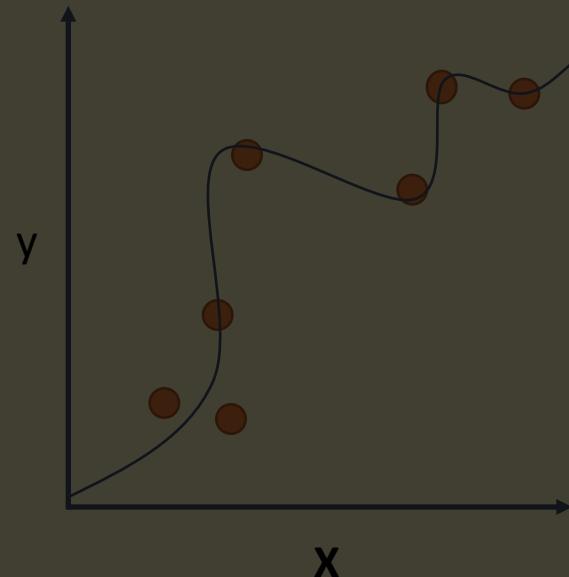


# Classification Example

- **Data:** Human Activity Recognition Using Smartphone Sensors (2012)
- **Source:**  
<https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones>

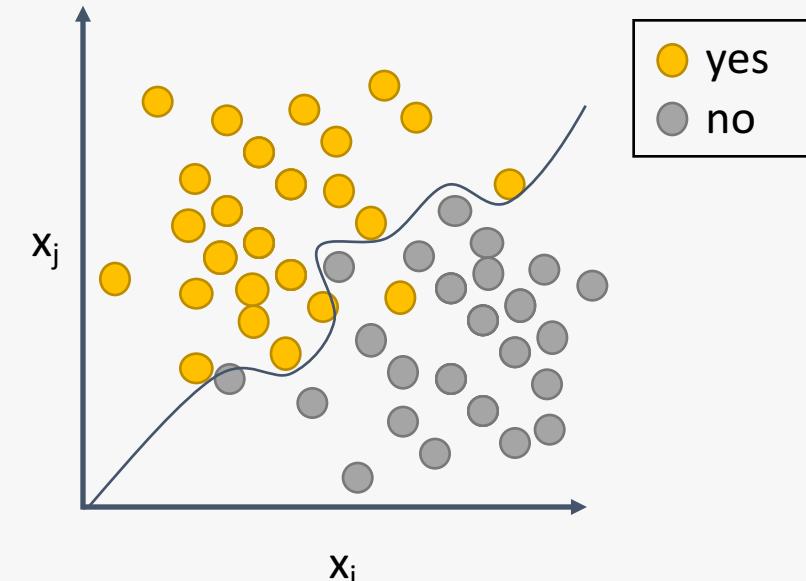
# Supervised Learning – You Already Have Target Data

Regression:  
How much will a customers spend?



H<sub>2</sub>O algos:  
Penalized Linear Models  
Random Forest  
Gradient Boosting  
Neural Networks  
Stacked Ensembles

Classification:  
Will a customer make a purchase? Yes or No



H<sub>2</sub>O algos:  
Penalized Linear Models  
Naïve Bayes  
Random Forest  
Gradient Boosting  
Neural Networks  
Stacked Ensembles

- Smartphone Sensor Data
  - 561 Features
  - 6 Activities
    - Walking
    - Walking Upstairs
    - Walking Downstairs
    - Sitting
    - Standing
    - Laying
  - Train (7k)
  - Test (3k)

[About](#) [Citation Policy](#) [Donate a Data Set](#) [Contact](#)

Repository  Web [Search](#)

[View ALL Data Sets](#)



**Machine Learning Repository**  
Center for Machine Learning and Intelligent Systems

### Human Activity Recognition Using Smartphones Data Set

[Download](#) [Data Folder](#) [Data Set Description](#)

**Abstract:** Human Activity Recognition database built from the recordings of 30 subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensors.

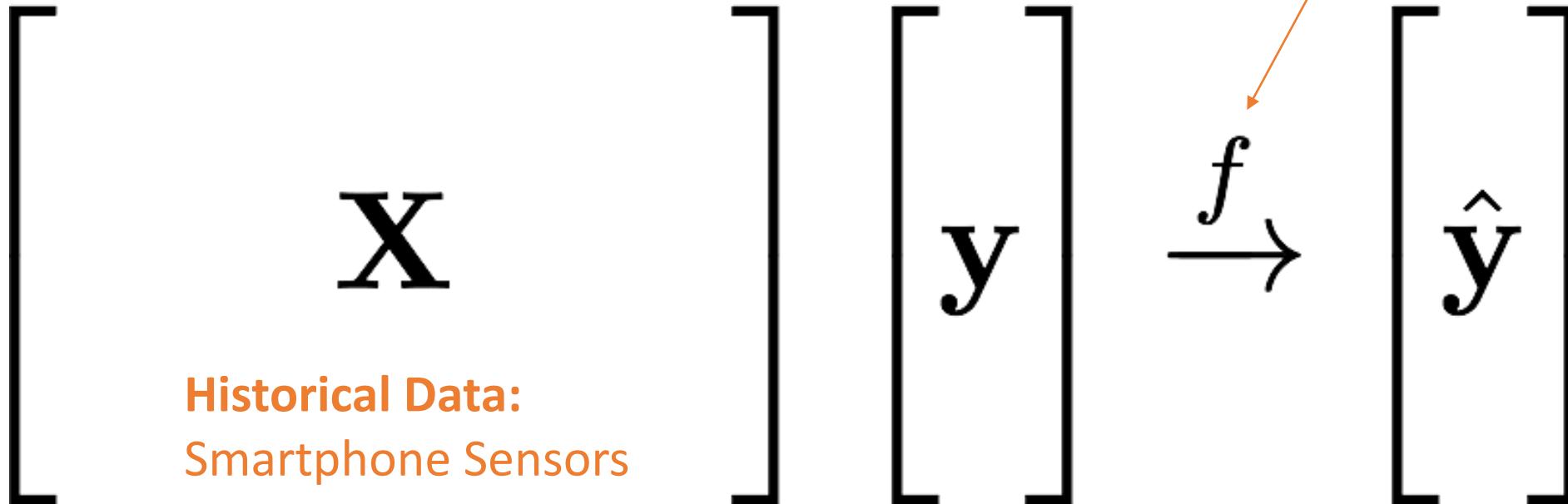
Data Set Characteristics:	Multivariate, Time-Series	Number of Instances:	10299	Area:	Computer
Attribute Characteristics:	N/A	Number of Attributes:	561	Date Donated:	2012-12-10
Associated Tasks:	Classification, Clustering	Missing Values?	N/A	Number of Web Hits:	505091

#### Source:

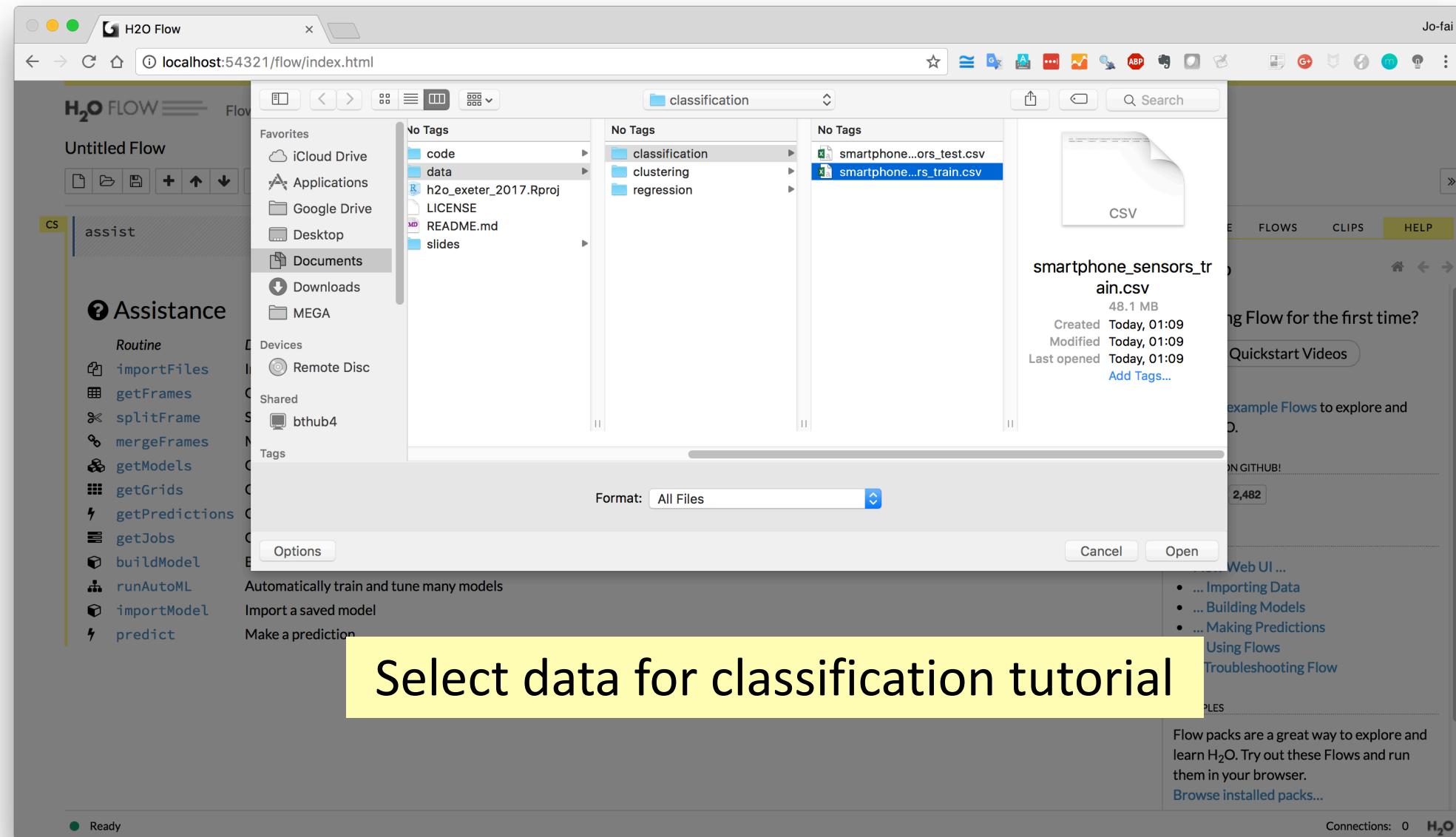
Jorge L. Reyes-Ortiz(1,2), Davide Anguita(1), Alessandro Ghio(1), Luca Oneto(1) and Xavier Parra(2)  
 1 - Smartlab - Non-Linear Complex Systems Laboratory  
 DITEN - Università degli Studi di Genova, Genoa (I-16145), Italy.  
 2 - CETeD - Technical Research Centre for Dependency Care and Autonomous Living  
 Universitat Politècnica de Catalunya (BarcelonaTech), Vilafranca i la Geltrú (08800), Spain  
 activityrecognition @' smartlab.ws



# Supervised Learning Example



Machine Learning:  
Learn Patterns  
from Data



**Enum = Categorical Value (in Java)**

The screenshot shows the H2O Flow web interface. In the top navigation bar, the title 'H2O FLOW' is visible along with various menu items: Flow, Cell, Data, Model, Score, Admin, and Help. The current page is 'Untitled Flow'. Below the navigation is a toolbar with icons for file operations like Open, Save, and Print, as well as other flow-related functions.

The main area is titled 'Setup Parse' under 'PARSE CONFIGURATION'. It includes the following settings:

- Sources: smartphone\_sensors\_train.csv
- ID: Key\_Frame\_\_smartphone\_sensors\_train.hex
- Parser: CSV (highlighted by a yellow callout box)
- Separator: ;,'44'
- Column Headers: Auto (radio button)
- First row contains column names (radio button, selected)
- First row contains data (radio button)
- Options: Enable single quotes as a field quotation character (checkbox)
- Delete on done (checkbox, checked)

Below this is the 'EDIT COLUMN NAMES AND TYPES' section. It lists 11 columns with their current types and values:

Index	Column Name	Type	Value 1	Value 2	Value 3	Value 4	Value 5	Value 6	Value 7	Value 8	Value 9
1	activity	Enum	STANDING	STANDING	STANDING	STANDING	STANDING	STANDING	STANDING	STANDING	STANDING
2	f1_tBodyAccmeanX	Numeric	0.28858451	0.27841883	0.27965306	0.27917394	0.27662877	0.27719877	0.27945388	0.27743247	0.27729342
3	f2_tBodyAccmeanY	Numeric	-0.020294171	-0.016410568	-0.019467156	-0.026200646	-0.016569655	-0.01009785	-0.019640776	-0.030488303	-0.021750698
4	f3_tBodyAccmeanZ	Numeric	-0.13290514	-0.12352019	-0.11346169	-0.12328257	-0.11536185	-0.10513725	-0.11002215	-0.12536043	-0.12075082
5	f4_tBodyAccstdX	Numeric	-0.9952786	-0.99824528	-0.99537956	-0.99609149	-0.99813862	-0.99733496	-0.99692104	-0.99655926	-0.99732847
6	f5_tBodyAccstdY	Numeric	-0.98311061	-0.97530022	-0.96718701	-0.9834027	-0.98081727	-0.99048681	-0.96718593	-0.96672843	-0.96124532
7	f6_tBodyAccstdZ	Numeric	-0.91352645	-0.96032199	-0.97894396	-0.9906751	-0.99048163	-0.99542003	-0.98311783	-0.98158533	-0.98367156
8	f7_tBodyAccmadX	Numeric	-0.99511208	-0.99880719	-0.99651994	-0.99709947	-0.99832113	-0.9976274	-0.99700268	-0.99648525	-0.99759576

At the bottom left, there is a status message 'Ready' with a green dot. On the right side, it says 'Connections: 0' and has the H2O logo.

H2O Flow Jo-fai

localhost:54321/flow/index.html#

Untitled Flow

Key\_Frame\_smartphone\_sensors\_train.hex

Actions: View Data Split... Build Model... Predict Download Export Delete

	Rows	Columns	Compressed Size
	7352	562	25MB

▼ COLUMN SUMMARIES

label	type	Missing	Zeros	+Inf	-Inf	min	max	mean	sigma	cardinality	Actions
activity	enum	0	1407	0	0	0	5.0	.	.	6	Convert to numeric
f1_tBodyAccmeanX	real	0	0	0	0	-1.0	1.0	0.2745	0.0703	.	.
f2_tBodyAccmeanY	real	0	0	0	0	-1.0	1.0	-0.0177	0.0408	.	.
f3_tBodyAccmeanZ	real	0	0	0	0	-1.0	1.0	-0.1091	0.0566	.	.
f4_tBodyAccstdX	real	0	0	0	0	-1.0	1.0	-0.6054	0.4487	.	.
f5_tBodyAccstdY	real	0	0	0	0	-0.9999	0.9162	-0.5109	0.5026	.	.
f6_tBodyAccstdZ	real	0	0	0	0	-1.0	1.0	-0.6048	0.4187	.	.
f7_tBodyAccmadX	real	0	0	0	0	-1.0	1.0	-0.6305	0.4241	.	.
f8_tBodyAccmadY	real	0	0	0	0	-1.0	0.9677	-0.5269	0.4859	.	.
f9_tBodyAccmadZ	real	0	0	0	0	-1.0	1.0	-0.6062	0.4141	.	.
f10_tBodyAccmaxX	real	0	0	0	0	-1.0	1.0	-0.4686	0.5445	.	.
f11_tBodyAccmaxY	real	0	0	0	0	-1.0	1.0	-0.3060	0.2822	.	.
f12_tBodyAccmaxZ	real	0	0	0	0	-1.0	1.0	-0.5571	0.2939	.	.
f13_tBodyAccminX	real	0	0	0	0	-1.0	1.0	0.5236	0.3636	.	.
f14_tBodyAccminY	real	0	0	0	0	-1.0	1.0	0.3874	0.3436	.	.
f15_tBodyAccminZ	real	0	0	0	0	-1.0	1.0	0.5944	0.2978	.	.
f16_tBodyAccsma	real	0	0	0	0	-1.0	1.0	-0.5476	0.4718	.	.
f17_tBodyAccenergyX	real	0	0	0	0	-1.0	1.0	-0.8200	0.2596	.	.
f18_tBodyAccenergyY	real	0	0	0	0	-1.0	1.0	-0.9019	0.1263	.	.
f19_tBodyAccenergyZ	real	0	0	0	0	-1.0	1.0	-0.8458	0.2220	.	.

Ready Connections: 0 H2O

Exploratory analysis  
(click on any variable)

H2O Flow Jo-fai

localhost:54321/flow/index.html#

**Untitled Flow**

Flow ▾ Cell ▾ Data ▾ Model ▾ Score ▾ Admin ▾ Help ▾

File ▾ + New ▾ Open ▾ Save ▾ Undo ▾ Redo ▾ Copy ▾ Paste ▾ Delete ▾ Find ▾ Filter ▾

◀ Previous 20 Columns ▶ Next 20 Columns

▶ CHUNK COMPRESSION SUMMARY

▶ FRAME DISTRIBUTION SUMMARY

CS | getColumnSummary "Key\_Frame\_\_smartphone\_sensors\_train.hex", "activity"

133ms

**Summary: activity**

**Actions** [Impute](#) [Inspect](#)

**CHARACTERISTICS**

All

avg(percent)

**DOMAIN (MAX 1000 LEVELS)**

label

label	count
LAYING	~1300
STANDING	~1300
SITTING	~1200
WALKING	~1200
WALKING_UPSTAIRS	~1000
WALKING_DOWNSTAIRS	~1000

Connections: 0 H2O

H2O Flow Jo-fai

localhost:54321/flow/index.html#

H2O FLOW Flow Cell Data Model Score Admin Help

Untitled Flow

buildModel "drf"

Build a Model

Select an algorithm: Distributed Random Forest

PARAMETERS

model\_id my\_random\_forest  
training\_frame Key\_Frame\_smartphone\_sensors\_train.hex  
validation\_frame (Choose...)  
nfolds 0  
response\_column activity  
ignored\_columns Search...

GRID?

Destination id for this model; auto-generated if not specified.

Id of the training data frame (Not required, to allow initial validation of model parameters).

Id of the validation data frame.

Number of folds for N-fold cross-validation (0 to disable or >= 2).

Response variable column.

Showing page 1 of 6.

activity	ENUM(6)
f1_tBodyAccmeanX	REAL
f2_tBodyAccmeanY	REAL
f3_tBodyAccmeanZ	REAL
f4_tBodyAccstdX	REAL
f5_tBodyAccstdY	REAL

Ready Connections: 0 H2O

## Build Random Forest Model

- 1) Select "...\_train" dataset
- 2) Select "activity" as response

H2O Flow Jo-fai

localhost:54321/flow/index.html#

Untitled Flow

**Model**

Model ID: my\_random\_forest  
Algorithm: Distributed Random Forest

Actions: Refresh, Predict..., Download POJO, Download Model Deployment Package (MOJO), Export, Inspect, Delete, Download Gen Model

MODEL PARAMETERS

SCORING HISTORY - LOGLOSS

A line graph titled "SCORING HISTORY - LOGLOSS". The y-axis is labeled "training\_logloss" and ranges from 0.0 to 4.0. The x-axis is labeled "number\_of\_trees" and ranges from 0 to 50. The data points show a rapid decrease in logloss as the number of trees increases, starting around 3.5 at 0 trees and approaching 0.0 at 50 trees.

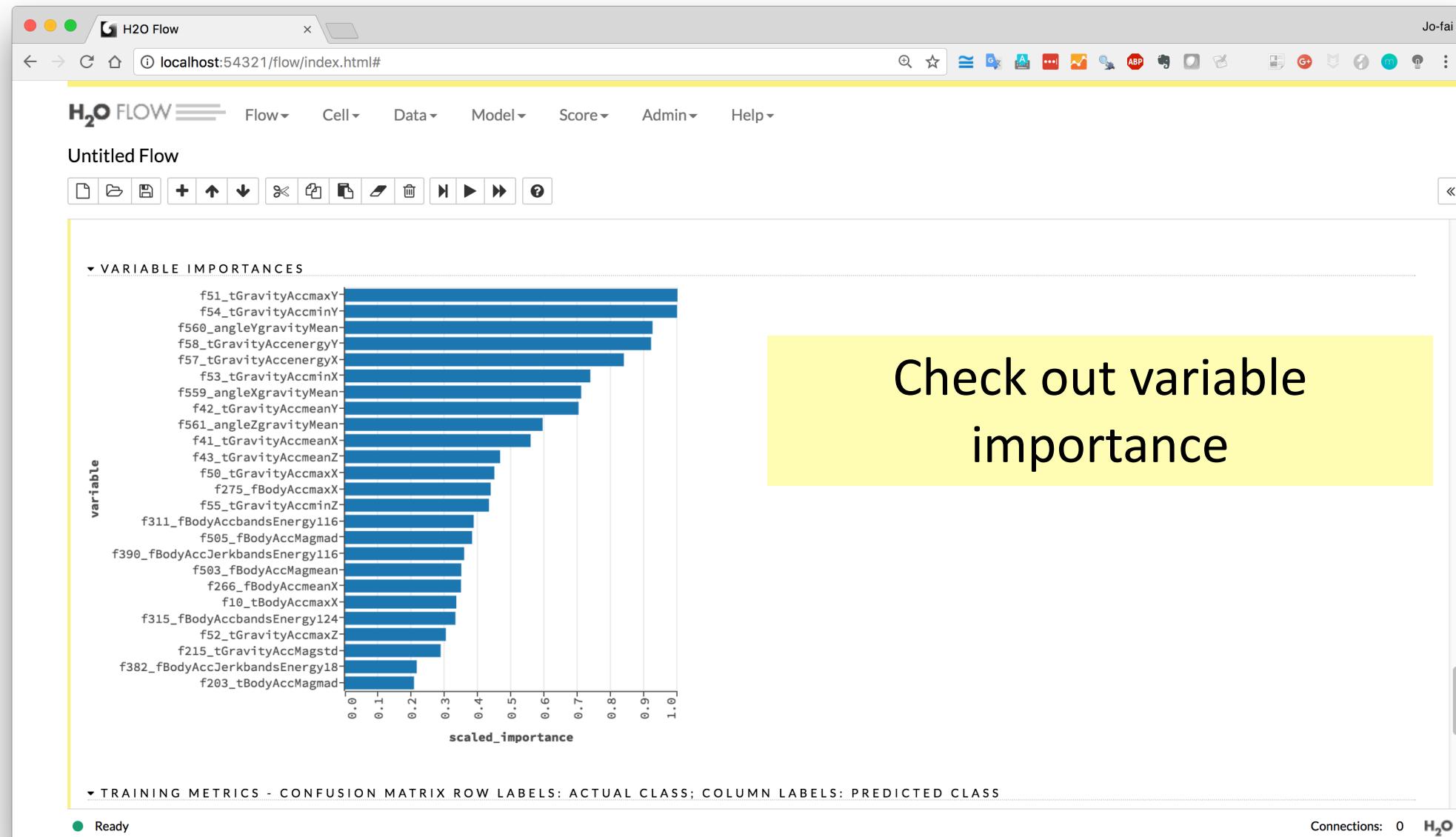
number_of_trees	training_logloss
0	3.5
1	3.0
2	2.5
3	2.0
4	1.6
5	1.3
6	1.1
7	0.9
8	0.7
9	0.5
10	0.4
15	0.2
20	0.15
25	0.1
30	0.08
35	0.06
40	0.05
45	0.04
50	0.03

VARIABLE IMPORTANCES

Ready

Connections: 0 H2O

Performance metric  
Logarithmic Loss (logloss)  
A loss function for classification



