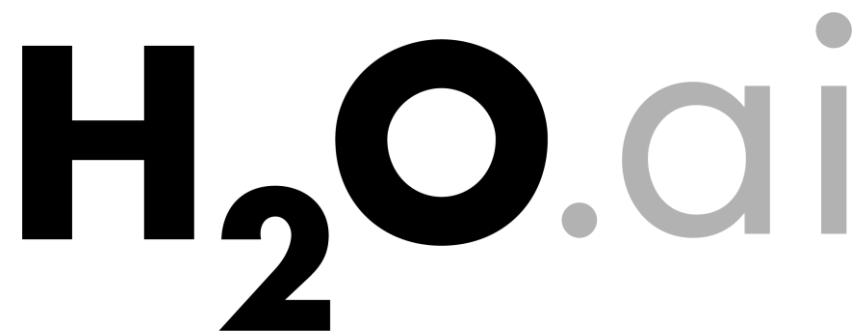


# Simple Machine Learning Example with H<sub>2</sub>O Flow (Web) Interface



Jo-fai (Joe) Chow

Data Scientist

joe@h2o.ai

@matlabulous

# About this Presentation

- Data
  - Credit Card Default Payment
- H<sub>2</sub>O Flow Example
  - Download/Install H<sub>2</sub>O
  - H<sub>2</sub>O Flow Interface
  - Import Dataset
  - Exploratory Analysis
  - Train a Model
  - Make Predictions



# Credit Card Default Payment Data

[https://github.com/woobe/credit\\_card\\_demo](https://github.com/woobe/credit_card_demo)

**default\_payment\_training\_data.csv**

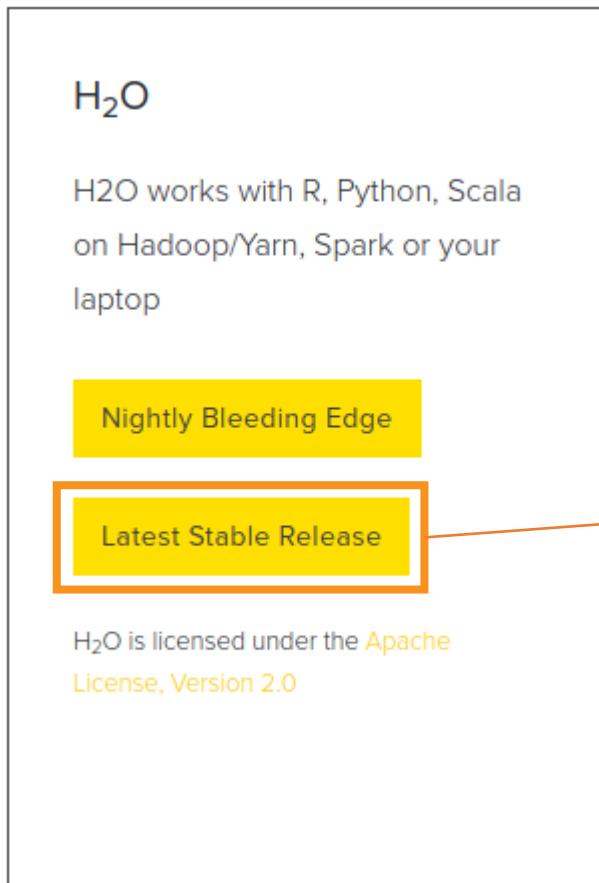
**default\_payment\_test\_data.csv**

default_payment_training_data.csv - Excel																										
default_payment_training_data.csv																										
AB99																										
1	DEFAULT_PAYMENT_NEXT_MONTH	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X		
2	Yes	20000	Female	2	1	24	2	2	-1	-1	-2	-2	3913	3102	689	0	0	0	0	689	0	0	0	0	0	0
3	Yes	120000	Female	2	2	26	-1	2	0	0	0	0	2682	1725	2682	3272	3455	3261	0	1000	1000	1000	1000	0	2000	
4	No	90000	Female	2	2	34	0	0	0	0	0	0	29239	14027	13559	14331	14948	15549	1518	1500	1000	1000	1000	1000	5000	
5	No	50000	Female	2	1	37	0	0	0	0	0	0	46990	48233	49291	28314	28959	29547	2000	2019	1200	1100	1069	1000	1000	1000
6	No	50000	Male	2	1	57	-1	0	-1	0	0	0	8617	5670	35835	20940	19146	19131	2000	36681	10000	9000	689	679	679	
7	No	100000	Female	2	2	23	0	-1	-1	0	0	-1	11876	380	601	221	-159	567	380	601	0	581	1687	1542	1542	
8	No	140000	Female	3	1	28	0	0	2	0	0	0	11285	14096	12108	12211	11793	3719	3329	0	432	1000	1000	1000	1000	1000
9	No	20000	Male	3	2	35	-2	-2	-2	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
10	No	200000	Female	3	2	34	0	0	2	0	0	-1	11073	9787	5535	2513	1828	3731	2306	12	50	300	3738	66	66	
11	No	630000	Female	2	2	41	-1	0	-1	-1	-1	-1	12137	6500	6500	6500	2870	1000	6500	6500	6500	2870	0	0	0	0
12	Yes	70000	Male	2	2	30	1	2	2	0	0	2	65802	67369	65701	66782	36137	36894	3200	0	3000	3000	1500	0	0	0
13	No	250000	Male	1	2	29	0	0	0	0	0	0	70887	67060	63561	59696	56875	55512	3000	3000	3000	3000	3000	3000	3000	3000
14	No	50000	Female	3	3	23	1	2	0	0	0	0	50614	29173	28116	28771	29531	30211	0	1500	1100	1200	1300	1100	1100	
15	Yes	20000	Male	1	2	24	0	0	2	2	2	2	15376	18010	17428	18338	17905	19104	3200	0	1500	0	1650	0	0	0
16	No	320000	Male	1	1	49	0	0	0	-1	-1	-1	253286	246536	194663	70074	5856	195599	10358	10000	75940	20000	195599	50000	50000	
17	No	360000	Female	1	1	49	1	-2	-2	-2	-2	-2	0	0	0	0	0	0	0	0	0	0	0	0	0	
18	No	180000	Female	1	2	29	1	-2	-2	-2	-2	-2	0	0	0	0	0	0	0	0	0	0	0	0	0	
19	No	130000	Female	3	2	39	0	0	0	0	0	-1	38358	27688	24489	20616	11802	930	3000	1537	1000	2000	930	33764		
20	Yes	120000	Female	2	1	39	-1	-1	-1	-1	-1	-1	316	316	316	0	632	316	316	0	632	316	0	0	0	
21	Yes	70000	Female	2	2	26	2	0	0	2	2	2	41087	42445	45020	44006	46905	46012	2007	3582	0	3601	0	1820	0	0
22	Yes	450000	Female	1	1	40	-2	-2	-2	-2	-2	-2	5512	19420	1473	560	0	0	19428	1473	560	0	0	0	1128	
23	No	90000	Male	1	2	23	0	0	0	0	-1	0	17111	7070	0	5398	6360	8292	5757	0	5398	1200	2045	2000	2000	
24	No	50000	Male	3	2	23	0	0	0	0	0	0	17111	7070	0	5398	6360	8292	5757	0	1426	1001	1432	1062	997	
25	Yes	60000	Male	1	2	27	1	-2	-2	-2	-2	-2	0	0	0	0	0	0	0	0	0	0	0	0	0	
26	No	50000	Female	3	2	30	0	0	0	0	0	0	17111	7070	0	5398	6360	8292	5757	0	1000	1000	1500	1000	1012	
27	No	50000	Female	3	1	47	-1	-1	-1	-1	-1	-1	17111	7070	0	5398	6360	8292	5757	0	3421	2044	30430	257	0	
28	No	50000	Male	1	2	26	0	0	0	0	0	0	17111	7070	0	5398	6360	8292	5757	0	1500	1000	1000	1600	0	
29	No	230000	Female	1	2	27	-1	-1	-1	-1	-1	-1	17111	7070	0	5398	6360	8292	5757	0	13281	15339	14307	37292	0	
30	No	100000	Male	1	2	32	0	0	0	0	0	0	186503	181328	180422	170410	173901	177413	8026	8060	6300	6400	6400	6737	2504	
31	No	500000	Female	2	1	54	-2	-2	-2	-2	-2	-2	15054	9806	11068	6026	-28335	18660	1500	1518	2043	0	47671	617		
32	No	500000	Male	1	1	58	-2	-2	-2	-2	-2	-2	15054	9806	11068	6026	-28335	18660	1500	1518	2043	0	47671	617		
33	No	160000	Male	1	2	30	-1	-1	-2	-2	-2	-2	30265	-131	-527	-923	-1488	-1884	131	396	396	565	792	0	0	
34	No	280000	Male	2	1	40	0	0	0	0	0	0	186503	181328	180422	170410	173901	177413	8026	8060	6300	6400	6400	6737	2504	
35	No	60000	Female	2	2	22	0	0	0	0	0	0	15054	9806	11068	6026	-28335	18660	1500	1518	2043	0	47671	617		

