

# Klasifikasi Produktivitas Pekerja Garmen Menggunakan Algoritma Random Forest

**Fiqki Haidar Amrulloh<sup>\*1</sup>, Ghilman Farhani Putra Aji<sup>2</sup>, Rio Ghaniy S.<sup>3</sup>, Vincentius Sagi  
Alban Anindyajati<sup>4</sup>, Luthfi Rakan Nabila<sup>5</sup>, Henri Tantyoko<sup>6</sup>**

<sup>1,2,3,4,5</sup>Sains Data Institut Teknologi Telkom Purwokerto,

<sup>6</sup>Institut Teknologi Telkom Purwokerto

E-mail: <sup>\*1</sup>[20110005@ittelkom-pwt.ac.id](mailto:20110005@ittelkom-pwt.ac.id), <sup>2</sup>[20110006@ittelkom-pwt.ac.id](mailto:20110006@ittelkom-pwt.ac.id),  
<sup>3</sup>[20110016@ittelkom-pwt.ac.id](mailto:20110016@ittelkom-pwt.ac.id), <sup>4</sup>[21110037@ittelkom-pwt.ac.id](mailto:21110037@ittelkom-pwt.ac.id), <sup>5</sup>[henri@ittelkom-pwt.ac.id](mailto:henri@ittelkom-pwt.ac.id)

## **Abstrak**

*Penelitian ini membahas mengenai penerapan algoritma machine learning dalam melakukan klasifikasi produktivitas pekerja garmen. Penelitian ini menggunakan dataset produktivitas garmen yang didapatkan dari situs UC Irvine Machine Learning Repository dengan rentang waktu dari tanggal 1 januari 2015 sampai 11 maret 2015 dengan total data sebanyak 1197 baris. Algoritma yang diterapkan pada penelitian ini adalah random forest dengan hyperparameter tuning untuk melakukan klasifikasi produktivitas pekerja garmen. Metodologi penelitian ini melibatkan pengolahan data seperti pemilihan fitur yang relevan, transformasi data, dan normalisasi guna mendapatkan hasil evaluasi terbaik. pada penelitian ini juga dilakukan percobaan dengan decision tree dan algoritma svm sebagai pembandingnya. Algoritma random forest mengungguli algoritma lain dengan akurasi sebesar 94.36% di mana akurasi tersebut sudah cukup bagus dalam untuk mengklasifikasi produktivitas.*

**Kata Kunci :** Klasifikasi, Garmen, Produktivitas, Hyperparameter Tuning, Random Forest

## **Abstract**

*This research discusses the application of machine learning algorithms in classifying garment worker productivity. This research uses a garment productivity dataset obtained from the UC Irvine Machine Learning Repository website with a time span from January 1 2015 to March 11 2015 with a total of 1197 rows of data. The algorithm applied in this research is random forest with hyperparameter tuning to classify garment worker productivity. This research methodology includes data processing such as selecting relevant features, data transformation, and normalization in order to obtain the best evaluation results. In this research, experiments were also carried out with the decision trees and SVM algorithm as a comparison. The random forest algorithm outperforms other algorithms with an accuracy of 94.36%, which is quite good accuracy to classify productivity.*

**Keywords :** Classification, Garment, Productivity, Hyperparameter Tuning, Random Forest

## **1. PENDAHULUAN**

Dalam hidup sehari-hari, manusia membutuhkan tiga kebutuhan pokok yang biasa dikenal dengan kebutuhan primer, yaitu makan, tempat tinggal, dan pakaian. Dengan itu, industri garmen menempatkan

dirinya sebagai salah satu industri yang dibutuhkan oleh manusia. Produk garmen saat ini diproduksi dalam volume yang lebih besar dibandingkan sebelumnya, bahkan setelah memperhitungkan pertumbuhan populasi, dan produk-produk yang dibuang dalam jumlah besar [1]. Sementara itu, adapun tantangan yang dihadapi oleh industri garmen fokus terhadap tingginya

kompetisi antar perusahaan dan aktor yang terlibat seperti berbagai jenis stakeholder yang memiliki prioritas yang berbeda [2]. Stakeholder pada industri garmen sendiri dapat meliputi pemegang saham, pemilik usaha, dan pegawai di industri terkait. Oleh karena itu, dalam mendukung kemajuan suatu perusahaan, perlu adanya kolaborasi yang lebih tinggi antar stakeholder [2]. Hal tersebut dapat diraih dengan cara menerapkan struktur regulasi yang baik terutama dalam proses internal perusahaan seperti suatu indikator dalam melacak produktivitas dalam bekerja.

