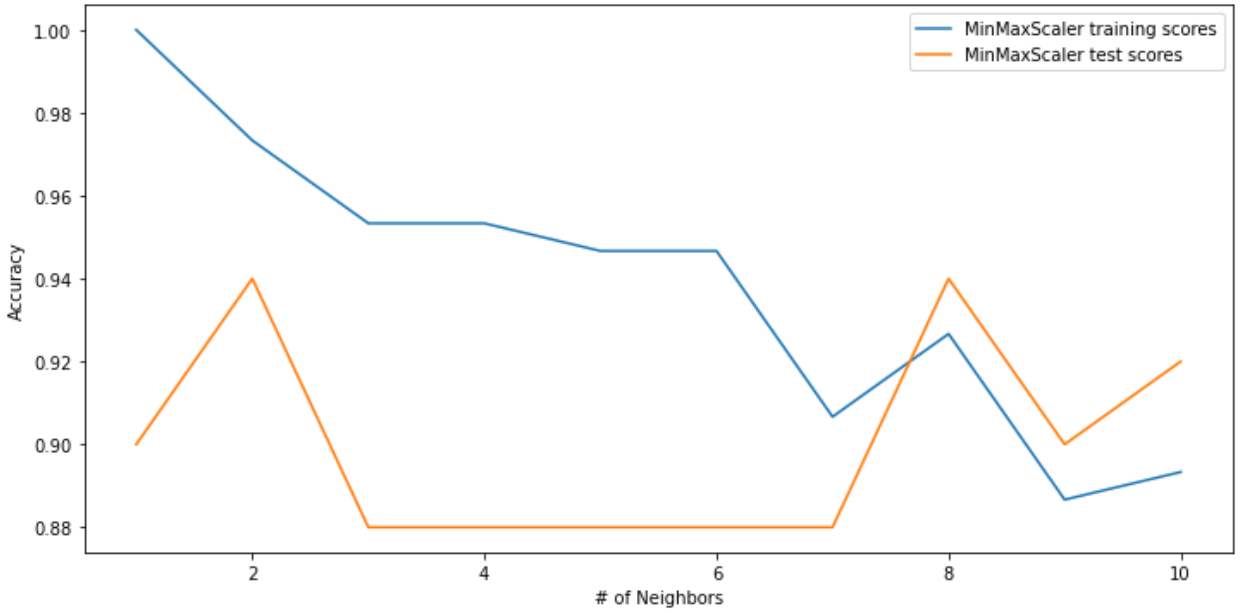
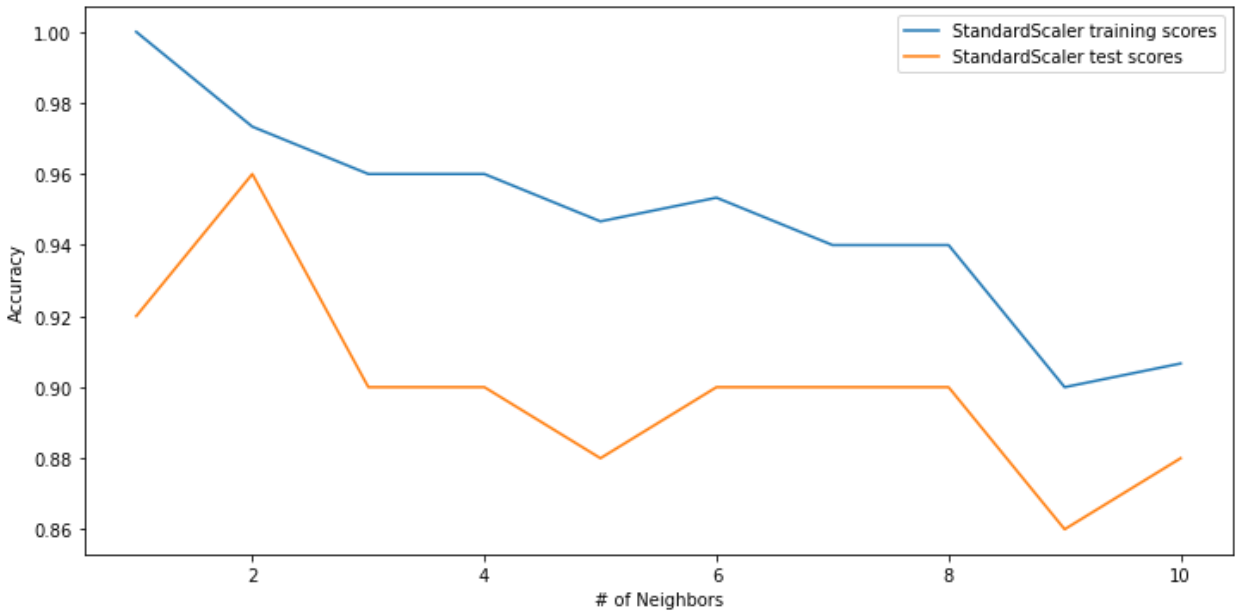
CURE PROJECT FOUR

Group Number 5

Group Members: Evan, Sayma, Karam, Dylan

(a) Graphs of model accuracy versus k values for different scenarios described in task 2:





(b) Final model parameters and scaler and accuracy:

K = 2 (for StandardScaler)

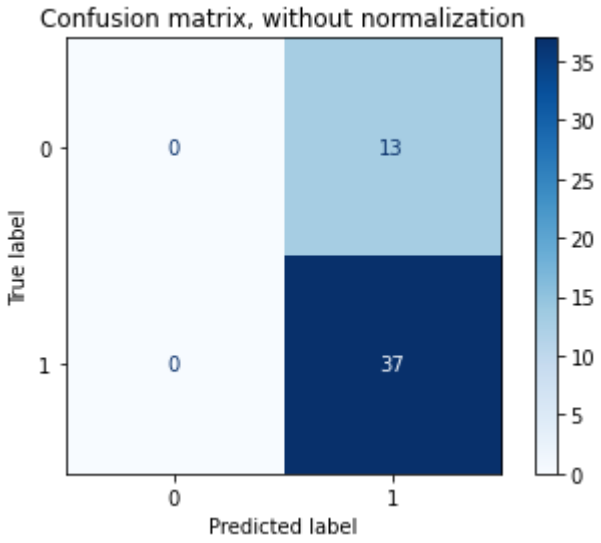
K = 8 (for MinMaxScaler)

Scaler used for attribute preprocessing: StandardScaler

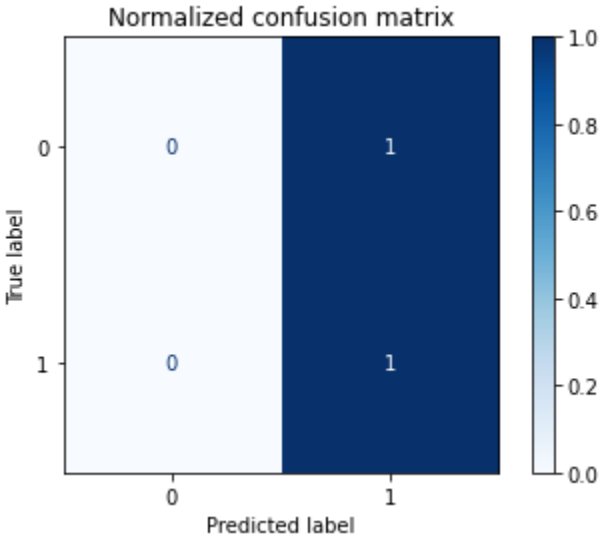
Model accuracy in the training set = about 0.97

Model accuracy in the test set = about 0.96

Confusion Matrix:



Co



Interpretations:

We can see from the first graph, that is a confused matrix without normalization that the accuracy is quite high. If we apply the formula, for this case, where accuracy = (TP) / (FP + TP) = (37) / (37 + 13) = (0.74), we may deduce that our machine learning model is pretty accurate. Recall/sensitivity, in this case, would be one since FN = 0, and TP = 1, thus giving greater reliability of the model to check correctly positive predicted outcomes out of the total number of positive outcomes. Precision would

be the same as accuracy in this case since we would only be using FP/TP for that. The F1 score for this model based on precision and recall came to 0.851, which is a fairly high score, thus further reinforcing the reliability of this machine-learning model. For the normalized confusion matrix, everything is represented as having a 0-1.00 sample. In this case, the normalized confusion matrix does a worse job of visually representing the quality of the machine learning model based on TP, FP, TN, and FN. We can deduce from the normalized confusion matrix that the flaw of our machine learning model is its tendency to predict that more samples are true when in reality they may be false, as denoted by a one for FP.

c)

New instance attribute values: (0.88, 0.92, -0.10)

Predicted class: 1