default_payment_test_data.csv																										
default_payment_test_data.csv																										
AC63																										
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	
1	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	PAY_5	PAY_6	BILL_AMT1	BILL_AMT2	BILL_AMT3	BILL_AMT4	BILL_AMT5	BILL_AMT6	PAY_AMT1	PAY_AMT2	PAY_AMT3	PAY_AMT4	PAY_AMT5	PAY_AMT6			
2	50000	Male		1	2	37	0	0	0	0	0	64400	57069	57608	19394	19619	20024	2500	1815	657	1000	1000	800			
3	500000	Male		1	2	29	0	0	0	0	0	367965	412023	445007	542653	483003	473944	55000	40000	38000	20239	13750	13770			
4	260000	Female		1	2	51	-1	-1	-1	-1	2	12261	21670	9966	8517	22287	13668	21818	9966	8583	22301	0	3640			
5	50000	Male		2	2	33	2	0	0	0	0	30518	29618	22102	22734	23217	23680	1718	1500	1000	1000	1000	716			
6	150000	Female		5	2	46	0	0	-1	0	0	-2	4463	3034	1170	1170	0	0	1013	1170	0	0	0	0		
7	20000	Male		1	2	24	0	0	0	0	0	17447	18479	19476	19865	20480	20063	1318	1315	704	928	912	1069			
8	130000	Female		2	1	51	-1	-1	-2	-2	-1	-1	99	0	0	0	2353	0	0	0	2353	0	0			
9	320000	Male		2	2	29	2	2	2	2	2	58267	59246	60184	58622	62307	63526	2500	2500	0	4800	2400	1600			
10	50000	Male		3	2	25	-1	0	0	0	0	42838	37225	36087	9636	9590	10030	1759	1779	320	500	1000	1000			
11	130000	Female		1	1	35	0	0	0	-1	-1	81313	117866	17740	1330	7095	1190	40000	5000	1330	7095	1190	2090			
12	20000	Male		3	2	44	2	2	0	0	0	8583	8303	9651	10488	12314	11970	0	1651	1000	2000	0	1500			
13	100000	Female		1	2	27	-2	-2	-2	-2	-2	-2000	5555	0	0	0	7555	0	0	0	0	0	0			
14	400000	Male		2	1	34	-1	-1	-1	-1	-1	19660	9666	11867	7839	14837	7959	9677	11867	7839	14837	7959	5712			
15	180000	Male		1	1	36	0	0	0	0	0	163736	116422	99278	95766	97753	95927	4655	2690	2067	2142	2217	1000			
16	260000	Female		1	1	60	1	-2	-1	-1	-1	-1100	-1100	21400	0	969	869	0	22500	0	969	1000	0			
17	140000	Male		2	1	32	0	0	0	0	0	86627	78142	68336	64648	58319	55251	3455	3110	5000	0	2100	2602			
18	210000	Male		3	1	45	2	3	4																	
19	370000	Male		1	2	50	-2	-2	-2																	
20	50000	Female		1	2	24	1	-2	-2																	
21	180000	Female		1	2	29	-1	-1	-1																	
22	120000	Male		2	2	26	0	0	0																	
23	470000	Male		2	2	27	2	2	2																	
24	50000	Male		2	2	23	2	0	0																	
25	20000	Male		2	2	23	1	2	0	0	0	20235	17132	16856	16875	13454	10104	0	1200	1000	0	1000	10000			
26	60000	Female		1	2	28	1	2	2	-2	-2	-1	21501	20650	0	0	0	2285	0	0	0	0	2285	0		
27	250000	Female		2	1	75	0	-1	-1	-1	-1	-1	52874	1631	1536	1010	5572	794	1631	1536	1010	5572	794	1184		
28	30000	Male		2	2	28	0	0	0	0	0	0	29242	29507	29155	25255	22001	0	5006	1244	851	955	0	0		
29	100000	Female		3	1	43	0	0	-2	-2	-2	-2	62170	0	0	0	0	0	0	0	0	0	0			
30	50000	Female		1	2	26	-1	-1	-1	-1	-1	-1	1156	316	316	316	316	316	316	316	316	316	316	316		
31	110000	Female		2	2	36	0	0	0	0	0	0	47819	48947	50330	50894	52175	53652	2200	2500	2000	2100	2500	2200		
32	180000	Male		2	2	29	1	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	0	0	0	0	0	0			
33	110000	Male		2	2	29	1	2	2	0	0	0	58362	56598	51908	48647	47862	47969	2500	0	2000	2000	1854	2000		

# Download, Install and Start H<sub>2</sub>O

# www.h2o.ai/download



**H<sub>2</sub>O**  
Version 3.14.0.1

Fast Scalable Machine Learning API  
For Smarter Applications

DOWNLOAD AND RUN    INSTALL IN R    INSTALL IN PYTHON    INSTALL ON HADOOP    USE FROM MAVEN

**DOWNLOAD H<sub>2</sub>O**

Get started with H<sub>2</sub>O in 3 easy steps

1. Download H<sub>2</sub>O. This is a zip file that contains everything you need to get started.
2. From your terminal, run:  

```
cd ~/Downloads
unzip h2o-3.14.0.1.zip
cd h2o-3.14.0.1
java -jar h2o.jar
```
3. Point your browser to <http://localhost:54321>

# H<sub>2</sub>O Flow (Web) Interface

Untitled Flow



CS

assist



168ms

## ❓ Assistance

Routine	Description
importFiles	Import file(s) into H <sub>2</sub> O
getFrames	Get a list of frames in H <sub>2</sub> O
splitFrame	Split a frame into two or more frames
mergeFrames	Merge two frames into one
getModels	Get a list of models in H <sub>2</sub> O
getGrids	Get a list of grid search results in H <sub>2</sub> O
getPredictions	Get a list of predictions in H <sub>2</sub> O
getJobs	Get a list of jobs running in H <sub>2</sub> O
buildModel	Build a model
importModel	Import a saved model
predict	Make a prediction

Check out more examples

Enter this URL **localhost:54321**  
and you should see this first page

OUTLINE FLOWS CLIPS HELP

## 💡 Help



Using Flow for the first time?

