

OPTIMASI ANALISIS KESUBURAN TANAH DENGAN PENDEKATAN SOFT VOTING ENSEMBLE

Yoga Mahendra Awaludin

Fakultas Ilmu Komputer, Program Studi Teknik Informatika

Universitas Dian Nuswantoro

Email: 111202012525@mhs.dinus.ac.id

Fikri Budiman

Fakultas Ilmu Komputer, Program Studi Teknik Informatika

Universitas Dian Nuswantoro

Email: fikri.budiman@dsn.dinus.ac.id

ABSTRAK

Penelitian ini mengusulkan Optimasi Algoritma *Decision Tree*, *K-Nearest Neighbor* dan *Support Vector Machine* menggunakan Metode *Soft Voting* dalam konteks kesuburan tanah pertanian. Metode *soft voting* adalah suatu pendekatan dimana hasil klasifikasi dari ketiga algoritma dikombinasikan dengan memberikan bobot-bobot pada masing-masing klasifikasi. Ini memungkinkan algoritma yang digunakan dapat berkolaborasi untuk memberikan solusi klasifikasi yang akurat dan handal. Hal ini didasari karena untuk klasifikasi kesuburan tanah masih dilakukan secara manual dan hal itu memungkinkan adanya suatu kesalahan dalam proses klasifikasi. Studi ini sangat relevan karena kesuburan tanah merupakan aspek kunci dalam pertanian yang dapat mempengaruhi hasil panen dan kualitas produk. Metode penelitian ini melibatkan pengumpulan data unsur hara dari Dinas Pertanian Kabupaten Grobogan dan analisis sistematis dengan Algoritma *Decision Tree*, *K-Nearest Neighbor* dan *SVM* kemudian diterapkan Metode Optimasi *Soft Voting* untuk meningkatkan akurasi. Hasil penelitian menunjukkan bahwa Metode Optimasi *Soft Voting* mampu mengatasi masalah kesuburan tanah dengan meningkatkan akurasi klasifikasi. Nilai akurasi dari Algoritma *Decision Tree* sebesar 88,7%, nilai akurasi *K-Nearest Neighbor* sebesar 86,7, nilai akurasi *SVM* sebesar 90,1%, dan akurasi penggabungan algoritma dengan Optimasi *Soft Voting* sebesar 90,4. Penelitian ini penting dalam bidang pertanian karena membantu petani dan ahli pertanian dalam mengambil keputusan mengenai klasifikasi kesuburan tanah. Dengan meningkatkan pemahaman penggunaan teknologi melalui optimalisasi algoritma, diharapkan produktivitas pertanian dan ketahanan pangan meningkat.

Kata kunci: Decision Tree, K-Nearest Neighbor, Soft Voting, pertanian, optimasi

ABSTRACT

This study focuses on optimizing Decision Tree, K-Nearest Neighbor and Support Vector Machine (SVM) algorithms using the Soft Voting Method in the context of agricultural soil fertility. The research is motivated by the manual process of soil fertility prediction, which can introduce errors. This research is highly relevant because soil fertility significantly affects crop yields and product quality, making it a critical aspect of agriculture. The research methodology involves collecting nutrient data from the Grobogan District Agriculture Office and systematically analyzing it using Decision Tree, K-Nearest Neighbor, and SVM algorithms. Subsequently, the Soft Voting Optimization Method is applied to improve prediction accuracy. The findings indicate that the Soft Voting Optimization Method effectively addresses soil fertility challenges, significantly enhancing prediction accuracy. The achieved accuracy values are as follows: Decision Tree algorithm at 88.7%, K-Nearest Neighbor at 86.7%, SVM at 90.1%, and the combined algorithm with Soft Voting Optimization achieving 90.4% accuracy. This research is crucial in agriculture, aiding farmers and agricultural experts in making informed decisions about soil fertility predictions. By optimizing algorithms and enhancing technology utilization, it is expected that agricultural productivity will increase, ultimately contributing to improved food security.

Keywords: *Decision Tree, K-Nearest Neighbor, Soft Voting, agriculture, optimization*

1. PENDAHULUAN

Indonesia dikenal dunia sebagai negara agraris yang sebagian besar penduduknya terlibat dalam kegiatan pertanian. Kondisi seperti inilah menjadikan sektor pertanian menjadi andalan utama mata pencaharian penduduk Indonesia sebesar 87,50% pada tahun 2019 [1][2]. Pertanian memiliki peranan penting dalam keberlangsungan hidup manusia [3]. Peranan tersebut meliputi pemenuhan kebutuhan pangan, penciptaan lapangan kerja, kontribusi terhadap ekonomi dan kesejahteraan masyarakat. Meskipun sektor pertanian menjadi andalan di negeri ini namun tidak dapat terlepas dengan berbagai permasalahan yang meliputi masalah ekologi, sosial, ekonomi dan lain sebagainya [3]. Kesuburan tanah merupakan penentu peningkatan jumlah produksi dalam pertanian. Keberagaman komposisi kimia tanah mempunyai kemampuan untuk menunjang kehidupan hasil pertanian dengan tersedianya unsur hara, sehingga terdapat jenis tanah yang disebut tanah subur dan sebaliknya [4]. Tentunya dibutuhkan penerapan teknologi untuk setidaknya mengatasi permasalahan yang terjadi [5].

Pada kondisi di lapangan pemanfaatan teknologi jarang diterapkan hanya daerah tertentu yang menerapkan teknologi dalam prakteknya karena sumber daya yang memadai baik dari sumber daya manusia dan sumber daya lainya yang menunjang diterapkannya sebuah teknologi didalamnya. Badan penyuluhan pertanian tiap daerah melakukan pengecekan di lapangan berkala untuk melakukan perhitungan variabel tertentu dan menganalisis kesuburan tanah. Hal tersebut tentunya menyulitkan proses analisis dan tingkat kesalahan relatif tinggi.

Teknik data mining merupakan elemen utama dalam perkembangan teknologi saat ini karena penggunaan teknik data mining dalam sebuah dataset tersebut mampu mengambil sebuah keputusan yang lebih baik dan mampu mengekstraksi sebuah informasi untuk mengungkap sebuah wawasan baru dari sebuah dataset mentah yang akan diolah [6][7]. Efisiensi proses ini dapat digunakan dalam meningkatkan keuntungan, memperkecil biaya pengeluaran dalam sebuah kegiatan [8]. Pemanfaatan teknik pembelajaran *ensemble* dalam dunia *machine learning* merupakan langkah yang efektif untuk diterapkan. Teknik *ensemble* ini bertujuan untuk meningkatkan akurasi dari model yang dibentuk dengan cara menggabungkan nilai hasil dari model yang berbeda untuk mengatasi sebuah permasalahan dan meningkatkan keandalan klasifikasi. Ini dilakukan karena pastinya setiap model yang dibentuk pasti memiliki kelebihan dan kelemahan tertentu dan dengan menggabungkannya kita dapat memperoleh hasil akhir yang akurat daripada menggunakan metode tunggal [9]. Penggunaan teknik *ensemble* seperti *bagging*, *boosting* dan *stacking* dalam sebuah penelitian sudah terbukti mampu mengoptimasi akurasi dari sebuah algoritma yang digunakan dalam penelitian. Selain itu, dalam konteks *ensemble*, Metode *Soft Voting* merupakan cara pendekatan yang efektif. Metode ini memungkinkan penggabungan hasil klasifikasi dari berbagai model dengan memberikan bobot pada masing-masing klasifikasi.. Bobot ini ditentukan berdasarkan performa dari masing- masing model yang telah dilatih atau dengan model validasi yang sesuai. Model validasi merujuk pada proses evaluasi model *machine learning* untuk memastikan bahwa model yang telah dibentuk mampu memberikan hasil klasifikasi yang akurat dan dapat diandalkan saat dihadapkan dengan data lainnya. Termasuk teknik *ensemble* yang lainya seperti *Bagging*, *Boosting*, dan *Stacking*, metode *soft voting* memberikan solusi klasifikasi yang lebih akurat dan handal dan dapat mengoptimalkan akurasi algoritma yang digunakan dalam penelitian.

