

고객 특성에 기반한 가입 심사 모델 개선 및 상품 추천을 통한 포빅생명 수익 극대화 전략

A2

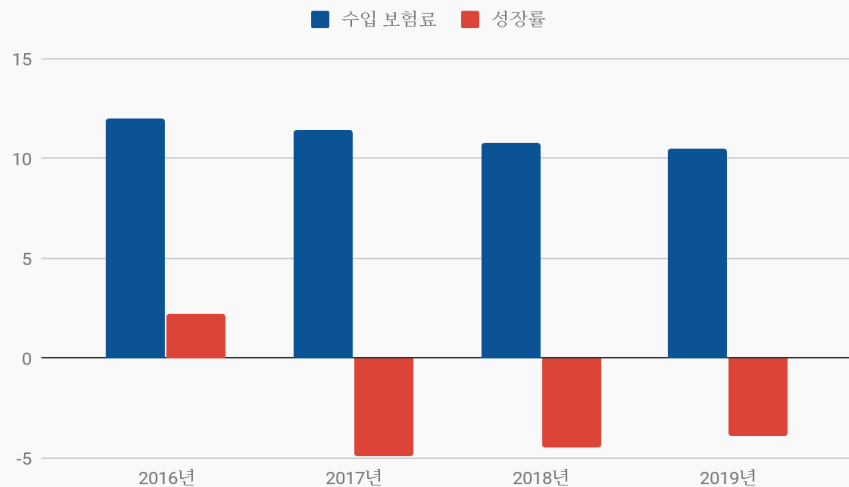
강지영, 김범수, 김효진, 배향운, 양혜지, 정지성, 최지영

1. 추진배경
2. 현상 파악
3. 데이터 분석
4. 개선 방안
5. 개별 소감

1. 추진배경 "생명보험 업계와 마찬가지로 수익 보험료가 감소하는 상황"

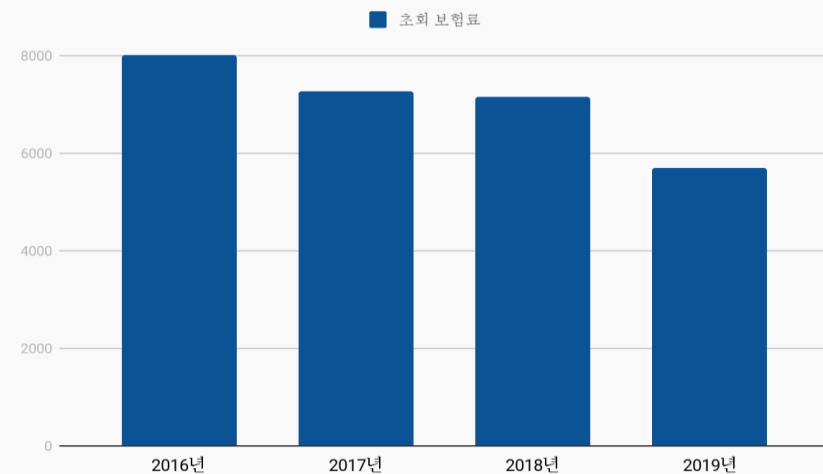
● 생명 보험 업계 상황

- 수입 보험료 계속 감소
- 영업력 지표인 초회 보험료 3년 사이 40.5% 급감
- 수입보험료는 감소했지만 지급보험금 증가



● 포빅 생명

- 2019년 초회보험료 5690억원
- 포빅 생명 또한 초회 보험료 감소
- 가입 거절 비율 18.9%



※ 초회 보험료: 신규 가입자가 낸 첫 보험료, 통상적으로 보험사의 영업력을 측정하는 척도로 사용됨

2. 현상 파악 “보험사의 수익성 하락원인 도출”

● 수익성 하락 원인

1 신규 고객 유치 한계



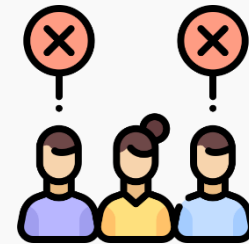
가구당 보험 가입률 98.4%
생명 보험 시장 포화 상태로
새로운 고객 유치 어려움

2 과지급 발생



고객 약 15000명 중
납입, 청구 대비 과다지급
30% 발생

3 기존 가입 심사 모델의 오류



혈액, 소변 검사 결과 및
과거 병력 등 엄격한 건강
심사 기준으로 보험 가입
거절 비율(18.9%)이 높음

2. 현상 파악 “프로젝트 목표와 수익성 하락 원인을 해결할 방안 제시”

● 개선 시나리오

1 신규 고객 유치 한계



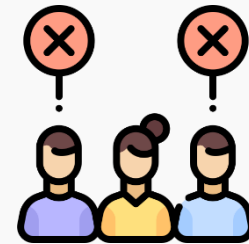
거절 고객을 개선된
분류 모델을 통해
신규 고객으로 유인

2 과지급 발생



과지급이 예상되는 고객에게
보험 상품 추천 및 할증

3 기존 가입 심사 모델의 오류



건강 검진 정보 등을
활용해 심사 항목을 보완하여
분류의 정확도 향상

목표 : 고객 분류 및 보험 상품 추천 모델을 통한 수익성 10% 향상

3. 데이터 분석 “잠재원인을 해결하기 위한 데이터를 수집하고 정제함”

