OPEN ACCESS

Volume 9 (4), October-December 2025, 1437-1444

E-ISSN:2580-1643

Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)

DOI: https://doi.org/10.35870/jtik.v9i4.4049

Klasifikasi Kualitas Tanah Berdasarkan Kandungan pH, Kelembapan, dan Suhu Menggunakan Algoritma K-Nearest Neighbors

Md. Wira Putra Dananjaya [1*], Gede Humaswara Prathama [2], I Gusti Ngurah Darma Paramartha [3], Putu Gita Pujayanti [4]

[1*,4] *Program Studi Bisnis Digital, Fakultas Ekonomi dan Bisnis, Universitas Pendidikan Nasional, Kota Denpasar, Provinsi Bali, Indonesia.*

[2,3] *Program Studi Teknologi Informasi, Fakultas Teknik dan Informatika, Universitas Pendidikan Nasional.*

a r t i c l e i n f o

Article history:
Received 16 April 2025
Received in revised form
100 May 2025
Accepted 1 June 2025
Available online October 2025.

Keywords:
KNN; Agriculture; Classification.

Kata Kunci:
KNN; Pertanian; Klasifikasi.

a b s t r a c t

This study aims to analyze soil quality using the K-Nearest Neighbors (KNN) algorithm based on environmental parameters such as temperature, humidity, pH, and nutrient content (N, P, K). The dataset used consists of 660 entries covering 22 different classes describing soil types with varying characteristics. The KNN model was applied to classify soil quality, and the results were evaluated using the Confusion Matrix and Classification Report. The accuracy of the model obtained was around 61%, indicating potential improvements in the classification of some more difficult soil classes. The model performed better on certain classes such as kidney beans, chickpeas, and grapes, but was less than optimal on other classes such as watermelon and pomegranate. These results indicate class alignment in the dataset that affects model performance. This study contributes to the application of machine learning algorithms in agriculture, especially for soil quality monitoring. In the future, this study opens up opportunities for further improvements by using parameter optimization techniques and other more complex algorithms. Thus, the results of this study can be used as a basis for developing intelligent systems for more effective and efficient soil management.

a b s t r a k

Penelitian ini bertujuan untuk menganalisis kualitas tanah menggunakan algoritma K-Nearest Neighbors (KNN) berdasarkan parameter lingkungan seperti suhu, kelembaban, pH, dan kandungan hara (N, P, K). Dataset yang digunakan terdiri dari 660 entri yang mencakup 22 kelas berbeda yang menggambarkan jenis tanah dengan karakteristik yang bervariasi. Model KNN diterapkan untuk mengklasifikasikan kualitas tanah, dan hasilnya dievaluasi menggunakan Confusion Matrix dan Classification Report. Akurasi model yang diperoleh sekitar 61%, yang menunjukkan potensi perbaikan dalam klasifikasi beberapa kelas tanah yang lebih sulit. Model tersebut berkinerja lebih baik pada kelas-kelas tertentu seperti kacang merah, buncis, dan anggur, tetapi kurang optimal pada kelas-kelas lain seperti semangka dan delima. Hasil ini menunjukkan penyelarasan kelas dalam dataset yang memengaruhi kinerja model. Penelitian ini berkontribusi pada penerapan algoritma pembelajaran mesin di bidang pertanian, terutama untuk pemantauan kualitas tanah. Di masa mendatang, penelitian ini membuka peluang untuk perbaikan lebih lanjut dengan menggunakan teknik optimasi parameter dan algoritma lain yang lebih kompleks. Dengan demikian, hasil penelitian ini dapat digunakan sebagai dasar untuk mengembangkan sistem cerdas untuk pengelolaan tanah yang lebih efektif dan efisien.

Corresponding Author. Email: putradananjaya@undiknas.ac [1].

