

Re: Choosing the right method

Source: <http://sci.tech-archive.net/Archive/sci.stat.math/2005-08/msg00074.html>

- *From:* Frank E Harrell Jr <f.harrell@xxxxxxxxxxxxxxxx>
 - *Date:* Thu, 04 Aug 2005 11:45:22 -0500
-

Eric wrote:

Dear all,

A friend of mine, who is ophthalmic surgeon needs to make a study regarding factors affecting a disease. He has got a dataset of approximately 300 samples (patients), with around 6 or 7 explanatory variables (quantitative or qualitative), and one response variable which is the occurring of the disease.

The problem is that in the dataset the disease is quite rare so that there are few cases of diseases.

What would be the best method in order to discriminate the populations.

We have been suggested : Artificial Neural Net, K nearest neighbors or Decision tree.

Thank you,

Eric

None of those methods will yield reliable predictions. You'll need something like data reduction (e.g. incomplete principal components logistic regression) or penalized maximum likelihood estimation. Above all don't attempt to find which of the variables are 'significant'

Frank

.