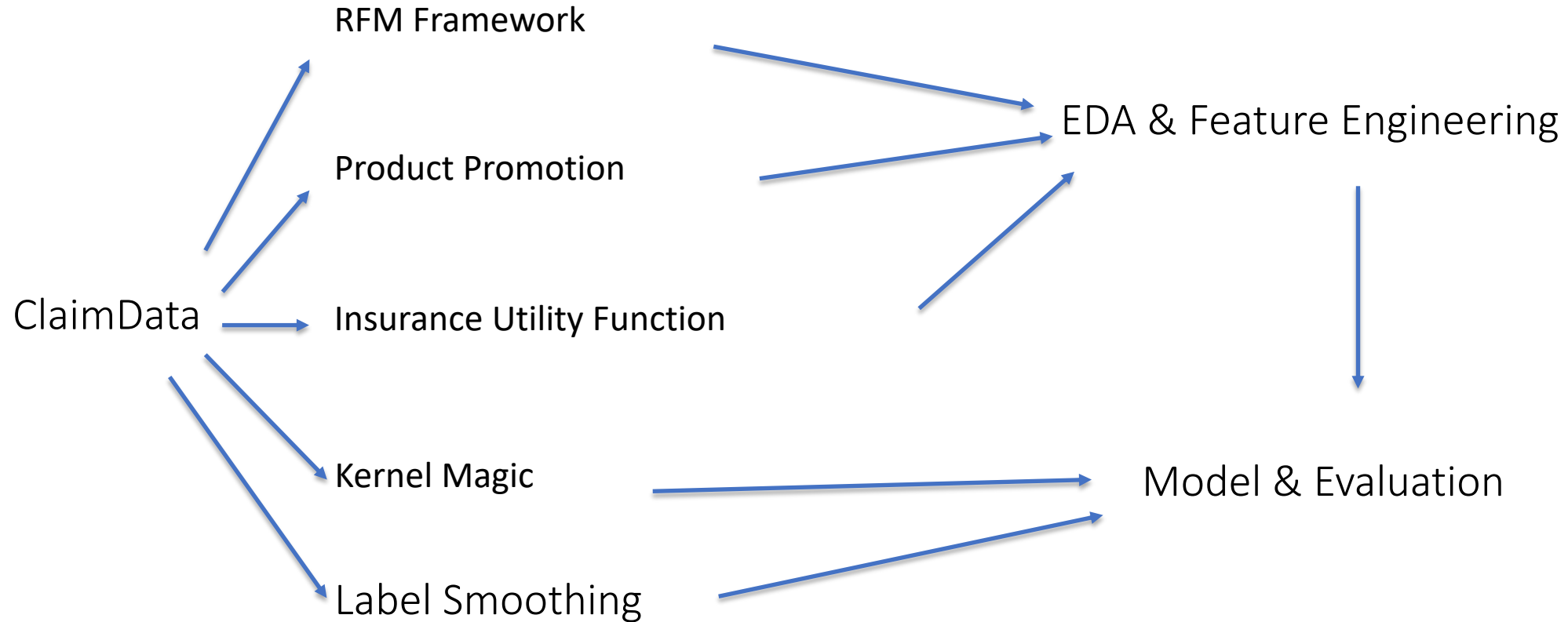


Pipeline



資料期間:

2016-12-27到2018-12-26

資料內容:

- * 理賠檔(CLAIM_ACCT_FIN)
- * 客戶屬性檔(CUST_PROPERTY_FIN)
- * 再購明細檔(COV_ACCT_FIN)
- * 客戶關係檔(CUST_RELATION_FIN)

Raw Data in 資料交換區

	INSURED_RK	Policy_RK	RIDER_CD	payment_period	EFFECTIVE_DT	SHORT_NAME	prod_detail2	POLICY HOLDER_RK	AFYP_NT
0	20	182361	1	年繳	2018-11-30	CAB	AHb	198568	3864.0
1	20	182362	0	年繳	2018-11-30	NPBBR	AHc	198568	3045.0
2	20	182362	0	年繳	2018-11-30	PAR	AHc	198568	472.5
3	20	182362	0	年繳	2018-11-30	AMN	AHc	198568	850.5
4	20	182362	0	年繳	2018-11-30	NHSA	AHa	198568	3675.0
5	20	182362	0	年繳	2018-11-30	HIR	AHa	198568	1690.5
6	20	182362	0	年繳	2018-11-30	ZDDR	AHb	198568	10878.0
7	20	182362	1	年繳	2018-11-30	20PCHI	AHa	198568	15582.0
8	248	58361	1	年繳	2018-02-22	6WPA	AHc	248	63000.0
9	248	179169	1	躉繳保費	2018-09-29	BVA	ILP	248	1050000.0

