

## Task 1 Documentation

SC\_H10 Team

NAME	ID
عبدالرحمن ياسر حلمي المحمدي	20191700348
احمد مجدي عبدالرحمن محمد	20191700056
ساره محمد عثمان سيف النصر	20191700279
احمد محمد سامي محمد	20191700065
عبدالرحمن محمد حسن احمد	20191700371

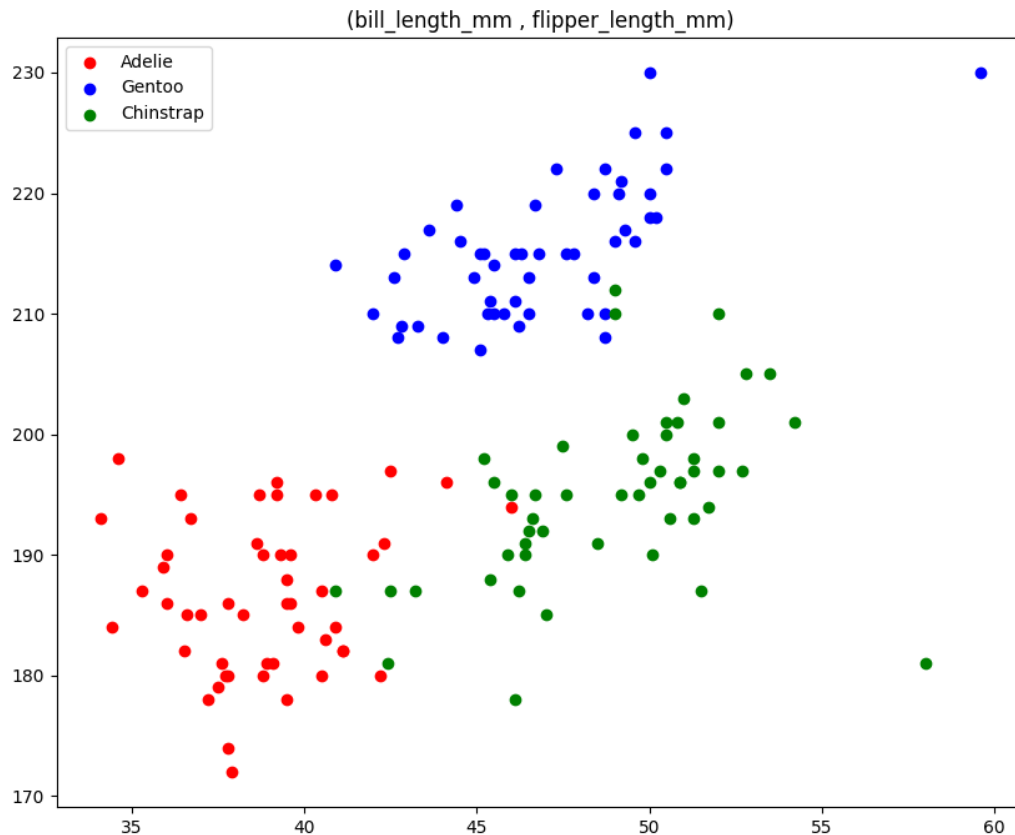
### AGENDA:

- Combinations Figures with analysis
- Preprocessing Technique
- Features that achieve higher accuracy

