

**PENGEMBANGAN ALGORITMA GENERATIVE
ADVERSARIAL NETWORK (GAN) UNTUK MENAMBAH
DATA CITRA MULTISPEKTRAL LAHAN SAWAH DALAM
PEMETAAN KESUBURAN LAHAN**

**ANDI MUHAMMAD ALIFIKRI
G64190005**

Draft Proposal Kolokium

**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA dan ILMU PENGETAHUAN
ALAM INSTITUT PERTANIAN BOGOR
BOGOR 2022**

I. PENDAHULUAN

1.1. Latar Belakang

Citra multispektral adalah pemisahan objek dalam hal fitur spesifik yang unik untuk setiap objek berdasarkan spektrum elektromagnetik (Kadyanan 2022). Dengan kata lain, citra multispektral dapat diartikan sebagai citra yang berisi data gelombang frekuensi yang detail yang direpresentasikan dalam spektrum elektromagnetik. Frekuensi penglihatan manusia terletak pada rentang panjang gelombang 400-700 nm, dan akibatnya manusia hanya dapat melihat dan membedakan warna mulai dari ungu hingga merah. Citra multispektral menampilkan spektrum cahaya lain yang tidak tampak oleh penglihatan manusia yang sangat berguna untuk berbagai jenis kebutuhan manusia. Citra multispektral dapat diperoleh dengan menggunakan sensor radiometer.

Dalam dunia pertanian, citra multispektral dapat digunakan untuk memetakan kesuburan lahan persawahan padi yang diukur menggunakan Bagan Warna Daun (BWD) atau Leaf Color Chart (LCC) dimana warna daun padi menjadi parameter kebutuhan pupuk tanaman padi. Data citra multispektral pula dapat digunakan untuk melatih sebuah model *machine learning* yang nantinya dapat digunakan untuk mengklasifikasikan tingkat kesuburan lahan pertanian, tetapi tentunya hal ini membutuhkan record data multispektral yang cukup banyak untuk melatih model tersebut.

Data citra multispektral lahan persawahan dapat diperoleh melalui kamera UAV multispektral, dengan masing-masing record data menunjukkan rata-rata fitur multispektral pada suatu grid sawah berukuran 4 x 4 meter. UAV (*Unmanned Aerial Vehicle*) merupakan sebuah pesawat tanpa awak yang dapat dikendalikan oleh manusia menggunakan remot kontrol. Citra multispektral dari tangkapan kamera UAV sering kali didapatkan jumlah record data kelas yang tidak seimbang. Contohnya, pada sebuah lahan persawahan, persebaran area lahan tanaman padi yang memiliki kebutuhan pupuk tinggi dan kebutuhan pupuk rendah tidak sama banyak. Akibatnya ketika melakukan pengambilan citra multispektral pada area sawah tersebut akan didapati jumlah record data yang tidak seimbang antara data kelas dengan kebutuhan pupuk tinggi dan data kelas dengan kebutuhan pupuk rendah. Sehingga peneliti harus melakukan pengambilan data secara berulang pada lokasi dan kondisi lahan sawah yang berbeda dalam mengumpulkan record data multispektral sampai jumlah record data setiap kelasnya sama banyak. Tentunya hal ini akan sangat sulit dilakukan karena membutuhkan pengambilan data yang sangat banyak dan memakan waktu yang lama.

Klasifikasi data dengan distribusi kelas yang tidak seimbang sangat mempengaruhi kinerja machine learning dalam membuat model klasifikasi sehingga pengklasifikasian menjadi bias terhadap record data kelas yang jumlahnya lebih banyak (Patel *et al.* 2020). Data yang tidak seimbang (imbalance) ini bisa menjadi masalah pada saat peneliti ingin melakukan training model untuk mendapatkan model klasifikasi pemetaan kesuburan lahan menggunakan machine learning, karena setiap record kelas data tidak memiliki jumlah yang sama sehingga membuat tingkat akurasi klasifikasi pada setiap kelas menjadi tidak maksimal.