Produktivitas menjadi salah satu hal yang dikhawatirkan dalam suatu organisasi, hal ini dikarenakan pada saat ini organisasi berurusan dengan berbagai proses manufaktur, spesifikasi teknis, dan tampilan dari produk yang mana merupakan faktor-faktor yang berhubungan dengan bagaimana suatu organisasi ingin memuaskan kebutuhan pelanggan [3]. Oleh karena itu penting adanya suatu sistem untuk mengklasifikasikan produktivitas pekerja, sehingga proses dalam suatu organisasi dapat lebih terukur.

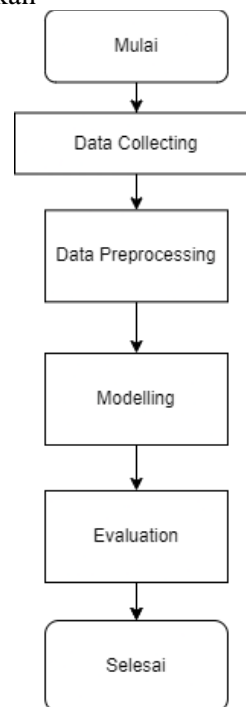
Beberapa model machine learning yang paling umum digunakan untuk klasifikasi adalah *Random Forest*, *Decision Tree*, dan *Support Vector Machine*. Hal tersebut ditunjukkan dengan ragamnya penelitian menggunakan model tersebut dalam beberapa tahun terakhir. Pada tahun 2023, Ferry Darmawan dkk berhasil memperoleh tingkat akurasi sebesar 82.5% pada model SVM untuk analisis sentimen kemungkinan depresi dan kecemasan pada Twitter [4]. Di tahun yang sama, Nining Nur Habibah dkk menggunakan Decision Tree C4.5 untuk mengklasifikasi faktor penyebab diabetes pada wanita, dan berhasil memperoleh akurasi sebesar 76.67% [5]. Satu tahun sebelumnya, Luthfiana Ratnawati dan Dwi Ratna Sulistyaningrum berhasil mendapatkan akurasi sebesar 75.3191% untuk memprediksi tingkat keparahan penyakit pada daun apel menggunakan *Random Forest* [6]. Pada penelitian ini, ketiga algoritma tersebut akan

dibandingkan berdasarkan perlakuan data *preprocessing* di tahap sebelumnya.

Dataset yang digunakan merupakan data sekunder yang didapat dari UC Irvine Machine Learning Repository dengan rentang waktu dari tanggal 1 januari 2015 sampai 11 maret 2015 dengan total data sebanyak 1197 baris. Dengan demikian penelitian ini diharapkan memberi kontribusi terhadap perkembangan teknologi dalam melakukan klasifikasi produktivitas dengan tujuan membantu pemegang kepentingan dalam keputusan yang tepat.

## 2. METODE PENELITIAN

Penelitian ini bertujuan untuk mendapatkan model yang memiliki hasil akurasi yang tinggi dalam mengklasifikasi dengan cara memaksimalkan tahapan *preprocessing data*. Oleh karena itu, dilakukan evaluasi performa pada berbagai macam tahap yang dilakukan serta implementasinya terhadap algoritma yang digunakan. Pada bagian ini dijelaskan bagaimana setiap tahapan penelitian dilakukan



Gambar 1. Diagram Alir Penelitian

## 2.1. Data Collecting

Dataset yang digunakan dalam penelitian ini adalah dataset Prediksi Produktivitas Pekerja Garmen pada situs UCI Machine Learning. Dataset ini terdiri dari 15 kolom dan 1.197 baris data

