1. Read pruning, entropy, and information gain in decision tree. Compute entropy manually from slide, put your steps and result in word file.

2. In the last class, I covered C4.5 decision tree classification technique where the dependent variable is categorical and splitting is based on information gain. Please read about decision tree regression where the dependent variable is continuous and splitting is based on standard deviation reduction: http://www.saedsayad.com/decision\_tree\_reg.htm. You will now have assignments on both. Implement decision tree classification with Titanic dataset (predict Survived) and decision tree regression Energy efficiency Dataset (outcome y1 or y2) in R. You can use rpart library but there are other options.

You can find dataset information from links below:  
https://www.kaggle.com/c/titanic  
<https://archive.ics.uci.edu/ml/datasets/Energy+efficiency>

3. Create report including your exploration about the data, data preprocessing, your models and the evaluation of your models.