In [25]: resumetable(data)

Dataset Shape: (180267, 24)

Out[25]:

	Name	dtypes	Missing	Uniques	First Value	Second Value	Third Value	Entropy
0	INSURED_RK	int64	0	48169	20	20	20	14.85690
1	Policy_RK_x	int64	0	74248	182361	182362	182362	15.57480
2	RIDER_CD	int64	0	2	1	0	0	0.99376
3	payment_period	object	0	5	年繳	年繳	年繳	1.19182
4	EFFECTIVE_DT	object	0	730	2018-11-30	2018-11-30	2018-11-30	9.26523
5	SHORT_NAME	object	0	171	CAB	NPBBR	PAR	5.79868
6	prod_detail2	object	0	7	AHb	AHc	AHc	2.52116
7	POLICY HOLDER_RK_x	int64	0	39559	198568	198568	198568	14.51697
8	AFYP_NT	float64	0	16173	3864	3045	472.5	10.67989
9	MATURITY_BENEFICIARY_RK_x	float64	60475	29599	NaN	198568	198568	14.16461
10	DEATH_BENEFICIARY_RK_x	float64	46972	30504	198568	198568	198568	14.15091
11	Y	int64	0	2	0	0	0	0.99645
12	INJURED_RK	float64	91850	16686	NaN	NaN	NaN	13.25710
13	Claim_RK	float64	91850	36480	NaN	NaN	NaN	14.69611
14	Policy_RK_y	float64	91850	19567	NaN	NaN	NaN	13.53732
15	BundleSubtype2	object	91850	6	NaN	NaN	NaN	1.07967
16	illness_code	object	91850	329	NaN	NaN	NaN	5.32387
17	illness_desc	object	91850	333	NaN	NaN	NaN	5.32544
18	DiagnosisCode_DESC	object	91850	21	NaN	NaN	NaN	3.10867
19	claim_settle_dt	object	91850	835	NaN	NaN	NaN	9.10911
20	REIMBURSED_YR_TW	float64	91850	16914	NaN	NaN	NaN	12.19505
21	POLICY HOLDER_RK_y	float64	91850	16735	NaN	NaN	NaN	13.25207
22	MATURITY_BENEFICIARY_RK_y	float64	141590	7555	NaN	NaN	NaN	12.16025
23	DEATH_BENEFICIARY_RK_y	float64	147755	6316	NaN	NaN	NaN	11.92530

Feature Engineering: RFM framework



RFM stands for Recency, Frequency, and Monetary.

指標	定義	量化標準
Recency	最近一次消費	<ul style="list-style-type: none">Recency : 計算train data 保單再購日期-理賠發生日過了幾個月/天/分鐘/秒/幾個工作天, 建立mapping table 然後 transform test data.
Frequency	消費頻率	<ul style="list-style-type: none">Counts:INSURED_RK(理賠+再購)資料筆數計算個人在當年度的移動平均購買件數, 總體購買件數
Monetary	消費金額	<ul style="list-style-type: none">REIMBURSED_YR_TW: 理賠金額(歸至該結案年度)AFYP_NT: 保單保費(&mean)

Feature Engineering: Promotion Effect

Question?

在Train data中,不同計績等行銷活動、激勵制度是否造成季節性熱賣商品?

