

Deskripsi Data

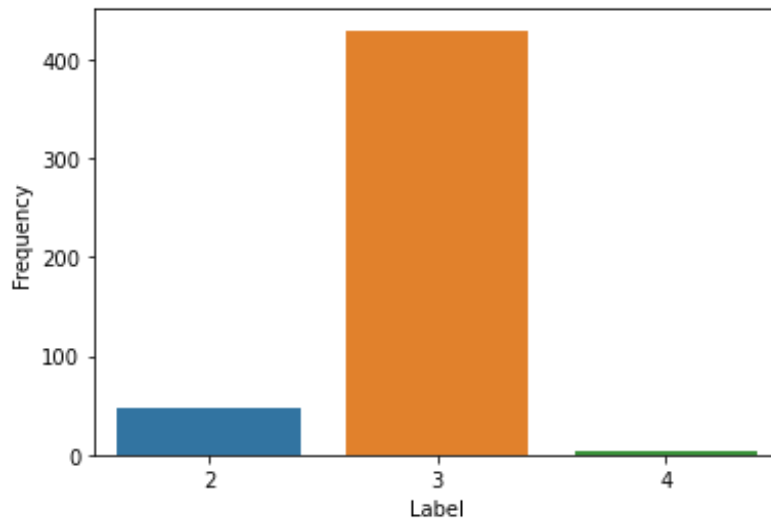
Data yang digunakan di sini adalah data citra multispectral yang diambil dari kamera UAV multispectral di atas lahan persawahan di Dramaga. Masing-masing record data menunjukkan rata-rata fitur multispectral pada suatu grid sawah berukuran 4 x 4 meter. Label data adalah level kebutuhan pupuk pada grid-grid 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.

Tujuan kegiatan ini adalah membuat model klasifikasi dari fitur citra multispektral yang diperoleh dari UAV menjadi level kebutuhan pupuk. Dengan menggunakan model klasifikasi tersebut, petani modern tidak lagi harus langsung terjun ke sawah untuk melakukan pengukuran kebutuhan pupuk secara manual menggunakan BWD. Kebutuhan pupuk dapat langsung dipetakan secara presisi menggunakan citra dari UAV multispektral.

Beberapa keuntungan kegiatan ini antara lain:

1. Proses budidaya padi menjadi lebih presisi, cost effective, dan environmental friendly karena jumlah pupuk yang digunakan dapat disesuaikan dengan kebutuhan real tanaman.
2. Proses pemetaan kebutuhan pupuk untuk budidaya padi presisi menjadi lebih efisien dari segi waktu dan biaya karena:
 - Petani tidak perlu lagi turun ke sawah dan mengecek warna daun padi satu per satu, kegiatan pemetaan kebutuhan pupuk cukup dilakukan oleh UAV multispectral yang terbang di atas lahan sawah dan mengambil citra multispectral lahan sawah (kurang lebih 15 menit, bergantung ukuran lahan).
 - Pemilik lahan tidak perlu menyewa buruh tani untuk mengukur warna daun tanaman padi satu per satu.

Plotting jumlah data per label:

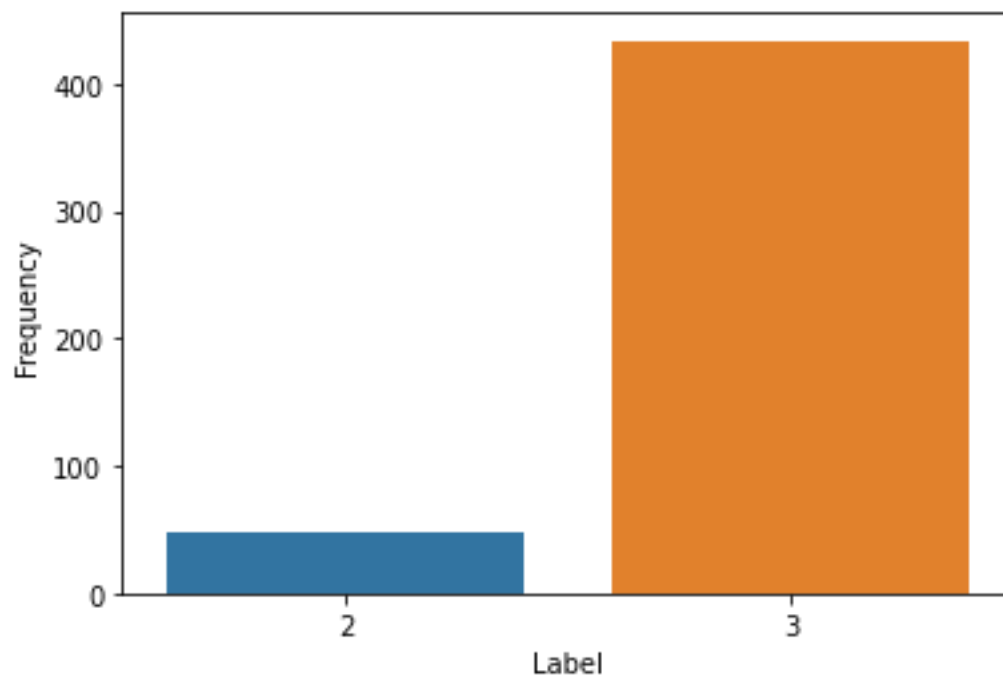


Dari hasil di atas dapat dilihat bahwa mayoritas data (430 record) diperoleh untuk label 3, sedangkan data label 2 hanya 47 record dan data label 4 hanya 4 record. Di sini, terjadi imbalance data.

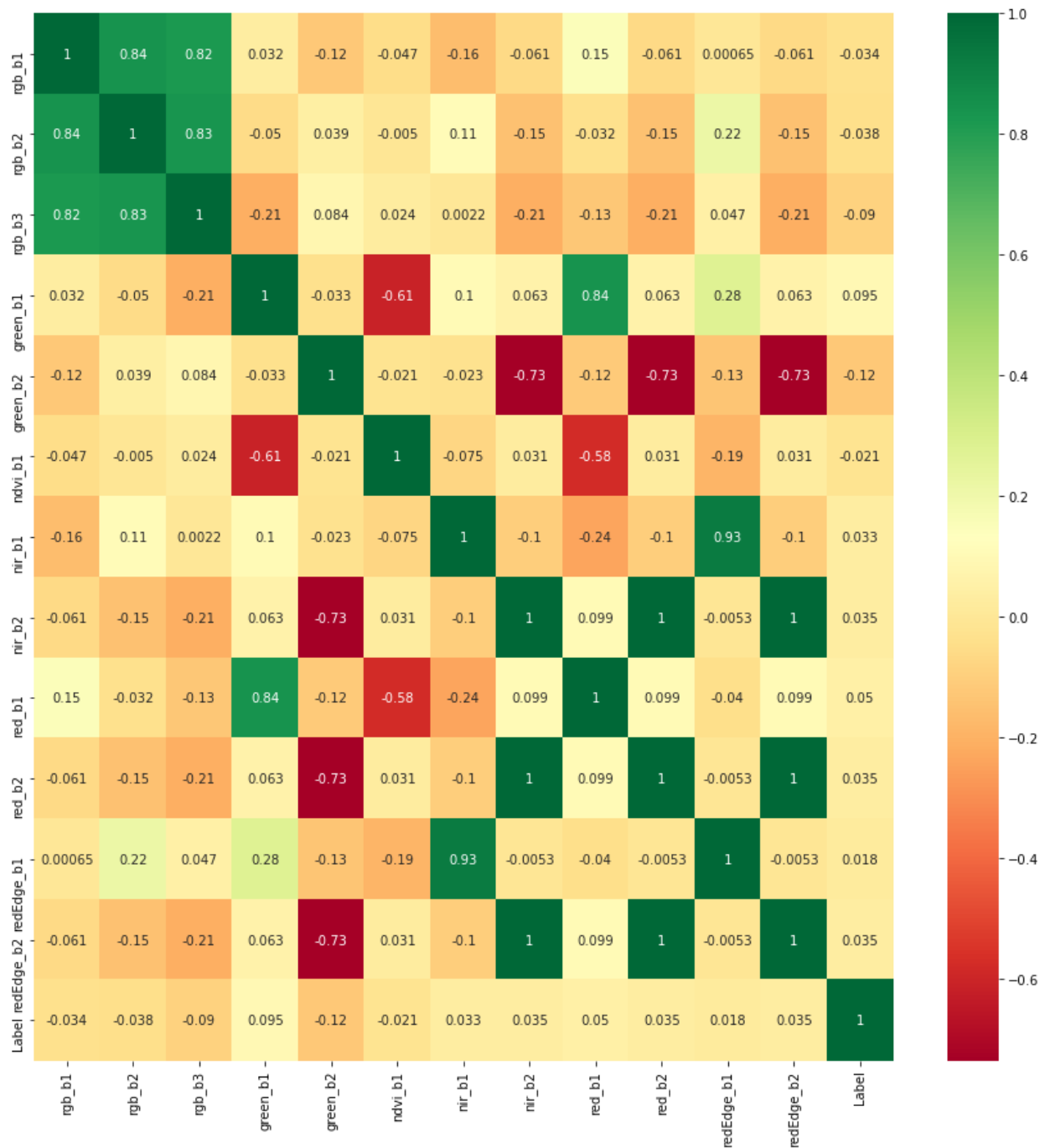
Cara mengatasi imbalance data: duplikasi data yang recordnya sedikit, atau klasifikasikan data menjadi 2 kelas saja. Di sini, label 4 dianggap sama dengan label 3, di mana label 3 dan 4 menyatakan perlu pupuk lebih sedikit (sekitar 50 ton/ha), sedangkan label 2 menyatakan perlu pupuk lebih banyak (sekitar 75 ton/ha).

