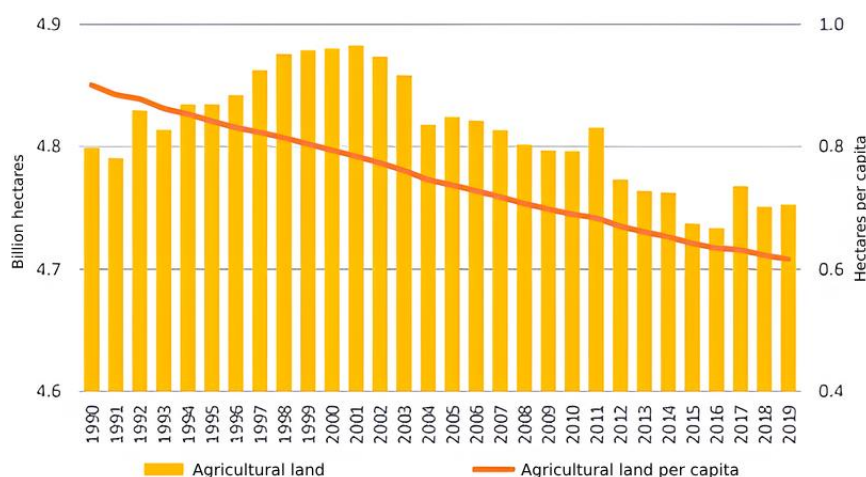


BAB I PENDAHULUAN

1.1 Latar Belakang

Peran sektor pertanian sangat penting terutama untuk mencukupi kebutuhan pangan di seluruh dunia. Luas lahan pertanian secara keseluruhan di muka bumi mencapai sekitar lima miliar hektar, atau 38 persen dari luas lahan global. Sekitar sepertiga dari luas lahan pertanian tersebut terdiri dari sawah dan perkebunan yang digunakan untuk menanam tanaman pangan, sedangkan sisanya merupakan padang rumput untuk menggembala ternak [1]. Lahan pertanian di negara-negara seluruh dunia tentunya terletak pada kondisi geografis yang beragam sehingga mempengaruhi kualitas lahan dan pertumbuhan jenis tanaman yang beragam pula.

Statistik menunjukkan bahwa luas lahan pertanian global per kapita tahun 1990-2019 menunjukkan grafik penurunan, ditunjukkan pada Gambar 1.1 [2]. Selain itu produktivitas beberapa tanaman pangan di beberapa negara cenderung melandai pada beberapa tahun terakhir [3]. Beberapa hal yang dapat mempengaruhi produktivitas tanaman pangan ialah faktor geografis yaitu berupa kelembaban, kadar pH tanah dan suhu udara [4]. Selain faktor geografis, faktor penting lainnya yang juga berpengaruh adalah kesuburan tanah. Kesuburan tanah ditentukan oleh unsur hara yang ada pada tanah, beberapa diantaranya adalah Nitrogen (N), *Phosphor* (P), serta Kalium (K) [5].



Gambar 1.1. Pertumbuhan luas lahan dan luas lahan per kapita dunia tahun 1990-2019 (Sumber: www.fao.org)

Tanaman memerlukan unsur hara dengan kadar yang berbeda-beda, bergantung pada beberapa kondisi lingkungan termasuk faktor geografis dan kesuburan tanah. Maka dari itu salah satu upaya yang dapat dilakukan untuk meningkatkan produktivitas yaitu dapat menanam jenis tanaman yang tepat dan sesuai dengan kondisi pada lingkungannya.

Salah satu solusi untuk merealisasikan hal tersebut adalah menggunakan *data mining*. *Data mining* merupakan bidang keilmuan yang menggabungkan teknik dari pembelajaran mesin, pengenalan pola, statistik, basis data, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari basis data yang besar [6]. Dengan menerapkan *data mining* maka diharapkan dapat melakukan klasifikasi tanaman pangan berdasarkan data kondisi lingkungan.

Terdapat beberapa penelitian sebelumnya yang terkait dengan hal ini. Salah satunya adalah sebuah penelitian yang berjudul *“Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms”* oleh Bondre D. dan Mahagaonkar S. melakukan perbandingan metode Random Forest dan SVM pada data tanaman yang memiliki fitur berupa nutrisi tanah, kelembaban, lokasi, cuaca, suhu udara, serta tekanan atmosfer. Pada penelitian ini didapatkan hasil bahwa ternyata Random Forest memiliki akurasi yang lebih baik yaitu sebesar 86%. [7]

Metode Random Forest juga menjadi metode terbaik pada penelitian *“A Machine Learning Approach to Recommend Suitable Crops and Fertilizers for Agriculture”* yang menghasilkan akurasi sebesar 99%. Penelitian tersebut juga meneliti metode Naive Bayes, *Logistic Regression* dan *Multilayer Perceptron*, namun hasilnya tidak lebih baik dari Random Forest. [8]