# Combination Figures with Analysis

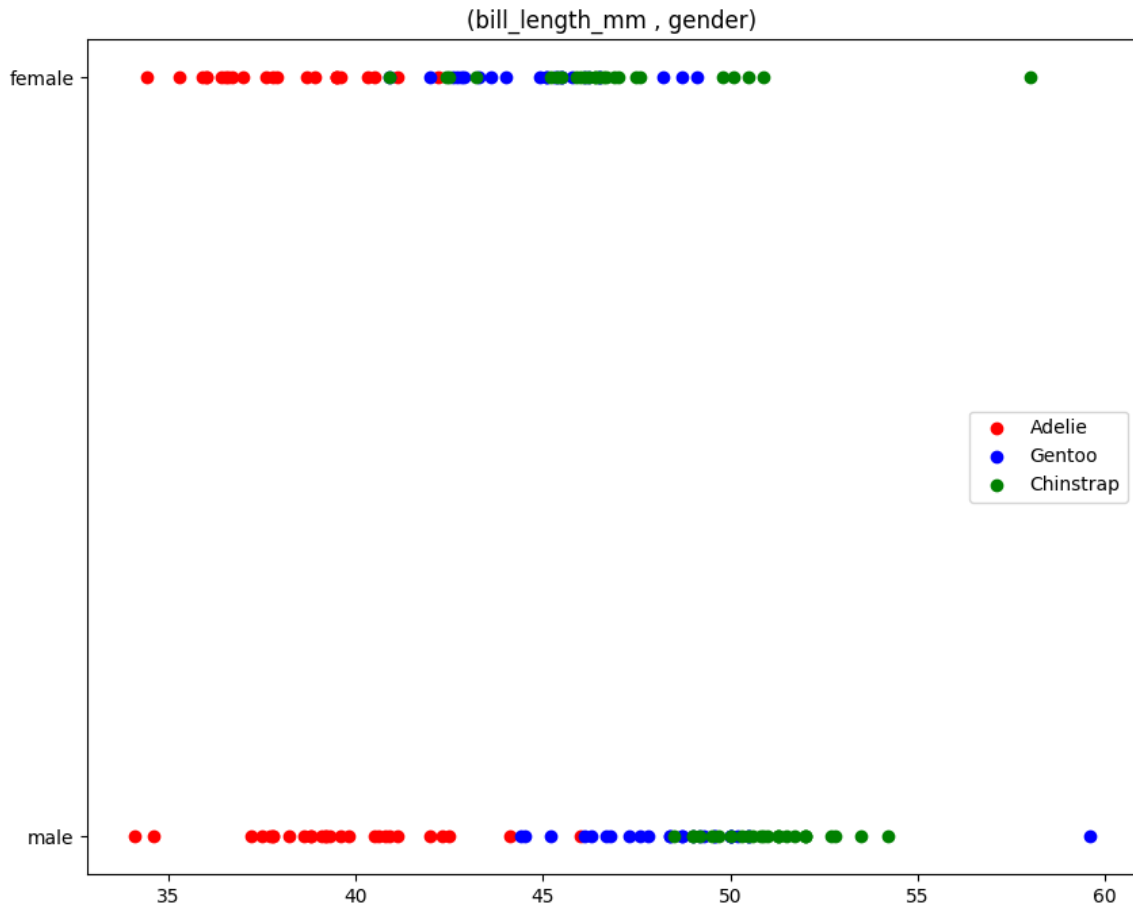
- Flipper Length and bill Length

---



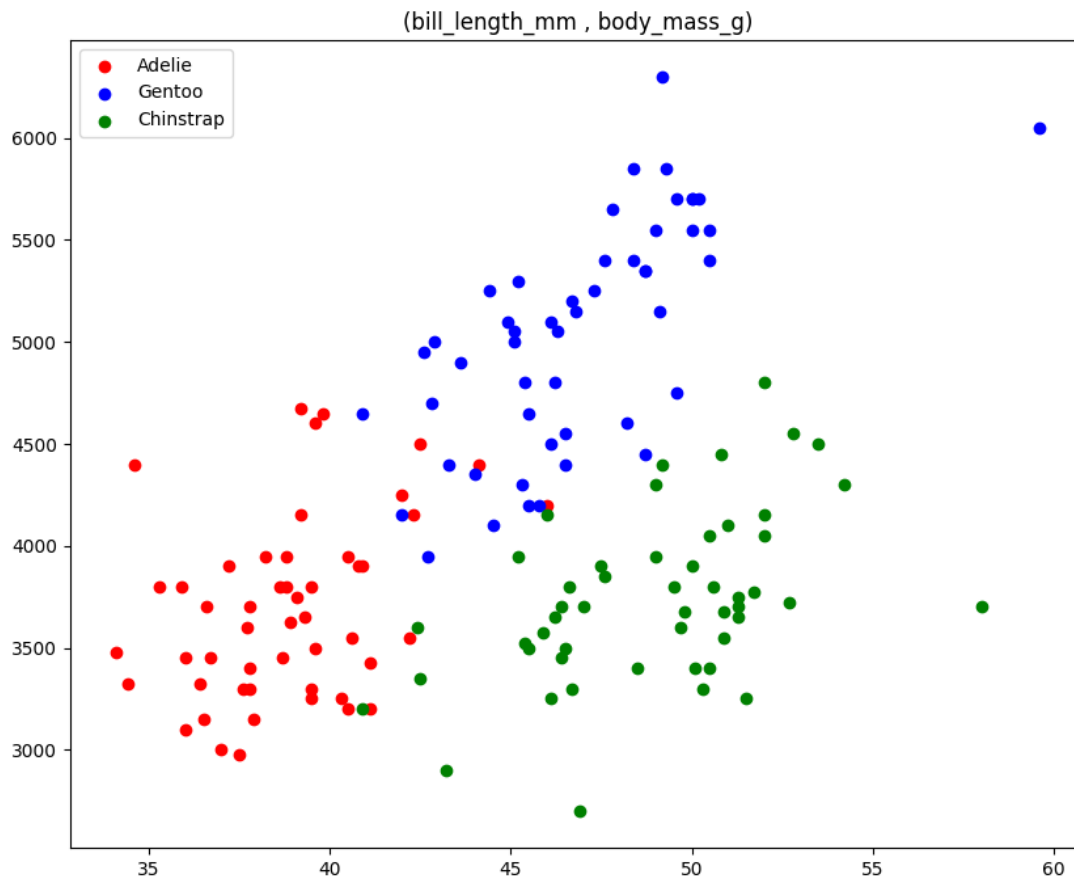
From this figure we can conclude that this two features will be good to classify between penguins types , because it's clear between Adelie and Gentoo classes there's a line can split data and achieve a great accuracy [horizontal line from 200 on y-axis] it may get 100% as accuracy for the model , but it will be not a good choice if we want to classify between chinstrap and any other class , because of the overlapping that has been shown in the figure , for that this combination will be good only to classify between Gentoo and Adelie.

- bill length and Gender:



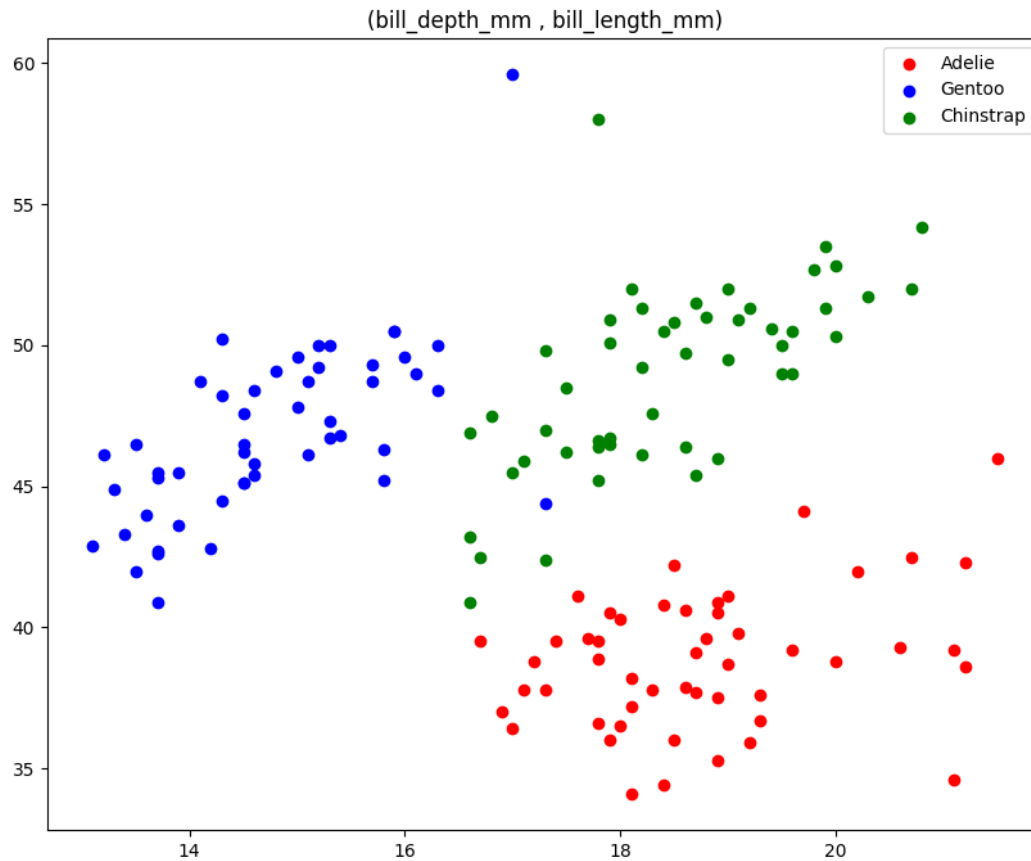
from this figure we can conclude that these two features is not the best way to classify between any types of penguins that the data cannot divided using a single line and achieve a good accuracy and there's an overlap between data and each other that cannot divided well , but also we can see that Adelie class has the least overlapping then this combination , may be good to classify between Adelie and any other class , but also this combination may not achieve the best accuracy.

