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

**ANDI MUHAMMAD ALIFIKRI**

**G64190005**



**DEPARTEMEN ILMU KOMPUTER**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**INSTITUT PERTANIAN BOGOR**

**BOGOR**

**2022**

ABSTRAK

ANDI MUHAMMAD ALIFIKRI. Implementasi Algoritme *Generative Adversarial Network* (GAN) Untuk Menambah Data Citra Multispektral Lahan Sawah Dalam Pemetaan Kesuburan Lahan. Dibimbing oleh Dr. KARLISA PRIANDANA, S.T., M.Eng.

DAFTAR ISI

[HALAMAN SAMPUL i](#_Toc118668382)

[DAFTAR ISI ii](#_Toc118668383)

[I PENDAHULUAN 1](#_Toc118668384)

[1.1 Latar Belakang 1](#_Toc118668385)

[1.2 Rumusan Masalah 3](#_Toc118668386)

[1.3 Tujuan 3](#_Toc118668387)

[1.4 Manfaat 3](#_Toc118668388)

[1.5 Ruang Lingkup 4](#_Toc118668389)

[II TINJAUAN PUSTAKA 5](#_Toc118668390)

[2.1 Citra Multispektral untuk Pemetaan Lahan Sawah 5](#_Toc118668391)

[2.2 Masalah dalam Imbalance Data 6](#_Toc118668392)

[2.3 Generative Adversarial Networks 6](#_Toc118668393)

[III METODE PENELITIAN 8](#_Toc118668394)

[3.1 Data Penelitian 8](#_Toc118668395)

[3.2 Tahapan Penelitian 8](#_Toc118668396)

[3.3 Lingkungan Pengembangan 10](#_Toc118668397)

[DAFTAR PUSTAKA 12](#_Toc118668398)

# PENDAHULUAN

## 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 secara detail yang direpresentasikan dalam spektrum elektromagnetik. Frekuensi penglihatan manusia hanya terletak pada rentang panjang gelombang 400-700 nm, dan akibatnya manusia hanya dapat melihat dan membedakan warna mulai dari ungu hingga merah (Santoso 2009). Sedangkan citra multispektral menangkap data dari berbagai panjang gelombang tertentu pada seluruh spektrum elektromagnetik sehingga memberikan lebih banyak informasi yang diperlukan dalam menganalisis karakteristik dan identifikasi komponen yang menyusun pada citra yang ditangkap (ElMasry *et al.* 2019). Oleh karena itu citra multispektral dapat menampilkan spektrum cahaya lain yang tidak tampak oleh penglihatan manusia yang sangat berguna untuk bermacam kebutuhan penelitian.

Dalam dunia pertanian, citra multispektral sering digunakan untuk memetakan kesuburan lahan persawahan padi dengan menggunakan Bagan Warna Daun (BWD) atau Leaf Color Chart (LCC) dimana warna daun padi menjadi parameter kebutuhan pupuk tanaman padi (Nasution *et al.* 2022). Citra multispektral dapat pula digunakan sebagai data latih untuk membuat sebuah model *machine learning* yang nantinya digunakan untuk mengklasifikasikan tingkat kesuburan lahan pertanian, tetapi tentunya hal ini membutuhkan *record* data yang cukup banyak untuk melatih model tersebut (Wijayanto *et al.* 2020).

Citra multispektral dapat diperoleh melalui alat UAV multispektral. UAV (*Unmanned Aerial Vehicle*) merupakan sebuah pesawat tanpa awak yang dapat dikendalikan oleh manusia menggunakan remot kontrol (Fahlstrom *et al.* 2022). UAV multispectral merupakan sebuah alat yang sangat efisien dalam melakukan monitoring tanaman padi di sawah (Wang *et al.* 2021). UAV memiliki berbagai sensor yang bisa menyamai sensor yang terdapat pada satelit sehingga dapat digunakan pada bidang pertanian dengan efisiensi yang tinggi untuk skala lapangan dalam pengambilan data (Shofiyanti 2011). Dengan efisiensi yang ditawarkan maka UAV multispectral banyak dimanfaatkan untuk memetakan dan menganalisis lahan persawahan.