Quickstart Videos

Or, view example Flows to explore and learn H<sub>2</sub>O.

STAR H<sub>2</sub>O ON GITHUB!

Star 2,304

### GENERAL

- Flow Web UI ...
- ... Importing Data
- ... Building Models
- ... Making Predictions
- ... Using Flows
- ... Troubleshooting Flow

### EXAMPLES

Flow packs are a great way to explore and learn H<sub>2</sub>O. Try out these Flows and run them

# Import Local CSV

## Untitled Flow



Import Files...

Upload File...

Split Frame...

Merge Frames...

List All Frames

Impute...

## CS Assistance

Routine	Description
importFiles	Import file(s) into H <sub>2</sub> O
getFrames	Get a list of frames in H <sub>2</sub> O
splitFrame	Split a frame into two or more frames
mergeFrames	Merge two frames into one
getModels	Get a list of models in H <sub>2</sub> O
getGrids	Get a list of grid search results in H <sub>2</sub> O
getPredictions	Get a list of predictions in H <sub>2</sub> O
getJobs	Get a list of jobs running in H <sub>2</sub> O
buildModel	Build a model
importModel	Import a saved model
predict	Make a prediction

OUTLINE FLOWS CLIPS HELP

## Help

## PACK

## examples

- [GBM\\_Example.flow](#)
- [DeepLearning\\_MNIST.flow](#)
- [GLM\\_Example.flow](#)
- [DRF\\_Example.flow](#)
- [K-Means\\_Example.flow](#)
- [Million\\_Songs.flow](#)
- [KDDCup2009\\_Churn.flow](#)
- [QuickStartVideos.flow](#)
- [Airlines\\_Delay.flow](#)
- [GBM\\_Airlines\\_Classification.flow](#)
- [GBM\\_GridSearch.flow](#)
- [RandomData\\_Benchmark\\_Small.flow](#)
- [GBM\\_TuningGuide.flow](#)

Untitled Flow



CS

assist

## ?

 Assistance

Routine	Description
importFiles	Import file(s) into H <sub>2</sub> O
getFrames	Get a list of frames in H <sub>2</sub> O
splitFrame	Split a frame into two or more frames
mergeFrames	Merge two frames into one
getModels	Get a list of models in H <sub>2</sub> O
getGrids	Get a list of grid search results in H <sub>2</sub> O
getPredictions	Get a list of predictions in H <sub>2</sub> O
getJobs	Get a list of jobs running in H <sub>2</sub> O
buildModel	Build a model
importModel	Import a saved model
predict	Make a prediction

## Upload Dataset...

Choose File default\_payment\_training\_data.csv

Cancel

Upload

Find **default\_payment\_training\_data.csv**  
and then click **Upload**

OUTLINE FLOWS CLIPS HELP

## ?

 Help

### PACK

examples

- [GBM\\_Example.flow](#)
- [DeepLearning\\_MNIST.flow](#)
- [GLM\\_Example.flow](#)
- [DRF\\_Example.flow](#)
- [K-Means\\_Example.flow](#)
- [Million\\_Songs.flow](#)
- [KDDCup2009\\_Churn.flow](#)
- [QuickStartVideos.flow](#)
- [Airlines\\_Delay.flow](#)
- [GBM\\_Airlines\\_Classification.flow](#)
- [GBM\\_GridSearch.flow](#)
- [RandomData\\_Benchmark\\_Small.flow](#)
- [GBM\\_TuningGuide.flow](#)



## Setup Parse

### PARSE CONFIGURATION

Sources [default\\_payment\\_training\\_data.csv](#)

ID Key\_Frame\_default\_payment\_training\_data.hex

Parser CSV

Separator `;\r\n`Column Headers  Auto First row contains column names First row contains dataOptions  Enable single quotes as a field quotation character Delete on done

### EDIT COLUMN NAMES AND TYPES

Search by column name...

1	DEFAULT_PAYMENT_NEXT_MONTH	Enum ▾	Yes	Yes	No	No	No	No	No	No	No
2	LIMIT_BAL	Numeric ▾	20000	120000	90000	50000	50000	100000	140000	20000	200000
3	SEX	Enum ▾	Female	Female	Female	Female	Male	Female	Female	Male	Female
4	EDUCATION	Numeric ▾	2	2	2	2	2	2	3	3	3
5	MARRIAGE	Numeric ▾	1	2	2	1	1	2	1	2	2
6	AGE	Numeric ▾	24	26	34	37	57	23	28	35	34
7	PAY_0	Numeric ▾	2	-1	0	0	-1	0	0	-2	0
8	PAY_2	Numeric ▾	2	2	0	0	0	-1	0	-2	0
9	PAY_3	Numeric ▾	-1	0	0	0	-1	-1	2	-2	2
10	PAY_4	Numeric ▾	-1	0	0	0	0	0	0	-2	0
11	PAY_5	Numeric ▾	-2	0	0	0	0	0	0	-1	0
12	PAY_6	Numeric ▾	-2	2	0	0	0	-1	0	-1	-1
13	BILL_AMT1	Numeric ▾	3913	2682	29239	46990	8617	11876	11285	0	11073
14	BILL_AMT2	Numeric ▾	3102	1725	14027	48233	5670	380	14096	0	9787
15	BILL_AMT3	Numeric ▾	689	2682	13559	49291	35835	601	12108	0	5535

[◀ Previous page](#) [▶ Next page](#)

Parse

**Hit Parse**

# Explore

## Untitled Flow



CS

```
parseFiles
source_frames: ["default_payment_training_data.csv"]
destination_frame: "Key_Frame__default_payment_training_data.hex"
parse_type: "CSV"
separator: 44
number_columns: 24
single_quotes: false
column_names:
["DEFAULT_PAYMENT_NEXT_MONTH","LIMIT_BAL","SEX","EDUCATION","MARRIAGE","AGE","PAY_0","PAY_2","PAY_3","PAY_4","PAY_5","PAY_6","BILL_AMT1","BILL_AMT2","BILL_AMT3","BILL_AMT4","BILL_AMT5","BILL_AMT6","PAY_AMT1","PAY_AMT2","PAY_AMT3","PAY_AMT4","PAY_AMT5","PAY_AMT6"]
column_types:
["Enum","Numeric","Enum","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric","Numeric"]
delete_on_done: true
check_header: 1
chunk_size: 4194304
```

1.1s

## Job

Run Time 00:00:00.619

Remaining Time 00:00:00.0

Type Frame

Key [Q Key\\_Frame\\_\\_default\\_payment\\_training\\_data.hex](#)

Description Parse

Status DONE

Progress 100%

Done.

