

4

Wednesday
Day (004-361)

January

Fraud Detection

Wells Fargo

Beetle

⇒ Problem:

- Data card transaction level fraud detection
- Subset of transaction are put on queue for human intervention
- Task was to prioritize queue
 - human agent should get fraud case
- Not a frontline fraud detection which is more challenging

⇒ Solution:-

- H2O GBM model

⇒ Model approach:-

- Binary classification
- Fraud or not fraud

FEBRUARY 2006

W	M	T	W	T	F	S	S
5		1	2	3	4	5	
6	6	7	8	9	10	11	12
7	13	14	15	16	17	18	19
8	20	21	22	23	24	25	26

Celebrating
100 years
of setting standards in banking

5

Thursday

Day (005-360)



January

- 08 a) Evaluation metrics (need to understand)
- Recall @ 40%
 - close to agent goal

- 10 a) Training
- HDFS
 - Data prep in pyspark
 - H2O and Mojo

=> Lesson learned

13 * Modeling challenges

14 -> Label generation

- Labeling declined auth
- which transaction was found out at many

16 -> Delayed labels

- fraud can be caught instantaneously or discovered later (customer calls after a week/month)

JANUARY							2006
W	M	T	W	T	F	S	S
52	30	31					1
1	2	3	4	5	6	7	8
2	9	10	11	12	13	14	15
3	16	17	18	19	20	21	22
4	23	24	25	26	27	28	29

6

Friday

Day (006-35)

08 -> You
att

=> C

* T

FEBRUARY

W	M	T	W
5		1	
6	6	7	8
7	13	14	15

6

Friday

Day (006-359)



January

→ You can't measure performance after putting on production

- This means delay in pushing this data for next training

⇒ Time series cross validation

- look ahead gap, to avoid data leakage
- in-and out-of-time samples

* Training challenges

- datasets change
- access restricted
- new platform mid project

* Added Rigor for production Model

- Model governance
- finance industry

7

Saturday

Day (007-358)



January

8

Sun
Day (0

08 - Know how your model works,

09 → Model monitoring

- 10 - regular input/output distribution monitoring
- 11 - performance monitoring
- 12 - trigger for when to retrain / turn off model.

13 → Production scoring challenges

- 14 - Pipeline for generating features close to production
- 15 - Cache historical features
- production scoring faster
 - instead of have put in memory (model params)
- 16 - H2O majo is good as well

→ Have clear regeneration document

JANUARY 2006						
T	W	T	F	S	S	
31						1
3	4	5	6	7	8	
10	11	12	13	14	15	
17	18	19	20	21	22	
24	25	26	27	28	29	

FEBRUARY

W	M	T	W	T	F
5			1	2	3
6	6	7	8	9	10
7	13	14	15	16	17
8	20	21	22	23	24
9	27	28			

Celebrating

8

Sunday

Day (008-357)



January

- 08 → Model versioning
- tag entire code
 - feature & training
 - Data tagging is also ~~needed~~ done by git hash
- 11 → Imbalance classes in time series
- 12 → Problem formulation
- ranking rather than classification
- 14 →
- formative model : milliseconds
 - cur model : seconds