- bill length and body mass:



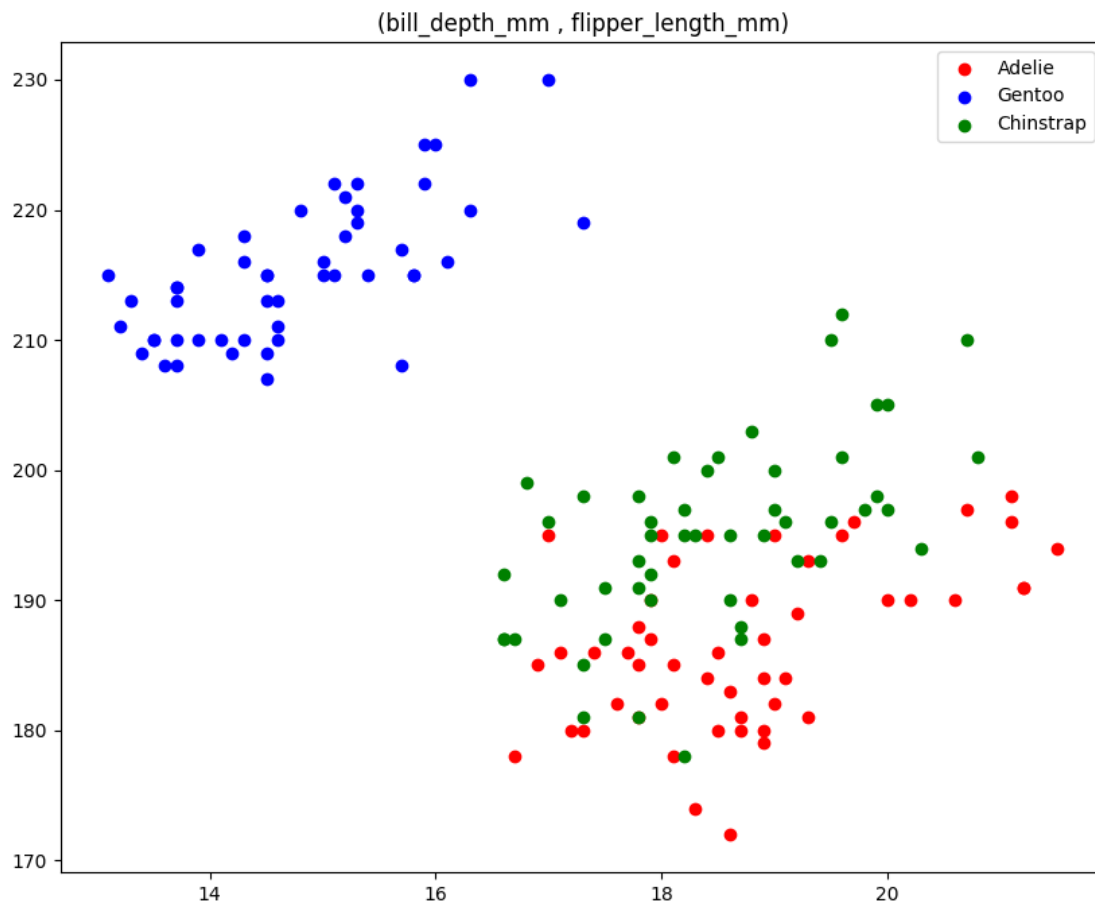
From this figure we can conclude that this combination is not a very good combination because, there's a lot of overlapping between values of three classes, which may make it hard to classify between them, it can be used to classify between gentoo and chinstrap class, but will not achieve a great accuracy.

Bill length and Bill depth:



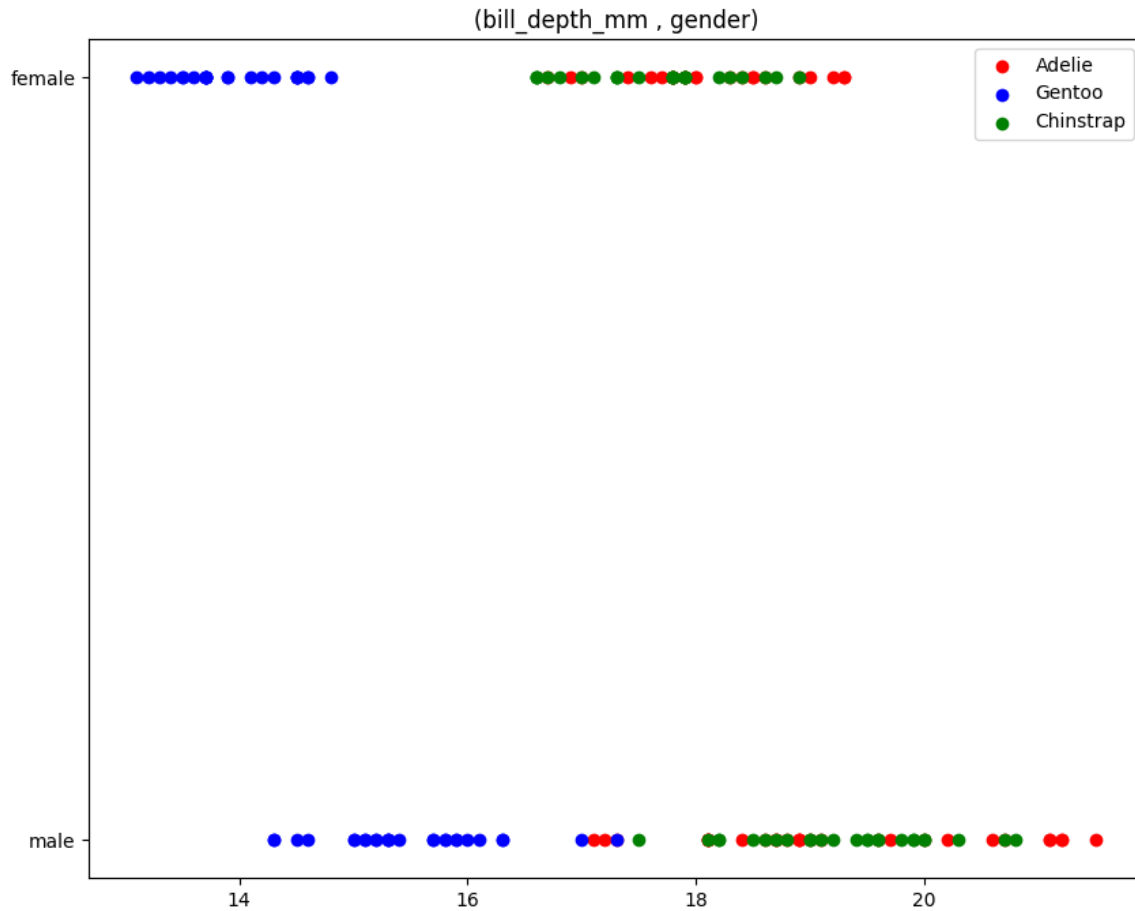
From this figure we can conclude that this combination is may be the best combination that can classify the data that the data distribution help to classify the data from this two features this combination can be used between any two classes and achieve a great accuracy for the model

