

**Klasifikasi Kualitas Tanah untuk Rekomendasi Jenis Tanaman dengan
Metode *Neural Network***



IPB University
— Bogor Indonesia —

Sindi Aprilianti	(G6401221037)
Muh Farid FB	(G6401221060)
Zhafran Agus	(G6401221073)
Aszriel Teddy Muhammad	(G6401221110)

DEPARTEMEN ILMU KOMPUTER
SEKOLAH SAINS DATA, MATEMATIKA, DAN INFORMATIKA
INSTITUT PERTANIAN BOGOR
BOGOR
2024

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam sektor pertanian, kualitas tanah merupakan salah satu penentu produktivitas tanaman. Jika tanah berada dalam kondisi yang baik dan tanaman yang digunakan sudah sesuai dengan jenis tanahnya, petani dapat mencapai hasil panen yang optimal (Astuti *et al.*, 2023). Kualitas tanah merujuk pada kemampuan tanah dalam mempertahankan produktivitas tanaman, menjaga kesediaan air, dan mendukung kegiatan manusia. Penilaian kualitas tanah dilakukan dengan menggunakan indikator-indikator yang menggambarkan proses-proses penting dalam tanah, yang dapat dilihat dari sifat fisik, kimia, dan biologis tanah (Kurniawan, *et al.*, 2021). Kualitas tanah tidak hanya mencakup kemampuan tanah dalam mendukung produktivitas tanaman menjadi optimal, tetapi juga memastikan keberlanjutan sistem pertanian serta menjaga kesehatan lingkungan dengan mencegah adanya pencemaran atau kontaminasi bahan berbahaya (Minarsih dan Hanudin, 2020).

Saat ini, masih banyak petani yang belum mempertimbangkan hubungan antara kondisi tanah dengan jenis tanaman yang dipilih, sehingga jika ternyata kualitas tanah tidak sesuai dengan jenis tanaman, hasil panen dapat mengalami kerugian yang cukup besar. Faktor kondisi tanah di lahan ini sangat memengaruhi hasil tanam (Khomsatun *et al.* 2020). Untuk mengatasi permasalahan tersebut, dibangunlah sebuah model neural network yang digunakan untuk menentukan klasifikasi kualitas tanah yang sesuai dengan jenis tanaman tertentu.

Neural network digunakan dalam berbagai bidang, termasuk pertanian, untuk melakukan prediksi dan juga klasifikasi berdasarkan data yang kompleks. Model yang dibangun menggunakan *Neural Network* akan bekerja dengan mempelajari pola dalam data yang memungkinkan sistem untuk mengidentifikasi hubungan antara kualitas tanah dengan jenis tanaman, menggunakan parameter seperti pH, kelembapan, dan kandungan nutrisi. Model ini dapat membantu menentukan jenis tanaman yang paling sesuai dengan kondisi tanah, sehingga akan mendapatkan hasil panen yang optimal.

1.2 Rumusan Masalah

1. Bagaimana membangun model klasifikasi untuk menentukan jenis tanaman yang sesuai berdasarkan kualitas tanah, dengan mempertimbangkan pH, kelembapan, dan kandungan nutrisi?
2. Bagaimana tingkat akurasi model *Neural Network* yang dihasilkan dalam memprediksi jenis tanaman berdasarkan dataset?

1.3 Tujuan

1. Membangun model klasifikasi yang dapat memprediksi jenis tanaman yang paling sesuai berdasarkan parameter kualitas tanah, seperti pH, kelembapan, dan kandungan nutrisi.
2. Mengukur akurasi model *Neural Network* yang telah dibangun dalam memprediksi jenis tanaman berdasarkan dataset yang digunakan.