Tabel 1. Dataset Produktivitas Garmen

date	.....	actual_prod uctivity
1/1/2015	.....	0.940725
1/1/2015	.....	0.886500
.....	.....	.....
3/11/2015	.....	0.505889
3/11/2015	.....	0.394722

Tabel 2. Spesifikasi Dataset Produktivitas Garmen

No	Spesifikasi	
1	Tugas	Klasifikasi
2	Jumlah Data	1197
3	Jumlah Kolom	15
4	Tipe Data	Numerik dan Kategorik
5	Missing Value	Ya
6	Bidang	Tekstil

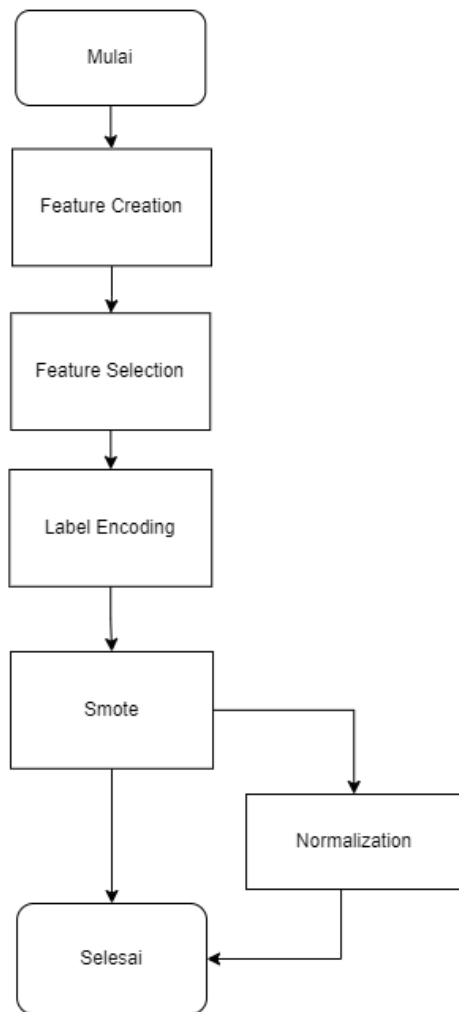
Kolom-kolom dalam dataset tersebut antara lain:

1. *date*, berisi format bulan, hari dan tahun.
2. *day*, jumlah hari kerja dalam seminggu.
3. *quarter*, proporsi per bulan dibagi menjadi 4 quarter.
4. *department*, departemen yang terkait dengan instansi.
5. *team\_no*, nomor tim yang terkait dengan instansi.

6. *no\_of\_workers*, jumlah pekerja tiap tim.
7. *no\_of\_style\_change*, jumlah perubahan gaya pada produk tertentu.
8. *targeted productivity*, target produktivitas.
9. *smv*, waktu yang dialokasikan pada sebuah penugasan.
10. *wip*, pekerjaan yang sedang berlangsung, termasuk jumlah item produk yang belum selesai.
11. *over\_time*, jumlah perpanjangan waktu masing-masing tim dalam hitungan menit.
12. *incentive*, jumlah insentif keuangan yang memungkinkan atau memotivasi suatu tindakan tertentu.
13. *idle\_time*, banyaknya waktu produksi yang terinterupsi karena beberapa alasan.
14. *idle\_men*, banyaknya pekerja yang *idle* ketika pekerjaan terinterupsi.
15. *actual\_productivity*, produktivitas pekerja dari range 0-1.

## 2.2. Data Preprocessing

Data *preprocessing* merupakan langkah penting dalam proses penambahan data. Hal ini untuk memastikan bahwa data bersih, lengkap, dan konsisten, serta dalam format yang mudah dianalisis [7]. Terdapat lima tahap data *preprocessing* yang akan dilakukan dalam penelitian ini: seleksi fitur, pelabelan data, normalisasi, imputasi data dan reduksi dimensi data. Tahapan tersebut dapat lebih jelas dilihat pada diagram berikut.



