**JURNAL MEDIA INFORMATIKA BUDIDARMA**
**Volume 7, Nomor 4, Oktober 2023, Page 1908-1919**
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i4.6856

# Analisis Optimasi Algoritma Decision Tree, Logistic Regression dan SVM Menggunakan Soft Voting

**Yosiko Aditya Pratama*, Fikri Budiman, Sri Winarno, Defri Kurniawan**

Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1,*]yosikoaditya@gmail.com, [2]fikri.budiman@dsn.dinus.ac.id, [3]sri.winarno@dsn.dinus.ac.id, [4]defri.kurniawan@dsn.dinus.ac.id
Email Penulis Korespondensi: yosikoaditya@gmail.com

**Abstrak**−Pertanian merupakan pilar utama dalam perekonomian sebuah negara. Salah satu kunci kesuksesan dalam pertanian adalah pemilihan lahan yang tepat. Identifikasi lahan yang subur atau tidak dapat dilakukan melalui pendekatan data mining yang dianggap lebih efisien. Hal ini karena data mining memiliki beberapa algoritma untuk mengekstraksi informasi penting dari sejumlah besar data melalui proses klasifikasi. Namun, algoritma klasifikasi dalam data mining sering menghadapi tantangan ketidakseimbangan data, yang dapat mengakibatkan tingkat akurasi yang rendah. Hasil pemrosesan data dengan model perhitungan yang memiliki tingkat akurasi rendah akan mengakibatkan banyaknya prediksi yang salah (fail prediction). Untuk mengatasi masalah ini, penelitian ini melakukan pengujian dan analisis perbandingan hasil Confusion Matrix dari empat model perhitungan, yaitu: algoritma Decision Tree, Logistic Regression, SVM, dan gabungan ketiga algoritma tersebut dengan teknik ensemble Soft Voting. Hasil pengujian menunjukkan bahwa pemrosesan data menggunakan algoritma Decision Tree, Logistic Regression, dan SVM, ditambah dengan optimasi model ensemble Soft Voting, memberikan akurasi tertinggi sebesar 91.53%. Hasil akurasi ini lebih tinggi jika dibandingkan dengan tiga model perhitungan lainnya, yaitu: algoritma Decision Tree dengan selisih 3.83%, Logistic Regression dengan selisih 2.66%, atau SVM dengan selisih 1.36%. Penelitian ini dapat memberikan kontribusi signifikan dengan mengidentifikasi solusi yang efisien untuk meningkatkan akurasi identifikasi lahan pertanian yang subur, yang merupakan langkah kunci dalam mendukung kesuksesan sektor pertanian dalam perekonomian negara.

**Kata Kunci:** Decision tree, Logistic Regression, Support Vector Machine, Soft Voting, Data Mining

**Abstract**−Agriculture constitutes a fundamental pillar of a nation's economy. One key to success in agriculture is the selection of suitable land. The prediction of whether land is fertile or not can be efficiently accomplished through a data mining approach. This is because data mining offers several algorithms for extracting crucial information from vast datasets through classification. However, classification algorithms in data mining often encounter the challenge of data imbalance, which can lead to low accuracy rates. Processing data with calculation models that have low accuracy rates can result in numerous erroneous predictions (fail predictions). To address this issue, this research conducts testing and comparative analysis of the confusion matrix results from four calculation models: the Decision Tree algorithm, Logistic Regression, SVM, and the combination of these three algorithms using the Soft Voting ensemble technique. The test results indicate that processing data using the Decision Tree, Logistic Regression, and SVM algorithms, along with the optimization of the Soft Voting ensemble model, achieves the highest accuracy rate of 91.53%. This accuracy rate is higher compared to the other three calculation models: the Decision Tree algorithm with a difference of 3.83%, Logistic Regression with a difference of 2.66%, and SVM with a difference of 1.36%. This research makes a significant contribution by identifying an efficient solution to improve the accuracy of identifying fertile agricultural land, which is a crucial step in supporting the success of the agricultural sector in the country's economy.

**Keywords**: Decision tree, Logistic Regression, Support Vector Machine, Soft Voting, Data Mining

## 1. PENDAHULUAN

Pertanian merupakan sektor penting dalam membangun perekonomian negara, yang menyimbang 10.9% PDB Nasional [1]. Pengelolaan lahan pertanian yang tepat dan optimal akan meningkatkan hasil pertanian di suatu daerah. salah satu kunci pengelolaan lahan pertanian adalah menentukan lahan mana yang bagus dan subur untuk digunakan sebagai lahan pertanian. Tanah subur adalah kondisi suatu tanah yang menyediakan unsur hara yang essensial bagi tumbuhan tanpa memberikan efek racun yang bersumber dari hara [2].

Dengan mengetahui tanah mana yang subur, akan mengurangi kemungkinan gagal panen yang diakibatkan tanah yang tidak memberikan pasokan unsur hara essensial yang optimal ke tanaman. Untuk mengetahui tanah tersebut subur ataupun sebaiknya, dapat dianalisis menggunakan pendekatan data mining. Dengan adanya analisis tanah, resiko bercocok tanam di lahan yang tidak subur akan berkurang, sehingga nantinya bisa menoptimalkan hasil panen.

Data mining merupakan proses komputasi dalam menggali informasi penting dan mengidentifikasi pola-pola tertentu dalam data yang berskala besar [3], [4]. Teknik analisis yang digunakan dalam data mining mencakup beberapa pendekatan kunci seperti asosiasi, estimasi, klasterisasi, dan klasifikasi. Setiap teknik ini memiliki pola pengolahan data yang unik serta menghasilkan informasi yang berbeda.

Teknik analisis asosiasi berfokus pada penemuan hubungan dan korelasi antara item-item dalam data. Ini memungkinkan untuk mengidentifikasi hubungan antara berbagai entitas atau produk dalam data, yang sangat bermanfaat dalam pemasaran dan rekomendasi produk [5]. Sementara itu, estimasi digunakan untuk menghitung atau memperkirakan nilai yang hilang atau tidak diketahui dalam dataset, memberikan pemahaman lebih dalam tentang data yang tidak lengkap [6]. Klasterisasi adalah teknik yang digunakan untuk mengelompokkan data ke


Submitted: **02/10/2023**; Accepted: **23/10/2023**; Published: **25/10/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 4, Oktober 2023, Page 1908-1919
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i4.6856

dalam kelompok-kelompok berdasarkan kesamaan fitur atau karakteristik tertentu. Hal ini membantu dalam pengelompokan data yang besar menjadi subkelompok yang lebih kecil, sehingga memudahkan analisis lebih lanjut [7]. Di sisi lain, klasifikasi adalah teknik yang digunakan untuk mengatribusikan label atau kategori tertentu ke data berdasarkan pola-pola yang telah ditemukan dalam dataset, yang membantu dalam prediksi dan pengambilan keputusan [8].

