Combustion Quality Classification?

Combustion quality can be determined by measuring the 'heat signature' of a combustion chamber in addition to its 'pressure signature'. These two signals are recorded over an interval of time to determine the combustion quality. A manufacturer wants to study the feasibility of determining the quality of combustion using one signal only, i.e., either pressure or heat. The manufacturer wants to start with using pressure data to determine if combustion is 'good' or 'bad'. To this end, the manufacturer prepares a data set of pressure signals and the associated class of the signal as 0 (bad) or 1 (good). The data is collected in the file 'pressure_data.xls', where the first column indicates the combustion quality (0 or 1) and the remaining columns are the recorded pressure signals. Each row represents a data sample.

You are hired to investigate this topic using machine learning. To this end, you are required to:

1) Split the data samples in the file 'pressure_data.xls' into a training set and a testing set;
2) Use the training set to train a classifier that maps the pressure data into a state of 0 or 1;
3) Evaluate the classifier on the training set using a confusion matrix;
4) Use the classifier to classify data samples in the testing set;
5) Compare the classification results with the ground truth (from the file) and generate the confusion matrix;
6) Repeat the same using the file 'pressure_data_raw.xls';
7) Discuss methods to improve your classification accuracy.

Note that the manufacturer is interested in the results, whether positive or negative. That is, the manufacturer intends this to be a feasibility study, and the answer should help the manufacturer know whether or not pressure data alone can be used to classify combustion into these two states.