2a)

The machine learning model predicts 90 as having X but the data shows only 10 actually have X and the rest are false positives. The information we know is that the victims shows no symptoms and it affects a small proportion of the population. This means to train a model to predict if someone has this disease the data set has too little information to distinguish between healthy and sick. A performance metric we can use to is classification accuracy. Here the accuracy is 10/90 is 11.11%.

Another method to measure performance metric we can use the confusion matrix. There are 10 true positives, 0 true negatives, 80 false positives and 0 false negatives. The accuracy is (TP+TN)/TOTAL=10/90=11.11%. and the misclassification rate is (FP+FN)/TOTAL=80/90=88.89%.

Models with accuracy levels of 11.11% are unsuitable for binary classification problems where accuracy greater than 95% is expected.

2b)

This is not the right approach as the features added could have very little to no relation to the desired output , provide no information gain and it might take the algorithm much longer to process the unnecessary features.. Ex- In the titanic model if we use the Name feature to model if a passenger survived or not, it could give unexpected results to different test cases. Then there is the introduction of variance error.This leads to the algorithm being highly sensitive to high degrees of variation in your training data, which can lead the model to overfit the data that is the model memorizes the data rather than generalize it. Adding more data without adding features in this case, can be helpful as increasing the data provided increases the accuracy and can help generalize the model but the data also needs to be free of outliers.