DACA

NGO 머신러닝 : 2주차

2021.1.1 ~ 2021.1.7

빅데이터 응용학과 석사 1기 양선욱

목차

이항 분석 문제 정의

* 가입 나이대별 이탈비율

가입 나이대	빈도 (a)	가입 나이대 비율 (a/a총합*100)	가입 나이대별 이탈자 수 (b)	가입 나이대별 이탈자 비율 (b/a*100)	전체 가입 나이대별 이탈자 비율 (b/a총합*100)	전체 가입 나이대별 이탈자 비율(100%) (b/b총합*100)
유아	25	2.88%	1	4%	0.11%	0.60%
10대	70	8.08%	27	38.57%	3.11%	16.36%
20대	177	20.43%	44	24.85%	5.08%	26.66%
30대	214	24.71%	26	12.14%	3.00%	15.75%
40대	215	24.82%	38	17.67%	4.38%	23.03%
50대	118	13.62%	22	18.64%	2.54%	13.33%
60대	35	4.04%	6	17.14%	0.69%	3.63%
70세 이상	12	1.38%	1	8.33%	0.11%	0.60%
총합	866	100%	165		19.05%	100%

- 10대 가입자의 38%가 이탈했다.
- 10대 가입자는 전체의 8%이지만 10대 가입자 이탈자는 전체 이탈자의 16%를 차지한다.

이항 분석 문제 정의

* 현재 나이대별 이탈비율

나이대	빈도 (a)	나이대 비율 (a/a총합*100)	나이대 이탈자 (b)	나이대별 이탈자 비율 (b/a*100)	전체 나이대별 이탈자 비율 (b/a총합*100)	전체 나이대별 이탈자 비율(100%) (b/b총합*100)
유아	13	1.85%	0	0	0	0
10대	27	3.85%	0	0	0	0
20대	93	13.26%	39	41.93%	4.50%	23.63%
30대	176	25.10%	35	22.15%	4.04%	21.21%
40대	251	35.80%	35	13.94%	4.04%	21.21%
50대	204	29.10%	37	18.13%	4.27%	22.42%
60대	75	10.69%	16	21.33%	1.84%	9.69%
70세 이상	27	3.85%	3	11.11%	0.34%	1.81%
총합	866	100%	165		19.05%	100%

- 20대의 41.9%가 이탈했다.

이항 분석 문제 정의

* 문제 발견

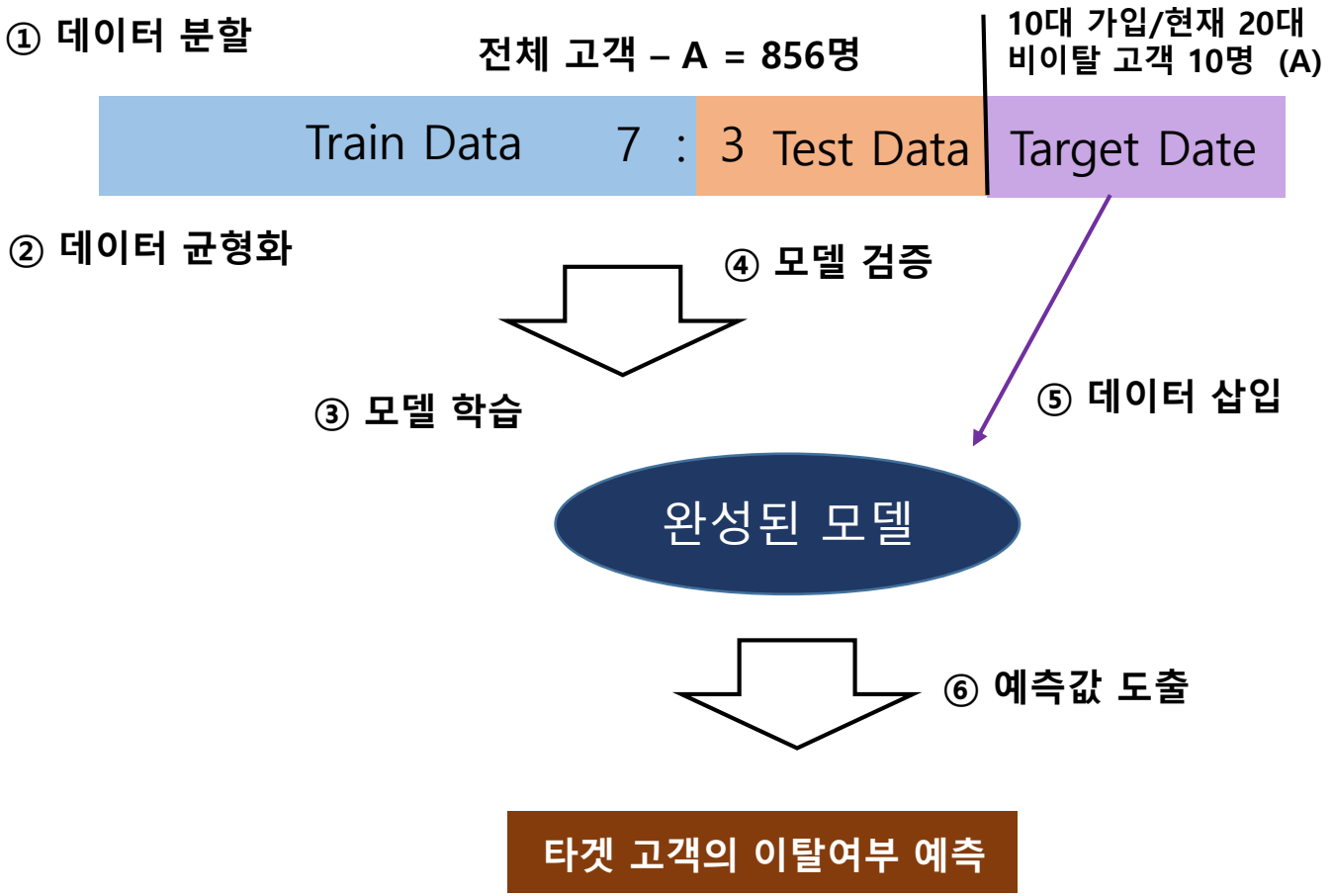
가입 연령대 10대 / 현재 연령대 20대

	인원 수(명)	비율(%)
비이탈자	26	52%
이탈자	24	48%
총합	50	100

* 분석 목적

- 이탈여부를 분류하는 이항분석 모델을 만든다.
- 가입 연령대10대/현재 연령대20대 비이탈자 중에 이탈로 예측되는 고객을 파악한다.

