

BAB II

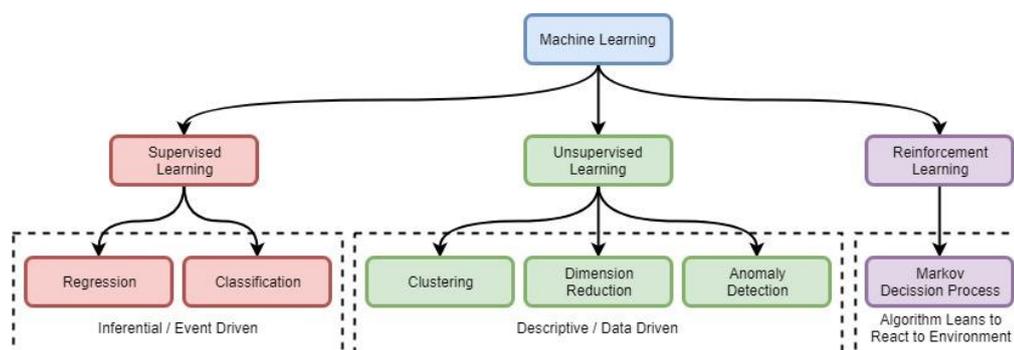
TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Machine Learning

Machine Learning merupakan salah satu teknik pembelajaran atau cabang dari kecerdasan buatan yang berfokus terhadap pembelajaran sendiri (*Self Learning*) (Afandi Nur Aziz Thohari, 2018). *Machine Learning* atau pembelajaran mesin adalah bagian dari *Artificial Intelligence* atau kecerdasan buatan. *Machine Learning* merupakan pembelajaran ilmiah mengenai algoritma dan model statistika dalam sistem komputer untuk melakukan tugas tertentu tanpa menggunakan perintah secara spesifik, dengan mengandalkan pola dan suatu Tindakan (Susi Yuliany, 2022).

Algoritma *Machine Learning* membangun model matematika berdasarkan data sampel (*Data Training*) untuk membuat keputusan atau prediksi. *Machine Learning* berfokus pada program komputer yang secara otomatis dapat meningkatkan kinerjanya berdasarkan pengalaman (Susi Yuliany, 2022).



Gambar 2. 1 Jenis *Machine Learning* (suprianto, 2018)

Terdapat beberapa jenis algoritma *machine learning*. Secara garis besar algoritma *machine learning* meliputi :

a. *Supervised Learning*

Supervised learning merupakan algoritma *machine learning* yang memiliki perkiraan fungsi pemetaan data *input* ke dapat *output* (Suprianto, 2018). *Supervised learning* banyak digunakan untuk memecahkan masalah klasifikasi ataupun masalah regresi.

b. *Unsupervised learning*

Unsupervised learning merupakan algoritma *machine learning* untuk menganalisis dan *clustering* (pengelompokkan) data yang tidak berlabel (Maulid, 2022).

c. *Reinforcement learning*

Reinforcement learning adalah algoritma *machine learning* yang belajar melalui interaksi dengan lingkungan (Maulid, 2022). *Reinforcement learning* merupakan jenis algoritma yang secara otomatis dapat menentukan kebijakan yang diperoleh melalui percobaan dan terus berinteraksi dengan lingkungan yang bersifat dinamis.

2.1.2 Prediksi

Menurut han (2007) prediksi adalah sebuah metodologi statistic yang dikembangkan oleh sir frances Galton seorang ahli matematika. Pada kenyataannya, regresi dan prediksi sebagai sinonim. Tetapi, beberapa teknik klasifikasi dapat diadaptasi untuk prediksi. Prediksi merupakan kombinasi dari beberapa teknik *data mining*. Pada umumnya, prediksi digunakan untuk menganalisis permasalahan atau kejadian di masalalu dalam urutan tertentu untuk memperkirakan kejadian di masa depan (Mustofa, 2021).