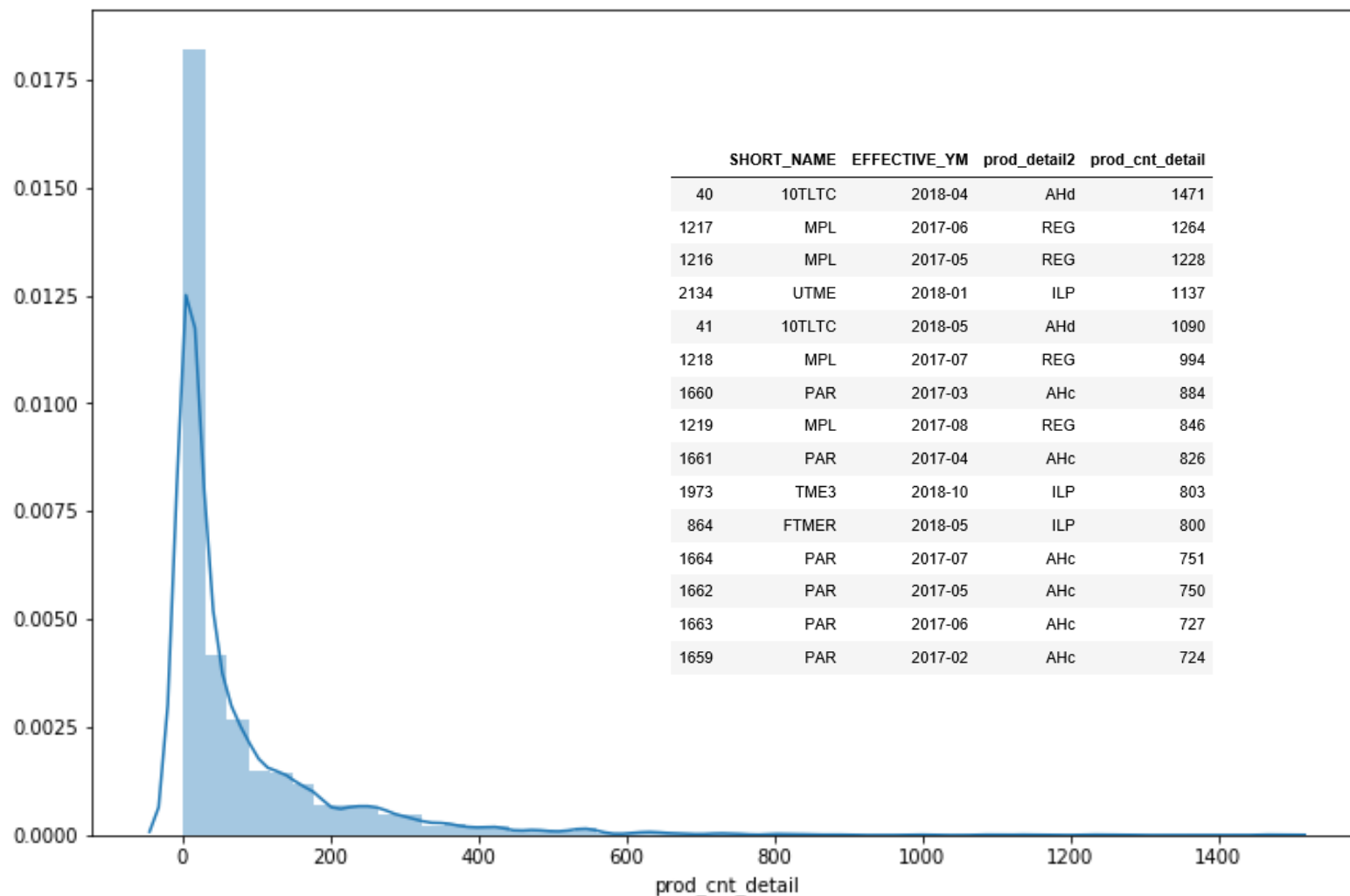
Answer?!

有! 而且分配非常長尾; 據此應該要計算客戶是受業務員影響才購買? 還是本身對壽險商品就有一定程度認知?