* 분석 과정



다항 분석 문제 정의

* 이탈여부x총 납입금액 구간

	전체		비이탈자			이탈자		
10만 미만	81	9.3%	46	5.3%	6.5%	35 *	4.0%	21.2%
10만 이상 30만 미만	326	37.9%	271	30.2%	38.7%	58	6.6%	35.1%
30만 이상 50만 미만	375	43.3%	316	36.4%	45.0%	59	6.8%	35.7%
50만 이상	81	9.3%	68	7.8%	9.7%	13	1.5%	7.8%
총합	866	100%	701	80.9%	100%	165	19.1%	100%

* 문제 발견

- 이탈 고객의 56.3%는 총 납입금액이 30만원 미만이다.
- 모든 이탈자에 대한 복귀 프로젝트는 실효성이 없어 보인다.

* 분석 목적

- 복귀 시 30만원 이상을 납입할 이탈 고객을 찾아 내고자 한다.

* : 이탈시점 기준

다항 분석 문제 정의

* 분석 과정

① 데이터 분할

비이탈 고객 701명 80.9%

이탈 고객 165명 19.1%

② 모델 학습

③ 모델 평가

④ 이탈 고객 데이터 삽입

완성된 모델

⑤ 이탈 고객의 총납입금액 구간 도출

⑥ 타겟 고객 도출

30만 이상

이탈 고객의 총납입구간

1. 이항 로지스틱 회귀분석

* 사용 데이터

- 독립변수 : 나이, 가입나이, 성별, 가입일수, 납부횟수, 총납입금액, 미납율
- 종속 변수 : 이탈여부

n=856

* 데이터 전처리

- 종속 변수인 이탈 여부의 비율이 4:1이라서 오버 샘플링으로 train data의 비율을 맞추어 준다.
- 독립변수 표준화와 명목형에 대한 원핫인코딩

* 최적의 모델 구하기

	C	학습용정확도	평가용정확도
16	1.7	0.648453608	0.711538462
17	1.8	0.648453608	0.711538462
115	1.7	0.648453608	0.711538462
116	1.8	0.648453608	0.711538462
214	1.7	0.648453608	0.711538462
215	1.8	0.648453608	0.711538462
313	1.7	0.648453608	0.711538462
314	1.8	0.648453608	0.711538462
412	1.7	0.648453608	0.711538462
413	1.8	0.648453608	0.711538462
511	1.7	0.648453608	0.711538462
512	1.8	0.648453608	0.711538462
610	1.7	0.648453608	0.711538462
611	1.8	0.648453608	0.711538462
709	1.7	0.648453608	0.711538462
710	1.8	0.648453608	0.711538462
808	1.7	0.648453608	0.711538462
809	1.8	0.648453608	0.711538462
907	1.7	0.648453608	0.711538462
908	1.8	0.648453608	0.711538462

* 모델 조건

C : 0.1~10 / 간격 0.1

10회 반복

데이터 분할 : 고정

오버 샘플링 : 고정

모델 random

학습용/평가용 정확도 최대

* 모델 결정

C = 1.7

학습용 정확도 : 0.64

평가용 정확도 : 0.71

모델 채택

1. 이항 로지스틱 회귀분석

* 모델평가&분석 결과

학습용 데이터 정확도 : 0.64

평가용 데이터 정확도 : 0.71

교차검증 10회 정확도 평균 : 0.80

ⓐ 모델 Confusion Matrix

	예측비이탈	예측이탈	합계
실제비이탈	168	48	216
실제이탈	27	17	44
합계	195	65	260

ⓑ 모델 Classification Report

	precition	recall	f1	support
비이탈	0.86	0.78	0.82	216
이탈	0.26	0.39	0.31	44
macro avg	0.56	0.58	0.56	260
weighted avg	0.76	0.71	0.73	260

ⓒ 가입10대/현재20대/비이탈 고객 10명에 대한 예측이탈

	AGE	SEX	LONGEVIT	PAY_RATE	PAY_NUM	PAY_SUM	CHURN	가입나이	가입나이인	연령대	예측이탈
43	22	2	57	0	12	360000	0	17	10대	20대	1
133	20	1	37	0.25	9	90000	0	17	10대	20대	1
175	27	2	126	0	24	180000	0	16	10대	20대	0
271	22	2	63	0	12	360000	0	17	10대	20대	1
273	27	2	128	0	12	360000	0	16	10대	20대	0
357	23	2	112	0	13	390000	0	14	10대	20대	0
560	27	2	124	0	12	240000	0	17	10대	20대	0
617	21	2	135	0	24	120000	0	10	10대	20대	0
638	21	2	75	0	12	120000	0	15	10대	20대	0
674	23	2	62	0	12	360000	0	18	10대	20대	1

- 타겟층 10명의 고객 중 이탈할 것이라 예측되는 4명의 고객을 찾았다.
- 이탈에 대한 precition은 26%, recall은 39%이다.

2. 다항 로지스틱 회귀분석

* 사용 데이터

- 독립변수 : 나이, 가입나이, 성별, 가입일수, 납부횟수, 미납율
- 종속 변수 : 총납입금액구간

n=701

* 데이터 전처리

- 독립변수 표준화와 명목형에 대한 원핫인코딩

* 최적의 모델 구하기

	C	solver	학습용정확도	평가용정확도
33	3.4	newton-cg	0.593877551	0.573459716
34	3.5	newton-cg	0.593877551	0.573459716
35	3.6	newton-cg	0.593877551	0.573459716
36	3.7	newton-cg	0.593877551	0.573459716
132	3.4	sag	0.593877551	0.573459716
133	3.5	sag	0.593877551	0.573459716
134	3.6	sag	0.593877551	0.573459716
135	3.7	sag	0.593877551	0.573459716
136	3.8	sag	0.593877551	0.573459716
235	3.8	saga	0.593877551	0.573459716
236	3.9	saga	0.593877551	0.573459716
237	4	saga	0.593877551	0.573459716
238	4.1	saga	0.593877551	0.573459716
239	4.2	saga	0.593877551	0.573459716
240	4.3	saga	0.593877551	0.573459716
241	4.4	saga	0.593877551	0.573459716
330	3.4	newton-cg	0.593877551	0.573459716
331	3.5	newton-cg	0.593877551	0.573459716
332	3.6	newton-cg	0.593877551	0.573459716
333	3.7	newton-cg	0.593877551	0.573459716

* 모델 조건

C : 0.1~10 / 간격 0.1

10회 반복

solver : newton-cg, sag, saga

multi_calss : multinomial

데이터 분할 : 고정

오버 샘플링 : 고정

모델 random

평가용 정확도 최대

* 모델 결정

C = 3.4

solver : newton-cg

multi_class : multinomial

학습용 정확도 : 0.593

평가용 정확도 : 0.573

모델 채택

2. 다항 로지스틱 회귀분석

* 모델 평가

학습용 데이터 정확도 : 0.593

교차검증 10회 정확도 평균 : 0.599

평가용 데이터 정확도 : 0.573

ⓐ 모델 Confusion Matrix

실제/예측	10만 이하	10만 ~ 30만	30만 ~ 50만	50만 이상	합계
10만 이하	4	6	4	0	14
10만~30만	3	30	42	1	76
30만~50만	0	25	73	4	102
50만 이상	0	2	3	14	19
합계	7	63	122	19	211

