

빅데이터 활용 금융 아이디어 제안서

제안명 비식별 헬스 빅데이터를 연계한 지능형 맞춤 보험상품 비교추천 서비스

1. 제안 개요

제안 배경	<ul style="list-style-type: none"> • 보험 공급자 및 가입자 모두에게 비대칭적 정보 교류가 발생하여 상품 거래에 대한 효과와 그로 인한 효용이 모두 떨어지는 상황임 • 보험시장은 다음과 같은 4가지 문제가 있음 <ul style="list-style-type: none"> - 보험시장의 역선택 문제 - 일률적인 보험 판매 방식의 문제 - 필요할 때 혜택 받지 못하는 문제 - 가입한 보험의 만족도가 낮은 문제
아이디어 내용	<ul style="list-style-type: none"> • 아이디어 소개 <ul style="list-style-type: none"> ✓ 보험 상품 공급자, 소비자 모두 만족할 수 있는 보험상품 개인 맞춤형 추천 및 판매 서비스 ✓ 비식별화된 공공 헬스 데이터와 민간 금융데이터를 연계하여 질병예측모형 및 상품 추천 알고리즘을 개발함으로 보험시장의 비대칭적 정보교류로 발생하는 사회경제적 비용을 줄일 수 있음 • 아이디어 작동 프로세스 <ol style="list-style-type: none"> 1) 개인 건강정보 입력 2) 비식별 처리 3) 예측모형을 통한 질병 유병 확률 계산 4) 보험상품의 질병 가중치 정보를 유병 확률과 계산 5) 추천 로직에 따라 연관 보험상품 리스트 6) 맞춤 보험상품 추천 • 활용 데이터 <ol style="list-style-type: none"> 1) 비식별 공공 헬스 데이터 <ul style="list-style-type: none"> - 국민건강보험공단에서 연구용으로 개방한 건강검진 데이터 2) 민간 금융 데이터 <ul style="list-style-type: none"> - 보험사에서 제공하는 보험 상품 정보 데이터
기대효과	<ul style="list-style-type: none"> • 실질적 수요 및 잠재적 필요를 반영한 최적화된 보험 상품 추천 • 기존 보험 비교추천 플랫폼의 한계 극복 • 대면 채널(설계사)의 전문성 강화에 활용 가능 • 이종 산업 간 데이터 연계를 통한 융합 서비스 발굴 및 사회문제 해결 • 비식별 조치 기반 개인정보 보안 이슈 해결

2. 배경 및 필요성

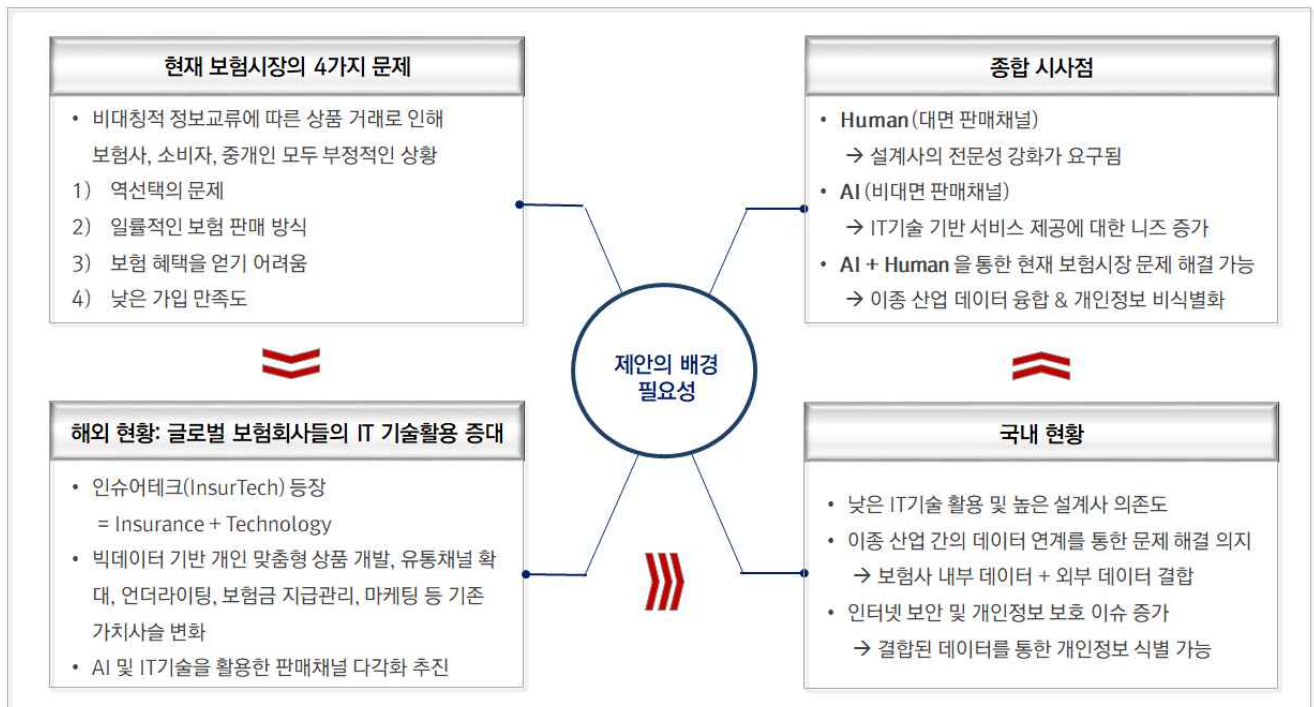


그림 1. 제안의 배경 및 필요성 개요

2.1. 현재 국내 보험시장의 4가지 문제

□ 문제 개요

- 보험 공급자 및 가입자 모두에게 비대칭적 정보 교류가 발생하여 상품 거래에 대한 효과와 그로 인한 효용이 모두 떨어지는 상황임
- 보험시장은 현재 다음과 같은 4가지 문제가 있으며 각 문제는 상호 연관성이 있음
 - 보험시장의 역선택 문제
 - 일률적인 보험 판매 방식의 문제
 - 필요할 때 혜택 받지 못하는 문제
 - 가입한 보험의 만족도가 낮은 문제

1) 역선택 문제

- 공급자와 소비자 사이에 정보 비대칭이 발생하고 불합리한 의사결정이 이루어져 보험사에게 손실이 발생하고 이는 소비자에게 전가되어 보험사, 소비자 모두에게 부정적임
- 대부분 사고 위험이나 건강상태가 좋지 못한 사람들이 가입을 많이 하며 이로 인해 보험사의 손실은 커지고 보험료는 계속해서 상승하는 악순환 구조임 (김대환&이봉주, 2014)
- 보험상품은 난해하고 전문적인 내용의 약관으로 구성되어 소비자는 주로 보험설계사의

설명에 의존하여 보험계약을 체결하게 되고 약관의 내용을 온전히 이해하지 못하는 경우가 많아 중도 해지가 많이 발생하여 보험사, 소비자 모두에게 비용이 발생함

2) 일률적인 보험 판매 방식

- 인구통계학적 관점에서 보험 상품이 제공되는 구조임. 예를 들면, 20대에게는 암 보험 보다는 실비가 중요하다고 판단하여 실비 보험을 추천
- 이러한 방식은 미래지향적으로 봤을 때 개인이 원하는 또는 개인 스스로도 모르는 자신의 실질적 필요를 반영하지 못함
- 근시안적 접근에 따른 보험상품 추천 및 구매가 이루어져 필요시에 혜택을 받지 못하게 됨

3) 보험 혜택을 얻기 어려움

- 스스로 어떤 보험이 필요한지 잘 모르며 막상 보험 혜택을 받고자 했을 때 해당이 되지 않는 경우가 많음
- 약관내용이 같아도 해석이 달라 보험금 지급을 안하는 경우도 많음 (“암보험, 정작 필요할 때 보험금 제대로 안줘 애를 먹고 있다” , 쿠키뉴스, 2017.02.03.)
- 건강보험 같은 경우 실제 보장 비율이 60%대로 낮기 때문에 국민 10명 가운데 6명이 실손보험을 가입 (“실손보험 솔림 왜? ‘건강보험만으로는 불안하니까’ ” , 한겨레, 2016.06.17.)