## Bill depth and Flipper Length:



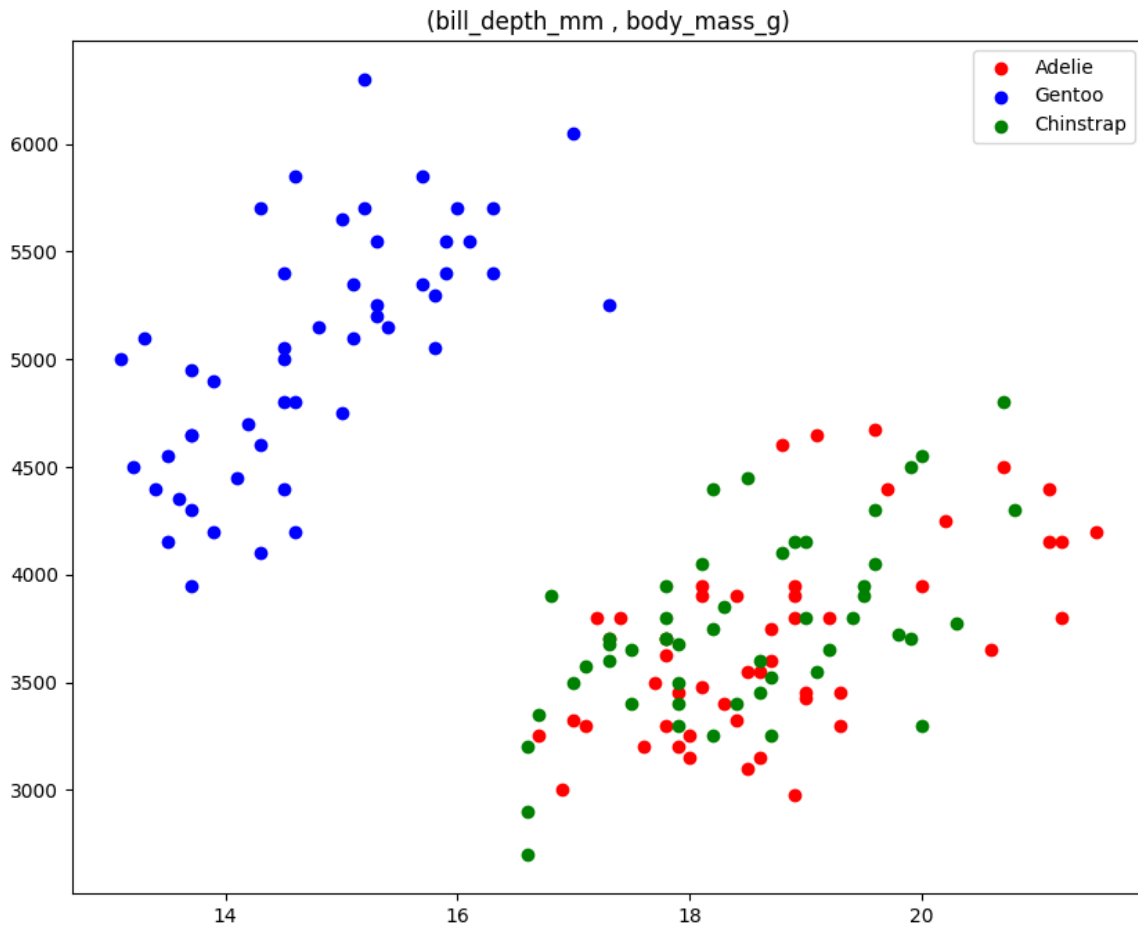
From this figure we can conclude that this combination is maybe the best to classify between Gentoo and any other class because the distribution of the data between Adelie and Chinstrap has a lot of overlapping that make it hard for the model to classify between these two classes

- Gender and bill depth:



From this figure we can conclude that these two feature's combination may be good to classify between Gentoo and any other class especially in female gender the distribution shown clearly there's a difference between Gentoo and [Adelie or Chinstrap] , may there's an overlap in male gender and bill depth between Gentoo and Adelie but it's a little overlap cannot affect the model with a high effect , for that this combination is good if we want to classify between Gentoo and any other class

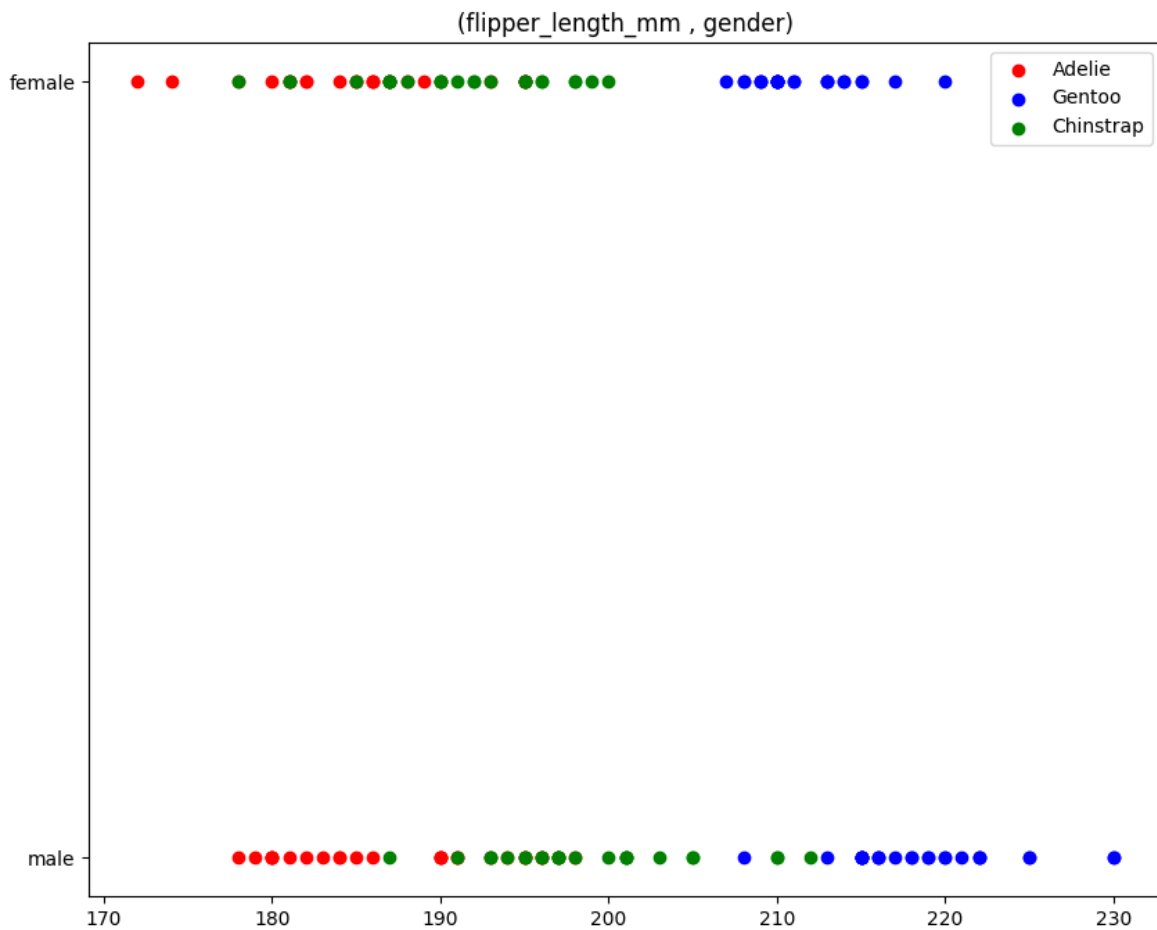
- Bill depth and Body mass:



From this figure we can conclude that this combination is a good to classify between Gentoo and any other class but it not the best choice , if we classify between Adelie and Chinstrap , because there's a lot of overlapping between these two classes in this combination which will make the model confused and cannot achieve a good accuracy to classify between these two classes.

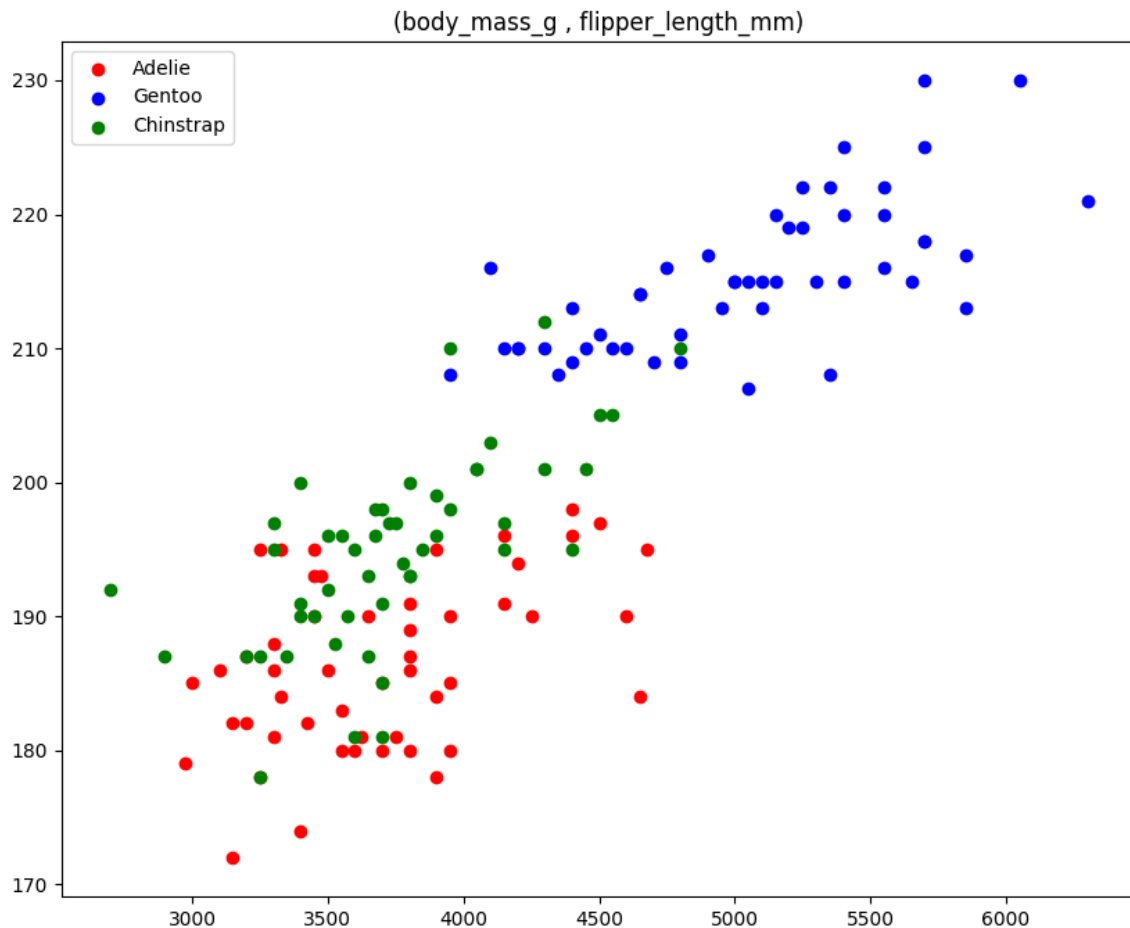


- Flipper Length and Gender:



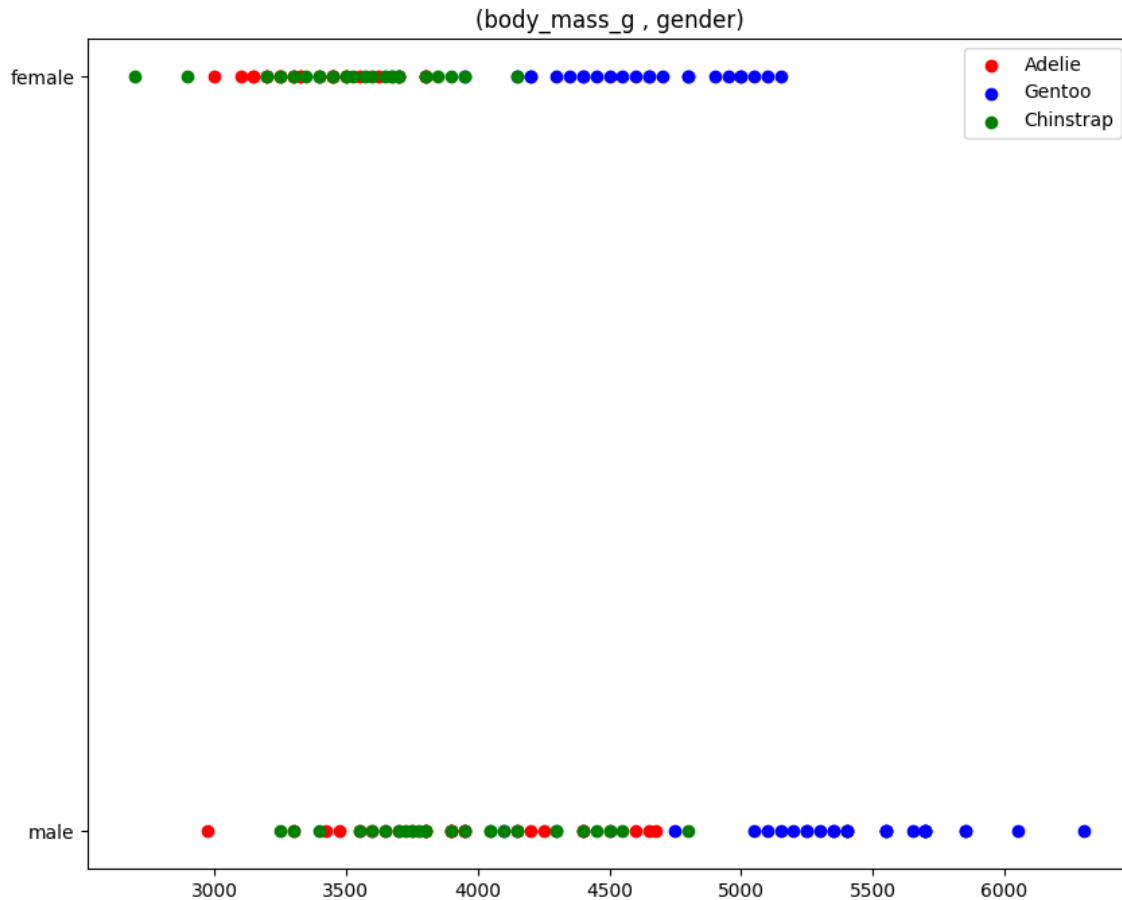
From this figure we can conclude that this combination may be a good combination to classify between Gentoo and any other class , because in female value there's some overlapping between Adelie and Chinstrap values which make this combination not the best to classify between these two classes.

- Body mass and flipper length:



From this figure we can conclude that this combination is not the best to classify between Chinstrap and any other class because the Chinstrap values distribution make a lot of overlapping between Gentoo and Adelie , which make it hard to classify between Chinstrap and any other class[Gentoo or Adelie], but this combination is a great choice to classify between Gentoo and Adelie , Since there's almost no overlapping between them on the figure , for that we can use this combination to classify between Gentoo and Adelie.

- Body mass and Gender:



From this figure we can conclude that this combination may be a good combination to classify between Gentoo and any other class , but it's not good choice if we want to classify between Adelie and Chinstrap , Since there's a lot of overlapping between the values of these two classes.

# Pre-Processing Technique

- In first we take the 50 values of the first class from the data and split it to 30 for train and 20 for testing , and apply that also for the other class
- We use train test split function to make a shuffle and split data randomly between train and test , then we combine the 30 values of class 1 and 30 of class 2 in one variable and shuffle the data again to make it randomly not on an arrange that the first 30 row for class 1 and second 30 row for class 2, we make also a combination between the 20 row of class 1 that needed in testing phase and the other 20 row of class 2 , then also we shuffle them randomly
- Then we send the data to the model to train as a numpy array after take only the two features selected from the GUI
- After that we test our model using predict function send to it the 40 rows
- Than We calculate the accuracy after comparing between the predicted values and the actual values
- After that we make a plot for the decision boundary using this equation

$$Y = - (b - w_1 * X) / w_2$$

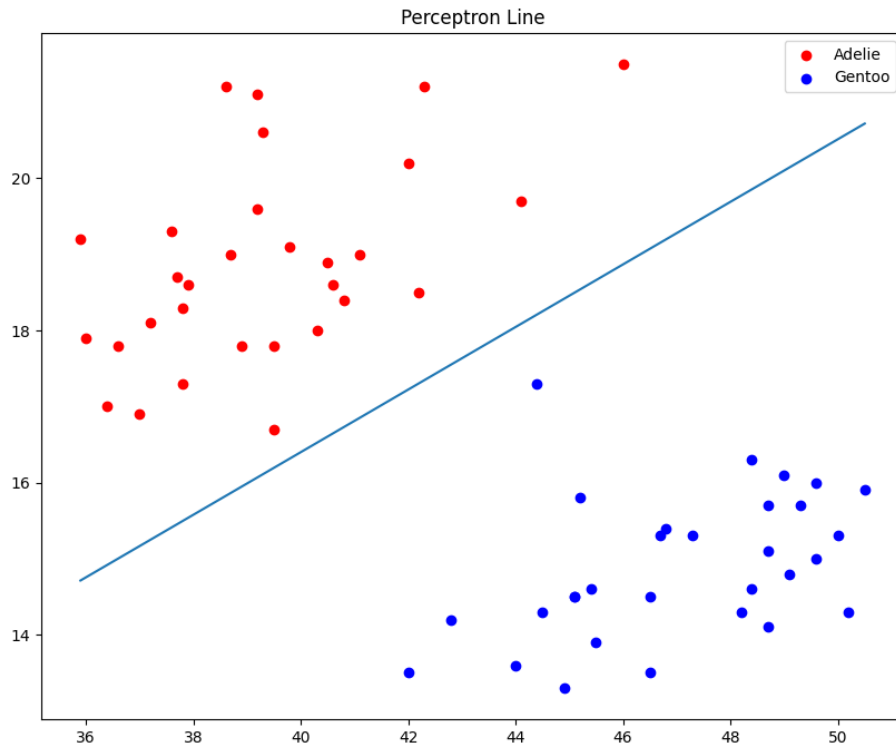
That y = feature 2 value , and x = feature 1 value

We send for this equation the minimum value of feature 1 column and the maximum value to get a two point (Xmin , Ymin) , (Xmax , Ymax) to draw the decision boundary line.