H2O Flow Jo-fai

localhost:54321/flow/index.html#

H2O FLOW Flow Cell Data Model Score Admin Help

Untitled Flow

TRAINING METRICS - CONFUSION MATRIX ROW LABELS: ACTUAL CLASS; COLUMN LABELS: PREDICTED CLASS

	LAYING	SITTING	STANDING	WALKING	WALKING_DOWNSTAIRS	WALKING_UPSTAIRS	Error	Rate
<b>LAYING</b>	1407	0	0	0	0	0	0	0 / 1,407
<b>SITTING</b>	0	1216	69	0	0	1	0.0544	70 / 1,286
<b>STANDING</b>	0	43	1331	0	0	0	0.0313	43 / 1,374
<b>WALKING</b>	0	0	2	1213	7	4	0.0106	13 / 1,226
<b>WALKING_DOWNSTAIRS</b>	0	1	0	5	970	10	0.0162	16 / 986
<b>WALKING_UPSTAIRS</b>	0	0	1	1	5	1066	0.0065	7 / 1,073
<b>Total</b>	1407	1260	1403	1219	982	1081	0.0203	149 / 7,352

OUTPUT

OUTPUT - MODEL SUMMARY

OUTPUT - SCORING HISTORY

OUTPUT - TRAINING\_METRICS

OUTPUT - TRAINING\_METRICS - TOP-6 HIT RATIOS

OUTPUT - VARIABLE IMPORTANCES

PREFVIEW POJO

Ready Connections: 0 H2O

Confusion Matrix  
(Yellow cells = correct predictions)

H2O Flow Jo-fai

localhost:54321/flow/index.html#

H2O FLOW Flow Cell Data Model Score Admin Help

Untitled Flow

OUTPUT - TRAINING\_METRICS - TOP-6 HIT RATIOS

OUTPUT - VARIABLE IMPORTANCES

PREVIEW POJO

Preview POJO

Make predictions on test set

CS predict 41ms

**Predict**

Name: my\_predictions

Model: my\_random\_forest

Frame: Key\_Frame\_smartphone\_sensors\_test.hex

Actions: ⚡ Predict

Ready Connections: 0 H2O

The screenshot shows the H2O Flow web application. At the top, there's a navigation bar with tabs for Flow, Cell, Data, Model, Score, Admin, and Help. Below the navigation is a toolbar with various icons for file operations like Open, Save, and Print, as well as navigation controls like back, forward, and search. A large yellow callout box on the right side contains the text "Make predictions on test set". In the main workspace, there's a step labeled "predict" with a "CS" icon. Below it, there's a "Predict" section with fields for "Name" (set to "my\_predictions"), "Model" (set to "my\_random\_forest"), "Frame" (set to "Key\_Frame\_smartphone\_sensors\_test.hex"), and an "Actions" button labeled "⚡ Predict". The bottom status bar indicates the step is "Ready" and shows connection information: "Connections: 0" and the H2O logo.

H2O Flow Jo-fai

localhost:54321/flow/index.html#

**Compare Results**

Untitled Flow

**combined-my\_predictions**

**Predictions (with prob. of each class)**

◀ Previous 20 Columns ▶ Next 20 Columns

Row	predict	LAYING	SITTING	STANDING	WALKING	WALKING_DOWNSTAIRS	WALKING_UPSTAIRS	activity	f1_tBodyAccmeanX	f2_tBodyAccmeanY	f3_tBodyAccmeanZ	f4_t
1	STANDING	0.0401	0.0401	0.9199	0	0	0	STANDING	0.2572	-0.0233	-0.0147	
2	STANDING	0	0.1698	0.8302	0	0	0	STANDING	0.2860	-0.0132	-0.1191	
3	STANDING	0.0182	0.2000	0.7818	0	0	0	STANDING	0.2755	-0.0261	-0.1182	
4	STANDING	0.0408	0.0816	0.8776	0	0	0	STANDING	0.2703	-0.0326	-0.1175	
5	STANDING	0	0.0579	0.9421	0	0	0	STANDING	0.2748	-0.0278	-0.1295	
6	STANDING	0	0.0394	0.9606	0	0	0	STANDING	0.2792	-0.0186	-0.1139	
7	STANDING	0	0.0965	0.9035	0	0	0	STANDING	0.2797	-0.0183	-0.1040	
8	STANDING	0	0.0446	0.9554	0	0	0	STANDING	0.2746	-0.0250	-0.1168	
9	STANDING	0	0.1429	0.8571	0	0	0	STANDING	0.2725	-0.0210	-0.1145	
10	STANDING	0	0.1569	0.8431	0	0	0	STANDING	0.2757	-0.0104	-0.0998	
11	STANDING	0	0.1667	0.8333	0	0	0	STANDING	0.2786	-0.0152	-0.0989	
12	STANDING	0.0179	0.1607	0.8214	0	0	0	STANDING	0.2792	-0.0219	-0.1097	
13	STANDING	0	0	1.0	0	0	0	STANDING	0.2745	-0.0231	-0.1125	
14	STANDING	0	0.1325	0.8675	0	0	0	STANDING	0.2691	-0.0277	-0.1102	
15	STANDING	0	0.1115	0.8885	0	0	0	STANDING	0.2756	-0.0189	-0.0974	