Dengan cara ini, diperoleh 434 record data untuk label 3 dan 47 record data untuk label 2.

Plot jumlah data setelah data dijadikan 2 label saja (label 2 dan label 3):



Penelusuran korelasi antar fitur untuk menentukan fitur yang penting



Dari grafik heatmap di atas, terlihat bahwa fitur yang memiliki korelasi tertinggi dengan label adalah fitur green_b2 (nilai korelasi -0.12), green_b1 (nilai korelasi 0.095), kemudian rgb_b3 (nilai korelasi -0.09).

Pada penelitian ini, seluruh fitur masih digunakan (belum dilakukan seleksi fitur).

Pembagian data latih, data validasi dan data uji

Jumlah total data = 481 data yang terdiri dari 434 data label 3 dan 47 data label 2. Note bahwa data masih imbalance di sini. Pada penelitian selanjutnya, dapat dibuat agar jumlah data latin menjadi balance dengan cara mengurangi sample data untuk label 3.

Pembagian data dilakukan dengan mengambil rasio sebagai berikut:

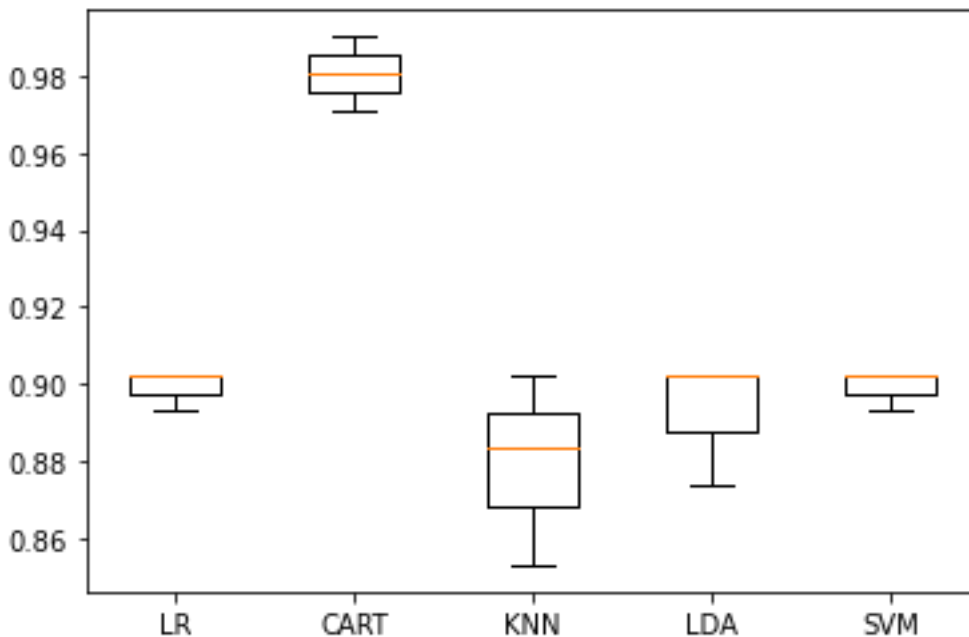
Data latih : data validasi : data uji = 64% : 16% : 20% = 307 : 77 : 97