Pada penelitian yang dilakukan oleh wang et al pada tahun 2021, mereka ingin memprediksi status ketersediaan nitrogen tanaman padi pada fase vegetatif menggunakan UAV multispectral. Di dalam penelitian ini mereka mendapati bahwa UAV multispektral cukup efisien ketika digunakan untuk memantau status nitrogen tanaman padi di sawah. Namun, ada beberapa gangguan latar belakang seperti banjir dan tanah gundul pada lahan sawah. Hal ini menciptakan masalah data yang tidak seimbang ketika mereka mencoba membuat sebuh model prediksi status nitrogen untuk tanaman padi pada data citra multispektral (Wang *et al.* 2021). Masalah serupa terjadi pada sebuah lahan persawahan di kecamatan Dramaga. Persebaran tanaman padi yang berkebutuhan pupuk tinggi dan berkebutuhan pupuk rendah tidak sama banyak dalam satu lahan sawah. Akibatnya ketika melakukan pengambilan data citra multispektral pada area sawah tersebut, jumlah data yang diperoleh tidak seimbang antara data kelas dengan kebutuhan pupuk tinggi dan data kelas dengan kebutuhan pupuk rendah.

Data yang tidak seimbang (imbalance) bisa menjadi masalah ketika para peneliti ingin melakukan training model untuk mendapatkan model klasifikasi pemetaan kesuburan lahan menggunakan machine learning, karena setiap kelas data tidak memiliki jumlah yang sama, sehingga membuat tingkat akurasi klasifikasi pada setiap kelas menjadi tidak maksimal (Zhang *et al.* 2021).  Distribusi kelas yang tidak seimbang akan sangat mempengaruhi kinerja machine learning dalam membuat model klasifikasi. Dampaknya adalah pengklasifikasian menjadi bias atau *overfitting* terhadap *record* data yang jumlah kelasnya lebih banyak (Patel *et al.* 2020).

Ada beberapa pendekatan dalam mengatasi data imbalance yaitu pendekatan pada tingkat data dan pendekatan pada tingkat algoritme. Pada pendekatan tingkat data menggunakan *oversampling* dan undersampling. Dikatakan pendekatan tingkat data karena mereka langsung memanipulasi data yang ada pada dataset demi menyeimbangkan sampel data dengan cara mengurangi sampel data mayoritas ataupun menghapus kelas minoritas. Kekurangan dari teknik *oversampling* ini yaitu bisa terjadi over fitting ketika dilakukan pelatihan model sedangkan untuk teknik undersampling dapat menyebabkan kehilangan beberapa sampel data yang (Rout *et al.* 2018). Lalu tipe yang kedua adalah pendekatan tingkat algoritme atau biasa disebut sebagai pendekatan internal karena memanfaatkan desain algoritme klasifikasi baru atau memodifikasi algoritme yang ada demi mengatasi bias akibat ketidakseimbangan data (Spelmen dan Porkodi 2018).

Fokus utama pada penelitian ini adalah ingin menambah jumlah sampel data. Maka dari itu pendekatan yang akan digunakan adalah pendekatan tingkat data (*Data-Level Approaches*) dengan teknik *oversampling*. Teknik *oversampling* yang populer digunakan saat ini adalah SMOTE (*Synthetic Minority Oversampling Technique*) (Douzas *et al.* 2019). Algoritme SMOTE telah terverifikasi efektif dan valid dalam memproduksi sampel data sintetis dan telah banyak variasi SMOTE yang dibuat oleh beberapa peneliti (Xie dan Zhang 2018). Teknik *oversampling* lainnya yaitu dengan menggunakan algoritme GAN (*Generative Adversarial Networks).* GAN merupakan sebuah algoritme deep learning yang yang digunakan untuk membuat data sintetis dari hasil pelatihan dua buah model algoritme yang saling bersaing (*adversarial*). Berbeda dengan SMOTE yang melakukan pendekatan pembelajaran berdasarkan informasi lokal dari dataset yang ada, GAN melakukan proses pembelajaran pada keseluruhan distribusi sampel data (Xie dan Zhang 2018). Perbedaan lainnya yaitu pada SMOTE memproduksi sampel data sintetis di sepanjang garis yang menghubungkan antar sampel data minoritas, sedangkan GAN bekerja dengan memperkirakan distribusi data aslinya lalu memproduksi data sintetis untuk kelas minor (Douzas dan Bacao 2018).

Secara umum, algoritme *Generative Adversarial Network* (GAN) digunakan untuk memproduksi citra sintetis/buatan dengan mempelajari data nilai pixel pada citra aslinya. Kita ketahui bersama bahwa nilai pixel pada sebuah citra mengikuti sebaran normal atau distribusi gauss, yang dapat dinormalisasi menjadi [-1,1] dengan menggunakan transformasi min-max. Sedangkan data tabular biasanya memiliki nilai kontinu yang bukan merupakan sebaran normal. Jadi ketika dilakukan transformasi min-max pada data tersebut maka menyebabkan *vanishing gradient problem* pada saat diimplementasikan keGAN(Xu *et al.* 2019).