2.1.3 *Random forest*

Random forest adalah salah satu metode *machine learning* untuk klasifikasi yang digunakan dengan membangun pohon keputusan (Demmarita Leni, 2022). Menurut Breiman algoritma *random forest* dilakukan setelah membuat pemodelan data latih. Algoritma *random forest* harus menentukan jumlah *variable predictor* yang diambil secara acak dan pohon yang akan dibentuk agar mendapatkan hasil yang optimal. (Afandi Nur Aziz Thohari, 2018)

Random forest adalah algoritma *supervised learning* yang dipublikasikan oleh Breiman pada tahun 2001. Algoritma *random forest* sering digunakan untuk penyelesaian suatu permasalahan yang berhubungan dengan klasifikasi, regresi, dan sebagainya. *Random forest* ini berupa algoritma kombinasi dari *decision tree* dimana setiap *tree* (pohon) bergantung pada pada nilai *random vector* yang dijadikan sampel bebas dan merata pada semua *tree* (pohon) dalam *forest*. Hasil prediksi dari *random forest* didapatkan melalui hasil terbanyak di setiap individual *Decision Tree*.

2.1.4 Unsur hara tanah

Unsur hara tanah adalah unsur hara tanah yang dibutuhkan dalam jumlah yang cukup besar. Terdapat beberapa unsur hara tanah yang termasuk kedalam golongan unsur hara tanah seperti Nitrogen, Pospor, Kalium, Kalsium, Magesium (Mg), dan Belerang/Sulfur (S). Tetapi yang akan di analisis dalam penelitian ini hanya beberapa unsur hara tanah yaitu Nitrogen, Posfor, Kalium (NPK). Dari ketiga unsur hara tanah tersebut memiliki kegunaan atau perannya masing-masing, diantaranya untuk unsur hara tanah nitrogen berperan sebagai pembentuk klorofil yang sangat penting dalam proses fotosintesis, selain itu unsur hara tanah nitrogen ini dapat merangsang vegetatif tanaman secara keseluruhan (pertumbuhan akar, batang dan daun). Sedangkan untuk Pospor, unsur hara tanah ini dapat mempercepat dalam pertumbuhan bunga serta proses pemasakan biji dan buah. Dan untuk unsur hara tanah kalium berperan dalam memperkuat tanaman sehingga bunga, buah dan daun tidak mudah gugur.

2.1.5 *Training dan testing*

Training (latihan) dan *testing* (pengujian) merupakan dua faktor penting dalam keberhasilan *machine learning* dan dua jenis data yang wajib ada pada jenis *supervised learning*, karena kedua data tersebut merupakan karakteristik dari *supervised learning*. *Training set* adalah bagian dataset yang dilatih untuk membuat prediksi atau menjalankan fungsi dari algoritma *machine learning* sesuai dengan tujuan. Sedangkan *testing set* adalah bagian data set yang diuji untuk menentukan keakuratan atau performa (Mufadhol, 2022).



Gambar 2. 2 *Training Set* dan *Testing Set*

Data *Training* (data latih) merupakan data yang digunakan untuk membuat algoritma *machine learning*. Dalam proses pembuatan model, data *training* di-*input*-kan ke dalam algoritma *machine learning* untuk mempelajari hubungan antara *input* dan *output*. Sedangkan, data *testing* (data uji) merupakan data yang digunakan untuk mengevaluasi kinerja algoritma *machine learning* yang telah dibuat. Data *testing* digunakan untuk untuk mengevaluasi seberapa baik model tersebut dapat mengatasi permasalahan yang dihadapi (Ripple, 2023).

2.1.6 *Python*

Python merupakan salah satu bahasa pemrograman tingkat tinggi yang mudah dipahami serta pengimplementasiannya (Novita, 2022). Bahasa pemrograman *python* mempunyai banyak *library* yang memudahkan dalam pembuatan program yang melibatkan banyak manipulasi vektor dan matrik (*tensorflow*, *numpy*, *pandas*), serta tampilan visual berbagai grafik yang mudah dibaca (*library scikit learn matplotlib*, dan *heatmap (Library Seaborn)* untuk menunjukkan kolerasi dalam bentuk peta warna dan numerik (Widi Hatomo, 2021).