# Features that achieve higher accuracy

Bill length and bill depth combination:

- Adelie and Gentoo



accuracy : 100.0%

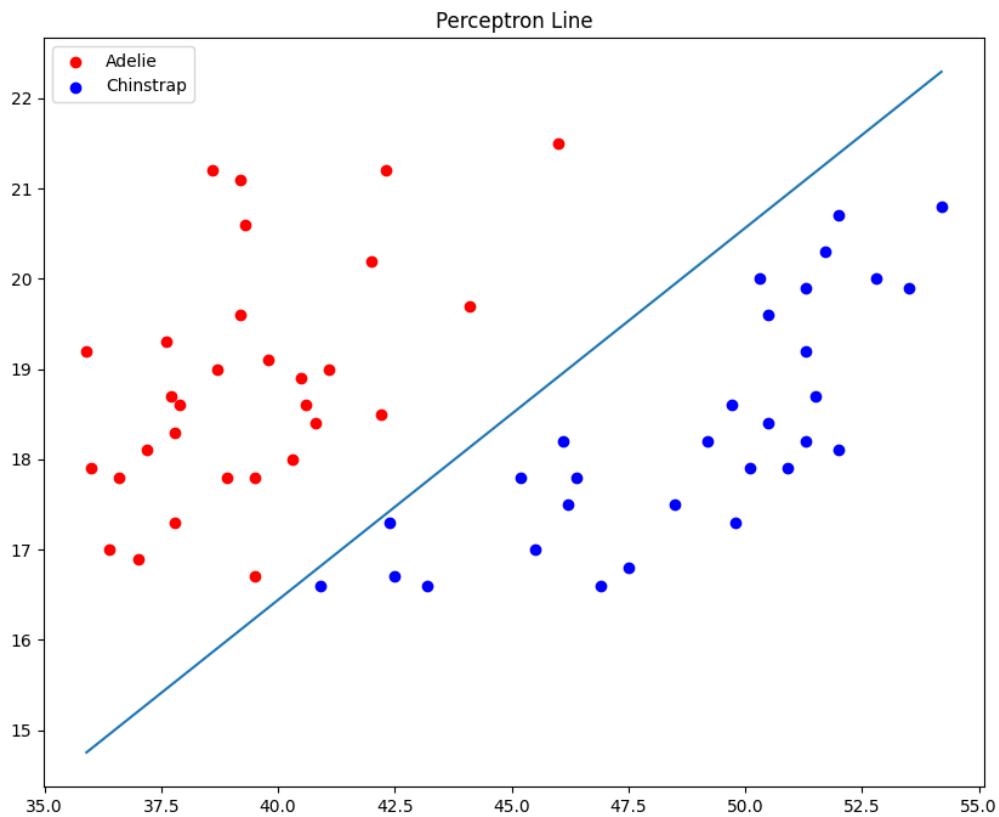
	Actual Values	
	positive	negative
Predicted [+ve]	20	0
Predicted [-ve]	0	20

As we mentioned before this combination will be good to classify between Gentoo and Adelie , with or without bias, we achieve an 100% accuracy for the model.

Learning rate = 0.01

Epochs = 100

- Adelie and Chinstrap



accuracy : 97.5%

	Actual Values	
	positive	negative
Predicted [+ve]	19	1
Predicted [-ve]	0	20

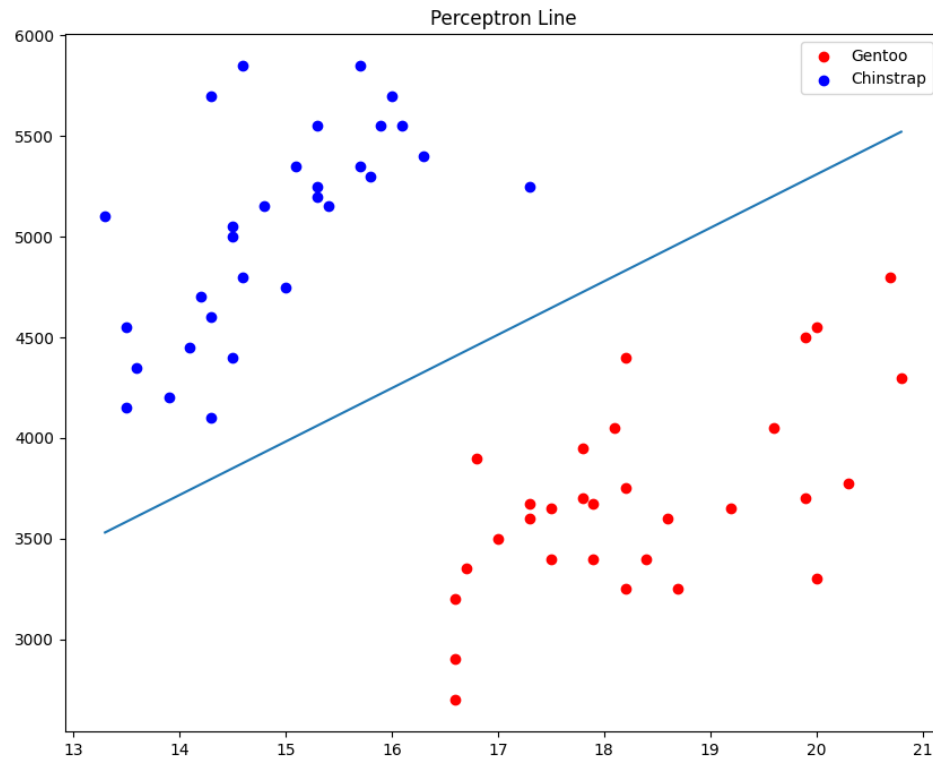
As we also mentioned before this combination will also be good to classify between Adelie and Chinstrap , while using it , it gives 97.5% accuracy [with or without bias], sometimes it gives 100% accuracy depending on the values that has been taken in the training phase [as mentioned before data is entered randomly].