4) 낮은 가입 만족도

- 보험상품 가입자들의 만족도는 낮은 수준이었고 주된 원인은 보험상품 정보의 안내 부족 및 부적합한 상품 추천이었음
- 2014년 세계 30개국 총 1만 5000명을 대상으로 보험만족도를 조사한 결과 한국의 가입자 보험만족도는 세계 꼴찌를 기록하였음. 설계사 관련 불만이 가장 많았고 보험 상품에 대한 설명이 부족하고 약관이 어려운 것이 불만이었고 보험사를 옮기는 이유였음. (“가입 때만 王.. 보험만족도, 세계 꼴찌” , 조선일보, 2014.03.30.)
- 금융감독원이 16개 생명보험사의 보험설계사 240명을 대상으로 미스터리 쇼핑을 실시한 결과, 고객 만족도가 미흡한 수준으로 나타남. 진단결과를 바탕으로 한 적합한 보험 권유 부족, 진단결과와 다른 상품 선택 시 부적합한 사실 안내 등이 원인이었음. (“보험설계사의 변액보험판매 만족도 낮아” , 한국소비자원 제공, 2012)

2.2. 해외 현황: 글로벌 보험회사들의 IT 기술 활용 증대

□ 인슈어테크 등장

- 인슈어테크(InsurTech)는 Insurance(보험)과 Technology(기술)의 합성어로 데이터 분석, 인공지능(AI), 사물인터넷(IoT) 등의 기술을 활용하여 기존 보험 산업을 혁신하는 서비스를 의미함 (윤일영, 2017)
- 인슈어테크의 등장과 함께 상품개발과 유통채널, 언더라이팅, 보험금 지급관리, 마케팅 및 고객관리의 영역에서 보험 산업과 기술의 융합으로 기존 가치사슬(Value Chain)이 무너지고 소비자를 중심으로 한 서비스 플랫폼 체계로 재편되고 있음 (삼정KPMG경제연구원, 2016)
- 그중 미국, 캐나다, 스위스, 중국, 인도 등의 해외 보험사들은 소비자의 개인용 모바일 기기를 활용한 온라인 상품 구매에 대응하기 위해 인슈어테크에 투자하고 있음 (윤일영, 2017)

□ AI 및 IT기술을 활용한 판매채널 다각화 추진

- 전 세계적으로 설계사, 중개인을 중심으로 한 전속판매채널(대면)뿐만 아니라 웹사이트와 같은 비전속판매채널(비대면)을 통한 보험 판매채널이 생겨나며 소비자 중심으로 보험산업의 유통구조가 변화하고 있음 (안철경, 2011)



그림 2. 보험시장 유통채널 트렌드 변화 (출처: 보험연구원)

- AI, 빅데이터와 같은 디지털 기술의 발달과 함께 보험 판매채널이 다원화되고 있으며 빅데이터 기반의 개인 맞춤형 상품 구매, 시간과 공간에 자유로운 O2O 기반의 온디맨드 서비스 등에 대한 소비자의 니즈가 증가하고 있음 (이선주, 2017)
- 미국의 보험사 Oscar는 개인에게 최적화된 건강관리서비스와 헬스케어서비스를 제공하며, 스위스의 보험사 Knip은 보험상품을 비교해 개인에게 제안하는 사업을 진행하고 있음 (윤일영, 2017)

2.3. 국내 현황

□ 낮은 IT기술 활용 및 높은 설계사 의존도

- 우리나라 보험산업은 금융산업 중에서도 설계사 의존도가 매우 높고 다른 금융 업종에 비해서 IT기술 활용이 부족한 혁신이 더딘 분야임 (윤경일, 2017)
- 국내 보험소비자를 대상으로 유통채널 선호도를 조사한 결과, 설계사가 선호도가 가장 높은 채널로, 이어서 중개인, 전화 순으로 선호하는 판매 채널로 조사되었음 (보험연구원, 2016)
- 게다가 인터넷 보안 문제로 인한 편의성이 낮아 보험 소비자가 보험회사 웹사이트에 대한 관심이 낮고 실제 이용 빈도가 낮은 것으로 드러남 (보험연구원, 2016)

□ 이종 산업 간의 데이터 연계를 통한 문제 해결 의지

- 보험회사 현직 전문가 인터뷰를 실시한 결과, 회사 내부 데이터만으로는 고객을 분류하고 예측하는 것이 어려운 상황이며 외부 데이터와의 연계 및 결합을 통해 이를 해결해야 한다고 하였음
- 보험회사 내부 데이터와 외부 산업 데이터를 결합을 통한 신규 가치 제공이 필요한 상황이며 이 과정에서 개인정보 보호 이슈가 대두되었음
- SKT는 최근 한화생명, 이지서티 등과 추진한 빅데이터 개인정보 비식별화 작업 추진 사례와 방법에 대한 결과를 제공하였고, 이종간 데이터 결합을 통한 신규 데이터 도출을 통한 가치를 마련하는 기틀을 마련하는 중이라고 하였음 (“개인정보 보호, 활용 동시 가능한 ‘비식별화’”. 채널i 산업뉴스. 2017.04.19.)

2.4. 요약 및 시사점

□ 제안 배경 요약

- 현재 국내 보험시장은 역선택의 문제, 일률적 보험 추천 및 판매 방식의 문제, 보험 혜택 저하 문제, 낮은 가입 만족도 등의 한계가 존재함
- 해외에서는 IT기술을 활용하여 판매채널을 다각화하고 상품 및 유통구조를 소비자 중심으로 개발하여 이와 같은 문제들을 해결하고 있음
- 이와 관련한 국내 현황을 살펴보면 설계사 중심의 대면판매가 주요채널로 작용하고 IT기술의 활용은 부족한 상황임. 그러나 소비자는 설계사를 선호하면서도 만족도가 낮고 IT기술을 통한 편의성 높은 서비스를 요구하고 있음
- 따라서 인슈어테크(InsurTech)로 대두되는 글로벌 트렌드로 현재 국내 보험산업의 문제를 해결하되 국내 여건과 소비자의 의견을 반영한 서비스를 제공할 필요가 있음
- 이때 이종 산업 간의 데이터 연계 및 결합이 요구되며 이 과정에서 개인정보를 보호할 수 있는 방법으로서 개인정보 비식별화 조치가 필요한 상황임