Bagging, *Boosting*, *Stacking*, dan *Soft Voting* merupakan jenis dari *ensemble learning* dan memiliki tujuan yang sama yaitu meningkatkan kinerja model dan mengoptimalkan hasil akurasi, namun memiliki perbedaan dari segi skema dan peran dalam sebuah penelitian. *Bagging* memiliki skema kinerja dengan melibatkan pembentukan beberapa model dengan sampel acak dari data *training*. Setiap model melakukan klasifikasi secara independen. *Boosting* memiliki skema kinerja dengan melibatkan pembentukan beberapa model secara berurutan, dimana setiap model berusaha memperbaiki model sebelumnya. Model-model yang lemah diberikan bobot yang lebih tinggi. *Stacking* memiliki skema kinerja dengan melibatkan penggunaan beberapa model untuk menghasilkan hasil klasifikasi. Metode ini memungkinkan penggabungan beberapa model dengan suatu pendekatan yang berbeda untuk menghasilkan klasifikasi yang lebih kuat. Peran penggunaan *Boosting* pada suatu penelitian adalah membantu mengurangi varians dan meningkatkan stabilitas model. Dengan membuat model yang berbeda, *bagging* mengurangi risiko

overfitting dan membantu model untuk lebih umum dan lebih stabil. Peran *Boosting* adalah memfokuskan pada peningkatan performa model dengan memberikan penekanan pada contoh klasifikasi yang salah yang dilakukan oleh model sebelumnya. Sedangkan peran dari *Stacking* dalam sebuah penelitian adalah memungkinkan penggabungan hasil prediksi dari beberapa model dengan pendekatan yang berbeda untuk menghasilkan klasifikasi akhir yang kuat dan andal.

Dalam upaya pada pengembangan metode ini, dilakukan uji coba untuk meningkatkan kinerja klasifikasi dalam sebuah dataset dengan menggabungkan 3 algoritma yang berbeda di antara *Decision Tree*, *K-Nearest Neighbor* dan *Support Vector Machine (SVM)* dengan metode optimasi yang diterapkan dalam klasifikasi kesuburan tanah untuk meningkatkan nilai akurasi dari model-model yang dibentuk dari ketiga algoritma tersebut. Banyak Teknik optimasi yang telah dikembangkan untuk diterapkan dalam mengatasi *Imbalance Class*. Masalah *Imbalance Class* dapat dikategorikan menjadi 3 jenis, yaitu pendekatan level data, pendekatan level algoritma, dan pendekatan *hybrid* [10]. Untuk menghasilkan nilai akurasi yang tinggi serta peningkatan hasil klasifikasi, 3 algoritma tersebut perlu dilakukan sebuah optimasi [11][12][13] dengan Teknik *ensemble*. Pada penelitian ini menggunakan metode optimasi *soft voting* yang mampu meningkatkan kinerja model. Optimasi *soft voting* merupakan pendekatan yang digunakan untuk mengoptimalkan bobot dari masing-masing model berdasarkan probabilitasnya [14]. Keunggulan menggunakan optimasi ini yaitu selain dapat meningkatkan akurasi model juga dapat memberikan suara untuk setiap kelas tanpa mempertimbangkan model yakin dengan klasifikasi tersebut.

Pada penelitian terdahulu telah dilakukan klasifikasi penyakit stroke menggunakan Algoritma *Naïve Bayes* yang dilakukan pada rumah sakit dan puskesmas di Banjarmasin. Penelitian ini menggunakan jumlah data 572 baris dan jumlah data training 28.500 [11]. Setelah dilakukan penelitian terbukti bahwa implementasi metode optimasi pada sebuah algoritma yang digunakan dalam sebuah penelitian itu mampu menaikkan akurasi dari model yang dibentuk. Meskipun kenaikan akurasinya tidak signifikan tetapi itu mempengaruhi hasil klasifikasi jangka panjang. Akurasi Algoritma *Naïve Bayes* sebelum dilakukan optimasi sebesar 0.976 dengan menggunakan *split* data 80-20. Setelah dilakukan metode optimasi untuk meningkatkan hasil akurasi Algoritma *Naïve Bayes* mengalami kenaikan 0,005 dari akurasi sebelumnya dengan *split* data sebesar 70-30. Kinerja Algoritma *Naïve Bayes* mempunyai waktu klasifikasi yang singkat sehingga mampu mempercepat dalam mengklasifikasikan sebuah dataset yang diolah [15].

Penelitian terdahulu yang lain menerapkan sebuah metode optimasi *Backward Elimination* untuk melakukan klasifikasi kepuasan pelanggan pada sebuah perusahaan. Penelitian ini menggunakan algoritma *k-Nearest Neighbor* dan *Naïve Bayes* [16]. Penggunaan optimasi *Backward Elimination* mampu menaikkan akurasi algoritma *k-NN* secara signifikan dari 63.54 hingga mencapai 97.28 dengan peningkatan sebesar 33,74 Sedangkan pada algoritma *Naïve Bayes* tidak menghasilkan peningkatan akurasi yang signifikan dari 98.92 menjadi 99.04 dengan peningkatan sebesar 0.12.

Meskipun telah ada upaya yang signifikan dalam menentukan klasifikasi kesuburan tanah, sedikit penelitian yang secara khusus menerapkan proses penggabungan beberapa algoritma dengan Optimasi *Soft Voting*. Masih terdapat ruang untuk dapat mengeksplorasi dan menguji Metode *Soft Voting* untuk diterapkan dalam beberapa algoritma dan berbagai permasalahan. Dalam beberapa kasus, mungkin faktor penerapan Optimasi *Soft Voting* mampu meningkatkan akurasi klasifikasi, tetapi perlu dilakukan penelitian lebih lanjut untuk memahami dampak yang lebih jelas dari penerapan Metode *Soft Voting*.

2. METODOLOGI PENELITIAN

2.1. Dataset

Dalam penelitian ini dataset yang digunakan berupa data *privat* yang diperoleh dari Dinas Pertanian Kabupaten Grobogan mengenai kesuburan tanah yang dipengaruhi beberapa aspek unsur hara yang terkandung. Dataset ini memiliki 5000 *record* data yang terdiri dari 16 atribut serta 2520 data yang mengklasifikasikan tanah subur dan 2480 data tanah yang tidak subur. Variabel tersebut saling berkaitan untuk menentukan hasil akhir kondisi tanah subur tidaknya. Untuk rincian data yang digunakan dalam penelitian ini dapat dilihat pada tabel 1 dibawah ini.

