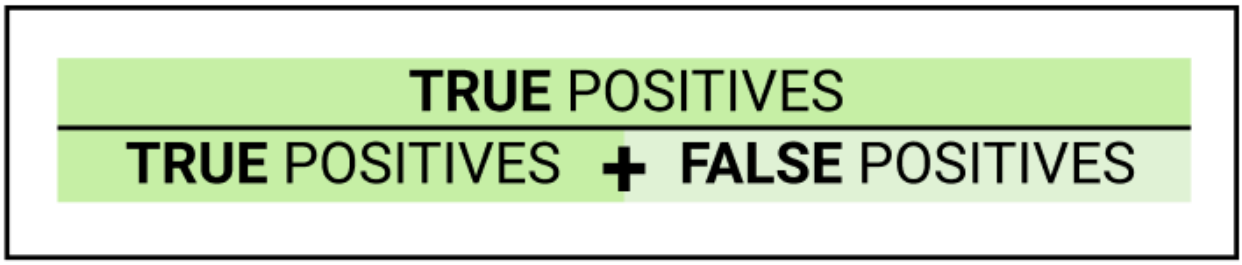
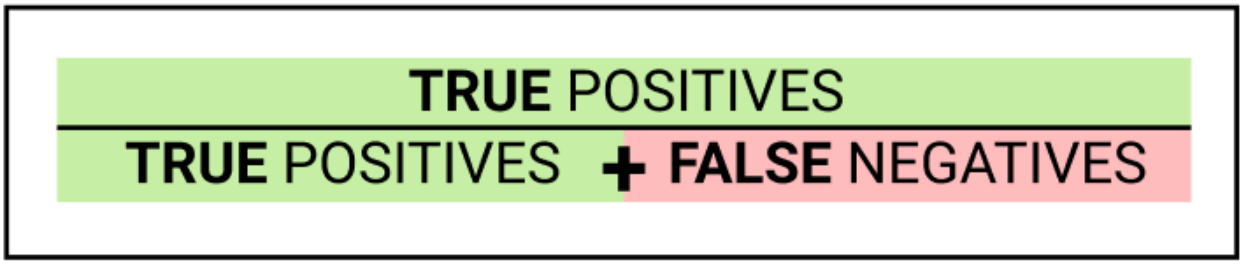
Metrics File: Twitter Sentiment Analysis

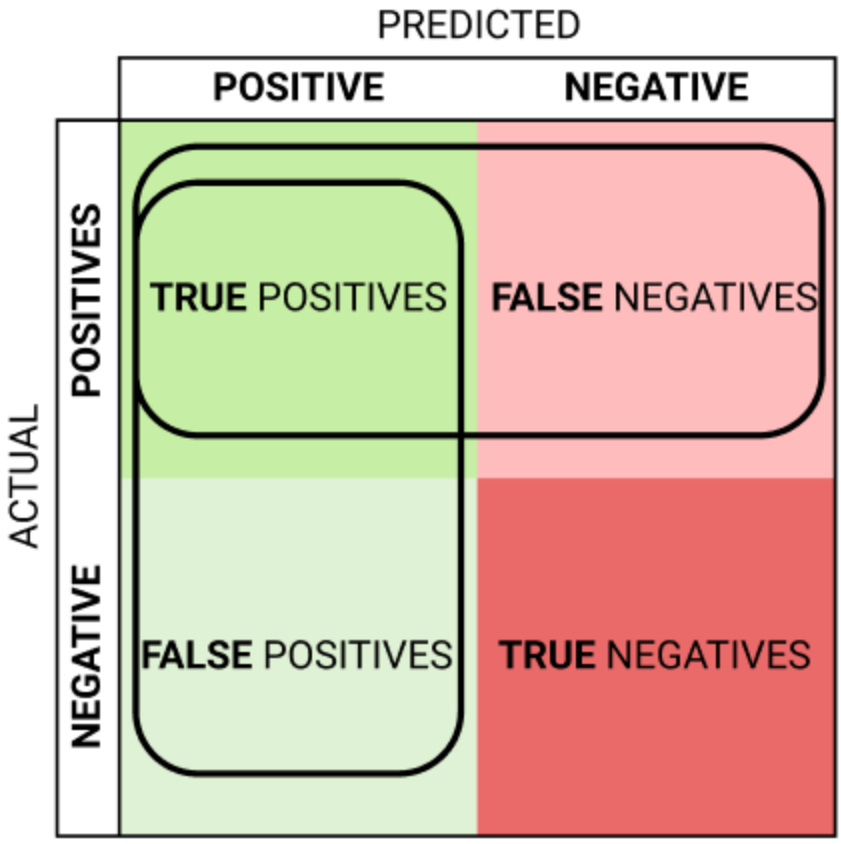
Due to Sentiment Analysis being a multilabel classification problems, the following metrics were used to evaluate the performance:

Performance Metrics

1. Accuracy: The proportion of correct predictions made by the model. It is the number of true positive and true negative predictions divided by the total number of predictions.
2. Precision: The proportion of true positive predictions among all positive predictions. It measures how many of the positive predictions made by the model are actually correct.
3.
4. 

5.
6. Recall (Sensitivity or True Positive Rate): The proportion of true positive predictions among all actual positive cases. It measures the ability of the model to identify all positive cases.
7. 

8. Confusion Matrix: A table that summarizes the true positive, true negative, false positive, and false negative predictions made by the model.



9.

Parameters/Hyperparameters

Model parameters are configuration variables that are internal to the model, and a model learns them on its own. Hyperparameters are those parameters that are explicitly defined by the user to control the learning process.

Logistic Regression

Parameter/Hyperparameter: logit(pi), solver, penalty, and regularization strength

Random Forest

Parameter/Hyperparameter: the number of trees, number of splits of each tree, number of features, bootstrap

Gradient Boost

Parameter/Hyperparameter: n_estimators , learning_rate , and max_depth or max_leaf_nodes

Naive Bayes

Parameter/Hyperparameter: prior probabilities of different classes, as well as the likelihood of different features for each class