

# Sistem Deteksi Konten Negatif pada Teks Website Menggunakan Metode Random Forest

Aldiva Wibowo<sup>1</sup>, Hermawan Syahputra<sup>2</sup>

<sup>1,2</sup>Program Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Universitas Negeri Medan Jl. William Iskandar Ps. V, Kenangan Baru, Kec. Percut  
Sei Tuan, Kabupaten Deli Serdang, Sumatera Utara 20221  
e-mail: [aldivaawibowo@gmail.com](mailto:aldivaawibowo@gmail.com)<sup>1</sup>, [hsyahputra@unimed.ac.id](mailto:hsyahputra@unimed.ac.id)<sup>2</sup>

## ABSTRAK

Internet saat ini telah mengalami perkembangan yang signifikan sebagai sarana untuk mencari informasi yang memberikan dampak positif bagi kehidupan sehari-hari masyarakat. Namun, di samping itu, internet juga menjadi tempat yang efektif untuk menyebarkan informasi yang memberikan dampak negatif, seperti penyebaran konten negatif kepada masyarakat. Untuk mengatasi masalah ini, pemerintah telah menetapkan sistem laporan dan DNS (Domain Name System) yang dikenal sebagai Trust Positif. Namun, upaya tersebut masih tergolong tidak efektif karena hanya didasarkan pada DNS dan laporan masyarakat. Oleh karena itu, tujuan dari penelitian ini adalah untuk menggunakan metode Machine Learning untuk membuat model klasifikasi konten negatif website berdasarkan isi teks website. Metode yang digunakan dalam penelitian ini adalah Random Forest. Proses penelitian dimulai dengan melakukan preprocessing pada data teks dengan pendekatan NLP (Natural Language Processing) dan kemudian mengaplikasikan metode Machine Learning. Hasil dari pengujian menunjukkan bahwa Random Forest cukup efektif dalam penelitian ini, dengan nilai accuracy sebesar 92%, precision sebesar 77%, dan recall sebesar 86%. Nilai tersebut didapatkan melalui pengujian dengan menggunakan 526 url website.

**Kata Kunci:** Machine Learning, Random Forest, Konten Negatif, Klasifikasi, Teks

## I. PENDAHULUAN

Internet saat ini memberikan banyak informasi pendidikan yang digunakan masyarakat global, namun juga menyebarkan informasi dari berbagai sumber yang positif dan negatif. Internet menjadi bagian tidak terpisahkan dari kehidupan modern. Keberhasilan teknologi ini telah memberikan hasil yang positif. Namun, ada dampak negatif yang juga muncul, yaitu adanya sumber data yang tidak baik di internet, seperti aktivitas penipuan, perjudian, dan konten yang tidak sesuai untuk semua kalangan [1]. Elemen negatif di internet menyulitkan pengguna untuk menentukan situs website yang positif atau negatif. Konten negatif dapat memberikan dampak buruk bagi pengguna.

Pemerintah melalui KOMINFO mengatasinya dengan membuat Blacklist TrustPositif yang dapat menyaring website yang berisikan konten negatif seperti penipuan, perjudian, dan konten dewasa. Trust Positif memblokir website yang dilaporkan dan pengguna dapat melaporkan website dengan konten negatif melalui fitur aduan di dalamnya. Trust Positif saat ini sudah melakukan pemblokiran pada banyak website yang berbahaya. Namun, masih kurang efektif karena proses penerimaan aduan dari masyarakat dan instansi terkait, disesuaikan dengan dataset dari server pusat, sebelum akhirnya dilakukan pemblokiran pada website yang diadukan. Metode machine learning dan deep learning telah banyak diaplikasikan di dalam teks oleh peneliti [2]–[5].

Inovasi untuk permasalahan deteksi website yang memiliki konten negatif dibutuhkan untuk mendeteksi website yang dicurigai terdapat konten teks negatif

didalamnya sekaligus mempermudah proses melakukan pemblokiran berdasarkan data aktual. Pendekatan yang digunakan untuk menyelesaikan permasalahan adalah membuat model machine learning yang dapat melakukan klasifikasi dari url website terhadap setiap isi konten pada website berdasarkan isi teks nya. Machine Learning adalah ilmu dan seni dari pemrograman komputer yang memberikan kemampuan komputer untuk dapat belajar dari data [6]. Machine Learning mempelajari dan mengekstrak pengetahuan dari data. Salah satu bidang di persimpangan statistik, kecerdasan buatan, dan ilmu komputer dan juga yang dikenal sebagai analisis prediktif atau pembelajaran statistik [7].