Actions

[Q View](#)

Hit View



CS

getFrameSummary "Key\_Frame\_\_default\_payment\_training\_data.hex"

181ms

## Key\_Frame\_\_default\_payment\_training\_data.hex

Actions: [View Data](#) [Split...](#) [Build Model...](#) [Predict](#) [Download](#) [Export](#)

[Delete](#)

Rows	Columns	Compressed Size
21600	24	1MB

### COLUMN SUMMARIES

label	type	Missing	Zeros	+Inf	-Inf	min	max	mean	sigma	cardinality	Actions
DEFAULT_PAYMENT_NEXT_MONTH	enum	0	16767	0	0	0	1.0	0.2238	0.4168	2	<a href="#">Convert to numeric</a>
LIMIT_BAL	int	0	0	0	0	10000.0	10000000.0	165471.4667	128853.3148	·	<a href="#">Convert to enum</a>
SEX	enum	0	13621	0	0	0	1.0	0.3694	0.4827	2	<a href="#">Convert to numeric</a>
EDUCATION	int	0	9	0	0	0	6.0	1.8500	0.7796	·	<a href="#">Convert to enum</a>
MARRIAGE	int	0	37	0	0	0	3.0	1.5558	0.5225	·	<a href="#">Convert to enum</a>
AGE	int	0	0	0	0	21.0	79.0	35.4053	9.2768	·	<a href="#">Convert to enum</a>
PAY_0	int	0	10563	0	0	-2.0	8.0	-0.0052	1.1267	·	<a href="#">Convert to enum</a>
PAY_2	int	0	11284	0	0	-2.0	8.0	-0.1224	1.2009	·	<a href="#">Convert to enum</a>
PAY_3	int	0	11309	0	0	-2.0	8.0	-0.1554	1.2073	·	<a href="#">Convert to enum</a>
PAY_4	int	0	11905	0	0	-2.0	8.0	-0.2106	1.1722	·	<a href="#">Convert to enum</a>
PAY_5	int	0	12148	0	0	-2.0	8.0	-0.2500	1.1450	·	<a href="#">Convert to enum</a>
PAY_6	int	0	11548	0	0	-2.0	8.0	-0.2781	1.1636	·	<a href="#">Convert to enum</a>
BILL_AMT1	int	0	1439	0	0	-165580.0	964511.0	50566.4500	72759.4178	·	<a href="#">Convert to enum</a>
BILL_AMT2	int	0	1848	0	0	-69777.0	983931.0	48656.5150	70553.0143	·	<a href="#">Convert to enum</a>
BILL_AMT3	int	0	2093	0	0	-157264.0	1664089.0	46412.0669	68567.1707	·	<a href="#">Convert to enum</a>
BILL_AMT4	int	0	2284	0	0	-170000.0	891586.0	42411.4054	63313.3971	·	<a href="#">Convert to enum</a>
BILL_AMT5	int	0	2557	0	0	-81334.0	927171.0	40025.8482	60523.1932	·	<a href="#">Convert to enum</a>
BILL_AMT6	int	0	2931	0	0	-339603.0	961664.0	38615.2118	59526.3095	·	<a href="#">Convert to enum</a>
PAY_AMT1	int	0	3850	0	0	0	505000.0	5591.2800	15306.3294	·	<a href="#">Convert to enum</a>
PAY_AMT2	int	0	3932	0	0	0	1684259.0	5827.2038	21146.6211	·	<a href="#">Convert to enum</a>

Click on any feature  
to see some graphs  
(default payment and  
education used in this  
example)

## Untitled Flow



BILL_AMT3	int	0	2093	0	0	-157264.0	1664089.0	46412.0669	68567.1707	• Convert to enum
BILL_AMT4	int	0	2284	0	0	-170000.0	891586.0	42411.4054	63313.3971	• Convert to enum
BILL_AMT5	int	0	2557	0	0	-81334.0	927171.0	40025.8482	60523.1932	• Convert to enum
BILL_AMT6	int	0	2931	0	0	-339603.0	961664.0	38615.2118	59526.3095	• Convert to enum
PAY_AMT1	int	0	3850	0	0	0	505000.0	5591.2800	15306.3294	• Convert to enum
PAY_AMT2	int	0	3932	0	0	0	1684259.0	5827.2038	21146.6211	• Convert to enum

[◀ Previous 20 Columns](#)[▶ Next 20 Columns](#)

## ► CHUNK COMPRESSION SUMMARY

## ► FRAME DISTRIBUTION SUMMARY

```
getColumnSummary "Key_Frame__default_payment_training_data.hex", "DEFAULT_PAYMENT_NEXT_MONTH"
```

138ms

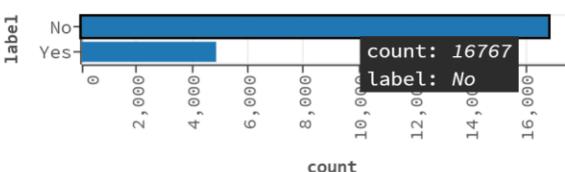
## Summary: DEFAULT\_PAYMENT\_NEXT\_MONTH

[Actions](#) [◀ Impute](#) [☰ Inspect](#)

## CHARACTERISTICS



## DOMAIN (MAX 1000 LEVELS)

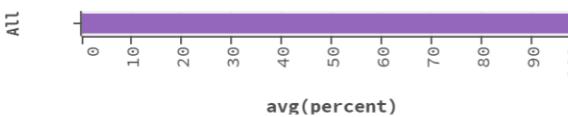




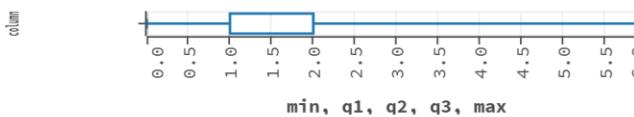
## Summary: EDUCATION

Actions ◀ Impute ☰ Inspect

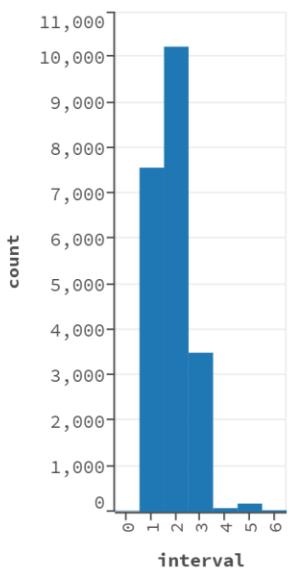
### CHARACTERISTICS



### SUMMARY



### DISTRIBUTION



# Train a Model

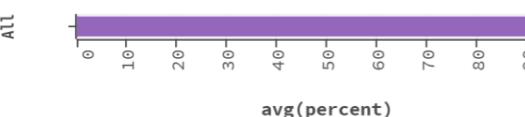
Untitled Flow



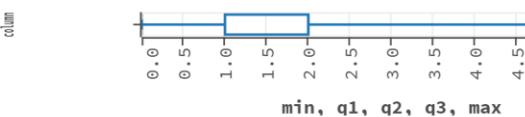
## Summary: EDUCATION

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

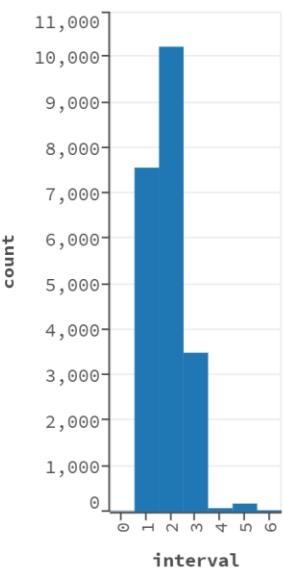
### CHARACTERISTICS



### SUMMARY



### DISTRIBUTION



Aggregator...

Deep Learning...

Distributed Random Forest...

Gradient Boosting Machine...

Generalized Linear Modeling...