1.4 Manfaat

Dengan membangun model klasifikasi tanah untuk jenis tanaman menggunakan *Neural Network*, diharapkan petani dapat menanam tanaman yang sesuai berdasarkan kondisi tanah dan meningkatkan produktivitas. Selain itu, model ini meningkatkan efisiensi pemilihan tanaman dengan memberi rekomendasi yang cepat dan akurat, mengurangi kebutuhan untuk percobaan yang dapat memakan banyak waktu. Model ini juga mengurangi risiko kerugian akibat ketidaksesuaian tanaman dengan tanah yang sering menyebabkan hasil panen yang buruk dan merugikan para petani.

BAB II

TINJAUAN PUSTAKA

Setiap jenis tanah dapat memberikan banyak fungsi, namun diketahui bahwa beberapa tanah justru lebih efektif untuk menjalankan fungsi tertentu dibandingkan yang lainnya, tergantung pada sifat intrinsik tanah tersebut (Vrebos *et al.* 2020).

2.1 Klasifikasi

Klasifikasi merupakan salah satu bidang ilmu pada *data mining* dimana proses klasifikasi dilakukan dengan melakukan *training* terlebih dahulu pada sebagai data training dilanjutkan dengan proses testing untuk data yang baru (Pradana *et al.*, 2022).

Klasifikasi merupakan bentuk dasar dari analisis data yang mempunyai teknik untuk menentukan keanggotaan kelompok berdasarkan data-data yang sudah ada. Konsep dasar dari klasifikasi adalah beberapa data yang memiliki struktur data yang mirip akan memiliki klasifikasi yang mirip pula (Apriliyani *et al.*, 2022).

2.2 Klasifikasi Tanah

Klasifikasi tanah yaitu ilmu yang mempelajari cara-cara membedakan sifat-sifat tanah satu sama lain, dan mengelompokkan tanah-tanah yang memiliki kesamaan sifat ke dalam kelas-kelas tertentu (Ara *et al.*, 2022). Klasifikasi tanah merupakan cara untuk membedakan sifat tanah dan mengelompokkannya kedalam beberapa kelas sesuai dengan kesamaan sifat yang dimiliki. Hal ini yang menjadi dasar dan acuan agar komoditas hasil pertanian dapat dikelola secara efektif dan mendapatkan hasil maksimal (Parti *et al.*, 2023).

Klasifikasi tanah merupakan pengelompokan tanah untuk mengorganisasi sifat-sifat dan produktivitas tanah sehingga pemanfaatan lahan dapat didasarkan pada jenis-jenis tanah untuk menjaga kelestarian dan meningkatkan produktivitas lahan tersebut (Siwi *et al.*, 2023). Klasifikasi tanah berguna sebagai alat komunikasi antar para pakar dan praktisi tanah di seluruh dunia, evaluasi atau penilaian lahan, berbagi teknik dan metode pengelolaan tanah antar wilayah, alat pemersatu dan ciri khas budaya bangsa, serta menjadi nilai tingkat kemajuan dan penguasaan ilmu pengetahuan dan teknologi tanah di suatu negara (Sembodo *et al.*, 2023).

2.3 Kualitas Tanah

Kualitas tanah adalah kemampuan tanah untuk menampilkan fungsi-fungsinya dalam penggunaan lahan atau ekosistem untuk menopang produktivitas biologi, mempertahankan

kualitas lingkungan, dan meningkatkan kesehatan tanaman, binatang, dan manusia (Soil Science Society of America, 1997). Sedangkan menurut Doran dan Parkin (1993) pengertian kualitas tanah harus mencakup: (1) produktivitas, kemampuan tanah untuk menghasilkan tanaman, (2) kualitas lingkungan, kemampuan tanah untuk menetralkan kontaminan-kontaminan lingkungan, patogen dan aspek-aspek merusak lainnya, (3) kesehatan, kemampuan tanah mempertahankan kesehatan tanaman, hewan dan manusia. Jadi, Tanah yang berkualitas, tidak saja tanah tersebut subur dan produktif akan tetapi harus mencakup aspek lingkungan dan kesehatan (Minarsih *et al.*, 2020).