Learning rate = 0.01

Epochs = 100

## Bill depth and body mass combination:

- Gentoo and Chinstrap



accuracy : 100.0%

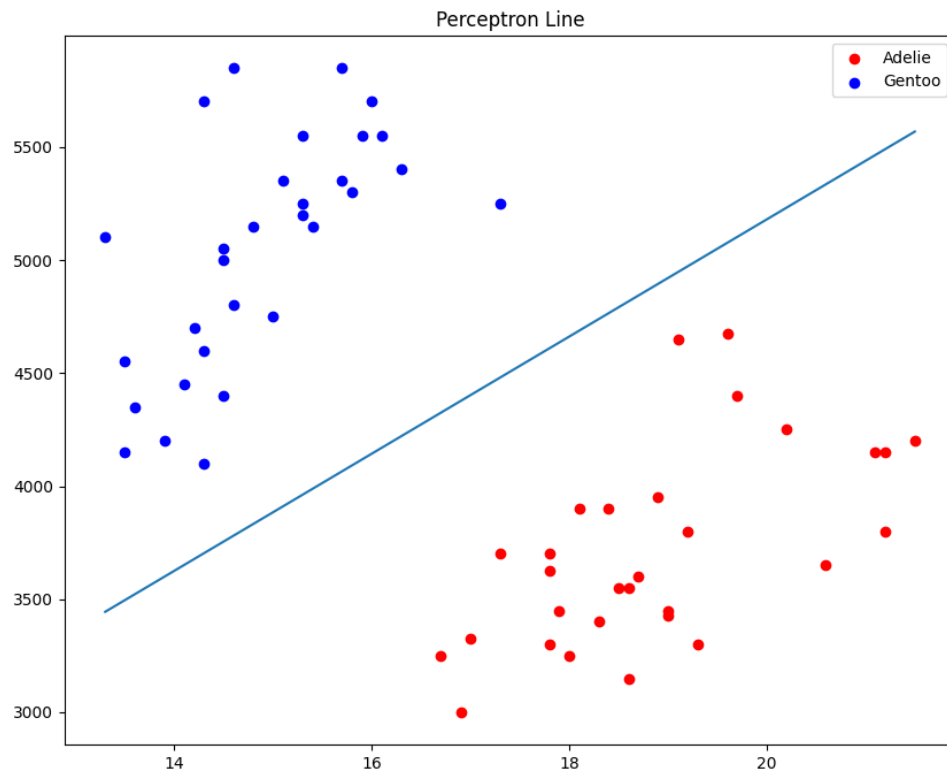
	Actual Values	
	positive	negative
Predicted [+ve]	20	0
Predicted [-ve]	0	20

As we mentioned before this combination will be great to classify between Gentoo or any other class , for that we used it to classify between Gentoo and Chinstrap , and it gives an 100% accuracy.

Learning rate = 0.01

Epochs = 200 or [100 with using for bias.]

- Gentoo and Adelie



```
accuracy : 100.0%
```

	Actual Values	
	positive	negative
Predicted [+ve]	20	0
Predicted [-ve]	0	20

As we mentioned before this combination will be great to classify between Gentoo or any other class , for that it can be used to classify between Gentoo and Adelie , and it gives an 100% accuracy.

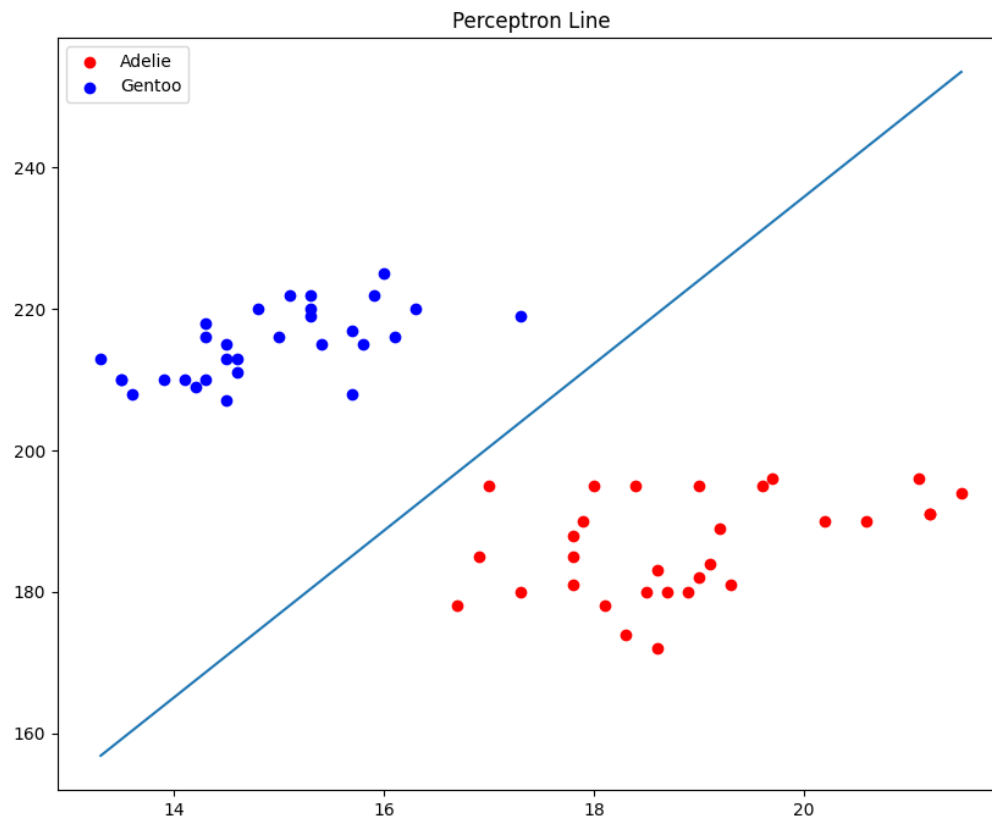
Learning rate = 0.01

Epochs = 200 or [100 with using for bias.]



## Flipper length and bill Depth

- Adelie and Gentoo



```
accuracy : 100.0%

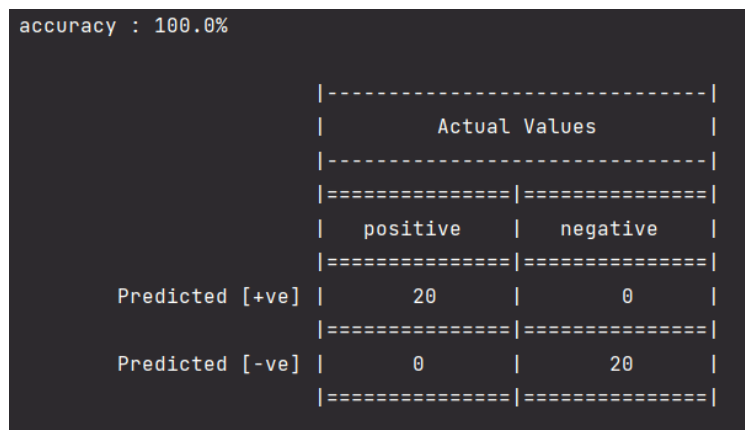
|-----|
|      Actual Values      |
|-----|
|=====|=====|
| positive | negative |
|=====|=====|
Predicted [+ve] |      20      |      0      |
|=====|=====|
Predicted [-ve] |      0       |      20     |
|=====|=====|
```

As we mentioned before this combination will be great to classify between Gentoo or any other class , for that it can be used to classify between Gentoo and Adelie , and it gives an 100% accuracy.

Learning rate = 0.01

Epochs = 100 with or without using for bias

- ## With or without using for bias



Epochs = 100