2.2 Penelitian Terkait

Penelitian terdahulu bertujuan untuk mendapatkan acuan dan bahan perbandingan untuk penelitian. Selain itu, untuk menghindari anggapan kesamaan antara penelitian terdahulu dan penelitian ini. Maka pada bagian kajian pustaka ini

peneliti mencantumkan hasil-hasil penelitian terkait atau *State of the art* sebagai berikut.

Tabel 2. 1 Penelitian Terkait (*State of The Art*)

Penelitian 1	
Judul Paper	<i>Crop Prediction Based on Charactersitics of the Agricultural Environment Using Various Feature Selection Tecniques and Classifiers</i>
Penulis	S. P. Raja, Barbara Sawicka, Zoran Stamenkovic, (Senior member, IEEE), dan Mariammal
Tahun	2022
Jurnal	IEEE (<i>Institute of Electrical and Electronics Engineers</i>)
Metode	<i>Naïve Bayes</i> <i>Decision Tree</i> <i>Support Vector Machine</i> <i>K Nearest Neighbor</i> <i>Bagging</i> <i>Random forest</i>
Hasil	Teknik <i>ansambel</i> memberikan hasil prediksi akurasi yang lebih baik dibandingkan dengan teknik klasifikasi. Hasil dari algoritma <i>random forest</i> apabila digunakan bersama dengan metode validasi lipatan nilai akurasi tertinggi yaitu 97,29, <i>precision</i> 98,94, <i>recall</i> 97,54 dan apabila bersama dengan validasi pemisahan data hasil akurasinya yaitu 97,11.
Penelitian 2	
Judul Paper	<i>Performance Evaluation of Naive Bayes Algorithm for Classification of Fertilizer Types</i>
Penulis	Rastri Prathivi, April Firman Daru, dan Sara Sharifzadeh
Tahun	2022

Jurnal	Telematika – Vol. 15, No.1, 57-69
Metode	<i>Naïve Bayes</i>
Hasil	<p>Pemodelan dengan algoritma <i>naïve bayes</i> telah diterapkan pada eksisting data penggunaan pupuk yang telah dihitung nilai probabilitas tiap kelasnya mendapatkan nilai probabilitas tertinggi dari suatu kelas. Pengukuran nilai akurasi dengan <i>naïve bayes</i> diukur dengan menggunakan metode <i>split validation</i>, dimana data latih akan dibagi menjadi dua bagian yaitu data latih dan data uji agar nilai yang diperoleh menggunakan model lebih akurat.</p> <p>Dari pemodelan yang diterapkan, di peroleh nilai akurasi sebesar 60% yang menunjukkan tingkat akurasi model yang diperoleh hasil klasifikasi berupa nama pupuk yang diharapkan untuk membantu dalam menentukan nama pupuk yang harus digunakan.</p>
Penelitian 3	
Judul Paper	<i>A machine Learning Approach to recommend Suitable Crops and Fertilizers For Agricultur</i>
Penulis	Govind Kumar Jha, Preetish Ranjan and Manish Gaur
Tahun	2020
Jurnal	<i>Recommender System with Machine Learning and Artificial Intelligence: Practical Tools and Applications in Medical, Agricultural and Other Industries</i>
Metode	<i>Naïve Bayes</i> <i>Bayes Net</i> <i>Logistic Regression</i> <i>MLP</i> <i>Random forest</i>
Hasil	Hasil perbandingan dari beberapa algoritma <i>machine learning</i> untuk menentukan kesehatan tanah pertambangan telah