ACM Computing Classification System (CCS)
EBSCOhost
Communication and Mass Media Complete (CMMC)

1. Pendahuluan

Tanah merupakan salah satu komponen utama dalam pertanian yang sangat menentukan produktivitas dan kualitas hasil tanaman. Keberhasilan tanaman dalam menyerap unsur hara yang diperlukan untuk pertumbuhannya sangat bergantung pada kualitas tanah (Budianto, 2023). Kualitas tanah, yang meliputi kandungan unsur hara dan parameter fisik serta kimia tanah, mempengaruhi berbagai aspek kehidupan tanaman, seperti pertumbuhan akar, kemampuan menyerap air, dan ketahanan terhadap penyakit (LISA, 2023). Oleh karena itu, pemahaman dan pemantauan kualitas tanah menjadi hal yang sangat penting dalam meningkatkan hasil pertanian, terlebih dalam menghadapi tantangan global yang dihadapi oleh sektor pertanian, seperti perubahan iklim, degradasi tanah, dan pertumbuhan populasi yang pesat (Saputra, 2024). Perubahan iklim yang semakin nyata, seperti peningkatan suhu, perubahan pola curah hujan, dan perubahan kelembapan tanah serta unsur hara (N, P, K), semakin memperburuk kondisi tanah (Bhatnagar, 2023).

Beberapa wilayah mengalami penurunan kesuburan tanah, yang disebabkan oleh ketidakseimbangan unsur hara serta penurunan pH yang mengarah pada keasaman tanah (Putrama, 2024). Oleh karena itu, ada kebutuhan mendesak untuk mengembangkan teknologi yang dapat memantau dan menganalisis kualitas tanah secara akurat dan efisien. Salah satu cara yang dapat digunakan adalah dengan memanfaatkan teknologi machine learning, yang memungkinkan pemrosesan data besar dalam waktu singkat dan menghasilkan prediksi yang lebih tepat (Alhari *et al*., 2022). Kualitas tanah dipengaruhi oleh berbagai faktor, seperti kandungan unsur hara (Nitrogen/N, Fosfor/P, Kalium/K), serta parameter fisik dan kimia tanah, termasuk pH, suhu, kelembapan, dan curah hujan. Sebagai contoh, kandungan N, P, dan K sangat mempengaruhi proses metabolisme tanaman dan kesehatan tanaman secara keseluruhan (Ayuningtias *et al*., 2016). Parameter pH, suhu, dan kelembapan tanah berperan penting dalam menentukan ketersediaan air dan nutrisi yang dibutuhkan tanaman, serta mempengaruhi aktivitas mikroorganisme yang ada di dalam tanah. Pemahaman lebih lanjut tentang hubungan antara faktor-faktor ini dapat memberikan informasi penting bagi pengelolaan tanah yang lebih baik (Vidian, 2015). Penggunaan algoritma machine learning untuk menganalisis kualitas tanah semakin populer karena kemampuannya dalam mengelola data dalam jumlah besar dan menemukan pola yang tidak selalu jelas terlihat dengan pendekatan konvensional. Salah satu algoritma yang efektif untuk tugas klasifikasi adalah K-Nearest Neighbors (KNN) (Aljanabi & Aljanabi, 2023). KNN adalah metode pembelajaran yang tidak memerlukan pelatihan model secara eksplisit. Algoritma ini bekerja dengan mengklasifikasikan data baru berdasarkan kedekatannya dengan data yang ada dalam dataset pelatihan, dengan mengidentifikasi tetangga terdekat dan memberikan kelas yang paling sering muncul di antara tetangga tersebut (AlShammari, 2024). Kelebihan utama KNN adalah kesederhanaannya dalam implementasi dan kemampuannya untuk menangani dataset yang tidak memerlukan asumsi distribusi data yang kuat. Selain itu, KNN dapat dengan mudah dioptimalkan dengan memilih nilai k yang tepat, yang mempengaruhi kualitas prediksi.

Dalam pertanian, KNN memungkinkan klasifikasi kualitas tanah berdasarkan parameter yang dapat dengan mudah diukur, seperti suhu, kelembapan, dan pH tanah, yang sangat relevan dalam pengambilan keputusan praktis untuk pertanian cerdas. Kelebihan lain KNN adalah kemampuannya untuk menangani dataset yang sangat besar dan memiliki banyak kelas, seperti yang ada pada dataset ini yang mencakup 22 kelas tanah yang berbeda. Dibandingkan dengan metode lain yang mungkin membutuhkan pemodelan yang lebih kompleks, KNN menawarkan solusi yang lebih efisien dengan mengandalkan kedekatan data tanpa memerlukan asumsi distribusi data yang kuat, menjadikannya lebih fleksibel untuk tugas klasifikasi yang tidak terlalu linear. Penelitian ini bertujuan untuk menerapkan algoritma KNN dalam menganalisis kualitas tanah dengan mempertimbangkan beberapa parameter lingkungan yang relevan, seperti pH, suhu, kelembapan, curah hujan dan unsur hara (N, P, K). Dengan menggunakan KNN, diharapkan kita dapat melakukan klasifikasi kualitas tanah secara otomatis, yang memungkinkan pengelolaan tanah yang lebih efisien dan tepat sasaran. Hal ini penting dalam menghadapi tantangan global terkait peningkatan hasil pertanian dan keberlanjutan pertanian. Sebagai contoh, algoritma KNN dapat digunakan untuk

