Case Study on Confusion Matrix

Individual Project Report Submitted as partial fulfillment of requirements
for the Predictive Analytics Course in the Business School
at University of Colorado Denver

By *Shashank Gubba*
On *12/11/2022*

Contents

I. Introduction

Confusion matrix is generally a matrix of size n*n that is used to measure the performance of classification models. It is a tabular way of visualizing the performance of your prediction model. Every entry in a confusion matrix indicates the number of predictions made by the model where they are classified the classes correctly and incorrectly.

II. Confusion Matrix for Binary Classification

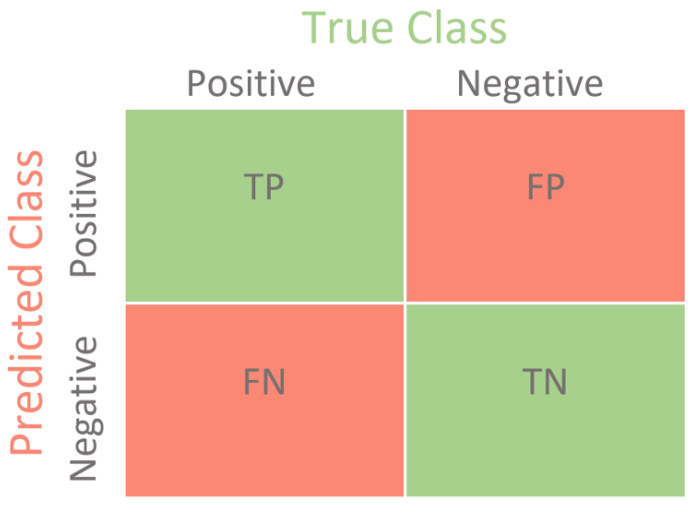
Generally, most of the confusion matrices that one sees is of the binary classification type, which has only two classes to classify, preferably a negative and a positive class. Following are the metrics of a binary classification matrix,

True Positive (TP), refers to the number of predictions where the classifier correctly predicts the positive class as positive.

True Negative (TN), refers to the number of predictions where the classifier correctly predicts the negative class as negative.

False Positive (FP), refers to number of predictions where the classifier incorrectly predicts the negative class as positive.

False Negative (FN), refers to number of predictions where the classifier incorrectly predicts the positive class as negative.



Confusion Matrix for a Binary Classification

Accuracy, gives us the accuracy of the model and is calculated using the following,

(TP+TN/TP+TN+FP+FN)

Precision, it tells us what fraction of predictions as a positive class were actually positive and is calculated using,

(TP/TP+FP)

Recall or True Positive Rate, tells us what fraction of all positive samples were correctly predicted as positive by the classifier and is calculated using,

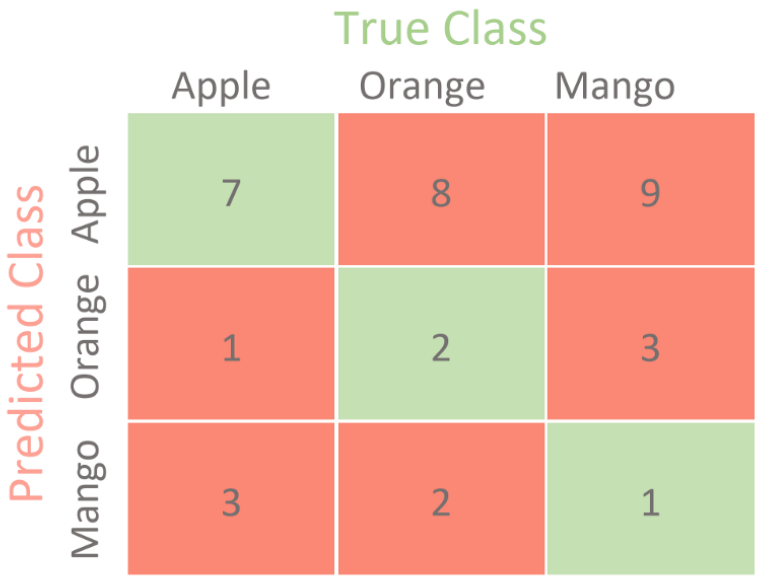
(TP/TP+FN)

F1-Score, it combines precision and recall into a single measure, which is given by,

2*(Precision*Recall/Precision+Recall) = (2TP/2TP+FP+FN)

III. Confusion Matrix for a Multi-Class Classification

Confusion Matrix for a Multi-Class Classification is used to know the performance of a prediction model, which gives us the comparison between actual and predicted values. The matrix is of the size N, where N is the number of outputs or number of classes. Let us consider our multi-class classification to be a 3-class classification problem, with a dataset that has three class labels, namely Apple, Orange and Mango. Confusion matrix is as follows:



Confusion matrix for a Multi-Class Classification

Using the above data, we can calculate TP, TN, FP, FN, Precision, Recall, Accuracy, F1-Score etc. All these factors play a key role in the prediction models. After combining all the measures, we can get the prediction of a given data. Ideally, a perfect model has precision 1, recall 1, which means F1 score of 1 which is a 100% accurate model.

Predictions can be calculated in different ways as well, by calculating Micro F1, Macro F1 and Weighted F1 scores.

IV. References

- https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826
- https://www.v7labs.com/blog/confusion-matrix-guide#h7
- https://medium.com/mlearning-ai/confusion-matrix-for-multiclass-classification-f25ed7173e66