Algoritma Machine Learning yang digunakan untuk membuat model mendeteksi konten negatif berdasarkan teks pada website pada penelitian ini adalah Random Forest. Berdasarkan beberapa penelitian yang sudah menggunakan Random Forest pada data teks sebelumnya [8]–[10]. Random Forest merupakan gabungan dari pohon keputusan yang digunakan untuk membuat model yang lebih kuat dan memiliki kinerja generalisasi yang lebih baik dengan cara menghindari overfitting dengan mengambil rata-rata beberapa pohon keputusan yang memiliki varians tinggi [11]. Pada dasarnya metode Random Forest ini membuat banyak pohon keputusan yang kemudian mengambil keputusan berdasarkan beberapa pohon keputusan yang sudah dibuat.

Input data yang digunakan untuk mendeteksi apakah konten didalam suatu website adalah Uniform Resource Locator (URL) atau link yang mengarahkan kepada website target untuk melakukan web scraping. Proses pengambilan data dari internet yang disebut sebagai Web Scraping, melibatkan ekstraksi dokumen

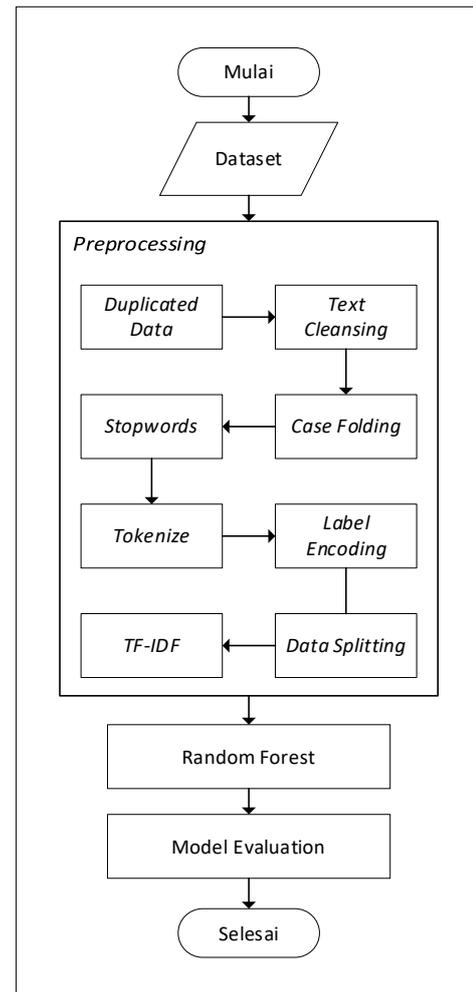
semi-terstruktur dari website. Dokumen tersebut umumnya ditulis dalam bahasa markup seperti HTML atau XHTML, yang menggunakan tag untuk mendefinisikan elemen dalam dokumen. Setelah di ambil, dokumen tersebut dianalisis untuk mengekstrak data yang diperlukan untuk tujuan tertentu, seperti analisis bisnis atau pengambilan keputusan [12].

Penelitian ini menggunakan pendekatan Natural Language Processing (NLP) untuk melakukan tahapan preprocessing pada data teks. NLP adalah metode analisis teks yang menggunakan komputer untuk memahami dan mengolah bahasa alami. Implementasi NLP meliputi pengumpulan pengetahuan tentang cara manusia memahami dan menggunakan bahasa, mengembangkan alat dan teknik yang tepat untuk memahami dan mengolah bahasa alami dan membuat sistem komputer yang dapat melakukan tugas-tugas yang ditetapkan. Tahap preprocessing yang dilakukan melalui NLP ini diharapkan dapat membantu dalam menganalisis dan mengolah data teks agar lebih mudah dipahami dan diolah oleh sistem komputer [13].