Generalized Low Rank Modeling...

K-means...

Naive Bayes...

Principal Components Analysis...

Stacked Ensemble...

Word2Vec...

List All Models

List Grid Search Results

Import Model...

Export Model...

Choose one of the supervised machine learning algorithms for training  
(Gradient Boosting Machine algo used in this example)



## Build a Model

Select an algorithm: Gradient Boosting Machine ▾

### PARAMETERS

model\_id **GBM-Demo-Model**

Give this model a name (optional)

training\_frame **Key\_Frame\_default\_payment\_training\_data.hex**

Choose the dataset for training

validation\_frame (Choose...)

nfolds 0

response\_column **DEFAULT\_PAYMENT\_NEXT\_MONTH**

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 &gt;= 2).

Response variable column.

ignored\_columns Search...

Showing page 1 of 1.

<input type="checkbox"/> DEFAULT_PAYMENT_NEXT_MONTH	ENUM(2)
<input type="checkbox"/> LIMIT_BAL	INT
<input type="checkbox"/> SEX	ENUM(2)
<input type="checkbox"/> EDUCATION	INT
<input type="checkbox"/> MARRIAGE	INT
<input type="checkbox"/> AGE	INT
<input type="checkbox"/> PAY_0	INT
<input type="checkbox"/> PAY_2	INT
<input type="checkbox"/> PAY_3	INT
<input type="checkbox"/> PAY_4	INT

 All     None

Only show columns with more than 0 % missing values.

ignore\_const\_cols 

Ignore constant columns.

Choose **DEFAULT\_PAYMENT\_NEXT\_MONTH** as the target response (this is the outcome you want to predict)

## Untitled Flow



stopping_metric	AUTO	for k:=stopping_rounds scoring events (0 to disable) Metric to use for early stopping (AUTO: logloss for classification, deviance for regression)	<input type="checkbox"/>
stopping_tolerance	0.001	Relative tolerance for metric-based stopping criterion (stop if relative improvement is not at least this much)	<input type="checkbox"/>
max_runtime_secs	0	Maximum allowed runtime in seconds for model training. Use 0 to disable.	<input type="checkbox"/>
learn_rate_annealing	1	Scale the learning rate by this factor after each tree (e.g., 0.99 or 0.999)	<input type="checkbox"/>
distribution	AUTO	Distribution function	<input type="checkbox"/>
huber_alpha	0.9	Desired quantile for Huber/M-regression (threshold between quadratic and linear loss, must be between 0 and 1).	<input type="checkbox"/>
checkpoint		Model checkpoint to resume training with.	<input type="checkbox"/>
col_sample_rate_per_tree	1	Column sample rate per tree (from 0.0 to 1.0)	<input type="checkbox"/>
min_split_improvement	0.00001	Minimum relative improvement required to split a node	<input type="checkbox"/>
histogram_type	AUTO	What type of histograms to use	<input type="checkbox"/>
categorical_encoding	AUTO	Encoding scheme for categorical features	<input type="checkbox"/>

You can ignore all other settings for now  
(so this is like an out-of-box model)

## EXPERT

## GRID?

build_tree_one_node	<input type="checkbox"/>	Run on one node only; no network overhead but fewer cpus used. Suitable for small datasets.	<input type="checkbox"/>
sample_rate_per_class		A list of row sample rates per class (relative fraction for each class, from 0.0 to 1.0), for each tree	<input type="checkbox"/>
col_sample_rate_change_per_level	1	Relative change of the column sampling rate for every level (from 0.0 to 2.0)	<input type="checkbox"/>
max_abs_leafnode_pred	1.7976931348623157e+308	Maximum absolute value of a leaf node prediction	<input type="checkbox"/>
pred_noise_bandwidth	0	Bandwidth (sigma) of Gaussian multiplicative noise ~N(1,sigma) for tree node predictions	<input type="checkbox"/>
calibrate_model	<input type="checkbox"/>	Use Platt Scaling to calculate calibrated class probabilities. Calibration can provide more accurate estimates of class probabilities.	<input type="checkbox"/>
calibration_frame	(Choose...)	Calibration frame for Platt Scaling	<input type="checkbox"/>

Build Model

Scroll down and hit **Build Model**

## Untitled Flow



pred\_noise\_bandwidth 0

Bandwidth (sigma) of Gaussian multiplicative noise ~N(1,sigma) for tree node predictions

calibrate\_model 

Use Platt Scaling to calculate calibrated class probabilities. Calibration can provide more accurate estimates of class probabilities.

calibration\_frame (Choose...)

▼ Calibration frame for Platt Scaling

Build Model

CS

```
buildModel 'gbm', {"model_id":"GBM-Demo-Model","training_frame":"Key_Frame__default_payment_training_data.hex","nfolds":0,"response_column":"DEFAULT_PAYMENT_NEXT_MONTH","ignored_columns":[],"ignore_const_cols":true,"ntrees":50,"max_depth":5,"min_rows":10,"nbins":20,"seed":-1,"learn_rate":0.1,"sample_rate":1,"col_sample_rate":1,"score_each_iteration":false,"score_tree_interval":0,"balance_classes":false,"nbins_top_level":1024,"nbins_cats":1024,"r2_stopping":1.7976931348623157e+308,"stopping_rounds":0,"stopping_metric":"AUTO","stopping_tolerance":0.001,"max_runtime_secs":0,"learn_rate_annealing":1,"distribution":"AUTO","huber_alpha":0.9,"checkpoint":"","col_sample_rate_per_tree":1,"min_split_improvement":0.0001,"histogram_type":"AUTO","categorical_encoding":"AUTO","build_tree_one_node":false,"sample_rate_per_class":[],"col_sample_rate_change_per_level":1,"max_abs_leafnode_pred":1.7976931348623157e+308,"pred_noise_bandwidth":0,"calibrate_model":false}
```

6.2s

## Job

Run Time 00:00:05.183

Remaining Time 00:00:00.0

Type Model

Key GBM-Demo-Model

Description GBM

Status DONE

Progress 100%

Done.

Actions

View

The model building process should not take too long (say, < 3 mins depends on hardware)

When it is done, you can hit View to look at model summary

## Untitled Flow



## Model

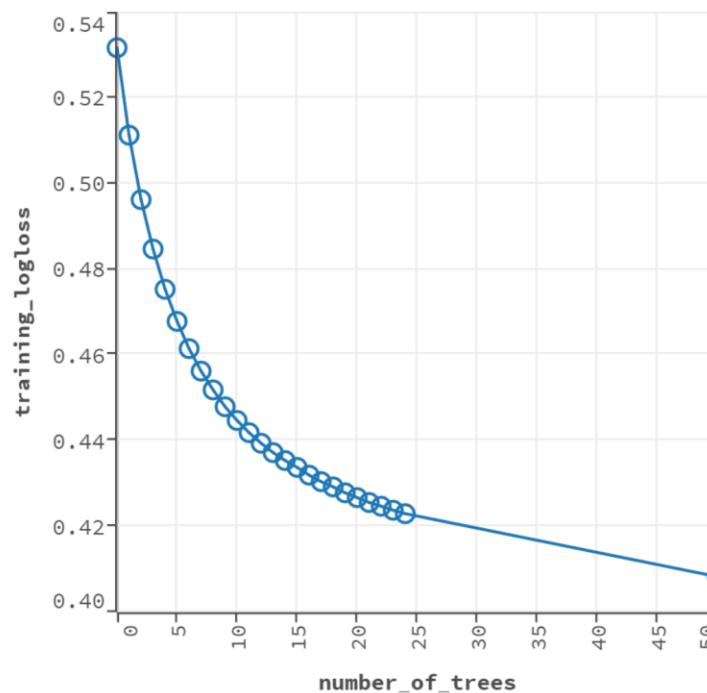
Model ID: GBM-Demo-Model

Algorithm: Gradient Boosting Machine

Actions: [Refresh](#) [Predict...](#) [Download POJO](#) [Download Model Deployment Package](#) [Export](#) [Inspect](#) [Delete](#) [Download Gen Model](#)