Untuk mengatasi masalah tersebut, Xu *et al. (*2019) melakukan penelitian dengan judul “*Modeling Tabular Data using Conditional GAN*”. Mereka ingin mengembangkan algoritme *Conditional* GAN agar dapat digunakan untuk memproduksi data sintetis pada data tabular. Kemudian mereka berhasil mengembangkan algoritme GANyang mereka sebut sebagai CTGAN *(Conditional Tabular Generative Adversarial Network).* Mereka menunjukkan bahwa CTGAN mampu mempelajari distribusi data secara lebih baik dibandingkan *Bayesian Network*. Sedangkan mereka mengamati bahwa sejauh ini belum ada *deep learning generative model* yang mampu mengungguli Bayesian Network selain CTGAN ini.

Oleh karena itu, penulis tertarik untuk meneliti apakah dengan menerapkan algoritme CTGAN tersebut dapat menghasilkan data sintetis yang valid dan dapat digunakan untuk menambah data *imbalance* pada citra multispektral lahan sawah untuk pemetaan kesuburan lahan.

## Rumusan Masalah

Data citra lahan sawah umumnya imbalance sehingga tidak cukup untuk digunakan dalam pemodelan pemetaan kesuburan lahan sawah, maka penulis merumuskan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan algoritme GAN *(Generative Adversarial Network)* untuk memproduksi data sintetis demi mengatasi *Imbalance Data* Citra Multispektral Lahan Sawah dalam pemetaan Kesuburan Lahan?
2. Apakah data sintetis yang dihasilkan oleh model algoritme GAN ini dapat digunakan sebagai data training yang valid?

## Tujuan

Berdasarkan latar belakang dan rumusan masalah, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan algoritme GAN *(Generative Adversarial Network)* untuk membangkitkan data tambahan/sintesis citra multispektral lahan sawah.
2. Membuat model klasifikasi kesuburan lahan sawah menggunakan data citra multispektral yang sudah ditambah dengan data sintetis.
3. Menganalisis pengaruh penambahan data sintetis terhadap akurasi model klasifikasi.

## 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 ketidakseimbangan data.

## Ruang Lingkup

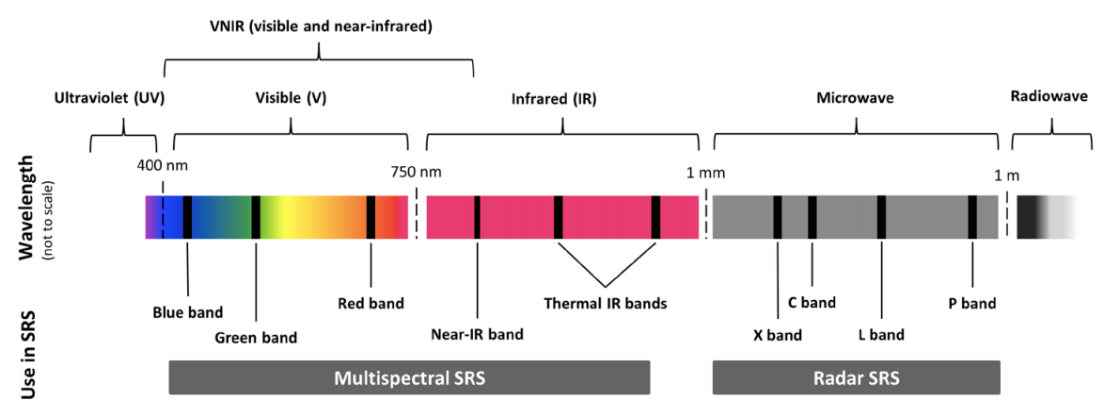
Ruang lingkup pada penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan data sekunder citra multispektral yang diambil dari kamera UAV multispectral di atas lahan persawahan di Kecamatan Dramaga yang telah diambil dari penelitian sebelumnya.
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 algoritme GAN (Generative Adversarial Network) menggunakan bahasa pemrograman python dengan memanfaatkan library tensorflow.