Tabel 1. Data awal

<i>Data</i>	<i>Kelas</i>	<i>Atribut</i>	<i>Record Data</i>
<i>Pertanian</i>	<i>1</i>	<i>16</i>	<i>5000</i>

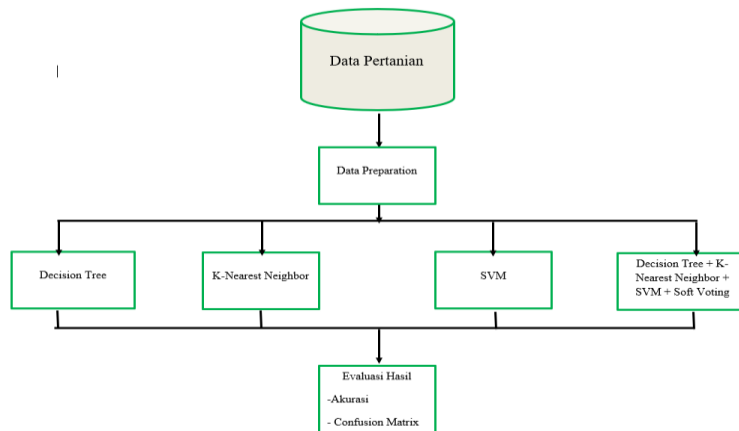
Adapun spesifikasi tipe data atribut dapat dilihat pada tabel 2 dibawah ini.

Tabel 2. Tipe Data

<i>Atribut</i>	<i>Tipe data</i>
<i>pH</i>	<i>Numeric</i>
<i>EC</i>	<i>Numeric</i>
<i>OC</i>	<i>Numeric</i>
<i>OM</i>	<i>Numeric</i>
<i>N</i>	<i>Numeric</i>
<i>P</i>	<i>Numeric</i>
<i>K</i>	<i>Numeric</i>
<i>Zn</i>	<i>Numeric</i>
<i>Fe</i>	<i>Numeric</i>
<i>Cu</i>	<i>Numeric</i>
<i>Mn</i>	<i>Numeric</i>
<i>Sand</i>	<i>Numeric</i>
<i>Silt</i>	<i>Numeric</i>
<i>Clay</i>	<i>Numeric</i>
<i>CaCO3</i>	<i>Numeric</i>
<i>CEC</i>	<i>Numeric</i>

2.2. Skema Penelitian

Penelitian ini secara umum terbagi menjadi 3 bagian utama yaitu *pre-processing*, proses, dan evaluasi. *Pre-processing* bertujuan untuk menyiapkan data yang layak untuk dilakukan pemrosesan. Menyiapkan data yang layak itu meliputi pembersihan data, pengurangan dimensi data, penanganan data yang tidak seimbang dan standarisasi data. Pada bagian proses bertujuan untuk implementasi algoritma pada sebuah dataset guna menghasilkan output yang diinginkan berupa sebuah informasi klasifikasi. Setelah melakukan *pre-processing* dan proses maka langkah terakhir yang harus dilakukan adalah bagian evaluasi. Pada bagian ini dilakukan proses evaluasi yang menggunakan konsep *confusion matrix*s untuk menghitung akurasi kinerja dari algoritma yang diterapkan.



Gambar 1. Skema Penelitian

Berdasarkan gambar 1 terdapat 5 proses komputasi diantaranya:

1. Pre-processing Data yang didalamnya terdapat proses class balancing dan standarisasi data.
2. Penerapan algoritma Decision Tree pada sebuah dataset pertanian yang digunakan dalam penelitian ini.
3. Penerapan algoritma K-NN pada sebuah dataset pertanian yang digunakan dalam penelitian ini
4. Penerapan algoritma SVM pada sebuah dataset pertanian yang digunakan dalam penelitian ini
5. Penerapan algoritma decision tree, K-NN, SVM menggunakan metode optimasi soft voting untuk meningkatkan akurasi.
6. Melakukan evaluasi akhir dengan menentukan akurasi baik dari per algoritma maupun yang sudah di optimasi menggunakan metode soft voting.

Keenam langkah tersebut dilakukan guna menunjang keberhasilan penelitian ini. Langkah pertama menerapkan algoritma yang digunakan dalam penelitian ini dan dihitung akurasinya. Selanjutnya menerapkan metode optimasi soft voting untuk meningkatkan akurasi dari algoritma yang digunakan dalam penelitian ini. Selain meningkatkan akurasi penggunaan metode soft voting juga berfungsi untuk memaksimalkan kinerja dari algoritma yang digunakan karena metode ini mampu menangani ketidakpastian dalam melakukan klasifikasi dalam sebuah dataset. Langkah terakhir yaitu melakukan evaluasi akhir dengan menelaah hasil akurasi per algoritma dengan yang sudah dilakukan optimasi. Kesimpulan apa yang didapatkan dari langkah-langkah yang sudah dilakukan sebelumnya.

2.3. Data Preparation

Kondisi sebuah data itu tidak bisa di klasifikasi tanpa dasar harus melakukan analisis data terlebih dahulu untuk menentukan Langkah apa yang harus dilakukan terhadap data tersebut. Langkah ini dalam pembelajaran mesin sering disebut dengan *data preparation*. Tujuan melakukan Langkah ini adalah untuk menyiapkan dan memastikan sebuah data yang digunakan dalam analisis, pemodelan adalah data yang berkualitas tinggi, akurat dan siap untuk digunakan. Dalam penelitian ini terdapat 3 langkah data preparation diantaranya *mapping data*, *oversampling data* dan *scaling data*.

Mapping data merupakan sebuah Teknik dalam pembelajaran mesin yang digunakan untuk mengubah dataset menjadi bentuk yang dapat digunakan dalam *machine learning* yaitu bentuk numerik. Tujuan *mapping data* ini adalah untuk memungkinkan penggunaan algoritma *machine learning*, meningkatkan kualitas data karena mudah untuk diproses dan mengoptimalkan kinerja model. Dalam penelitian ini *class data* yang memungkinkan dilakukan *mapping data* karena nilai dari atribut tersebut berbentuk nominal dan harus untuk diubah. Setelah melampaui Langkah ini data yang awalnya berbentuk nominal akan menjadi numerik yang selanjutnya akan dilanjutkan dengan proses lain untuk menyiapkan data yang berkualitas tinggi.

Oversampling data merupakan suatu langkah yang umum dilakukan dalam pembelajaran mesin karena mampu mengatasi ketidakseimbangan jumlah data. Dalam sebuah data awal kondisi jumlah data yang akan diproses itu beragam ada yang sudah seimbang ada juga jumlah data yang tidak seimbang jika hal tersebut tidak ditangani dengan perlakuan khusus maka akan mempengaruhi hasil klasifikasi karena terdapat jumlah data yang dominan. Dalam penelitian ini terdapat 2 *class* pada data yang digunakan dalam penelitian ini yaitu *fertile* dan *non-fertile* yang disimbolkan "1" untuk class *fertile* dan "2" untuk class *non-fertile*. Kondisi data sebelum dilakukan proses *pre-processing* itu beragam tetapi kebanyakan class data belum seimbang yang dikhawatirkan akan mempengaruhi hasil klasifikasi karena ada salah satu *class* yang dominan. Pada data ini jumlah *class fertile* dan *non-fertil* memiliki *range* nilai yang cukup jauh oleh karena itu perlu dilakukan *oversampling class*. Setelah dilakukan proses ini jumlah *class* menjadi setara dari jumlah *class fertile* 2520 dan *non-fertile* 2480 menjadi sama-sama 2480.

