**Step 4: Model Comparison**

For this assignment I used two different algorith which are Random Forest Classifier and Logistic Regression. We see that the Random Forest model achieved a higher accuracy score. Additionally, when we look at the confusion matrices, it is evident that we obtained higher accuracy values in the prediction section. To achieve better results, other machine learning models can also be used, or statistically, other measurements can be conducted to improve the values we obtained on our own.

Prediction metrics for Random Forest Classifier:

Accuracy predicts: 84.65%

Precision: 73.97%

Recall: 45.07%

F1\_score: 56.01%

A screenshot of a computer

Description automatically generated

Prediction metrics results for Logistic Regression:

Accuracy predicts: 92.92%

Precision: 95.59%

Recall: 70.60%

F1\_score: 81.22%

A screenshot of a computer

Description automatically generated

**F1 Score**

*F1 Score is used to measure a test’s accuracy.*

F1 Score is the Harmonic Mean between precision and recall. The range for F1 Score is [0, 1]. It tells you how precise your classifier is (how many instances it classifies correctly), as well as how robust it is (it does not miss a significant number of instances).

High precision but lower recall gives you an extremely accurate, but it then misses many instances that are difficult to classify. The greater the F1 Score, the better the performance of our model. Mathematically, it can be expressed as:



F1 Score

F1 Score tries to find the balance between precision and recall.

* **Precision:**It is the number of correct positive results divided by the number of positive results predicted by the classifier.



Precision

* **Recall:**It is the number of correct positive results divided by the number of ***all***relevant samples (all samples that should have been identified as positive).



Recall