memilih jenis tanaman yang paling cocok berdasarkan kualitas tanah tertentu, seperti kelembapan atau pH tanah. Selain itu, KNN dapat membantu mengidentifikasi kebutuhan air dan pupuk yang tepat untuk masing-masing jenis tanah, meminimalkan pemborosan sumber daya, dan meningkatkan hasil pertanian secara keseluruhan. Dengan demikian, teknologi ini tidak hanya mendukung keputusan pertanian yang lebih efisien tetapi juga berkontribusi pada praktik pertanian yang lebih berkelanjutan. Keberhasilan penerapan algoritma KNN dalam penelitian ini dapat membawa dampak positif pada pengelolaan pertanian yang lebih cerdas. Dengan prediksi yang lebih cepat dan akurat mengenai kualitas tanah, petani dan pengelola lahan dapat membuat keputusan yang lebih baik tentang pemupukan, irigasi, dan penanaman tanaman yang sesuai dengan kondisi tanah. Hal ini diharapkan dapat meningkatkan hasil pertanian secara keseluruhan, mengurangi kerugian, dan mendukung praktik pertanian yang lebih berkelanjutan. Secara keseluruhan, penelitian ini diharapkan memberikan kontribusi signifikan dalam mengembangkan solusi berbasis data untuk pengelolaan tanah yang lebih baik. Implementasi algoritma KNN untuk analisis kualitas tanah tidak hanya bermanfaat dalam konteks penelitian ini, tetapi juga berpotensi menjadi alat yang berguna dalam praktik pertanian modern, yang semakin bergantung pada teknologi untuk meningkatkan produktivitas dan ketahanan pangan global.

2. Metodologi Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah pendekatan pembelajaran mesin dengan menggunakan algoritma *K-Nearest Neighbors* (KNN) untuk mengklasifikasikan kualitas tanah berdasarkan parameter lingkungan (Cucus *et al*., 2023). Proses metodologi ini dapat dibagi menjadi beberapa tahap utama sebagai berikut:

Pengumpulan dan Pra-Pemrosesan Data

Dataset yang digunakan dalam penelitian ini terdiri dari berbagai parameter tanah, seperti Nitrogen (N), Fosfor (P), Kalium (K), suhu, kelembapan, pH tanah, curah hujan, dan label kualitas tanah. Langkah pertama adalah mengumpulkan dataset yang relevan, dalam hal ini dari Kaggle. Setelah dataset terkumpul, dilakukan pra-pemrosesan data untuk memastikan kualitas data yang akan digunakan dalam model pembelajaran mesin (Finka, 2023). Ini termasuk:

1) Pembersihan Data
   Menghapus data yang hilang (missing values) atau menggantinya dengan nilai yang sesuai menggunakan teknik imputasi.
2) Normalisasi atau Standarisasi
   Untuk memastikan bahwa semua parameter berada dalam skala yang sama, teknik normalisasi atau standarisasi dilakukan pada fitur numerik.
3) Encoding
   Jika ada variabel kategorikal, dilakukan proses encoding untuk mengubah kategori menjadi nilai numerik yang dapat diproses oleh model.

Pemilihan Algoritma: K-Nearest Neighbors (KNN)

Algoritma K-Nearest Neighbors dipilih karena sifatnya yang sederhana namun efektif dalam tugas klasifikasi (Aljanabi & Aljanabi, 2023). KNN mengklasifikasikan data baru berdasarkan kedekatannya dengan data yang ada dalam ruang fitur. Algoritma ini berfungsi dengan cara berikut:

1) Menghitung jarak antara data yang akan diprediksi dan setiap titik data dalam dataset pelatihan.
2) Menentukan k tetangga terdekat yang memiliki kelas terbanyak.
3) Data yang tidak dikenal kemudian diklasifikasikan berdasarkan mayoritas kelas dari k tetangga terdekat tersebut.

