

2조 분석_1

헤지 향운

보험사 수익 감소

- 원인 1 : 고객 승인 모델 오류(가입 거절 비율이 높음)

 청구 대비 지급을 기준으로 고객 위험도 분류 모델 재정의

- 원인 2 : 신규 가입자 감소, 시장 포화

 월별 신규 가입자 추세 파악

원인 1) 청구 대비 지급 고객 위험도 분류 모델 재정의

1분류 : CUSTOMER_ID 당

Y1 = 1차 : (청구 - 지급)

2차 : ratio(본인 부담률) (정액 X)

Y2 = 지급액(정액)

2분류 : 위험 고객 집단 중 저, 고 위험 판단

Y1 = (청구 - 지급 = 본인부담금) / 총 청구액 = 본인 부담률 (정액 X)

Y2 = 지급액 (정액)

1분류:수익 고객과 위험 고객 분류

청구-지급=본인부담금, 비정액 보험의 경우, 병원비의 일정 금액, 일정 비율은 본인이 부담하고 있음!
만약 본인부담금이 음수인 경우(과 지급)
EX) 병원비 50000을 청구, 보험에서 40000원을 지급한 경우
본인 부담금은 10000원

본인 부담금이 존재하지만, 300원, 3원, 1원 등 거의 본인이 부담하지 않는 경우도 존재 ->
청구-지급 내에서 일정 수준 이상을 부담 한 경우
수익 고객으로 분류(본인 부담률 ratio 도입)
Ratio = (청구-지급)/총 청구액

전체 청구 고객
4525

정액 X 4048

593

청구-지급>0

Ratio >= 10%

수익 350

Ratio < 10%

위험 243

청구-지급<=0
(청구 대비
동일 or 과 지급된 경우)

위험 3455

정액 477

지급<median

수익 93

지급>=median

위험 384

1분류 : 모델링 결과_혈액검사, 일반검사 분리(data 활용 ↑)

일반검진

RF, 초기

Accuracy on training set	1.000
Accuracy on validation set	0.889
Accuracy on test set	0.865

일반 검진을 한 거절 고객 483 명 → RF 모델에 적용

Risk 0(수익성 고객) : 7명 -> 가입 유도(할증 X)

Risk 1(위험 고객) : 476명 -> 2분류 진행

혈액 검사

GB, 초기

Accuracy on training set	0.902
Accuracy on validation set	0.910
Accuracy on test set	0.913

혈액 검사를 한 거절 고객 2344 명 → GB 모델에 적용

Risk 0(수익성 고객) : 97명 -> 가입 유도(할증 X)

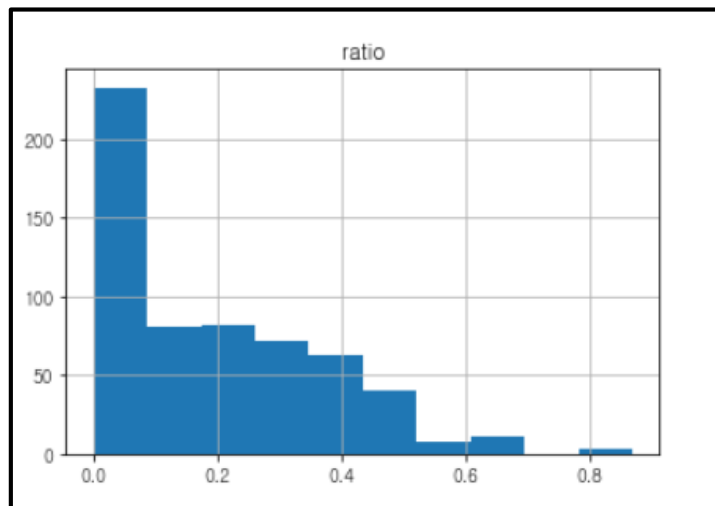
Risk 1(위험 고객) : 2344명 -> 2분류 진행

1분류 : 추가 설명

비 정액

dif_amount	
count	4.048000e+03
mean	-4.454558e+05
std	3.289536e+06
min	-7.266000e+07
25%	-2.000000e+05
50%	0.000000e+00
75%	0.000000e+00
max	6.947112e+06

청구-지급 = 0 인 고객이 다수 존재
-> 현 상황을 고려하면 청구 = 지급은
수익보다는 위험 우려를 가진 요소로 판
단 -> 위험으로 분류



히스토그램을 통해, 본인부담
금이 존재하지만 10%도 안되
는 경우가 다수 존재
-> 데이터 분포가 특정 비율
에 밀집
-> 10% 미만의 가입자는 위
험 우려가 존재한다고 판단하
여 위험으로 분류