● 데이터 수집 및 정제

* BMI = 키/체중²

* Whtr = 허리둘레/신장

	데이터 수	내용	목적	결측치 변수		이상치 변수	파생변수
보험 가입 검진 정보	14,939	보험 가입, 거절 고객 신체 정보	고객 분류 모델 수립	혈액 多	일반 6	-	BMI Whtr
보험 청구 정보	49,450	고객의 보험 지급, 청구 목록	고객 분류 및 질병 예측 모델 수립	52		-	비율 금액
국민 건강 검진 정보	177,347	공공 데이터	분류 정확도 향상을 위한 추가 변수 선정	-		-	-
보험 상품별 보험료	30	보험 상품 정보	보험 상품 추천 시스템	-		-	-
상병 정보	13,756	병에 대한 정보	질병 파악 및 예측 모델 수립	-		-	-

3. 데이터 분석 “분석 목적과 데이터 셋에 맞추어 계획을 수립”

● 데이터 분석 계획

목적	데이터	분석 방법	주요내용	담당자
전체 데이터의 분포 특성 및 변수 간의 관련성 확인	보험 가입 검진 정보	막대그래프 분석	가입 거절 고객의 연령대 분석	김효진
	보험 청구 정보	꺾은선 그래프 분석	월별 가입 인원과 보험사 수익의 연관 분석	최지영
	전체 데이터	Box Plot 분석	전체 데이터의 이상치 존재 여부 확인	김범수
	보험 가입 검진 정보	상관분석	혈액검사 판정결과와 BMI지수 간 상관관계 분석	김범수
상병 정보 분석	보험 청구 정보, 상병 정보	군집 분석	청구 금액이 높은 질병 확인	최지영
		연관 분석	질병 별 청구 금액과 청구 빈도 확인	정지성
영향 인자 분석	보험 가입 검진 정보, 보험 청구 정보	회귀 분석 Decision Tree Random Forest Gradient Boosting	청구 - 지급 금액에 영향을 주는 영향 인자 분석	양혜지 배향운 강지영 김효진
			납입 - 지급 금액에 영향을 주는 영향 인자 분석	
건강검진 정보 분석	국민 건강 검진 정보	상관 분석	음주, 흡연 여부와 혈액검사 판정결과 간 상관관계 분석	배향운

3. 데이터 분석 “납입 금액을 기준으로 위험 고객 분류”

● 고객 분류 모델링 분석 결과(납입 기준)

고객 특성

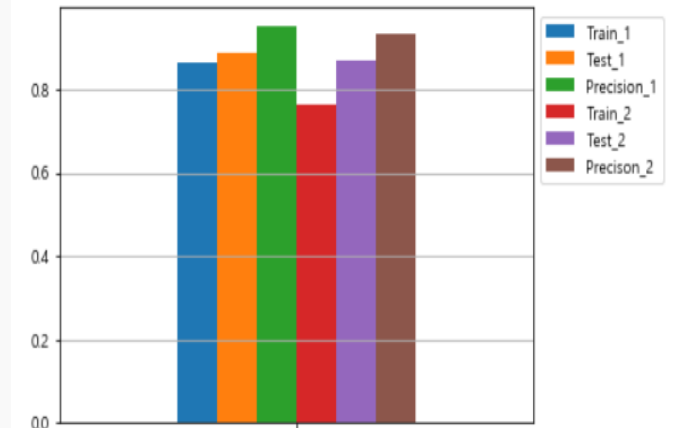
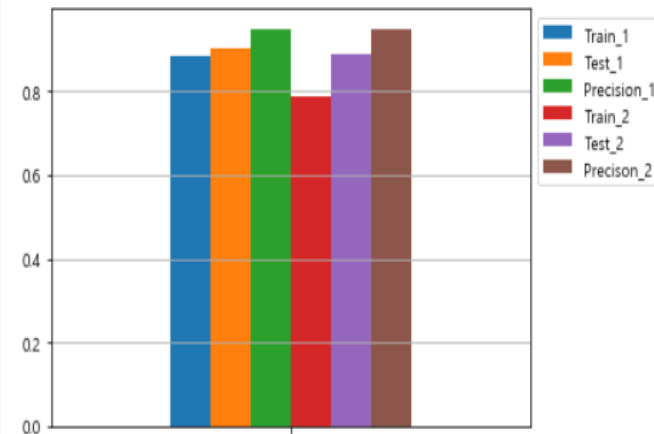


납입 이력

- 목표변수 : 개인별 총 납입액-지급액
- 보험 가입 검진 정보, 보험 청구 데이터 이용
- 일반(985명)과 혈액 검진(2456명)으로 분리
- 1분류 : 위험 / 비위험 고객 분류
- 2분류 : 할증 / 거절 고객 분류
- 신규 고객 유치를 위해 가입 거절했던 고객들을 대상으로 위험도 재 산정 후 고객 유인/할증/거절

거절 고객	수익	할증	거절	합계
일반	452	25	6	483
혈액	1458	850	11	2319

- 위험 고객 분류 Train, Test 정확도 및 정밀도 비교



일반	TRAIN	TEST	PRECISION
1분류	0.885	0.905	0.950
2분류	0.789	0.890	0.950

혈액	TRAIN	TEST	PRECISION
1분류	0.865	0.887	0.950
2분류	0.763	0.869	0.931

Gradient Boosting 모델을 이용해 수익, 할증, 거절 고객으로 분류

3. 데이터 분석 “청구 금액을 기준으로 위험 고객 분류 ”

● 고객 분류 모델링 분석 결과(청구 기준)

