

Overview

Brian Kim

The Model Building Process

- **Explore** the data. (Descriptive statistics, visualizations)
- **Prepare** the data. (Data cleaning, generate features and labels)
- **Split** the data into training and test sets. (cross-validation, holdout samples)
- **Train** the model on the training set. (decision tree algorithm)
- **Predict** on the test set and **evaluate** based on certain metrics. (precision, recall)
- **Repeat**

Machine Learning

This general structure applies to more than just decision trees — you can use it with any machine learning algorithm, or just anything that outputs a prediction.

Other models/algorithms to try are:

- Logistic Regression
- Support Vector Machines
- Random Forests
- Boosting
- ... and more!