**Question 1**

The two methods implemented in Question 1 are micro-averaging and macro-averaging. For micro-averaging, precision and recall are calculated by making the sum of TP, FP, FN of each class and compute the formation.

For macro-averaging, we calculate precision and recall for each class and take the mean.

F-score is calculated by taking the value of beta as 1.

The difference between micro and macro averaging is that micro-averaging weights each sample equally while macro-averaging weights each class equally.

|  |  |  |
| --- | --- | --- |
|  | Micro-averaging | Macro-averaging |
| Precision | 0.75 | 0.719 |
| Recall | 0.75 | 0.736 |
| F1-score | 0.75 | 0.727 |

According to the table above, micro-averaging gives better results than micro-averaging. The reason behind this is because ‘Mountain’ has larger proportion among other classes in the dataset and its good evaluation averages up the score in micro-averaging. Similarly, classes other than ‘Mountain’ have relatively smaller proportion in the dataset and their results have averaged down the overall result in micro-averaging.

The micro and macro averaging have similar performance is because FN and FP have both low proportion in every class.

In this assignment, macro-averaging will be preferred over micro-averaging because the dataset is slightly imbalanced as ‘Mountain’ has larger proportion in the dataset.

**Question 2**

The assumption that the numeric data comes from a Gaussian distribution is not always true in this dataset. There are a bunch of data in training set having values that skew the data distribution in some features for each class. We group every feature of each class, discard missing values and plot the histograms. In figure 1 (x7 in bridge), the data is skewed by the value round -1400 which is explicitly an outlier. There are points that beyond the endpoint of the pose and those values would skew the distribution and make the class more probably to be selected to predict the test set. In figure 2 (x11, downwarddog), if we discard the right value around 300 and all the other value are under the range, the distribution is not a normal distribution as it has two peaks. The size of dataset for each feature in class is not large enough to generalise a Gaussian distribution.

Chart, line chart

Description automatically generated

Figure 2.1: Mountain

Chart

Description automatically generated

Figure 2.2 Downwarddog

Chart, scatter chart

Description automatically generatedFigure 2.3 Childs