고객 특성

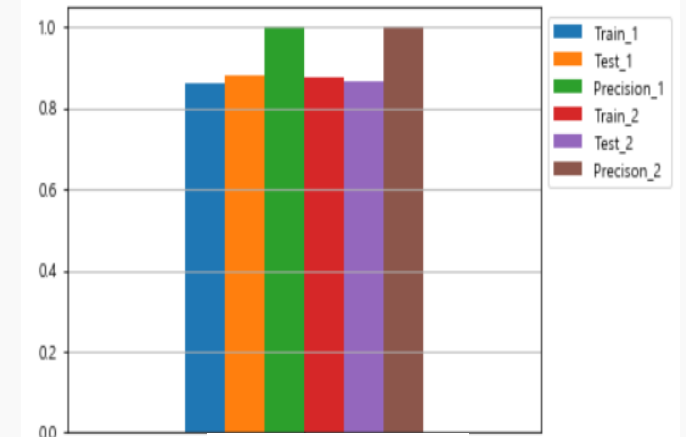
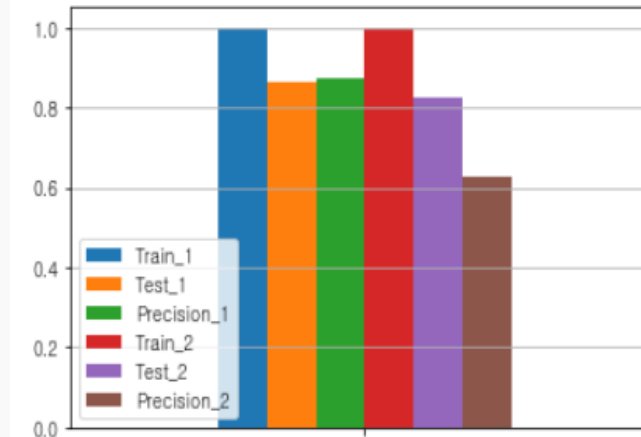


청구 이력

- 목표 변수 : 개인별 (지급-청구)/지급
- 보험 가입 검진 정보, 보험 청구 데이터 이용
- 일반 검진(989명)과 혈액 검진(3535명)으로 분리하여 개별 모델링 진행
- 기존 고객의 건강 특성 및 청구, 지급 패턴을 토대로 거절 고객에 대한 위험도 산정
- 기존 분류 모델 정확도(0.840) 대비 10% 향상

거절 고객	수익	할증	거절	합계
일반	7	469	7	483
혈액	76	2204	64	2319

- 위험 고객 분류 Train, Test 정확도 및 정밀도 비교



일반	TRAIN	TEST	PRECISION
1분류	1	0.865	0.871
2분류	1	0.825	0.625

혈액	TRAIN	TEST	PRECISION
1분류	0.907	0.916	0.919
2분류	0.867	0.864	0.778

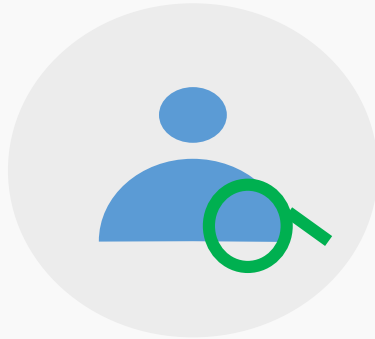
일반 검진은 Random Forest, 혈액 검진은 Gradient Boosting 모델을 이용해 고객 분류

3. 데이터 분석 “보험 상품 추천 및 할증 모델 제시”

● 보험 상품 추천 프로세스

- 목표 변수 : 해당 질병으로 인한 보험상품별 평균 지급액
- 선정 이유 : 같은 질병에 걸렸더라도 병원비를 평균보다 높게 지급 받은 사람과 낮게 지급 받은 사람의 특성을 이용하여 상품 추천 또는 할증 가능

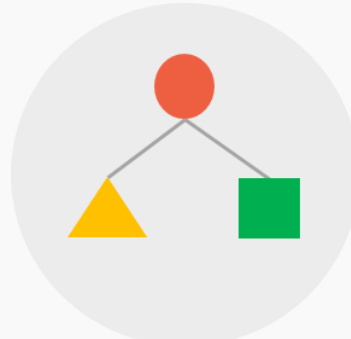
고객 정보 입력



- 개인정보 입력 : 10개
예) 성별(남:1 / 여:2), 키, 몸무게
- 질병 코드 입력 : 1개

M79	기타 연조직장애
S33	요추 및 골반의 관절 및 인대의 탈구
J00	급성 비인두염[감기]
R10	복부 및 골반 통증

모델링



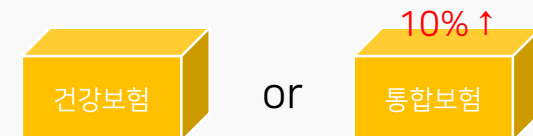
- 총 21가지 보험상품
- 각 보험상품 별 10가지 모델링
- 최대 210가지 모델링 가능

Naive Bayes	ANN
Logistic Regression	KNN
Linear SVC	Decision Tree
Support Vector Machine	Random Forest
Perceptron	Gradient Boosting

최종 상품 선정



- 분류(1 : 수익, 0 : 손해) 결과를 바탕으로
- 추천 : 수익성이 제일 높은 상품
 - 할증 : 보험료 10% 인상

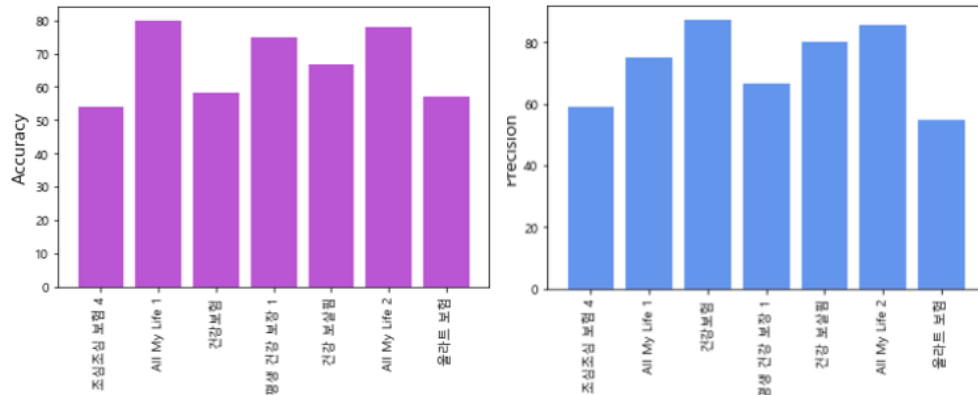


3. 데이터 분석 “보험 상품 추천 및 할증 모델 분석 결과”

