Deploying Predictive Models in the Cloud using Yhat

Luke Sheneman
NKN Technology and Data Services Manager

NKN Tech Talk
December 17, 2015

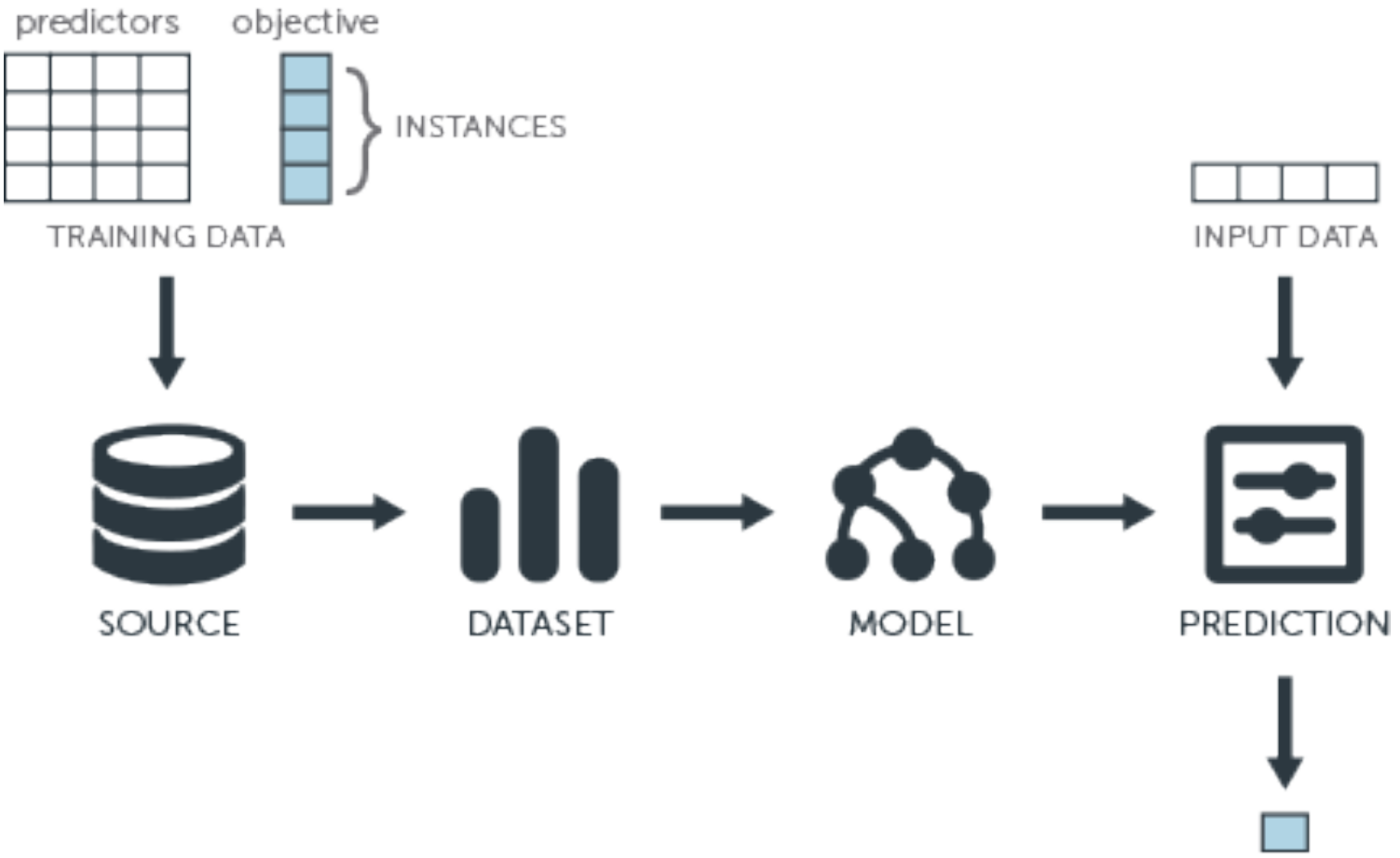

Statistic
Visualisation
Pattern recognit.
DATA science
Data Mining
Machine Learning
Neurocomputing
Database & Data procss