BAB III

METODOLOGI

Bagian metode penelitian menjelaskan metode yang diterapkan dalam penelitian ini, termasuk objek penelitian dan tahapan-tahapan yang dilakukan untuk mencapai tujuan. Selain itu, bagian ini juga menguraikan tentang *Neural Network* yang telah dikembangkan, beserta gambaran umum sistem yang dihasilkan.

3.1 Jenis Penelitian

Penelitian ini menggunakan metode kuantitatif. Secara harfiah, pendekatan kuantitatif berarti segala sesuatu dapat diamati dan diukur. Penelitian kuantitatif bersifat terukur, menghasilkan data berupa angka, dan dianalisis menggunakan metode statistik deskriptif maupun inferensial (Surya, 2019). Dalam penelitian ini, angka-angka tersebut dijelaskan melalui proses pengumpulan data, interpretasi, hingga pencapaian hasil yang diharapkan. Penelitian dilakukan dengan melakukan peninjauan lebih mendalam terhadap objek penelitian, diikuti oleh penarikan kesimpulan berdasarkan hasil penelitian yang mencakup pengembangan model serta sistem yang dirancang.

3.2 Objek Penelitian

Dataset klasifikasi kondisi tanah berdasarkan tanaman yang paling cocok diperoleh dari **Crop Recommendation Dataset**, yang diunggah oleh Siddharth Sharma di platform Kaggle. Dataset ini merupakan hasil augmentasi berdasarkan parameter-parameter tanah secara fisik maupun kimiawi dan terdiri atas 8 kolom dengan total 2200 data. Jenis tanaman yang terdapat dalam dataset meliputi apel, pisang, kedelai hitam, kacang chickpea, kelapa, kopi, kapas, anggur, rami, kacang merah, kacang lentil, jagung, mangga, kacang ngengat, kacang hijau, melon, jeruk, pepaya, kacang gude, delima, padi, serta semangka. Informasi lebih rinci mengenai fitur dan label pada dataset, yang mencakup kondisi tanah secara fisik dan kimiawi serta nama tanaman, dapat dilihat pada **Tabel 3.1**.

Tabel 3.1 Penjelasan Kolom Objek Penelitian

No.	Nama Atribut	Keterangan	Tipe Data
1	N	Rasio kandungan Nitrogen dalam tanah yang diteliti	Int64
2	P	Rasio kandungan Fosfor dalam tanah yang diteliti	Int64

3	K	Rasio kandungan Kalium dalam tanah yang diteliti	Int64
4	tempretaure	Temperatur tanah dalam satuan Celcius	Float64
5	humidity	Kelembapan relatif tanah dalam satuan persen	Float64
6	pH	Nilai pH tanah	Float64
7	rainfall	Curah hujan tanah dalam satuan milimeter	Float64
8	label	Nama tanaman yang direkomendasikan untuk tanah tersebut	String

3.3 Tahap Penelitian

Untuk membangun model klasifikasi kualitas tanah menggunakan Neural Network, awalnya dilakukan ekstraksi data untuk memudahkan proses pengembangan. Berikutnya masuk ke tahap *preprocessing* data yang dilakukan dengan beberapa proses normalisasi agar mendapatkan hasil yang lebih akurat. Beberapa langkah dalam *processing* meliputi pengecekan data untuk mengetahui kondisi dataset yang abnormal, hal ini dilakukan dengan perulangan sampai seluruh data telah diobservasi dan akan digunakan dalam pembangunan model. Tahap berikutnya adalah menggunakan label encoding untuk mengubah data yang memiliki tipe kategorik menjadi numerik. Datasest akan dibagi dengan proporsi data latih sebesar 80% dan data uji 20%. Model berikutnya dikembangkan dengan menggunakan TensorFlow dan dievaluasi menggunakan metrik accuracy.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Data Understanding

Data yang digunakan dalam analisis ini mencakup berbagai variabel penting yang memengaruhi kualitas tanah dan pemilihan jenis tanaman yang sesuai. Data ini diperoleh dari dataset terkait tanah dan tanaman, yang terdiri dari informasi mengenai kandungan nutrisi tanah, kelembapan, suhu, curah hujan, pH, dan jenis tanaman yang cocok. Dalam analisis ini, variabel tambahan seperti kebutuhan air, penggunaan air, dan kualitas tanah dapat ditambahkan untuk mendukung klasifikasi yang lebih mendalam. Berikut adalah deskripsi variabel dalam dataset:

1. **N (Nitrogen Content)**

Nitrogen adalah nutrisi penting dalam tanah yang membantu tanaman tumbuh sehat, terutama untuk menghasilkan daun hijau dan proses fotosintesis.