● 보험 상품 추천 모델 분석 결과

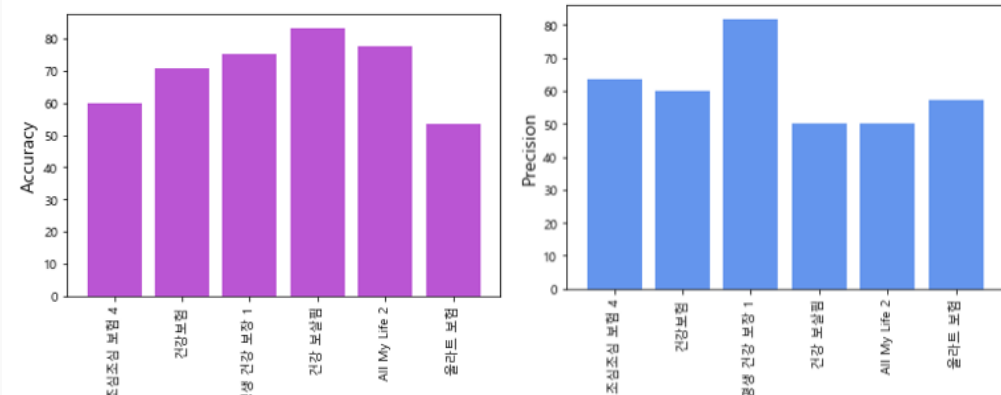
- INPUT : 고객 정보 → OUTPUT : 보험 상품 추천 (분류 결과 1 : 수익, 0 : 손해)
- 추천모델 : 수익(1)으로 분류되었을 때, 그에 맞는 보험 상품 추천
- 할증모델 : 손해(0)로 분류되었을 때, 상품에 대해서 10% 할증

추천 모델



보험	모델	정확도	정밀도
조심조심 보험 4	Naive Bayes	54%	59%
All My Life 1	KNN	80%	75%
건강보험	Support Vector Machines	58%	87%
평생 건강 보장 1	Perceptron	75%	66%
건강 보살핌	Random Forest	66%	80%
All My Life 2	Random Forest	77%	85%
올라트 보험	Random Forest	57%	55%

할증 모델

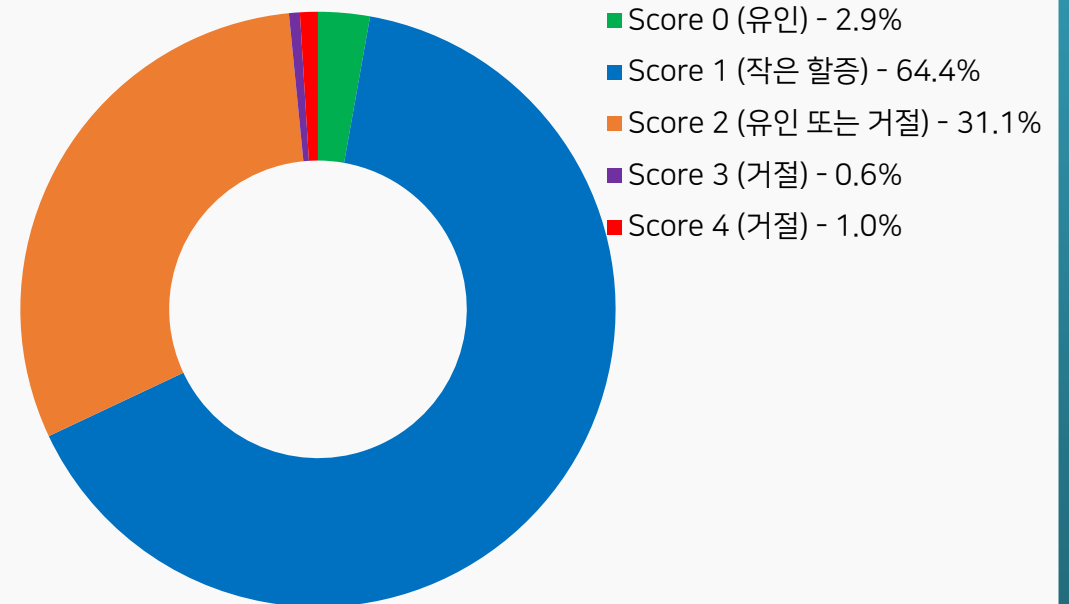
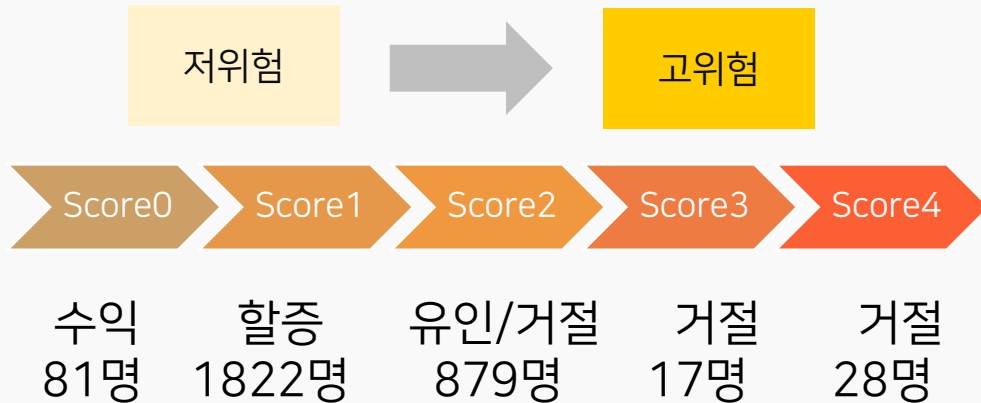


