```
#  Pre processing data


# Dealing with Class Imbalances
#

# Over fitting


# Skewness

# Multi colinearity

# Missing Values

# Noise


# Feature Engineering

# Centering and Scaling




# Strong AI
# Black box modeling
# In many fields, the choice of a model is not particularly important.
# For example, in retail credit risk (think FICO score) logistic regression, CHAID, random forest,
# gradient boosting, etc. are all pretty close in terms of separating power and stability.
# As long as the model is capable of capturing the essence of the data ("people with a lot #
# of debt who were delinquent before tend to be riskier"), it can work well. The difference is
# in the third sigfig. Enough to separate first place from tenth place in Kaggle competition, but not as important in
practice.

# DataRobot does not have priors, so cannot identify overfitting that would be obvious to a human.
# In credit scoring example above, a human would check the signs of regression coefficients. DataRobot
# can't do that because it does not know what those signs should be.
```