Gambar 2. Tahapan Data Preprocessing

### 2.2.1. Feature Creation

*Feature creation* adalah pembuatan fitur baru atau mengekstraksi fitur berdasarkan fitur-fitur yang sudah ada untuk meningkatkan performa model pembelajaran mesin [8]. Fitur baru membantu model untuk belajar dari data yang lebih kompleks. Fitur baru yang dibuat yaitu fitur kategori\_produktif, fitur ini dibuat berdasarkan fitur actual\_productivity yang dibagi menjadi dua kelas, kelas pertama berisi value di bawah 0.5 dengan label “tidak\_produktif” dan kelas kedua berisi value sama dengan atau di atas 0.5 menjadi label “produktif”.

Tabel 3. Spesifikasi Dataset Produktivitas

Actual_Productivity	Kategori_Produktif
$\geq 0.5$	Produktif
$< 0.5$	Tidak Produktif

### 2.2.2. Feature Selection

Tahap pertama yang dilakukan adalah seleksi fitur. Seleksi fitur merupakan salah satu metode dari feature engineering, dimana pada metode ini, subset dari fitur yang paling relevan atau penting dipilih untuk digunakan dalam analisis atau pemodelan. Tujuannya yakni untuk mengurangi kompleksitas model dengan mempertahankan fitur-fitur yang paling informatif dan relevan [7]. Hal ini membantu dalam meningkatkan kinerja model dan mengurangi overfitting. Kolom yang digunakan dalam pemodelan diantaranya, targeted\_productivity, smv, over\_time, incentive, no\_of\_workers, dan actual\_productivity

### 2.2.3. Label Encoding

Dalam machine learning dan analisis data, label encoding mengubah variabel kategori menjadi format numerik. Salah satu jenis label encoding yang digunakan adalah One-Hot Encoding, yang mengubah semua nilai kategori menjadi vektor biner dengan nilai 1 untuk kategori yang sesuai dan 0 untuk kategori lainnya [9]. Penelitian ini akan menggunakan *one-hot encoding* menjadi format yang dapat dipahami oleh algoritma pembelajaran mesin. Pengaplikasian *one-hot encoding* hanya diterapkan pada kolom *quarter, department, day*.

### 2.2.4. SMOTE

Salah satu teknik *oversampling* yang kerap digunakan dalam menangani

data dengan kelas yang tidak seimbang ialah SMOTE atau *Synthetic Minority Oversampling Technique*. SMOTE diyakini mampu menangkan masalah *imbalance* dengan cara menghasilkan data sintesis pada kelas minoritas melalui metode *k-Nearest Neighbour* (k-NN) [10]. Teknik *oversampling* menghasilkan data yang semula *imbalance* berjumlah 1197 baris data dengan proporsi 1 : 8, menjadi sejumlah 2128 baris dengan proporsi seimbang yaitu 1 : 1 untuk value pada variabel targetnya.

### 2.2.5. Normalization

Normalisasi merupakan proses perubahan skala pada suatu atribut dari dataset sehingga bisa dirubah menjadi suatu rentang tertentu [11]. Metode normalisasi yang diterapkan pada penelitian ini adalah normalisasi z-score, supaya dapat menangani data outlier dengan lebih baik. Berikut rumus normalisasi z-score.

$$X_{stand} = \frac{x - mean(x)}{std_{dev}(x)}$$

$$std_{dev}(x) = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x - mean(x))^2}$$

Kolom yang dilakukan normalisasi pada penelitian ini adalah kolom *over\_time* dan *incentive*. Tahapan ini merupakan tahapan pembandingan. Artinya, akan dibandingkan akurasi model berdasarkan dilakukan atau tidaknya normalisasi pada data.

## 2.3. Modelling

Pada penelitian ini, akan melakukan klasifikasi menggunakan tiga jenis model: *Random Forest*, *Decision Tree*, dan *Support Vector Machine*[12]. Ketiga model tersebut akan digunakan untuk mencari model yang terbaik dalam mengklasifikasikan produktivitas pekerja garmen

### 2.3.1. Random Forest

*Random forest* merupakan algoritma yang menggabungkan hasil dari

banyak pohon keputusan dengan membangkitkan atribut secara acak pada setiap *node*. *Random forest* bisa digunakan untuk berbagai macam *task*, salah satunya untuk keperluan klasifikasi [13]. Cara kerja *random forest* cukup sederhana, pertama algoritma *random forest* akan membangun kumpulan pohon keputusan. Setiap pohon keputusan dibangun dari sampel data yang diambil acak dari data asli, kemudian sampel data acak ini akan digunakan untuk melatih pohon keputusan, dari setiap pohon keputusan akan menghasilkan prediksi setiap data uji, terakhir semua prediksi dari semua pohon keputusan akan digabungkan untuk menghasilkan prediksi akhir [14].