보험	모델	정확도	정밀도
조심조심 보험 4	KNN	60%	63%
건강보험	ANN	70%	60%
평생 건강 보장 1	Perceptron	75%	81%
건강 보살핌	KNN	83%	50%
All My Life 2	Random Forest	77%	50%
올라트 보험	Random Forest	53%	57%

4. 개선 방안 “청구 기준과 납입 기준 고객 분류 결과를 종합적으로 고려하여 거절 고객을 분류”

● 거절 고객군 최종 분류 결과

- Score = 청구기준 위험도 + 납입기준 위험도 (0: 수익, 1: 할증, 2: 거절)
- 예시 Score 1 : 청구기준 위험도 0 + 납입기준 위험도 1
또는 청구기준 위험도 1 + 납입기준 위험도 0



4. 개선 방안 “건강 검진을 통한 가입 심사 항목을 강화함으로써 분류 모델의 정확도 향상”

● 가입 심사 항목 강화

- 국민 건강 검진 정보 추가 이용
- 약 35000명의 데이터 랜덤 추출

보험 가입 검진 정보

기존 존재 항목
Bp_max
Bp_min
Age
Gender 등



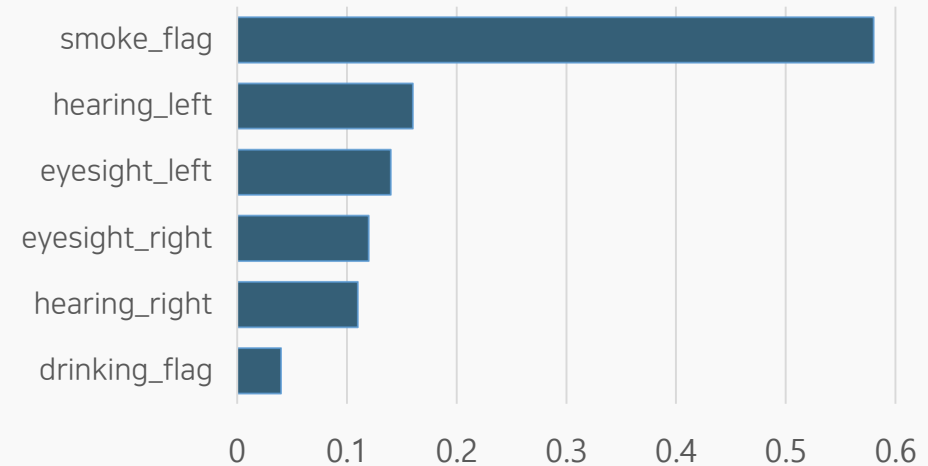
국민 건강 검진 정보

추가 고려 항목
Smoke
Hearing L/R
Drinking
Eyesight L/R



국민 건강 검진 정보에서 고객 분류 모델링의 정확도를
높여줄 수 있는 변수들을 추가적으로 선정

- 모델링 결과 및 변수 중요도
- 과 적합 방지 및 TEST 정확도를 고려한 GB 모델 선정



일반	TRAIN	TEST
1분류	0.958	0.955
2분류	0.875	0.884

4. 개선 방안 “개인의 검진 정보와 사전 고지한 과거 병력에 따라 보험 상품이 추천되거나 할증됨”

● 보험 상품 및 할증 예시



69세 남성, 김범수
키 : 160cm
몸무게 : 59kg
가슴둘레 : 84
허리둘레 : 72
Bmi : 23.05
Whtr : 0.45
맥박 : 66
혈압 판정 결과 : 1(비정상)



상품 추천

“ 질병 코드를 입력하세요 ” : J00(급성 비인두염)

보험	모델	정확도	정밀도	분류 결과
건강 보살핌	KNN	82.61	90.00	0
All My Life1	KNN	77.78	85.71	0
조심조심 보험 4	ANN	81.25	83.33	1
건강보험	Random Forest	90.00	88.89	1
평생 건강 보장 1	KNN	88.64	92.68	1

상품 할증

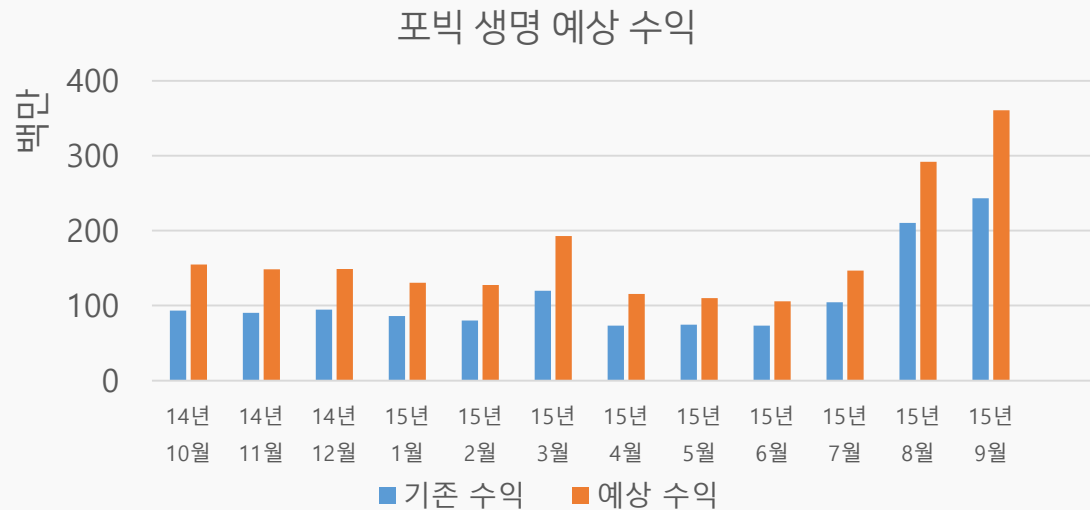
“ 질병 코드를 입력하세요 ” : S33(요추 및 골반 관절 및 인대 탈구)