Dalam menganalisis tingkat kesuburan tanah, teknik yang paling cocok dan sangat relevan adalah klasifikasi. Klasifikasi memungkinkan pengelompokan data secara sistematis dan presisi, dengan mengkategorikan informasi tanah ke dalam kelompok atau kelas berdasarkan beragam parameter, atribut, dan kriteria yang ada [9].

Dengan menggunakan teknik klasifikasi ini, data tanah dapat diorganisir dengan lebih baik dan diberikan label yang menggambarkan karakteristiknya. Hal ini mempermudah pengolahan data yang lebih lanjut dan memungkinkan identifikasi kelas atau kategori yang tepat secara lebih efisien. Teknik klasifikasi juga memberikan kerangka kerja yang kuat untuk menganalisis dan memahami perbedaan dalam tingkat kesuburan tanah, yang merupakan elemen kunci dalam pengelolaan pertanian yang berkelanjutan dan efektif [10]. Dengan demikian, klasifikasi menjadi alat penting dalam upaya memaksimalkan hasil panen dan memastikan pemanfaatan sumber daya pertanian yang optimal.

Ada banyak algoritma klasifikasi data mining yang bisa diterapkan untuk membantu memprediksi kesuburan tanah, yaitu algoritma Priori, K-means, Logistic Regression, KNN, Naive Bayes, Decision Tree dan SVM [11]. Masalah yang sering ada di algoritma klasifikasi data mining adalah ketidakseimbangan data (imbalanced data), yang disebabkan adanya kelas yang memiliki jumlah sampel yang lebih banyak secara signifikan, sehingga algoritma klasifikasi lebih cenderung mengabaikan kelas dengan sampel sedikit, sehingga hasil proses pengolahan data akan kurang akurat [12]. Selain itu, penggunaan algoritma untuk membantu menentukan kesuburan tanah juga terkadang memiliki nilai akurasi yang kurang maksimal, sehingga perlu dioptimasi lebih lanjut untuk mengatasi kemungkinan fail prediction.

Untuk mengatasi ketidakseimbangan data dan fail prediction, perlu adanya proses pengolahan data menggunakan lebih dari satu algoritma klasifikasi dan penggunaan metode ensemble. Dengan demikian, proses pengolahan data akan menggabungkan kelebihan dari tiap algoritma yang nantinya akan menutupi kekurangan proses, sehingga akan didapatkan hasil akurasi yang lebih baik.

Disini penulis akan menggunakan algoritma Decision Tree, Logistic Regression dan Support vector Machine yang kemudian ketiganya dioptimasi menggunakan metode Soft Voting. Kelebihan algoritma Decision tree adalah sifatnya yang fleksibel, mudah diinterpretasikan dan lebih mudah mengolah hubungan non-linear [13]. Lalu Logistic Regression memiliki kelebihan dalam memprediksi masalah klasifikasi biner [14].

Kemudian Support Vector Machine (SVM) memiliki kekuatan dalam mengolah masalah klasifikasi yang lebih kompleks, serta memiliki nilai akrurasi yang tinggi jika dibandingkan dengan Naïve Bayes dan KNN [15]. Dengan menggabungkan ketiga algoritma tersebut dengan metode esemble Soft Voting, memungkinkan untuk menggabungkan prediksi dari ketiga model algoritma tersebut yang nantinya akan menghasilkan model perhitungan yang cenderung lebih akurat dan andal, karena mampu memanfaatkan kontribusi masing-masing model secara proporsional, meningkatkan stabilitas, dan mengurangi risiko overfitting atau bias yang mungkin terjadi pada satu model perhitungan tunggal [16].

Sejumlah penelitian terdahulu dalam bidang model perhitungan maupun kesuburan tanah dengan menggunakan data mining telah dilakukan. Dalam penelitian Fandi Yulian Pamuji dan Viry Puspaning Ramadhan, peneliti mengggunakan algoritma Decision Tree dan Random Forest untuk memprediksi keberhasilan Immunotheraphy. Namun, masalah ketidakseimbangan data dan nilai akurasi yang tergolong belum cukup tinggi, dengan nilai 84.4% untuk Decision Tree dan 85.5% untuk Random Forest [17].

Penelitian lainnya, Rina Kartika, telah mencoba metode Analytical Hierarchy Process untuk memprediksi kesuburan tanah [18]. Selanjutnya, dalam penelitian Wahyu Nugraha dan Raja Sabaruddin, Teknik Resampling untuk Mengatasi Ketidakseimbangan data pada C4.5, Random Forest, dan SVM [19]. Di sisi lain, Mochammad Ilham Aziz menganalisis metode Ensemble pada algoritma Decision Tree untuk Klasifikasi penyakit jantung [20]. Selain itu, dalam penelitian Andrea Manconi, peneliti mengkaji optimasi Soft Voting dalam memprediksi pasien yang terkena COVID-19 [21].

Penelitian ini memiliki tujuan untuk mengetahui berapa akurasi yang dihasilkan oleh algoritma Decision Tree, Logistic Regression dan Support Vector Machine sesudah dilakukan proses optimasi soft voting untuk mengatasi permasalahan ketidakseimbangan kelas, serta mengavaluasi hasil akurasi jika dibandingkan dengan model perhitungan algoritma Decision Tree, Logistic Regression dan Support Vector Machine secara individu.

## 2. METODOLOGI PENELITIAN

### 2.1 Dataset

Data yang digunakan untuk penelitian ini merupakan data privat yang diambil di Dinas Pertanian Kabupaten Grobogan. Jumlah data yang ada di dataset ini berjumlah 5000 record, 2 Kelas dengan atribut sebanyak 17 buah.


Submitted: **02/10/2023**; Accepted: **23/10/2023**; Published: **25/10/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 4, Oktober 2023, Page 1908-1919
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i4.6856

**Tabel 1.** Identifikasi model data

| Data | Kelas | Atribut | Record |
|---|---|---|---|
| Data Kesuburan | 2 | 17 | 5000 |

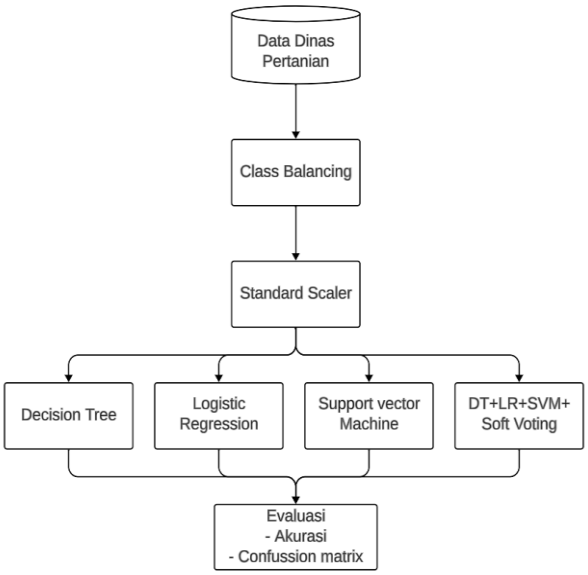
Adapun variabel input dan outputnya terdiri dari total 17 atribut. Terdapat 16 atribut bertipe data Numeric yang berisi parameter kesuburan tanah, serta 1 atribut bertipe data Binary yang berisi biner apakah tanah subur atau tidak.

