Scikit-Learn GUI

NETSI Team: Abhilash Nair, Sean Dai, Graham Wright, Rohit Kale

Client: Dr. Olufisayo Omojokun

Presentation Overview

- Introduction to Machine Learning
- Importance of Machine Learning
- Feasibility Report
- Statement of Work
- Prototype Walkthrough
- Reflection

Introduction to Machine Learning

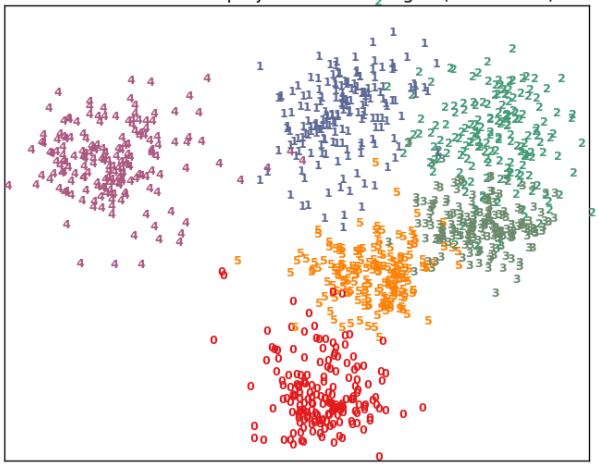
- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

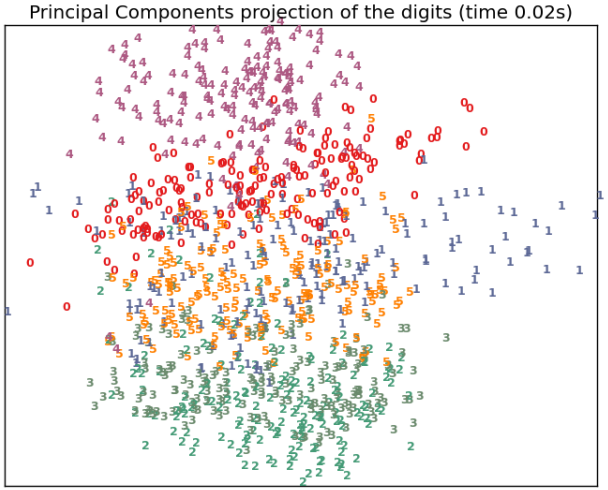
Importance of Machine Learning

- Growing data sets
- Complexity of data
- Replacement of manual human processes
- Cool applications!

Data Visualization & Machine Learning

The 64-dimensional digits dataset





Source code: http://scikit-learn.org/stable/_downloads/plot_lle_digits.py

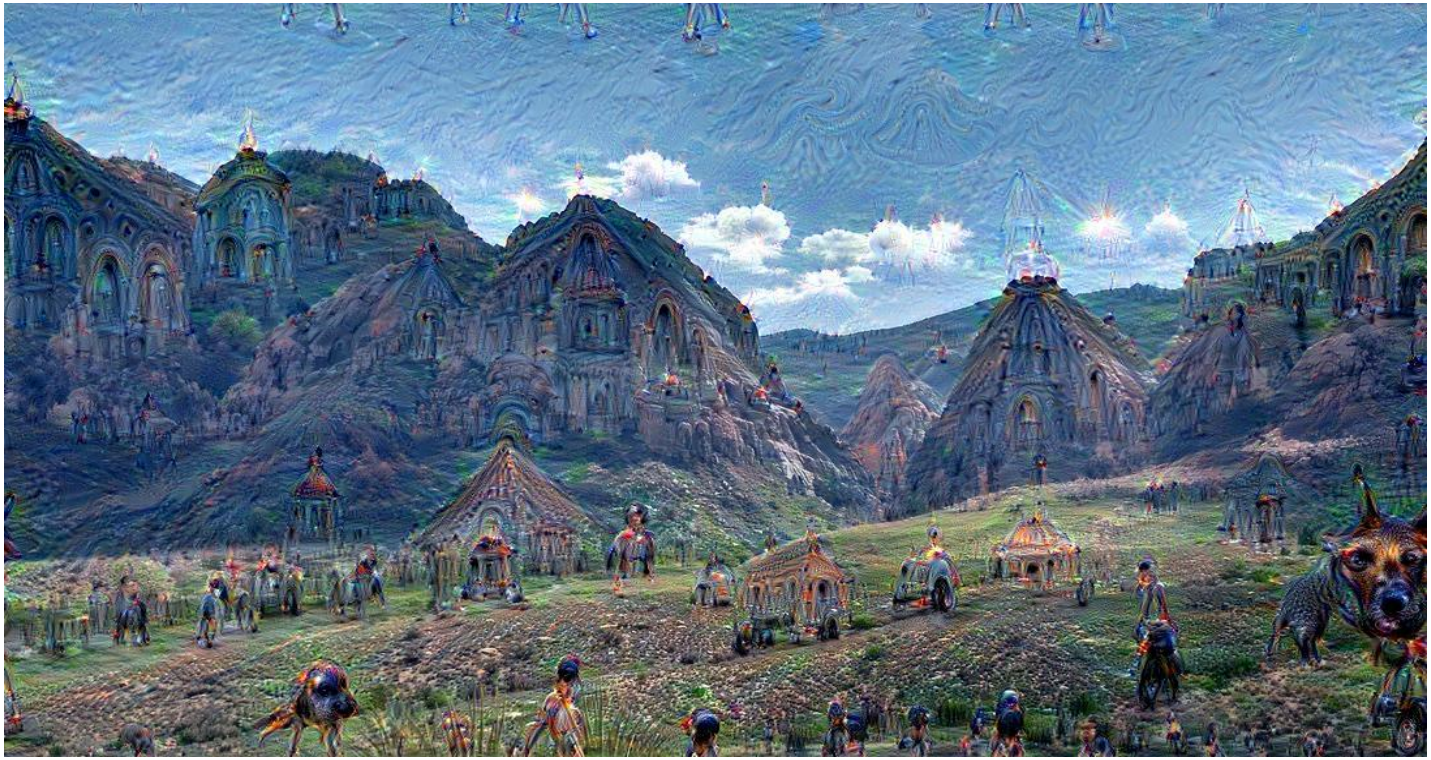
ML: Google Deep Dream

- Artificial Neural Network (ANN)
- MIT Computer Science and AI Laboratory Places data set
- Results and training data

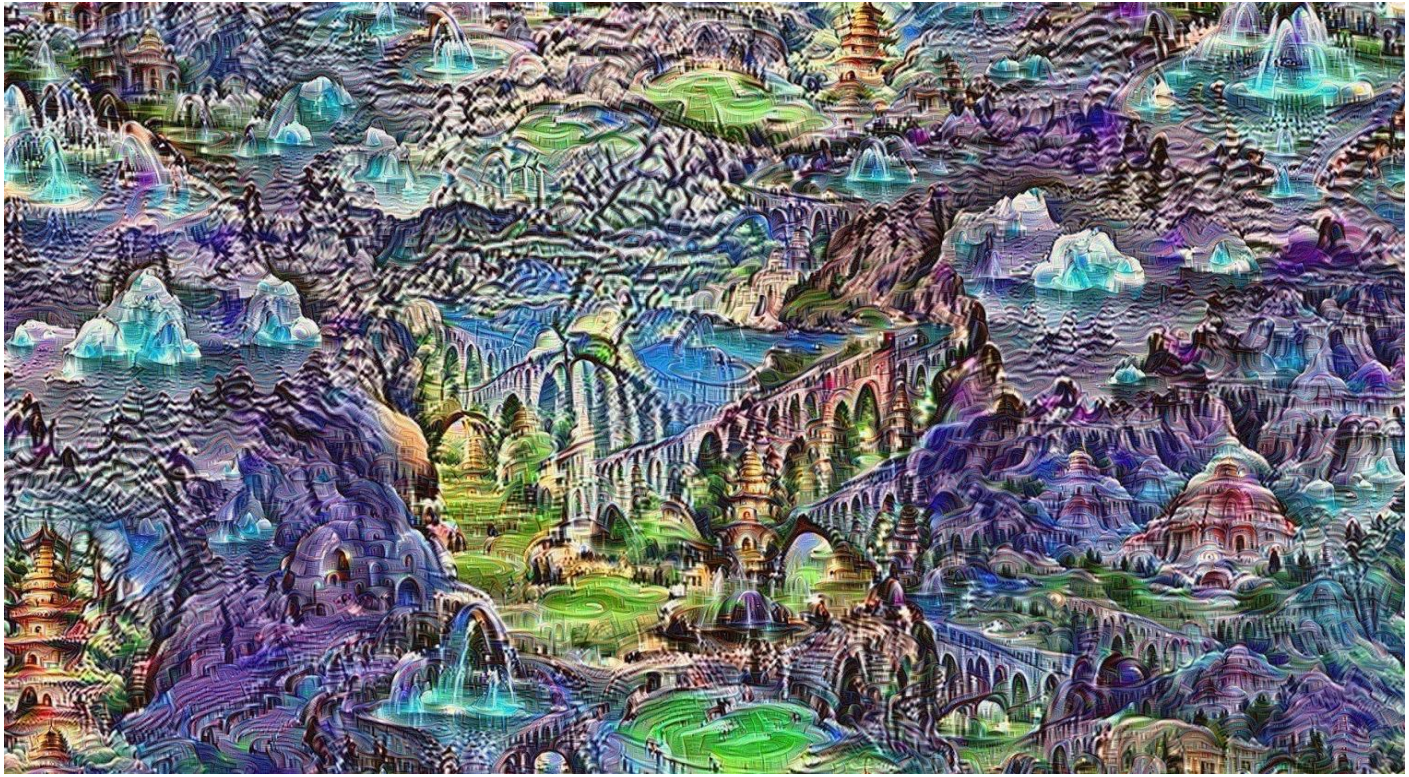
Source: http://googleresearch.blogspot.co.uk/2015/06/inceptionism-going-deeper-into-neural.html

Before

After



Random Noise



Feasibility Report

Alternatives

1. Standalone GUI
   - PyQt and *scikit-learn*
   - Python visualization libraries
   - Client side computation

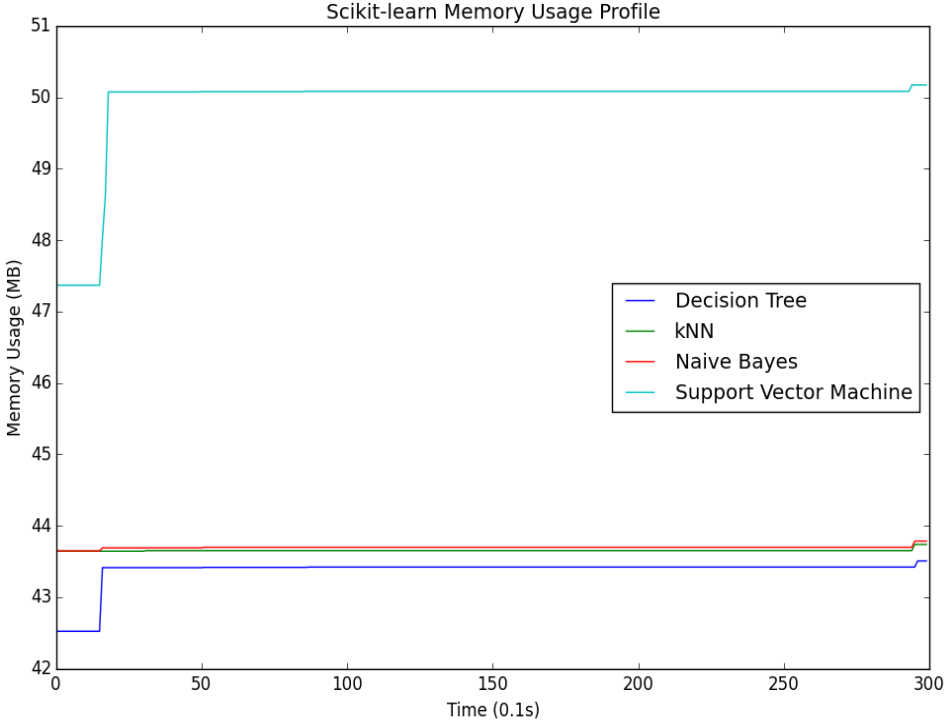
2. Web application
   - Javascript visualization libraries
   - Server side computation

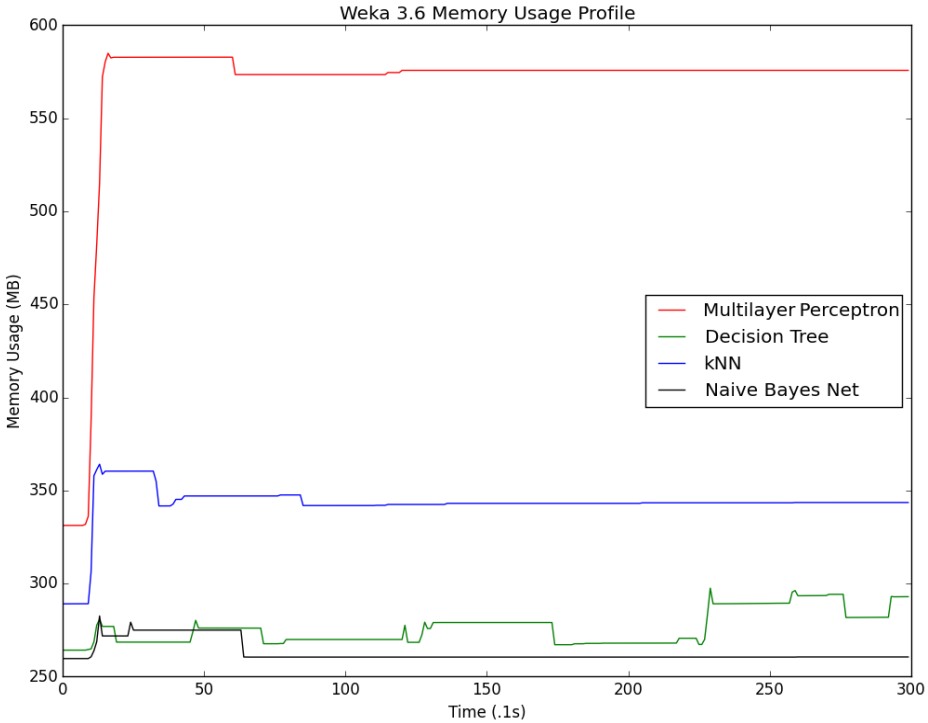
3. Tutorials and improved documentation for Weka
   - YouTube tutorials
   - Extended documentation

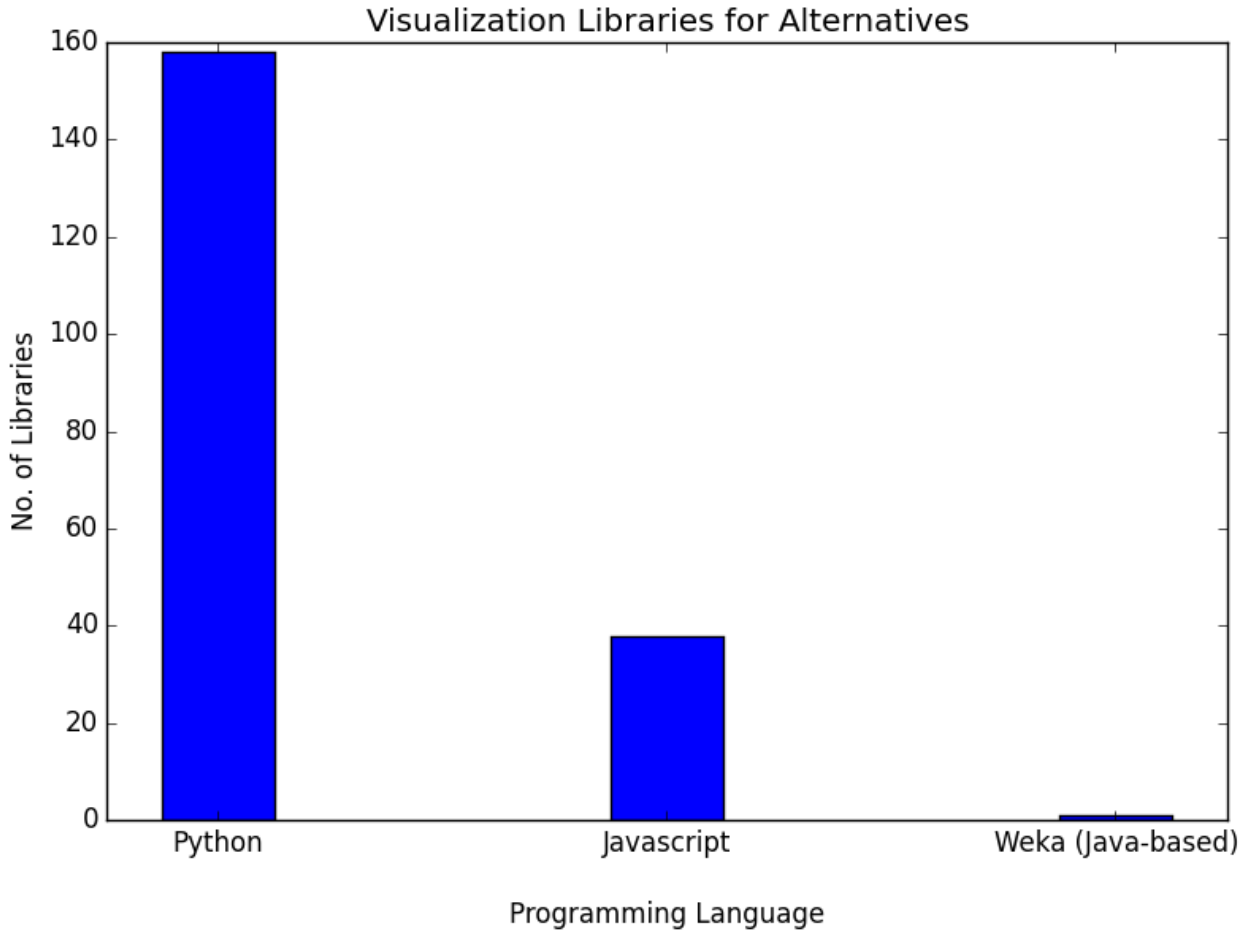
Technical Evaluations

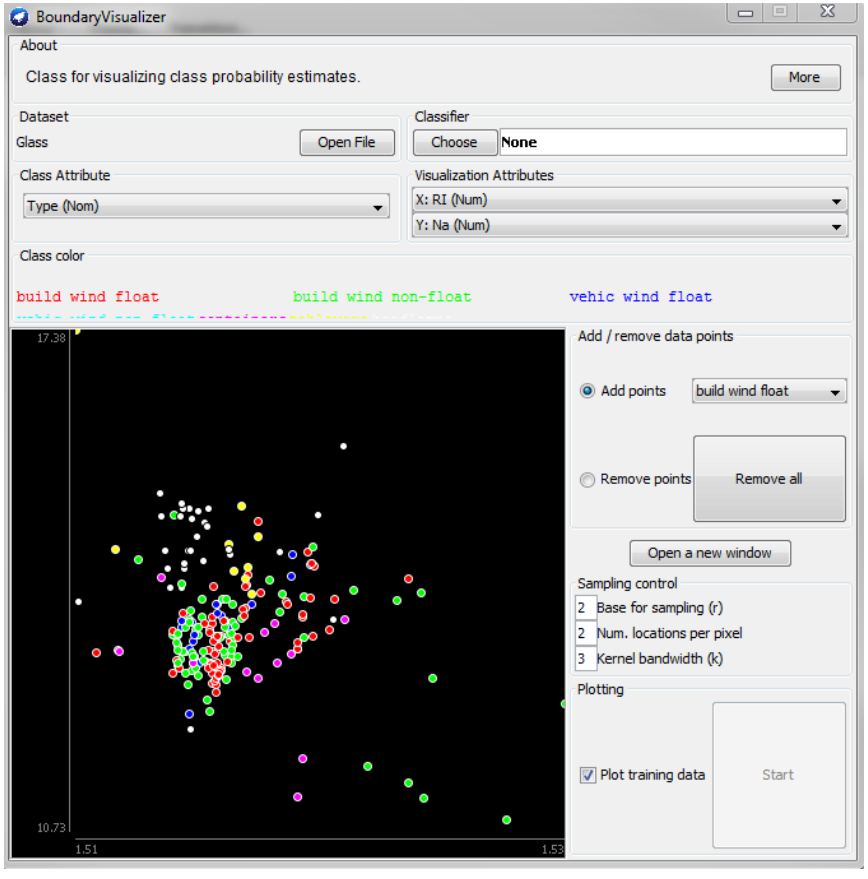
Alternatives	Memory Consumption	Hard Disk Space	Visualization Libraries
1. Standalone GUI	• > 143 MB	12.9 MB (PYQT) 80 MB Python 37.7 MB scipy 10.3 MB scikit-learn 100 MB numpy 16.6 MB matplotlib • > 257.5 MB Total	• 158
2. Web Application	• > 75 MB	• > 95 MB Firefox/Chrome	• 38
3. Weka Improved	• > 250 MB	160 MB Java 92.8 MB Weka • > 252.8 MB Total	• 1

Scikit-Learn Vs. Weka

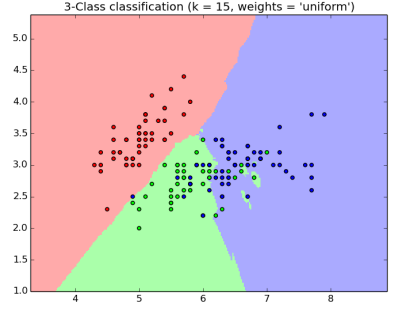


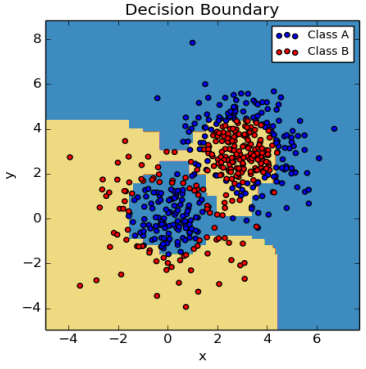


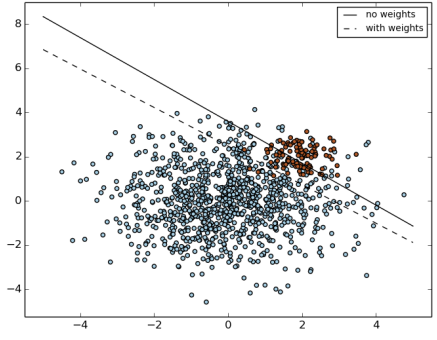

Visualization Libraries for Alternatives
160
140
120
100
80
60
40
20
0
No. of Libraries
Python
Javascript
Weka (Java-based)
Programming Language

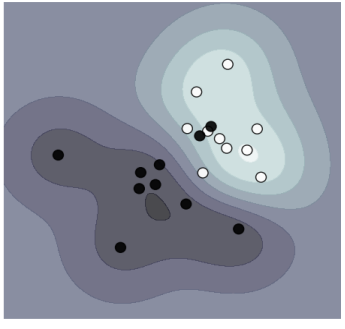

BoundaryVisualizer
About
Class for visualizing class probability estimates.
More
Dataset
Glass
Open File
Classifier
Choose
None
Class Attribute
Type (Nom)
Visualization Attributes
X: RI (Num)
Y: Na (Num)
Class color
build wind float
build wind non-float
vehic wind float
17.38
10.73
1.51
1.53
Add / remove data points
Add points
build wind float
Remove points
Remove all
Open a new window
Sampling control
2 Base for sampling (r)
2 Num. locations per pixel
3 Kernel bandwidth (k)
Plotting
Plot training data
Start

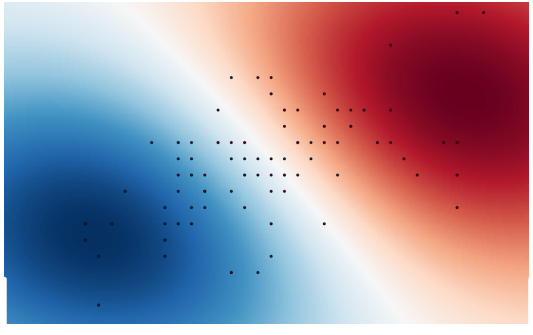
Vs.


3-Class classification (k = 15, weights = 'uniform')


Decision Boundary
Class A
Class B
x
y


no weights
with weights





Management Evaluations

Alternatives	Implementation Duration	Ease of Maintenance and Distribution
1. Standalone GUI	• 5 months	• Medium - Hard
2. Web Application	• 7 months	• Medium - Hard
3. Weka Improved	• 2 months	• Easy

Economic Evaluations

Alternatives	Total Project Cost
1. Standalone GUI	• $0
2. Web Application	• Initial $5.99/month via GoDaddy (Price increase to scale)
3. Weka Improved	• $0

Political/Cultural Evaluations

Alternatives	Accessibility to Documentation	User Experience
1. Standalone GUI	• Offline, online, and in the GUI	• 9.00/10.00
2. Web Application	• Online and in the GUI	• 6.50/10.00
3. Weka Improved	• Offline and on YouTube	• 7.00/10.00

Conclusion of Evaluations

1. Standalone Python GUI
    - Accessible documentation
    - Visualization capability

2. Web Application
    - Team unfamiliarity with web-dev
    - Monthly server cost

3. Weka
    - Outdated implementation framework
    - High memory consumption
    - Limited visualization capability

Team’s Decision

- Standalone GUI
  - Python
  - Clear documentation
  - GUI hints
  - No-cost development

Statement of Work

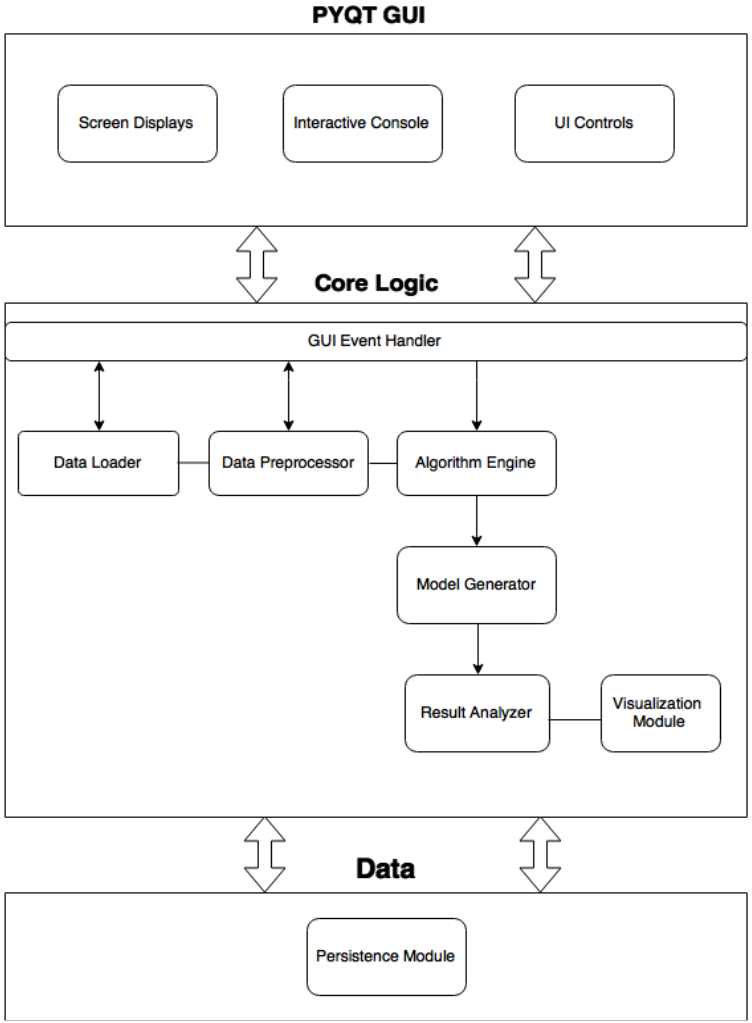
Problem

- Current tool in market
- Effectiveness and efficiency
- Learning curve

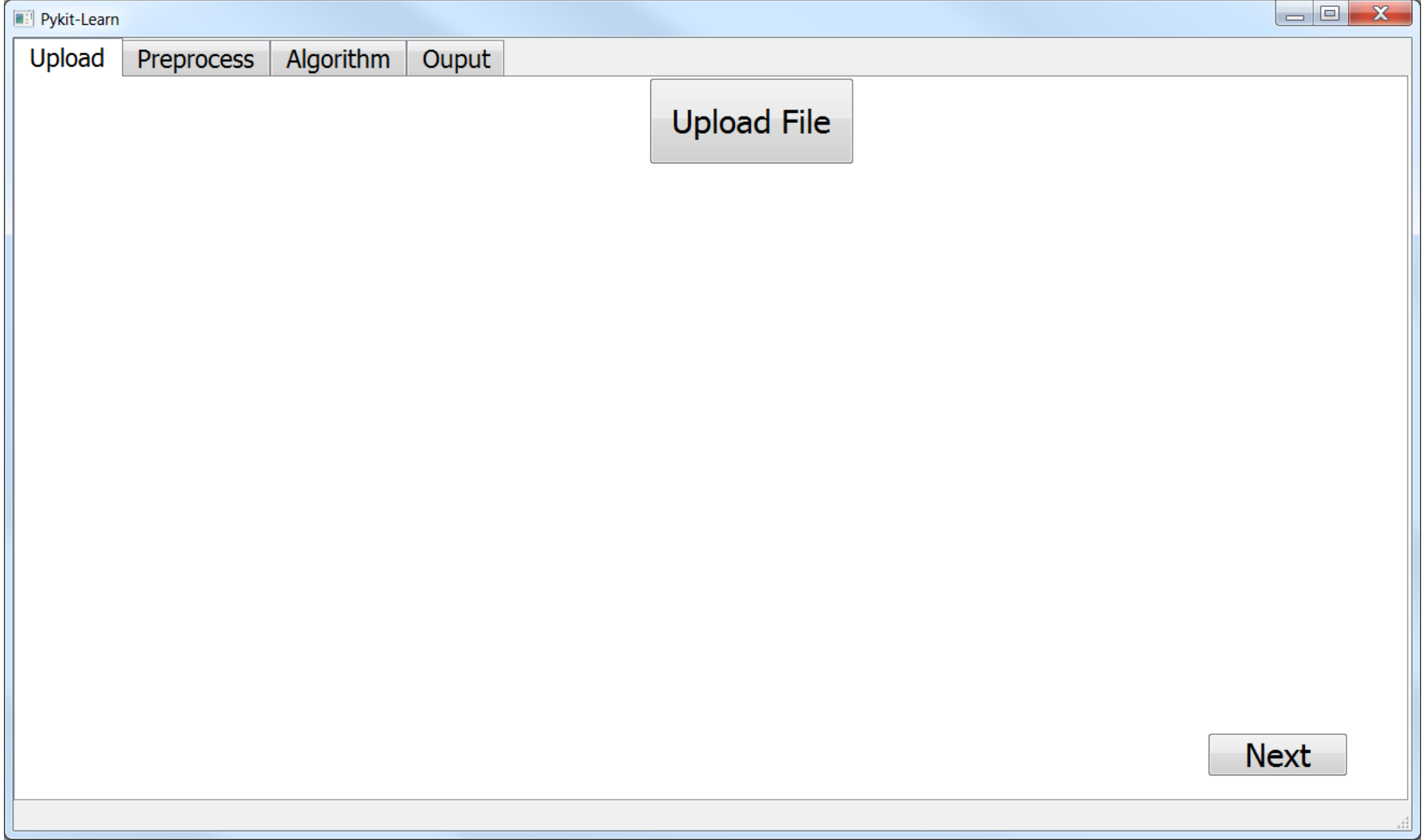
Requirements/Features

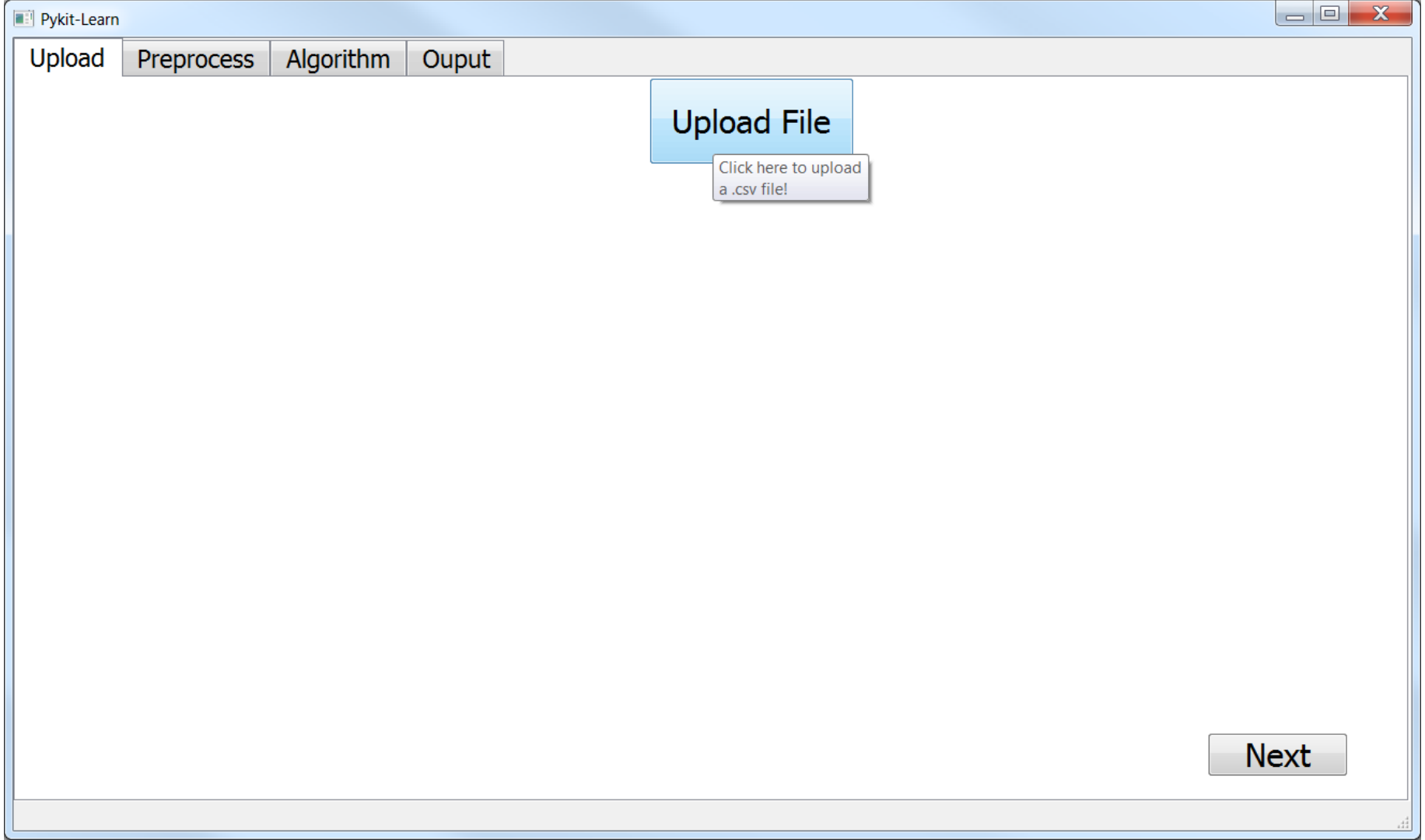
Functional Requirements	Nonfunctional Requirements
FR 1 - File Support	NFR 1 - Documentation
FR 2 - Raw Input Visualization	NFR 2 - Task Responsiveness
FR 3 - Preprocessing	NFR 3 - Runtime Stability
FR 4 - Algorithm Support	NFR 4 - Ease of Use
FR 5 - Output Metrics	NFR 5 - Help and Support
FR 6 - Output Visualization	NFR 6 - Windows Compatibility

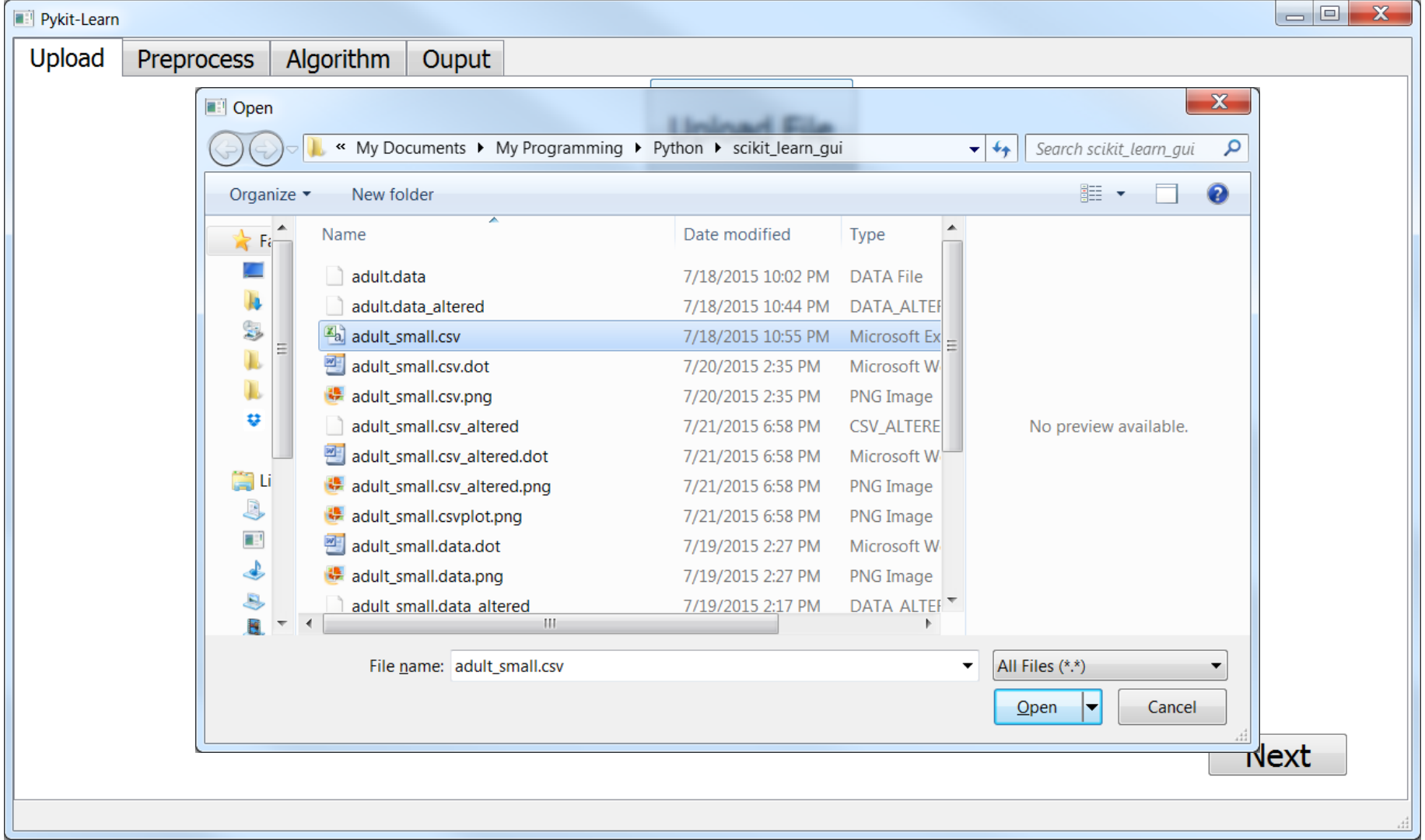
System Architecture



Prototype Walkthrough


Pykit-Learn
Upload
Preprocess
Algorithm
Ouput
Upload File
Next

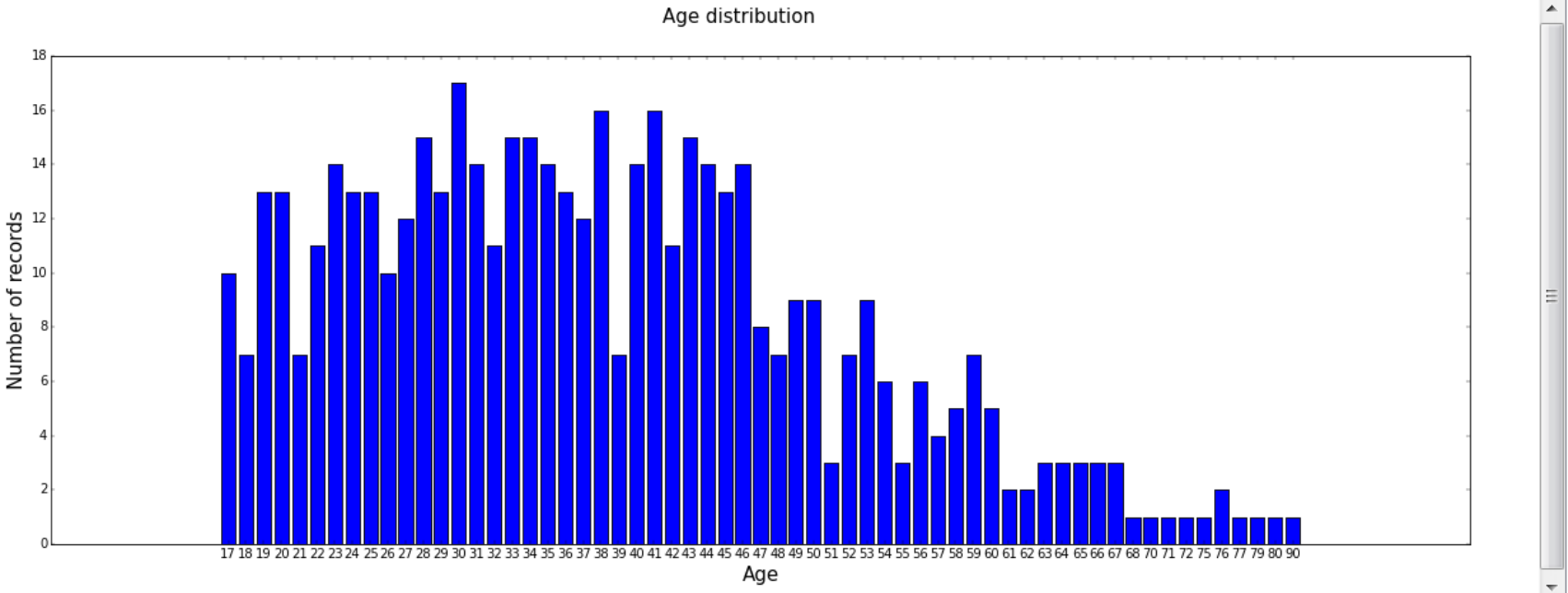

Pykit-Learn
Upload
Preprocess
Algorithm
Ouput
Upload File
Click here to upload a .csv file!
Next


Pykit-Learn
Upload
Preprocess
Algorithm
Ouput
Open
My Documents ▸ My Programming ▸ Python ▸ scikit_learn_gui
Search scikit_learn_gui
Organize
New folder
Name
Date modified
Type
adult.data 7/18/2015 10:02 PM DATA File
adult.data_altered 7/18/2015 10:44 PM DATA_ALTE
adult_small.csv 7/18/2015 10:55 PM Microsoft Ex
adult_small.csv.dot 7/20/2015 2:35 PM Microsoft W
adult_small.csv.png 7/20/2015 2:35 PM PNG Image
adult_small.csv_altered 7/21/2015 6:58 PM CSV_ALTERE
adult_small.csv_altered.dot 7/21/2015 6:58 PM Microsoft W
adult_small.csv_altered.png 7/21/2015 6:58 PM PNG Image
adult_small.csvplot.png 7/21/2015 6:58 PM PNG Image
adult_small.data.dot 7/19/2015 2:27 PM Microsoft W
adult_small.data.png 7/19/2015 2:27 PM PNG Image
adult_small.data_altered 7/19/2015 2:17 PM DATA_ALTE
No preview available.
File name: adult_small.csv
All Files (*.*)
Open
Cancel
Next

Pykit-Learn

Upload | Preprocess | Algorithm | Ouput

Upload File

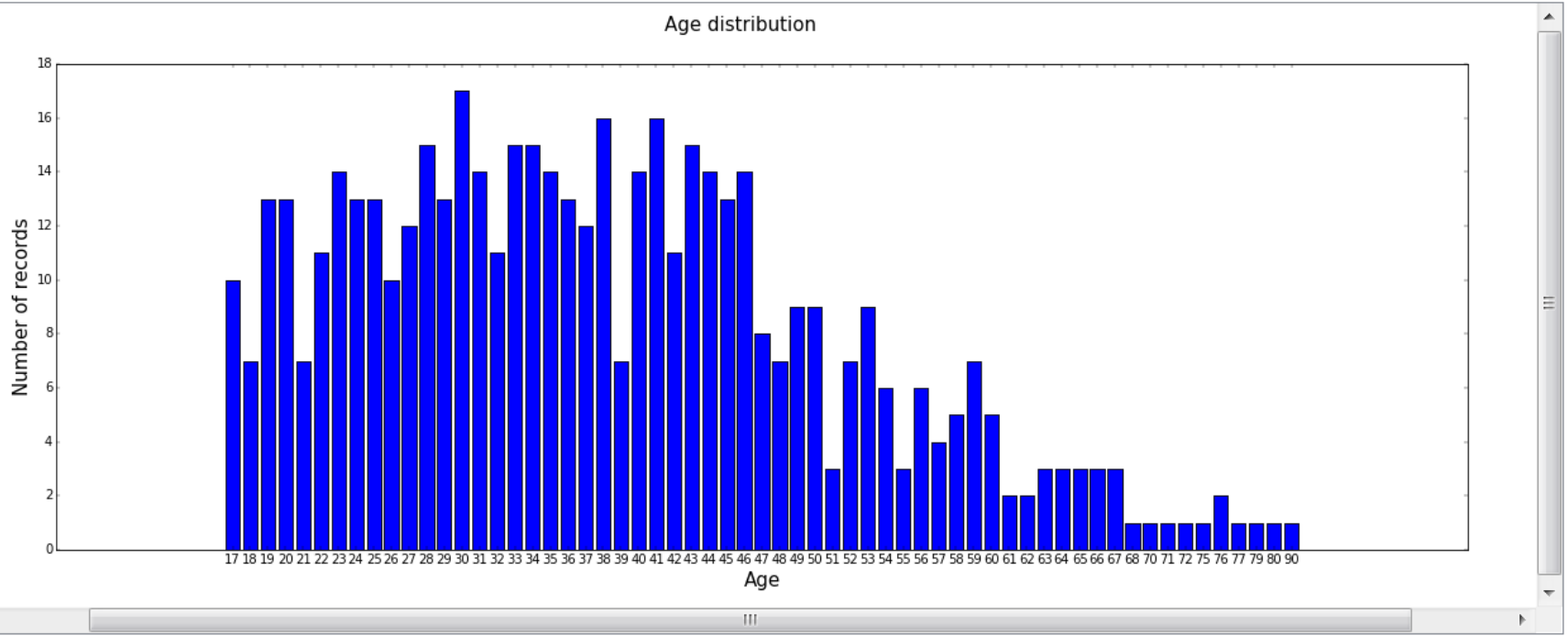


Next

Pykit-Learn

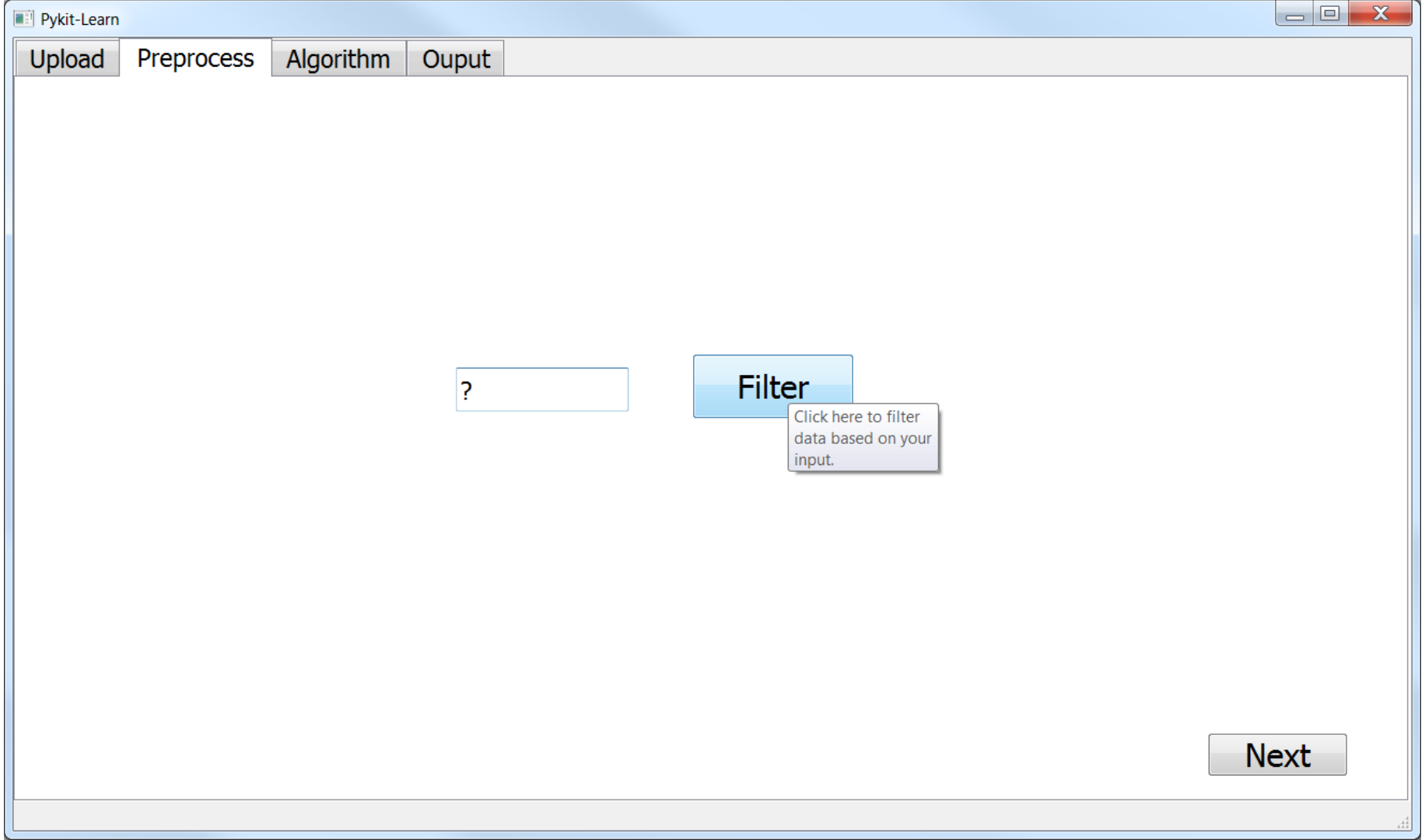
Upload | Preprocess | Algorithm | Ouput

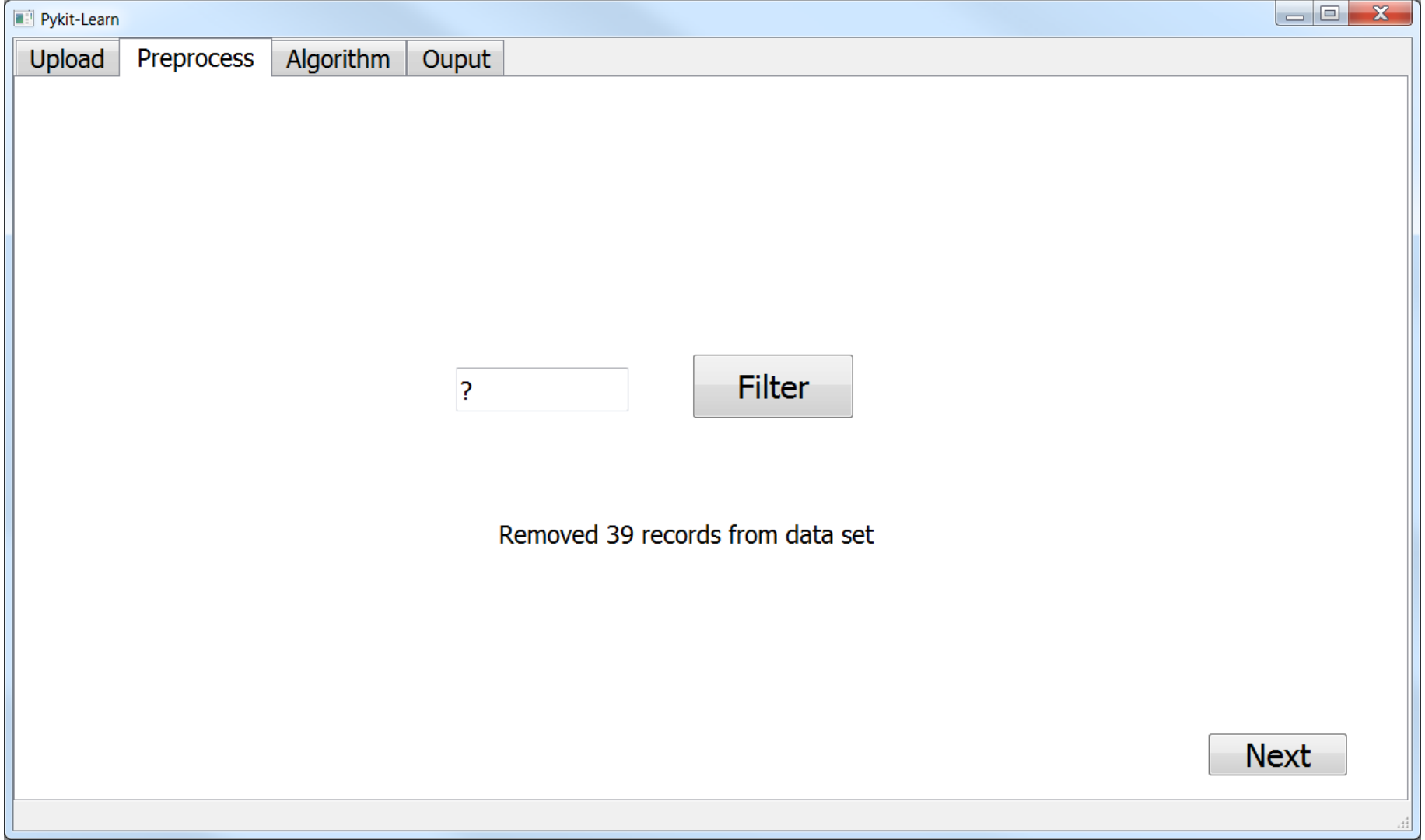
Upload File

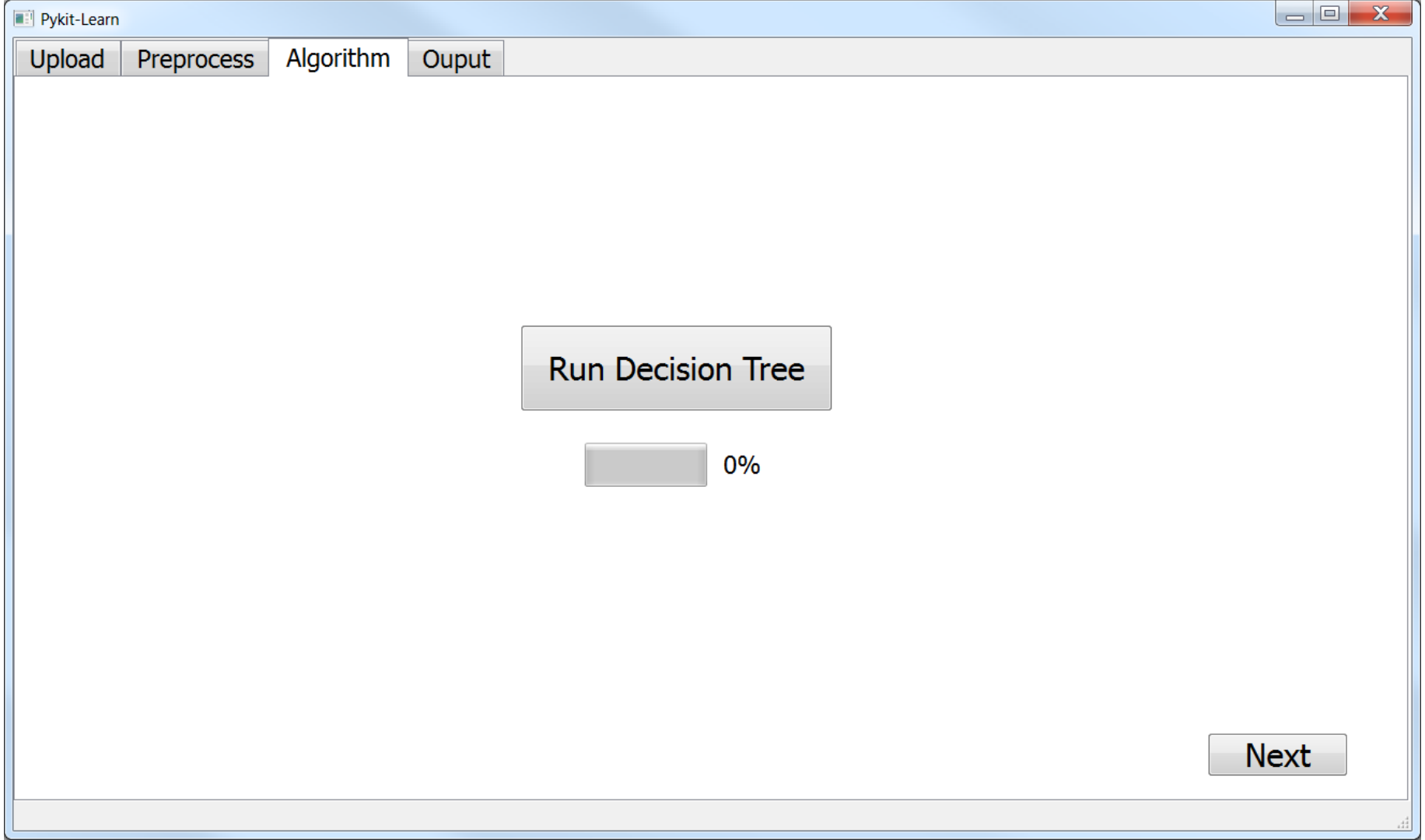


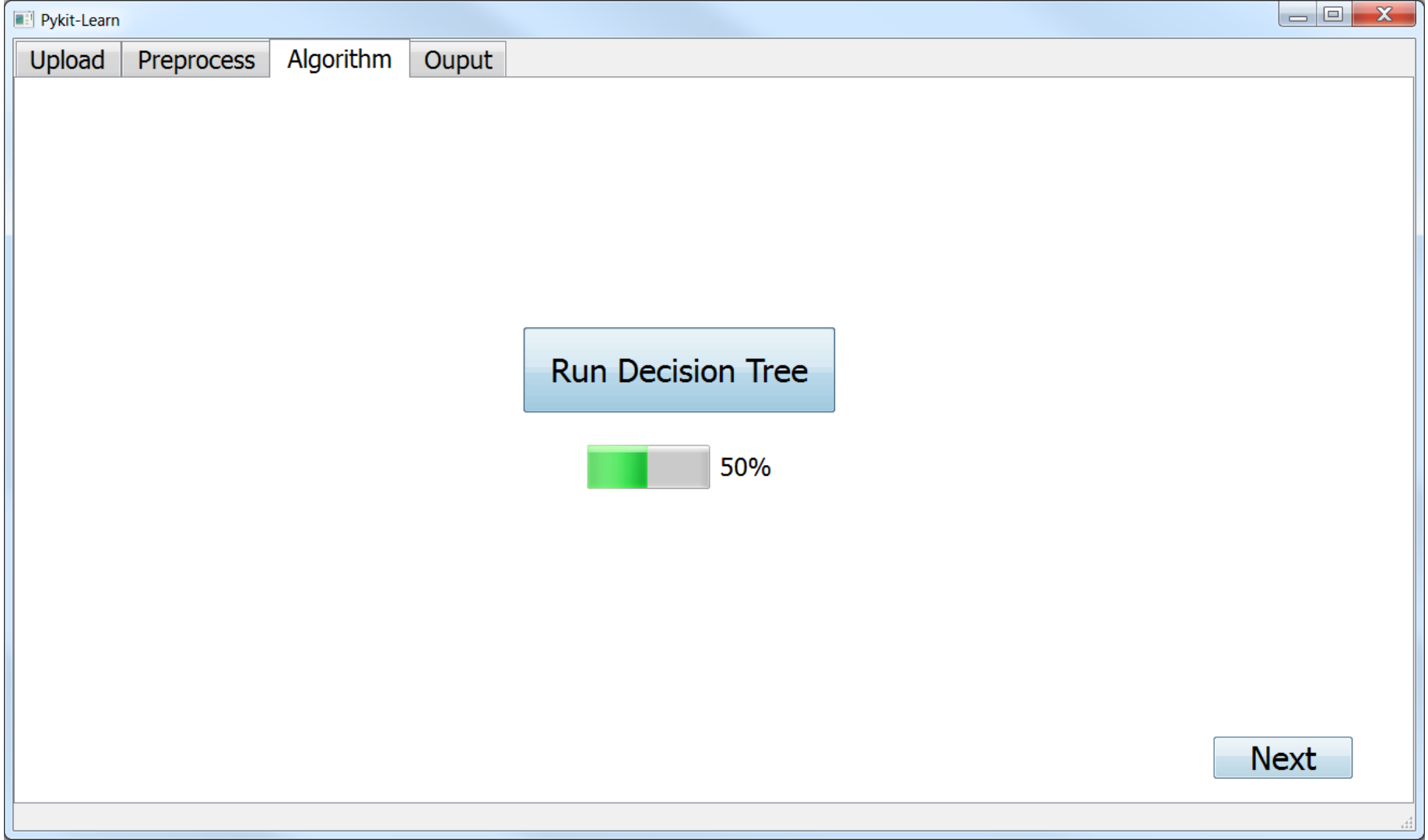
Next

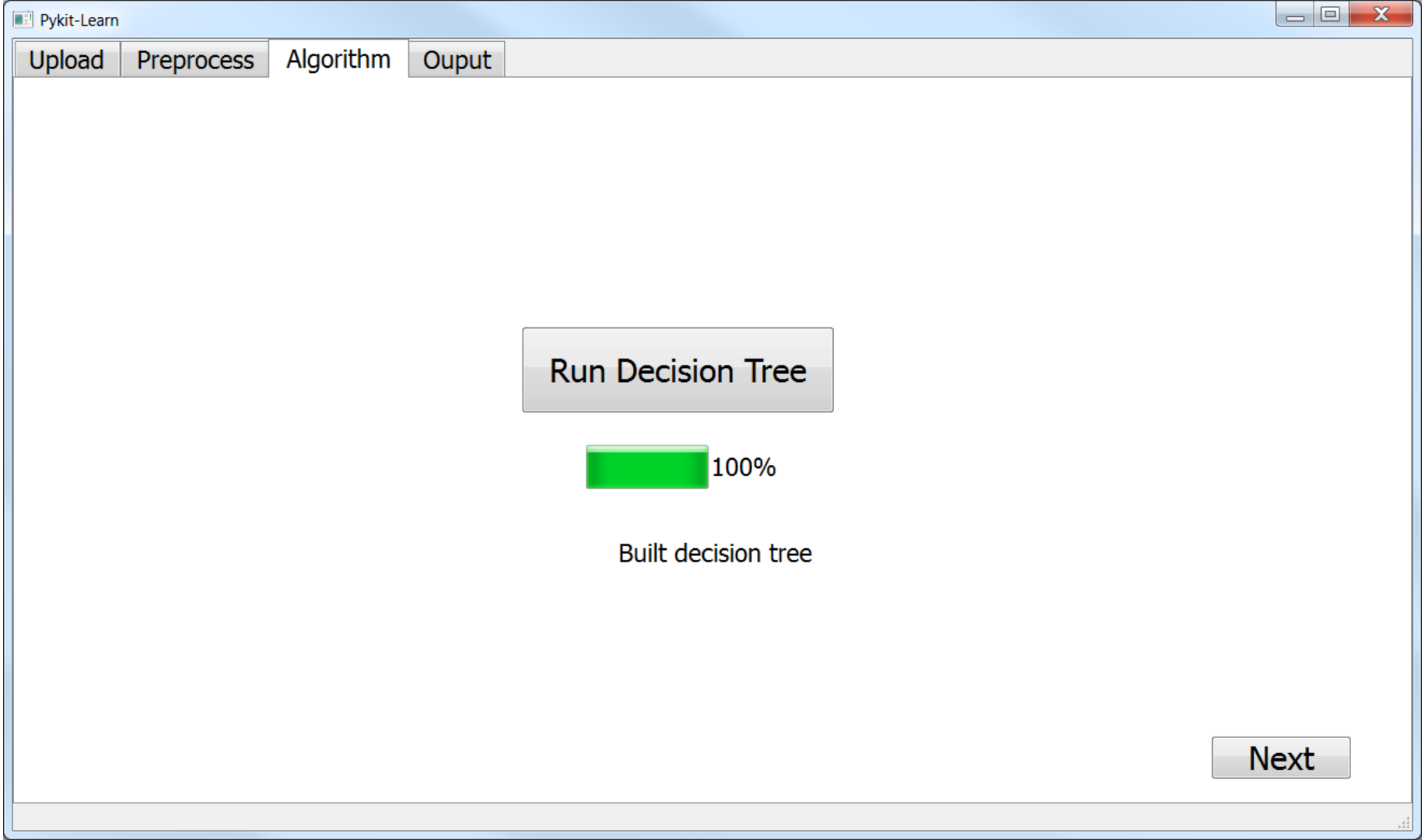
Move to the next screen


Pykit-Learn
Upload
Preprocess
Algorithm
Ouput
?
Filter
Click here to filter data based on your input.
Next


Pykit-Learn
Upload
Preprocess
Algorithm
Ouput
?
Filter
Removed 39 records from data set
Next

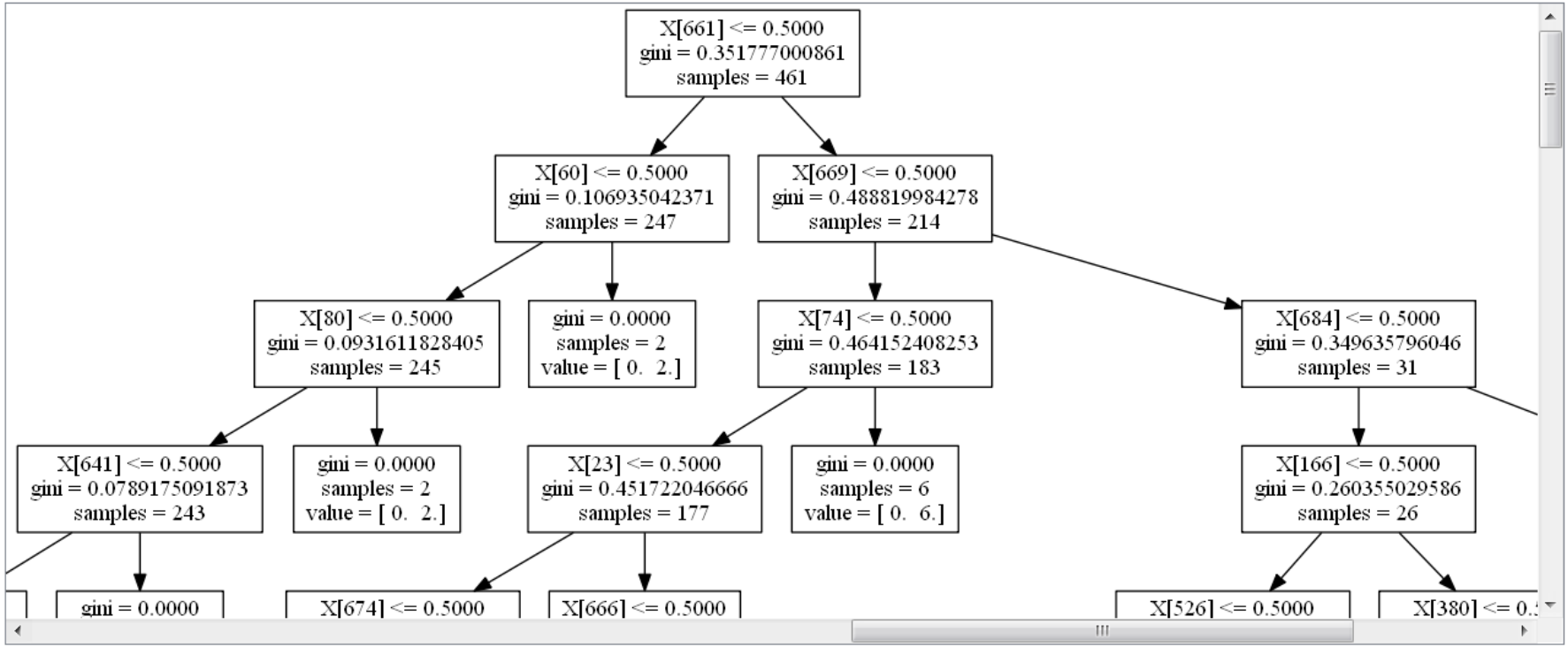

Pykit-Learn
Upload
Preprocess
Algorithm
Ouput