□ 종합 시사점

- 대면 판매채널의 경우, 기존 전속 보험설계사에 대한 만족도가 떨어지는 만큼 기존 설계사의 고능률화가 요구되는 상황임
- 비대면 판매채널의 경우, IT기술을 기반으로 한 인터넷 채널의 성장과 함께 비교추천 서비스의 기능강화와 구매절차의 간소화에 대한 소비자의 요구가 증가하였음
- Havard Medical School에 따르면, AI 단독(92.5%) 혹은 Human 단독(96.5%)에 비교하여 AI와 Human의 조합이 가장 높은 정확도(99.5%)로 암 발병 유무를 예측할 수 있었음 (Prescott, 2016)
- 보험 산업에서도 비대면 채널(AI)과 대면 채널(Human)을 융합하여 높은 정확도의 소비자 맞춤형 보험상품 비교추천을 통해 상품을 제공하는 공급자 및 설계사와 소비자 모두가 만족하고 사회경제적 비용을 줄일 수 있음

3. 제안 아이디어

3.1 아이디어 소개

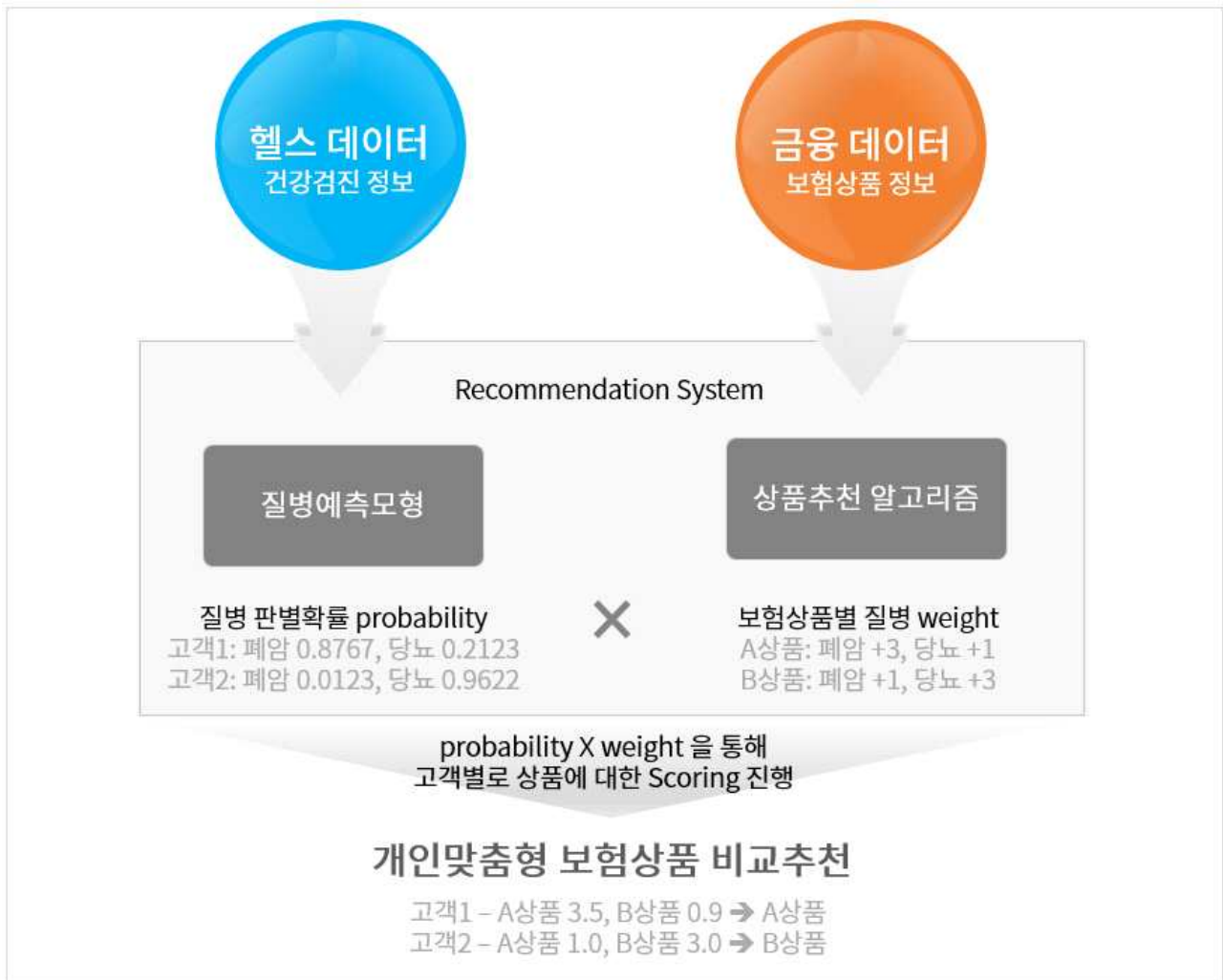


그림 3. 아이디어 전체 개념도

- 보험상품 공급자, 소비자 모두 만족하는 AI 기반 보험상품 개인 맞춤형 비교추천 서비스
- 비식별화된 공공 헬스 데이터와 민간 금융데이터를 연계하여 질병예측모형 및 상품 비교추천 알고리즘을 개발함으로 소비자 개인 맞춤 보험상품을 비교추천 할 수 있기 때문에 보험시장의 비대칭적 정보교류로 발생하는 사회경제적 비용을 줄일 수 있음