Berdasarkan dari permasalahan yang ada, Penelitian ini dirancang untuk mengatasi permasalahan yang ada dengan mengembangkan sistem pendeteksi konten negatif pada teks website. Algoritma Machine Learning Random Forest digunakan sebagai mesin klasifikasi pada setiap input data teks dari website yang ditentukan. Input yang digunakan dalam sistem ini adalah URL atau link yang mengarah pada website target untuk proses web scraping. Proses ini meliputi tahapan preprocessing yang akan mengklasifikasikan apakah website yang diinput diklasifikasikan sebagai website yang memiliki konten negatif atau website yang aman dari konten negatif. Implementasi teknologi Machine Learning ini diharapkan dapat membantu dalam mendeteksi suatu website yang memiliki konten negatif atau website yang aman dengan cara menganalisis URL website tujuan.

## II. METODE PENELITIAN

Metode penelitian yang diterapkan pada penelitian ini diawali dengan melakukan akuisisi data untuk dataset yang digunakan dalam pelatihan *machine learning* yang digunakan dalam penelitian ini yaitu *Random Forest*. Setelah itu melakukan preprocessing pada dataset teks yang telah disiapkan dengan tujuan untuk membersihkan data teks dari simbol, angka, dan karakter yang tidak dibutuhkan sehingga dapat dilakukan pembobotan pada kata. Dilanjutkan dengan tahap melakukan pelatihan menggunakan *Random Forest* setelah itu tahap terakhir adalah melakukan evaluasi dan melakukan test pada data.



Gambar 1. Diagram Alir

### 1. Dataset

Dataset yang digunakan pada penelitian ini berupa dataset teks yang didapatkan dari hasil webscraping link pada website TrustPositif yang menyimpan link yang tergolong kedalam website yang memiliki konten negatif didalamnya. Data yang digunakan akan menjadi fondasi dari machine learning yang dirancang untuk dapat melakukan klasifikasi website yang memiliki konten negatif didalamnya berdasarkan dari isi teks. Jumlah dataset teks yang digunakan adalah 2628 dengan 4 kelas yang terdiri dari kelas Negatif yaitu : Penipuan, Pornografi, dan Perjudian, dan Kelas Aman yaitu Whitelist. dataset terdiri dari kolom teks website, label, link, dan isi teks dari website.

### 2. Preprocessing

Preprocessing adalah tahap penting dalam pemrosesan teks. Sebuah teks dapat terdiri dari kata-kata, kalimat, dan paragraf. Sebuah urutan karakter yang bermakna dianggap sebagai teks. Untuk memberikan data teks ke algoritma pembelajaran mesin dalam bentuk yang lebih baik daripada dalam bentuk alami [14]. Sesuai dari diagram alir metode penelitian pada gambar 1. Alur tahapan preprocessing yang digunakan :

#### 2.1. Duplicated Data

Data duplikat adalah data yang ada lebih dari satu dalam dataset. Hal ini akan menyebabkan masalah dalam training model dan harus dihapus untuk menjaga kualitas model.

## 2.2. Text Cleansing

Adalah tahap untuk menyingkirkan karakter yang tidak diperlukan dari dataset teks. Karakter-karakter seperti angka, non-ascii, dan simbol akan dihapus, sehingga hanya kata-kata yang tersisa.

## 2.3. Case Folding

Case Folding merupakan proses pre-processing yang penting untuk dilakukan pada dataset teks. Dengan menggunakan Case Folding, seluruh teks diubah menjadi huruf kecil yang sama, sehingga tidak ada perbedaan nilai antara huruf besar dan kecil yang dapat mempengaruhi hasil prediksi.

## 2.4. Stopwords

Stopwords adalah kata-kata yang diabaikan dalam proses machine learning karena terlalu umum digunakan dan perlu dikeluarkan. Stopwords dapat ditemukan dalam daftar stoplist yang sudah tersedia untuk setiap bahasa atau dapat ditambahkan sesuai kebutuhan.

## 2.5. Tokenize

Tokenisasi digunakan untuk memecah kalimat, paragraf, atau dokumen menjadi token-token individual yang akan diberi nilai pada dataset teks.

## 2.6. Label Encoding

Label encoding adalah teknik preprocessing yang digunakan untuk mengubah data kategorikal nomor berurutan. Sebagai inputan label pada *machine learning*.

## 2.7. Data Splitting

Merupakan Proses pembagian data menjadi data latih dan data uji dengan rasio yang ditentukan yaitu 70/30. Data akan digunakan untuk melatih dan validasi pada model.

## 2.8. TF-IDF

TF-IDF (Term Frequency — Inverse Document Frequency) adalah metode yang digunakan untuk menghitung bobot kata dalam dataset teks dengan menghitung Term Frequency (TF) dan Inverse Document Frequency (IDF) pada setiap token di setiap dokumen dalam korpus.