ⓑ 모델 Classification Report

	precition	recall	f1	support
10만 이하	0.57	0.29	0.38	14
10만~30만	0.48	0.39	0.43	76
30만~50만	0.60	0.72	0.65	102
50만 이상	0.74	0.74	0.74	19
macro avg	0.60	0.53	0.55	211
weighted avg	0.57	0.57	0.56	211

* 분석 결과

ⓒ 이탈 고객 165명 중 복귀 시 30만 이상 납부 예측 고객

구간	인원수	비율
10만 이하	39	23.6%
10만~30만	36	21.8%
30만~50만	78	47.2%
50만 이상	12	7.2%
합계	165	100%

Target

ⓓ 타겟 고객 90명의 prime key인 CONTACT_ID 리스트 확보

	CONTACT_ID
701	498C2EE8-3991-4E47-A5B8-8D77C45A5786
702	321FC308-C6E5-4756-B69D-B214889F64E9
704	7301099D-EE83-4A0D-80E3-984B9AA10937
712	ED70415F-A2EA-4E1B-840C-ACF50C37AB2E
713	00E65724-2054-4935-B8DE-73FB54157E3B
714	E59FAA61-D279-4DD0-8BFC-5CD5BB9112C5
718	8033B55E-EF6D-4D76-AEA9-AECFEE6F4214
719	26C84742-AB86-44FE-98F4-72A41F21C05D
722	5CA0DED2-89D4-4974-9083-D2827D0B4BE7

3. 가우시안 나이브 베이즈 - 이항분석

* 사용 데이터

- 독립변수 : 나이, 가입나이, 성별, 가입일수, 납부횟수, 총납입금액, 미납율
- 종속 변수 : 이탈여부

n=856

* 데이터 전처리

- 종속 변수인 이탈 여부의 비율이 4:1이라서 오버 샘플링으로 train data의 비율을 맞추어 준다.
- 독립변수 표준화와 명목형에 대한 원핫인코딩

* 최적의 모델 구하기

priors	학습용정확도	평가용정확도
[0.9, 0.1]	0.637113402	0.846153846
[0.9, 0.1]	0.637113402	0.846153846
[0.9, 0.1]	0.637113402	0.846153846
[0.9, 0.1]	0.637113402	0.846153846
[0.9, 0.1]	0.637113402	0.846153846
[0.9, 0.1]	0.637113402	0.846153846
[0.9, 0.1]	0.637113402	0.846153846
[0.9, 0.1]	0.637113402	0.846153846
[0.9, 0.1]	0.637113402	0.846153846
[0.9, 0.1]	0.637113402	0.846153846
[0.9, 0.1]	0.637113402	0.846153846
[0.89, 0.11]	0.635051546	0.846153846
[0.88, 0.12]	0.635051546	0.846153846
[0.89, 0.11]	0.635051546	0.846153846
[0.88, 0.12]	0.635051546	0.846153846
[0.89, 0.11]	0.635051546	0.846153846
[0.88, 0.12]	0.635051546	0.846153846
[0.89, 0.11]	0.635051546	0.846153846
[0.88, 0.12]	0.635051546	0.846153846
[0.89, 0.11]	0.635051546	0.846153846
[0.88, 0.12]	0.635051546	0.846153846
[0.89, 0.11]	0.635051546	0.846153846

* 모델 조건

사전확률 : 0.10~0.60 / 0.01간격

var_smoothing : default

10회 반복

데이터 분할 : 고정

오버 샘플링 : 고정

모델 random

평가용 정확도 최대

* 모델 결정

사전확률 : [0.9,0.1]

var_smoothing : default

학습용 정확도 : 0.637

평가용 정확도 : 0.846

모델 채택

3. 가우시안 나이브 베이즈 - 이항분석

* 모델 평가

학습용 데이터 정확도 : 0.638

평가용 데이터 정확도 : 0.846

교차검증 10회 정확도 평균 : 0.808

ⓐ 모델 Confusion Matrix

	예측비이탈	예측이탈	합계
실제비이탈	208	8	216
실제이탈	32	12	44
합계	203	57	260

ⓑ 모델 Classification Report

	precition	recall	f1	support
비이탈	0.87	0.96	0.91	216
이탈	0.60	0.27	0.37	44
macro avg	0.73	0.62	0.64	260
weighted avg	0.82	0.85	0.82	260

* 분석 결과

ⓒ 가입10대/현재20대/비이탈 고객 10명에 대한 예측이탈

AGE	SEX	LONGEVIT	PAY_RATE	PAY_NUM	PAY_SUM	CHURN	가입나이	가입나이인	연령대	예측이탈
22	2	57	0	12	360000	0	17	10대	20대	0
20	1	37	0.25	9	90000	0	17	10대	20대	1
27	2	126	0	24	180000	0	16	10대	20대	0
22	2	63	0	12	360000	0	17	10대	20대	0
27	2	128	0	12	360000	0	16	10대	20대	0
23	2	112	0	13	390000	0	14	10대	20대	0
27	2	124	0	12	240000	0	17	10대	20대	0
21	2	135	0	24	120000	0	10	10대	20대	0
21	2	75	0	12	120000	0	15	10대	20대	0
23	2	62	0	12	360000	0	18	10대	20대	0

ⓓ 이탈 예측 고객 prime key인 CONTACT_ID 리스트 확보

AF463972-42A0-4CAD-AB2D-4B0455933059

3. 가우시안 나이브 베이즈 - 다항분석

* 사용 데이터

- 독립변수 : 나이, 가입나이, 성별, 가입일수, 납부횟수, 미납율
- 종속 변수 : 총납입금액구간

n=701

* 데이터 전처리

- 독립변수 표준화와 명목형에 대한 원핫인코딩

* 최적의 모델 구하기

priors	학습용정확도	평가용정확도
None	0.606122449	0.601895735
[0.065, 0.387, 0.45, 0.098]	0.608163265	0.597156398

* 모델 조건
사전확률 :
[0.065,0.387,0.45,0.098], default
var_smoothing : default
10회 반복
데이터 분할 : 고정
오버 샘플링 : 고정
모델 random

평가용 정확도 최대

* 모델 결정
사전확률 : default
var_smoothing : default
학습용 정확도 : 0.606
평가용 정확도 : 0.601

모델 채택

3. 가우시안 나이브 베이즈 - 다항분석

* 모델 평가

학습용 데이터 정확도 : 0.606　　교차검증 10회 정확도 평균 : 0.586

평가용 데이터 정확도 : 0.601

ⓐ 모델 Confusion Matrix

실제/예측	10만 이하	10만 ~ 30만	30만 ~ 50만	50만 이상	합계
10만 이하	5	6	3	0	14
10만~30만	6	20	48	2	76
30만~50만	0	9	88	5	102
50만 이상	0	1	4	14	19
합계	11	36	143	21	211