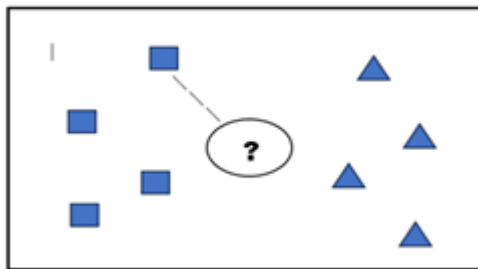
Langkah data preparation yang terakhir adalah *scaling data*. *Scaling data* merupakan sebuah proses untuk mengubah nilai atribut sehingga memiliki *mean* sebesar 0 dan standar deviasi sebesar 1. Hal ini dilakukan karena dalam pembelajaran mesin nilai atribut dari data awal itu mempengaruhi kinerja mesin sebab memiliki skala nilai atribut yang besar. Maka dari itu dibutuhkan proses *scaling data* atau standarisasi data untuk mengatasi ini semua dan mempermudah kinerja mesin menjadi lebih ringan.

2.4. *Klasifikasi Algoritma Decision Tree*

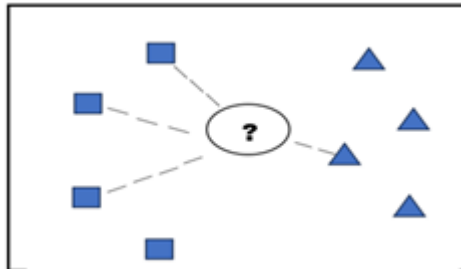
Algoritma *Decision Tree* merupakan sebuah metode yang umum dilakukan pada pembelajaran mesin khususnya klasifikasi data. Metode *Decision Tree* mengubah sejumlah besar fakta menjadi pohon keputusan mewakili aturan. Pohon keputusan juga berguna untuk eksplorasi data, menemukan hubungan tersembunyi antara beberapa kandidat variabel masukan dan suatu variabel target. Setiap pohon pada pohon keputusan tersebut memiliki cabang yang mewakili atribut-atribut di dalamnya yang harus dipenuhi syarat atau aturan untuk menuju ke *node* atau cabang selanjutnya hingga sampai cabang yang tidak memiliki cabang lagi atau bisa disebut daun pada pohon keputusan. Keunggulan menggunakan algoritma ini dalam pembelajaran mesin adalah algoritma ini mudah untuk dipahami dan diinterpretasikan dalam sebuah data karena mampu menganalisis proses pengambilan keputusan yang kompleks menjadi lebih sederhana dengan memanfaatkan konsep *tree* atau pohon keputusan. Cocok juga untuk diterapkan dalam dataset yang memiliki dimensi kecil maupun besar dan konsep pohon keputusannya dapat divisualisasikan menjadi lebih sederhana sehingga mudah untuk di analisis.

2.5. *Algoritma K-Nearest Neighbor*

Algoritma *K-NN*, atau *K-Nearest Neighbors*, adalah metode pengenalan pola yang berguna untuk memahami data tanpa pengetahuan model sebelumnya. Ini adalah salah satu teknik klasifikasi paling dasar yang dapat digunakan ketika kita memiliki sedikit atau tidak ada informasi tentang bagaimana data didistribusikan. *K-NN* bekerja dengan mencari tetangga terdekat dari sampel data, dan aturan paling sederhana adalah ketika $K=1$, artinya kita hanya mempertimbangkan satu tetangga terdekat. Selain digunakan untuk klasifikasi, *K-NN* juga berguna dalam *regresi*, dimana kita mencari mean dari k tetangga terdekat untuk mengklasifikasi nilai atribut suatu objek. Pendekatan ini memungkinkan untuk lebih menekankan kontribusi tetangga terdekat dalam klasifikasi.



Gambar 2. Cara Kerja KNN dengan $k = 1$



Gambar 3. Cara Kerja KNN dengan $k = 4$

Gambar 2 dan gambar 3 di atas memperlihatkan konsep Algoritma K-NN. Pada gambar 2 memperlihatkan konsep algoritma KNN jika nilai K yang digunakan 1 sedangkan gambar 3 memperlihatkan konsep algoritma KNN jika nilai K yang digunakan 4.

2.6. Algoritma Support Vector Machine

Algoritma *Support Vector Machine* merupakan sebuah algoritma yang digunakan dalam sebuah pembelajaran mesin yang menggunakan spekulasi berupa fungsi *linier* dalam sebuah data yang berdimensi tinggi dan menghasilkan sebuah model yang dilatih dengan algoritma yang didasarkan dengan Teknik optimasi. Nilai akurasi yang dihasilkan algoritma ini sangat bergantung pada nilai kernel dan parameter yang digunakan dalam sebuah penelitian. *Support Vector Machine* dikategorikan menjadi 2 jenis berdasarkan karakteristiknya yaitu *support vector machine linier* dan *non-linier*. Dalam pembagian datanya istilah dasar yang harus dipahami adalah *hyperplane*. *Hyperplane* mengacu pada sebuah sub ruang yang memiliki ukuran 1 unit lebih kecil daripada ukuran ruang keseluruhannya dan mampu membagi data menjadi dua bagian yang setara. *Support vector machine linear* merupakan sebuah metode yang digunakan untuk memisahkan dan membagi data dari *hyperplane* menggunakan *penalized margin*. Sedangkan *support vector machine non-linier* merupakan sebuah metode untuk melakukan pemetaan data dari data input yang berdimensi rendah ke ruang fitur yang berdimensi tinggi dan biasanya Teknik ini sering disebut dengan fungsi kernel *trick*.

Keunggulan algoritma ini adalah mampu menangani keadaan data yang tidak seimbang dan tahan terhadap *overfitting* karena *SVM* ini berusaha untuk menentukan *hyperplane* yang terbaik dengan margin maksimum yang meminimalisir resiko *overfitting*. Namun *Support Vector Machine* juga memiliki kelemahan yaitu rentan terhadap sebuah parameter yang telah di *tuning* buruk di awalnya. Buruknya *t* dalam sebuah parameter mencakup beberapa aspek diantaranya seleksi fitur yang tidak tepat, skala fitur yang tidak sesuai, kesalahan dalam merepresentasikan data dan pengambilan sampel yang tidak seimbang. Selain itu algoritma ini juga sulit diterapkan pada kernel yang kompleks. Jenis kernel yang dimaksud adalah kernel RBF (*Radial Basis Function*). Jenis kernel ini memiliki dimensi yang kompleks sehingga jenis kernel ini sulit untuk diterapkan dalam algoritma ini. Untuk parameter yang digunakan, *SVM* memiliki beberapa parameter diantaranya parameter *penalty*, *gamma* (parameter kernel), *degree* (degree dari kernel polynomial) dan lain sebagainya. Untuk menentukan kernel dan parameter yang digunakan dalam algoritma ini harus menganalisis karakteristik data yang digunakan. Pemilihan model juga harus disesuaikan dengan karakteristik yang diinginkan dari sebuah penelitian.

