

# **Data Analytics-1**

## **Assignment-4**

### **Report and Analysis**

#### **Team-4**

**(a) Theoretical question: There are very less samples in the above dataset , How do you deal with that?**

**Cross-Validation:** Use techniques like k-fold cross-validation to assess your model's performance. This helps in getting a better estimate of how well your model will generalize to unseen data, even with a small dataset.

**(b) Plot the confusion matrix, and include the precision, recall, f1-score metrics in the report.**

#### **One-vs-One**

Confusion Matrix:

```
[[23  0  0]
 [ 1 13  0]
 [ 0  0 18]]
```

Classification Report:

	precision	recall	f1-score	support
Adelie Penguin (Pygoscelis adeliae)	0.96	1.00	0.98	23
Gentoo penguin (Pygoscelis papua)	1.00	0.93	0.96	14
Chinstrap penguin (Pygoscelis antarctica)	1.00	1.00	1.00	18

#### **One-vs-All**

Confusion Matrix:

```
[[27  0  0]
 [ 1  9  0]
 [ 0  0 18]]
```

## Classification Report:

	precision	recall	f1-score	support
Adelie Penguin ( <i>Pygoscelis adeliae</i> )	0.96	1.00	0.98	27
Gentoo penguin ( <i>Pygoscelis papua</i> )	1.00	0.90	0.95	10
Chinstrap penguin ( <i>Pygoscelis antarctica</i> )	1.00	1.00	1.00	18

## Random Forest Classifier:

### Confusion Matrix:

```
[[27  0  0]
 [ 0 10  0]
 [ 0  0 18]]
```

## Classification Report:

	precision	recall	f1-score	support
Adelie Penguin ( <i>Pygoscelis adeliae</i> )	1.00	1.00	1.00	27
Chinstrap penguin ( <i>Pygoscelis antarctica</i> )	1.00	1.00	1.00	10
Gentoo penguin ( <i>Pygoscelis papua</i> )	1.00	1.00	1.00	18

## (c) Compare the results obtained for one-vs-one and one-vs-all (which according to you performs better for the above dataset)

Certainly, let's compare the results between the One-vs-One (OvO) and One-vs-All (OvA) classification approaches:

### One-vs-One (OvO):

- Adelie Penguin (*Pygoscelis adeliae*): Precision = 0.96, Recall = 1.00, F1-score = 0.98, Support = 23
- Gentoo Penguin (*Pygoscelis papua*): Precision = 1.00, Recall = 0.93, F1-score = 0.96, Support = 14
- Chinstrap Penguin (*Pygoscelis antarctica*): Precision = 1.00, Recall = 1.00, F1-score = 1.00, Support = 18

### One-vs-All (OvA):

- Adelie Penguin (*Pygoscelis adeliae*): Precision = 0.96, Recall = 1.00, F1-score = 0.98, Support = 27
- Gentoo Penguin (*Pygoscelis papua*): Precision = 1.00, Recall = 0.90, F1-score = 0.95, Support = 10
- Chinstrap Penguin (*Pygoscelis antarctica*): Precision = 1.00, Recall = 1.00, F1-score = 1.00, Support = 18

Here are some observations:

1. **Adelie Penguin (*Pygoscelis adeliae*):** Both OvO and OvA approaches show similar results for this class, with high precision, recall, and F1-score. In both cases, it correctly identifies Adelie Penguins.
2. **Gentoo Penguin (*Pygoscelis papua*):** In OvO, it has a slightly higher recall (0.93) compared to OvA (0.90), which means that in OvO, it correctly identifies more Gentoo Penguins. However, the precision is the same in both cases, indicating that when it predicts Gentoo Penguin, it is often correct.
3. **Chinstrap Penguin (*Pygoscelis antarctica*):** Both OvO and OvA correctly identify Chinstrap Penguins with high precision, recall, and F1-score. There is no significant difference between the two approaches for this class.

Overall, both OvO and OvA approaches perform well in classifying penguin species.