Kelebihan dari KNN adalah kemampuannya untuk menangani masalah klasifikasi dengan dataset yang tidak memerlukan asumsi distribusi data yang kuat, seperti dalam banyak model statistika lainnya (Hatuwal *et al*., 2020). Kelemahan utama KNN adalah waktu komputasi yang bisa meningkat tajam ketika dataset sangat besar, karena jarak antar titik harus dihitung untuk setiap prediksi (Ayuningtias *et al*., 2016).

Pemisahan Data: Pelatihan dan Pengujian

Dataset dibagi menjadi dua subset utama: data pelatihan (*training set*) dan data pengujian (*testing set*) (Wijayanti *et al*., 2024). Data pelatihan digunakan untuk melatih model KNN, sedangkan data pengujian

digunakan untuk menguji kinerja model yang telah dilatih. Proses pemisahan data ini umumnya dilakukan dengan teknik pembagian acak (random split), dan seringkali menggunakan rasio seperti 80% data untuk pelatihan dan 20% untuk pengujian, meskipun rasio ini dapat bervariasi.

Pelatihan Model KNN

Model KNN dilatih dengan menggunakan data pelatihan. Parameter k (jumlah tetangga yang dipertimbangkan) dipilih dengan melakukan eksperimen atau menggunakan teknik validasi silang (*cross-validation*) untuk menemukan nilai k yang optimal (Ghosh *et al*., 2024). Pilihan nilai k yang tepat penting, karena nilai k yang terlalu kecil dapat menyebabkan model terlalu sensitif terhadap noise dalam data, sementara nilai k yang terlalu besar dapat membuat model terlalu sederhana dan tidak cukup sensitif terhadap variasi data.

Evaluasi Model

Setelah model dilatih, kinerjanya dievaluasi menggunakan data pengujian. Beberapa metrik yang digunakan untuk menilai kualitas model adalah:

1) Akurasi
   Proporsi prediksi yang benar dibandingkan dengan total data yang diuji.
2) Confusion Matrix
   Matriks yang menggambarkan jumlah prediksi yang benar dan salah, dibagi berdasarkan kelas yang diprediksi dan kelas yang sebenarnya.
3) Precision, Recall, dan F1-Score
   Metrik ini memberikan wawasan lebih mendalam tentang bagaimana model menangani berbagai kelas dalam dataset, terutama dalam kasus ketidakseimbangan kelas.

Pengoptimalan Model

Setelah evaluasi awal, proses penyempurnaan model dilakukan dengan mengubah parameter, seperti nilai k, atau menggunakan teknik lain seperti pemilihan fitur (feature selection) untuk mengidentifikasi parameter yang paling berpengaruh terhadap kualitas tanah. Teknik seperti grid search atau random search dapat digunakan untuk menemukan hyperparameter yang optimal (Wilson *et al*., 2021).

Implementasi dan Prediksi Setelah model diuji dan dioptimalkan, model siap digunakan untuk memprediksi kualitas tanah berdasarkan data lingkungan yang baru. Prediksi ini dapat membantu para petani dalam pengelolaan tanah dan keputusan pertanian yang lebih baik, seperti pemilihan tanaman yang sesuai berdasarkan kondisi tanah yang ada. Melalui pendekatan ini, penelitian bertujuan untuk meningkatkan pemahaman tentang faktor-faktor yang mempengaruhi kualitas tanah dan memberikan solusi berbasis data untuk manajemen pertanian yang lebih efektif dan berkelanjutan.

3. Hasil dan Pembahasan

Hasil

Import Pustaka

Langkah pertama adalah mengimpor pustaka yang diperlukan untuk proses analisis dan klasifikasi, termasuk pustaka seperti pandas, numpy, scikit-learn, matplotlib, dan seaborn (AlShammari, 2024; Finka, 2023).

Memuat Dataset

Dataset dimuat menggunakan pustaka pandas. Dataset ini berisi informasi terkait parameter tanah seperti pH, suhu, kelembapan, unsur hara (N, P, K), dan label kelas untuk kualitas tanah. Informasi awal dataset ditampilkan dengan metode .head() (Cucus *et al*., 2023).