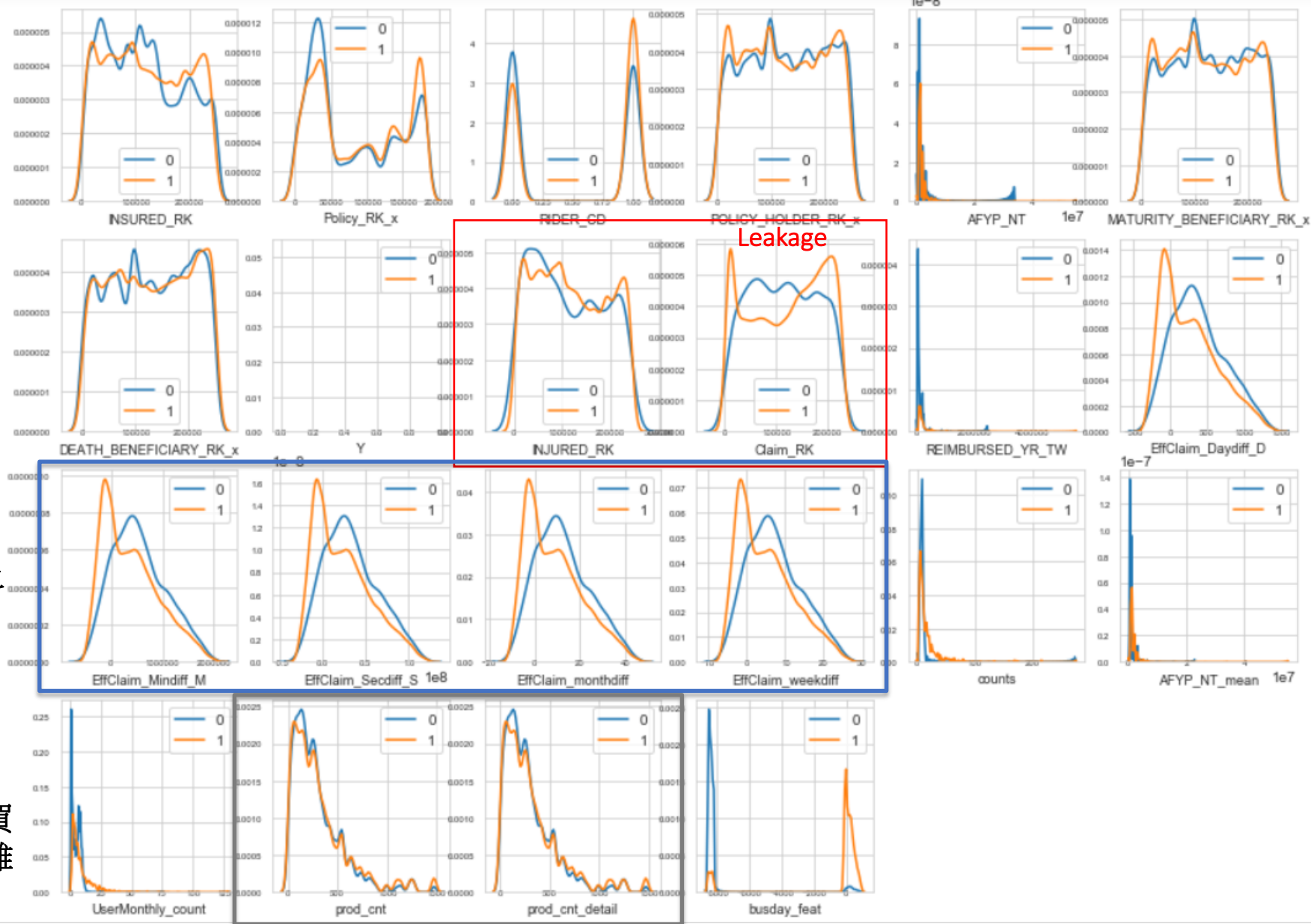
Baseline Solution:

Label Smoothing計算客戶對各商品的偏好程度, 再以行銷力度加權, 分離在相同計績力度下, 各個客戶之

Insurance Utility Function



EDA on continuous variable



再購與理賠日越近的資料
再購情況顯著不同

當月該產品是否熱買
僅在尾部有顯著分離

Dependent Variable Distribution

2016-12-27

2018-08-23

2018-12-26

Train data

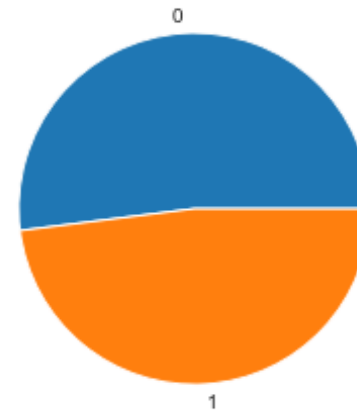
Test data

54.2%

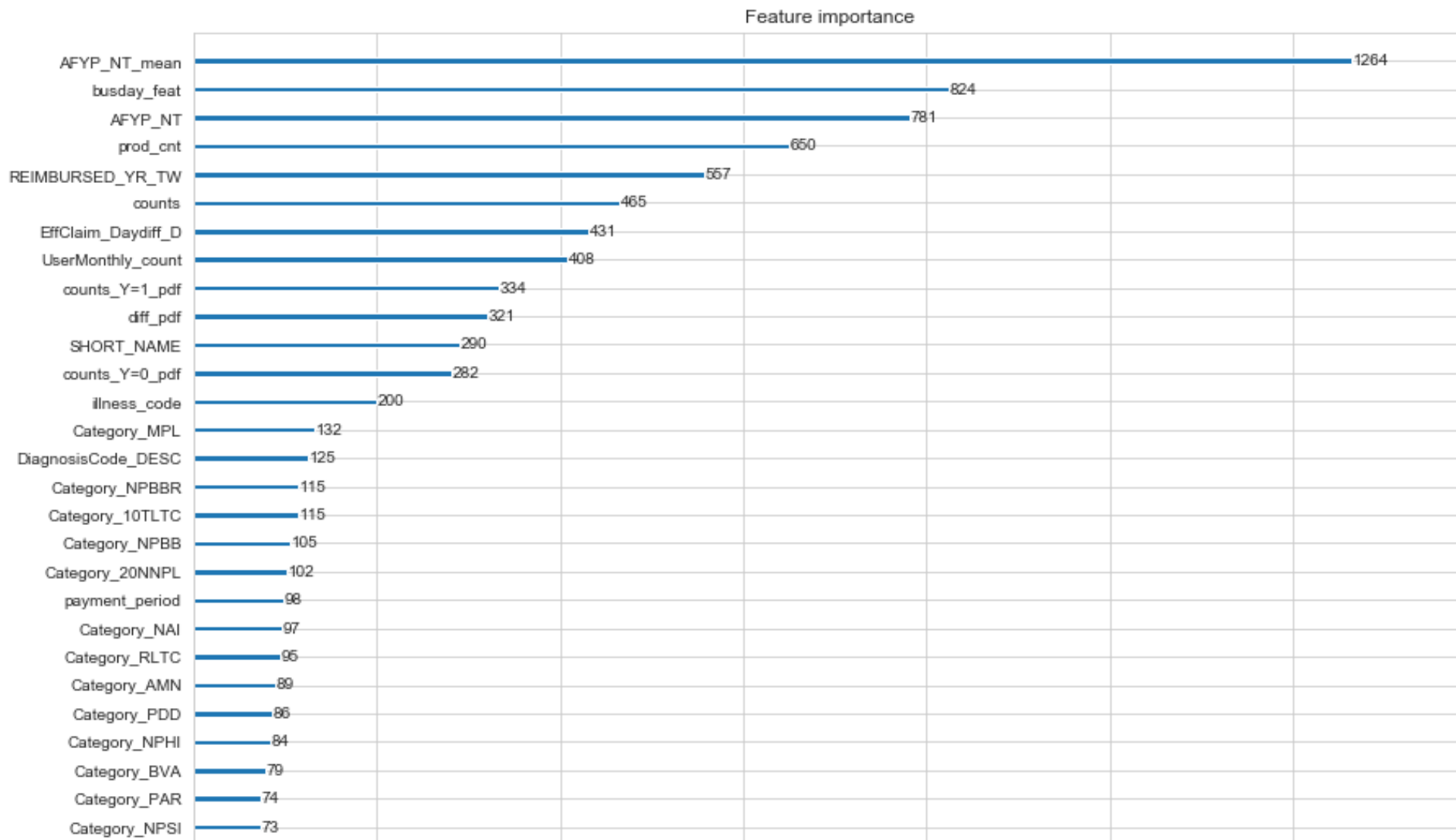


Predict

48%



Model Evaluation



Precision: 0.88589

Recall: 0.88335

F1: 0.88462

	train-logloss-mean	train-logloss-std	test-logloss-mean	\
0	0.509749	0.000202	0.510293	
1	0.410081	0.000028	0.411230	
2	0.349685	0.000200	0.351340	
3	0.311568	0.000240	0.313744	
4	0.285990	0.000012	0.288544	
..	
382	0.110492	0.003730	0.171053	
383	0.110398	0.003783	0.171042	
384	0.110190	0.003600	0.170998	
385	0.109975	0.003438	0.170870	
386	0.109908	0.003398	0.170862	

	test-logloss-std
0	0.000243
1	0.000054
2	0.000225
3	0.000244
4	0.000052
..	...
382	0.000393
383	0.000392
384	0.000349
385	0.000273
386	0.000275

[387 rows x 4 columns]

Summary

1. 在RFM架構下, 日期與金額為客戶分群重要變數, 並就不同分類給予以下建議:

Recency	Frequency	Monetary	建議
高	高	高	運用客戶壽險效用值排名、結合官網保障缺口推薦商品，更全面的推薦適合客戶的商品
低	高	高	運用歷史資料進行Categorical Smoothing，了解客戶壽險偏好並做個體行銷，促進近期末購買高價值客戶的回流

2. 行銷機制對再購行為本來就有影響, 若提供業務員針對個別客戶的Insurance Utility Function相關資訊，將有利於對個別客戶做出差異化銷售，提升再購商機。
3. 透過Kernel可以分辨資料有意義值域與沒意義值域，降低資料探索偏差，提升資訊準確度。
4. 透過此Pipeline試算test data名單內外交易再購率, 名單內為88%, 名單外為10%, **名單內再購率為名單外8倍**。
5. 實務上捨棄Recall拉高Precision，以達資源使用效率最大化。