Berdasarkan referensi yang telah diuraikan di atas, penulis memilih metode Random Forest karena dapat menghasilkan akurasi yang cukup baik pada kasus klasifikasi data tanaman. Namun terdapat gap yang ada penelitian-penelitian tersebut, yaitu terbatas pada perbandingan metode dan menghasilkan model machine learning saja. Sehingga pada penelitian ini penulis ingin melakukan pengembangan dengan pencarian parameter terbaik serta menerapkan hasil modelnya dalam sebuah sistem. Untuk itu dalam laporan skripsi ini penulis akan melakukan penelitian dengan judul *“Penerapan Metode Random Forest Untuk Klasifikasi Tanaman Berdasarkan Kondisi Lingkungan”* dengan harapan bahwa dapat menghasilkan model dengan akurasi yang lebih baik dan dapat memberikan klasifikasi tanaman yang cocok ditanam pada kondisi lingkungan tertentu.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Berdasarkan latar belakang yang telah diuraikan sebelumnya, maka permasalahan yang ingin diselesaikan dalam penelitian ini adalah bagaimana menggunakan metode yang tepat untuk menghasilkan model yang dapat mengklasifikasikan tanaman berdasarkan kondisi lingkungan. Selain itu perlu dilakukan juga pencarian nilai parameter terbaik, diantaranya *n_estimators*, *max_depth*, dan *min_samples_leaf* sehingga hasil akurasi dapat optimal dan klasifikasi yang dihasilkan lebih baik. Dalam jangka panjang, hal ini dapat membantu meningkatkan produktivitas pertanian dan membantu dalam memenuhi kebutuhan pangan global. Dengan demikian, penerapan metode klasifikasi yang sesuai dapat memberikan manfaat besar bagi industri pertanian dan kesejahteraan manusia.

1.2.2 Metode Usulan

Metode yang diusulkan pada penelitian ini adalah dengan menerapkan Random Forest untuk klasifikasi jenis tanaman berdasarkan kondisi lingkungan. Untuk mengukur performa model maka akan digunakan metrik *Accuracy*, *F1-Score*, *Precision*, dan *Recall*. Dengan begitu dapat diketahui performa model yang dihasilkan.

1.2.3 Pertanyaan Penelitian

1. Berapa nilai parameter terbaik *n_estimators*, *max_depth*, dan *min_samples_leaf* pada metode Random Forest dalam kasus klasifikasi tanaman berdasarkan kondisi lingkungan?
2. Berapa tingkat *precision*, *recall*, *f1 score*, dan akurasi pada model yang dihasilkan menggunakan metode Random Forest dalam kasus klasifikasi tanaman berdasarkan kondisi lingkungan?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan Penelitian

1. Mengetahui nilai parameter terbaik *n_estimators*, *max_depth*, dan *min_samples_leaf* pada metode Random Forest untuk klasifikasi tanaman berdasarkan kondisi lingkungan.

2. Mengetahui tingkat precision, recall, f1 score, dan akurasi pada model yang dihasilkan menggunakan metode Random Forest dalam kasus klasifikasi tanaman berdasarkan kondisi lingkungan.

1.3.2 Manfaat Penelitian

Manfaat penelitian ini adalah untuk mengetahui jenis tanaman yang direkomendasikan untuk ditanam sesuai dengan kondisi lingkungan tertentu.

1.4 Batasan Masalah

Berikut adalah batasan masalah yang digunakan dalam penelitian ini agar lebih terarah dan tidak melebar:

1. Raw material dataset yang digunakan untuk klasifikasi adalah data publik yang berjudul *Crop Recommendation Dataset* yang tersedia di <https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset>
2. Dataset memiliki 7 fitur, diantaranya adalah kandungan Nitrogen (N), Phospor (P), Kalium (K), temperatur, kelembaban, tingkat keasaman (pH), serta curah hujan.
3. Dataset memiliki kelas berjumlah 22, diantaranya adalah padi, jagung, rami, kapas, kelapa, pepaya, jeruk, apel, melon, semangka, anggur, mangga, pisang, delima merah, kacang lentil, blackgram, kacang hijau, mothbean, kacang gude, kacang merah, kacang arab, dan kopi.
4. *Raw material* dataset berjumlah 2200 baris data, yang mana tiap kelasnya berjumlah 100 data.

1.5 Sistematika Penulisan

Laporan skripsi ini disusun dengan sistematika penulisan sebagai berikut.

BAB I : PENDAHULUAN

Pada bab ini berisi uraian latar belakang terkait penelitian “Penerapan Metode Random Forest Untuk Klasifikasi Tanaman Berdasarkan Kondisi Lingkungan”. Selanjutnya juga memuat rumusan masalah, tujuan dan manfaat penelitian, batasan masalah serta sistematika penulisan laporan skripsi ini.

BAB II : TINJAUAN PUSTAKA

Pada bab ini memuat landasan teori yang digunakan pada penelitian ini. Teori-teori yang dimaksud diantaranya adalah penjelasan terkait data mining, klasifikasi, kesesuaian lahan dan kondisi tanah,

library scikit-learn, random forest, k-fold cross validation, metrik akurasi, presisi, recall, dan f1-score serta contoh perhitungan random forest. Kemudian juga menjelaskan mengenai label tanaman yang digunakan pada data, diantaranya adalah tanaman padi, jagung, rami, kapas, kelapa, pepaya, jeruk, apel, melon, semangka, anggur, mangga, pisang, delima merah, kacang lentil, blackgram, kacang hijau, mothbean, kacang gude, kacang merah, kacang arab, kopi, serta pembahasan mengenai penelitian terkait.

BAB III : METODE PENELITIAN

Pada bab ini menguraikan metode yang dilakukan penulis dalam melakukan penelitian. Bab ini juga menjelaskan tahapan-tahapan yang dilakukan dalam penelitian ini. Diantaranya mengenai metode penelitian, alur penelitian, tahapan penelitian serta skenario pengujian yang dilakukan.

BAB IV : HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan mengenai hasil dan pembahasan dari proses penelitian yang dilakukan meliputi kebutuhan sistem, proses penyiapan data, label encoding, splitting, cross-validation, *training* dengan random forest, *testing* pengukuran performa, serta menghasilkan model yang diterapkan pada sebuah sistem.

BAB V : PENUTUP

Pada bab ini merumuskan kesimpulan dari hasil penelitian yang dilakukan serta memberikan saran peneliti selanjutnya.

DAFTAR PUSTAKA