### 2.3.2. Hyperparameter Tuning

Untuk mendapatkan hasil evaluasi yang terbaik maka diperlukan suatu metode dalam mencari parameter yang optimal pada suatu model. Oleh karena itu proses *hyperparameter tuning* menggunakan *grid search*. *Grid Search* sendiri merupakan salah satu teknik *hyperparameter tuning* yang bekerja dengan cara mengevaluasi performa model secara sistematis dengan kombinasi parameter yang berbeda-beda untuk mencari kelompok parameter yang terbaik.

Dalam penelitian ini terdapat beberapa parameter yang akan dicari kombinasi terbaiknya sebagai berikut.

- a. *n\_estimator*, menspesifikasikan jumlah model *decision tree* di dalam model *random forest*
- b. *max\_depth*, menspesifikasikan kedalaman maksimum setiap model *decision tree* pada model
- c. *min\_samples\_split*, menspesifikasikan jumlah minimum dari sampel yang dibutuhkan untuk memisahkan *node* yang ada di dalam model
- d. C, menspesifikasikan seberapa jauh model SVM dalam menerapkan regularisasi

- e. gamma, menspesifikasikan nilai yang digunakan pada fungsi kernel untuk mengukur pengaruh suatu titik data terhadap titik data yang lain

Tabel 4. Hyperparameter Random Forest

Parameter	Nilai Grid Search	Parameter Terbaik (Tanpa Normalisasi)	Parameter Terbaik (Dengan Normalisasi)
n_estimator	50, 100, 200	100	100
max_depth	None, 10, 20, 30	20	None
min_samples_split	2, 5, 10	5	5

Tabel 5. Hyperparameter Decision Tree

Parameter	Nilai Grid Search	Parameter Terbaik (Tanpa Normalisasi)	Parameter Terbaik (Dengan Normalisasi)
max_depth	None, 10, 20, 30	20	20
min_samples_split	2, 5, 10	10	2

Tabel 6. Hyperparameter SVM

Parameter	Nilai Grid Search	Parameter Terbaik (Tanpa Normalisasi)	Parameter Terbaik (Dengan Normalisasi)

C	0.1, 1, 10	10	10
gamma	0.01, 0.1, 1	0.01	0.1

## 2.4. Evaluation

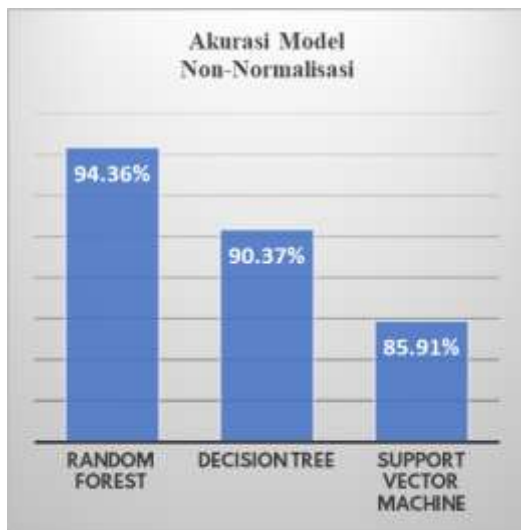
Model evaluasi pada pemodelan klasifikasi adalah proses untuk mengukur kinerja model klasifikasi. Model evaluasi digunakan untuk menilai seberapa baik model dapat memprediksi kelas target baru. Metode evaluasi yang digunakan pada penelitian ini adalah *accuracy* dan *confusion matrix*. *Accuracy* adalah persentase jumlah prediksi benar dengan jumlah keseluruhan data prediksi, sedangkan *confusion matrix* merupakan matriks ordo 2x2 yang berisi perbandingan antara hasil prediksi dengan nilai sebenarnya [15].