2. **P (Phosphorous Content)**

Fosfor membantu tanaman memiliki akar yang kuat, berbunga dengan baik, dan menghasilkan panen yang lebih bagus. Fosfor juga membantu tanaman mendapatkan energi

3. **K (Potassium Content)**

Kalium membantu tanaman melawan penyakit, membuka dan menutup pori-pori daun (stomata), serta membuat hasil panen lebih enak dan berkualitas.

4. **Temperature (Suhu)**

Suhu udara di sekitar tanaman, diukur dalam derajat Celsius (°C). Suhu yang terlalu panas atau terlalu dingin bisa mengganggu pertumbuhan tanaman.

5. **Humidity (Kelembapan)**

Tingkat kelembapan udara di sekitar tanaman, diukur dalam persentase (%). Udara lembap membantu tanaman menyimpan air, sedangkan udara kering bisa membuat tanaman kekurangan air.

6. **pH (Tingkat Keasaman Tanah)**

Keasaman tanah, diukur dengan pH. Sebagian besar tanaman tumbuh dengan baik di tanah yang pH-nya antara 6–7. Tanah yang terlalu asam atau basa bisa mengganggu tanaman.

7. Rainfall (Curah Hujan)

Jumlah air hujan yang turun dalam periode tertentu, diukur dalam milimeter (mm). Air hujan adalah sumber utama air untuk tanaman.

4.2 Pemilihan Model

4.2.1 Persiapan Data

Pada tahap awal, persiapan data dilakukan untuk memastikan dataset sudah sesuai dengan kebutuhan pembuatan model. Parameter input dalam dataset yang dipilih untuk proses pembuatan dan pelatihan model meliputi ‘**humidity**’, ‘**pH**’, ‘**P**’, ‘**K**’, dan ‘**N**’. Sedangkan parameter target, yaitu kolom ‘**label**’ yang berisi informasi terkait jenis tanaman apa yang cocok ditanam pada kondisi tanah tersebut. Dataset kemudian dibagi menjadi data latih sebesar 80% dan data uji sebesar 20% menggunakan fungsi **train_test_split**. Berikutnya, karena data target merupakan data kategorik, maka akan diubah menjadi numerik menggunakan **label encoding**.

4.2.2 Pembuatan Model dan Data Training

Pembuatan model dan *training* data dilakukan dengan menggunakan dataset yang disiapkan pada proses sebelumnya. Pembuatan model *Neural Network* menggunakan Keras dari TensorFlow. Keras ini merupakan library deep learning yang memungkinkan pembuatan dan juga pelatihan model neural network dengan mudah dan cepat.

Model dibangun menggunakan objek *sequential*, yang berarti *layers* akan disusun secara berurutan. Lapisan pertama merupakan Dense(128), artinya lapisan ini memiliki 128 neuron dan menerima data input dengan dimensi yang sesuai dengan jumlah fitur dataset. Fungsi aktivasi ReLU digunakan untuk memberikan output non-linear. Berikutnya, Dropout(0.5) digunakan untuk mencegah terjadinya overfitting dan secara acak menonaktifkan 50% neuron di lapisan ini pada setiap iterasi training. Kemudian, hidden layer kedua ditambahkan dengan Dense(64) yang memiliki 64 neuron dan fungsi aktivasi ReLU.