2.7. Optimasi Soft Voting

Optimasi Soft voting merupakan salah satu metode dalam *ensemble learning* yang mampu meningkatkan kinerja model *ensemble* dan meningkatkan akurasi dari algoritma yang digunakan. Optimasi ini umum digunakan dalam pembelajaran mesin khususnya dalam klasifikasi data yang dimana kita memiliki beberapa model yang sudah dilatih untuk membuat akurasi yang lebih baik. Akurasi yang lebih baik tentunya akan menghasilkan klasifikasi yang baik juga pada data yang dianalisis. Dalam metode ini pemilihan model pembelajaran mesin yang digunakan dilakukan dengan berfokus pada hasil klasifikasi probabilitasnya dari masing-masing model yang telah dibentuk. Langkah-langkah metode Optimasi *Soft Voting* dalam meningkatkan hasil akurasi dari algoritma *Decision Tree*, *K-Nearest-Neighbor*, dan *SVM* pada kasus kesuburan tanah adalah sebagai berikut:

- a. Pemilihan algoritma: Pada proses ini akan dipilih algoritma yang digunakan. Dalam penelitian ini menggunakan algoritma *Decision Tree*, *K-Nearest-Neighbor*, dan *SVM* sebagai model yang akan digunakan dalam *ensemble* untuk meningkatkan akurasi dalam klasifikasi kesuburan tanah.

- b. Pelatihan Model: Proses ini melatih masing-masing model yang dibentuk dari algoritma yang digunakan menggunakan data pelatihan yang relevan. Selama proses pelatihan, model-model yang dibentuk akan belajar dari data dan merepresentasikan menjadi model yang mampu melakukan klasifikasi.
- c. Prediksi: Model-model akan digunakan dalam mengklasifikasi kesuburan tanah pada data uji atau data yang belum pernah dilihat sebelumnya.
- d. Penggabungan Hasil: Pada proses ini menggabungkan hasil klasifikasi dari masing-masing model. Dalam metode optimasi soft voting akan diberikan bobot pada hasil klasifikasi dari masing-masing model yang dibentuk berdasarkan performa algoritma yang digunakan.
- e. Perhitungan Klasifikasi Akhir: Menghitung hasil klasifikasi dengan menggabungkan hasil klasifikasi dari masing-masing model dengan bobot yang sesuai
- f. Penentuan Hasil Klasifikasi: Menggunakan perhitungan klasifikasi akhir untuk menentukan kelas atau label kategori kesuburan tanah. Menentukan kelas dalam proses ini yang dimaksud adalah mengidentifikasi keadaan tanah dengan acuan unsur hara. Tanah tersebut dalam kategori subur atau tidak subur.
- g. Evaluasi dan Validasi: Dalam proses ini akan dilakukan evaluasi dan validasi dari hasil yang diperoleh dengan menggunakan konsep confusion matrix seperti menghitung nilai akurasi, presisi dan recall. Evaluasi ini dapat memastikan bahwa penggunaan metode optimasi soft voting memberikan hasil yang lebih baik daripada model Tunggal.
- h. Pengaturan Bobot: Langkah ini opsional untuk dilakukan karena harus menyesuaikan dengan keadaan data. Pada beberapa kasus mungkin perlu mengatur ulang bobot yang diberikan pada hasil klasifikasi dari model-model yang telah dibentuk berdasarkan performa actual pada data uji atau data validasi.
- i. Optimasi lebih lanjut: Langkah ini juga opsional untuk dilakukan karena dalam langkah ini akan dilakukan eksperimen dengan berbagai bobot atau bahkan mencoba metode optimasi lain untuk mencapai hasil yang lebih baik.

2.8. Confusion Matrix

Confusion matrix merupakan sebuah matriks yang berukuran 2x2 untuk merepresentasi hasil klasifikasi pada sebuah data. Hasil klasifikasi ini diubah menjadi bilangan biner agar mudah untuk dianalisis. Dengan menggunakan *confusion matrix* itu memudahkan kita untuk melakukan analisis dan pemahaman sejauh mana kinerja model yang telah kita bentuk untuk mengklasifikasi data yang akurat dan mampu mengetahui tingkat dan jenis kesalahan dari model yang telah dibentuk. Hasil nilai akurasi, presisi dan recal akan ditampilkan dalam bentuk persentase.

- a. Akurasi. Akurasi merupakan jumlah rasio klasifikasi yang benar dari sebuah data. Rumus akurasi dapat dilihat dari rumus 1 dibawah ini:

$$\text{Akurasi} = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

- b. Presisi. Presisi merupakan sebuah matriks evaluasi yang digunakan untuk mengukur sejauh mana keberhasilan model klasifikasi dalam mengidentifikasi *instance* positif dengan akurat dari semua *instance* yang bernilai positif. Berikut ini merupakan rumus dari presisi:

$$\text{Presisi} = \frac{TP}{TP+FP} \quad (2)$$

- c. Recall. Recall merupakan sebuah matriks evaluasi yang digunakan untuk mengukur sejauh mana keberhasilan model klasifikasi dalam menganalisis *instance* positif yang sebenarnya. Berikut ini merupakan rumus dari recall:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

3. HASIL DAN PEMBAHASAN

3.1. Data Preparation

Data yang digunakan dalam penelitian ini memiliki 2 kondisi yang harus diperbaiki diantaranya *class* data yang masih berbentuk nominal dan keadaan data yang tidak seimbang serta dibutuhkan langkah lain dalam data preparation yaitu *scaling data* yang tidak kalah pentingnya dengan langkah yang lain karena mampu meningkatkan kinerja mesin menjadi lebih efisien dan tentunya cepat. Langkah awal yang dilakukan pada *data preparation* adalah *mapping data*. *Mapping data* mampu mengatasi *class* data yang masih berbentuk nominal diubah menjadi bentuk numerik sehingga memungkinkan dapat diterapkan algoritma *machine learning* di dalamnya. Berikut ini data sebelum diproses dengan *mapping data*.

Tabel 3. Data sebelum mapping

	<i>pH</i>	<i>EC</i>	<i>OC</i>	<i>OM</i>	<i>N</i>	<i>P</i>	<i>K</i>	<i>Zn</i>	<i>Fe</i>	<i>Cu</i>	<i>Mn</i>	<i>Sand</i>	<i>Silt</i>	<i>Clay</i>	<i>CaCO3</i>	<i>CEC</i>	<i>Output</i>
3	9.48	0.13	0.10	0.18	107	14.65	136	0.65	41	0.12	75	91.9	10.5	5.1	0.00	8.86	Fertile
4	8.71	0.25	0.01	0.08	232	35.16	112	0.19	24	0.34	16	81.8	8.4	9.8	0.01	5.04	Fertile

Berdasarkan tabel 3 diatas terdapat *class* yang memiliki bentuk nominal yaitu *class output*. Maka dari itu dibutuhkan langkah ini untuk mengubah nilai yang berbentuk nominal menjadi bentuk numerik sehingga dapat diproses oleh mesin. Berikut ini adalah tabel data yang sudah diproses dengan *mapping data*.