Machine Learning:
Training Statistical Predictive Models in R

topepo.github.io/caret/

- caret (*Classification And Regression Training*) is a powerful R package for machine learning
  - Data Splitting
  - Training
  - Cross Validation
  - Prediction

Model Training and Prediction



Real Example – Wearable Sensor Data

github.com/sheneman/machinelearning

- Use *quantitative* raw data from wearable sensors to *qualitatively* classify human motion.
  - Velloso, E.; Bulling, A.; Gellersen, H.; Ugulino, W.; Fuks, H. Qualitative Activity Recognition of Weight Lifting Exercises. Proceedings of 4th International Conference in Cooperation with SIGCHI (Augmented Human '13) . Stuttgart, Germany: ACM SIGCHI, 2013. http://groupware.les.inf.puc-rio.br/har

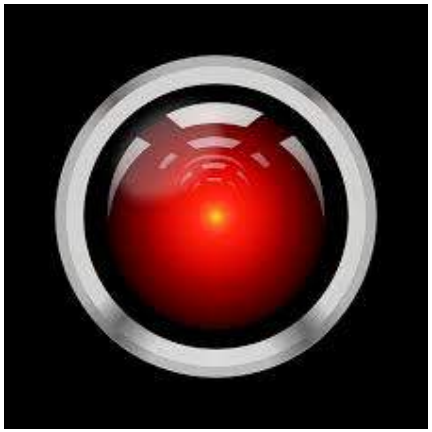
- 160 variables, 20K observations
  - Predictors:
    - Subject ID, roll, pitch, yaw, acceleration, time
      - mean, stdev, kurtosis, skewness
  - Outcomes/objectives:
    - Classification: A,B,C,D,E