ⓑ 모델 Classification Report

	precition	recall	f1	support
10만 이하	0.45	0.36	0.40	14
10만~30만	0.56	0.26	0.36	76
30만~50만	0.62	0.86	0.72	102
50만 이상	0.67	0.74	0.70	19
macro avg	0.57	0.55	0.54	211
weighted avg	0.59	0.60	0.57	211

* 분석 결과

ⓒ 이탈 고객 165명 중 복귀 시 30만 이상 납부 예측 고객

	구간	인원수	비율
Target	10만 이하	55	33.3%
	10만~30만	17	10.3%
	30만~50만	81	49.0%
	50만 이상	12	7.2
	합계	165	100%

ⓓ 타겟 고객 93명의 prime key인 CONTACT_ID 리스트 확보

CONTACT_ID
498C2EE8-3991-4E47-A5B8-8D77C45A5786
321FC308-C6E5-4756-B69D-B214889F64E9
7301099D-EE83-4A0D-80E3-984B9AA10937
1119C4BD-8E41-464C-9AFF-ADAC92DB0EBA
ED70415F-A2EA-4E1B-840C-ACF50C37AB2E
00E65724-2054-4935-B8DE-73FB54157E3B
E59FAA61-D279-4DD0-8BFC-5CD5BB9112C5
AA310335-6F8D-4210-8DAF-A1E840B131AD
8033B55E-EF6D-4D76-AEA9-AECFEE6F4214
26C84742-AB86-44FE-98F4-72A41F21C05D
5CA0DED2-89D4-4974-9083-D2827D0B4BE7
97C93A72-B42C-484E-B407-27CEEB200111

4. 의사결정나무 - 이항분석

* 사용 데이터

- 독립변수 : 나이, 가입나이, 성별, 가입일수, 납부횟수, 총납입금액, 미납율
- 종속 변수 : 이탈여부

n=856

* 데이터 전처리

- 종속 변수인 이탈 여부의 비율이 4:1이라서 오버 샘플링으로 train data의 비율을 맞추어 준다.
- 의사결정나무 특성상 별도의 독립변수 표준화와 명목형에 대한 원핫인코딩을 하지 않음.

* 최적의 모델 구하기

	max_depth	criterion	학습용정확도	평가용정확도
69	7	entropy	0.805154639	0.780769231
89	7	entropy	0.805154639	0.780769231
99	7	entropy	0.805154639	0.780769231
9	7	entropy	0.805154639	0.776923077
19	7	entropy	0.805154639	0.776923077
29	7	entropy	0.805154639	0.776923077
39	7	entropy	0.805154639	0.776923077
49	7	entropy	0.805154639	0.776923077
59	7	entropy	0.805154639	0.776923077
79	7	entropy	0.805154639	0.776923077
3	6	gini	0.782474227	0.773076923
33	6	gini	0.782474227	0.773076923
8	6	entropy	0.779381443	0.769230769
18	6	entropy	0.779381443	0.769230769
28	6	entropy	0.779381443	0.769230769

* 모델 조건

max_depth : 3~7

criterion : gini, entrophy

10회 반복

min_samples_split=2

데이터 분할 : 고정

오버 샘플링 : 고정

모델 random

평가용 정확도 최대

* 모델 결정

max depth : 7

criterion : entrophy

min_samples_split=2

학습용 정확도 : 0.805

평가용 정확도 : 0.780

모델 채택

4. 의사결정나무 - 이항분석

* 모델 평가

학습용 데이터 정확도 : 0.593

평가용 데이터 정확도 : 0.573

교차검증 10회 정확도 평균 : 0.860

ⓐ 모델 Confusion Matrix

	예측비이탈	예측이탈	합계
실제비이탈	181	35	216
실제이탈	22	22	44
합계	203	57	260

ⓑ 모델 Classification Report

	precition	recall	f1	support
비이탈	0.89	0.84	0.86	216
이탈	0.39	0.50	0.44	44
macro avg	0.64	0.67	0.65	260
weighted avg	0.81	0.78	0.79	260

* 분석 결과

ⓒ 가입10대/현재20대/비이탈 고객 10명에 대한 예측이탈

AGE	SEX	LONGEVIT	PAY_RATE	PAY_NUM	PAY_SUM	CHURN	가입나이	가입나이연	연령대	예측이탈
22	2	57	0	12	360000	0	17	10대	20대	0
20	1	37	0.25	9	90000	0	17	10대	20대	1
27	2	126	0	24	180000	0	16	10대	20대	0
22	2	63	0	12	360000	0	17	10대	20대	0
27	2	128	0	12	360000	0	16	10대	20대	0
23	2	112	0	13	390000	0	14	10대	20대	0
27	2	124	0	12	240000	0	17	10대	20대	0
21	2	135	0	24	120000	0	10	10대	20대	0
21	2	75	0	12	120000	0	15	10대	20대	0
23	2	62	0	12	360000	0	18	10대	20대	0

ⓓ 이탈 예측 고객 prime key인 CONTACT_ID 리스트 확보

AF463972-42A0-4CAD-AB2D-4B0455933059

4. 의사결정나무 - 다항분석

* 사용 데이터

- 독립변수 : 나이, 가입나이, 성별, 가입일수, 납부횟수, 미납율
- 종속 변수 : 총납입금액구간

n=701

* 데이터 전처리

- 의사결정나무 특성상 별도의 독립변수 표준화와 명목형에 대한 원핫인코딩을 하지 않음.

* 최적의 모델 구하기

	max_depth	criterion	학습용정확도	평가용정확도
5	3	entropy	0.659183673	0.687203791
25	3	entropy	0.659183673	0.687203791
75	3	entropy	0.659183673	0.687203791
95	3	entropy	0.659183673	0.687203791
15	3	entropy	0.659183673	0.682464455
35	3	entropy	0.659183673	0.682464455
45	3	entropy	0.659183673	0.682464455
55	3	entropy	0.659183673	0.682464455
65	3	entropy	0.659183673	0.682464455
85	3	entropy	0.659183673	0.682464455
17	5	entropy	0.697959184	0.658767773
27	5	entropy	0.697959184	0.658767773
37	5	entropy	0.697959184	0.658767773
67	5	entropy	0.697959184	0.658767773
87	5	entropy	0.697959184	0.658767773
97	5	entropy	0.697959184	0.658767773
0	3	gini	0.659183673	0.654028436
7	5	entropy	0.697959184	0.654028436
10	3	gini	0.659183673	0.654028436