Run Decision Tree
0%
Next


Pykit-Learn
Upload
Preprocess
Algorithm
Ouput
Run Decision Tree
50%
Next


Pykit-Learn
Upload
Preprocess
Algorithm
Ouput
Run Decision Tree
100%
Built decision tree
Next

Pykit-Learn

Upload | Preprocess | Algorithm | Ouput



Home

Schedule: Release 1

Feature	Requirements Met
Upload input data	FR 1
Visualize input data	FR 2
PyQT UI layout	FR 1, FR4, NFR 4, NFR 6
Preprocess input data	FR 3

Schedule: Release 2

Feature	Requirements Met
Provide algorithm options	FR 4
Full algorithmic support	FR 4
Progress bars added	NFR 2
Informational messages added to GUI	NFR 5

Schedule: Release 3

Feature	Requirements Met
Error handling	NFR 3
Display Result Statistics	FR 5
Visualize Output Data	FR 6
Complete documentation	NFR 1

Reflection: What We Learned

- Research process
- Extensive documentation
- Time management

What We Should Do Differently

- Regular meeting schedule
- Elimination of indecisiveness
- Precise research methodology
- Better communication

Next Semester

- Pivotal Tracker for project management
- Delegation of responsibilities
- Version control software

Concluding Remarks

Take aways:

- Machine learning and the future.
- The layman/technical gap

Sources

1. http://googleresearch.blogspot.co.uk/2015/06/inceptionism-going-deeper-into-neural.html
2. http://scikit-learn.org/stable/_downloads/plot_lle_digits.py
3. http://scikit-learn.org/stable/_images/

Questions?