# TINJAUAN PUSTAKA

## Citra Multispektral untuk Pemetaan Lahan Sawah

Hanya sebagian kecil dari gelombang elektromagnetik yang dapat dilihat oleh mata manusia. Berbeda dengan mata manusia, sensor kamera multispektral yang biasanya dijumpai pada satelit dan/atau pesawat tanpa awak (UAV) dapat menangkap lebih banyak informasi gelombang elektromagnetik misalnya gelombang inframerah, ultraviolet, dan bahkan gelombang mikro.



Gambar 1 Spektrum Gelombang Elektromagnetik (Prerona 2020)

Di dalam sebuah penelitian yang dilakukan oleh Rokhmatuloh *et al.* (2019) dengan judul “Pemetaan Sawah Padi menggunakan citra UAV Multispektral”. Penelitian ini ingin meneliti seberapa efektif sensor citra multispektral yang ada pada UAV untuk menghasilkan nilai NDVI (Normalized Difference Vegetation Index) dalam memetakan persawahan. UAV Multispektral yang digunakan memiliki sensor multispectral empat bands yaitu green, red, red edge, dan near-infrared. Dalam melakukan pemetaan sawah padi mereka menggunakan nilai NDVI sebagai alat untuk membedakan berbagai jenis dan karakteristik tanaman. Untuk menghitung nilai NDVI ini digunakan formula perhitungan sebagai berikut:

(1)

di mana *red* dan NIR merupakan nilai radiasi cahaya merah tampak dan near-infrared yang ditangkap oleh sensor. Dasil penelitian ini didapatakan kesimpulan bahwa sensor multispektral pada UAV telah berhasil menghasilkan nilai NDVI untuk membantu memetakan lahan sawah dengan dapat membedakan jenis jenis tanaman yang ada pada lahan sawah yang diteliti. Selain itu nilai NDVI yang dihasilkan pula cukup akurat dalam membedakan umur tanaman padi, tetapi NDVI yang diperoleh cukup bervariasi dan tidak jarang beberapa tanaman yang sama memiliki nilai NDVI yang berbeda.

## Masalah pada Imbalance Data

Ketidakseimbangan data kelas pada sebuah dataset terjadi karena adanya kelas yang jumlah datanya sedikit (kelas minoritas) dibandingkan kelas lain (kelas mayoritas). Sebagai contoh dari 1000 jumlah sampel terdapat 99% data yang tergolong kelas mayoritas dan hanya terdapat 1% data yang tergolong kelas minoritas, lalu ketika melakukan pelatihan model didapatkan akurasi 99% yang sebenarnya akurasi itu hanya berasal dari kelas mayoritas sedangkan kelas minoritas didominasi oleh kelas mayoritas, bisa saja 1% data minoritas tersebut mengandung inforamsi yang penting. Sehingga ketika menjalankan model klasifikasi, sebagian besar akan salah mengklasifikasikan *(missclassification)* sampel data kelas minoritas, sedangkan sampel data pada kelas mayoritas akan jarang untuk terjadi kesalahan klasifikasi (Ali *et al.* 2019). Kesalahan klasifikasi pada model seperti ini memberikan dampak pada kerugian biaya dan waktu (Longadge *et al.* 2013).

Data yang tidak seimbang *(imbalance)* dapat menjadi masalah pada saat peneliti ingin melakukan training model untuk mendapatkan model klasifikasi menggunakan machine learning, karena setiap kelas data tidak memiliki jumlah yang sama, sehingga membuat tingkat akurasi klasifikasi pada setiap kelas menjadi tidak maksimal (Zhang *et al.* 2021).  Ketika melakukan klasifikasi data dengan distribusi kelas yang tidak seimbang akan sangat mempengaruhi kinerja machine learning dalam membuat model klasifikasi. sehingga akibatnya adalah pengklasifikasian menjadi bias terhadap *record* data yang jumlah kelasnya lebih banyak (Patel *et al.* 2020)

Pendekatan yang dapat dilakukan dalam mengatasi data imbalance yaitu pendekatan pada tingkat data dan pendekatan pada tingkat algoritme. Pada pendekatan tingkat data menggunakan *oversampling* dan undersampling. Dikatakan pendekatan tingkat data karena mereka langsung memanipulasi data yang ada pada dataset demi menyeimbangkan sampel data dengan cara mengurangi sampel data mayoritas ataupun menghapus kelas minoritas. Kekurangan dari teknik *oversampling* ini yaitu bisa terjadi over fitting ketika dilakukan pelatihan model sedangkan untuk teknik undersampling dapat menyebabkan kehilangan beberapa sampel data yang (Rout *et al.* 2018). Lalu tipe yang kedua adalah pendekatan tingkat algoritme atau biasa disebut sebagai pendekatan internal karena memanfaatkan desain algoritme klasifikasi baru atau memodifikasi algoritme yang ada demi mengatasi bias akibat ketidakseimbangan data (Spelmen dan Porkodi 2018).