Pada lapisan terakhir, Dense (len(y.unique()), activation='softmax'), digunakan untuk dapat menghasilkan prediksi kelas sebanyak jumlah kelas yang berada pada kolom target, yaitu “label”. Fungsi aktivasi **softmax** digunakan untuk masalah klasifikasi multiclass. Setelah model dibangun, berikutnya dipilih adam optimizer yang

dapat menyesuaikan laju learning secara otomatis, dan **sparse_categorical_crossentropy** yang digunakan sebagai loss function. Model kemudian akan melacak akurasi menggunakan metrik 'accuracy'.

Berikutnya model akan di-*training* menggunakan data latih **x_train** dan target **y_train** menggunakan metode **model.fit()**. Selama training, data akan dibagi menjadi training data dan validation data. **Validation_split=0.2** menunjukkan bahwa 20 dari data latih akan digunakan untuk validasi agar menghindari overfitting. *Training* dilakukan sebanyak 500 epoch dan **batch_size=32**, yang digunakan untuk menentukan jumlah data yang digunakan untuk pembaruan bobot pada setiap iterasinya. **EarlyStopping** digunakan untuk menghentikan pelatihan lebih awal jika *loss function* pada data validasi tidak menunjukkan adanya perbaikan setelah sejumlah epoch. Setelah selesai di *training*, evaluasi akan dilakukan menggunakan data uji **x_test** dan target **y_test** menggunakan **model.evaluate()**. Model ini akan mengukur akurasi dari hasil prediksi pada data uji dan menghasilkan nilai akurasi untuk mengevaluasi performa model.

4.3 Simulasi

Simulasi dilakukan menggunakan terminal, di mana pengguna memasukkan parameter secara manual, seperti pH, kadar nitrogen (N), fosfor (P), kalium (K), dan presentase kelembapan. Data yang dimasukkan oleh pengguna ini diolah oleh program Python yang memuat model machine learning (**my_model.h5**). Model tersebut digunakan untuk memprediksi jenis tanaman yang direkomendasikan berdasarkan parameter yang diberikan. Hasil prediksi langsung ditampilkan di terminal, dan pengguna dapat mengulangi simulasi untuk sampel tanah lainnya sesuai kebutuhan.

```
=== Soil to Plant Type Classification ===
Enter soil parameters for classification:
  Soil pH (e.g., 6.5): 6.5
  Nitrogen (N) content (e.g., 50): 50
  Phosphorus (P) content (e.g., 40): 40
  Potassium (K) content (e.g., 60): 60
  Moisture percentage (e.g., 20): 20
1/1 ————— 0s 63ms/step

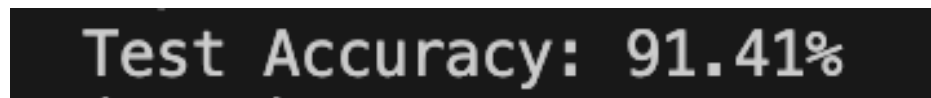
Recommended Plant Type: Rice
```

Gambar 4.1 Hasil simulasi program

Gambar 4.1 Menunjukkan hasil simulasi program dengan *model machine learning my_model.h5* untuk memprediksi jenis tanaman berdasarkan parameter tanah yang dimasukkan melalui terminal.

4.4 Evaluasi

Neural Network Model dievaluasi menggunakan data uji setelah *preprocessing* (penghapusan duplikasi, pemilihan fitur, *encoding* label, *sampling*, dan normalisasi). Model Sequential dengan lapisan Dense (128, ReLU), Dropout (0.5), Dense (64, ReLU), Dropout (0.3), dan output Softmax dikompilasi dengan Adam, *sparse_categorical_crossentropy*, dan metrik akurasi. *Early stopping* dengan *patience* 10 digunakan. Pelatihan dengan *batch size* 32 dan maksimal 500 epoch (validasi 20%) menghasilkan akurasi pada data uji sebesar 91.41% . Meskipun akurasi cukup tinggi, evaluasi lanjutan dengan metrik lain misal nya *recall*, F1-Score, *Confusion Matrix* yang lebih disarankan untuk pemahaman performa yang lebih mendalam.



