

Your Assignment, part 1 due by Lesson #7

- Perform data analysis, what is the data telling you
- What were your top 3 – 5 observations from initial data exploration;
- Describe and provide supporting screen shots;
- What features look most promising for modeling?
- Describe how you selected these features, screen shots to support your discussion, if appropriate;
- Do you need to deal with missing values?
- Scale the attributes?
- Make initial attempt to model and predict, set your own baseline

Your Assignment, part 2 due by Lesson #8

- Construct initial model, post on Kaggle to set your team baseline
 - Screenshot of position on leaderboard;
- Compare with different model settings, different models.
- Opportunity for feature selection, feature transformation such as binning or clustering, etc.

Describe the data cleaning, transformation steps you selected and describe the why and how. Run at least two to three new model evaluations, submit at least one new submission to Kaggle contests site

Your Assignment, part 3 due by Lesson #10

Repeat Part 2 as necessary during 3 part of assignment

- Compare with multiple classifiers, different model settings
 - Write up describing modeling strategy, screen shots of both results in WEKA or tool of choice to illustrate exploration phase, screen shots of all submissions to Kaggle contest site;
- Can you identify an opportunity for model ensembles, if so describe and evaluate;