## 3. Random Forest

*Random Forest* adalah algoritma *Machine Learning* yang menggabungkan teknik pohon keputusan dan bagging untuk membuat prediksi, dengan cara membagi data menjadi cabang-cabang sampai kriteria stop terpenuhi dan membuat prediksi dengan bergerak melalui simpul dan cabang [15]. *Random Forest* adalah metode yang menggabungkan banyak model pohon keputusan dalam satu hutan pohon. Model ini lebih mudah diterjemahkan dibandingkan dengan model parametrik

lainnya. Dengan mengkombinasikan hasil dari beberapa model pohon keputusan, diperoleh model prediksi yang lebih komprehensif dan akurat [16]. Tahapan perhitungan diawali dengan mencari fitur yang akan menjadi akar dari pohon keputusan. Untuk mencari fitur yang menjadi akar tahap pertama adalah mendapatkan nilai *entropy* setelah mendapatkan nilai dari *entropy* dilanjutkan dengan menghitung *information gain* dari hasil *information gain* pada data maka dapat melakukan generalisasi pohon keputusan. Secara umum formula untuk *entropy* dan *information gain* adalah sebagai berikut:

$$Entropy(S) = \sum_{i=1}^n P_i \log_2 P_i \quad (1)$$

$S$  = Himpunan Kasus

$P_i$  = Proporsi dari  $S_i$  terhadap  $S$

$n$  = Jumlah partisi  $S$

$$Gain(S, A) = E(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} E(S_i) \quad (2)$$

$E$  = Entropy

$S$  = Himpunan Kasus

$A$  = Atribut

$n$  = Jumlah partisi atribut  $A$

$|S_i|$  = Jumlah kasus pada partisi ke  $i$

$|S|$  = Jumlah kasus dalam  $S$

## 4. Model Evaluation

Untuk menentukan atau mengukur performa dari algoritma pembelajaran yang digunakan, hasil yang diperoleh dari algoritma tersebut perlu di evaluasi dan dianalisis dengan benar. Evaluasi ini dapat dilakukan dengan menggunakan berbagai metrik yang berbeda, seperti accuracy, sensitivity, dan specificity. Metrik ini dapat diperoleh dari nilai scalar *confusion matrix* yang diwakili dari hasil klasifikasi yang diperoleh dari algoritma tersebut. Beberapa metrik yang dapat digunakan untuk mengevaluasi performa dari algoritma pembelajaran yang digunakan yaitu precision, recall, F1-score, ROC, Informedness, Markedness dan metode penilaian Correlation [17]. *Confusion Matrix* adalah sebuah tabel yang digunakan untuk mengevaluasi performa model klasifikasi. Tabel ini menyajikan jumlah prediksi yang benar dan salah dari model dibandingkan dengan nilai sebenarnya. *Confusion matrix* menampilkan dua dimensi yaitu prediksi yang benar dan salah (True Positive, False Positive, True Negative, False Negative) yang digunakan untuk menghitung metrik seperti accuracy, precision, recall, F1-score. Dengan menggunakan *confusion matrix*, kita dapat mengetahui tingkat kinerja model dalam mengklasifikasikan data dan juga dapat menemukan kelemahan model dalam mengklasifikasikan data tertentu [18].

Tabel 1. Confusion Matrix

		Data Prediksi	
		Kelas Positif	Kelas Negatif
Data Aktual	Kelas Positif	TP True Positive	FN False Negative
	Kelas Negatif	FP False Positive	TN True Negative

- 1.1. True Positive (TP): model benar memprediksi kelas positif.
- 1.2. True Negative (TN): model benar memprediksi kelas negatif.
- 1.3. False Positive (FP): model salah memprediksi kelas positif.
- 1.4. False Negative (FN): model salah memprediksi kelas negatif.

Nilai accuracy, recall, precision, dan F1-score dapat diketahui dengan rumus berikut:

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

- Precision =  $\frac{TP}{(TP+FP)}$  (4)

- Recall =  $\frac{TP}{(TP+FN)}$  (5)

- F1 score =  $2 \times \frac{(Recall \times Precision)}{(Recall+Precision)}$  (6)

### III. HASIL DAN PEMBAHASAN

Bagian ini membahas mengenai hasil penelitian yang dilakukan dalam penelitian.

