

Peter McKay, Daniel Walinsky  
2015-02-19

## Project Proposal

For our final project, we intend to tackle the following Kaggle problem: the Forest Cover Type Prediction. In brief, we wish to, given a collection of geographic and cartographic variables that apply to a given chunk of forest, determine what kind of trees will be most common in that forest region.

Technically speaking, this is a multi-class classification problem.

Random forests and decision trees would be poetically appropriate, but possibly less than optimal.

intro

related work

methodology We plan to carry out most of our research using Python. More specifically, Python3, with the help of a number of libraries designed to assist in machine learning and data analysis.

**pandas: Python Data Analysis Library:** We will make use of data structures and input mechanisms introduced in pandas.

**scikit-learn:** In an effort to avoid duplicating the wheel (and running out of RAM while we try to get that wheel to load), we will be making use of scikit-learn to actually carry out our machine algorithms.

**numpy:** numpy provides a number of useful

results

reflection

conclusion