Tabel 4. Data setelah mapping

	<i>pH</i>	<i>EC</i>	<i>OC</i>	<i>OM</i>	<i>N</i>	<i>P</i>	<i>K</i>	<i>Zn</i>	<i>Fe</i>	<i>Cu</i>	<i>Mn</i>	<i>Sand</i>	<i>Silt</i>	<i>Clay</i>	<i>CaCO3</i>	<i>CEC</i>	<i>Output</i>
3	9.48	0.13	0.0	0.18	107	14.65	136	0.65	4.1	0.12	7.5	91.9	10.5	5.1	0.00	8.86	1
4	8.71	0.25	0.1	0.08	232	35.16	112	0.19	2.4	0.34	1.6	81.8	8.4	9.8	0.01	5.04	1

Bentuk *class output* setelah mengalami proses *mapping data* berubah menjadi bentuk numerik. 1 itu menandakan bahwa kondisi tanahnya subur atau *fertile* dan 0 menandakan kondisi tanah yang kurang subur atau *non-fertile*. Langkah selanjutnya dalam *data preparation* adalah *oversampling data*. Langkah ini harus dilakukan karena data awal mengalami ketidakseimbangan jumlah data yang dikhawatirkan akan mempengaruhi hasil klasifikasi karena terdapat jumlah data yang dominan. Data mentah yang diperoleh terdapat 5000 *record* dengan 2 *output fertile* dan *non-fertile*. Kedua *class* tersebut harus diubah menjadi angka karena pada pembelajaran mesin data yang diolah harusnya berbentuk angka. Perbandingan jumlah data dapat dilihat dari tabel 5 dibawah ini.

Tabel 5. Data Sebelum Oversampling

	<i>fertile (1)</i>	<i>non fertile (0)</i>
Jumlah Data	2520	2480

Dapat dilihat dari tabel diatas jumlah data antara *fertile* dan *non fertile* tidak seimbang. Dengan demikian perlu dilakukan proses *oversampling data* agar jumlah data yang dihasilkan seimbang atau tidak ada jumlah data *class* yang dominan. *Oversampling data* mengubah jumlah data *class* yang dominan menjadi sama dengan jumlah data yang *subordian*. Jumlah data yang *subordian* adalah jumlah data yang dilihat secara jumlah di bawah dari jumlah data yang dominan. Berikut ini adalah data dari hasil proses *oversampling data* dapat dilihat pada tabel 6 dibawah ini.

Tabel 6. Data setelah oversampling

	<i>fertile (1)</i>	<i>non fertile (0)</i>
<i>Jumlah Data</i>	2480	2480

Jumlah data sudah seimbang jadi untuk penentuan klasifikasi data lebih *fair* karena tidak ada lagi data yang tidak seimbang. Proses selanjutnya adalah proses *scaling data* yang bertujuan untuk mempercepat proses kerja mesin karena ukuran data telah diubah ke dalam *range* 0 hingga 1. Langkah terakhir dalam *data preparation* adalah *scalling data*. *Scalling data* bertujuan untuk mencegah nilai atribut dalam skala besar yang tentunya akan mendominasi perhitungan dan menghambat kinerja mesin. Hal ini jika dibiarkan kinerja mesin akan berat dan dapat mempengaruhi hasil analisis klasifikasi suatu data. *Scaling data* mengubah skala nilai yang besar menjadi dalam *range* 0 hingga 1 sehingga mudah untuk dilakukan pemahaman dan interpretasi data. Untuk lebih memahami proses *scaling data*, berikut disajikan tabel 7 yang berisi sebuah data sebelum dilakukan *scaling data* dengan tabel yang sudah dilakukan proses *scaling data*.

Tabel 7. Data sebelum Scaling

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

Data dari tabel 7 diatas terlihat jelas terdapat jarak antara data satu dengan data yang lainnya. Seperti contoh nilai 388 dibandingkan dengan nilai 0.01 tentunya terdapat jarak yang jauh. Dalam pembelajaran mesin hal ini menjadi masalah karena nilai dari data tersebut dianggap besar oleh mesin. Maka dari itu dibutuhkan Langkah ini untuk mengecilkan skala dari nilai data. Berikut merupakan data yang sudah dilakukan proses *scaling data* dapat dilihat dari tabel 8 dibawah ini.

Tabel 8. Data setelah scaling

<i>Atribut</i>	<i>Nilai</i>
<i>pH</i>	0.828095782
<i>EC</i>	-0.318343675
<i>OC</i>	-0.192919002
<i>OM</i>	-0.188486833
<i>N</i>	1.672786853
<i>P</i>	-0.383982762
<i>K</i>	1.962261632
<i>Zn</i>	-0.038325819
<i>Fe</i>	-0.151271996
<i>Cu</i>	-0.079174971
<i>Mn</i>	-0.026125838
<i>Sand</i>	1.384107169
<i>Silt</i>	-0.982856116
<i>Clay</i>	-0.969005053
<i>CaCO₃</i>	-0.475793405
<i>CEC</i>	-0.058853868

Bandingkan dengan data sebelum dilakukan proses *scaling data* terlihat jelas nilainya berubah menjadi *range* 0 -1 dan mesin dapat memproses data itu dengan cepat dan mudah karena skala nilai yang kecil.

3.2. Hasil Penelitian

Pada penelitian ini menggunakan dataset berupa data kesuburan tanah dengan 16 parameter acuan yang diperoleh dari Dinas Pertanian Kabupaten Grobogan. Algoritma yang digunakan dalam penelitian ini adalah Algoritma *Decision Tree*, Algoritma *K-Nearest Neighbor* dan *Support Vector Machine* yang dioptimasi kinerja dan hasil akurasi dengan menerapkan metode *soft voting* pada model yang telah dibentuk dan bergantung pada probabilitas setiap modelnya. Setelah dilakukan riset dan analisis dari algoritma yang digunakan pada penelitian ini serta menerapkan optimasi pada model yang telah dibentuk akan menghasilkan akurasi yang nantinya akan dianalisis lebih lanjut untuk menentukan akurasi yang lebih tinggi.

Tabel 9. Hasil akurasi

<i>Algoritma</i>	<i>Akurasi</i>
<i>Algoritma Decision Tree</i>	88,7%
<i>Algoritma K-Nearest Neighbor</i>	86,7%
<i>Algoritma SVM</i>	90,1%
<i>Decision Tree + K-Nearest Neighbor + SVM + Soft Voting</i>	90,4%

Berdasarkan Tabel 9 diatas diperoleh nilai akurasi dari algoritma yang digunakan dalam penelitian ini serta implementasi dari metode optimasi *soft voting*. Dapat disimpulkan bahwa algoritma yang digunakan dalam penelitian ini jika diimplementasikan metode optimasi *soft voting* akan menghasilkan akurasi lebih tinggi daripada akurasi per algoritma. Meskipun kenaikan akurasi tidak signifikan tetapi itu juga berpengaruh terhadap klasifikasi sebuah data yang nantinya akan dianalisis.