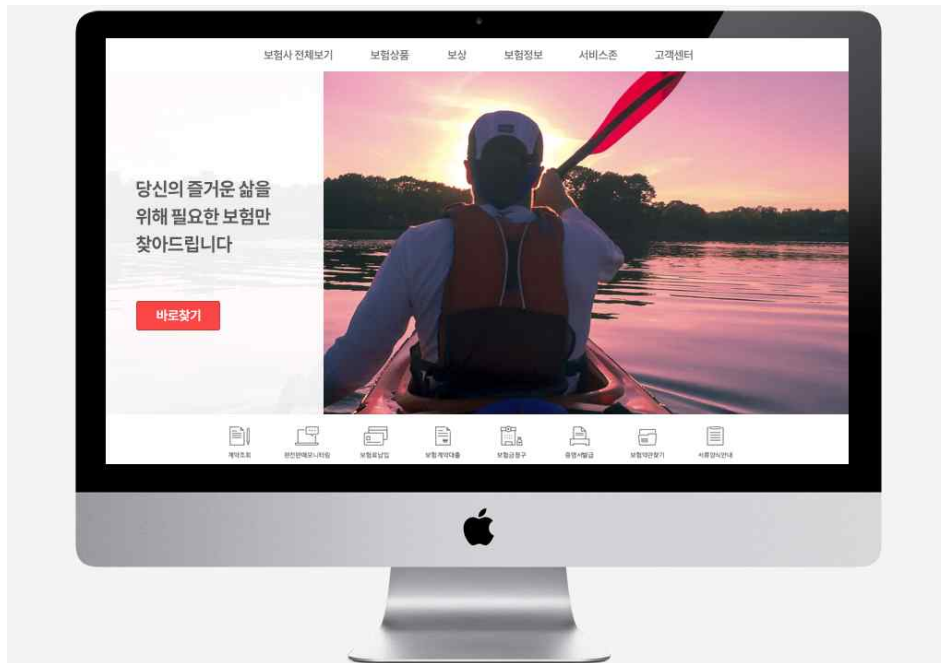


그림 4. 본 아이디어의 홈페이지 서비스 메인 화면

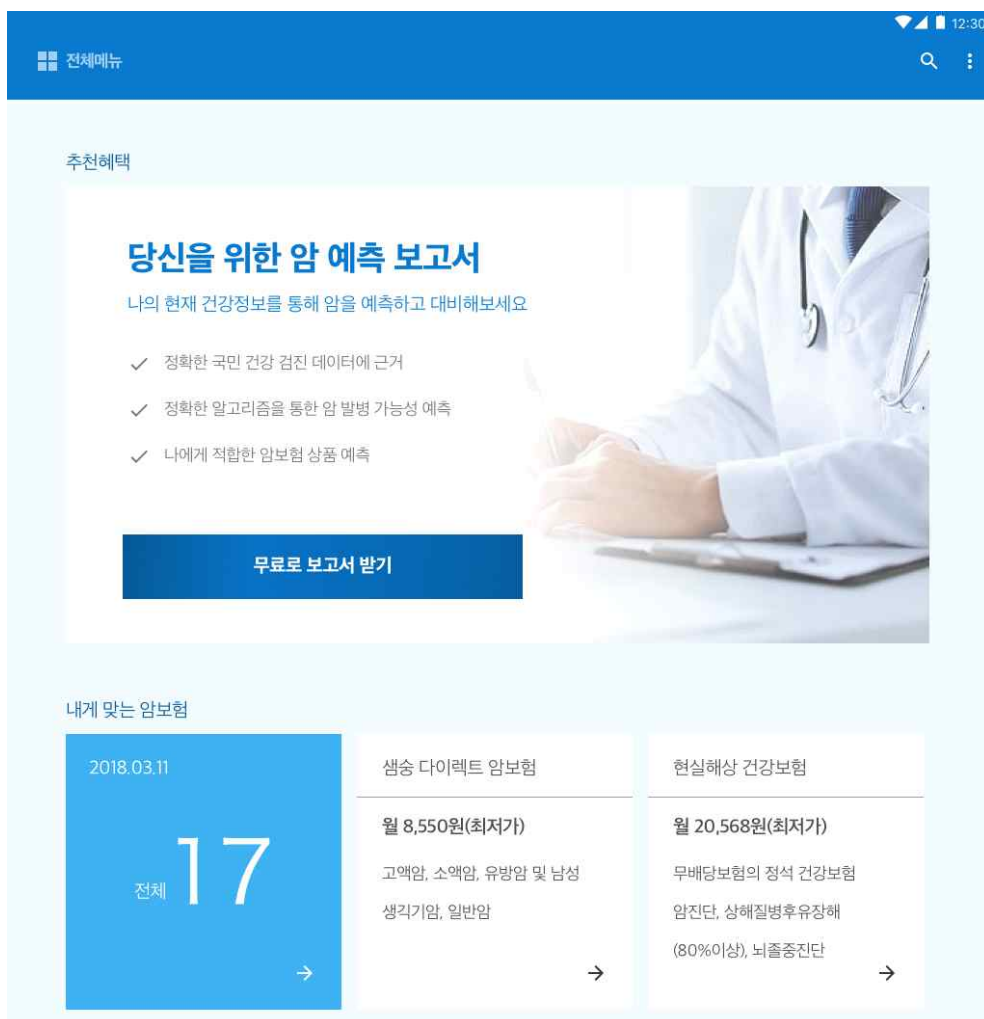


그림 5. 아이디어 프로토타입 화면 - 서브메인

전체메뉴

12:30

기입하신 정보는 저장되지 않으니 안심하세요.

정확한 예측을 위해 정보를 입력해주세요.

1. 기본정보

2. 건강정보

귀하의 이름은?

김승권

성별은?

✓ 남자

여자

생년월일은 언제이신가요?

19861126

결혼 하셨나요?

✓ 기혼

미혼

부양가족이 있으신가요?

유

✓ 무

다음 단계로 이동

김승권님 안녕하세요.

귀하는 남자이며, 31세의,

기혼이시며, 부양가족은 없습니다.

그림 6. 아이디어 프로토타입 화면 - 개인 건강정보 입력

전체메뉴

12:30

결과를 저장하기

✓ 회원가입이 필요한 서비스입니다.

김승권님의 암 예측 보고서입니다.

암 발병률

69%

나이에 비해 암 발병률이 높습니다.

흡연을 상당히 많이 하시며, 고혈압과 당뇨가 있고

체지방이 높습니다.

지금부터 대비하세요.

원하신다면, 적합한 암 보험 상품을 추천드립니다.