정액

```
[ ]:
```

insu_pay_amount	
count	4.770000e+02
mean	1.133858e+06
std	6.486067e+06
min	3.000000e+04
25%	2.000000e+05
50%	2.000000e+05
75%	4.000000e+05
max	1.000000e+08

요약 통계량을 통해 20만원 다수 확인
-> 대부분 20만원을 지급일 때, 그 20만원보다
적은 지급액을 받은 고객을 수익이라 판정
-> 현 상황은 보험사의 수익 개선이 필요하기에
20만원은 수익성보다는 위험 우려를 가진
요소로 보는 것이 옳다 판단(위험으로 분류)

2분류: 저 위험, 고 위험 분류 $\text{Ratio} = (\text{청구} - \text{지급}) / \text{총 청구액}$

Ratio는 본인 부담률 ->
클수록 보험사에 수익이
되는 고객
Ratio가 작을수록 청구
대비 과 지급된 금액이
많다.

1분류 위험 고객
4082

정액 X 3698

$\text{Ratio} \geq Q1 - 1.5IQR$
3132

저

$\text{Ratio} < Q1 - 1.5IQR$
566

고

Ratio가 작을수록 위험
저 위험을 할증으로 최대
한 확보 필요
-> 최소 한계치인 Box
plot 내 $Q1 - 1.5IQR$ 사용

정액 384

$\text{지급} \leq Q3 + 1.5IQR$
331

저

$\text{지급} > Q3 + 1.5IQR$
53

고

지급액이 많을수록 손해
저 위험을 최대한 확보하
여 할증으로 유도 필요
Boxplot 내 MAX 한계
인 $Q3 + 1.5IQR$ 사용

2분류 : 모델링 결과_혈액검사, 일반검사 분리(data 활용 ↑)

일반 검진

GB, 초기

Accuracy on training set	0.928
Accuracy on validation set	0.810
Accuracy on test set	0.817

거절 고객 적용

일반 검진 1분류 위험 고객 476명 → RF 모델에 적용

- Risk 0(저위험 고객) : 462명
- Risk 1(고위험 고객) : 14명

일반 검진에서 거절을 당한 고객인만큼 고위험군 고객일 가능성이 높음

혈액 검사

GB, 초기

Accuracy on training set	0.856
Accuracy on validation set	0.834
Accuracy on test set	0.857

거절 고객 적용

혈액 검사 1분류 위험 고객 2274 명 → GB 모델에 적용

- Risk 0(저위험 고객) : 2200명 ->할증 도입
- Risk 1(고위험 고객) : 47명
Boxplot 밖 고위험 (이상치)와 유사 특성 -> 위험 우려가 크므로
거절 유지

향후 계획

할증 및 상품 추천

할증 고객의 수(Train 3400 거절 Predict 2500)가 생각보다 많음

-> **군집**을 나눠 군집별 가입 보험, 상병, 가입 시기, 납부 기간, 납부 액을 분석하여 적절한 **상품 추천**이나 **할증 정도**를 결정하면 좋을 것 같음

1분류에서 수익으로 판정된 고객의 경우,

-> 수익성 고객 집단의 상병, 보험, 가입 시기 및 납부 기간, 납부 액을 분석하여 **상품 추천** 제시

건강검진_screen

본인부담금을 기준으로 위험도를 산정_모델1
질병을 기준으로 위험도를 산정_모델2
~~을 기준으로 위험도를 산정_모델3
=> 위험도에 대한 복합적인 평가 => 최종 위험도 산정

1. 분류 기준 변수들로(AGE, GENDER, HEIGHT 등)으로 건강검진 데이터를 위험군일지 예측해보는 것도 좋을 듯

-> 국민 건강으로 특히 위험이 높은 연령은 **엄격한 모델**을 도입, 다양한 평가 기준 적용
-> 위험이 낮은 연령은 주 타겟으로 선정하는 것도 좋을 것 같음!

2. 월별 신규 가입자 분석 -> 주로 어떤 연령의 가입자가 늘고 있는지?
기존 가입자 중 연령의 분포가 어떻게 되는지 파악

EX) 국민 건강 데이터에선 20대의 위험이 적다고 판정이 된 경우,
우리 보험의 20대가 적다고 나온다면 => 20대를 유인하는 대안 제시
기존 가입자 중 20대가 선호하는 보험 상품의 종류는? 주 상병은?

3. 흡연, 음주 등 추가적인 데이터

Screen 데이터에 분류 모델을 적용해 위험도를 산정(0,1)

분류 모델을 통해 나온 값이 흡연, 음주 여부와 연관이 존재하는지 파악

⇒ 상관관계가 존재한다고 판단되는 경우, 건강 검진 항목에 흡연/음주도 추가

⇒ 흡연, 음주가 유의한 경우 -> 흡연, 음주와 연관이 깊은 설명 변수도 파악 해보기!