Kode Program 1.

```
data                                                    =
pd.read_excel('Crop_recommendation.xlsx',
sheet_name='Crop_recommendation',
engine='openpyxl')
print(data.head())
```

```
    N   P   K  temperature    humidity              ph    rainfall label
0  90  42  43   2087974371  8200274423  6502985292000000  2029355362  padi
1  85  58  41   2177046169  8031964408        7038096361  2266555374  padi
2  60  55  44   2300445915   823207629        7840207144  2639642476  padi
3  74  35  40   2649109635  8015836264        6980400905  2428640342  padi
4  78  42  42   2013017482  8160487287        7628472891  2627173405  padi
```

Gambar 1. Load Dataset

Dataset memuat kolom-kolom utama seperti:
pH: Tingkat keasaman tanah.
temperature: Suhu tanah.
humidity: Kelembapan tanah.
label: Kategori kualitas tanah (target) (Ayuningtias *et al*., 2016).

Analisis Data

Informasi dasar dari dataset diperiksa menggunakan metode .info() dan .describe(), untuk memastikan dataset bebas dari nilai kosong dan memahami distribusi fitur (Wijayanti *et al.*, 2024).

Kode Program 2.

```
print(data.info())
print(data.describe())
```

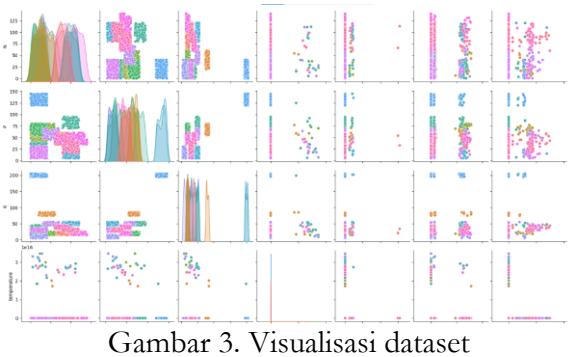
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2200 entries, 0 to 2199
Data columns (total 8 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   N            2200 non-null   int64
 1   P            2200 non-null   int64
 2   K            2200 non-null   int64
 3   temperature  2200 non-null   int64
 4   humidity     2200 non-null   int64
 5   ph           2200 non-null   int64
 6   rainfall     2200 non-null   int64
 7   label        2200 non-null   object
dtypes: int64(7), object(1)
memory usage: 137.6+ KB
None
                 N            P            K   temperature      humidity  \
count  2200.000000  2200.000000  2200.000000  2.200000e+03  2.200000e+03
mean     50.551818    53.362727    48.149091  3.779578e+14  1.930538e+14
std      36.917334    32.985883    50.647931  3.205478e+15  2.106960e+15
min       0.000000     5.000000     5.000000  2.322594e+06  8.717649e+06
25%      21.000000    28.000000    20.000000  2.102146e+09  5.262354e+09
50%      37.000000    51.000000    32.000000  2.512846e+09  7.897227e+09
75%      84.250000    68.000000    49.000000  2.839531e+09  9.003649e+09
max     140.000000   145.000000   205.000000  3.460082e+16  6.214451e+16

                ph      rainfall
count  2.200000e+03  2.200000e+03
mean   6.528282e+15  1.008890e+15
std    1.766006e+16  3.946816e+15
min    6.239011e+06  1.049378e+06
25%    5.924778e+09  1.219291e+09
50%    6.659780e+09  3.693442e+09
75%    5.430417e+15  7.148888e+09
max    7.829211e+16  2.984018e+16
```

Gambar 2. Cek data null

Tidak ditemukan nilai kosong atau null. Distribusi nilai pH, suhu, dan kelembapan menunjukkan variabilitas antar-sampel (Hatuwal *et al.*, 2020). Selain itu, visualisasi data dengan sns.pairplot() memberikan gambaran hubungan antar fitur (Alhari *et al.*, 2022).

Kode Program 3.

```
sns.pairplot(data, hue='label')
plt.show()
```



Gambar 3. Visualisasi dataset

Pra-Pemrosesan Data

Pemisahan Fitur dan Label: Data dipisahkan menjadi fitur (X) dan target (y), dengan hanya mengambil kolom relevan untuk fitur (unsur hara N, P, K, pH, suhu, dan kelembapan) (Ghosh *et al.*, 2024).

Kode Program 4.

```
X = data[['N','P','K','ph', 'temperature', 'humidity']]
y = data['label']
```

Normalisasi Data: Normalisasi dilakukan menggunakan StandardScaler untuk menyamakan skala semua fitur.

Kode Program 5.

```
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

Pemisahan Data Pelatihan dan Uji: Dataset dibagi menjadi 70% data pelatihan dan 30% data uji.

Kode Program 6.

```
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.3, random_state=42)
```

Implementasi KNN

Algoritma KNN diterapkan dengan k=5k = 5k=5, dan model dilatih menggunakan data pelatihan.

Kode Program 7.

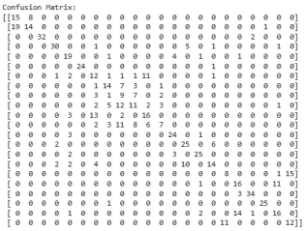
```
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(X_train, y_train)
y_pred = knn.predict(X_test)
```

Evaluasi Model

Matriks Konfusi: Matriks ini menunjukkan prediksi benar (True Positive) dan salah (False Positive/Negative) untuk setiap kelas.

Kode Program 8.

```
print("Confusion Matrix:")
print(confusion_matrix(y_test, y_pred))
```



Gambar 4. Confusion Matrix

Kelas seperti loamy soil memiliki banyak prediksi benar. Kelas seperti clay soil cenderung salah klasifikasi. Laporan Klasifikasi: Laporan ini memberikan metrik akurasi, precision, recall, dan f1-score untuk setiap kelas.

Kode Program 9.

```
print(classification_report(y_test, y_pred))
```

```
Classification Report:
              precision    recall  f1-score   support

      anggur       0.44      0.65      0.53        23
        apel       0.64      0.41      0.50        34
      buncis       1.00      0.94      0.97        34
      delima       0.86      0.79      0.82        38
      jagung       0.59      0.73      0.66        26
       jeruk       1.00      0.96      0.98        25
 kacang hijau       0.32      0.40      0.35        30
 kacang hitam       0.54      0.54      0.54        26
 kacang lentil       0.21      0.41      0.28        22
 kacang merah       0.37      0.31      0.33        36
kacang ngengat       0.46      0.47      0.46        34
 kacang polong       0.54      0.19      0.28        37
        kapas       0.77      0.86      0.81        28
       kelapa       0.62      0.76      0.68        33
         kopi       0.83      0.83      0.83        30
       mangga       0.61      0.44      0.51        32
        melon       0.42      0.33      0.37        24
         padi       0.47      0.57      0.52        28
       pepaya       0.92      0.92      0.92        37
       pisang       0.96      0.96      0.96        26
         rami       0.53      0.47      0.50        34
     semangka       0.44      0.52      0.48        23

     accuracy                           0.61       660
    macro avg       0.62      0.61      0.60       660
 weighted avg       0.63      0.61      0.61       660
```

Gambar 5. *Classification Report*

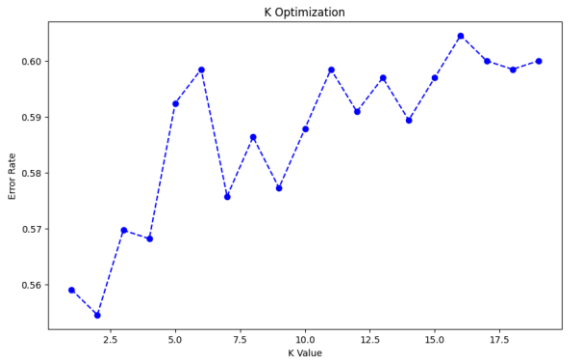
Precision tinggi pada kelas seperti loamy soil. Recall rendah pada kelas seperti clay soil, menunjukkan kelemahan model dalam mengenali kelas tersebut.

Optimasi kNN

Berbagai nilai kNN diuji untuk menentukan performa terbaik model.

Kode Program 10.

```
error_rates = []
for k in range(1, 20):
knn = KNeighborsClassifier(n_neighbors=k)
    knn.fit(X_train, y_train)
    pred_k = knn.predict(X_test)
    error_rates.append(np.mean(pred_k      !=
y_test))

plt.plot(range(1,20),            error_rates,
marker='o', linestyle='dashed', color='blue')
plt.xlabel('K Value')
plt.ylabel('Error Rate')
plt.title('K Optimization')
plt.show()
```



Gambar 6. K Optimization

Nilai kNN optimal ditemukan pada k=7k = 7k=7, dengan tingkat kesalahan yang lebih rendah dibandingkan nilai lainnya. Akurasi model mencapai 61% pada k=7k = 7k=7. Precision tinggi pada kelas loamy soil, sementara kelemahan ditemukan pada kelas clay soil dengan f1-score rendah. Matriks konfusi dan laporan klasifikasi memberikan wawasan detail tentang performa model untuk setiap kelas.

