In machine learning problems with skewed classes, accuracy is not a good metric to analyze the performance of our learning model. For example, Suppose we were tasked with the problem with cancer classification and we had a model that predicted y = 1 if a person had cancer and 0 otherwise. Since patients that have cancer is much rarer than patients that do not we have a skewed class representation. Suppose in our dataset that 99% of the patients do not have cancer. A model trained on this problem would have a very high accuracy. In fact a model that always predicted that a person did not have cancer would have have an accuracy of 99%. Accuracy clearly is not a good indicator in this scenario and thus we turn to different metrics to analyze the performance of our learning model. The metrics that we use to analyze the performance of our model will be derived from a confusion matrix. A confusion matrix is a table that is used to describe the performance of a classification model. There are four cells in a confusion matrix true positive, true negatives, false positives and false negatives. A true positive is a when a entity is predicted as positive and the actual class label is positive. A true negative is when a entity is predicted as negative and the actual class label is negative. False positive is when a class is predicted as positive when it is actually negative. A false negative is when a class is predicted as negative when it is in fact false. From these four variables we can formally four metrics to analyze our models recall, precision and F1 score.

$$precision \quad = \quad \frac{\#true\,positives}{\#no\,of\,predicted\,positives} = \frac{TP}{TP + FP}$$

For our cancer example we can define our precision as of all the patients we predicted have cancer what fractioned actually have cancer.

$$recall \quad = \quad \frac{True\,positives}{Acutal\,positives} = \frac{TP}{TP + FN}$$

For our cancer example we can define our recall as of all the patients that have cancer what fraction did we correctly detect as having cancer.
There is a precision/recall trade off, increasing one generally decreases the other. It can be difficult to evaluate the performance of multiple classifiers if we have to view both their recall and precision scores. We combine this scores to obtain an F1 score.

$$F_1 \quad = \quad \frac{2 \cdot precision \cdot recall}{precision + recall}$$

The F1 score is the harmonic mean of precision and recall. Using the standard mean would treat all values equally, The F1 score gives more weight to low

values. As a result, the classifier will only get a high $F_1$ score if both recall and precision are high.