## Generative Adversarial Networks

Algoritme Generative Adversarial Networks (GAN) merupakan salahsatu jenis algoritme *oversampling*. Dikatakan *oversampling* karena algoritme ini berusaha membuat atau *generate* data sintetis baru menyerupai data yang dimiliki dengan tujuan menambah jumlah sampel data. GAN bekerja dengan mempelajari distribusi data asli baik data mayoritas maupun data minoritas yang jumlahnya terbatas, lalu memanfaatkan distribusi yang telah dipelajari untuk menghasilkan data sintetis (Sampath *et al.* 2021).

Algoritme gan pertamakali diperkenalkan oleh Goodfellow *et al.* (2014). Sesuai namanya, GAN merupakan dua jaringan machine learning yang saling bersaing *(adversarial)* untuk mengoptimalkan kinerja masing masing. Kedua jaringan itu ialah *Generator* dan *Discriminator. Generator* akan membuat data sintetis dengan tujuan untuk menipu diskriminator sedangkan diskriminator akan mencoba membedakan antara dara asli dengan data yang dibuat oleh diskriminator. Kedua jaringan ini memiliki kemampuan untuk mengoptimalkan kinerja masing-masing dengan konsep backpropagation. Ketika *discriminator* berhasil membedakan antara data asli dengan data sintetis maka dikembalikan sebagai feedback ke *generator* dan generator akan berusaha meningkatkan kinerjanya. Begitu pula sebaliknya diskriminator akan terus belajar untuk membedakan data palsu dan data asli secara akurat. Proses pelatihan selesai ketika diskriminator tidak lagi dapat membedakan antara data asli dan data sintetis. Secara matematis Goodfellow *et al.* (2014) menggambarkan GAN dalam persamaan (2).

(2)

dimana z adalah variabel input dengan sampel data berupa random noise untuk melatih jaringan *Generator,*  merepresentasikan probabilitas x merupakan data asli bukan merupakan data hasil dari *Generator* . Diskriminator (D) dilatih untuk memaksimalkan probabilitas penentuan sampel data asli dan data sintetis. Secara simultan, Generator (G) dilatih untuk meminimalkan nilai agar D dapat terkecoh. Dengan kata lain, D dan G memainkan permainan minimax-game.

Di dalam paper tersebut pula disebutkan metode *Conditional Generative model* untuk mendapatkan data sintetis sesuai kelas yang diinginkan. Dengan secara sederhana menambahkan variabel y sebagai input tambahan pada D dan G.

(3)

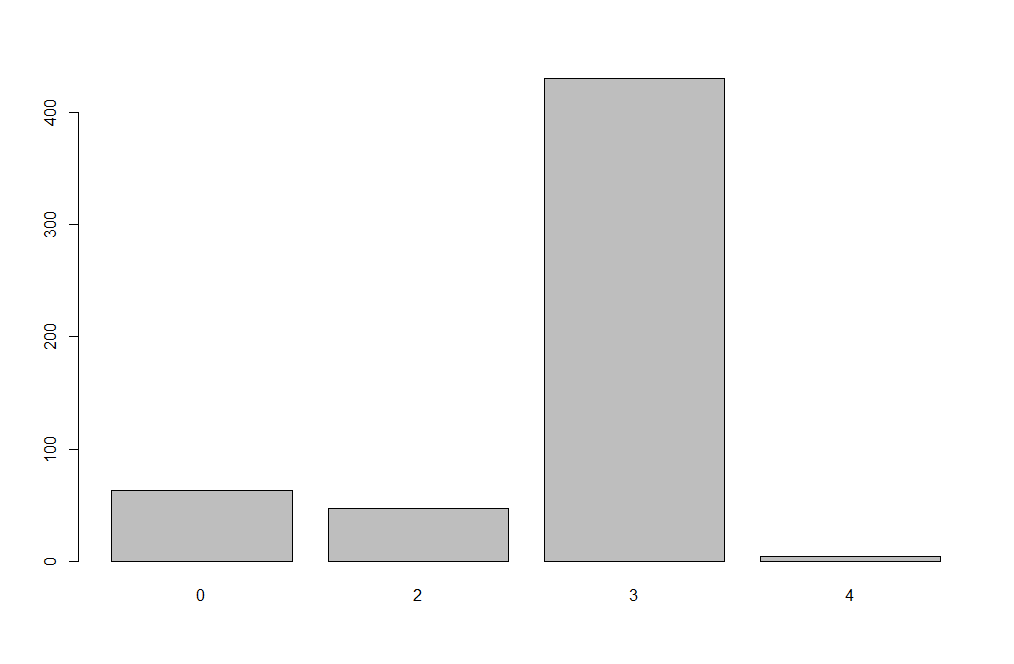
*Conditional model* ini mampu mengarahkan proses generasi data sehingga output dari G dapat dikontrol sesuai apa yang diinginkan. Nilai y ini merupakan informasi tambahan dapat berupa label kelas setiap data (Mirza dan Osindero 2014).