보험	모델	정확도	정밀도	분류 결과
가족 만족 보험 2	Random Forest	50.00	66.67	0
All My Life2	Random Forest	70.00	75.00	0
조심조심 보험 4	Decision Tree	71.43	94.12	0
올라트 보험	Random Forest	60.00	64.29	0
평생 건강 보장 1	Decision Tree	70.45	90.32	0

4. 개선 방안 “현 수준과 비교하여 기대 수익 산정”

● 기대 성과

- 총 수익 = 보험료 - (병원비 지불 금액 + 혈액 검사비용)



기간(월)	14년 10월	14년 11월	14년 12월	15년 1월	15년 2월	15년 3월	15년 4월	15년 5월	15년 6월	15년 7월	15년 8월	15년 9월
기존 수익 (만원)	9,336	9,055	9,470	8,625	8,007	11,971	7,336	7,453	7,317	10,461	21,054	24,322
예상 수익 (만원)	15,739	14,829	14,897	13,069	12,735	19,309	11,571	11,006	10,576	14,667	29,219	36,085

- 현재 연 수익(14년 10월 1일 ~15년 9월 30일)
: 7,758,975,450원 (약 77억원)

- 예상 연 수익
: 8,364,569,850원 (약 83억원)

- 고객 분류 모델 예상 추가 이익
: 310,334,400원 (약 3억 1천만원)

- 질병 예측 모델 예상 추가 이익
: 295,260,000만원 (약 2억 9천만원)

총 605,594,400원 (약 6억원), 7.8% 의 매출 증대 예상

4. 개선 방안 “분석 목적과 데이터 셋에 맞추어 계획을 수립”