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1		user_name	raw_timestamp_part_	raw_timesta	cvtd_timestamp	new_wi	num_window	roll_belt	pitch_belt	yaw_belt	total_accel_b	kurtosis_roll	kurtosis_pict	kurtosis_yaw	skewness_ro	skewness_ro	skewness_ya	max_roll_bel	max_picth_b	max_yaw_be	min_roll_bel	min_pitch_b	min_yaw_be	amplitude_r	amplitude_p	amplitude_y	var_total_
2	1	carlitos	1323084231	788290	5/12/11 11:23	no	11	1.41	8.07	-94.4	3							NA	NA		NA	NA		NA	NA		NA
3	2	carlitos	1323084231	808298	5/12/11 11:23	no	11	1.41	8.07	-94.4	3							NA	NA		NA	NA		NA	NA		NA
4	3	carlitos	1323084231	820366	5/12/11 11:23	no	11	1.42	8.07	-94.4	3							NA	NA		NA	NA		NA	NA		NA
5	4	carlitos	1323084232	120339	5/12/11 11:23	no	12	1.48	8.05	-94.4	3							NA	NA		NA	NA		NA	NA		NA
6	5	carlitos	1323084232	196328	5/12/11 11:23	no	12	1.48	8.07	-94.4	3							NA	NA		NA	NA		NA	NA		NA
7	6	carlitos	1323084232	304277	5/12/11 11:23	no	12	1.45	8.06	-94.4	3							NA	NA		NA	NA		NA	NA		NA
8	7	carlitos	1323084232	368296	5/12/11 11:23	no	12	1.42	8.09	-94.4	3							NA	NA		NA	NA		NA	NA		NA
9	8	carlitos	1323084232	440390	5/12/11 11:23	no	12	1.42	8.13	-94.4	3							NA	NA		NA	NA		NA	NA		NA
10	9	carlitos	1323084232	484323	5/12/11 11:23	no	12	1.43	8.16	-94.4	3							NA	NA		NA	NA		NA	NA		NA
11	10	carlitos	1323084232	484434	5/12/11 11:23	no	12	1.45	8.17	-94.4	3							NA	NA		NA	NA		NA	NA		NA
12	11	carlitos	1323084232	500302	5/12/11 11:23	no	12	1.45	8.18	-94.4	3							NA	NA		NA	NA		NA	NA		NA
13	12	carlitos	1323084232	528316	5/12/11 11:23	no	12	1.43	8.18	-94.4	3							NA	NA		NA	NA		NA	NA		NA
14	13	carlitos	1323084232	560359	5/12/11 11:23	no	12	1.42	8.2	-94.4	3							NA	NA		NA	NA		NA	NA		NA
15	14	carlitos	1323084232	576390	5/12/11 11:23	no	12	1.42	8.21	-94.4	3							NA	NA		NA	NA		NA	NA		NA
16	15	carlitos	1323084232	604281	5/12/11 11:23	no	12	1.45	8.2	-94.4	3							NA	NA		NA	NA		NA	NA		NA
17	16	carlitos	1323084232	644302	5/12/11 11:23	no	12	1.48	8.15	-94.4	3							NA	NA		NA	NA		NA	NA		NA
18	17	carlitos	1323084232	692324	5/12/11 11:23	no	12	1.51	8.12	-94.4	3							NA	NA		NA	NA		NA	NA		NA
19	18	carlitos	1323084232	732306	5/12/11 11:23	no	12	1.55	8.08	-94.4	3							NA	NA		NA	NA		NA	NA		NA
20	19	carlitos	1323084232	740353	5/12/11 11:23	no	12	1.57	8.06	-94.4	3							NA	NA		NA	NA		NA	NA		NA
21	20	carlitos	1323084232	788335	5/12/11 11:23	no	12	1.59	8.07	-94.4	3							NA	NA		NA	NA		NA	NA		NA
22	21	carlitos	1323084232	876301	5/12/11 11:23	no	12	1.6	8.1	-94.4	3							NA	NA		NA	NA		NA	NA		NA
23	22	carlitos	1323084232	892313	5/12/11 11:23	no	12	1.57	8.09	-94.4	3							NA	NA		NA	NA		NA	NA		NA
24	23	carlitos	1323084232	932285	5/12/11 11:23	no	12	1.56	8.1	-94.3	3							NA	NA		NA	NA		NA	NA		NA
25	24	carlitos	1323084232	996313	5/12/11 11:23	yes	12	1.51	8.1	-94.4	3	5.587755	#DIV/0!	#DIV/0!	2.713152	#DIV/0!	#DIV/0!	-94.3	3	5.6	-94.4	3	5.6	0.1	0	0	
26	25	carlitos	1323084233	28311	5/12/11 11:23	no	13	1.53	8.11	-94.4	3							NA	NA		NA	NA		NA	NA		NA
27	26	carlitos	1323084233	56286	5/12/11 11:23	no	13	1.55	8.09	-94.4	3							NA	NA		NA	NA		NA	NA		NA
28	27	carlitos	1323084233	72305	5/12/11 11:23	no	13	1.54	8.11	-94.4	3							NA	NA		NA	NA		NA	NA		NA
29	28	carlitos	1323084233	120363	5/12/11 11:23	no	13	1.53	8.14	-94.4	3							NA	NA		NA	NA		NA	NA		NA