▼

그림 7. 아이디어 프로토타입 화면 - 맞춤 보험상품 추천

3.2 분석활용 아이디어

□ 아이디어 작동 프로세스

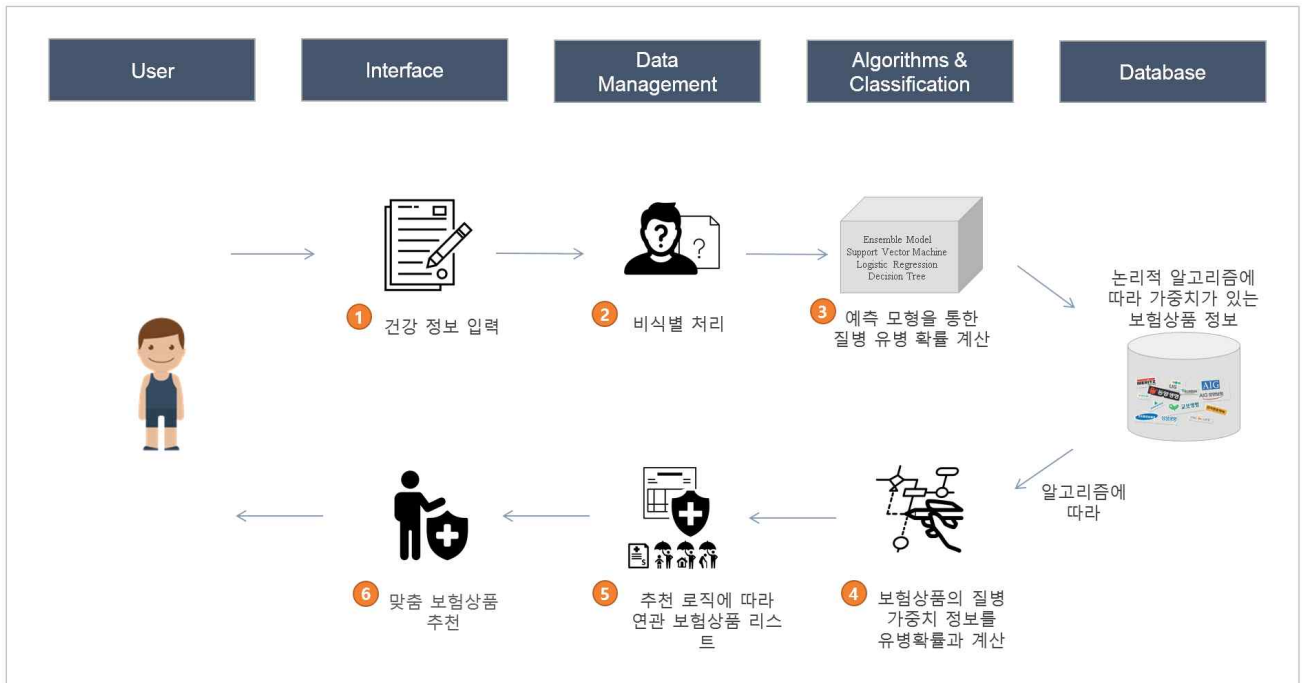


그림 8. 아이디어 작동 프로세스 개념도

1) 개인 건강정보 입력

- 질병 예측 분류 모형 개발의 독립변수(feature)에 해당하는 정보를 기입하는 설문 형식의 폼(form) 제공

2) 개인건강정보 비식별 처리

- 모형 개발 시 사용된 데이터의 비식별처리 방식으로 입력된 건강정보 비식별화

3) 예측분류 모형을 통한 질병 유병확률 계산

- 비식별처리된 데이터를 추천 시스템 내 예측 분류 모형에 돌려서 질병 예측 분류하고 유병확률을 계산

4) 질병 가중치 정보를 유병확률과 계산

- 보험상품마다 부여된 질병 가중치에 따라 유병확률이 높은 질병과 연관성 계산

5) 추천 로직에 따른 연관 보험상품 리스트

- 연관성이 높은 상품들을 거리로 계산하여 데이터를 추출

6) 맞춤 보험상품 추천

- 회원정보 및 기타 정보를 결합하여 연관성이 높은 보험상품 중 상위 상품들을 추천

3.3 활용대상 데이터

□ 건강검진 정보를 담은 헬스 데이터(공공 연구 데이터)와 보험상품 정보를 담은 금융데이터(민간 데이터)의 연계

□ 헬스 데이터 (비식별 공공 데이터) *별첨1 참조

- 질병예측모형 개발에 사용하는 데이터
- 국민건강보험공단에서 연구용으로 개방한 것으로 국민건강보험에 가입한 대한민국 국민으로 대상으로 하여 건강검진 받은 사람들을 무작위로 추출한 대표성을 가진 자료
- 약 51만 명의 동일한 대상자에 대해서 사회적, 경제적, 건강정보 변수가 포함된 자격정보, 진료내역 및 건강검진자료 3개의 테이블을 대체키를 통해 결합한 데이터
- 데이터 특징
 - 개인 식별이 불가능한 비식별 처리 데이터이면서 동시에 모집단의 대표성을 잘 반영한 데이터
- 개인의 민감성 정보를 식별자 대체 및 그룹화, 마스킹을 통해 비식별화
 - 1단계: 대체변수 사용
 - 2단계: 그룹화
 - 3단계: 마스킹

단계	비식별 조치 내용
1단계 대체변수 사용	<ul style="list-style-type: none"> 주민등록번호(13자리) → 개인일련번호(8자리) 명세서번호(23자리 5개 변수조합) → 청구일련번호(12자리 연도+일련번호)
2단계 그룹화	<ul style="list-style-type: none"> 연령 → 5세 단위 주거주소 → 시 단위(17개) 소득금액 → 소득단위(10분위) 장애정도 → 장애중증도 3단위
3단계 마스킹	<ul style="list-style-type: none"> 공단에서 규정한 식별가능성 있는 민감상병 114개 민감상병 대분류표시. 첫째 앞자리만 표시 ex) E1014 → E10

표 1. 비식별 조치 단계별 내용

○ 데이터 항목 상세

- 모형개발 및 검증에 사용한 test data 및 training data

항목	변수명	상세설명
성별	SEX	여자0 남자1
연령	AGE	연속형 숫자
소득분위	CTRB_PT_TYPE_CD	지역 및 직장가입자 10분위(1~10)
체질량지수	BMI	체중(kg) / (신장*신장)m
혈압	BP_HIGH	최대 혈압수치, 단위 mmHg
혈당	BLDS	식전에 공복인 상태에서 췌 혈당 수치
⋮	⋮	⋮
가족력(고혈압)	FMLY_HPRTS_PATIEN_YN	가족 중에 고혈압 환자가 있었는지 유무
흡연유무	SMK_STAT_TYPE_RSPS	흡연여부
운동횟수	MOV_FEQ	1주일에 약 30분정도 운동하는 일 수 (1~7)
특정질병유무	DISEASE	만드는 질병예측모형에 따라 질병유무 변수는 변화

표 2. 예측모형 개발에 사용하는 training/test 데이터 항목 상세

□ 보험상품 데이터 (민간 데이터)

- 다수의 민간 보험사에서 제공하는 보험상품 정보가 있고 상품과 연관된 질병, 그리고 상품 비용이나 보장 범위 등에 따라 서로 다른 가중치(weight)를 가지고 있음
- 가중치는 비즈니스 로직에 따라 그 값이 정해지며 학습을 위해 사전적으로 가중치 설정이 필요
- 전처리 과정이 용이한 주요 공시 정보를 바탕으로 의사결정을 할 수 있도록 구성
- 데이터 항목 상세
 - 자체 데이터베이스에 저장한 보험상품 데이터로 고객의 정보(질병 판별확률, 기타 비식별 개인 정보)와의 연계를 용이하게 하기 위한 형태로 데이터 항목 구성

항목	변수명	상세설명
원청사	COMPANY	동부, 삼성, 교보 등 60여개 원청사
종류	IC	실비, 태아, 암, 3대질병, 사망, 어린이, 치아, 운전자
종류2	IC2	일반사망, 상해, 질병, 고액암, 소액암 등
납입기간	YEAR	2~20년 만기
⋮	⋮	⋮
만기(재가입)	MAX_RE	만기0, 재가입1
납입액	M-1	300만원~2억
보장액	M-2	300만원~50억
갱신	RE-T	1년갱신0, 비갱신1

표 3. 예측모형 개발에 사용하는 Scraping 데이터 항목 상세

3.4 데이터 분석 및 서비스 개발 방안

□ 본 서비스 개발 방향은 크게 두 축으로 구성되며 하나는 질병 예측모형 개발이고, 다른 하나는 보험상품 추천 알고리즘 설계임

○ 질병 예측모형 개발 (별첨1 참조)

- 분석 소프트웨어: 오픈소스인 R과 SAS-U 사용
- 분석 방법: SEMMA 방법론 기반 모델링
- 모형 개발 절차

1) 데이터 탐색

2) 데이터 전처리

- 비식별 처리된 데이터를 활용하고 진료 기록을 통해 질병별로 사용자 식별번호를 추출하여 이를 기준으로 사회경제적 정보를 담은 자격 데이터와 건강검진 기록을 담은 데이터를 조인
- 결측값 및 환자별로 질병 분류 변수들을 추가

3) 데이터 셋 생성

- 질병별 사용자 데이터와 통제 그룹의 데이터를 통합하여 분석용 데이터 셋 생성
- Data Partitioning 작업을 통해 train 데이터와 test 데이터로 분할

4) 모델링

- SVM, Logistic Regression, Decision Tree와 같은 분류 알고리즘을 활용하여 모델링

5) 평가 및 배치

- 다수의 모델 교차검증 및 모델 간 평가 (Cross-validation, f-measure 활용)
- Test Data를 통한 모형 성능 평가 (ROC, AUC 활용)
- 질병별로 적절한 모형 선택 및 배치하고 질병 판별확률이 계산될 수 있도록 함