# METODE PENELITIAN

## Deskripsi Data

Data yang digunakan pada penelitian ini merupakan data sekunder yang diambil pada sebuah penelitian oleh Kahfi Gunardi seorang mahasiswa Program Magister IPB University yang meneliti tentang Perbandingan Algoritme Klasifikasi untuk Mendeteksi Kebutuhan Nitrogen Tanaman Padi berdasarkan Data Citra Multi-spektral Drone. Pengambilan data citra multispektral ini menggunakan sensor kamera *Unmanned Aerial Vehicle* (UAV) atau Drone pada lahan sawah di desa Margajaya Kecamatan Dramaga, Kabupaten Bogor.

Dataset ini memiliki total 544 record data dan memiliki 13 atribut yaitu “rgb\_b1, rgb\_b2, rgb\_b3, green\_b1, green\_b2, ndvi\_b1, nir\_b1, nir\_b2, red\_b1, red\_b2, redEdge\_b1, redEdge\_b2, Label”. Salah satu atributnya adalah label yang terdiri dari tiga kelas yaitu 2, 3, dan 4 menggambarkan jumlah pupuk yang dibutuhkan pada tanaman padi.

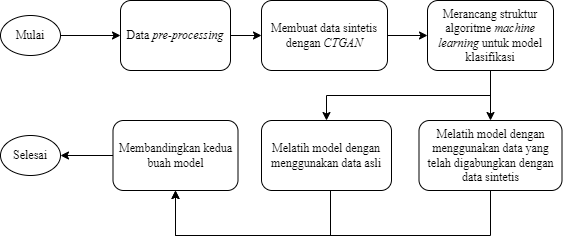


Gambar 2 Plot jumlah data pada setiap label kelas

Dari gambar diatas dapat kita lihat dari total 544 record, data paling banyak terdapat pada label 3 sebanyak 430 data, label 2 sebanyak 47 data, dan yang paling sedikit terdapat pada label 4 sebanyak 4 data. Sedangkan yang berlabel 0 merupakan data yang tidak diketahui label sebenarnya.

## Tahapan Penelitian

Penelitian ini mencakup beberapa tahapan, yang ditunjukan pada diagram alir berikut:



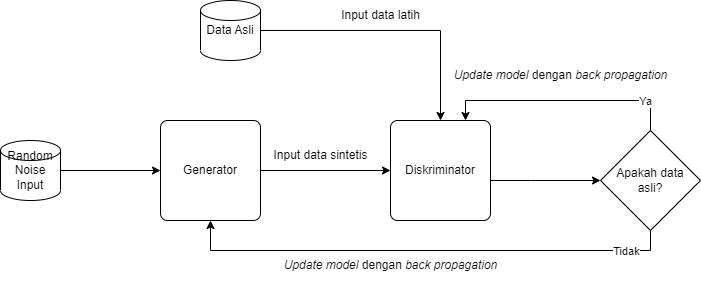
Gambar 3 Skema metode penelitian untuk mengetahui pengaruh penambahan data sintetis

### Data Pre-processing

Melakukan data *pre-processing* untuk membuat sampel data yang akan diteliti lebih optimal dan siap digunakan. Tahap *pre-processing* pada penelitian ini meliputi pembersihan data (data *cleaning)* dan reduksi data. Pembersihan data ditujukan untuk menghilangkan *mising value* dan menghapus data yang tidak konsisten. Sedangkan reduksi data ditujukan untuk menghapus baris data yang tidak memiliki label kelas.