* 모델 조건

max_depth : 3~7

criterion : gini, entropy

10회 반복

min_samples_split=2

데이터 분할 : 고정

오버 샘플링 : 고정

모델 random

평가용 정확도 최대

* 모델 결정

max depth : 3

criterion : entropy

min_samples_split=2

학습용 정확도 : 0.659

평가용 정확도 : 0.687

모델 채택

4. 의사결정나무 - 다항분석

* 모델 평가

학습용 데이터 정확도 : 0.659

교차검증 10회 정확도 평균 : 0.657

평가용 데이터 정확도 : 0.687

ⓐ 모델 Confusion Matrix

실제/예측	10만 이하	10만 ~ 30만	30만 ~ 50만	50만 이상	합계
10만 이하	4	7	3	0	14
10만~30만	0	48	27	1	76
30만~50만	0	19	79	4	102
50만 이상	0	0	5	14	19
합계	4	64	114	19	211

ⓑ 모델 Classification Report

	precition	recall	f1	support
10만 이하	1.0	0.29	0.44	14
10만~30만	0.65	0.63	0.64	76
30만~50만	0.69	0.77	0.73	102
50만 이상	0.74	0.74	0.74	19
macro avg	0.77	0.61	0.64	211
weighted avg	0.70	0.69	0.68	211

* 분석 결과

ⓒ 이탈 고객 165명 중 복귀 시 30만 이상 납부 예측 고객

Target

구간	인원수	비율
10만 이하	22	13.3%
10만~30만	57	34.5%
30만~50만	71	43.0%
50만 이상	15	9.0%
합계	165	100%

ⓓ 타겟 고객 86명의 prime key인 CONTACT_ID 리스트 확보

CONTACT_ID
498C2EE8-3991-4E47-A5B8-8D77C45A5786
321FC308-C6E5-4756-B69D-B214889F64E9
7301099D-EE83-4A0D-80E3-984B9AA10937
283B57D7-52E8-47E6-8569-D9A38FD1304F
31AB51F9-A216-4AAB-9418-A9F4C4A6CDAC
1119C4BD-8E41-464C-9AFF-ADAC92DB0EBA
ED70415F-A2EA-4E1B-840C-ACF50C37AB2E
00E65724-2054-4935-B8DE-73FB54157E3B
E59FAA61-D279-4DD0-8BFC-5CD5BB9112C5
AA310335-6F8D-4210-8DAF-A1E840B131AD
5CA0DED2-89D4-4974-9083-D2827D0B4BE7

5. 보팅 앙상블 - 이항분석

* 사용 데이터

- 독립변수 : 나이, 가입나이, 성별, 가입일수, 납부횟수, 총납입금액, 미납율
- 종속 변수 : 이탈여부

n=856

* 데이터 전처리

- 종속 변수인 이탈 여부의 비율이 4:1이라서 오버 샘플링으로 train data의 비율을 맞추어 준다.
- 독립변수 표준화와 명목형에 대한 원핫인코딩

* 최적의 모델 구하기

1. 이항 로지스틱 회귀분석

C = 1.7

2. 가우시안 나이브 베이즈

*사전확률 : [0.9,0.1]

var_smoothing : default

3. 의사결정 나무

max depth : 7

criterion : entrophy

min_samples_split=2

* 모델 조건

이항 로지스틱 회귀분석 &
가우시안 나이브 베이즈 &
의사결정나무

voting = hard, soft

데이터 분할 : 고정

오버 샘플링 : 고정

	학습용 정확도	평가용 정확도
Soft	0.754	0.765
Hard	0.668	0.707

평가용 정확도 최대

* 모델 결정

voting = soft

학습용 정확도 : 0.637

평가용 정확도 : 0.846

모델 채택

5. 보팅 앙상블 - 이항분석

* 모델 평가

학습용 데이터 정확도 : 0.754

평가용 데이터 정확도 : 0.765

교차검증 10회 정확도 평균 : 0.839

ⓐ 모델 Confusion Matrix

	예측비이탈	예측이탈	합계
실제비이탈	176	40	216
실제이탈	21	23	44
합계	197	63	260

ⓑ 모델 Classification Report

	precition	recall	f1	support
비이탈	0.89	0.81	0.85	216
이탈	0.37	0.52	0.43	44
macro avg	0.63	0.67	0.64	260
weighted avg	0.80	0.77	0.78	260

* 분석 결과

ⓒ 가입10대/현재20대/비이탈 고객 10명에 대한 예측이탈

AGE	SEX	LONGEVIT	PAY_RATE	PAY_NUM	PAY_SUM	CHURN	가입나이	가입나이인	연령대	예측이탈
22	2	57	0	12	360000	0	17	10대	20대	1
20	1	37	0.25	9	90000	0	17	10대	20대	1
27	2	126	0	24	180000	0	16	10대	20대	0
22	2	63	0	12	360000	0	17	10대	20대	1
27	2	128	0	12	360000	0	16	10대	20대	0
23	2	112	0	13	390000	0	14	10대	20대	0
27	2	124	0	12	240000	0	17	10대	20대	0
21	2	135	0	24	120000	0	10	10대	20대	0
21	2	75	0	12	120000	0	15	10대	20대	1
23	2	62	0	12	360000	0	18	10대	20대	1

ⓓ 이탈 예측 고객 5명의 prime key인 CONTACT_ID 리스트 확보

CONTACT_ID
A510F2F4-5B22-46C6-8789-219B075D60B5
AF463972-42A0-4CAD-AB2D-4B0455933059
1C8F9C34-206C-4E90-A444-BDA69130E4CB
420108F3-7C0C-488D-91F4-7211C07DED80
DA88D58F-6775-4B14-B97A-847B5D375AFC

5. 보팅 앙상블 - 다항분석

* 사용 데이터

- 독립변수 : 나이, 가입나이, 성별, 가입일수, 납부횟수, 미납율
- 종속 변수 : 총납입금액구간

n=701

* 데이터 전처리

- 독립변수 표준화와 명목형에 대한 원핫인코딩

* 최적의 모델 구하기

1. 이항 로지스틱 회귀분석

C = 3.4

solver : newton-cg

multi_class : multinomial

2. 가우시안 나이브 베이즈

*사전확률 : default

var_smoothing : default

3. 의사결정 나무

max depth : 3

criterion : entropy

min_samples_split=2

* 모델 조건

이항 로지스틱 회귀분석 &
가우시안 나이브 베이즈 &
의사결정나무

voting = hard, soft

데이터 분할 : 고정

오버 샘플링 : 고정

	학습용 정확도	평가용 정확도
Soft	0.636	0.616
Hard	0.636	0.644

평가용 정확도 최대

* 모델 결정

voting = Hard

학습용 정확도 : 0.636

평가용 정확도 : 0.616

모델 채택

5. 보팅 앙상블 - 다항분석

* 모델 평가

학습용 데이터 정확도 : 0.636

교차검증 10회 정확도 평균 : 0.627

평가용 데이터 정확도 : 0.644

ⓐ 모델 Confusion Matrix

실제/예측	10만 이하	10만 ~ 30만	30만 ~ 50만	50만 이상	합계
10만 이하	4	7	3	0	14
10만~30만	2	30	42	2	76
30만~50만	0	10	88	4	102
50만 이상	0	1	4	14	19
합계	6	48	137	20	211