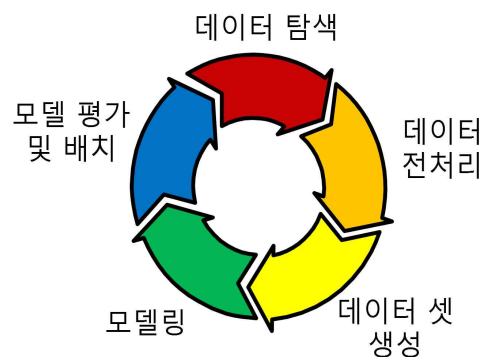


그림 9. 모형 개발에 활용한 SEMMA 방법론

○ 보험상품 추천 알고리즘 설계

1) 보험상품별로 rating(가중치) 작업

- 예를 들어, 폐암관련 보험상품의 경우 폐암질병과 관련된 정보에 대해서 상품추천 확률이 올라가도록 점수를 매김
- ‘무배당 흥국생명 온라인암보험(비갱신형)’은 다른 암보험과는 다르게 폐암 같은 특정암에 대한 보장내용이 있기 때문에 다른 상품과는 다르게 폐암에 대한 가중치 점수가 높게 설정되어 있어 폐암 판별확률이 높은 고객에게 우선적으로 추천이 될 것임

2) 예측 분류 모형의 확률 값과 상품별 가중치를 계산하여 유사도를 계산 모듈 개발

3) 콘텐츠 기반 추천방식과 협업필터링 방식을 혼합한 hybrid 방식의 추천 시스템으로 각 추천 방식의 한계 보완

3.5 데이터 확보 방안

□ 건강검진 데이터 (비식별 공공 데이터)

- 국민건강보험공단에서 운영하는 NHISS(National Health Insurance Sharing Service)를 통해 예측모형 개발에 필요한 자료를 신청
- 정책 개발 및 학술연구 등의 목적으로 공단에 자료제공을 요청하여 승인을 받으면 필요 데이터 범위에 따라 무료로 혹은 비용을 납부한 후에 자료를 받고 사용 가능

자료제공 처리절차 안내

자료제공은 온라인상의 "데이터신청"신청하기"를 시작으로 다음과 같은 절차를 거쳐 진행이 되며 "MY서비스"를 통해 진행 상태를 확인 하실 수 있습니다.



그림 10. NHSS에서 데이터 신청 프로세스 (출처:NHSS)

□ 보험상품 데이터 (민간 데이터)

- 각 보험사로부터 보험상품 데이터베이스 API를 요청하여 승인을 받은 후에 보험상품 정보를 이용할 수 있음
- 승인이 불가한 항목에 대해선, 보험 가입자 승인을 통해, 원청사(보험사) 웹사이트 Front-Web Scraping 방식으로 사용 가능 (기존 보험 플랫폼 ‘레몬클립’, ‘봄애플’ 등이 사용하고 있는 방식)
- 많은 보험상품 비교 사이트에서 이러한 방식으로 보험사로부터 정보를 받아 상품을 유통하여 판매하고 있음. 다음은 보험상품 비교 사이트 ‘보험몰’ 예시 그림



그림 11. 보험사에서 보험상품 정보를 제공받아 판매하는 사이트 예시 (출처: 보험몰)

3.5 아이디어 타당성

□ 예측 분류 모형의 기술적 타당성

○ 폐암 환자 예측모형 및 판별 확률을 통해 질병 유병률 확인 *별첨2 참조

- 개발한 프로토타입의 폐암 환자 예측모형 예측 정확도가 약 90%로 상당히 정확한 확률로 질병을 예측하여 활용 데이터의 모델 개발 사용의 유의미한 가능성 확인

○ 통풍 환자 예측모형 개발을 통한 신규 비즈니스 모델 개발 *별첨3 참조

- 통풍같이 예측하기 어려운 질병도 본 데이터를 활용하여 예측모형을 개발한 결과 약 95%의 정확도를 보여줘 이를 활용한 신규 서비스 발굴의 유의한 가능성 확인
- 실제 민간 병원 관계자 인터뷰 결과 통풍 환자 예측모형 개발을 통한 신규 비즈니스 모델 개발의 가능성 확인

4. 기대효과

본 데이터가 가진 유전정보, 진료기록, 생활습관 뿐만 아니라 기후 환경정보, SNS, 정밀검사기록(XRAY, CT 사진 등) 등 다양한 정형, 비정형 데이터의 연계를 통한 PRECISION MEDICINE 질병예측 및 맞춤 보험 상품 추천의 가능성을 보여줌



그림 12. 본 아이디어 시사점

1) 실질적 수요 및 잠재적 필요를 반영한 최적화된 보험 상품 추천

- 비식별화된 건강정보를 바탕으로 개인에게 맞춤 보험 상품을 비교추천하여 사회경제적 비용을 감소시킬 수 있음
- 정보 비대칭으로 인해 발생하는 역선택의 문제로 필요에 의한 상품 거래가 이뤄지지 않아 높은 사회경제적 비용이 발생하였음. 판매자, 구매자 모두 피해 입는 불합리한 구조를 해결할 것으로 기대됨

2) 기존 보험 비교추천 플랫폼의 한계 극복

- 마이리얼플랜, 봄애플, 레몬클립 등 기존 보험 추천 플랫폼들은 단순히 개인의 인구통계학적 데이터나, 설문 항목, 기존 보험 정보만을 스크랩-활용하여 실제로 사용자에게 꼭 필요한 보험에 접근하거나, 새로운 수요를 창출하지 못한다는 한계가 있었음
- 상기 아이디어는 비식별 데이터를 활용하여 기존 플랫폼이 가지고 있던 한계를 보완-발전시켜, 실제적인 니즈와 수요를 창출해낼 수 있음

3) 대면채널(설계사)의 전문성 강화에 활용 가능

- 본 아이디어는 AI기반의 비대면 채널로써 전문성이 떨어지는 기존 설계사를 보조하여 그들의 전문성을 높여주어 부적합한 보험상품 추천을 발생한 낮은 보험가입 만족도 문제를 해결할 것으로 기대됨
- 비식별 조치를 통해 개인 건강정보를 활용하고 이를 활용할 수 있기 때문에 기존 인력을 보조하는 채널로써 빅데이터 기반으로 새로운 비즈니스 모델을 발굴하였으며 이를 통한 일자리 창출 등의 효과를 기대할 수 있음

4) 이중 산업 간 데이터 연계를 통한 융합 서비스 발굴 및 사회문제 해결

- 특정 산업군의 데이터(민간 금융데이터)뿐만 아니라 유관 산업 및 이중 산업의 데이터를 연계(헬스 분야 공공데이터)하여 새로운 융합 서비스의 발굴이 가능한 점을 보임
- 또한, 본 아이디어를 발전시켜 다른 금융 산업으로 확장시켜 비슷한 문제를 융합적 관점에서 바라보고 해결할 수 있는 가능성 제공

5) 비식별 조치 기반 개인정보 보안 이슈 해결

- 비식별 처리를 기반으로 민감한 개인의 건강 정보 활용의 가능성을 높이고 이를 통해 데이터 활용영역의 폭을 넓혀 그동안 문제로 지적되었던 보험 시장의 문제를 해결하고 나아가 다른 영역에서도 이를 적용할 수 있을 것으로 기대됨
- 개인 정보의 비식별 조치에 대해 고객이 인지하고 신뢰할 수 있도록 기반을 마련하는 것이 필요하며 이를 위한 모범 사례로서 사람들의 인식 제고에 기여할 수 있음

5. 참고자료

Prescott, Bonnie. “Artificial Intelligence Approach Improves Accuracy in Breast Cancer Diagnosis.” Better Together, 22 June 2016, <http://hms.harvard.edu/news/better-together>.