### Membangkitkan Data Sintetis Dengan CTGAN

Pada tahap ini data yang telah dibersihkan akan digunakan pada algortima CTGAN untuk menghasilkan data sintetis yang mirip dengan data aslinya. Algortime CTGAN akan melatih dua model yaitu diskriminator dan generator. Data yang asli akan dipelajari oleh diskrimnator, slanjutnya generator akan membuat data sintetis untuk diklasifikasikan oleh diskriminator apakah itu merupakan data asli atau data sintetis. Ketika diskriminator berhasil menebak data tersebut sebagai data sintetis maka generator akan meningkatkan performanya untuk memproduksi data sintetis yang semirip mungkin dengan data aslinya. Proses ini akan terus berulang hingga diskriminator tidak lagi dapat membedakan antara data asli dengan data palsu. Selanjutnya generator akan memproduksi data sintetis sebanyak yang diinginkan dan hasil dari generator ini akan digabungkan dengan data aslinya untuk digunakan pada tahap pelatihan model *decision tree*.



Gambar 4 Ilustrasi alur kerja algoritma CTGAN

### Merancang Struktur Model Klasifikasi

Model kalsifikasi yang akan digunakan pada penelitian ini yaitu *decission tree* dengan memanfaatkan library dari sklearn pada python yaitu sklearn.tree.DecisionTreeClassifier(). Merancang struktur model *decision tree* dengan menentukan parameter dan atribut yang akan digunakan.

### Melatih Model

Pada tahap ini, akan dibuat dua buah model *decision tree*. Model yang pertama menggunakan data sebagai data latih. Model kedua menggunakan data campuran di mana data ini merupakan perpaduan antara data asli dan data yang telah dihasilkan oleh CTGAN.

### Membandingkan Kedua Buah Model

Tahap ini merupakan tahap membandingkan performa hasil pengujian antara kedua buah model. Hasil pengujian kemudian dilakukan analisis untuk diambil kesimpulan. Pada tahapan penelitian ini menggunakan confusion matrik untuk menghitung akurasi, precision, F1 score, dan recall untuk mengukur kinerja suatu model. Dari hasil ini kemudian dapat diambil kesimpulan apakah dengan menggunakan CTGAN dapat menghasilkan data sintetis yang valid dan bisa digunakan untuk menambah data citra multispektral lahan sawah dalam pemetaan kesuburan lahan.

## Lingkungan Pengembangan

Spesifikasi perangkat keras dan perangkat lunak yang digunakan untuk penelitian ini adalah sebagai berikut:

### Perangkat keras dengan spesifikasi:

* Processor AMD Reyzen 5 4600H CPU @ 3.0Ghz.
* Graphic Processor NVIDIA® GeForce® GTX 1050.
* RAM 16 GB.
* SSD 512 GB.

### Perangkat lunak yang digunakan:

* Sistem Operasi Windows 11 64-bit
* Bahasa Pemrograman phyton 3.7.15
* Lingkungan Pengembangan Google Colabs.
* Library python SKLEARN
* Library python CTGAN

DAFTAR PUSTAKA

Ali H, Salleh M, Saedudin R, Hussain K, Mushtaq M. 2019. Imbalance class problems in data mining: A review. *Indonesian Journal of Electrical Engineering and Computer Science*. 14.doi:10.11591/ijeecs.v14.i3.pp1552-1563.

Douzas G, Bacao F. 2018. Effective data generation for imbalanced learning using conditional generative adversarial networks. *Expert Systems with Applications*. 91:464–471.doi:10.1016/j.eswa.2017.09.030.

Douzas G, Bacao F, Fonseca J, Khudinyan M. 2019. Imbalanced Learning in Land Cover Classification: Improving Minority Classes’ Prediction Accuracy Using the Geometric SMOTE Algorithm. *Remote Sensing*. 11(24):3040.doi:10.3390/rs11243040.

ElMasry G, Mandour N, Al-Rejaie S, Belin E, Rousseau D. 2019. Recent Applications of Multispectral Imaging in Seed Phenotyping and Quality Monitoring—An Overview. *Sensors*. 19(5):1090.doi:10.3390/s19051090.

Fahlstrom PG, Gleason TJ, Sadraey MH. 2022. *Introduction to UAV Systems*. John Wiley & Sons. Ed ke-Google-Books-ID: s8Z6EAAAQBAJ.

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. 2014. Generative Adversarial Nets. Di dalam: *Advances in Neural Information Processing Systems*. [internet] Vol. 27. Curran Associates, Inc. [diunduh 2022 Okt 14]. Tersedia pada: https://proceedings.neurips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html

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.

Longadge MR, Dongre SS, Malik DL. 2013. Class Imbalance Problem in Data Mining: Review. 2(1):6.