ⓑ 모델 Classification Report

	precition	recall	f1	support
10만 이하	0.67	0.29	0.40	14
10만~30만	0.62	0.39	0.48	76
30만~50만	0.64	0.86	0.74	102
50만 이상	0.70	0.74	0.72	19
macro avg	0.66	0.57	0.58	211
weighted avg	0.64	0.64	0.62	211

* 분석 결과

ⓒ 이탈 고객 165명 중 복귀 시 30만 이상 납부 예측 고객

Target

구간	인원수	비율
10만 이하	43	26.0%
10만~30만	30	18.1%
30만~50만	80	48.4%
50만 이상	12	7.2%
합계	165	100%

ⓓ 타겟 고객 92명의 prime key인 CONTACT_ID 리스트 확보

CONTACT_ID
498C2EE8-3991-4E47-A5B8-8D77C45A5786
321FC308-C6E5-4756-B69D-B214889F64E9
7301099D-EE83-4A0D-80E3-984B9AA10937
1119C4BD-8E41-464C-9AFF-ADAC92DB0EBA
ED70415F-A2EA-4E1B-840C-ACF50C37AB2E
00E65724-2054-4935-B8DE-73FB54157E3B
E59FAA61-D279-4DD0-8BFC-5CD5BB9112C5
AA310335-6F8D-4210-8DAF-A1E840B131AD
8033B55E-EF6D-4D76-AEA9-AECFEE6F4214
26C84742-AB86-44FE-98F4-72A41F21C05D
5CA0DED2-89D4-4974-9083-D2827D0B4BE7
97C93A72-B42C-484E-B407-27CEEB200111

6. 랜덤 포레스트 - 이항분석

* 사용 데이터

- 독립변수 : 나이, 가입나이, 성별, 가입일수, 납부횟수, 총납입금액, 미납율
- 종속 변수 : 이탈여부

n=856

* 데이터 전처리

- 종속 변수인 이탈 여부의 비율이 4:1이라서 오버 샘플링으로 train data의 비율을 맞추어 준다.
- 랜덤 포레스트 특성상 별도의 독립변수 표준화와 명목형에 대한 원핫인코딩을 하지 않음.

* 최적의 모델 구하기

	n_estimators	max_depth	학습용정확도	평가용정확도
47	109	5	0.811340206	0.807692308
136	127	4	0.781443299	0.807692308
267	153	5	0.808247423	0.807692308
1046	309	4	0.770103093	0.807692308
107	121	5	0.811340206	0.803846154
221	144	4	0.786597938	0.803846154
631	226	4	0.772164948	0.803846154
666	233	4	0.780412371	0.803846154
1087	317	5	0.812371134	0.803846154
1156	331	4	0.775257732	0.803846154
1282	356	5	0.807216495	0.803846154
1461	392	4	0.77628866	0.803846154
1961	492	4	0.784536082	0.803846154
2	100	5	0.810309278	0.8
56	111	4	0.782474227	0.8
67	113	5	0.815463918	0.8

* 모델 조건

n_estimator = 100~500 / 1간격

max_depth = 3~7 / 1간격

데이터 분할 : 고정

오버 샘플링 : 고정

모델 : 랜덤

평가용 정확도 최대

* 모델 결정

n_estimator = 109

max_depth = 5

학습용 정확도 : 0.811

평가용 정확도 : 0.807

모델 채택

6. 랜덤 포레스트 - 이항분석

* 모델 평가

학습용 데이터 정확도 : 0.804

평가용 데이터 정확도 : 0.803

교차검증 10회 정확도 평균 : 0.849

ⓐ 모델 Confusion Matrix

	예측비이탈	예측이탈	합계
실제비이탈	192	24	216
실제이탈	27	17	44
합계	219	41	260

ⓑ 모델 Classification Report

	precition	recall	f1	support
비이탈	0.88	0.89	0.88	216
이탈	0.41	0.39	0.40	44
macro avg	0.65	0.64	0.64	260
weighted avg	0.80	0.80	0.80	260

* 분석 결과

ⓒ 가입10대/현재20대/비이탈 고객 10명에 대한 예측이탈

AGE	SEX	LONGEVIT	PAY_RATE	PAY_NUM	PAY_SUM	CHURN	가입나이	가입나이연	연령대	예측이탈
22	2	57	0	12	360000	0	17	10대	20대	1
20	1	37	0.25	9	90000	0	17	10대	20대	1
27	2	126	0	24	180000	0	16	10대	20대	0
22	2	63	0	12	360000	0	17	10대	20대	1
27	2	128	0	12	360000	0	16	10대	20대	1
23	2	112	0	13	390000	0	14	10대	20대	1
27	2	124	0	12	240000	0	17	10대	20대	1
21	2	135	0	24	120000	0	10	10대	20대	0
21	2	75	0	12	120000	0	15	10대	20대	1
23	2	62	0	12	360000	0	18	10대	20대	1

ⓓ 이탈 예측 고객 8명의 prime key인 CONTACT_ID 리스트 확보

CONTACT_ID
A510F2F4-5B22-46C6-8789-219B075D60B5
AF463972-42A0-4CAD-AB2D-4B0455933059
1C8F9C34-206C-4E90-A444-BDA69130E4CB
5BA9E6A5-AEDD-41CC-B92C-74B3A97B9840
D6BAB8AD-F93B-470F-A6F8-5C8CD3B682A8
E71A9BB7-728D-40F8-A0E1-BE9CD29916F2
420108F3-7C0C-488D-91F4-7211C07DED80
DA88D58F-6775-4B14-B97A-847B5D375AFC

6. 랜덤 포레스트 - 다항분석

* 사용 데이터

- 독립변수 : 나이, 가입나이, 성별, 가입일수, 납부횟수, 미납율
- 종속 변수 : 총납입금액구간

n=701

* 데이터 전처리

- 랜덤 포레스트 특성상 별도의 독립변수 표준화와 명목형에 대한 원핫인코딩을 하지 않음.