	disimpulkan bahwa algoritma <i>random forest</i> menghasilkan prediksi yang paling akurat pada kondisi tanah. Hal tersebut dapat membantu para petani dengan merekomendasikan tanaman yang sesuai berdasarkan jenis tanah dan lahan serta membantu memberikan keputusan yang lebih baik mengenai pertanian seperti menunjukkan rekomendasi dosis nutrisi yang dibutuhkan dan memberikan saran kepada petani mengenai pupuk dan jumlah pupuk yang harus diterapkan.
Penelitian 4	
Judul Paper	<i>Spatial distribution of soil nutrient content for sustainable rice agriculture using geographic information system and Naïve Bayes classifier</i>
Penulis	Anton Yudhana, Andreyan Dwi Cahyo, Liya Yusrina Sabila, Arsyad Cahya Subrata and Ilham Mufan
Tahun	2024
Jurnal	<i>International Journal On Smart Sensing And Intelligent Systems</i>
Metode	<i>Geographic information system</i> <i>Naïve bayes classifier</i>
Hasil	Tujuan dari penelitian ini untuk membantu petani memantau unsur hara tanah terutama posfor. Hasil dari penelitian ini menunjukkan bahwa algoritma <i>bayes</i> dapat mengklasifikasikan status posfor tanah dataran rendah dengan probabilitas 0,34 untuk kondisi posfor sedang dan 0,66 untuk kondisi posfor tinggi. Hasil pengujian sampel menunjukkan tingkat kesalahan 3% dan tingkat keberhasilan 97%.
Penelitian 5	
Judul Paper	Impelemntasi Algoritma <i>Random forest</i> untuk Memprediksi Hasil Panen Padi di Desa Minanga
Penulis	Nahya Nur, Farid Wajidi, Sulfayanti, dan Wildayani

Tahun	2023
Jurnal	Jurnal Politeknik Caltex Riau, Vol 9910 : 58-64
Metode	<i>Random forest</i>
Hasil	Tujuan dari penelitian ini yaitu dapat membantu dalam memprediksi hasil panen padi sesuai dengan kriteria dan data yang ada sebelumnya seperti luas lahan, jumlah bibit, jenis pupuk, curah hujan dan lainnya. Hasil dari algoritma <i>Random forest</i> yang di dapatkan dari 9 pohon, variabel yang memiliki nilai paling tinggi adalah variabel luas lahan. Sehingga nilai akurasi yang diperoleh dengan menggunakan algoritma <i>random forest</i> yaitu 95,11%.
Penelitian 6	
Judul Paper	<i>Prediction of forest nutrient and moisture regimes from understory vegetation with random forest classification models</i>
Penulis	Lisein Jonathan, Fayolle Adeline, Legrain Andyne, Prevot Celine, dan Claessens Hugues
Tahun	2022
Jurnal	Ecological Indicators 144, 109446
Metode	<i>Random forest Classification</i>
Hasil	Penelitian ini bertujuan untuk memprediksi lokasi hutan, dan khususnya lokasi hutan rezim nutrisi dan kelembapan dari tumbuhan bawah hingga menyediakan keahlian flora dan fauna dari para ahli. Prediksi algoritma <i>random forest</i> , nutrisi mencapai tingkat kesalahan 30% pada validasi silang, sedangkan prediksi kelembapan memiliki kinerja yang sedikit lebih rendah baik dengan kesalahan <i>out-of-bag</i> sebesar 44%.
Penelitian 7	
Judul Paper	<i>Soil Analysis and Crop Fertility Prediction Using Machine Learning</i>