#### 1. Data Preprocessing

Proses preprocessing data teks dilakukan mulai dari menghapus data ganda, lalu melakukan *text cleansing* dari proses ini dihasilkan data teks yang tidak memiliki angka ataupun simbol.

Tabel 2. Teks Sebelum Cleansing

	Category	Text
0	Perjudian	Lihat kumpulan daftar nama situs agen judi pok...
1	Perjudian	Togel Online   Bandar Togel   Judi Togel   Age...
2	Perjudian	Akun Togel SG45   SG49   HK   KL   Klik4d Daft...
3	Whitelist	MEDAN, Sebuah warung internet (warnet) yang...

Tabel 3. Teks setelah tahap *Cleansing*

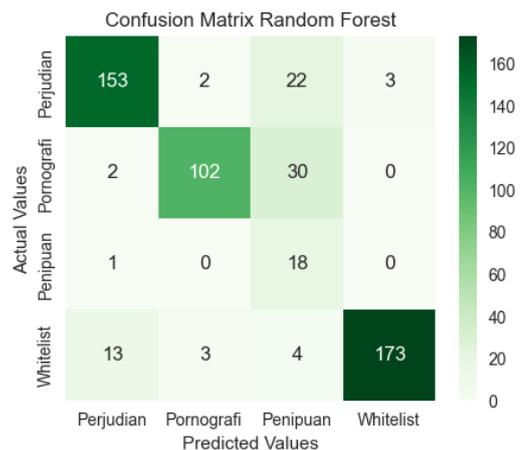
	Category	Text
0	Perjudian	lihat kumpul daftar nama situs agen judi poker...
1	Perjudian	togel online bandar togel judi togel agen toge...
2	Perjudian	akun togel sg sg hk kl klikd daftar akun togel...
3	Whitelist	medan warung internet warnet ada jalan juang l...

#### 2. Tahap Training

Training model Random Forest dengan dataset dilakukan setelah melewati proses yang harus dilakukan sebelumnya yaitu preprocessing pada teks, pembobotan pada kata, pembagian data untuk data latih dan data uji. Setelah tahap training selesai selanjutnya adalah tahap evaluasi.

#### 3. Evaluasi

Evaluasi dilakukan setelah tahap Training selesai dilakukan. Hasil evaluasi yang didapatkan dari uji coba model dengan data test diambil dari dataset berjumlah 526 ditampilkan kedalam confusion matrix pada gambar 2.



Gambar 2. Confusion Matrix Random Forest

Berdasarkan dari confusion matrix untuk perjudian, pornografi, dan whitelist menunjukkan jumlah yang cukup besar dibandingkan penipuan. Hal ini dikarenakan data untuk kelas penipuan lebih sedikit dibandingkan kelas lain. Walau begitu jumlah prediksi yang benar untuk penipuan cukup baik dari total data uji untuk kelas penipuan itu sendiri yaitu berjumlah 19. Kondisi ini terjadi dikarenakan jumlah data yang tidak seimbang antara kelas penipuan dengan kelas lainnya. Nilai metrik accuracy, precision, recall, dan f1-score yang didapatkan model setelah evaluasi dapat dilihat pada tabel berikut.

Tabel 4. Nilai Metrik Model

	Metrik	Nilai
1	Accuracy	0.923954373
2	Precision	0.771198565
3	Recall	0.863733877
4	F1-Score	0.762007503

#### IV. PENUTUP

Berdasarkan dari hasil analisis, perancangan sistem, dan evaluasi yang dilakukan. Mengaplikasikan machine learning dalam mendeteksi konten negatif berdasarkan isi teks pada website dapat bekerja pada website berbahasa indonesia. Random Forest terbukti dapat melakukan klasifikasi pada website yang di dalamnya terdapat konten negatif berdasarkan isi teks website. Dari evaluasi yang dilakukan pada model didapatkan nilai metrik Random Forest untuk kasus mendeteksi konten negatif yaitu Accuracy 92%, Precision 77%, Recall 86%, dan F1-Score 76%. Sistem yang dirancang dapat membantu dalam mendeteksi konten negatif berdasarkan isi teks pada website.