● Web Design 시연

최지영	2002.06.20	010-5166-7111	
남	30	180	20
145	40	30	30

yms21888@naver.com

질병 여부(선택)
M51 (기타 추간판장애) ▼

상품 추천 결과 : 평생 건강 보장 1 [10% 할증] → 월 330,000원

상품 검색

바로가기



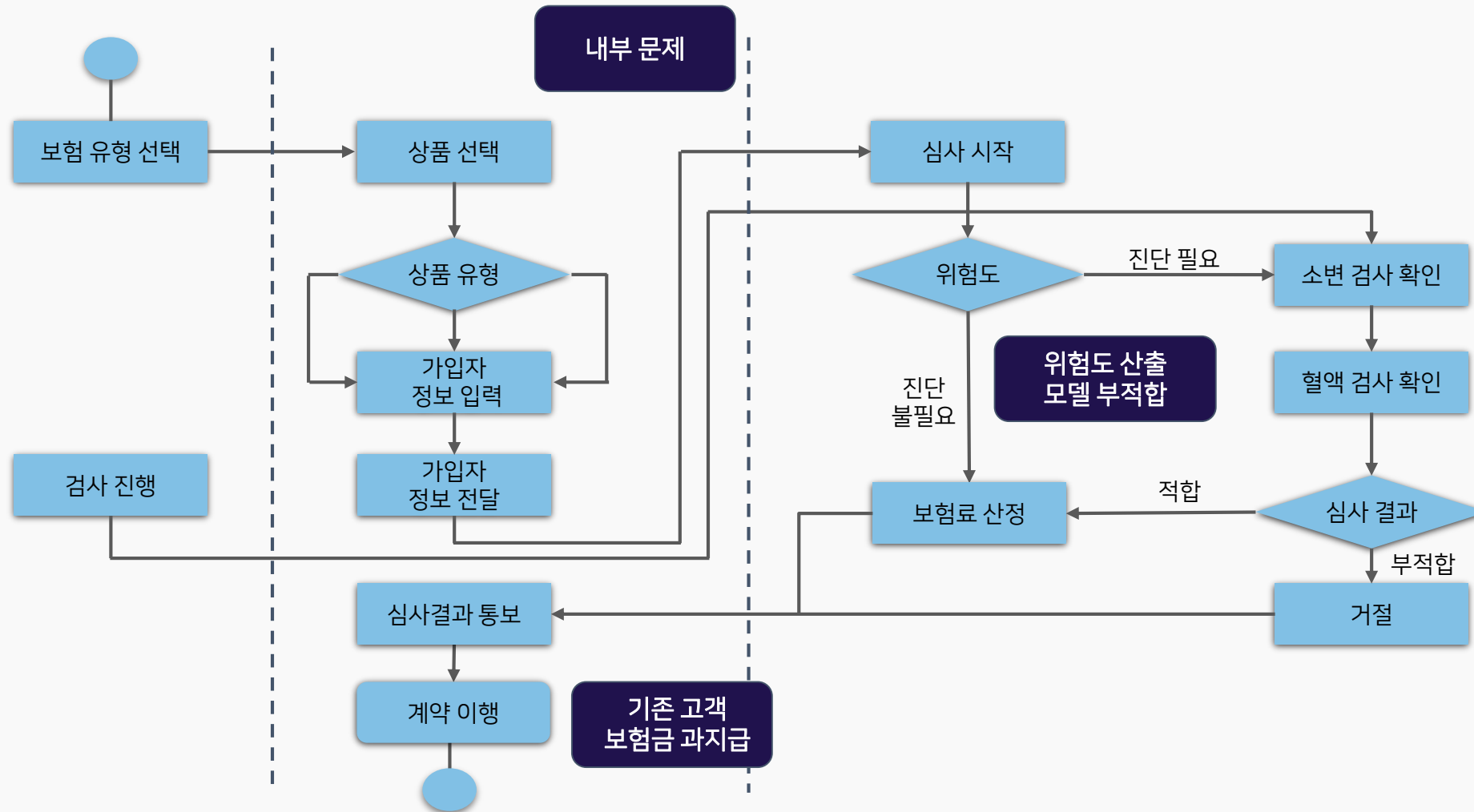
5. 개별 소감

팀원	소감
강지영	처음부터 끝까지 많은 노력을 쏟았던 프로젝트였기 때문에 큰 애착이 갑니다. 정말 많은 시간 동안 함께 회의를 하며 더 좋은 의견을 내고, 많은 시간 동안 분석을 진행하며 더 정확한 모델을 만들기 위해 힘썼습니다. 조원 모두의 열정과 노력을 담았기에 프로젝트 또한 이렇게 성공적으로 끝마칠 수 있지 않았나 생각합니다. 개인적으로는, 이번 프로젝트를 통해 데이터 분석 분야가 어렵지만 굉장히 흥미로운 분야임을 느끼게 되었습니다. 빅데이터 교육과 프로젝트를 통해 한 단계 성장할 수 있어 정말 뜻깊은 경험이었습니다.
김범수	저는 이번 빅 데이터 분석 프로젝트가 첫 데이터 분석 경험이었습니다. 처음이라 그런지 데이터를 처음 받았을 때는 어떤 방향으로 분석을 시작해야 될지 혼란스러웠지만, 인터넷 검색과 실제 보험업에 종사하시는 분들에게 물어보는 등 팀원들의 도움을 받아 도메인 지식을 습득하고 보니 여러 가지 아이디어가 떠올랐습니다. 팀원들 모두가 노력하면서 분석한 결과를 모아 피드백을 진행하다 보니 좋은 결과물이 나왔습니다. 분석한 결과를 가지고 웹 코딩을 진행하면서 관심이 없었던 웹 프로그래밍 지식을 쌓을 수 있게 되었고, 만족할 결과물이 나와 기분이 좋습니다. 우여곡절이 많았지만, 무사히 프로젝트를 마칠 수 있게 도와주신 교수님들과 팀원들에게 감사합니다.
김효진	프로젝트를 진행하면서 느낀 점이 많습니다. 처음 실제 고객과 유사한 보험 데이터 셋을 받았을 때는 많은 데이터 양에 당황했었습니다. 계획을 수립하고 분석을 시작하면서는 생각한 대로 결과가 나오지 않아 포기하고 싶은 마음이 들기도 했습니다. 코로나19로 일정이 연기된 뒤 다시 분석을 진행해야 했을 때는 막막함이 밀려왔습니다. 혼자였다면 이러한 고비들을 잘 넘기지 못했을 것입니다. 다행히 조원들과 함께하는 프로젝트였기에 서로 의견을 나누며 더 좋은 아이디어와 분석 방법을 생각해낼 수 있었고 프로젝트를 잘 마칠 수 있었습니다. 이번 팀 활동은 '혼자 간다면 빨리 갈 수 있지만 함께 간다면 멀리 갈 수 있다'라는 말을 제대로 이해할 수 있는 계기였습니다.