김대환, 이봉주. 실손의료보험의 역선택 분석. 보험학회지. 96(0): 25-50. 2014.

윤일영. 보험과 기술의 융합, 인슈어테크(InsurTech). 융합연구정책센터. 2017.03

보험연구원 동향분석실. 2017년도 보험산업 전망과 과제. 보험연구원. 2016.12

이선주. 디지털 혁명에 따른 보험 판매채널의 변화. 보험연구원 리포트. 2017.04

안철경. 보험유통채널의 평가 및 신타렌드. 보험연구원. 2011.

“가입 때만 王.. 보험만족도, 세계 꼴찌”, 조선일보, 2014.03.30.

“개인정보 보호, 활용 동시 가능한 ‘비식별화’”. 채널i 산업뉴스. 2017.04.19.

“보험설계사의 변액보험판매 만족도 낮아”, 한국소비자원 제공, 2012

“실손보험 쏠림 왜? ‘건강보험만으로는 불안하니까’”, 한겨레, 2016.06.17.

“암보험, 정작 필요할 때 보험금 제대로 안줘 애를 먹고 있다”, 쿠키뉴스, 2017.02.03.

*별첨1. 건강검진 데이터 상세 설명

표본연구DB란?					
표본 연구 DB는 전국민 건강보험 빅데이터를 기반으로 수요도가 높은 데이터를 표본 추출하여 정보주체를 알 수 없도록 비식별 조치한 후 정책 및 학술연구용으로 제공하기 위해 규격화한 데이터셋을 말하며, 현재 표본 코호트 DB, 건강검진 코호트 DB, 노인 코호트 DB, 영유아검진 코호트DB, 직장여성 코호트 DB 등 5종의 데이터셋이 있습니다.					
구분	표본수(천명)	구축기간	자료건수(천건)	내용	
표본코호트DB	1,025	2002~15(14개년)	2,619,397	· 전 국민을 대표하는 약 100만 명의 표본연구DB(전국민의 2%)	
건강검진코호트DB	515	상동	2,087,629	· 만40~79세의 건강 검진 수검자 중심으로 의료이용, 검진결과 등을 분석할 수 있는 표본연구DB (모집단의 10%)	
노인코호트DB	558	상동	2,749,045	· 2002년 기준 만60세 이상 노령층을 중심으로 구축한 표본연구DB (모집단의 10%)	
영유아검진코호트DB	84	2008~15(8개년)	233,688	· 2008~2012년(5년) 출생자 중 영유아건강검진 1~2차를 한번 이상 수검한 영유아를 모집단으로 하여 각 연도별 5% 단순 무작위추출	
직장여성코호트DB	185	2007~15(9개년)	368,226	· '07.12월 말 기준 건강보험 자격 유지자 중 만15~64세의 여성 직장가입자 약 360만 명의 5% 무작위 추출	

그림 13. 질병 예측 분류 모형 개발에 활용한 데이터 정보



그림 14. NHSS에서 제공하는 비식별화된 공공 헬스 데이터

*별첨2. 폐암환자 예측 분류 모형 개발 예시 (SAS-U 활용)

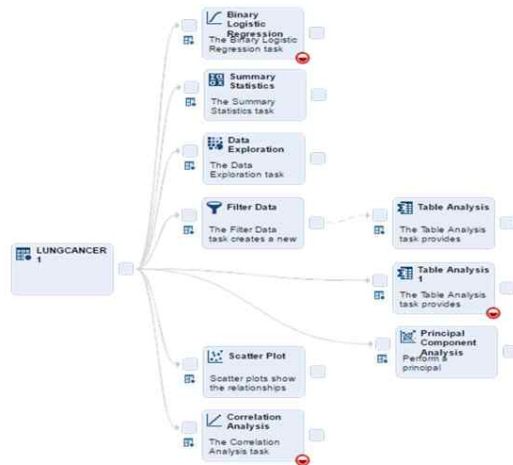


그림 15. SAS-U를 활용한 모형 개발과정

Logistic Regression (로지스틱 회귀분석)

유의한 변수만 선택하여 로지스틱 회귀식 도출

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	4.6386	0.6201	55.1845	<.0001
SEX	9	-0.8247	0.0935	97.6125	<.0001
SEX	1	0	0		
SIDO	11	-0.7113	0.2848	7.6285	0.0058
SIDO	26	-0.6441	0.2754	5.4705	0.0193
SIDO	27	-0.5615	0.2885	4.1185	0.0424
SIDO	28	-0.4340	0.2849	2.3201	0.1277
SIDO	29	-0.5501	0.2834	0.9312	0.6397
SIDO	30	-0.3652	0.2948	1.4276	0.2321
SIDO	31	-0.5270	0.3095	2.8994	0.0987
SIDO	35	164.2	1.974835	0.0000	1.0000
SIDO	41	-0.6103	0.2625	8.4043	0.0021
SIDO	42	-0.2284	0.2902	0.9194	0.4312
SIDO	43	-0.2245	0.2873	0.8105	0.4345
SIDO	44	-0.0841	0.2730	0.0993	0.9429
SIDO	45	-0.3590	0.2858	1.6562	0.1978
SIDO	46	-0.2775	0.2819	1.0660	0.3019
SIDO	47	-0.3285	0.2719	1.4238	0.2328
SIDO	48	-0.1719	0.2708	0.4034	0.5253
SIDO	49	0	0		
CTRB_PT_TYPE_CD	9	0.5475	0.3981	1.9124	0.1695
CTRB_PT_TYPE_CD	1	-0.1231	0.1173	1.1010	0.2940
CTRB_PT_TYPE_CD	2	-0.0578	0.1269	0.2080	0.6499
CTRB_PT_TYPE_CD	3	-0.20001	0.1207	0.0229	0.9803
CTRB_PT_TYPE_CD	4	-0.0313	0.1184	0.0097	0.9302
CTRB_PT_TYPE_CD	5	-0.0476	0.1187	0.1612	0.6881
CTRB_PT_TYPE_CD	6	0.30199	0.1149	0.0002	0.9835
CTRB_PT_TYPE_CD	7	-0.0822	0.1095	0.5634	0.4528
CTRB_PT_TYPE_CD	8	-0.0445	0.1080	0.1771	0.6739
CTRB_PT_TYPE_CD	9	-0.0773	0.0979	0.6231	0.4299

$$\begin{aligned} \text{logit}(p) &= -4.8386 - 0.8247 \times (\text{SEX}_{\text{Female}}) - 0.65 \times (\text{SMOKE}_{\text{No}}) \\ &\quad - 0.368 \times (\text{FAMILY}_{\text{Cancer}}) + 0.0838 \times (\text{AGE}) - 0.148 \times (\text{HMG}) \\ &\quad - 0.034 \times (\text{BMI}) - 0.004 \times (\text{TOT.CHO}) + 0.133 \\ &\quad \times (\text{OLIG.PROTE.CD}) - 0.0068 \times (\text{BP.HIGH}) \end{aligned}$$

그림 15. 로지스틱 회귀분석으로 개발한 폐암 예측 분류 모형

Odds Ratio (오즈비)

여성인 경우, 담배를 피는 경우, 가족력(암)이 있는 경우, 나이가 많을 경우, 혈액소가가 높은 경우엔 폐암 발생 유무에 대한 반응률이 높은 것을 알 수 있음