3.3. Representasi Confusion Matrix

Dari algoritma yang digunakan dalam penelitian ini perlu untuk dilakukan evaluasi hasil dari kinerja model yang telah dibentuk menggunakan konsep *confusion matrix*. Dengan menggunakan

confusion matrix memudahkan kita untuk melakukan analisis sejauh mana kinerja model yang telah dibentuk untuk melakukan klasifikasi dari data yang diolah. Berikut ini penjelasan mengenai representasi data menggunakan konsep *confusion matrix* yang dapat dilihat dari Tabel 10.

Tabel 10. Confusion matrix kinerja Decision Tree

		Prediction		Recall
		Non-fertile	Fertile	
Actual	non-fertile	434	72	0.92
	fertile	40	446	0.86
Precision		0.85	0.91	

Tabel 10 menampilkan hasil evaluasi kinerja algoritma *decision tree* yang diterapkan pada sebuah data. Dari 5000 data yang diuji, 434 data kelas *non-fertile* di klasifikasikan dengan benar, dan 40 data dari kelas *no-fertile* yang di klasifikasikan sebagai kelas *fertile*. Kemudian 446 data *fertile* yang di klasifikasikan benar, dan 72 data *fertile* yang di klasifikasikan sebagai kelas *non-fertile*. Dari tabel diatas dapat dilihat nilai presisi dan recall dari masing-masing kelas. Nilai presisi untuk *non-fertile* sebesar 0.85 dan untuk kelas *fertile* sebesar 0.91 dengan rata-rata presisi sebesar 0.88. Sedangkan untuk nilai recall dari kelas *non-fertile* sebesar 0.92 dan untuk kelas *fertile* sebesar 0.86 dengan rata-rata sebesar 0.89. Selanjutnya evaluasi *confusion matrix* untuk kinerja algoritma *K-Nearest Neighbor* yang ditampilkan pada tabel 11 dibawah. Ini.

Tabel 11. Confusion matrix kinerja KNN

		Prediction		Recall
		Non-fertile	Fertile	
Actual	non-fertile	451	55	0.85
	fertile	76	410	0.88
Precision		0.89	0.84	

Tabel 11 menampilkan hasil evaluasi kinerja algoritma *decision tree* yang diterapkan pada sebuah data. Dari 5000 data yang diuji, 451 data kelas *non-fertile* di klasifikasikan dengan benar, dan 76 data dari kelas *no-fertile* yang di klasifikasikan sebagai kelas *fertile*. Kemudian 410 data *fertile* yang di klasifikasikan benar, dan 55 data *fertile* yang di klasifikasikan sebagai kelas *non-fertile*. Dari taabel diatas dapat dilihat nilai presisi dan recall dari masing-masing kelas. Nilai presisi untuk *non-fertile* sebesar 0.89 dan untuk kelas *fertile* sebesar 0.84 dengan rata-rata presisi sebesar 0.865. Sedangkan untuk nilai recall dari kelas *non-fertile* sebesar 0.85 dan untuk kelas *fertile* sebesar 0.88 dengan rata-rata sebesar 0.865. Selanjutnya evaluasi *confusion matrix* untuk kinerja algoritma *SVM* yang ditampilkan pada tabel dibawah. Ini.

Tabel 12. Confusion matrix kinerja SVM

		Prediction		Recall
		Non-fertile	Fertile	
Actual	non-fertile	461	45	0.89
	fertile	53	433	0.90
Precision		0.91	0.89	

Tabel 12 menampilkan hasil evaluasi kinerja algoritma *decision tree* yang diterapkan pada sebuah data. Dari 5000 data yang diuji, 461 data kelas *non-fertile* di klasifikasikan dengan benar, dan 53 data dari kelas *no-fertile* yang diklasifikasikan sebagai kelas *fertile*. Kemudian 433 data *fertile* yang di klasifikasikan benar, dan 45 data *fertile* yang di klasifikasikan sebagai kelas *non-fertile*. Dari taabel diatas dapat dilihat nilai presisi dan recall dari masing-masing kelas. Nilai presisi untuk *non-fertile* sebesar 0.85 dan untuk kelas *fertile* sebesar 0.91 dengan rata-rata presisi sebesar 0.88. Sedangkan untuk nilai recall dari kelas *non-fertile* sebesar 0.92 dan untuk kelas *fertile* sebesar 0.86 dengan rata-rata sebesar 0.89. Selanjutnya evaluasi *confusion matrix* untuk kinerja algoritma *k-Nearest Neighbor* yang ditampilkan pada tabel 13.

Tabel 13. Confusion matrix kinerja Algoritma Decision Tree, KNN, dan SVM dengan Soft Voting

		Prediction		Recall
		Non-fertile	Fertile	
Actual	non-fertile	450	56	0.92
	fertile	39	447	0.88
Precision		0.88	0.91	

Confusion matrix diatas merupakan hasil representasi data biner dari representasi metode optimasi *soft voting* pada ketiga algoritma. Dari matriks 2x2 diatas menghasilkan nilai *true* positif sebesar 450, *true* negative sebesar 447, *false* positif sebesar 56 dan *false* negative sebesar 39.

3.4. Analisa Hasil Pengujian

Berdasarkan hasil dari pengujian model klasifikasi pada penelitian ini diperoleh hasil bahwa implementasi metode optimasi *soft voting* menghasilkan akurasi sebesar 90,4. Akurasi ini mengalami peningkatan dari akurasi per algoritma yang digunakan meskipun tidak signifikan. Hasil dari pengujian dapat diidentifikasi secara kuantitatif maupun kualitatif. Secara kuantitatif hasil pengujian ini menggunakan konsep dari *confusion matrix* yang didalamnya dapat diimplementasikan matriks kinerja seperti presisi, recall, dan akurasi. Matriks kinerja tersebut mampu mengukur sejauh mana kinerja algoritma dan metode optimasi yang digunakan dalam klasifikasi kesuburan tanah berdasarkan parameter kandungan unsur zat hara. Seperti contoh ditinjau dari hasil akurasi pengujian Algoritma *Decision Tree* dalam klasifikasi kesuburan tanah menghasilkan nilai akurasi sebesar 88,77 %, algoritma *k-Nearest Neighbor* menghasilkan akurasi sebesar 86,7%, dan algoritma *SVM* menghasilkan nilai akurasi sebesar 90,1%. Tujuan utama dari penelitian ini adalah mengoptimasi atau meningkatkan hasil akurasi klasifikasi gabungan algoritma dengan sebuah metode optimasi. Pada penelitian ini menggunakan metode optimasi *soft voting*. Penggunaan metode *soft voting* berhasil membuktikan bahwa nilai akurasi mengalami peningkatan dari nilai akurasi per algoritma. Hasil ini akan lebih baik dalam klasifikasi dan meminimalisir hasil klasifikasi yang tidak tepat karena telah mengalami peningkatan akurasi dari algoritma yang digunakan. Untuk presisi dan recall pada setiap algoritma dengan gabungan algoritma yang sudah dioptimasi mengalami keadaan yang fluktuatif atau data yang hasilnya naik turun. Berdasarkan hasil tersebut penambahan metode optimasi *soft voting* hanya mempengaruhi nilai akurasi saja tidak dengan presisi maupun recall. Hasil pengujian juga dapat diidentifikasi secara kualitatif. Hasil penelitian ini berkontribusi dalam bidang pertanian khususnya pada kesuburan tanah dengan acuan kandungan unsur zat hara. Langkah ini tentunya membantu kinerja badan penyuluh pertanian terkait dan dinas pertanian suatu daerah untuk mengklasifikasi kesuburan tanah dengan akurat karena proses yang dilakukan sudah terkomputerisasi dan dapat meminimalisir kesalahan dalam hasil klasifikasi. Model-model machine learning, seperti *Decision Tree*, *k-Nearest Neighbor* dapat berinteraksi secara sinergis untuk memberikan kontribusi dalam menghasilkan klasifikasi kesuburan tanah. *Decision Tree* mampu mengestrak pola dalam data mungkin lebih efektif dalam mengidentifikasi karakteristik tanah yang memiliki kandungan unsur hara tertentu. Sementara itu *SVM* dengan kemampuannya dapat memisahkan data yang kompleks dan dapat dioperasikan dalam situasi dimana kandungan tanah memiliki variasi yang tinggi. Melalui interaksi algoritma-algoritma yang digunakan dalam penelitian ini dengan metode optimasi *soft voting* dapat memaksimalkan kelebihan masing-masing algoritma untuk memberikan solusi klasifikasi terbaik yang lebih adaptif dengan kandungan unsur hara dalam tanah. Analisis kualitatif juga berperan penting dalam memberikan deskripsi yang lebih rinci tentang hasil klasifikasi dalam penelitian ini. Hasil klasifikasi mengalami peningkatan setelah diimplementasikan metode optimasi *soft voting* dalam kemampuan untuk klasifikasi kesuburan tanah. Hal ini memiliki implikasi positif bagi petani, badan penyuluh terkait dan dinas terkait