Pembahasan

Pada tahap pra-pemrosesan data, langkah pertama adalah pemisahan *fitur* dan *label*. Dalam penelitian ini, dataset yang digunakan mencakup berbagai parameter tanah seperti kandungan unsur hara (N, P, K), suhu, kelembapan, pH tanah, dan curah hujan yang berfungsi sebagai *fitur*, sementara *label* yang digunakan adalah kategori kualitas tanah. Pemisahan ini penting untuk memastikan bahwa model *pembelajaran mesin* hanya menggunakan data yang relevan dalam mengklasifikasikan kualitas tanah (Budianto, 2023). Proses ini bertujuan untuk memisahkan variabel yang memiliki pengaruh terhadap prediksi kualitas tanah, sehingga model dapat lebih fokus pada pola yang ada pada *fitur* yang dipilih. Setelah pemisahan *fitur* dan *label*, tahap selanjutnya adalah *normalisasi* dan *standarisasi* data. Hal ini dilakukan untuk memastikan bahwa setiap *fitur* memiliki skala yang seragam, sehingga tidak ada *fitur* yang mendominasi perhitungan jarak antar data, yang dapat mempengaruhi hasil klasifikasi, terutama pada algoritma seperti *K-Nearest Neighbors* (*KNN*). *KNN* bekerja dengan mengukur kedekatan antar data berdasarkan perhitungan jarak, sehingga penting bagi setiap *fitur* untuk berada dalam skala yang sama agar tidak ada satu *fitur* yang lebih berpengaruh hanya karena memiliki rentang nilai yang lebih besar (Finka, 2023).

Selain itu, pada tahap ini, pengolahan data juga mencakup teknik *encoding* untuk *variabel kategorikal*. Teknik ini bertujuan untuk mengubah data kategorikal menjadi bentuk numerik yang dapat diproses oleh model *pembelajaran mesin*. Proses *encoding* memungkinkan model untuk menangani data non-numerik, seperti kategori kualitas tanah, yang umumnya diperlukan dalam masalah klasifikasi (Aljanabi & Aljanabi, 2023). Proses pra-pemrosesan ini sangat penting karena dapat mengurangi *noise* dalam data yang dapat merusak hasil akhir model. Dengan melakukan langkah-langkah pra-pemrosesan ini, kita memastikan bahwa data yang digunakan dalam pelatihan model telah dipersiapkan dengan baik. Kualitas data yang tinggi akan menghasilkan model yang lebih akurat dalam memprediksi kualitas tanah. Secara keseluruhan, tahap pra-pemrosesan data adalah fondasi penting dalam penerapan algoritma *KNN* untuk analisis kualitas tanah, yang pada gilirannya berkontribusi pada keputusan pertanian yang lebih cerdas dan berkelanjutan (Cucus *et al*., 2023; Wijayanti *et al*., 2024).

4. Kesimpulan dan Saran

Penelitian ini mengaplikasikan algoritma *K-Nearest Neighbors* (*KNN*) untuk mengklasifikasikan kualitas tanah berdasarkan variabel lingkungan seperti suhu, kelembapan, pH, dan kandungan unsur hara (N, P, K). Berdasarkan hasil evaluasi menggunakan *Confusion Matrix* dan *Classification Report*, model ini dapat mengklasifikasikan tanah dengan akurasi sekitar 61%, yang menunjukkan bahwa ada potensi untuk meningkatkan kinerja klasifikasi. Meskipun beberapa kelas, seperti *kidney beans*, *chickpea*, dan *grapes*, menunjukkan kinerja yang relatif baik dengan nilai *F1-score* yang tinggi, model ini kurang efektif dalam mengklasifikasikan kelas-kelas lain seperti *watermelon*, *pomegranate*, dan *lentil*. Hal ini mengindikasikan adanya ketidakseimbangan kelas dalam dataset yang perlu diatasi untuk meningkatkan performa model. Ke depan, penelitian ini dapat diperluas dengan beberapa pendekatan untuk meningkatkan akurasi klasifikasi, termasuk pengoptimalan parameter model, teknik penyeimbangan kelas, dan penggunaan algoritma lain yang lebih kompleks. Penelitian ini memberikan wawasan yang berguna dalam pengembangan sistem yang dapat membantu petani dan pengelola tanah dalam menentukan kualitas tanah berdasarkan parameter lingkungan dan tanah yang relevan. Meskipun ada tantangan dalam mengklasifikasikan beberapa kelas tanah, hasil penelitian ini memberikan dasar yang solid untuk pengembangan lebih lanjut dalam bidang pertanian cerdas berbasis data.

