



Python Interface Overview

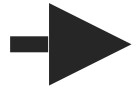
Action	Pandas or scikit-learn	H2O
Reading data	<code>pandas.read_csv(data_path)</code>	<code>h2o.import_file(data_path)</code>
Summarizing data	<code>pandas_frame.describe()</code>	<code>h2o_frame.describe()</code>
Summary statistics	<code>pandas_frame.mean()</code>	<code>h2o_frame.mean()</code>
Combining rows	<code>pandas.concat(list[frame1,frame2])</code>	<code>h2o_frame.rbind(h2o_frame2)</code>
Combining columns	<code>pandas.concat(list[frame1,frame2],axis = 1)</code>	<code>h2o_frame.cbind(h2o_frame2)</code>
Data selection	<code>pandas_frame[:, :]</code>	<code>h2o_frame[:, :]</code>
Transforming columns	<code>np.log(pandas_frame[x])</code> <code>np.sqrt(pandas_frame[x])</code>	<code>h2o_frame[x].log()</code> <code>h2o_frame[x].sqrt()</code>
Building Random Forest	<code>model = RandomForestClassifier(n_estimators = 100)</code> <code>model = model.fit(x_frame, y_frame)</code>	<code>model = H2ORandomForestClassifier(n_trees = 100)</code> <code>model = model.train(x, y, train_frame)</code>
Model Prediction	<code>model.predict</code>	<code>model.predict</code>
Model Metrics	<code>metrics.auc</code>	<code>metrics = model.model_performance(frame)</code> <code>metrics.auc()</code>

Reading Data into H2O with Python

STEP 1



Python
user



```
h2o_df = h2o.import_file("../data/allyears2k.csv")
```

Python Interface Overview

Action	Pandas or scikit-learn	H2O
Reading data	<code>pandas.read_csv(data_path)</code>	<code>h2o.import_file(data_path)</code>
Summarizing data	<code>pandas_frame.describe()</code>	<code>h2o_frame.describe()</code>
Summary statistics	<code>pandas_frame.mean()</code>	<code>h2o_frame.mean()</code>
Combining rows	<code>pandas.concat(list[frame1,frame2])</code>	<code>h2o_frame.rbind(h2o_frame2)</code>
Combining columns	<code>pandas.concat(list[frame1,frame2],axis = 1)</code>	<code>h2o_frame.cbind(h2o_frame2)</code>
Data selection	<code>pandas_frame[:, :]</code>	<code>h2o_frame[:, :]</code>
Transforming columns	<code>np.log(pandas_frame[x])</code> <code>np.sqrt(pandas_frame[x])</code>	<code>h2o_frame[x].log()</code> <code>h2o_frame[x].sqrt()</code>
Building Random Forest	<code>model = RandomForestClassifier(n_estimators = 100)</code> <code>model = model.fit(x_frame, y_frame)</code>	<code>model = H2ORandomForestClassifier(n_trees = 100)</code> <code>model = model.train(x, y, train_frame)</code>
Model Prediction	<code>model.predict</code>	<code>model.predict</code>
Model Metrics	<code>metrics.auc</code>	<code>metrics = model.model_performance(frame)</code> <code>metrics.auc()</code>