**Tabel 2.** Variabel, atribut dan tipe data

| Variabel | Atribut | Tipe Data |
|---|---|---|
| **Input** | pH | numeric |
| | EC | numeric |
| | OC | numeric |
| | OM | numeric |
| | N | numeric |
| | P | numeric |
| | K | numeric |
| | Zn | numeric |
| | Fe | numeric |
| | Cu | numeric |
| | Mn | numeric |
| | Sand | numeric |
| | Silt | numeric |
| | Clay | numeric |
| | CaCO3 | numeric |
| | CEC | numeric |
| **Output** | Fertile / Non Fertile | Binary |

### 2.2 Skema Penelitian

Proses penelitian ini terbagi menjadi tiga, yaitu pre-processing, process dan evaluasi. pada pre-prosessing terdapat tahap Class Balancing dan Standard Scaler. kemudian pada bagian process, data diuji sebanyak empat kali, yaitu: 1. Pengujian dengan Decision Tree, 2. Pengujian dengan Logistic Regression, 3. Pengujian dengan Suport vector Machines, 4. pengujian dengan Decision Tree, Logistic Regression, Suport Vector Machines dan Soft Voting. kemudian hasil uji akan dievaluasi dengan Confussion Matrix.




**Gambar 1.** Skema Penelitian

### 2.3 Class Balancing

Data biner yang dipakai di penelitian ini terdiri dari dua kelas, yaitu kelas "1" (fertile) dan kelas "0" (non Fertile), dan jumlah sampel atau instans dari kedua kelas tersebut tidak seimbang, artinya salah satu kelas memiliki jumlah yang jauh lebih banyak daripada kelas yang lain. Penyeimbangan kelas dilakukan untuk membuat jumlah sampel dari kedua kelas menjadi seimbang, yaitu dengan cara menurunkan jumlah sampel dari kelas yang terlalu dominan.


Submitted: **02/10/2023**; Accepted: **23/10/2023**; Published: **25/10/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 4, Oktober 2023, Page 1908-1919
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i4.6856

### 2.4 StandardScaler

Standard Scaler adalah salah satu teknik pemrosesan data yang digunakan dalam data mining untuk melakukan normalisasi atau standarisasi data. Tujuannya adalah untuk mengubah data mentah menjadi bentuk yang memiliki mean (rerata) nol dan deviasi standar (standar deviasi) satu. Ini membantu dalam menghilangkan perbedaan skala antar variabel dalam data, sehingga mencegah variabel dengan skala yang besar mendominasi perhitungan di dalam algoritma pemrosesan data. Rumusnya sebagai berikut:

$$Z = \frac{X-\mu}{\sigma} \quad (1)$$

Dimana:
$Z$: nilai baru dari data setelah di-scaling.
$X$ : nilai asli dari data.
$\mu$ : rerata (mean) dari data.
$\sigma$ : deviasi standar (standard deviation) dari data.

### 2.5 Klasifikasi Decision Tree

Decision Tree (Pohon Keputusan) adalah salah satu algoritma machine learning yang digunakan untuk tugas klasifikasi. Konsep dasar algoritma ini menggambarkan alur pengambilan keputusan seperti struktur pohon dengan node (simpul) yang mewakili keputusan, cabang-cabang yang mewakili kondisi berdasarkan fitur-fitur input, dan daun-daun yang mewakili hasil klasifikasi atau nilai regresi.

Ada empat langkah dalam pemrosesan algoritma Decision tree. Pertama pemilihan atribut terbaik dengan memilih atribut yang menghasilkan Information Gain tertinggi atau Gini Impurity terendah. Setelah atribut terbaik dipilih, data dibagi menjadi subkelompok berdasarkan nilai-nilai atribut tersebut. Proses ini diulangi untuk setiap subkelompok, hingga membentuk pohon keputusan secara hierarkis. Pembentukan pohon berhenti ketika mencapai kondisi terminasi, seperti mencapai kedalaman maksimum pohon atau tidak ada perbedaan yang signifikan dalam nilai target.

### 2.6 Klasifikasi Logistic regression

Logistic Regression adalah algoritma klasifikasi yang digunakan untuk memprediksi probabilitas sukses atau gagalnya suatu peristiwa dengan menghasilkan output dalam bentuk probabilitas antara 0 dan 1. Algoritma ini berfungsi dengan memodelkan hubungan antara variabel independen (fitur) dengan variabel dependen biner (target) menggunakan fungsi logistik. Rumus logistik (sigmoid) yang digunakan dalam Logistic Regression adalah:

$$P(Y=1) = \frac{1}{1+e^{-(b_0+b_1X_1+b_2X_2+\cdots+b_kX_k}} \quad (2)$$

Dimana:
$(P(Y=1))$ : probabilitas bahwa target (Y) adalah 1
$(b_0, b_1, b_2, \ldots, b_k)$ : koefisien regresi yang harus dipelajari dari data pelatihan
$(X_1, X_2, \ldots, X_k)$: fitur-fitur yang digunakan dalam model

### 2.7 Klasifikasi Support Vector Machine

Support Vector Machine (SVM) adalah algoritma klasifikasi yang digunakan untuk mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan fitur-fitur yang ada dalam dataset. SVM bekerja dengan mencari hyperplane (bidang pemisah) terbaik yang memaksimalkan margin antara dua kelas yang ingin dipisahkan. Dalam hal ini, SVM mencoba untuk memahami pola atau struktur yang ada dalam data sehingga dapat memprediksi kategori atau kelas yang sesuai untuk data yang belum terlihat. Rumus SVM mencari Hyperplane terbaik dalam bentuk:

$$f(x) = sign(w.x.b) \quad (3)$$

Dimana:
$f(x)$ : fungsi prediksi yang menghasilkan output untuk data x
$w$ : vektor bobot yang digunakan untuk menentukan orientasi Hyperplane
$x$ : vektor fitur dari data yang ingin diprediksi
$b$ : bias, yang merupakan pergeseran Hyperplane dari titik asal

### 2.8 Soft Voting

Soft Voting adalah metode ensemble learning yang digunakan untuk meningkatkan kualitas prediksi dengan menggabungkan hasil prediksi dari beberapa model machine learning yang berbeda. Cara kerjanya cukup sederhana, pertama, setiap model dalam ensemble membuat prediksi terhadap data yang sama. Kemudian, setiap prediksi diberi bobot berdasarkan performa relatif model tersebut. Bobot ini mencerminkan tingkat kepercayaan terhadap masing-masing model. Setelah memberikan bobot, prediksi akhir dihitung dengan mengambil rata-rata


Submitted: **02/10/2023**; Accepted: **23/10/2023**; Published: **25/10/2023**

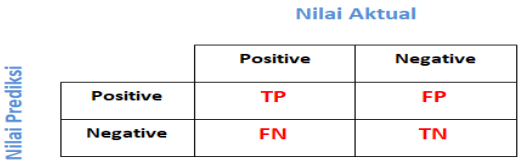
JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 4, Oktober 2023, Page 1908-1919
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i4.6856

tertimbang dari prediksi semua model. Dalam konteks klasifikasi, ini bisa berarti menghitung probabilitas prediksi kelas tertentu dengan mempertimbangkan bobot masing-masing model.