* 최적의 모델 구하기

	n_estim	max_dep	학습용정확도	평가용정확도	차이
1568	324	3	0.737149533	0.611374408	0.125775
1638	334	3	0.740654206	0.611374408	0.12928
1603	329	3	0.731308411	0.601895735	0.129413
1120	260	3	0.74182243	0.611374408	0.130448
742	206	3	0.737149533	0.606635071	0.130514
1694	342	3	0.742990654	0.611374408	0.131616
77	111	3	0.738317757	0.606635071	0.131683
2331	433	3	0.744158879	0.611374408	0.132784
35	105	3	0.751168224	0.616113744	0.135054
2415	445	3	0.751168224	0.616113744	0.135054
406	158	3	0.747663551	0.611374408	0.136289
161	123	3	0.738317757	0.601895735	0.136422
126	118	3	0.753504673	0.616113744	0.137391
273	139	3	0.748831776	0.611374408	0.137457
1939	377	3	0.744158879	0.606635071	0.137524
658	194	3	0.739485981	0.601895735	0.13759
2310	430	3	0.739485981	0.601895735	0.13759
568	181	4	0.768691589	0.630331754	0.13836
2444	449	4	0.764018692	0.625592417	0.138426
547	178	4	0.759345794	0.620853081	0.138493
602	186	3	0.740654206	0.601895735	0.138758
2499	457	3	0.740654206	0.601895735	0.138758
2625	475	3	0.735981308	0.597156398	0.138825

- 학습용 0.9, 평가용 0.6 같은 과잉적합이 의심되는 모델들 발견

* 모델 조건

n_estimator = 100~500 / 1간격

max_depth = 3~9 / 1간격

데이터 분할 : 고정

오버 샘플링 : 고정

모델 : 랜덤

학습용 – 평가용 정확도 최소

* 모델 결정

n_estimator = 324

max_depth = 3

학습용 정확도 : 0.737

평가용 정확도 : 0.611

모델 채택

6. 랜덤 포레스트 - 다항분석

* 모델 평가

학습용 데이터 정확도 : 0.740

교차검증 10회 정확도 평균 : 0.593

평가용 데이터 정확도 : 0.597

ⓐ 모델 Confusion Matrix

실제/예측	10만 이하	10만 ~ 30만	30만 ~ 50만	50만 이상	합계
10만 이하	7	4	3	0	14
10만~30만	9	31	34	2	76
30만~50만	0	21	74	7	102
50만 이상	0	3	2	14	19
합계	16	59	113	23	211

ⓑ 모델 Classification Report

	precition	recall	f1	support
10만 이하	0.44	0.50	0.47	14
10만~30만	0.53	0.41	0.46	76
30만~50만	0.65	0.73	0.69	102
50만 이상	0.61	0.74	0.67	19
macro avg	0.56	0.59	0.57	211
weighted avg	0.59	0.60	0.59	211

* 분석 결과

ⓒ 이탈 고객 165명 중 복귀 시 30만 이상 납부 예측 고객

Target

구간	인원수	비율
10만 이하	54	32.7%
10만~30만	32	19.3%
30만~50만	62	37.5%
50만 이상	17	10.3%
합계	165	100%

ⓓ 타겟 고객 79명의 prime key인 CONTACT_ID 리스트 확보

CONTACT_ID
321FC308-C6E5-4756-B69D-B214889F64E9
7301099D-EE83-4A0D-80E3-984B9AA10937
00E65724-2054-4935-B8DE-73FB54157E3B
E59FAA61-D279-4DD0-8BFC-5CD5BB9112C5
AA310335-6F8D-4210-8DAF-A1E840B131AD
8033B55E-EF6D-4D76-AEA9-AECFEE6F4214
26C84742-AB86-44FE-98F4-72A41F21C05D
EFDB8367-99F4-4784-8CFC-891570B4A2E0
00031553-2563-4D99-A4AB-1800882885FE
9A84A716-2755-4A6C-906C-B92036242BF5

7. 그래디언트 부스팅 - 이항분석

* 사용 데이터

- 독립변수 : 나이, 가입나이, 성별, 가입일수, 납부횟수, 총납입금액, 미납율
- 종속 변수 : 이탈여부

n=856

* 데이터 전처리

- 종속 변수인 이탈 여부의 비율이 4:1이라서 오버 샘플링으로 train data의 비율을 맞추어 준다.
- 그래디언트 부스팅 특성상 별도의 독립변수 표준화와 명목형에 대한 원핫인코딩을 하지 않음.

* 최적의 모델 구하기

	n_estim	max_de	learning	학습용정확도	평가용정확도	차이
1694	167	6	0.5	0.998969072	0.823076923	0.175892
1719	168	6	0.5	0.998969072	0.823076923	0.175892
18	100	6	0.4	0.998969072	0.819230769	0.179738
43	101	6	0.4	0.998969072	0.819230769	0.179738
193	107	6	0.4	0.998969072	0.819230769	0.179738
218	108	6	0.4	0.998969072	0.819230769	0.179738
243	109	6	0.4	0.998969072	0.819230769	0.179738
268	110	6	0.4	0.998969072	0.819230769	0.179738
293	111	6	0.4	0.998969072	0.819230769	0.179738
318	112	6	0.4	0.998969072	0.819230769	0.179738
343	113	6	0.4	0.998969072	0.819230769	0.179738
612	124	5	0.3	0.998969072	0.819230769	0.179738
637	125	5	0.3	0.998969072	0.819230769	0.179738
943	137	6	0.4	0.998969072	0.819230769	0.179738
968	138	6	0.4	0.998969072	0.819230769	0.179738
993	139	6	0.4	0.998969072	0.819230769	0.179738
1018	140	6	0.4	0.998969072	0.819230769	0.179738
1087	143	5	0.3	0.998969072	0.819230769	0.179738
1112	144	5	0.3	0.998969072	0.819230769	0.179738
1137	145	5	0.3	0.998969072	0.819230769	0.179738
1162	146	5	0.3	0.998969072	0.819230769	0.179738
1444	157	6	0.5	0.998969072	0.819230769	0.179738
1544	161	6	0.5	0.998969072	0.819230769	0.179738

* 모델 조건

n_estimator = 100~500 / 1간격

max_depth = 3~7 / 1간격

learning_rate = 0.1~0.5 / 0.1간격

데이터 분할 : 고정

오버 샘플링 : 고정

모델 : 고정

평가용 정확도 최대

* 모델 결정

n_estimator = 167

max_depth = 6

learning_rate = 0.5

학습용 정확도 : 0.998

평가용 정확도 : 0.823

모델 채택

7. 그래디언트 부스팅 - 이항분석

* 모델 평가

학습용 데이터 정확도 : 0.998

평가용 데이터 정확도 : 0.823

교차검증 10회 정확도 평균 : 0.851

ⓐ 모델 Confusion Matrix

	예측비이탈	예측이탈	합계
실제비이탈	194	22	216
실제이탈	24	20	44
합계	218	42	260

ⓑ 모델 Classification Report

	precition	recall	f1	support
비이탈	0.89	0.90	0.89	216
이탈	0.48	0.45	0.47	44
macro avg	0.68	0.68	0.68	260
weighted avg	0.82	0.82	0.82	260

* 분석 결과

ⓒ 가입10대/현재20대/비이탈 고객 10명에 대한 예측이탈

AGE	SEX	LONGEVIT	PAY_RATE	PAY_NUM	PAY_SUM	CHURN	가입나이	가입나이연	연령대	예측이탈
22	2	57	0	12	360000	0	17	10대	20대	0
20	1	37	0.25	9	90000	0	17	10대	20대	0
27	2	126	0	24	180000	0	16	10대	20대	0
22	2	63	0	12	360000	0	17	10대	20대	0
27	2	128	0	12	360000	0	16	10대	20대	0
23	2	112	0	13	390000	0	14	10대	20대	0
27	2	124	0	12	240000	0	17	10대	20대	0
21	2	135	0	24	120000	0	10	10대	20대	0
21	2	75	0	12	120000	0	15	10대	20대	0
23	2	62	0	12	360000	0	18	10대	20대	1

ⓓ 이탈 예측 고객 1명의 prime key인 CONTACT_ID 리스트 확보

DA88D58F-6775-4B14-B97A-847B5D375AFC

7. 그래디언트 부스팅 - 다항분석

* 사용 데이터

- 독립변수 : 나이, 가입나이, 성별, 가입일수, 납부횟수, 미납율
- 종속 변수 : 총납입금액구간

n=701

* 데이터 전처리

- 그래디언트 부스팅 특성상 별도의 독립변수 표준화와 명목형에 대한 원핫인코딩을 하지 않음.