Dalam hal ini, data validasi yang digunakan adalah 20% keseluruhan dari data latih.

Kemudian dilakukan 5-fold cross validation untuk membangun model menggunakan lima algoritme klasifikasi yang akan dibandingkan yaitu:

1. Logistic regression (LR) classifier,
2. Decision Tree (CART) classifier,
3. K-Nearest Neighbor (KNN) classifier,
4. Linear Discriminant Analysis (LDA) classifier,
5. Support Vector Machine (SVM) classifier.

Hasil akurasi pembangunan model untuk kelima classifier di atas ditunjukkan pada gambar berikut:



Dari hasil di atas, model terbaik diperoleh untuk metode klasifikasi Decision Tree dengan akurasi tertinggi 98,97%.

Penjabaran rinci mengenai kinerja masing-masing model classifier di atas adalah sebagai berikut:

LOGISTIC REGRESSION

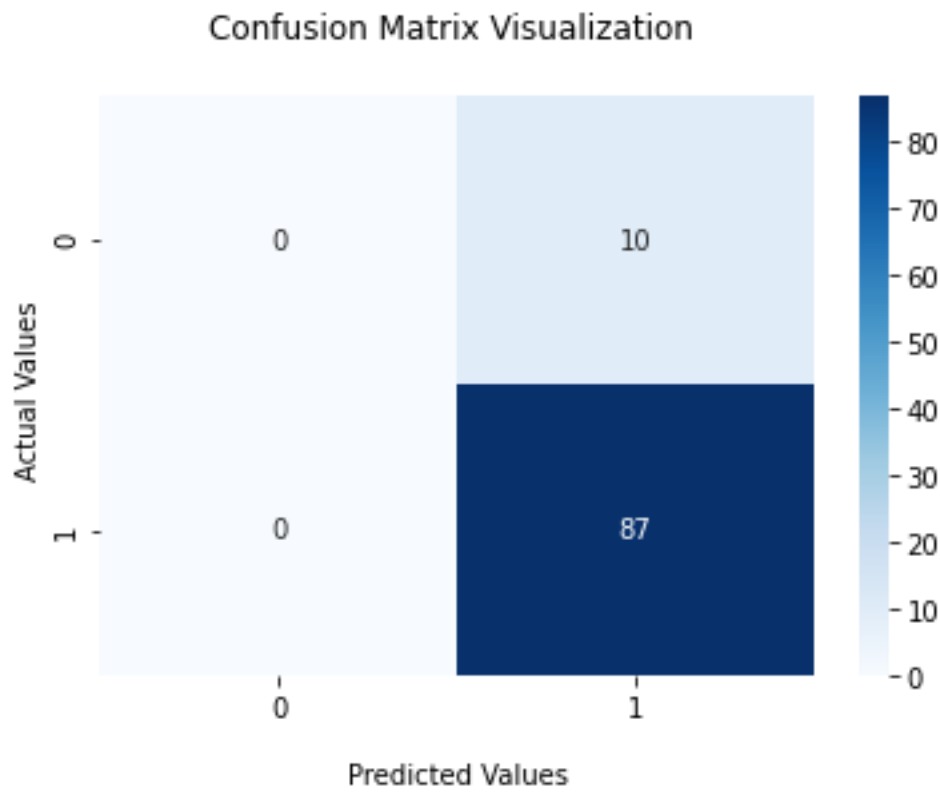
The accuracy: 0.8969072164948454

The Classification Report:

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 2 | 0.00 | 0.00 | 0.00 | 10 |
| 3 | 0.90 | 1.00 | 0.95 | 87 |
| accuracy | | | 0.90 | 97 |
| macro avg | 0.45 | 0.50 | 0.47 | 97 |
| weighted avg | 0.80 | 0.90 | 0.85 | 97 |

The Confusion Matrix:

```
[[ 0 10]
 [ 0 87]]
```



DECISION TREE

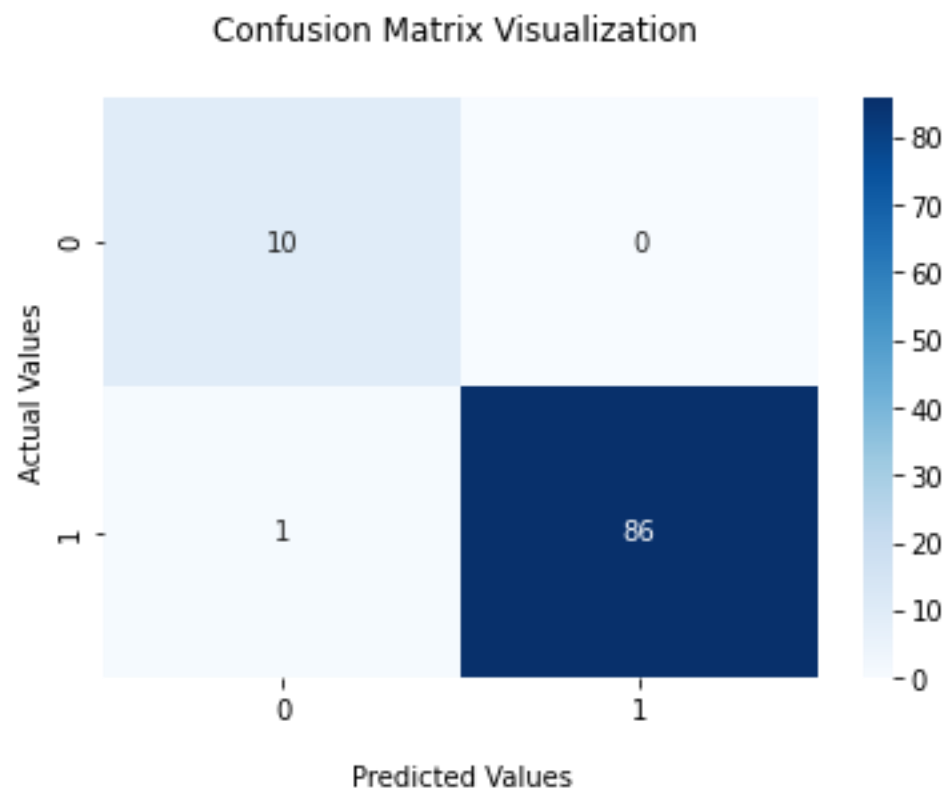
The accuracy: 0.9896907216494846

The Classification Report:

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 2 | 0.91 | 1.00 | 0.95 | 10 |
| 3 | 1.00 | 0.99 | 0.99 | 87 |
| accuracy | | | 0.99 | 97 |
| macro avg | 0.95 | 0.99 | 0.97 | 97 |
| weighted avg | 0.99 | 0.99 | 0.99 | 97 |