### 2.9 Confusion Matrix

Confusion Matrix adalah alat evaluasi yang penting dalam analisis dan evaluasi model klasifikasi. Tabel ini mengorganisir hasil prediksi model dengan jelas, dan memudahkan untuk memahami seberapa baik model tersebut dapat membedakan antara kelas target yang berbeda dalam dataset yang diberikan. Dalam Confusion Matrix, terdapat empat elemen utama yang memberikan wawasan tentang performa model, yaitu:

a. True Positive (TP): Ini adalah kasus di mana model dengan benar memprediksi sampel sebagai positif (kelas yang diinginkan).
b. False Positive (FP): Ini adalah kasus di mana model salah memprediksi sampel sebagai positif, padahal seharusnya negatif (kelas yang tidak diinginkan).
c. True Negative (TN): Ini adalah kasus di mana model dengan benar memprediksi sampel sebagai negatif.
d. False Negative (FN): Ini adalah kasus di mana model salah memprediksi sampel sebagai negatif, padahal seharusnya positif.




**Gambar 2.** Tabel Confussion Matrix

Dengan menggunakan informasi ini, kita dapat menghitung berbagai metrik evaluasi performa model, yaitu:

1. Akurasi (Accuracy), yaitu sejauh mana model dapat memprediksi dengan benar seluruh kelas. Rumusnya yaitu

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \quad (4)$$

2. Presisi (Precision) adalah sejauh mana model dapat memprediksi kelas positif dengan benar dari semua prediksi positif yang dilakukan. Adapun cara menghitungnya yaitu:

$$Presisi = \frac{TP}{TP+FP} \quad (5)$$

3. Recall (Sensitivitas) adalah sejauh mana model dapat mendeteksi semua kasus positif yang sebenarnya. rumusnya sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

4. F1-Score yaitu adalah ukuran yang menggabungkan Precision dan Recall untuk memberikan pemahaman yang lebih lengkap tentang performa model. berikut rumusnya:

$$F1 - Score = \frac{2 . Presisi . Recall}{Presisi + Recall} \quad (7)$$

## 3. HASIL DAN PEMBAHASAN

### 3.1 Class Balancing

pada data Dinas Pertanian kabupaten Grobogan, ada 5000 record dengan output hasil biner fertile (nilai 1) atau Non fertile (nilai 0). Dari 5000 data yang diuji, terdapat perbandingan output yang bisa dilihat di tabel dibawah

**Tabel 3.** output data sebelum Class Balancing

| | fertile (1) | non fertile (0) |
|---|---|---|
| **Jumlah Data** | 2520 | 2480 |

Data yang tidak seimbang akan meningkatkan kemungkinan masalah imbalance data yang akan mengurangi akurasi hasil. Dengan demikian dilakukanlah penyetaraan data, dimana jumlah hasil output 1 dan 0 harus sama. Data yang bernilai output 1 akan dikurangi secara acak sampai jumlahnya sama dengan data hasil output 0. Berikut ini adalah hasil dari data balancing

**Tabel 4.** Output data setelah Class Balancing

| | fertile (1) | non fertile (0) |
|---|---|---|
| **Jumlah Data** | 2480 | 2480 |


Submitted: **02/10/2023**; Accepted: **23/10/2023**; Published: **25/10/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 4, Oktober 2023, Page 1908-1919
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i4.6856

Selanjutnya, proses penyetaraan data ini dapat berdampak pada karakteristik asli dataset, sehingga memerlukan langkah-langkah lanjutan dalam analisis data untuk memastikan kualitas dan reliabilitas hasil yang dihasilkan. Untuk itu, data akan diolah lagi menggunakan metode standard scaler

### 3.2 Standard Scaler

Selanjutnya, setelah melakukan class balancing, data yang akan diolah juga harus mengikuti proses Standar Scaler. Hal ini diperlukan untuk mencegah variabel dengan skala besar mendominasi perhitungan, yang dapat memengaruhi hasil analisis. Seiring dengan itu, banyak algoritma pemrosesan data bergantung pada asumsi bahwa data terdistribusi secara normal dengan mean dan deviasi standar yang sama. Oleh karena itu, penggunaan fungsi Standard Scaler menjadi penting untuk membantu memenuhi asumsi ini dan meningkatkan kinerja algoritma. Untuk lebih memahami proses ini, berikut adalah contoh data sebelum diproses menggunakan standard scaler:

**Tabel 5.** Nilai sebelum Standard Scaler

| Atribut | Nilai |
|---|---|
| pH | 8.68 |
| EC | 0.08 |
| OC | 0.03 |
| OM | 0.01 |
| N | 86 |
| P | 23.84 |
| K | 388 |
| Zn | 0.37 |
| Fe | 8.4 |
| Cu | 0.32 |
| Mn | 2.2 |
| Sand | 89.9 |
| Silt | 4.1 |
| Clay | 4.2 |
| CaCO3 | 0 |
| CEC | 4.6 |

Nilai "388" dalam data tersebut, memiliki nilai yang jauh lebih besar dibandingkan dengan variabel lainnya seperti "0.03" atau "4.6". Jika tidak dilakukan normalisasi, variabel "388" ini mungkin mendominasi perhitungan dan membuat fitur-fitur lainnya tampak kurang berpengaruh, yang bisa mengakibatkan hasil analisis yang bias atau tidak akurat

Dalam proses standarisasi data ini, langkah awalnya adalah memahami dataset dengan baik untuk memastikan pemahaman yang tepat tentang variabel-variabel yang terlibat. Selanjutnya, dilakukan perhitungan rerata (mean) dan deviasi standar (standard deviation) dari setiap fitur dalam dataset. Proses inti dari standarisasi adalah scaling data, di mana setiap nilai dalam fitur diubah dengan mengurangkan mean dan membaginya dengan deviasi standar. Hasilnya adalah data yang telah dinormalisasi, dengan mean 0 dan deviasi standar 1. Data yang telah di-scaling ini siap digunakan dalam berbagai analisis atau pemodelan, membantu menghilangkan perbedaan skala antar fitur, meningkatkan akurasi algoritma, dan mempermudah interpretasi hasil. Berikut ini adalah hasil data setelah diolah menggunakan metode Standard Scaler:

**Tabel 6.** Nilai setelah Standar Scaler

| Atribut | Nilai |
|---|---|
| pH | - 0.07086721 |
| EC | -0.64549601 |
| OC | 0.15915766 |
| OM | -0.71080559 |
| N | -0.37410705 |
| P | 136.858.555 |
| K | 19.250.128 |
| Zn | -0.38236917 |
| Fe | 0.77246524 |
| Cu | -0.25177615 |
| Mn | 0.63224742 |
| Sand | -0.80875193 |
| Silt | -0.54113854 |
| Clay | -0.28343365 |
| CaCO3 | -0.86331243 |


Submitted: **02/10/2023**; Accepted: **23/10/2023**; Published: **25/10/2023**

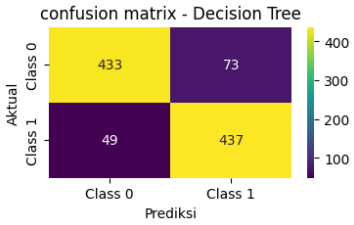
JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 4, Oktober 2023, Page 1908-1919
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i4.6856

| Atribut | Nilai |
|---|---|
| CEC | -0.11857233 |

Proses Standar Scaler dalam penelitian ini telah berhasil mengubah variabel-variabel dalam dataset menjadi format yang konsisten dan terstandardisasi dengan mean 0 dan deviasi standar 1. Hal ini membantu menghindari dominasi variabel-variabel dengan skala besar dan meningkatkan akurasi algoritma pemrosesan data. Hasilnya adalah data yang lebih seimbang dalam pengaruh masing-masing fitur, yang berdampak positif pada kinerja algoritma klasifikasi seperti Decision Tree, Logistic Regression, dan Support Vector Machine, serta meningkatkan hasil evaluasi seperti akurasi, presisi, recall, dan F1 Score.

### 3.3 Evaluasi Confusion Matrix Pada Perhitungan Decision tree

Data yang sudah melalui proses Class Balancing dan Standard Scaler kemudian telah diuji dengan berulang selama empat kali untuk memastikan kestabilan hasil, lalu dievaluasi menggunakan Confusion Matrix. Perhitungan evaluasi Confusion Matrix untuk algoritma klasifikasi decision tree menghasilkan hasil seperti yang diperlihatkan dalam gambar berikut:




**Gambar 3.** Confussion Matrix Decision tree

Gambar di atas menampilkan evaluasi dari pemrosesan data menggunakan algoritma Decision Tree dengan menggunakan metode confusion matrix. Pada kolom True Positive (TP), tercatat dengan nilai 433, sedangkan pada True Negative (TN), angkanya adalah 43. Kemudian, pada bagian False Positive (FP), terdapat 73, dan False Negative (FN) sebanyak 49. Dari hasil Confusion Matrix di atas, dapat dilakukan perhitungan akurasi, presisi, recall, dan F1 Score.

**Tabel 7.** Hasil evaluasi algoritma Decision Tree

| Parameter | Nilai |
|---|---|
| Akurasi | 87.7 % |
| Presisi | 85.57 % |
| Recall | 89.83 % |
| F1 Score | 87.78 % |

Hasil evaluasi confusion matrix untuk akurasi mendapatkan nilai sebesar 87.7%, kemudian hasil presisi mendapatkan nilai 85.57%. Pada perhitungan recall mendapatkan 89.83%, dan pada F1 score Confusion Matrix pada Decision Tree mendapatkan nilai sebesar 87.78%. Dari nilai-nilai ini dapat disimpulkan bahwa model Decision Tree memiliki kinerja yang baik dalam mengklasifikasikan data dengan akurasi yang layak. Presisi yang tinggi menunjukkan kemampuan model dalam menghindari kesalahan False Positive, sedangkan recall yang baik menunjukkan kemampuan model dalam mengidentifikasi dengan benar instance positif dalam dataset. Selain itu, nilai F1 score mencerminkan keseimbangan antara presisi dan recall, yang juga menunjukkan hasil yang memuaskan

### 3.4 Evaluasi Confusion Matrix pada perhitungan Logistic regression

Selanjutnya, data juga diproses menggunakan Logistic Regression. Hasil evaluasi berupa akurasi, presisi, recall, dan F1 score dapat dihitung dari Confusion Matrix di gambar berikut:


**Gambar 4.** Confussion Matrix Logistic Regression

Dalam hasil evaluasi Confusion Matrix yang diterapkan pada perhitungan Logistic Regression, ditemukan True Positive (TP) dengan nilai 458, sementara True Negative (TN) memperoleh nilai 424. Selanjutnya, False


Submitted: **02/10/2023**; Accepted: **23/10/2023**; Published: **25/10/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 4, Oktober 2023, Page 1908-1919
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i4.6856

Positive (FP) mendapatkan nilai 48 dan False Negative (FN) memiliki nilai 62. Semua nilai ini berkontribusi pada evaluasi akurasi, presisi, recall, dan F1 score sebagai berikut:

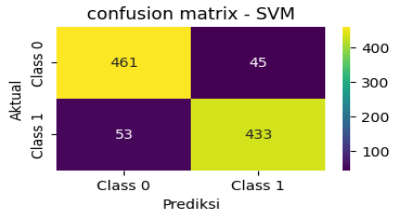
**Tabel 8.** Hasil evaluasi algoritma Logistic Regression

| Parameter | Nilai |
|---|---|
| Akurasi | 88.87 % |
| Presisi | 90.55 % |
| Recall | 88.02 % |
| F1 Score | 89.27 % |

Dari hasil perhitungan evaluasi Confusion Matrix pada proses menggunakan algoritma Logistic Regression, dapat disimpulkan bahwa model ini memiliki akurasi dengan nilai 88.87%. Lalu pada perhitungan presisi mendapatkan nilai yang tinggi, yaitu sekitar 90.55%, menunjukkan bahwa model mampu mengklasifikasikan positif dengan tepat. Selain itu, recall yang mencapai 88.02% dan F1 score sekitar 89.27% yang mana nilai ini menunjukkan keseimbangan yang baik antara presisi dan recall yang mana menandakan kinerja model yang solid.

### 3.5 Evaluasi Confusion matrix Pada Perhitungan Support Vector Machine

Selanjutnya, performa perhitungan data Dinas Pertanian Kabupaten Grobogan dengan algoritma klasifikasi Support vector Machine (SVM) dapat dilihat melalui perhitungan hasil Confusion Matrix berikut ini.




**Gambar 5.** Confussion Matrix SVM

Hasil Confusion Matrix menunjukkan kinerja model yang baik. Didapatkan nilai 461 pada bagian True Positive (TP) dan pada bagian False Positive (FP) mendapat nilai 45. Kemudian, didapatkan nilai 53 pada False Negative (FN) yang mana untuk mengindikasi data positif yang tidak terdeteksi. Terakhir, hasil pada True Negative (TN) mendapatkan nilai 433. Dari hasil Confussion Matrix tersebut, maka didapatkanlah hasil matrix akurasi, presisi, recall dan F1 Score sebagai evaluasi, dengan nilai sebagai berikut:

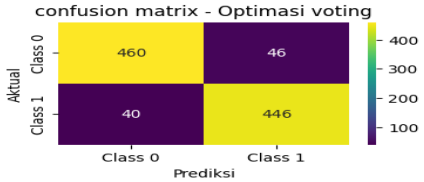
**Tabel 9.** Hasil evaluasi algoritma SVM

| Parameter | Nilai |
|---|---|
| Akurasi | 90.17 % |
| Presisi | 91.12 % |
| Recall | 89.69 % |
| F1 Score | 90.40 % |

Algoritma SVM yang diberikan memiliki performa yang sangat baik. Dengan akurasi sebesar 90.17%, model ini mampu dengan tepat mengklasifikasikan sebagian besar data. Presisi yang tinggi, yaitu 91.12%, menunjukkan bahwa ketika model SVM memberikan prediksi positif, alias cenderung benar. Recall sebesar 89.69% menunjukkan bahwa model mampu menemukan sebagian besar kasus positif yang sebenarnya. F1 Score yang mencapai 90.40% menggambarkan keseimbangan yang baik antara presisi dan recall, menunjukkan kemampuan algoritma SVM dalam menghadapi kasus ketidakseimbangan kelas. Dengan performa ini, model SVM ini dapat dianggap efektif untuk tugas klasifikasi yang sesuai.

### 3.6 Evaluasi Confussion Matrix pada perhitungan ketiga algoritma + Soft Voting

Perhitungan ini menunjukan hasil performa ketika algoritma Decision Tree, Logistic Regression dan SVM digabung dan dioptimasi menggunakan Soft Voting




**Gambar 6.** Confussion Matrix ensemble Soft Voting


Submitted: **02/10/2023**; Accepted: **23/10/2023**; Published: **25/10/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 4, Oktober 2023, Page 1908-1919
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i4.6856

Pengujian Confusion Matrix ini menunjukan performa yang sangat baik. Pada True Positive (TP) mendapat nilai 460 yang mengindikasikan kemampuan model dalam mengidentifikasi dengan tepat kelas positif. Hanya ada 46 False Positive (FP), yang menunjukkan model memiliki tingkat kesalahan positif palsu yang rendah. Kemudian jumlah False Negative (FN) sebanyak 40, yang menunjukan instance positif. Jumlah True Negative (TN) memiliki nilai 446 yang menunjukkanperforma baik untuk mengklasifikasikan negatif dengan benar.

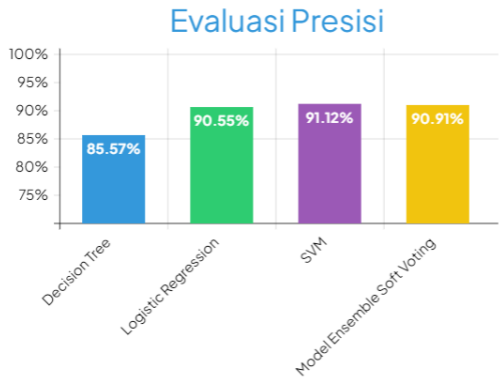
**Tabel 10.** Hasil evaluasi algoritma ensemble Soft Voting

| Parameter | Nilai |
|---|---|
| Akurasi | 91.53 % |
| Presisi | 90.91 % |
| Recall | 92.03 % |
| F1 Score | 91.47 % |

Performa dalam proses perhitungan dengan algoritma Decision Tree, Logistic Regression dan SVM yang dioptimasi menggunakan metode ensemble soft voting memberikan nilai akurasi 91.53%, yang mana dalam mengklasifikasikan keseluruhan data memiliki indikasi sangat baik. Nilai presisi mendapatkan 90.91% dan recall sebanyak 92.03%, yang mana keduanya memiliki hasil yang sangat tinggi. Kemudian F1 score mendapat nilai performa sebanyak 91.47%, yang berarti presisi dan recall memiliki keseimbangan yang tinggi.

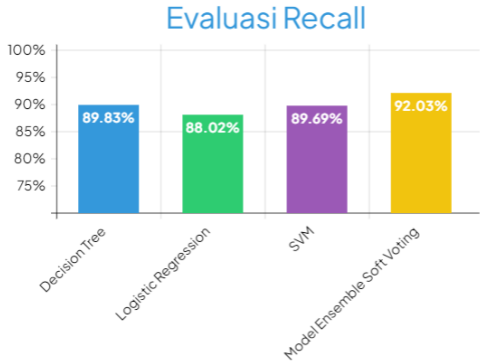
### 3.7 Analisa Hasil Pengujian

Dalam analisis model klasifikasi yang telah dilakukan, kita akan mengevaluasi performa empat model yang berbeda: Decision Tree, Logistic Regression, Support Vector Machine (SVM), dan Ensemble Soft Voting yang akan disajikan dalam bentuk diagram batang sebagai perbandingan performa keempat model perhitungan tersebut.




**Gambar 7.** Perbandingan hasil presisi

Pertama, pada hasil evaluasi presisi, yang mengukur sejauh mana model mampu mengklasifikasikan positif dengan benar tanpa memberikan banyak False Positives. Model Logistic Regression dan SVM memiliki tingkat presisi yang tinggi, dengan Logistic Regression mencapai sekitar 90.55% dan SVM sekitar 91.12%. Ensemble Soft Voting juga memiliki presisi yang tinggi juga, sekitar 90.91%. Meskipun Decision Tree memiliki presisi yang lebih rendah, sekitar 85.57%, masih memiliki kemampuan yang layak dalam mengklasifikasikan positif.



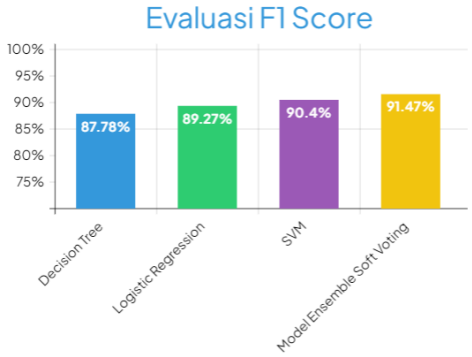

**Gambar 8.** Perbandingan hasil recall

Kemudian, hasil evaluasi recall, yang mengukur sejauh mana model dapat mengidentifikasi semua instance positif dalam dataset. Ensemble Soft Voting memiliki nilai recall tertinggi, yaitu sekitar 92.03%, yang mana ini menunjukkan kemampuannya dalam mengidentifikasi nilai positif dengan sangat baik. Decision Tree juga memiliki recall yang tidak kalah tinggi, dengan nilai sekitar 89.83%, kemudian diikuti oleh Algoritma SVM


Submitted: **02/10/2023**; Accepted: **23/10/2023**; Published: **25/10/2023**

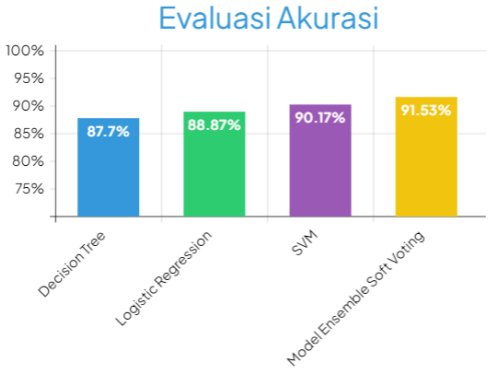
JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 4, Oktober 2023, Page 1908-1919
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i4.6856

dengan hasil nilai sekitar 89.69%, dan Logistic Regression mendapat nilai recall terendah jika dibandingkan dengan ketiga model perhitungan lainya, dengan mendapatkan sekitar 88.02%. Meskipun begitu, nilai dari Logistic Regression masih bisa dibilang mampu mengidentifikasi nilai positif dengan baik.