Penulis	Jagdeep Yadav, Shalu Chopra, dan Vijayalakshmi
Tahun	2021
Jurnal	<i>International Journal of Innovative Research in Advanced Engineering (IJIRAE)</i> . Issue 03, Volume 8
Metode	<i>Support Vector Machine</i> <i>Random forest</i> <i>Naïve Bayes</i> <i>Logitic Regressi</i> <i>Multilayer Perceptron</i> <i>ANN</i>
Hasil	Penelitian ini dilakukan untuk memprediksi kesuburan tanah dan hasil tanaman dengan jenis tanaman yang dapat tumbuh pada tanah subur. Beberapa model telah dilakukan pengujian. Model <i>multiplayer perceptron</i> dan <i>random forest</i> menunjukkan akurasi yang baik di antara semua pengklasifikasian tetapi <i>ANN</i> telah memberikan akurasi tertinggi dalam prediksi kesuburan tanah, prediksi tanaman dan hasil panen.
Penelitian 8	
Judul Paper	<i>Estimation of Soil Cohesion Using Machine Learning Method: A Random forest Approach</i>
Penulis	Hai-Bang Ly , Thuy-Anh Nguyen, and Binh Thai Pham
Tahun	2021
Jurnal	Hindawi, <i>Advances in Civil Engineering</i> . Article ID 8873993, 14 pages
Metode	<i>Random forest</i> <i>Support Vector Machine (SVM)</i> <i>Gaussian Regression Process (GPR)</i>
Hasil	Hasil penelitian menunjukkan bahwa algoritma <i>Random forest</i> yang diusulkan dapat secara akurat memprediksi kohesi

	tanah dengan akurasi tinggi $R=0,891$ dan rendah <i>error</i> $RMSE=3,323$ dan $MAE=2,511$ dan kemampuan prediksinya lebih baik dibandingkan dengan model SVM dan GPR. Oleh karena itu, algoritma <i>random forest</i> dapat digunakan sebagai pendekatan hemat biaya dalam memprediksi gaya kohesi tanah yang digunakan dalam desain dan inspeksi konstruksi.
Penelitian 9	
Judul Paper	<i>Analysis of Tree Species Suitability for Plantation Forests in Beijing (China) Using an Optimal Random forest Algorithm</i>
Penulis	Yuan Wang, Zhongke Feng, dan Wenyuan Ma
Tahun	2022
Jurnal	MPDI. Forest 13, 820.
Metode	<i>Random forest</i>
Hasil	<p>Hasil penelitian menunjukkan bahwa faktor iklim merupakan faktor yang paling mempengaruhi pertumbuhan pohon. Model prediksi memiliki akurasi yang baik, dengan AUC $0,75 - 0,85$. Diantara enam spesies pohon utama yang diteliti, pinus <i>tabulaeformis carr</i> cocok untuk seluruh lahan hutan di beijing. <i>Platyclusus orientalis</i>, <i>Robinia pseudoacacia lin</i>, dan <i>salix matsudana koids</i> cocok di tanam di daerah pegunungan sedangkan <i>sophora</i> dan <i>populus</i> cocok ditanam di daerah dataran tenggara Beijing.</p> <p>Model <i>random forest</i> yang dioptimalkan diterapkan dalam penelitian ini memberikan wawasan tentang kesesuaian distribusi spesies pohon di Beijing, dan dapat menjadi acuan dalam rancangan penghijauan.</p>
Penelitian 10	
Judul Paper	<i>Classification of crops based on macronutrients and weather data using machine learning techniques</i>
Penulis	Ritesh Dash , Dillip Ku Dash dan G.C. Biswal

Tahun	2020
Jurnal	<i>Elsevier. Result in Engineering</i>
Metode	<i>Support Vector Machine</i> <i>Decision Tree</i>
Hasil	Dalam penelitian ini telah menunjukkan cara memprediksi tanaman yang cocok untuk tanah tertentu dengan data tanah. Tiga jenis tanaman yang dipertimbangkan seperti padi, <i>whea</i> , dan tebu. Berdasarkan dengan ciri-ciri tanah dan parameter masukan lainnya, dengan perkiraan model SVM tanaman yang cocok yaitu dengan tingkat akurasi 91% (skor kebugaran). Model SVM menunjukkan kinerja yang lebih baik dibandingkan lainnya.