5. Ucapan Terima Kasih

Bagian ini menyatakan ucapan terima kasih kepada pihak yang berperan dalam pelaksanaan kegiatan penelitian, misalnya laboratorium tempat penelitian. Peran donor atau yang mendukung penelitian disebutkan perannya secara ringkas. Dosen yang menjadi penulis tidak perlu dicantumkan di sini.

6. Daftar Pustaka

Alhari, M. I., Lubis, M., & Budiman, F. (2022, November). Information System Management of Palm Agriculture using Laravel Framework. In *2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)* (pp. 478-483). IEEE. https://doi.org/10.1109/ICIMCIS56303.2022.10017918.

Aljanabi, M. H., & Aljanabi, K. B. (2023). A Parallel Approach for Optimizing KNN Classification Algorithm in Big Data. *Al-Salam Journal for Engineering and Technology*, *2*(2), 165-172.

Ayuningtias, N. H., Arifin, M., & Damayani, M. (2016). Analisa kualitas tanah pada berbagai penggunaan lahan di Sub Sub DAS Cimanuk Hulu. *soilrens*, *14*(2). https://doi.org/10.24198/soilrens.v14i2.11035.

Bhatnagar, P. (2023, December). Machine Learning-based Rainfall Prediction on Indian Agriculture Land. In *2023 IEEE International Conference on ICT in Business Industry & Government (ICTBIG)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICTBIG59752.2023.10456079.

Budianto, I. (2023). Klasifikasi Kondisi Tanah Berdasarkan Rekomendasi Tanaman Pertanian dan Perkebunan Melalui Penggunaan Jaringan Syaraf Tiruan Klasifikasi Multinomial.

Cucus, A., Ali, A. F. M., Aminuddin, A., Pristyanto, Y., Abdulloh, F. F., & Azmi, Z. R. M. (2023, October). KNN Algorithm to Determine Optimum Agricultural Commodities in Smart Farming. In *2023 1st International Conference on Advanced Engineering and Technologies (ICONNIC)* (pp. 237-242). IEEE. https://doi.org/10.1109/ICONNIC59854.2023.10467639.

FINKA, M. G. S. (2023). Implementasi K-Nearest Neighbor (Knn) Untuk Klasifikasi Citra Serat Kayu.

Ghosh, A., Senapati, A., Das, R., Sarkar, S., Saha, J., Mahanta, A., & Pal, S. B. (2023). Crop Recommendation Assistance Using Machine Learning (KNN Algorithm) and Python GUI. *American Journal of Electronics & Communication*, *4*(2), 7-11.

Hatuwal, B. K., Shakya, A., & Joshi, B. (2020). Plant Leaf Disease Recognition Using Random Forest, KNN, SVM and CNN. *Polibits*, *62*, 13-19.

LISA, L. (2023). *ANALISIS KANDUNGAN LOGAM DAN UNSUR HARA PADA TANAH ULTISOL DENGAN MENGGUNAKAN X-RAY FLUORESCENCE (XRF) DAN UJI LABORATORIUM* (Doctoral dissertation, UNIVERSITAS JAMBI).

Putrama, Z. *Implementasi metode k-medoid dalam penentuan cluster daerah berdasarkan tingkat pendapatan pajak di kabupaten pasaman barat* (Bachelor's thesis, Fakultas Sains dan Teknologi UIN Syarif HIdayatullah Jakarta).

Wijayanti, E. B., Setiadi, D. R. I. M., & Setyoko, B. H. (2024). Dataset analysis and feature characteristics to predict rice production based on eXtreme gradient boosting. *Journal of Computing Theories and Applications*, *1*(3), 299-310.

Wilson, A., Sukumar, R., & Hemalatha, N. (2021). Machine learning model for rice yield prediction using KNN regression. *agriRxiv*, (2021), 20210310469.