

# **Laporan Proyek Akhir**

## **Location-Based Species Presence Prediction**



**Disusun Oleh:**

**12S22003 – Yohana Natalia Siahaan**

**12S22016 – Desri Stevie Natalie Dabukke**

**12S22017 – Lenna Febriana**

**INSTITUT TEKNOLOGI DEL**  
**FAKULTAS INFORMATIKA DAN TEKNIK**  
**ELEKTRO**

**TAHUN AJARAN 2024/2025**

## BAB 1 PENDAHULUAN

Keanekaragaman hayati merupakan salah satu aspek penting dalam menjaga keseimbangan ekosistem dan keberlanjutan lingkungan hidup. Salah satu tantangan utama dalam pengelolaan keanekaragaman hayati adalah mengetahui spesies apa saja yang ada atau berpotensi ada di suatu wilayah geografis. Proses ini biasanya memerlukan survei lapangan yang intensif dan memakan banyak waktu, tenaga, serta biaya. Namun, dengan kemajuan teknologi informasi dan ketersediaan data lingkungan berskala besar, kini terbuka peluang untuk mengembangkan sistem prediksi kehadiran spesies secara otomatis.

Salah satu pendekatan yang digunakan dalam konteks ini adalah *Species Distribution Modelling (SDM)*, yaitu metode yang memodelkan hubungan antara kondisi lingkungan dengan keberadaan spesies tertentu. Pendekatan ini sangat berguna untuk menghasilkan peta distribusi spesies, mendukung konservasi, dan membantu proses identifikasi spesies. Melalui kompetisi GeoLifeCLEF 2025, tersedia data observasi kehadiran spesies serta data lingkungan seperti citra satelit, tutupan lahan, iklim, dan intensitas aktivitas manusia, yang dapat dimanfaatkan untuk membangun model prediktif berbasis lokasi.

Proyek ini dirancang untuk menjawab tantangan tersebut dengan mengembangkan model *machine learning* yang mampu memprediksi spesies tumbuhan yang mungkin hadir di suatu titik lokasi berdasarkan data GPS dan atribut lingkungan sekitarnya. Sistem ini diharapkan tidak hanya memberikan kontribusi bagi penelitian dan konservasi, tetapi juga mendorong keterlibatan masyarakat luas dalam pengamatan lingkungan melalui teknologi berbasis data.

## **BAB 2 BUSSINESS UNDERSTANDING**

### **2.1 Objektif Bisnis**

Proyek ini bertujuan untuk mengembangkan sistem prediksi kehadiran spesies tumbuhan pada suatu lokasi berbasis data lokasi (GPS) dan berbagai prediktor lainnya seperti citra satelit, data iklim, tutupan lahan, serta jejak aktivitas manusia. Sistem ini diharapkan dapat:

- a. Membantu upaya konservasi dan manajemen keanekaragaman hayati dengan memetakan keberadaan spesies secara akurat.
- b. Menyediakan alat pendukung pengambilan keputusan bagi peneliti ekologi, pemerintah, dan organisasi lingkungan.
- c. Meningkatkan efisiensi alat identifikasi spesies otomatis (contoh: Pl@ntNet) dengan menyaring spesies berdasarkan lokasi.
- d. Mendorong partisipasi masyarakat umum (*citizen scientists*) melalui rekomendasi spesies berbasis lokasi.
- e. Mempercepat anotasi dan validasi data observasi spesies untuk memperkaya data set keanekaragaman hayati.

### **2.2 Tujuan Teknis**

Untuk memenuhi tujuan bisnis di atas, tujuan teknis proyek ini mencakup:

1. Membangun model Species Distribution Modelling (SDM) berbasis data spasial dan lingkungan.
2. Menggunakan data observasi seperti:
  - a. Koordinat GPS
  - b. Citra satelit (Sentinel-2)
  - c. Deret waktu satelit (Landsat ARD)
  - d. Deret waktu iklim (CHELSA)
  - e. Raster lingkungan (bioclimatic, tanah, dll)
3. Melatih model klasifikasi multikelas atau multilabel untuk memprediksi satu atau lebih spesies yang mungkin hadir pada titik lokasi tertentu.
4. Evaluasi performa model menggunakan metrik seperti Top-k Accuracy, Mean Average Precision (mAP), atau Recall@k

### 2.3 Rencana Proyek

Proyek ini dirancang untuk diselesaikan dalam jangka waktu 5 minggu, dimulai pada minggu ke-11 hingga minggu ke-15 kalender akademik kampus. Setiap minggunya difokuskan pada fase berbeda dalam siklus Data Science untuk memastikan alur kerja yang sistematis dan hasil yang optimal. Berikut adalah timeline dari pengerjaan proyek ini :

Task	Minggu				
Bussines Understanding					
Data Understanding					
Data Cleaning & Preparation					
Label Construction					
Modeling					
Evaluation & Improvement					
Final Integration & Reporting					
Presentation					

Dalam perencanaan proyek ini, algoritma Random Forest dipilih sebagai model utama untuk membangun sistem klasifikasi berdasarkan metadata pengamatan tanaman. Pemilihan model ini didasarkan pada beberapa pertimbangan, antara lain kemampuannya yang baik dalam menangani data tabular, robust terhadap outlier dan noise, serta minimnya kebutuhan preprocessing yang kompleks. Random Forest juga memiliki kelebihan dalam menangani feature non-linear dan interaksi antar fitur, yang sangat relevan dalam konteks data spasial dan temporal seperti latitude, longitude, tanggal, dan waktu pengamatan. Selain itu, model ini menyediakan interpretasi yang baik melalui *feature importance*, yang dapat membantu dalam menganalisis fitur mana yang paling berpengaruh dalam prediksi. Dengan kemudahan implementasi dan performa yang kompetitif, Random Forest menjadi pilihan yang tepat untuk digunakan dalam proyek ini, terutama untuk membangun baseline model yang dapat dikembangkan lebih lanjut di tahap evaluasi dan pengujian.