karena dapat menentukan apa yang harus dilakukan kedepannya untuk dapat meningkatkan produktivitas di area sebelumnya yang dianggap kurang subur.

4. KESIMPULAN

Berdasarkan hasil dan pembahasan dari penelitian ini dapat disimpulkan bahwa penggunaan metode optimasi soft voting dalam mengklasifikasi kesuburan tanah ternyata dapat meningkatkan performa akurasi klasifikasi dari gabungan algoritma yang digunakan. Hasil akurasi Algoritma Decision Tree sebesar 88,7%, akurasi Algoritma K-Nearest Neighbor sebesar 86,7%, hasil akurasi dari Algoritma SVM sebesar 90,1% dan Hasil akurasi dari proses penggabungan Algoritma Decision Tree, K-Nearest Neighbor dan SVM menggunakan Metode Optimasi Soft Voting sebesar 90,4%. Meskipun hasil akurasi tidak begitu signifikan kenaikannya tetapi itu juga dapat berpengaruh pada hasil klasifikasi. Penelitian selanjutnya diharapkan mampu menerapkan teknik optimasi lain untuk dikombinasikan dengan metode soft voting pada algoritma yang ingin digunakan dalam sebuah penelitian.

DAFTAR PUSTAKA

- [1] Khairad, F. (2020). "Sektor Pertanian di Tengah Pandemi Covid-19 Ditinjau dari Aspek Agribisnis." *Journal Agriuma*, 2(2), 82–9.
- [2] Heru Sandi, G., & Fatma, Y. (2023). "Pemanfaatan Teknologi Internet of Things (IoT) Pada Bidang Pertanian." *JATI (Jurnal Mhs Tek Inform)*, 7(1), 1–5. Heru Sandi G, Fatma Y. Pemanfaatan Teknologi Internet of Things (IoT) Pada Bidang Pertanian. *JATI (Jurnal Mhs Tek Inform*. 2023;7(1):1–5.
- [3] Chodijah, S. (2018). "Strategi Komunikasi Penyampaian Informasi Iklim Stasiun Klimatologi Sampali Medan Dalam Upaya Meminimalkan Kegagalan Panen Padi Sawah Akibat Iklim Ekstrim." *Persepsi Commun J*, 1(1), 55–69.
- [4] Trisnawati, A. (2022). "Analisis Status Kesuburan Tanah Pada Kebun Petani Desa Ladogahar Kecamatan Nita Kabupaten Sikka." *J Locus Penelit dan Pengabd*, 1(2), 68–80.
- [5] Siregar, M. A. R. (2023). "Peningkatan Produktivitas Tanaman Padi Melalui Penerapan Teknologi Pertanian Terkini," 1–11.
- [6] Aini, D. N., Oktavianti, B., Husain, M. J., Sabillah, D. A., Rizaldi, S. T., & Mustakim, M. (2022). "Seleksi Fitur untuk Prediksi Hasil Produksi Agrrikultur pada Algoritma K-Nearest Neighbor (KNN)." *J Sist Komput dan Inform*, 4(1), 140.
- [7] Alkhairi, P., & Windarto, A. P. (2019). "Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara." *Semin Nas Teknol Komput Sains*, 762–7.
- [8] Rofiq, H., Pelangi, K. C., & Lasena, Y. (2020). "Penerapan Data Mining Untuk Menentukan Potensi Hujan Harian Dengan Menggunakan Algoritma Naive Bayes." *J Manaj Inform dan Sist Inf*, 3(1), 8–15.
- [9] Adriano, B., Xia, J., Baier, G., Yokoya, N., & Koshimura, S. (2019). "Multi-source data fusion based on ensemble learning for rapid building damage mapping during the 2018 Sulawesi earthquake and Tsunami in Palu, Indonesia." *Remote Sens*, 11(7).
- [10] Aridas, C. K., Karlos, S., Kanas, V. G., Fazakis, N., & Kotsiantis, S. B. (2020). "Uncertainty Based Under-Sampling for Learning Naive Bayes Classifiers under Imbalanced Data Sets." *IEEE Access*, 8, 2122–33.

- [11] Byna, A., & Basit, M. (2020). "Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes." *J Sisfokom (Sistem Inf dan Komputer)*, 9(3), 407–11.
- [12] Prasetio, R. T., & Ripandi, E. (2019). "Optimasi Klasifikasi Jenis Hutan Menggunakan Deep Learning Berbasis Optimize Selection." *J Inform*, 6(1), 100–6.
- [13] Liklikwatil, R. D., Noersasongko, E., & Supriyanto, C. (2018). "Optimasi K-Nearest Neighbor Dengan Particle Swarm Optimization Untuk Memprediksi Harga Komoditi Karet." *e-Jurnal JUSITI (Jurnal Sist Inf dan Teknol Informasi)*, 7–2(2), 172–82.
- [14] Mahabub, A. (2020). "A robust technique of fake news detection using Ensemble Voting Classifier and comparison with other classifiers." *SN Appl Sci*, 2(4), 1–9.
- [15] Rahayu, A. S., Fauzi, A., & Rahmat, R. (2022). "Komparasi Algoritma Naïve Bayes Dan Support Vector Machine (SVM) Pada Analisis Sentimen Spotify." *J Sist Komput dan Inform*, 4(2), 349.
- [16] Yunitasari, Hopipah, H. S., & Mayasari, R. (2021). "Optimasi Backward Elimination untuk Klasifikasi Kepuasan Pelanggan Menggunakan Algoritme k-nearest neighbor (k-NN) and Naive Bayes." *Technomedia J*, 6(1), 99–110.