(4), 1292–1303. <https://doi.org/10.1016/J.JRMGE.2022.03.002>
- Zhou, F., Zhang, Z., & Chen, D. (2021). Real-time fault diagnosis using deep fusion of features extracted by PeLSTM and CNN. *Fault Diagnosis and Prognosis Techniques for Complex Engineering Systems*, 353–401. <https://doi.org/10.1016/B978-0-12-822473-1.00003-3>