
Data Analysis Pitch

By Jeff Lee and Clay Wadman

Coding Skillsets:

- Using OOP and Procedural Programming
 - Working with public and private variables
 - Analyzing statistics and translating to matplotlib graphs
 - Experience working with large-scale datasets
 - Filtering data to get to important information
 - Working with packaged and custom decision trees
 - Code organization with classes and comments
-

Approach/Work Management Skills:

- Ability to identify difficult versus less strenuous tasks
 - Experience dividing up “grunt” work between group members to be completed alone
 - Experience communicating via phone/email concerning work division and individual coding questions
 - Ability to seamlessly combine code done by separate individuals, using most effective portions from each draft
 - Completing “expert” work within necessary timelines
-

Areas of Work... :

We Are Proud Of:

- Smooth User Interface
- Statistics Graphing with Labeling
- Document Reader Function
- Scikit-Learn Tree Functionality

We Could Improve Upon:

- myTree Custom Decision Tree Implementation
-

Types of Improvements We Could Make to Improve Analysis:

- More accurate predictions could be made if we analyzed more exact data
 - Most popular words each user uses rather than word count
 - Time emails were sent; do specific users tend to email at certain hours?
 - Analyze email chains
 - We ignored data in replied and forwarded emails
 - Access to more data
 - The last thing we could do is simply have more training data. The more training we have, the more precise our decision tree will be
-