The Confusion Matrix:

```
[[10  0]
 [ 1 86]]
```



KNN

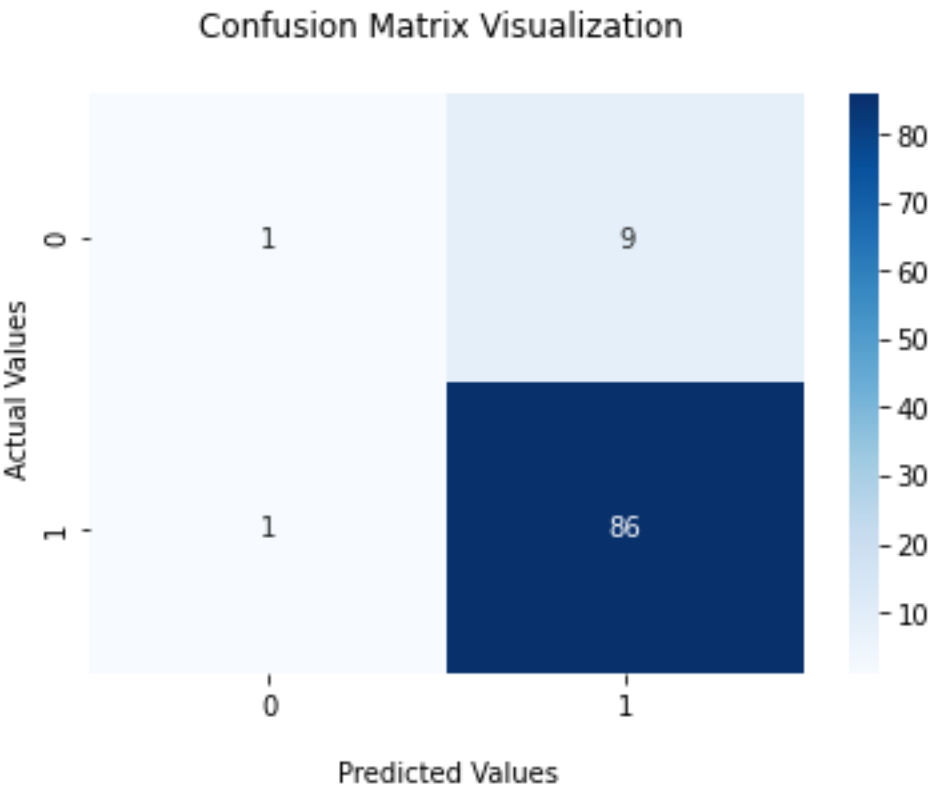
The accuracy: 0.8969072164948454

The Classification Report:

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 2 | 0.50 | 0.10 | 0.17 | 10 |
| 3 | 0.91 | 0.99 | 0.95 | 87 |
| accuracy | | | 0.90 | 97 |
| macro avg | 0.70 | 0.54 | 0.56 | 97 |
| weighted avg | 0.86 | 0.90 | 0.86 | 97 |

The Confusion Matrix:

```
[[ 1  9]
 [ 1 86]]
```



LDA

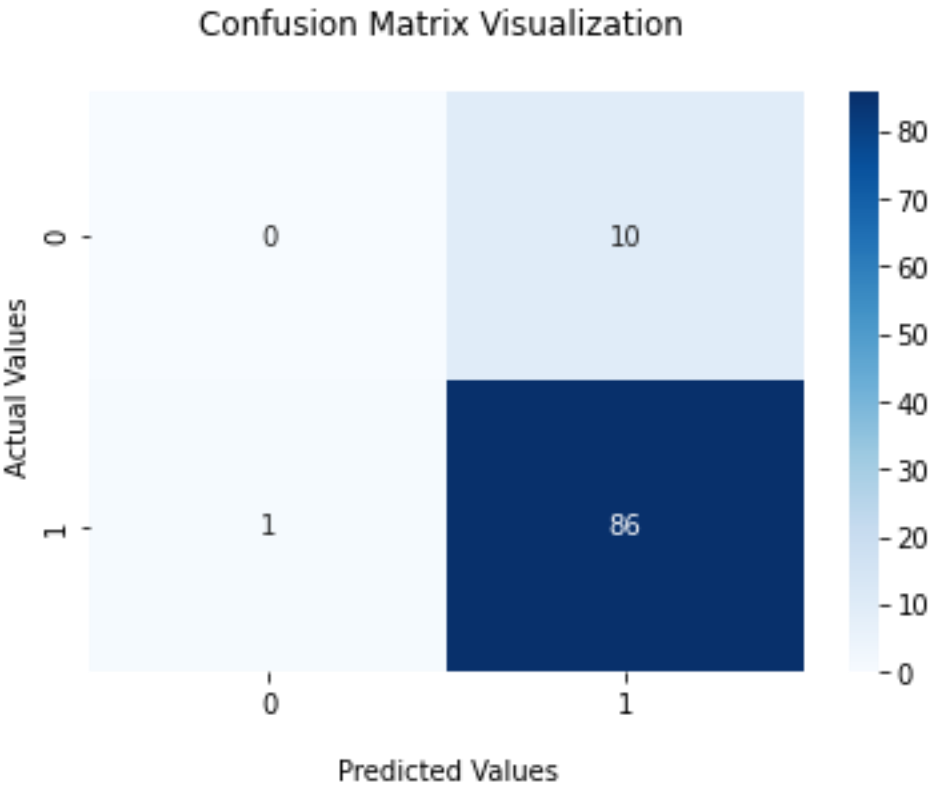
The accuracy: 0.8865979381443299

The Classification Report:

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 2 | 0.00 | 0.00 | 0.00 | 10 |
| 3 | 0.90 | 0.99 | 0.94 | 87 |
| accuracy | | | 0.89 | 97 |
| macro avg | 0.45 | 0.49 | 0.47 | 97 |
| weighted avg | 0.80 | 0.89 | 0.84 | 97 |

The Confusion Matrix:

```
[[ 0 10]
 [ 1 86]]
```



SVM

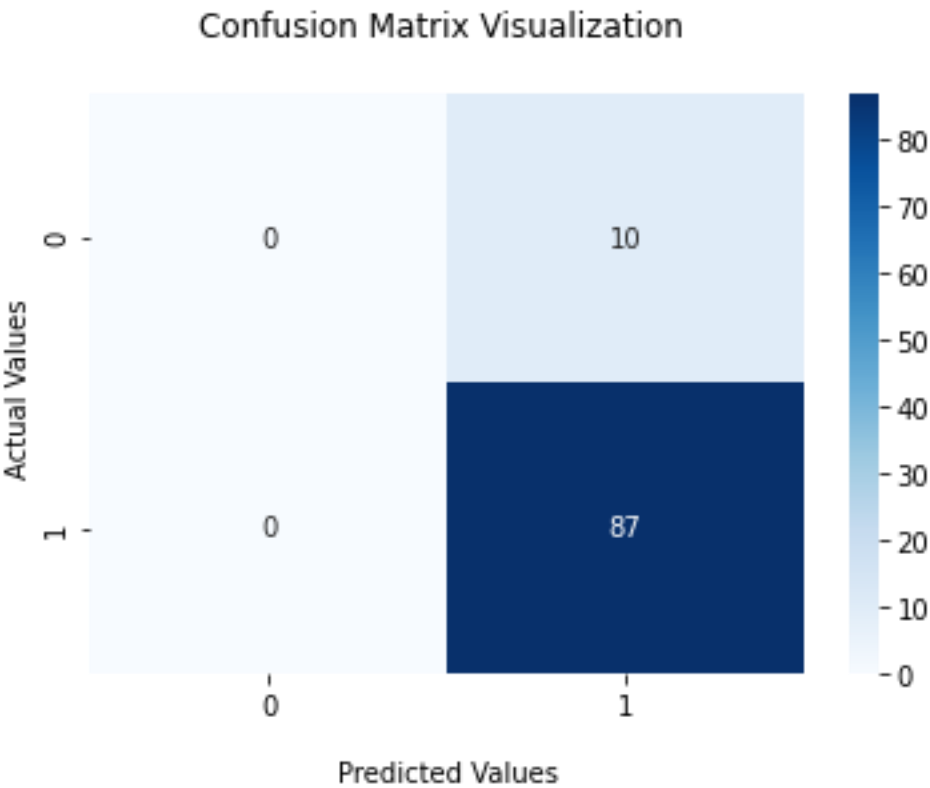
The accuracy: 0.8969072164948454

The Classification Report:

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 2 | 0.00 | 0.00 | 0.00 | 10 |
| 3 | 0.90 | 1.00 | 0.95 | 87 |
| accuracy | | | 0.90 | 97 |
| macro avg | 0.45 | 0.50 | 0.47 | 97 |
| weighted avg | 0.80 | 0.90 | 0.85 | 97 |

The Confusion Matrix:

```
[[ 0 10]
 [ 0 87]]
```



Hasil di atas menunjukkan bahwa kinerja kalsifikasi terbaik diperoleh untuk model klasifikasi Decision Tree dengan hasil akurasi sebesar 98,97%, precision sebesar , recall sebesar ... dan F1 score sebesar

Visualisasi untuk model decision tree yang dibangun adalah sebagai berikut:

