Problem Statement:

A data set is provided: raw_dataset_01.csv

About this data: this data set is real data, as it is from an experimental measurement. Two properties were collected, and the features were different adjustable processing parameters. However, we only need to know that the properties are the responses, while the features are the controllable parameters. For the first ten rows of the data, we do not have property measurements, and the prediction of those values is our objective.

From a design perspective, we want to identify which (if any) of the systems corresponding with the first ten rows have the following characteristic: relatively large value for Property1 and 'Yes' for Property2. The question : which of these systems (if any) meet this criteria? Justify your selection.

Required steps:

You are confined to the use of Linear and Multi-linear regression, Principal Component Regression, Logistic Regression, Support Vector Machine, and Support Vector Regression (some methods may only be applicable for one of the properties).

Compare these models in terms of accuracy and robustness, and when applicable in the lens of accuracy, sensitivity, and specificity, when deployed on an unseen test set.

Also indicate how generalizable each model is, and how sensitive it is to the parameters used in the regression model.

Which model(s) will you suggest to deploy? How confident are you in the predicted values? Justify.

Why do the methods you selected work better than others; that is, what in the mathematics of the methods allow it to be preferrable to the others (or conversely, what about the others make them preferrable).