		Actual Values	
		1 (Positive)	0 (Negative)
Predicted Values	1 (Positive)	<b>TP</b> (True Positive)	<b>FP</b> (False Positive) <small>Type I Error</small>
	0 (Negative)	<b>FN</b> (False Negative) <small>Type II Error</small>	<b>TN</b> (True Negative)

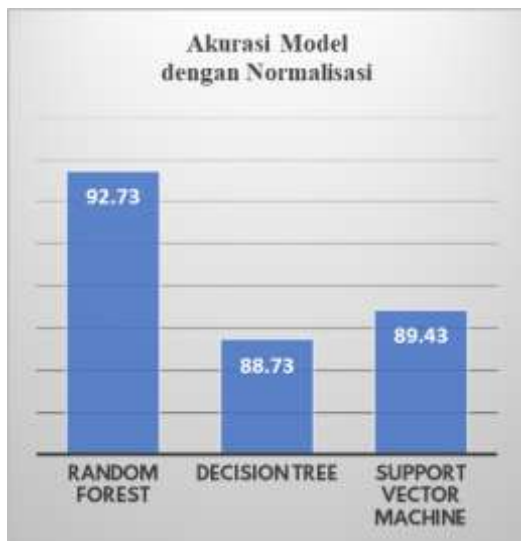
Gambar 3. Tabel *Confusion Matrix*

## 3. HASIL DAN PEMBAHASAN

Hasil yang diperoleh dari rangkaian proses mulai dari *preprocessing*, hingga *modelling*, telah didapat nilai evaluasi dari ketiga model dengan parameter terbaiknya masing-masing. Tabel berikut dapat menjelaskan terkait hasil performa model yang didapat.



Gambar 4. Grafik Bar-Chart Akurasi 3 Model Non-Normalisasi



Gambar 5. Grafik Bar-Chart Akurasi 3 Model dengan Normalisasi

Dari proses training algoritma yang dilakukan, didapatkan hasil akurasi yang tertinggi adalah dengan menggunakan metode algoritma *random forest* tanpa normalisasi dengan hasil akurasi sebesar 94.36%, dan tertinggi kedua adalah algoritma *random forest* dengan normalisasi yang menghasilkan nilai akurasi sebesar 92.73%. Kemudian akurasi terendah ada pada algoritma SVM tanpa normalisasi yang menghasilkan nilai akurasi sebesar 85.91%.

#### 4. KESIMPULAN

Semua tahap yang sudah dilakukan menghasilkan satu model terbaik yaitu model *random forest* yang dilatih menggunakan data tanpa proses normalisasi dihasilkan akurasi sebesar 94.36%, dengan demikian bisa dikatakan bahwa performa yang didapat sudah baik. Akan tetapi, performa model *random forest* tidak terbatas pada angka tersebut, mengingat adanya jenis parameter lain yang dapat diubah dengan nilai yang bermacam-macam, serta kondisi data yang digunakan dalam melatih model.

#### 5. SARAN

Untuk penelitian selanjutnya bisa ditekankan lebih dalam terkait pencarian parameter terbaik seperti menambah jenis parameter yang dicari beserta nilainya. Selain itu bisa juga dengan mengubah tahapan *preprocessing* pada data sehingga bisa menghasilkan data yang lebih layak untuk dimodelkan. Pemanfaatan algoritma lain untuk memprediksi produktivitas pun layak untuk diuji coba.

