Day 1

Module 1 – Supervised Learning

- You Are the Teacher
  - What Is Machine Learning?
  - Supervised Learning
  - Classification vs. Regression
- Why Decision Trees?
  - Ease of Use
  - Optimal for Many Business Problems

Module 2 – Classification Trees

- Basic Intuition
  - Trees Are Rules
  - Sample Decision Tree
- Overfitting Intuition
  - The Bugbear of Machine Learning
  - The Model Is Good! Or Is It?

Module 3 – Classification Tree Math

- Gini Impurity
- Gini Change
- Many Categories Impurity
- Numeric Feature Impurity

Module 4 – Using Classification Trees

- Classification Trees, tidymodels Style
  - The tidymodels Universe
  - Recipes
  - Model Specifications
  - Workflows
  - Model Fitting
- Hands-on Lab #1
Module 5 – Introducing the Bias-Variance Tradeoff

- Under/Overfitting
  - The Goldilocks Zone
  - Controlling Complexity
- The Bias-Variance Tradeoff
  - Intuitive Example
  - Model Example

Module 6 – Model Tuning

- Supervising the Data
  - Splitting the Data
  - Cross-Validation
- Model Tuning Intuition
  - Making an Intuitive Example Real
  - Estimating Generalization Error
  - What About the Test Set?
- Pruning Classification Trees
  - Pruning Intuition
  - Pre-Pruning
  - Post-Pruning

Module 7 – Model Tuning with tidymodels

- Measuring Model Accuracy
  - Accuracy
  - Confusion Matrices
  - Sensitivity
  - Specificity
- Model Tuning with Tidymodels
  - Setting Up Cross-Validation
  - Cross-Validation Results
  - Tuning the Tree
  - Tuning Results
- Hands-on Lab #2
Day 2

Module 8 – Feature Engineering

- Intuition
  - What Is Feature Engineering?
  - An Example
  - Extracting Features
  - Row vs Column Features

- Data Leakage
  - What Is It?
  - An Example
  - tidymodels to the Rescue

- Engineering Features for Decision Trees
  - Decision Boundaries
  - Visualizing Decision Boundaries
  - Concepts to Remember

- Missing Data
  - Why Is Data Missing?
  - Dealing with Missing Data
  - What Is Imputation?
  - Imputation in tidymodels

- Hands-on Lab #3

Module 9 – Regression Trees

- The Basics
  - Regression Trees Minimize SSE
  - Calculating SSE

- Numeric Feature SSE

- Many Categories SSE

- Regression Trees with tidymodels
  - Measuring Accuracy
  - Model Specification
  - Regression Trees in Practice
Module 10 – The Mighty Random Forest
- Bad, Tree! Bad!
  - Decision Tree Variance
  - High Variance Leads to Overfitting
  - Real-World Decision Trees
- Ensembles
  - Wisdom of the Crowd
  - Manufacturing Independence
- Bagging
  - Randomizing Rows
  - Bagging in Action
  - The Power of Bagging
- Feature Randomization
  - Intuition
  - Randomizing Columns
  - Feature Randomization in Action

Module 11 – Using the Random Forest
- Tuning Random Forests
  - The Bias-Variance Tradeoff
  - Random Forest Hyperparameters
- Feature Importance
  - Out of Bag (OOB) Data
  - Permutation Importance
  - An Example
- Random Forests with tidymodels
  - Comparing a Decision Tree to a Random Forest
- Hands-on Lab #4

Module 12 – Course Wrap-Up
- Want to Kaggle?
- Additional Resources