Test Accuracy: 91.41%

Gambar 4.1 Hasil *Test Accuracy*

BAB V

KESIMPULAN

Kualitas tanah memiliki peran yang sangat penting dalam hal menentukan produktivitas tanaman. Parameter seperti pH, kelembapan, dan kandungan nutrisi sangat memengaruhi jenis tanaman yang dapat tumbuh dengan optimal. Dengan membangun model sistem klasifikasi kualitas tanah menggunakan Neural Network, sistem ini dapat menentukan jenis tanaman yang sesuai dengan kualitas tanah berdasarkan parameter tersebut. Hasil dari evaluasi menunjukkan model memiliki akurasi yang baik. Penerapan model ini diharapkan dapat membantu petani dalam menentukan jenis tanaman yang sesuai kondisi tanah, sehingga dapat meningkatkan hasil panen dan juga mendukung keberlanjutan pertanian yang lebih efisien dan ramah lingkungan karena dapat mengurangi penggunaan pupuk.

DAFTAR PUSTAKA

- Astuti YP, Wardatunizza I, Subhiyakto ER, Kartikadarma E. 2023. Implementasi algoritma Convolutional Neural Network (CNN) untuk klasifikasi jenis tanah berbasis Android. *Jurnal Pengembangan IT (JPIT)* [Internet]. [diunduh 2024 Des 15]; 8(3): 220-225. <https://doi.org/10.30591/jpit.v8i3.5026>
- Apriliyani, E., & Salim, Y. (2022). Analisis performa metode klasifikasi Naïve Bayes Classifier pada Unbalanced Dataset. *Indonesian Journal of Data and Science*, 3(2), 47-54.
- Gayo, A. A. P., Zainabun, Z., & Arabia, T. (2022). Karakterisasi Morfologi dan Klasifikasi Tanah Aluvial Menurut Sistem Soil Taxonomy di Kabupaten Aceh Besar. *Jurnal Ilmiah Mahasiswa Pertanian*, 7(3), 503-508.
- Khomsatun, Ikhsan D, Ali M, Kusriani. 2020. Sistem pengambilan keputusan pemilihan lahan tanam di Kabupaten Wonosobo dengan K-Means Clustering dan Topsis. *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI* [Internet]. [diunduh 2024 Des 20]; 9(1):55-63.
- Kurniawan MF, Rayes ML, Agustina C. 2021. Analisis kualitas tanah pada lahan tegalan berpasir di Das Mikro Supituring, Kabupaten Kediri, Jawa Timur. *Jurnal Tanah dan Sumberdaya Lahan* [Internet]. [diunduh 2024 Des 13]; 8(2): 527-537. <http://10.21776/ub.jtsl.2021.008.2.23>
- Minarsih, S., & Hanudin, E. (2020, June). Kualitas tanah pada beberapa tipe penggunaan lahan. In Seminar Nasional Pertanian Peternakan Terpadu (Vol. 2, No. 03, pp. 146-157).
- Pradana, D., Alghifari, M. L., Juna, M. F., & Palaguna, D. (2022). Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network. *Indonesian Journal of Data and Science*, 3(2), 55-60.
- Siwi, R. S., Nurcholis, M., & Virgawati, S. (2023). Morfologi dan Klasifikasi Tanah Pada Formasi Waturanda Dengan Penggunaan Lahan Hutan dan Tegalan di Desa Lebakwangi, Banjarnegara, Jawa Tengah. *Jurnal Tanah Dan Sumberdaya Lahan*, 10(2), 307-318.
- Vrebo D, Jones A, Lugato E, O'Sullivan L, Schulte R, Staes J, Meire P. 2020. Spatial evaluation and trade-off analysis of soil functions through Bayesian networks. *Wiley European Journal of Soil Science* [Internet]. [diunduh 2024 Des 12]; 1-15. <https://10.1111/ejss.13039>