1. Sampul
2. Judul
3. Salah satu maslalah yang paling sering dihadapi ketika seorang peneliti melakukan kegiatan pengambilan data yaitu permasalahan data yang tidak seimbang. Seperti pada dataset berikut ini, data ini merupakan sebuah dataset citra **Multispektral** pada sebuah **Lahan Sawah pada salah satu desa di kecamatan dramaga.** Permasalahan pada dataset ini yaitu memiliki label data yang tidak seimbang. Apabila kita perhatikan pada diagram disamping ini, data yang berlabel tiga menjadi mayoritas dan sangat mendominasi dari keseluruhan data sedangkan label 4 menjadi minoritas. Dampaknya adalah …..
4. Lalu bagaimana cara kita mengatasi masalah tersebut?. Salah satu pendekatan dalam menangani imbalance data adalah data level approaches, metode pendekatan ini secara langsung memanipulasi data yang ada pada dataset, demi menyeimbangkan kelas data. Ada dua teknik populer yang sering digunakan yaitu Oversampling dan undersampling, Oversampling bekerja dengan menambah data minoritas, sedangkan undersampling dengan menghapus beberapa data mayoritas.

Untuk melatih sebuah model kalsifikasi kesuburan lahan, maka diperlukan data yang sangat banyak, oleh karena itu saya menggunakan teknik oversampling untuk menambah data.

1. Ada beberapa teknik over sampling diantaranya yaitu ….. Cara kerja dari smote ini adalah membuat data sintetik disepanjang garis lurus diantara data minoritas, seperti kita lihat pada ilustrasi disini. Menurut Xie dan Zhang 2018 ….
2. Teknik oversampling yang selanjutnya adalah GAN, Gan merupakan sebuah Teknik oversampling yang bisa dibilang tergolong masih baru karena Teknik ini baru diciptakan pada tahun 2014 oleh godfellow. Cara kerja dari GAN ini yaitu….

Berbeda dengan SMOTE, Gan …..

1. Hingga pada tahun 2019…

Oleh karena itu saya tertarik untuk menerapkan CTGAN ini untuk menangani kasus Imbalance data

1. Rumusan masalah, dan tujuan penelitian
2. Manfaat
3. Pada gambar Berikut merupakan Contoh citra multispektral
4. GAN
5. Pengembangan GAN
6. Random Forrest

Random forest ini yang nantinya saya akan gunakan untuk membuat model klasifikasi dari data asli dan data palsu yang nanti akan dibandingkan keduanya

1. Confusion Matrix

Setelah nanti modelnya dilatih maka akan dilakukan evaluasi menggunakan confusion matrix

Confusion matrix merupakan salah satu metode populer untuk mengukur performa dari suatu model klasifikasi dan mengevaluasi model

Untuk mengetahui tingkat keberhasilan suatu model dapat dilakukan dengan menghitung tingkat akurasi, presisi, recall serta F1 Score-nya.

1. Deskripsi data
2. Tahapan Penelitian
   1. 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.
   2. Pada tahap ini data yang telah dibersihkan akan digunakan pada algortima CTGAN untuk menghasilkan data sintetis yang mirip dengan data aslinya.
   3. Model kalsifikasi yang akan digunakan pada penelitian ini yaitu Random Forest dengan memanfaatkan library pada python yaitu sklearn.
   4. Pada tahap ini, akan dibuat dua buah model Random Forest. 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
   5. Melakukan evaluasi dan membandingkan performa hasil pengujian antara kedua buah model Random Forest. Hasil pengujian kemudian dilakukan analisis untuk diambil kesimpulan. Pada tahapan penelitian ini menggunakan confusion matrix dalam menghitung akurasi, precision, F1 score, dan recall untuk mengukur kinerja smodel. 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.
3. Lingkungan Pengembangan
4. Jadwal Penelitian

Contents