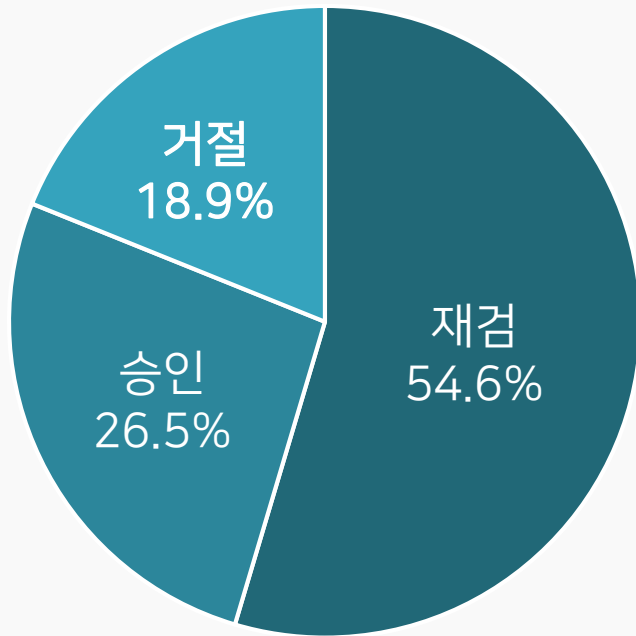
5. 개별 소감

팀원	소감
배향운	분석을 실제로 한 경험은 처음이기에 자료 조사, 목표 설정 등 처음부터 어려움이 많이 존재했지만, 2개월간 프로젝트로 배운 내용을 꼼꼼하게 정리하며 분석이라는 역량을 많이 배운 것 같습니다. 또한, 원본 데이터부터 시작하며 우리만의 결과를 도출해가고 회의를 통해 수정하고, 나아지는 결과물을 보며 성취감도 많이 느낄 수 있었습니다. 그만큼 애정도 많이 생겼고, 이번 프로젝트 외에도 이 분야로 더 배워보고 싶다고 느꼈습니다. 무엇보다 팀원들과 함께하며 더 많은 배움을 얻고, 벽찬 부분은 나눠서 도와주고 실패했을 때 힘내서 다시 시도할 수 있는 응원을 많이 받은 것 같아 기억에 오래 남을 것 같습니다.
양혜지	제대로 된 데이터 분석은 처음이라 감회가 깊습니다. 계획 수립부터 개선 방안까지 배웠던 모든 내용들을 프로젝트에 적용하면서 데이터 분석에 큰 매력을 느꼈고, 어떤 부분이 부족한지 깨닫게 되었습니다. 상황이 여의치 않아 다 같이 모여 프로젝트를 진행하지 못했지만 조원 모두 최선을 다해서 진행한 것에는 거짓이 없습니다. 서로 피드백을 통해서 분류 모델의 정확도를 높이고, 우리가 할 수 있는 개선 방안을 착안해내어 큰 성취감을 느낄 수 있었습니다. 다른 조의 분석 내용 또한 궁금했기에 실제로 모여서 발표 시간을 가질 수 없는 것이 아쉽지만 직접 실무에서 일하시는 교수님들이 피드백 해주시는 좋은 기회를 갖게 되어 감사함을 느낍니다.
정지성	데이터 분석 주제를 선정하는 것의 중요성을 알게 되었습니다. 실제 데이터는 '시그널'과 '노이즈'를 구분하기 어렵다. 동일한 데이터라도 여러 가지 측면으로 해석될 수 있습니다. 때문에 '어떤 분석을 선택할 것인가'가 중요하다고 생각합니다. 분석 주제를 먼저 선정한다면, 선택 기준을 세울 수 있고 이를 통해 불필요한 분석을 최소화할 수 있습니다. 프로젝트 초반에는 이러한 생각이 없어서 시행착오가 있었지만, 분석 주제를 세운 후로 좀 더 수월하게 진행할 수 있었습니다.
최지영	데이터 분석에 앞서서 보험이라는 분야는 저에게 생소한 분야였습니다. 데이터를 분석하면서 잘 몰라서 중요한 포인트를 놓치게 될까 열심히 노력했습니다. 열심히 하고 고생한 만큼, 보고 싶은 것만 보고 싶은 유혹이 생겼습니다. 숫자를 보지 않거나, 봤더라도 제대로 분석하지 않을 때도 있었습니다. 하지만 프로젝트를 마치며 모든 힌트는 데이터에 있고 답을 찾기 위해 항상 데이터를 근거로 의사결정을 해야 한다는 것을 느꼈습니다.

● 보험 가입 승인 프로세스



● 수익성 하락 원인 : 기존 가입 심사 모델의 오류



건강 검진

- 과거의 병력 확인
- 방문 진단 검사 도입 :
진단 나이를 41세→20
세로 변경하여 기준 강화

혈액 검사

- 간 수치로 간 기능 파악
- 혈당, 당화혈색소, 및
콜레스테롤 수치 등 평가

소변 검사

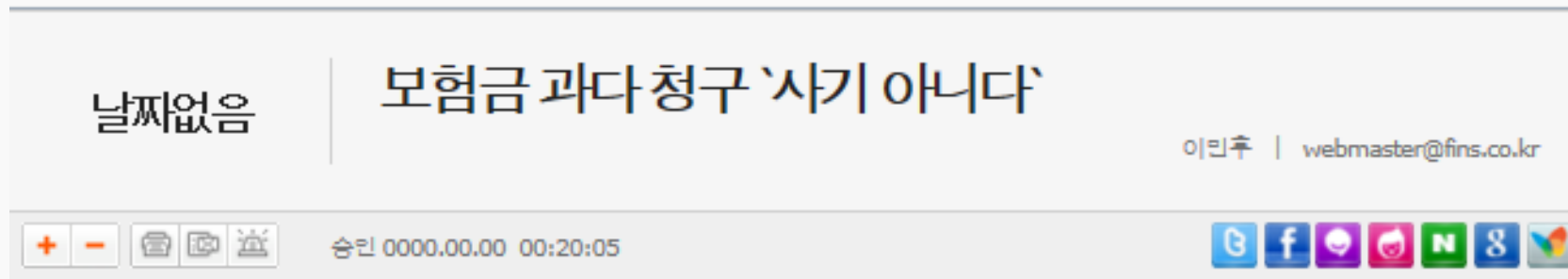
- 뇨당, 뇨단백, 뇨장혈
수치를 측정하여 평가
- 음주, 흡연 여부 평가

Check

건강검진, 혈액 검사, 소변 검사에 대한 판정 결과로 가입 거절

- 수익성 하락 원인 : 기존 가입 심사 모델의 오류

- 고 위험군의 환자이나 오분류로 인해 저 위험군으로 보험 가입, 청구·납입 대비 과다 지급되는 경우 발생



● 데이터 수집 계획

잠재 원인	데이터 수집계획						
	데이터명	속성	발생주기	수집방법	담당자	수집 가능성	주요특성
잘못된 심사 기준	개인 정보	범주형	일일	고객 정보	김범수	0	자가 진단
	건강 정보	연속형	일일	고객 정보/ 공공 데이터	양혜지	0	진단 결과
보험 홍보 부족	분기별 보험 가입률	연속형	분기	사내 데이터	정지성	0	자동 측정
	자사 상품 인지 도	범주형	반기	설문 조사	배향운	△	수동 측정
가격 경쟁력 약 화	자사 상품 가격	연속형	변동	사내 데이터	강지영	0	자동 측정
	타사 상품 가격	연속형	변동	협조 요청	김효진	△	수동 측정
보험료 과다 지급	납입액	연속형	수시	사내 데이터	김범수	0	자동 측정
	지급액	연속형	수시	사내 데이터	최지영	0	자동 측정

● ERD 다이어그램

