Lab 2- Hagay Ringel

My motivation to research diabetes stems from many reasons. Today there are many people who suffer from diabetes including my friends and relatives and each one has a different story. For some it started from birth and for some it came with the years. I think a lot of research in this area along with proper data analysis will lead in the future to reducing the prevalence of people with diabetes. The goal is to better understand the data and use it to reach conclusions. Today, there is a lot of information in the world about diabetes and unfortunately this information is lost or no one knows how to use it. With the help of programming and statistical tools it is possible to understand trends related to diabetes and spread these conclusions to medical professionals and the general public.

In total, there are 231 predictions. Among them, in 36 cases the classifier predicted "yes" and in 195 cases the prediction was "no". In reality, 87 cases have the disease while 144 don't have it. Here are some statistics related to this research:

Accuracy: In order to check how often the classifier is right, we need to calculate (130+22)/231= 0.66.

In 66% of the cases the classifier is correct.

Misclassification rate: how often the classifier wrong? We need to calculate (14+65)/231 = 0.34.

In 34% of the cases the classifier is wrong.

True positive rate: In order to understand how often the classifier predict yes, we have to calculate 22/87= 0.25.

In 25% of the cases the classifier will predict yes (Recall).

False positive rate: To figure out when it's actually no but predict yes, the calculation is 14/144= 0.1.

In 10% of the cases the prediction will be yes while the real answer is no.

True negative rate: when does it no and the prediction is also no? This is the calculation 130/144 = 0.90.

In 90% of the cases the prediction and real answer will be no.

Precision: when then prediction is correct and the answer is yes, the calculation is 22/36= 0.61.

In 61% of the cases the prediction is yes and it's correct.

Prevalence: How often the yes condition occur in the sample, 87/231= 0.38.

In 38% cases the condition of yes is occur in this sample.

F1 score- the weighted average of the recall and precision which is 0.36- a pretty low level of recall and precision.