► MODEL PARAMETERS

▼ SCORING HISTORY - LOGLOSS

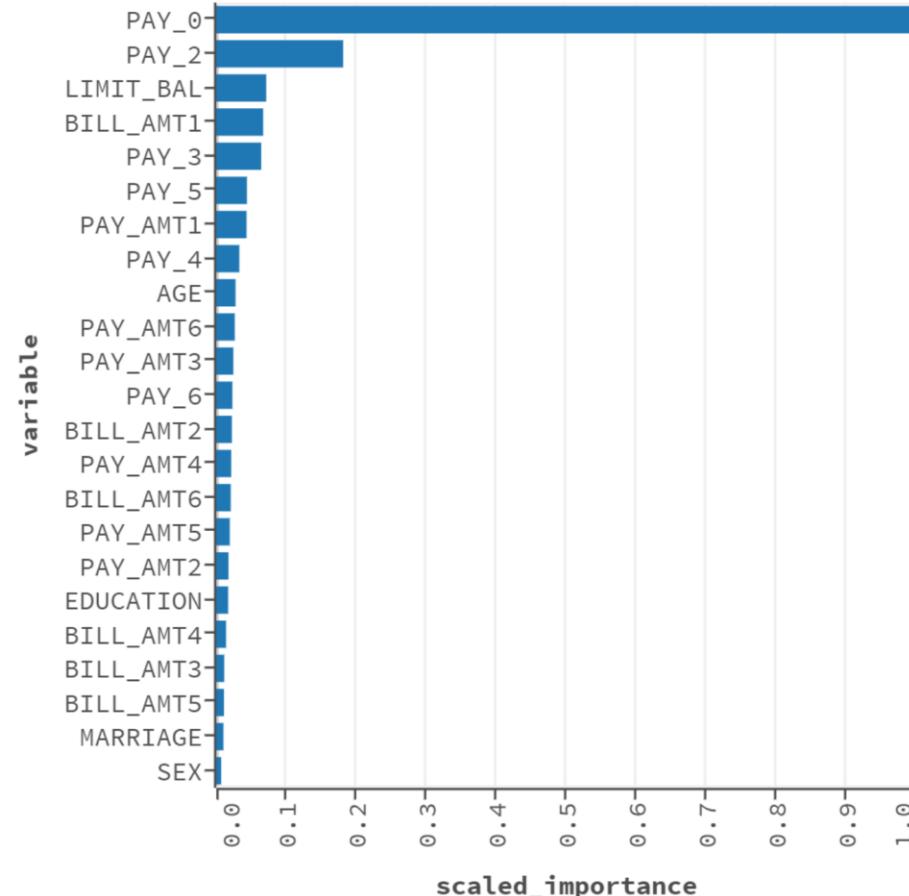


Logloss is one of the measures of model performance (less is better)

## Untitled Flow



## ▼ VARIABLE IMPORTANCES



The **Variable Importances** graph shows you which features are the main drivers (more/less important to outcomes)

# Make Predictions

Untitled Flow



CS

Expression...

## Upload Dataset...



Choose File default\_payment...test\_data.csv

Cancel

Upload

Now we have a model, it is time to use it and make some predictions.

First, Find '**default\_payment\_test\_data.csv**' and then click **Upload**



## Setup Parse

### PARSE CONFIGURATION

Sources

Parser

Separator

Column Headers  Auto

- First row contains column names
- First row contains data

Options  Enable single quotes as a field quotation character

Delete on done

### EDIT COLUMN NAMES AND TYPES

Search by column name...

1	LIMIT_BAL	<input type="button" value="Numeric▼"/>	50000	500000	260000	50000	150000	20000	130000	320000	50000
2	SEX	<input type="button" value="Enum ▼"/>	Male	Male	Female	Male	Female	Male	Female	Male	Male
3	EDUCATION	<input type="button" value="Numeric▼"/>	1	1	1	2	5	1	2	2	3
4	MARRIAGE	<input type="button" value="Numeric▼"/>	2	2	2	2	2	2	1	2	2
5	AGE	<input type="button" value="Numeric▼"/>	37	29	51	33	46	24	51	29	25
6	PAY_0	<input type="button" value="Numeric▼"/>	0	0	-1	2	0	0	-1	2	-1
7	PAY_2	<input type="button" value="Numeric▼"/>	0	0	-1	0	0	0	-1	2	0
8	PAY_3	<input type="button" value="Numeric▼"/>	0	0	-1	0	-1	0	-2	2	0
9	PAY_4	<input type="button" value="Numeric▼"/>	0	0	-1	0	0	0	-2	2	0
10	PAY_5	<input type="button" value="Numeric▼"/>	0	0	-1	0	0	0	-1	2	0
11	PAY_6	<input type="button" value="Numeric▼"/>	0	0	2	0	-2	0	-1	2	0
12	BILL_AMT1	<input type="button" value="Numeric▼"/>	64400	367965	12261	30518	4463	17447	99	58267	42838
13	BILL_AMT2	<input type="button" value="Numeric▼"/>	57069	412023	21670	29618	3034	18479	0	59246	37225
14	BILL_AMT3	<input type="button" value="Numeric▼"/>	57608	445007	9966	22102	1170	19476	0	60184	36087
15	BILL_AMT4	<input type="button" value="Numeric▼"/>	19394	542653	8517	22734	1170	19865	0	58622	9636

[◀ Previous page](#) [▶ Next page](#)

Hit Parse



CS

getFrameSummary "Key\_Frame\_\_default\_payment\_test\_data.hex"

67ms

## Key\_Frame\_\_default\_payment\_test\_data.hex

Actions: [View Data](#) [Split...](#) [Build Model...](#) [Predict](#) [Download](#) [Export](#)

[Delete](#)