**Gambar 9.** Perbandingan hasil F1 Score

Selanjutnya, F1 Score, yang mana nilai ini mencerminkan keseimbangan antara presisi dan recall. Pada model perhitungan Ensemble Soft Voting memiliki F1 Score tertinggi, yaitu sekitar 91.47%, hal ini menunjukkan keseimbangan yang baik antara kemampuan model dalam mengklasifikasikan positif dengan benar dan mengidentifikasinya dengan baik. Kemudian disusul oleh algoritma SVM yang memiliki F1 Score dengan nilai sekitar 90.40%, diikuti oleh Logistic Regression dengan mendapatkan nilai F1 Score sekitar 89.27%. Decision Tree memiliki F1 Score terendah dibanding dengan tiga model perhitungan lainya, dengan mendapat nilai sekitar 87.78%, yang mana nilai ini juga tetap mencerminkan keseimbangan yang baik antara presisi dan recall.




**Gambar 10.** Perbandingan hasil akurasi

Terakhir, pada hasil evaluasi akurasi, yang mengukur seberapa baik model cocok dengan data secara keseluruhan. Hasilnya adalah model perhitungan ensemble soft voting memiliki nilai akurasi tertinggi, dengan mendapatkan nilai sebesar 91.53%, hal ini menunjukkan kemampuannya dalam mengklasifikasikan data dengan benar secara keseluruhan, lebih baik jika dibandingkan dengan ketiga nilai model perhitungan lainya. Kemudian disusul SVM yang memiliki akurasi sekitar 90.17%, diikuti oleh Logistic Regression dengan hasil akurasinya sekitar 88.87%, dan terakhir Decision Tree mendapat nilai akurasi terendah dengan nilai sebesar 87.7%. Nilai akurasi dari decision tree memiliki selisih yang cukup jauh jika dibandingkan dengan model perhitungan esemble soft voting dengan perbedaan 3.83%

Singkatnya, dalam pengujian algoritma Decision Tree, Logistic Regression, dan SVM, ketiga model menunjukkan kinerja yang baik. Decision Tree memiliki akurasi 87.7% dan F1 Score 87.78%, menandakan kemampuan klasifikasi yang baik. Logistic Regression mencapai akurasi 88.87% dan F1 Score 89.27%, dengan presisi yang sangat tinggi. SVM juga mengesankan dengan akurasi 90.17% dan F1 Score 90.40%, menunjukkan kemampuan baik dalam mengklasifikasikan data. Namun, perpaduan ketiganya dengan Soft Voting menghasilkan performa terbaik dengan akurasi 91.53% dan F1 Score 91.47%, menunjukkan bahwa ensemble model dapat meningkatkan kinerja secara signifikan. Dari segi presisi, Logistic Regression memiliki unggulan, sementara SVM memiliki kombinasi baik antara presisi dan recall. Keseluruhan, Soft Voting memberikan keseimbangan yang optimal antara akurasi, presisi, dan recall.

Keseluruhan, penelitian ini memberikan wawasan yang berharga dalam penggunaan algoritma Machine Learning dalam konteks klasifikasi data. Hasil evaluasi model machine learning memberikan bukti bahwa penggabungan model dengan ensemble Soft Voting dapat menjadi solusi efektif untuk meningkatkan ketepatan klasifikasi


Submitted: **02/10/2023**; Accepted: **23/10/2023**; Published: **25/10/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
Volume 7, Nomor 4, Oktober 2023, Page 1908-1919
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i4.6856

## 4. KESIMPULAN

Hasil pengujian menunjukkan bahwa algoritma Machine Learning, seperti Decision Tree, Logistic Regression, dan SVM, memiliki potensi besar dalam mengatasi permasalahan klasifikasi pada dataset ini. Masing-masing model memberikan tingkat akurasi yang sangat baik, dengan SVM mencapai akurasi tertinggi pada 90.17%. Ini mengindikasikan kemampuan model dalam mengklasifikasikan data dengan tepat. Namun, perlu diperhatikan bahwa penggunaan ensemble model dengan Soft Voting menghasilkan performa terbaik dengan akurasi 91.53%. Hal ini menunjukkan bahwa penggabungan berbagai model-machine learning dapat meningkatkan kinerja secara signifikan dan mendukung kesimpulan bahwa ensemble model dapat digunakan untuk meningkatkan ketepatan klasifikasi. Penting untuk dicatat bahwa presisi dan recall juga memiliki peran penting dalam evaluasi model. Logistic Regression menonjol dalam hal presisi dengan 90.55%, sementara SVM menunjukkan keseimbangan yang baik antara presisi dan recall dengan F1 Score 90.40%. Harapan ke depan dalam penelitian ini adalah untuk terus mengembangkan dan mengoptimalkan model ini serta menjalankan eksperimen lebih lanjut dengan berbagai teknik preprocessing dan tuning parameter. Selain itu, pemahaman yang lebih mendalam tentang fitur-fitur yang paling berpengaruh dalam pengklasifikasian juga dapat digali untuk meningkatkan akurasi model. Selain itu, penggunaan model ini dalam konteks nyata dapat memberikan pandangan berharga dalam mendukung keputusan dan pemecahan masalah di bidang pertanian Kabupaten Grobogan, memungkinkan pemangku kepentingan untuk mengambil tindakan yang lebih tepat waktu dan efektif dalam mengelola sumber daya pertanian. Dengan pembaruan dan peningkatan yang berkelanjutan, model klasifikasi ini memiliki potensi besar untuk memberikan kontribusi yang berarti dalam mendukung sektor pertanian yang berkelanjutan dan efisien.

## UCAPAN TERIMAKASIH

Yang pertama dan utama, penulis ingin berterimakasih kepada Allah SWT yang telah memberikan kelancarandan petunjuk dalam pembuatan jurnal. Selanjutnya penulis juga ingin memberikan berterimakasih kepada bapak dosen pembimbing, Fikri Budiman, serta teman-teman saya, Ridlo, Yoga, Khaliq, yang juga memberikan bantuan wawasan dalam jurnal ini. Tidak lupa terimakasih juga ditujukan kepada Band Dewa19, Avenged Sevenfold, Queen, yang selalu memberikan suasana nyaman dengan lagu-lagunya.