#### DAFTAR PUSTAKA

- [1] K. A. Schumacher and A. L. Forster, "Textiles in a circular economy: An assessment of the current landscape, challenges, and opportunities in the United States," *Front. Sustain.*, vol. 3, p. 1038323, Nov. 2022, doi: 10.3389/frsus.2022.1038323.
- [2] J. Nikam, "Gaps, challenges and drivers for environmentally sustainable textile and garment manufacturing in India," Stockholm Environment Institute, May 2023. doi: 10.51414/sei2023.033.
- [3] M. Salehi, H. Shirouyehzad, and R. Dabestani, "Labour productivity measurement through classification and standardisation of products," *Int. J. Product. Qual. Manag.*, vol. 11, no. 1,

- p. 57, 2013, doi: 10.1504/IJPQM.2013.050568.
- [4] F. Darmawan, M. Joe, Y. I. Kurniawan, and L. Afuan, "Analisis Sentimen Kemungkinan Depresi dan Kecemasan pada Twitter Menggunakan Support Vector Machine," *J. Eksplora Inform.*, vol. 13, no. 1, pp. 24–36, Sep. 2023, doi: 10.30864/eksplora.v13i1.854.
- [5] N. N. Habibah, A. Nazir, I. Iskandar, F. Syafria, L. Oktavia, and I. Syurfi, "Pemodelan Klasifikasi Untuk Menentukan Penyakit Diabetes dengan Faktor Penyebab Menggunakan Decision Tree C4.5 Pada Wanita," *J. Sist. Komput. Dan Inform. JSON*, vol. 4, no. 4, p. 654, Jun. 2023, doi: 10.30865/json.v4i4.6202.
- [6] L. Ratnawati and D. R. Sulistyaningrum, "Penerapan Random Forest untuk Mengukur Tingkat Keparahan Penyakit pada Daun Apel," *J. Sains Dan Seni ITS*, vol. 8, no. 2, pp. A71–A77, Jan. 2020, doi: 10.12962/j23373520.v8i2.48517.
- [7] Imran, F. Qayyum, D.-H. Kim, S.-J. Bong, S.-Y. Chi, and Y.-H. Choi, "A Survey of Datasets, Preprocessing, Modeling Mechanisms, and Simulation Tools Based on AI for Material Analysis and Discovery," *Materials*, vol. 15, no. 4, p. 1428, Feb. 2022, doi: 10.3390/ma15041428.
- [8] K. Jain, "A SURVEY ON FEATURE SELECTION TECHNIQUES," vol. 4, no. 5.
- [9] S. Ana, R. Kurniawan, and A. Nazir, "PENGKLASTERAN RISIKO COVID-19 DI RIAU MENGGUNAKAN TEKNIK ONE HOT ENCODING DAN ALGORITMA K-MEANS CLUSTERING," *J. Inf. Dan Komput.*, vol. 10, no. 1, pp. 154–163, Apr. 2022, doi: 10.35959/jik.v10i1.291.
- [10] B. Santoso, H. Wijayanto, K. A. Notodiputro, and B. Sartono, "A Comparative Study of Synthetic Oversampling Method to Improve the Classification of Poor Households in Yogyakarta Province," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 187, p. 012048, Nov. 2018, doi: 10.1088/1755-1315/187/1/012048.
- [11] R. A. Maula *et al.*, "Handling Missing Value dengan Pendekatan Regresi pada Dataset Akuakultur Berukuran Kecil," *J. Rekayasa Elektr.*, vol. 18, no. 3, Sep. 2022, doi: 10.17529/jre.v18i3.25903.
- [12] M. Guia, R. Silva, and J. Bernardino, "Comparison of Naïve Bayes, Support Vector Machine, Decision Trees and Random Forest on Sentiment Analysis," in *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, Vienna, Austria: SCITEPRESS - Science and Technology Publications, 2019, pp. 525–531. doi: 10.5220/0008364105250531.
- [13] V. Y. Kulkarni and D. P. K. Sinha, "Random Forest Classifiers: A Survey and Future Research Directions," *Int. J. Adv. Comput.*
- [14] I. Kurniawan, D. C. P. Buani, A. Abdussomad, W. Apriliah, and R. A. Saputra, "Implementasi Algoritma Random Forest Untuk Menentukan Penerima Bantuan Raskin," *J. Teknol. Inf. Dan Ilmu Komput.*, vol. 10, no. 2, pp. 421–428, Apr. 2023, doi: 10.25126/jtiik.20231026225.
- [15] Ž. Đ. Vujovic, "Classification Model Evaluation Metrics," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, 2021, doi: 10.14569/IJACSA.2021.0120670.