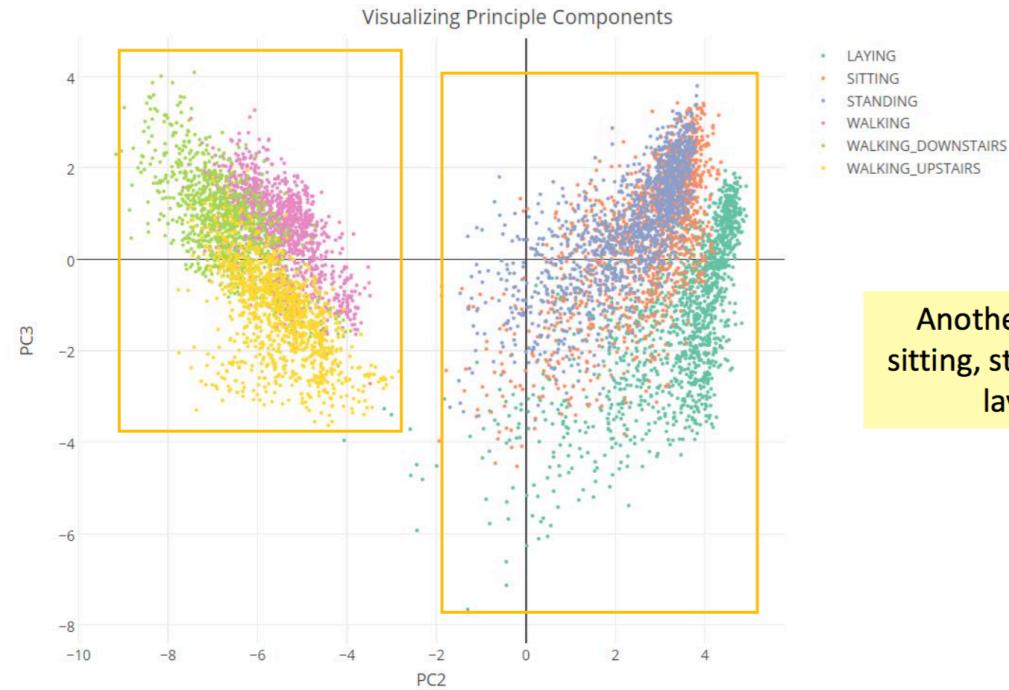
● Ready

Connections: 0 H2O

**Ground Truth**

# H<sub>2</sub>O Demo at IBM Conference

First, we can see two major clusters. One for movement (walking, walking upstairs and walking downstairs)



From the graph above, we can see that:

- it could be difficult to distinguish between **Standing** and **Sitting** as there are large overlaps in their sensor data.
- Laying has its own cluster so it should be easy to classify.
- Walking, Walking Upstairs and Walking Downstairs are understandably closer to each other yet they are quite different to Sitting, Standing and Laying.

Another one for sitting, standing and laying.

IBM Fast Track Your Data (ML Conference)  
Munich



[https://github.com/woobe/h2o\\_demo\\_for\\_ibm\\_dsx](https://github.com/woobe/h2o_demo_for_ibm_dsx)