Rows	Columns	Compressed Size
2399	23	138KB

### COLUMN SUMMARIES

label	type	Missing	Zeros	+Inf	-Inf	min	max	mean	sigma	cardinality	Actions
LIMIT_BAL	int	0	0	0	0	10000.0	760000.0	165744.0600	131629.3715	· Convert to enum	
SEX	enum	0	1457	0	0	0	1.0	0.3927	0.4884	2	Convert to numeric
EDUCATION	int	0	2	0	0	0	6.0	1.8245	0.7838	· Convert to enum	
MARRIAGE	int	0	4	0	0	0	3.0	1.5702	0.5182	· Convert to enum	
AGE	int	0	0	0	0	21.0	75.0	35.1605	9.2185	· Convert to enum	
PAY_0	int	0	1175	0	0	-2.0	8.0	0.0158	1.0939	· Convert to enum	
PAY_2	int	0	1259	0	0	-2.0	7.0	-0.1334	1.1983	· Convert to enum	
PAY_3	int	0	1267	0	0	-2.0	7.0	-0.1492	1.1750	· Convert to enum	
PAY_4	int	0	1345	0	0	-2.0	7.0	-0.2213	1.1150	· Convert to enum	
PAY_5	int	0	1372	0	0	-2.0	6.0	-0.2793	1.0625	· Convert to enum	
PAY_6	int	0	1328	0	0	-2.0	6.0	-0.2776	1.1082	· Convert to enum	
BILL_AMT1	int	0	168	0	0	-4370.0	581775.0	50891.3585	71673.6349	· Convert to enum	
BILL_AMT2	int	0	201	0	0	-22960.0	572677.0	48571.8070	68667.3832	· Convert to enum	
BILL_AMT3	int	0	232	0	0	-20320.0	565550.0	45980.2718	64757.2083	· Convert to enum	
BILL_AMT4	int	0	263	0	0	-20320.0	572805.0	41995.9241	60861.2765	· Convert to enum	
BILL_AMT5	int	0	283	0	0	-23003.0	823540.0	39790.6090	58735.6251	· Convert to enum	
BILL_AMT6	int	0	327	0	0	-51443.0	501370.0	38115.5727	55726.1698	· Convert to enum	
PAY_AMT1	int	0	420	0	0	0	235728.0	5109.2797	12727.4165	· Convert to enum	
PAY_AMT2	int	0	430	0	0	0	361560.0	5710.4068	17343.1463	· Convert to enum	
PAY_AMT3	int	0	502	0	0	0	221876.0	4771.8416	13915.0787	· Convert to enum	

[◀ Previous 20 Columns](#) [▶ Next 20 Columns](#)

### CHUNK COMPRESSION SUMMARY

### FRAME DISTRIBUTION SUMMARY

Now we have the **test** dataset in H<sub>2</sub>O ready for making predictions

## Untitled Flow



Predict...

Partial Dependence Plots...

List All Predictions

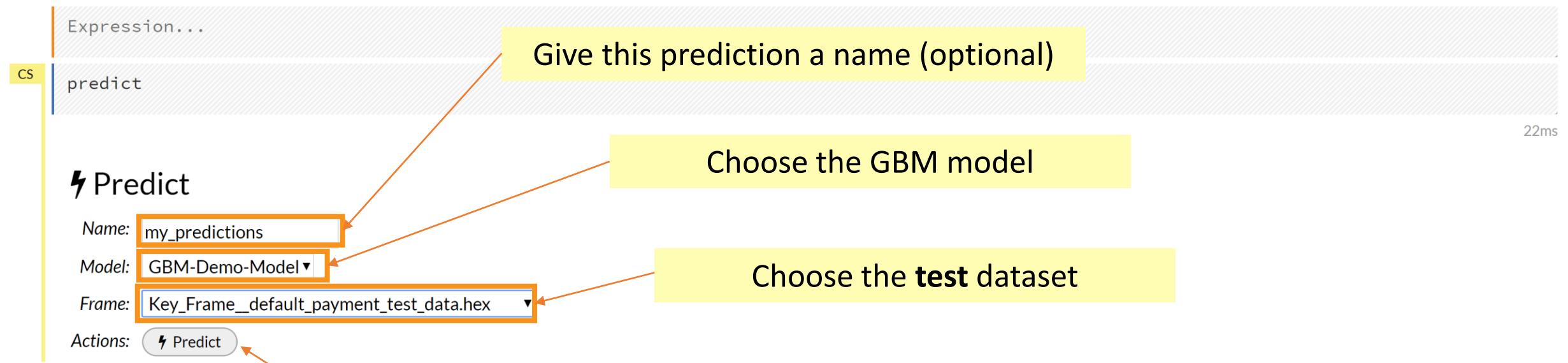
Score -&gt; Predict

CS

Expression...



## Untitled Flow



## Untitled Flow



EXPRESSIONS...

CS

predict

22ms



## ⚡ Predict

Name: my\_predictions

Model: GBM-Demo-Model▼

Frame: Key\_Frame\_\_default\_payment\_test\_data.hex ▼

Actions: ⚡ Predict

predict model: "GBM-Demo-Model", frame: "Key\_Frame\_\_default\_payment\_test\_data.hex", predictions\_frame: "my\_predictions"

64ms



## ⚡ Prediction

### PREDICTION

prediction\_frame my\_predictions

Combine predictions with frame

Hit **Combine predictions with frame**

## Untitled Flow



ACTIONS: Predict

```
predict model: "GBM-Demo-Model", frame: "Key_Frame__default_payment_test_data.hex", predictions_frame: "my_predictions"
```

64ms

## ⚡ Prediction

### ▼ PREDICTION

```
prediction_frame my_predictions
```

Combine predictions with frame

CS

```
bindFrames "combined-my_predictions", [ "my_predictions", "Key_Frame__default_payment_test_data.hex" ]
```

83ms

## grid Frames Combined

The specified frames were combined successfully.

View Frame

Hit View Frame

## Untitled Flow



CS

getFrameSummary "combined-my\_predictions"

56ms

Hit View Data

## grid combined-my\_predictions

Actions: [View Data](#) [Split...](#) [Build Model...](#) [Predict](#) [Download](#) [Export](#)[Delete](#)

Rows	Columns	Compressed Size
2399	26	177KB

## ▼ COLUMN SUMMARIES

label	type	Missing	Zeros	+Inf	-Inf	min	max	mean	sigma	cardinality	Actions
predict	enum	0	1751	0	0	0	1.0	0.2701	0.4441	2	<a href="#">Convert to numeric</a>
No	real	0	0	0	0	0.1717	0.9708	0.7786	0.1897	· ·	
Yes	real	0	0	0	0	0.0292	0.8283	0.2214	0.1897	· ·	
LIMIT_BAL	int	0	0	0	0	10000.0	7600000.0	165744.0600	131629.3715	·	<a href="#">Convert to enum</a>
SEX	enum	0	1457	0	0	0	1.0	0.3927	0.4884	2	<a href="#">Convert to numeric</a>
EDUCATION	int	0	2	0	0	0	6.0	1.8245	0.7838	·	<a href="#">Convert to enum</a>
MARRIAGE	int	0	4	0	0	0	3.0	1.5702	0.5182	·	<a href="#">Convert to enum</a>
AGE	int	0	0	0	0	21.0	75.0	35.1605	9.2185	·	<a href="#">Convert to enum</a>
PAY_0	int	0	1175	0	0	-2.0	8.0	0.0158	1.0939	·	<a href="#">Convert to enum</a>
PAY_2	int	0	1259	0	0	-2.0	7.0	-0.1334	1.1983	·	<a href="#">Convert to enum</a>
PAY_3	int	0	1267	0	0	-2.0	7.0	-0.1492	1.1750	·	<a href="#">Convert to enum</a>
PAY_4	int	0	1345	0	0	-2.0	7.0	-0.2213	1.1150	·	<a href="#">Convert to enum</a>
PAY_5	int	0	1372	0	0	-2.0	6.0	-0.2793	1.0625	·	<a href="#">Convert to enum</a>
PAY_6	int	0	1328	0	0	-2.0	6.0	-0.2776	1.1082	·	<a href="#">Convert to enum</a>



## combined-my\_predictions

### DATA

[← Previous 20 Columns](#) [→ Next 20 Columns](#)