30	29	carlitos	1323084233	136333	5/12/11 11:23	no	13	1.52	8.16	-94.4	3							NA	NA		NA	NA		NA	NA		NA
31	30	carlitos	1323084233	144318	5/12/11 11:23	no	13	1.52	8.17	-94.4	3							NA	NA		NA	NA		NA	NA		NA
32	31	carlitos	1323084233	152353	5/12/11 11:23	no	13	1.53	8.17	-94.4	3							NA	NA		NA	NA		NA	NA		NA
33	32	carlitos	1323084233	244310	5/12/11 11:23	no	13	1.44	8.19	-94.4	3							NA	NA		NA	NA		NA	NA		NA
34	33	carlitos	1323084233	304292	5/12/11 11:23	no	13	1.43	8.17	-94.4	3							NA	NA		NA	NA		NA	NA		NA
35	34	carlitos	1323084233	332328	5/12/11 11:23	no	13	1.44	8.18	-94.3	3							NA	NA		NA	NA		NA	NA		NA
36	35	carlitos	1323084233	340386	5/12/11 11:23	no	13	1.41	8.18	-94.4	3							NA	NA		NA	NA		NA	NA		NA
37	36	carlitos	1323084233	412295	5/12/11 11:23	no	13	1.42	8.12	-94.3	3							NA	NA		NA	NA		NA	NA		NA
38	37	carlitos	1323084233	448310	5/12/11 11:23	no	13	1.41	8.11	-94.3	3							NA	NA		NA	NA		NA	NA		NA
39	38	carlitos	1323084233	468406	5/12/11 11:23	no	13	1.4	8.04	-94.3	3							NA	NA		NA	NA		NA	NA		NA
40	39	carlitos	1323084233	504293	5/12/11 11:23	no	13	1.4	8.04	-94.3	3							NA	NA		NA	NA		NA	NA		NA
41	40	carlitos	1323084233	512347	5/12/11 11:23	no	13	1.4	8.05	-94.3	3							NA	NA		NA	NA		NA	NA		NA
42	41	carlitos	1323084233	512367	5/12/11 11:23	no	13	1.4	8.06	-94.3	3							NA	NA		NA	NA		NA	NA		NA
43	42	carlitos	1323084233	572298	5/12/11 11:23	no	13	1.39	8.05	-94.3	3							NA	NA		NA	NA		NA	NA		NA
44	43	carlitos	1323084233	572351	5/12/11 11:23	no	13	1.34	8.05	-94.3	3							NA	NA		NA	NA		NA	NA		NA
45	44	carlitos	1323084233	604309	5/12/11 11:23	no	13	1.3	8	-94.2	3							NA	NA		NA	NA		NA	NA		NA
46	45	carlitos	1323084233	660300	5/12/11 11:23	no	13	1.3	7.85	-94.2	3							NA	NA		NA	NA		NA	NA		NA
47	46	carlitos	1323084233	676353	5/12/11 11:23	no	13	1.29	7.81	-94.2	3							NA	NA		NA	NA		NA	NA		NA
48	47	carlitos	1323084233	688317	5/12/11 11:23	no	13	1.33	7.76	-94.2	3							NA	NA		NA	NA		NA	NA		NA
49	48	carlitos	1323084233	736343	5/12/11 11:23	no	13	1.33	7.69	-94.2	3							NA	NA		NA	NA		NA	NA		NA
50	49	carlitos	1323084233	784365	5/12/11 11:23	no	13	1.31	7.69	-94.2	3							NA	NA		NA	NA		NA	NA		NA
51	50	carlitos	1323084233	840318	5/12/11 11:23	no	13	1.3	7.68	-94.2	3							NA	NA		NA	NA		NA	NA		NA
52	51	carlitos	1323084233	937221	5/12/11 11:23	no	13	1.29	7.58	-94.1	3							NA	NA		NA	NA		NA	NA		NA
53	52	carlitos	1323084233	948290	5/12/11 11:23	yes	13	1.27	7.56	-94.1	3	-0.99713	#DIV/0!	#DIV/0!	0.007518	#DIV/0!	#DIV/0!	-94.1	3	-1	-94.4	3	-1	0.3	0	0	
54	53	carlitos	1323084234	399	5/12/11 11:23	no	14	1.26	7.54	-94.1	3							NA	NA		NA	NA		NA	NA		NA
55	54	carlitos	1323084234	64349	5/12/11 11:23	no	14	1.25	7.46	-94.1	3							NA	NA		NA	NA		NA	NA		NA
56	55	carlitos	1323084234	128295	5/12/11 11:23	no	14	1.26	7.47	-94.2	3							NA	NA		NA	NA		NA	NA		NA
57	56	carlitos	1323084234	140346	5/12/11 11:23	no	14	1.27	7.46	-94.1	3							NA	NA		NA	NA		NA	NA		NA
58	57	carlitos	1323084234	248292	5/12/11 11:23	no	14	1.25	7.45	-94.1	3							NA	NA		NA	NA		NA	NA		NA
59	58	carlitos	1323084234	268317	5/12/11 11:23	no	14	1.23	7.43	-94.1	3							NA	NA		NA	NA		NA	NA		NA
60	59	carlitos	1323084234	340471	5/12/11 11:23	no	14	1.24	7.43	-94.1	3							NA	NA		NA	NA		NA	NA		NA