## REFERENCES

[1] S. I. Kusumaningrum, “Pemanfaatan Sektor Pertanian Sebagai Penunjang Pertumbuhan Perekonomian Indonesia,” Jurnal Transaksi, Vol. 11, No. 1, Hlm.80, 2019.
[2] L. Apriatin Dan L. Kamelia, “Pemanfaatan Tanah Subur Melalui Pendampingan Budidaya Sayuran Secara Organik,” Jurnal Abdimu : Pengabdian Kepada Masyarakat, Vol. 1, No. 2, Hlm. 39–47, 2021, Doi: 10.32627.
[3] L. Setiyani, M. Wahidin, D. Awaludin, Dan S. Purwani, “Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review,” Faktor Exacta, Vol. 13, No. 1, Hlm. 35, Jun 2020, Doi: 10.30998/Faktorexacta.V13i1.5548.
[4] Z. Nabila, A. Rahman Isnain, Dan Z. Abidin, “Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means,” Jurnal Teknologi Dan Sistem Informasi (Jtsi), Vol. 2, No. 2, Hlm. 100, 2021.
[5] A. R. Wibowo Dan A. Jananto, “Implementasi Data Mining Metode Asosiasi Algoritma Fp-Growth Pada Perusahaan Ritel,” Jurnal Teknologi Informasi dan Komunikasi, Vol. 10, No. 2, Hlm. 200–212, 2020.
[6] Y. L. Nainel, E. Buulolo, Dan I. Lubis, “Penerapan Data Mining Untuk Estimasi Penjualan Obat Berdasarkan Pengaruh Brand Image Dengan Algoritma Expectation Maximization (Studi Kasus: Pt. Pyridam Farma Tbk),” Jurikom (Jurnal Riset Komputer), Vol. 7, No. 2, Hlm. 214, Apr 2020, Doi: 10.30865/Jurikom.V7i2.2097.
[7] A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, Dan W. Gata, “Analisis Sentimen Aplikasi Ruang Guru Di Twitter Menggunakan Algoritma Klasifikasi,” Jurnal Teknoinfo, Vol. 14, No. 2, Hlm. 115, Jul 2020, Doi: 10.33365/Jti.V14i2.679.
[8] T. Novika, P. Poningsih, H. Okprana, A. P. Windarto, Dan H. Siahaan, “Penerapan Data Mining Klasifikasi Tingkat Pemahaman Siswa Pada Pelajaran Matematika,” Jurnal Media Informatika Budidarma, Vol. 5, No. 1, Hlm. 9, Jan 2021, Doi: 10.30865/Mib.V5i1.2498.
[9] K. F. Irnanda, D. Hartama, Dan A. P. Windarto, “Analisa Klasifikasi C4.5 Terhadap Faktor Penyebab Menurunnya Prestasi Belajar Mahasiswa Pada Masa Pandemi,” Jurnal Media Informatika Budidarma, Vol. 5, No. 1, Hlm. 327, Jan 2021, Doi: 10.30865/Mib.V5i1.2763.
[10] S. Sunardi, A. Fadlil, Dan N. M. P. Kusuma, “Comparing Data Mining Classification For Online Fraud Victim Profile In Indonesia,” Intensif: Jurnal Ilmiah Penelitian Dan Penerapan Teknologi Sistem Informasi, Vol. 7, No. 1, Hlm. 1–17, Feb 2023, Doi: 10.29407/Intensif.V7i1.18283.
[11] F. Handayani Dkk., “Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network Dalam Prediksi Penyakit Jantung”, Jepin (Jurnal Edukasi Dan Penelitian Informatika), Vol. 7, No. 3, hlm. 329, Des 2021.
[12] N. Sulistiyowati Dan M. Jajuli, “Integrasi Naïve Bayes Dengan Teknik Sampling Smote Untuk Menangani Data Tidak Seimbang”, Jurnal Nuansa Informatika, Vol. 14, No. 1, Hlm. 34, 2020.
[13] A. Prayoga Permana, K. Ainiyah, Dan K. Fahmi Hayati Holle, “Analisis Perbandingan Algoritma Decision Tree, Knn, Dan Naive Bayes Untuk Prediksi Kesuksesan Start-Up”, JISKa (Jurnal Informatika Sunan Kalijaga), Vol. 6, No. 3, Hlm. 178, 2021.


Submitted: **02/10/2023**; Accepted: **23/10/2023**; Published: **25/10/2023**

JURNAL MEDIA INFORMATIKA BUDIDARMA
**Volume 7, Nomor 4, Oktober 2023, Page 1908-1919**
ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online)
Available Online at https://ejurnal.stmik-budidarma.ac.id/index.php/mib
DOI: 10.30865/mib.v7i4.6856

[14] P. : Amset, I. Batusangkar, I. B. Press, N. Putu, N. Hendayanti, Dan M. Nurhidayati, "Regresi Logistik Biner Dalam Penentuan Ketepatan Klasifikasi Tingkat Kedalaman Kemiskinan Provinsi-Provinsi Di Indonesia", Sainstek : Jurnal Sains Dan Teknologi, Vol. 12, No. 02, Hlm. 63-70, 2020.
[15] O. Bangun, H. Mawengkang, Dan S. Efendi, "Metode Algoritma Support Vector Machine (Svm) Linier Dalam Memprediksi Kelulusan Mahasiswa," Jurnal Media Informatika Budidarma, Vol. 6, No. 4, Hlm. 2006, Okt 2022, Doi: 10.30865/Mib.V6i4.4572.
[16] H. B. Kibria, M. Nahiduzzaman, M. O. F. Goni, M. Ahsan, Dan J. Haider, "An Ensemble Approach For The Prediction Of Diabetes Mellitus Using A Soft Voting Classifier With An Explainable Ai," Sensors, Vol. 22, No. 19, Okt 2022, Doi: 10.3390/S22197268.
[17] F. Yulian Pamuji, V. Puspaning Ramadhan, Dan R. Artikel, "Jurnal Teknologi Dan Manajemen Informatika Komparasi Algoritma Random Forest Dan Decision Tree Untuk Memprediksi Keberhasilan Immunotheraphy Info Artikel Abstrak," Vol. 7, No. 1, Hlm. 46–50, 2021.
[18] R. Kartika, S. Adi, Dan A. Murnomo, "Implementasi Metode Analytical Hierarchy Process Untuk Prediksi Tingkat Kesuburan Tanah", Edu Komputika Journal, Vol. 6, No. 1, Hlm. 8, 2019
[19] W. Nugraha Dan R. Sabaruddin, "Teknik Resampling Untuk Mengatasi Ketidakseimbangan Kelas Pada Klasifikasi Penyakit Diabetes Menggunakan C4.5, Random Forest, Dan Svm", Vol. 20, No. 3, Hlm. 352-361, 2021.
[20] M. Ilham Aziz Dan A. Zainul Fanani, "Analisis Metode Ensemble Pada Klasifikasi Penyakit Jantung Berbasis Decision Tree", Techno.COM, Vol. 7, No. 1, 2023, Doi: 10.30865/Mib.V7i1.5169.
[21] A. Manconi, G. Armano, M. Gnocchi, Dan L. Milanesi, "A Soft-Voting Ensemble Classifier For Detecting Patients Affected By Covid-19," Applied Sciences (Switzerland), Vol. 12, No. 15, Agu 2022, Doi: 10.3390/App12157554.


Submitted: **02/10/2023**; Accepted: **23/10/2023**; Published: **25/10/2023**