Oleh karena itu saya sebagai penulis ingin mengatasi kasus imbalance data ini dengan menggunakan algoritma Generative Adversarial Network (GAN).

Algoritma GAN merupakan algoritma yang dapat digunakan untuk membuat data sintetis (Fake) sehingga para peneliti tidak lagi perlu untuk melakukan pengambilan data baru di lapangan untuk menambah data dan sebagai gantinya menggunakan GAN untuk membuat/generate data sintetis yang sangat mirip dengan data aslinya (Aziira 2020). Algoritma GAN terdiri dari dua bagian yaitu generator untuk menghasilkan data sintetis dan diskriminator untuk membedakan data asli dan data fake, kedua bagian ini akan saling bersaing dalam melatih model, diskriminator akan meningkatkan akurasi dalam membedakan data sedangkan generator akan meningkatkan akurasi untuk menghasilkan data yang sangat mirip dengan data asli sehingga pada akhirnya diskriminator tidak dapat membedakan data asli dan data palsu. Ketika ini terjadi maka diperoleh model GAN yang optimal dalam membuat data sintesis/palsu untuk menambah data citra multispektral.

1.2. Rumusan Masalah

Berdasarkan latar belakang tersebut, maka penulis merumuskan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan algoritma GAN (Generative Adversarial Network) dalam memproduksi Data “Fake” untuk mengatasi *Imbalance Data* Citra Multispektral Lahan Sawah dalam Pemetaan Kesuburan Lahan?
2. Apakah data "fake" yang dihasilkan oleh model algoritma GAN ini dapat digunakan sebagai data training yang valid?

1.3. Tujuan

Tujuan dari penelitian ini adalah menghasilkan sebuah model yang dapat memproduksi (*Generate*) data “Fake” citra multispektral untuk menangani permasalahan jumlah data yang tidak seimbang.

1.4. Manfaat

Dengan menggunakan model dari GAN tersebut dapat membantu para peneliti dalam menambah data citra multispektral Lahan Sawah yang tidak seimbang sehingga para peneliti tidak mengulang pengambilan data ketika mengalami ketidak seimbangan data.

1.5. Ruang Lingkup

Ruang lingkup pada penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan data citra multispektral yang diambil dari kamera UAV multispectral di atas lahan persawahan di Kecamatan Dramaga.
2. Masing-masing record data menunjukkan rata-rata fitur multispectral pada suatu grid sawah berukuran 4 x 4 meter.
3. Label data adalah level kebutuhan pupuk pada grid-grid sawah tersebut, yang diukur menggunakan Bagan Warna Daun (BWD) atau Leaf Color Chart (LCC), di mana warna daun padi menjadi parameter kebutuhan pupuk tanaman padi.
4. Implementasi algoritma GAN (Generative Adversarial Network) menggunakan bahasa pemrograman python dengan memanfaatkan library tensorflow.

Daftar Pustaka

- Aziira AH. 2020. Pembuatan Data Numerik Sintetis Menggunakan Generative Adversarial Networks untuk Pengembangan Metode Dissolved Gas Analysis. Universitas Gadjah Mada. [diunduh 2022 Sep 16]. Tersedia pada: <http://etd.repository.ugm.ac.id/penelitian/detail/189995>
- Kadyanan IGAGA. 2022. Deteksi Perubahan Penggunaan Lahan dan Penutup Lahan Pada Citra Multispectral Berdasarkan Pendekatan Codebook dan Keyblock. *Jurnal Nasional Pendidikan Teknik Informatika : JANAPATI*. 11(2):145–155.doi:10.23887/janapati.v11i2.49102.
- Patel H, Singh Rajput D, Thippa Reddy G, Iwendi C, Kashif Bashir A, Jo O. 2020. A review on classification of imbalanced data for wireless sensor networks. *International Journal of Distributed Sensor Networks*. 16(4):1550147720916404.doi:10.1177/1550147720916404.