오즈비 해석

Odds Ratio Estimates				Odds Ratio Estimates			
Effect	Ratio	Estimate	95% Wald Confidence Limits	Effect	Ratio	Estimate	95% Wald Confidence Limits
SEX	Female	0.439	0.371 0.576	SIDO	46 vs 48	1.058	0.818 1.363
CTRB_PT_TYPE_CD	9 vs 10	1.728	0.795 3.759	SIDO	45 vs 48	0.691	0.384 1.212
CTRB_PT_TYPE_CD	1 vs 10	0.884	0.702 1.113	SIDO	46 vs 48	1.000	0.719 1.399
CTRB_PT_TYPE_CD	2 vs 10	0.944	0.736 1.211	SIDO	47 vs 48	0.771	0.462 1.294
CTRB_PT_TYPE_CD	3 vs 10	0.944	0.740 1.201	SIDO	48 vs 48	0.802	0.485 1.331
CTRB_PT_TYPE_CD	4 vs 10	1.031	0.807 1.304	DRUG_HABIT_CD	9 vs 1	1.127	1.120 1.135
CTRB_PT_TYPE_CD	5 vs 10	0.953	0.756 1.203	DRUG_HABIT_CD	9 vs 7	0.807	0.549 1.191
CTRB_PT_TYPE_CD	6 vs 10	1.000	0.800 1.250	DRUG_HABIT_CD	9 vs 7	0.802	0.549 1.191
CTRB_PT_TYPE_CD	7 vs 10	1.000	0.800 1.250	DRUG_HABIT_CD	9 vs 7	1.000	0.754 1.366
CTRB_PT_TYPE_CD	8 vs 10	0.958	0.777 1.177	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
CTRB_PT_TYPE_CD	9 vs 10	0.949	0.769 1.173	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.523	0.404 0.687	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	1.215	0.807 1.827	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	1.493	0.908 2.458	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.702 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.803	0.694 0.932	DRUG_HABIT_CD	9 vs 7	1.171	1.120 1.223
SMOKE_STATUS_CD	9 vs 1	0.815	0.694 0.959	DRUG_HABIT_CD	9 vs 7	1.171	1

*별첨3. 통풍환자 예측 분류 모형 개발을 통한 신규 비즈니스 모델 개발

국가지정 건강검진센터를 운영하면서 동시에 통풍 치료를 하는 “@@ 병원” 으로부터
통풍 환자 예측 모형 개발을 의뢰 받았다.
 병원에서 건강검진을 받는 사람들에게 추가 검진을 제안하여 **추가적인 진료 및 치료비로 수익을 올리는**
 새로운 비즈니스 모델 개발에 활용하기 위함이다.

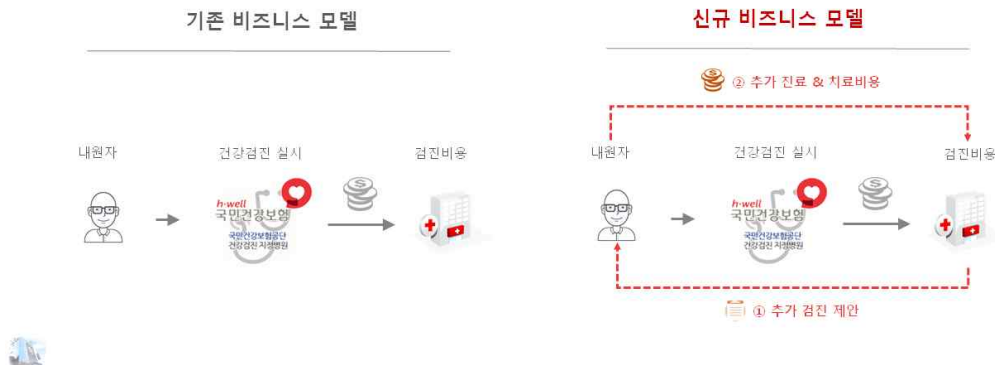


그림 17. 예측분류 모형 개발을 통한 신규 비즈니스 모델 발굴 가능

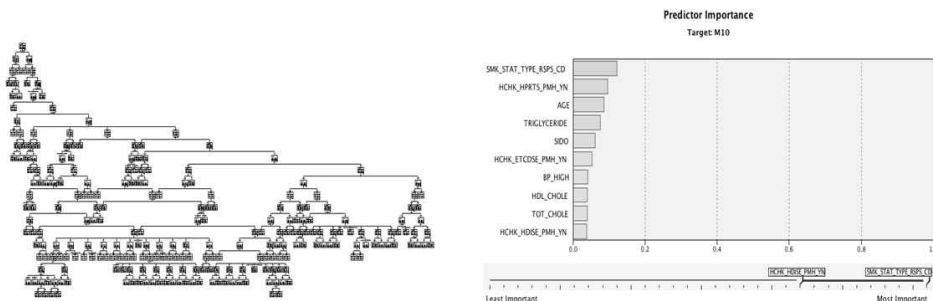
(2) 교차검증(Cross Validation)을 실시하여 모델 비교

Fold	정확도 (%)		
	C&RT	C5.0	Logistic Regression
1	66.9	95.4	82.8
2	66.6	94.0	82.4
3	71.5	94.8	84.3
평균	68.3	94.7	83.2
표준편차	2.75	0.70	1.00

3 모델의 교차검증을 통한 비교 결과, C5.0 모델이 **성능(94.7%)**, **안정성(0.7)** 측면 모두에서 다른 모델보다 뛰어남을 확인할 수 있다

그림 18. 교차검증을 통한 성능이 좋은 분류 모델 선택

종합적인 평가 결과, **C5.0 모델**이 가장 적합한 예측 모형으로 나타남



흡연자 > 본인 고혈압 과거병력 있는 자 > 나이가 많은 사람 > 높은 트리글리세라이드(지방) 수치
 > 본인 심장질환 과거병력 있는 자 > 높은 콜레스테롤 수치 의 순으로 모델에서 중요한 변수로 나타난다

그림 19. 개발한 통풍 예측 모형에 대한 설명

C5.0 모델 도입에 따른 경제적 수익 기대효과

해민병원의 경우, 연간 10만 여명의 사람들이 건강검진을 위해 병원을 방문한다
통풍 유병률은 약 0.5%(*1 국내 2008년 기준)으로 검진 내원자 중 500명 이상은 통풍환자일 가능성이 높다
매년 통풍환자가 늘어나는 만큼 실제 환자는 그 이상일 것. 게다가 해민병원은 관절염을 전문적으로 다루는 병원이므로 통풍환자의 방문이 더 높을 수 있을

일반 민영병원에서 진료 및 수술 권유 시, 공공병원 대비 수술치료 비율이 최대 6배 이상 증가한다.*2

$$500명 \times 646,248원 = 323,124,000원$$

신규 비즈니스 모델 수립 & 예측모형 개발에 따른 기대수익은 매년 약 **3억 2천만 원**이다.



*1. 이한희 & 성나영, "국민건강보험공단 자료를 이용한 국내 거주 통풍 환자의 유병률과 특성에 관한 연구", 대한류마티스학회, Journal of Rheumatic Diseases, 18권2호 (2011), pp.94-100
*2. 한겨레21, 의료상업화 보고서, http://h21.hani.co.kr/arti/cover/cover_general/31927.html, 2012.05

그림 20. 예측모형 개발에 따른 신규 비즈니스 모델 발굴과 그에 대한 기대효과

2017년 08월 27일

참가자 대표 : 오성우

(서명)

(사)금융보안원장 귀하