Row	predict	No	Yes	LIMIT_BAL	SEX	EDUCATION	MARRIAGE	AGE	PAY_0	PAY_2	PAY_3	PAY_4	PAY_5	PAY_6	BILL_AMT1	BILL_AMT2	BILL_AMT3	BILL_AMT4	BILL_AMT5	BILL_AMT6			
1	No	0.8564	0.1436	50000.0	Male	1.0	2.0	37.0	0	0	0	0	0	0	64400.0	57069.0	57608.0	19394.0	19619.0	20024.0			
2	No	0.9165	0.0835	500000.0	Male	1.0	2.0	29.0	0	0	0	0	0	0	367965.0	412023.0	445007.0	542653.0	483003.0	473944.0			
3	No	0.9294	0.0706	260000.0	Female	1.0	2.0	51.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	12261.0	21670.0	9966.0	8517.0	22287.0	13668.0			
4	Yes	0.3858	0.6142	50000.0	Male	2.0	2.0	33.0	2.0	0	0	0	0	0	30518.0	29618.0	22102.0	22734.0	23217.0	23680.0			
5	No	0.9274	0.0726	150000.0	Female	5.0	2.0	46.0	0	0	-1.0	0	0	-2.0	4463.0	3034.0	1170.0	1170.0	0	0			
6	No	0.8135	0.1865	20000.0	Male	1.0	2.0	24.0	0	0	0	0	0	0	17447.0	18479.0	19476.0	19865.0	20480.0	20063.0			
7	No	0.7781	0.2219	130000.0	Female	2.0	1.0	51.0	-1.0	-1.0	-2.0	-2.0	-1.0	-1.0	99.0	0	0	0	2353.0	0			
8	Yes	0.2608	0.7392	320000.0	Male	2.0	2.0	29.0	2.0	2.0	2.0	2.0	2.0	2.0	58267.0	59246.0	60184.0	58622.0	62307.0	63526.0			
9	No	0.8557	0.1443	50000.0	Male	3.0	2.0	25.0	-1.0	0	0	0	0	0	42838.0	37225.0	36087.0	9636.0	9590.0	10030.0			
10	No	0.9205	0.0795	130000.0	Female	1.0	1.0	35.0	0	0	0	-1.0	-1.0	-1.0	81313.0	117866.0	17740.0	1330.0	7095.0	1190.0			
11	Yes	0.3046	0.6954	20000.0	Male	3.0	2.0	44.0	2.0	2.0	0	0	0	2.0	8583.0	8303.0	9651.0	10488.0	12314.0	11970.0			
12	No	0.8553	0.1447	100000.0	Female	1.0	2.0	27.0	-2.0	-2.0	-2.0	-2.0	-2.0	-2.0	-2000.0	5555.0	0	0	0	0			
13	No	0.9581	0.0419	400000.0	Male	2.0	1.0	34.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	19660.0	9666.0	11867.0	7839.0	14837.0	7959.0			
14	No	0.8889	0.1111	180000.0	Male	1.0	1.0	36.0	0	0	0	0	0	0	163736.0	116422.0	99278.0	95766.0	97753.0	95927.0			
15	Yes	0.6848	0.3152	260000.0	Female	1.0	1.0	60.0	1.0	-2.0	-1.0	-1.0	-1.0	-1.0	-1100.0	-1100.0	21400.0	0	969.0	869.0			
16	No	0.8214	0.1786	140000.0	Male	2.0	1.0	32.0	0	0	0	0	2.0	0	86627.0	78142.0	68336.0	64648.0	58319.0	55251.0			
17	Yes	0.3657	0.6343	210000.0	Male	3.0	1.0	45.0	2.0	3.0	4.0	4.0	5.0	6.0	115785.0	122904.0	129847.0	137277.0	145533.0	154105.0			
18	No	0.9433	0.0567	370000.0	Male	1.0	2.0	50.0	-2.0	-2.0	-2.0	-2.0	-2.0	-2.0	6093.0	15130.0	8204.0	15398.0	4792.0	13453.0			
19	Yes	0.6832	0.3168	50000.0	Female	1.0	2.0	Features from the test dataset												-709.0	-2898.0	-3272.0	-3272.0
20	Predictions with probs of Yes/No																						
21	No	0.8953	0.1047	180000.0	Female	1.0	2.0	Features from the test dataset												0	0	17227.0	17042.0
22	No	0.8571	0.1429	120000.0	Male	2.0	2.0	Features from the test dataset												13736.0	116000.0	119131.0	122135.0
23	No	0.8953	0.1047	470000.0	Male	2.0	2.0	27.0	2.0	2.0	2.0	2.0	0	0	296573.0	303320.0	307843.0	479978.0	305145.0	309959.0			
24	No	0.8571	0.1429	50000.0	Male	2.0	2.0	23.0	2.0	0	0	0	0	0	49758.0	48456.0	44116.0	21247.0	20066.0	18858.0			
25	No	0.8953	0.1047	20000.0	Male	2.0	2.0	23.0	1.0	2.0	0	0	2.0	0	20235.0	17132.0	16856.0	16875.0	13454.0	10104.0			
26	No	0.8953	0.1047	60000.0	Female	1.0	2.0	28.0	1.0	2.0	2.0	-2.0	-2.0	-1.0	21501.0	20650.0	0	0	0	2285.0			
27	No	0.8571	0.1429	250000.0	Female	2.0	1.0	75.0	0	-1.0	-1.0	-1.0	-1.0	-1.0	52874.0	1631.0	1536.0	1010.0	5572.0	794.0			
28	No	0.8953	0.1047	30000.0	Male	2.0	2.0	28.0	0	0	0	0	0	0	29242.0	29507.0	29155.0	25255.0	22001.0	0			

## Untitled Flow



CS

getFrameSummary "combined-my\_predictions"

56ms

## combined-my\_predictions

Download the predictions as CSV

Actions:

View Data

Split...

Build Model...

Predict

Download

Export

Delete

Rows

2399

Columns

26

Compressed Size

177KB

### COLUMN SUMMARIES

label	type	Missing	Zeros	+Inf	-Inf	min	max	mean	sigma	cardinality	Actions
predict	enum	0	1751	0	0	0	1.0	0.2701	0.4441	2	Convert to numeric
No	real	0	0	0	0	0.1717	0.9708	0.7786	0.1897	· ·	
Yes	real	0	0	0	0	0.0292	0.8283	0.2214	0.1897	· ·	
LIMIT_BAL	int	0	0	0	0	10000.0	7600000.0	165744.0600	131629.3715	·	Convert to enum
SEX	enum	0	1457	0	0	0	1.0	0.3927	0.4884	2	Convert to numeric
EDUCATION	int	0	2	0	0	0	6.0	1.8245	0.7838	·	Convert to enum
MARRIAGE	int	0	4	0	0	0	3.0	1.5702	0.5182	·	Convert to enum
AGE	int	0	0	0	0	21.0	75.0	35.1605	9.2185	·	Convert to enum
PAY_0	int	0	1175	0	0	-2.0	8.0	0.0158	1.0939	·	Convert to enum
PAY_2	int	0	1259	0	0	-2.0	7.0	-0.1334	1.1983	·	Convert to enum
PAY_3	int	0	1267	0	0	-2.0	7.0	-0.1492	1.1750	·	Convert to enum
PAY_4	int	0	1345	0	0	-2.0	7.0	-0.2213	1.1150	·	Convert to enum
PAY_5	int	0	1372	0	0	-2.0	6.0	-0.2793	1.0625	·	Convert to enum
PAY_6	int	0	1328	0	0	-2.0	6.0	-0.2776	1.1082	·	Convert to enum