RESULTS BY MODELS :

**WITH ALL THE PREDICTORS :**

Linear regression :

MSE = 0.1540146

Logistic regression :

MSE = 0. 1540146

Cross Validation =

- for K = 10 : 0.1341752

- for leave-one-out (more precise) : 0.1319878

Polynomial regression + splines : *later in the tests*

Tree-based method :

«trees generally do not have the same level of predictive accuracy as some of the other regression and classification approaches seen »

BUT we’ll use it to have a general impression and view of the situation.

Classification tree :

Prunning tree :  
«  A smaller tree with fewer splits might lead to lower variance and better interpretation at the cost of a little bias.»

Size = 9

**AFTER BEST SUBSET SELECTION :**

**Subset of predictor selected :**

*Explain and show why…*

Non-linear regression :

Polynomial regression + splines :

Tree-based method :