Mirza M, Osindero S. 2014. Conditional Generative Adversarial Nets. [diunduh 2022 Okt 8]. Tersedia pada: http://arxiv.org/abs/1411.1784

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.

Prerona. 2020. A technical deep-dive into Satellite Imaging, Multispectral, SAR and GAN. *AiDash*. [diunduh 2022 Nov 18]. Tersedia pada: https://www.aidash.com/a-technical-deep-dive-into-satellite-imaging-multispectral-sar-and-gan/

Rokhmatuloh, Supriatna, Giok Pin T, Hernina R, Ardhianto R, Putut Ash Shidiq I, Adi W. 2019. Paddy Field Mapping Using Uav Multi-Spectral Imagery. *International Journal of GEOMATE*. 17(61):242–247.doi:https://doi.org/10.21660/2019.61.icee408.

Rout N, Mishra D, Mallick MK. 2018. Handling Imbalanced Data: A Survey. Di dalam: Reddy MS, Viswanath K, K.M. SP, editor. *International Proceedings on Advances in Soft Computing, Intelligent Systems and Applications*. Singapore. Singapore: Springer. (Advances in Intelligent Systems and Computing). hlm. 431–443.

Sampath V, Maurtua I, Aguilar Martín JJ, Gutierrez A. 2021. A survey on generative adversarial networks for imbalance problems in computer vision tasks. *J Big Data*. 8(1):27.doi:10.1186/s40537-021-00414-0.

Santoso I. 2009. *Interaksi Manusia dan Komputer*. Ed ke-2. C.V. Andi Offset. Ed ke-Google-Books-ID: \_pXa7CvwTC0C.

Shofiyanti R. 2011. TEKNOLOGI PESAWAT TANPA AWAK UNTUK PEMETAAN DAN PEMANTAUAN TANAMAN DAN LAHAN PERTANIAN. *Informatika Pertanian*. 20(2):58–64.

Spelmen VS, Porkodi R. 2018. A Review on Handling Imbalanced Data. Di dalam: *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*. *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*;. hlm. 1–11.

Wang Y-P, Chang Y-C, Shen Y. 2021. Estimation of nitrogen status of paddy rice at vegetative phase using unmanned aerial vehicle based multispectral imagery. *Precision Agric*. 23(1):1–17.doi:10.1007/s11119-021-09823-w.

Wijayanto AW, Wahyu Triscowati D, Marsuhandi AH. 2020. Maize field area detection in East Java, Indonesia: An integrated multispectral remote sensing and machine learning approach. Di dalam: *2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE)*. *2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE)*;. hlm. 168–173.

Xie Y, Zhang T. 2018. Imbalanced Learning for Fault Diagnosis Problem of Rotating Machinery Based on Generative Adversarial Networks. Di dalam: *2018 37th Chinese Control Conference (CCC)*. *2018 37th Chinese Control Conference (CCC)*;. hlm. 6017–6022.

Xu L, Skoularidou M, Cuesta-Infante A, Veeramachaneni K. 2019. Modeling Tabular data using Conditional GAN. Di dalam: *Advances in Neural Information Processing Systems*. [internet] Vol. 32. Curran Associates, Inc. [diunduh 2022 Nov 10]. Tersedia pada: https://papers.nips.cc/paper/2019/hash/254ed7d2de3b23ab10936522dd547b78-Abstract.html

Zhang C, Li J, Zhao Y, Li T, Chen Q, Zhang X, Qiu W. 2021. Problem of data imbalance in building energy load prediction: Concept, influence, and solution. *Applied Energy*. 297:117139.doi:10.1016/j.apenergy.2021.117139.

JADWAL PENELITIAN

Penelitian ini akan berlangsung dari awal Januari 2023 hingga akhir Juni 2023 dengan jadwal kegiatan seperti yang ditampilkan oleh tabel berikut.

Tabel 1 Jadwal kegiatan penelitian

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| No | Kegiatan | Tahun 2020 | | | | | | | | | | | | | | | | | | | | | | | |
| Januari | | | | Februari | | | | Maret | | | | April | | | | Mei | | | | Juni | | | |
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 1 | Komunikasi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | Perancangan Cepat |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | Pemodelan Perancangan Cepat |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | Pembuatan Prototipe |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | Penyebaran, pengiriman, dan umpan balik |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 6 | Penulisan laporan tugas akhir |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 7 | Seminar |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 8 | Revisi |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 9 | Sidang |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |