

TERM PROJECT PROPOSAL

TROY RAEN

1. PROBLEM OUTLINE

2. DATA ANALYSIS

	Correlations with Redshift
redshift	1.0
tu	0.4551
tg	0.1438
tr	0.2717
ti	0.3933
tz	0.4234
ty	0.4136
u10	0.3932
uerr10	0.518
g10	0.1435
gerr10	0.06651
r10	0.2716
rerr10	0.1222
i10	0.3931
ierr10	0.3822
z10	0.423
zerr10	0.4605
y10	0.4109
yerr10	0.45

???

Rongpu: what are the data columns? how to incorporate the errors am I using the correct: $NMAD = \text{median}(\text{abs}(\text{photz} - \text{specz}) / (1 + \text{specz}))$ out10 = sum(NAD \geq 0.1)/N

Jeff: general algorithms vs specific codes algorithms: neural net random forest relevance vector machine nearest neighbor random forest/NN classification?

???

To Do: create data files with colors run RF and NN for several sample sizes and plot

fitrensemble optimization: Method Bag NumLearningCycles 495 LearnRate NaN MinLeafSize 1