Training the Models

Method #1: Recursive Partitioning and Regression Trees (RPART):

```
control <- trainControl(method="cv", number=5)
modelfit <- train(classe ~ ., data=new_training_data,
                  trControl=control, method="rpart")
```

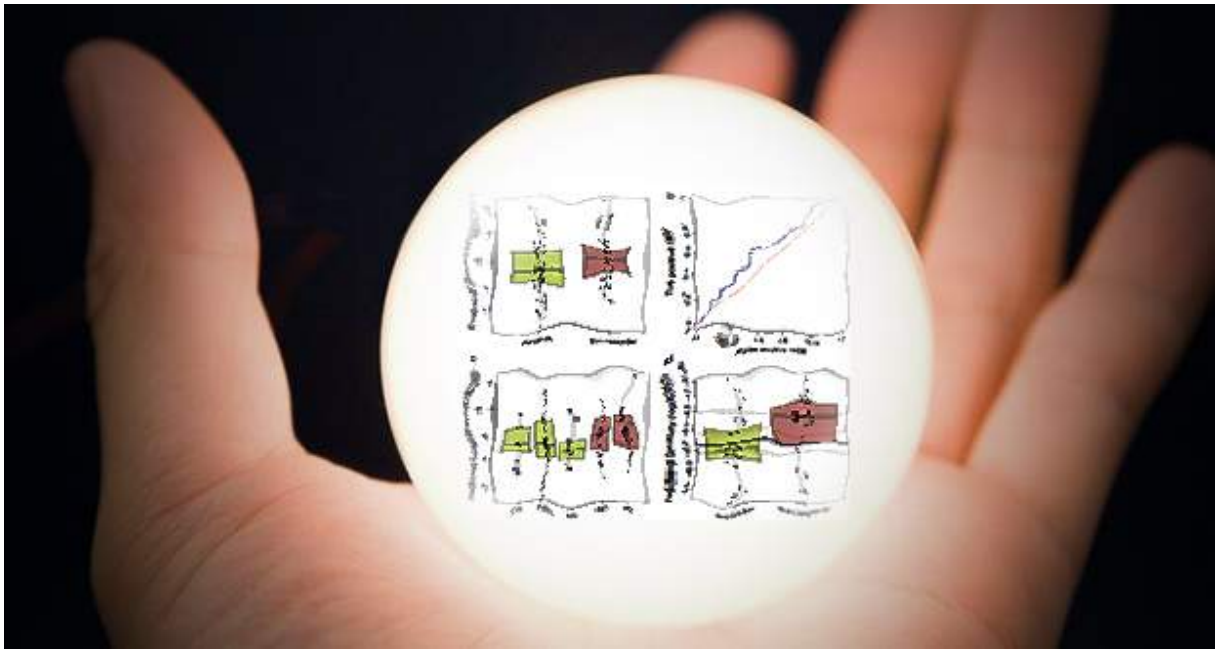


Method #2: Bootstrap Aggregation (BAGGING):

```
control <- trainControl(method="cv", number=5, verboseIter=TRUE)
modelfit <- train(classe ~ ., data=new_training_data,
                  trControl=control, method="treebag")
```

Prediction from Trained Models

predictions <- ***predict***(modelfit, new_training_data)



Results

RPART Method

Aggregating results
Selecting tuning parameters
Fitting cp = 0.0395 on full training set
Confusion Matrix and Statistics

Prediction	Reference A	B	C	D	E
A	5080	1581	1587	1449	524
B	81	1286	108	568	486
C	405	930	1727	1199	966
D	0	0	0	0	0
E	14	0	0	0	1631

Overall Statistics

Accuracy : 0.4956
95% CI : (0.4885, 0.5026)
No Information Rate : 0.2844
P-Value [Acc > NIR] : < 2.2e-16

BAGGING Method

Aggregating results
Fitting final model on full training set
Confusion Matrix and Statistics

Prediction	Reference A	B	C	D	E
A	5580	0	0	0	0
B	0	3797	**1**	0	0
C	0	0	3421	0	0
D	0	0	0	3216	0
E	0	0	0	0	3607

Overall Statistics

Accuracy : 0.9999
95% CI : (0.9997, 1)
No Information Rate : 0.2844
P-Value [Acc > NIR] : < 2.2e-16