Berdasarkan beberapa penelitian sebelumnya dengan menggunakan algoritma *random forest* menunjukkan hasil akurasi yang baik dan dapat diterapkan untuk prediksi. Setiap algoritma regresi tentu memiliki kekurangan dan kelebihan masing-masing. Salah satu kelebihan mampu mengatasi masalah *overfitting* jika data training yang dimiliki berukuran kecil. Kekurangan yang dimiliki algoritma *random forest* yakni waktu komputasi untuk kumpulan data besar relatif lambat.

Berdasarkan studi literatur yang telah dilakukan, *machine learning* telah banyak digunakan pada beberapa bidang salah satu dalam bidang pertanian adalah klasifikasi prediksi unsur hara tanah. Pada penelitian ini akan melakukan prediksi unsur hara tanah dengan menggunakan algoritma *random forest*. Mengacu pada beberapa penelitian sebelumnya, algoritma *random forest* menghasilkan nilai akurasi tertinggi dibandingkan dengan beberapa algoritma *machine learning*

lainnya. Pada penelitian ini proses evaluasi model dilakukan dengan melakukan perhitungan nilai R2, dan MAPE.

Berdasarkan analisis dari beberapa penelitian sebelumnya yang telah diuraikan. Maka pada penelitian ini model atau metode yang akan digunakan adalah algoritma *random forest* untuk memprediksi unsur hara tanah (Nitrogen, Pospor, dan Kalium) berdasarkan temperatur, *humidity*, pH, dan *rainfall*.

2.3 Matrik Penelitian

Pada matrik penelitian ini menjelaskan antara hubungan penelitian terkait dengan penelitian yang dilakukan. Dalam matrik penelitian ini menunjukkan beberapa pendekatan yang berbeda dalam menemukan dan menyelesaikan sebuah solusi dari permasalahan dalam penelitian. Tabel 2.2 menggambarkan perbedaan penelitian yang diusulkan dengan beberapa penelitian terkait.

Beberapa kontribusi dari penelitian sebelumnya yang telah disebutkan sebelumnya adalah bahan untuk menyusun *state of the art* terkait dengan referensi, teori-teori yang dapat mendukung penelitian ini. Terdapat 2 penelitian terkait yang dijadikan acuan penelitian ini, yaitu :

Penelitian yang kedua dengan judul “Implementasi Algoritma *Random forest* untuk Memprediksi Hasil Panen Padi di Desa Minanga” yang diteliti oleh (Nahya Nur d. , 2023), tujuan dari penelitian ini yaitu dapat membantu dalam memprediksi hasil panen padi sesuai dengan kriteria dan data yang ada sebelumnya seperti luas lahan, jumlah bibit, jenis pupuk, curah hujan dan lainnya. Hasil dari algoritma *random forest* yang didapatkan dari 9 pohon, variabel yang memiliki nilai paling tinggi adalah variabel luas lahan. Algoritma tersebut mampu memprediksi hasil secara akurat dengan nilai akurasi sebesar 95,11%, nilai MAPE sebesar 4,884%, nilai RMSE sebesar 0,250, dan nilai R^2 sebesar 0,99.

Penelitian selanjutnya diteliti oleh (Hai-Bang Ly, 2021) dengan judul “*Estimation of Soil Cohesion Using Machine Learning Method: A Random forest Approach*”. Hasil penelitian menunjukkan bahwa algoritma *random forest* yang diusulkan dapat secara akurat memprediksi kohesi tanah dengan akurasi tinggi $R=0,891$ dan rendah error $RMSE=3,323$ dan $MAE=2,511$ dan kemampuan prediksinya lebih baik dibandingkan dengan model SVM dan GPR. Oleh karena itu, algoritma *random forest* dapat digunakan sebagai pendekatan hemat biaya dalam memprediksi gaya kohesi tanah yang digunakan dalam desain dan inspeksi konstruksi.

Beberapa jurnal yang dikumpulkan tersebut bertujuan agar penelitian ini menjadi semakin konsisten dan kokoh, karena isi yang terdapat dari masing-masing jurnal penelitian yang telah dikumpulkan diatas dapat dijadikan acuan bagi penelitian yang dilakukan.