#### REFERENSI

- [1] R. Tullah, S. Maisaroh, and I. Sumpena, "Implementasi Machine Learning pada Aplikasi Pendeteksi Konten Pornografi," *Acad. J. Comput. Sci. Res.*, vol. 2, no. 2, 2020.
- [2] B. S. Sundar, "Emotion Detection on text using Machine Learning and Deep Learning Techniques," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 6. International Journal for Research in Applied Science and Engineering Technology (IJRASET), pp. 2277–2286, 2022. doi: 10.22214/ijraset.2022.44293.
- [3] C. Sharma, P. Kumar, and M. P. Singh, "Text-Mining Using Three Different Machine Learning Algorithms To Segregate Spam Messages From Non-Spam Messages," *SSRN Electronic Journal*. Elsevier BV, 2022. doi: 10.2139/ssrn.4159491.
- [4] R. C. Yeung and M. A. Fernandes, "Machine learning to detect invalid text responses: Validation and comparison to existing detection methods," *Behav. Res. Methods*, vol. 54, no. 6, pp. 3055–3070, Dec. 2022, doi: 10.3758/s13428-022-01801-y.
- [5] E. Asmawati, A. Saikhu, and D. Siahaan, "Sentiment Analysis of Text Memes: A Comparison Among Supervised Machine Learning Methods," in *2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, 2022, pp. 349–354. doi: 10.23919/EECSI56542.2022.9946506.
- [6] A. Gron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 1st ed. O'Reilly Media, Inc., 2017.
- [7] A. C. Muller and S. Guido, *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media, Incorporated, 2018. [Online]. Available: <https://books.google.co.id/books?id=jGdXswEACAAJ>
- [8] M. Zaroni, "Classification of NPL with a Random Forest approach," *Risk Management Magazine*, vol. 1, no. 2020. Italian Association of Financial Industry Risk Managers (AIFIRM), pp. 38–49, 2020. doi: 10.47473/2020rmm0007.
- [9] A. K. V. Mangapalli, and B. Malik, "Cucconi Feature Extracted Random Decision Forest Classification for Efficient Sentiment Analysis," *SSRN Electronic Journal*. Elsevier BV, 2022. doi: 10.2139/ssrn.4158220.
- [10] J. I. Sheeba, S. Pradeep Devaneyan, and V. G., "Detection of Spambot using Random Forest Algorithm." Rochester, NY, Oct. 02, 2019. doi: 10.2139/ssrn.3462968.
- [11] S. Raschka and V. Mirjalili, *Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-Learn, and TensorFlow 2, 3rd Edition*. Packt Publishing, 2019.
- [12] M. Turland, *Php-Architect's Guide to Web Scraping*. Marco Tabini & Associates, Incorporated, 2010.
- [13] S. Joseph, K. Sedimo, F. Kaniwa, H. Hlomani, and K. Letsholo, "Natural Language Processing: A Review," *Nat. Lang. Process. Rev.*, vol. 6, pp. 207–210, Mar. 2016.
- [14] D. Ramachandran and R. Parvathi, "Analysis of Twitter Specific Preprocessing Technique for Tweets," *Procedia Comput. Sci.*, vol. 165, pp. 245–251, 2019, doi: <https://doi.org/10.1016/j.procs.2020.01.083>.
- [15] A. Sekulić, M. Kilibarda, G. B. M. Heuvelink, M. Nikolić, and B. Bajat, "Random Forest Spatial Interpolation," *Remote Sens.*, vol. 12, no. 10, 2020, doi: 10.3390/rs12101687.
- [16] J. Zhou, E. Li, H. Wei, C. Li, Q. Qiao, and D. J. Armaghani, "Random Forests and Cubist Algorithms for Predicting Shear Strengths of Rockfill Materials," *Appl. Sci.*, vol. 9, no. 8, Art. no. 8, Jan. 2019, doi: 10.3390/app9081621.
- [17] S. Shaikh, "Measures Derived from a 2 x 2 Table for an Accuracy of a Diagnostic Test," *J. Biom. Biostat.*, vol. 02, Jan. 2011, doi: 10.4172/2155-6180.1000128.
- [18] D. A. Munawwaroh and A. H. Primandari, "Implementasi Crisp-Dm Model Menggunakan Metode Decision Tree Dengan Algoritma Cart Untuk Prediksi Lila Ibu Hamil Berpotensi Gizi Kurang," *Delta J. Ilm. Pendidik. Mat.*, vol. 10, no. 2, Art. no. 2, Aug. 2022, doi: 10.31941/delta.v10i2.2172.