Python IDE for Data Science

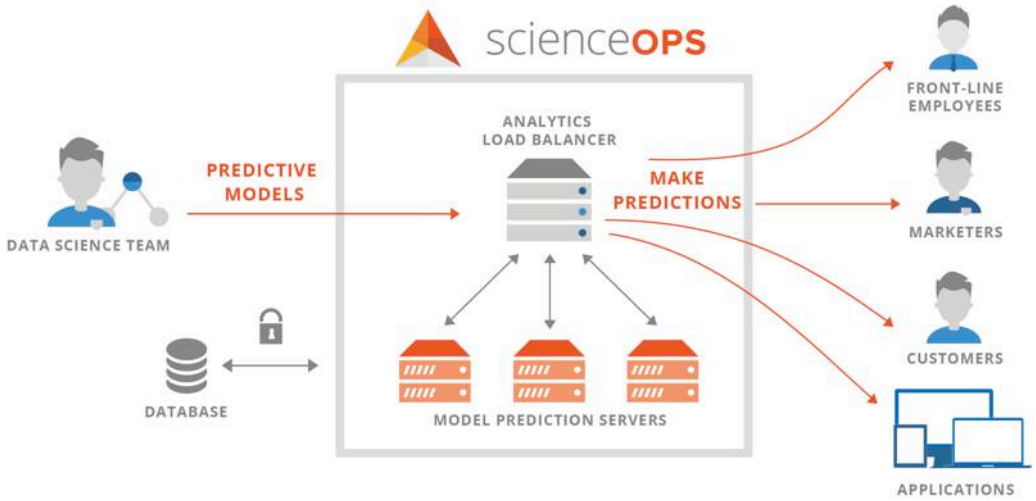
Publish Predictive Models

Distributed Computing Foo

scienceOPS

WHAT IS SCIENCEOPS?

ScienceOps is a platform for deploying, managing and scaling predictive models and advanced decision-making algorithms into production. The platform is designed for production-level settings and utilizes a distributed architecture to allocate and optimize requests to models across the cluster.



yhatr

- https://github.com/yhat/yhatr
  - implictly requires ***rjson*** package

- model.require()
  - specify library dependencies
- model.tranform()
  - transform input data
- model.predict()
  - call the prediction function

Hello World in yhat

R

```r
library(yhatr)

model.predict <- function(request) {
  me <- request$name
  greeting<-paste ("Hello", me, "!")
  data.frame(greeting)
}

yhat.config  <- c(
  username="YOUR_USERNAME",
  apikey="YOUR_APIKEY",
  env="https://sandbox.yhathq.com/"
)
yhat.deploy("HelloWorld")
```

Python

```python
from yhat import Yhat, YhatModel , preprocess

class HelloWorld(YhatModel):
    @preprocess(in_type=dict, out_type=dict)
    def execute(self, data):
        me = data['name']
        greeting = "Hello " + str(me) + "!"
        return { "greeting": greeting }

yh = Yhat("YOUR_USERNAME", "YOUR_APIKEY", "https://sandbox.yhathq.com/")
yh.deploy("HelloWorld", HelloWorld, globals())
```

* In R, all input and output to model.predict() must be a *Data Frame* type

Examples:

- Hello World!
- Custom NKN Example
- Beer Selector

ŷhat Applications

- MaaS – *Model as a Service* paradigm
- REST API
  - Build applications against standard API
  - Simulate 3rd party app development from your model
  - Deploy to one or millions of users
  - A Form of Data/Model Interoperability:
    - JSON + REST Web Service
    - Build interesting composite apps using predictions from multiple models!
- Auto re-deploy model as additional data streams in:
  - Weather, Climate, Precision Agriculture, Hazard Prediction, Climate Hot Spots, Social Media
- Horizontally Scalable on Cloud Infrastructure

*"Even if NKN does not use Yhat specifically, this kind of idea is a wonderful example of the **NKN Data Observatory** concept in action. Namely, Yhat can leverage existing data to do **new** science via **data/model <u>interoperability</u>**.*

Exposing standard web service APIs are the key."

-- Famous Olde Quote

Thank You!