* 최적의 모델 구하기

	n_estim	max_de	learning	학습용정확도	평가용정확도
5	50	4	0.1	0.918224299	0.658767773
0	50	3	0.1	0.877336449	0.654028436
25	60	3	0.1	0.885514019	0.654028436
30	60	4	0.1	0.926401869	0.6492891
50	70	3	0.1	0.897196262	0.644549763
75	80	3	0.1	0.90771028	0.639810427
1	50	3	0.2	0.919392523	0.639810427
100	90	3	0.1	0.917056075	0.63507109
26	60	3	0.2	0.933411215	0.630331754
80	80	4	0.1	0.956775701	0.630331754
15	50	6	0.1	0.988317757	0.630331754
150	110	3	0.1	0.931074766	0.625592417
175	120	3	0.1	0.936915888	0.625592417
55	70	4	0.1	0.949766355	0.625592417
40	60	6	0.1	0.994158879	0.625592417
125	100	3	0.1	0.925233645	0.620853081
200	130	3	0.1	0.943925234	0.620853081
105	90	4	0.1	0.961448598	0.620853081
65	70	6	0.1	0.996495327	0.620853081
51	70	3	0.2	0.942757009	0.616113744
90	80	6	0.1	0.996495327	0.616113744
225	140	3	0.1	0.947429907	0.611374408
12	50	5	0.3	0.996495327	0.611374408

* 모델 조건

n_estimator = 50~400 / 10간격

max_depth = 3~7 / 1간격

learning_rate = 0.1~0.5 / 0.1간격

데이터 분할 : 고정

오버 샘플링 : 고정

모델 : 고정

평가용 정확도 최대

* 모델 결정

n_estimator = 50

max_depth = 4

learning_rate = 0.1

학습용 정확도 : 0.918

평가용 정확도 : 0.658

모델 채택

7. 그래디언트 부스팅 - 다항분석

* 모델 평가

학습용 데이터 정확도 : 0.918

교차검증 10회 정확도 평균 : 0.629

평가용 데이터 정확도 : 0.658

ⓐ 모델 Confusion Matrix

실제/예측	10만 이하	10만 ~ 30만	30만 ~ 50만	50만 이상	합계
10만 이하	5	5	4	0	14
10만~30만	2	50	22	2	76
30만~50만	0	25	70	7	102
50만 이상	0	2	3	14	19
합계	7	82	99	23	211

ⓑ 모델 Classification Report

	precition	recall	f1	support
10만 이하	0.71	0.36	0.48	14
10만~30만	0.61	0.66	0.63	76
30만~50만	0.71	0.69	0.70	102
50만 이상	0.61	0.74	0.67	19
macro avg	0.66	0.61	0.62	211
weighted avg	0.66	0.66	0.66	211

* 분석 결과

ⓒ 이탈 고객 165명 중 복귀 시 30만 이상 납부 예측 고객

	구간	인원수	비율
	10만 이하	46	
	10만~30만	51	
Target	30만~50만	53	
	50만 이상	15	
	합계	165	100%

ⓓ 타겟 고객 68명의 prime key인 CONTACT_ID 리스트 확보

321FC308-C6E5-4756-B69D-B214889F64E9
7301099D-EE83-4A0D-80E3-984B9AA10937
31AB51F9-A216-4AAB-9418-A9F4C4A6CDAC
1119C4BD-8E41-464C-9AFF-ADAC92DB0EBA
00E65724-2054-4935-B8DE-73FB54157E3B
E59FAA61-D279-4DD0-8BFC-5CD5BB9112C5
AA310335-6F8D-4210-8DAF-A1E840B131AD
EFDB8367-99F4-4784-8CFC-891570B4A2E0
00031553-2563-4D99-A4AB-1800882885FE
1385CC3B-7F53-4E4B-97B7-0ACA2A5CACFD
1F5EAE34-D38C-4241-8671-40D6177F1475
2EAF8719-D7B0-48DD-833E-67104A35C48D
9969C89C-1354-4490-8537-420F7AD7967D
38B13B1B-3CAC-481E-AC47-1A4C5E8E9377

모델 비교 - 이항분석

*가입나이10대, 현재20대, 비이탈 고객 10명 중 이탈로 예측 된 고객에 O

분석 방법	학습용 정확도	평가용 정확도	교차검증	'이탈' F1	고객1	고객2	고객3	고객4	고객5	고객6	고객7	고객8	고객9	고객10
로지스틱 회귀분석	0.64	0.71	0.80	0.31	O	O		O						O
가우시안 나이브 베이즈	0.638	0.575	0.868	0.44		O								
의사결정나무	0.593	0.597	0.860	0.44		O								
보팅 앙상블	0.754	0.765	0.839	0.43	O	O		O					O	O
랜덤 포레스트	0.804	0.803	0.849	0.40	O	O		O	O	O	O		O	O
그래디언트 부스팅1	0.998	0.823	0.851	0.47										O
그래디언트 부스팅2	0.883	0.765	0.867	0.37	O	O		O	O	O				O

그래디언트 부스팅	n_estimators	max_depth	learning_rate
1 평가용 최대	167	5	0.5
2 학습용-평가용 최소	100	3	0.1

모델 비교 - 다항분석

분석 방법	학습용 정확도	평가용 정확도	교차검증	'30만~50만' F1	'50만 이상' F1	예측인원
로지스틱 회귀분석	0.593	0.573	0.599	0.35	0.74	90
가우시안 나이브 베이즈	0.606	0.601	0.586	0.72	0.70	93
의사결정나무	0.659	0.687	0.657	0.73	0.74	86
보팅 앙상블	0.636	0.644	0.627	0.74	0.72	92
랜덤 포레스트	0.740	0.597	0.593	0.69	0.67	79
그래디언트 부스팅	0.918	0.658	0.629	0.70	0.67	68

	6모델	5모델	4모델	3모델	2모델	1모델	합계
인원	45	26	21	3	3	9	107