# Experimental Evaluation

In this section we will describe the process of building and evaluation of predictive machine learning algorithms for prediction of k.

Hyper-parameters that are optimized in cross-validation process.

## Data exploration

## Experimental setup

Since our dataset contains relatively small number of samples, special attention was given to prevention of overfitting and obtaining misleading results in terms of predictive performances.

Feature selection algorithms:

## Feature selection

PCA

NMF

Select K Best – mutual info regression

m\_features=[5, 10, 15, 20, 25, 30]

## Algorithm parameter optimization

We used 8 algorithms ANN, GBT, K-nn, Lasso, RF. SVR, LR, Ridge

**Multiple Linear Regression, Ridge and Lasso Regression:**

In case of Multiple linear regression we didn’t optimize any hyper-parameters. In case of Lasso and Ridge regression we optimized *alpha* hyper-parameter that controls the magnitude of

**'alpha'**: [0.01, 0.03, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9],

SVR:

**'C'**: [0.01, 0.03, 0.05, 0.1, 0.5, 1, 10, 100, 500, 1000],

**'kernel'**: [**'linear'**, **'rbf'**]

**RF and GBT:**

**'n\_estimators'**: range(20, 101, 20),  
name + **'\_\_'** + **'max\_features'**: n\_features\_range(n\_samples, m\_features),  
name + **'\_\_'** + **'min\_samples\_leaf'**: np.arange(0.01, 0.03, 0.05),  
name + **'\_\_'** + **'max\_depth'**: range(2, 11, 2)

## Evaluation

In order to develop predictive models with good generalization performance (predictive performance on unseen cases), we used 10 fold cross validation procedure for model evaluation.

## Results