

우체국보험 CRM 강화를 위한 추가가입 유망고객에 관한 연구

김 소 연*

생명보험 시장이 포화되고 시장 내 경쟁이 심화됨에 따라 신규고객을 유치하는 것 만큼이나 기존고객을 활용하여 추가 가입을 유도하는 것에 대한 전략적 중요성이 강조되고 있다. 본 연구에서는 로지스틱 회귀분석모형과 의사결정트리모형을 활용하여 우체국보험 고객을 대상으로 어떠한 요인이 추가가입에 영향을 미치는지에 대해 살펴봄으로써 우체국보험 사업 CRM 강화 방안을 모색해 본다.

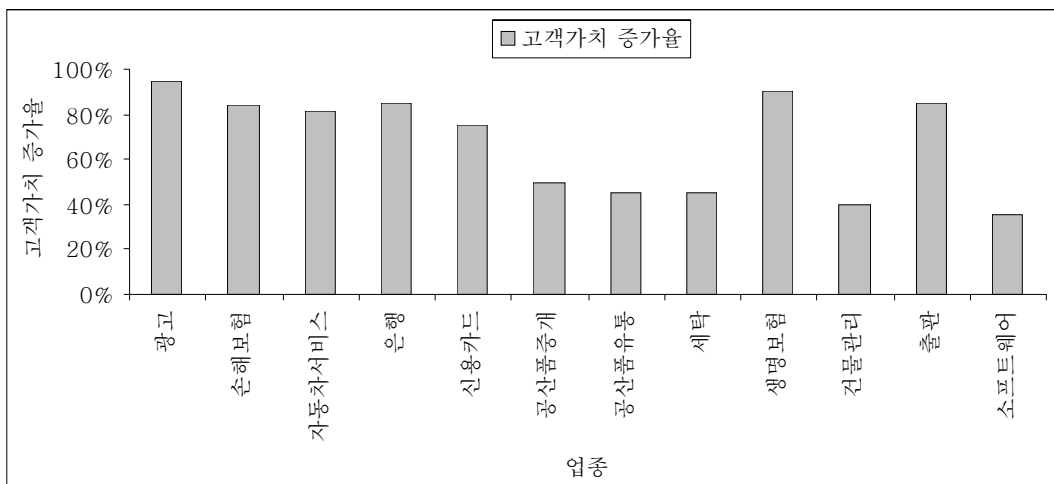
I. 머리말

1990년대 후반까지 생명보험 가입이 폭증한 데 기인하여 2008년 현재 국내 생명보험 가구 가입률은 90%에 육박하고 있으며 이처럼 시장이 포화상태에 도달함에 따라 보험에 대한 신규 수요는 크지 않은 상황이다. 더욱이 금융 자율화와 규제 완화로 인하여 보험시장 내 경쟁이 심화됨에 따라 기존의 마케팅 전략만으로는 효율적으로 고객을 유치하는 데 한계가 있다. 고객의 라이프스타일 변화에 따라 점차 다양한 형태의 위험보장이 요구되고 있으며 이에 따라 각 생명 보험사가 고객 요구에 맞추어 세분화된 상품개발을 추진하고 있는 지금 상품과 고객의 요구를 일치시키는 과학적인 마케팅 기술은 그 어느 때보다도 중시되고 있다. 이러한 관점에서 볼 때

* 우정사업본부 보험사업팀 전임연구원

보험시장의 가장 큰 이슈는 과거 보험사들이 앞 다투어 추진하던 신규고객 유치가 아니라 기존 가입 고객을 얼마나 잘 유지하며 그들에게 추가상품 가입을 유도할 수 있느냐 하는 부분이라 할 수 있다.

[그림 1] 업종별 고객유지율 증가(5%)에 따른 고객가치 변화



자료: 베인&컴퍼니.

최근 많은 기업들은 자사가 보유한 고객 데이터를 이용하여 시장에서의 경쟁력 확보를 위한 다양한 형태의 모델을 개발하는 데 데이터마이닝¹⁾ 방식을 적극 활용하고 있다. 특히, 은행, 카드, 보험 등 과거 고객 데이터를 확보하는데 심혈을 기울였던 금융회사들은 이 데이터를 적극적으로 활용할 수 있는 방법을 모색하고 있다. 생명보험 업계에서 이탈모형과 신상품에 대한 추가가입모형개발에 큰 관심을 가지고 적극 추진하는 이유도 이처럼 고객관계관리(Customer Relationship Management, CRM)에서 기존고객관리의 중요성이 강조되고 있기 때문이다. [그림 1]과 같이 고객유지율을 5% 증가시키는 경우, 평균 고객의 가치는 업종별로 35~100% 증가하는 것으로 조사되었다. 또한 신규고객 획득에 소요되는 비용은 기존고객 유지비용보다 5~10배가

1) 데이터마이닝: 대용량의 데이터에서 유용한 정보와 관계를 탐색하고 모형화한 후 지식으로 변환하여 기업의 의사결정에 활용하는 과정.

높은 것으로 조사되었다.

〈표 1〉에서 보는 바와 같이 Allied Insurance사가 자사의 고객들을 대상으로 실시한 조사에서도 고객이 자사로부터 구입한 보험 상품의 수가 증가할수록 유지율이 증가하는 경향이 있음을 보여주고 있다. 미국 보험대리점협회²⁾가 10년에 걸쳐 실시한 연구에서도 가입한 보험 상품의 수가 1개에서 2개로 증가할 때 유지율은 40%가 증가하고 2개에서 3개로 증가할 때에는 무려 90%가 증가함을 보여주고 있다. 또한 1996년 미국에서 실시된 조사도 비슷한 결과를 보이고 있다. 1개 상품에 가입한 고객의 연간 실효율은 10.4%의 높은 수준인 데 반하여, 가입상품의 개수가 2개, 3개, 4개로 증가함에 따라 실효율은 6.9%, 3.3%, 2.4%로 낮아지는 것으로 나타났다. 실효율 1%가 가져다주는 엄청난 이득을 감안하면 추가판매가 단순한 판매증진 이상의 큰 의미를 지니고 있음을 알 수 있다. 즉, 고객가치를 증대시키기 위해서는 고객유지율을 증가시켜야 하고, 고객유지율을 증대시키기 위해서는 가입한 보험 상품의 수를 증가시켜야 한다는 결론에 도달할 수 있다.

〈표 1〉 가입한 보험 상품별 평균 유지율

| 가입한 보험 상품별 고객 구분 | 평균 유지율 |
|---------------------------|--------|
| 자동차보험만 가입한 고객 | 77% |
| 주택보험만 가입한 고객 | 79% |
| 자동차보험 및 주택보험을 가입한 고객 | 96% |
| 자동차보험, 주택보험, 생명보험을 가입한 고객 | 97% |

2) 미국보험대리점협회(PIA): Professional Insurance Agents Association of America

II. 분석모형 설계와 자료탐색

1. 추가가입 모형 및 분석대상 정의

1) 모형 정의

추가 판매 모형은 우체국 보험 1건만을 보유한 고객 중 추가 가입 확률이 높은 것으로 여겨지는 고객의 속성을 발견하여 추가 가입 시에 정확한 타깃을 선정하고자 하는 데에 그 목적이 있다. 이에 따른 모형의 목표변수는 2007년 1월말 시점 1건 보유 고객 중 2007년 2월 1일에서 2007년 7월 31일 사이에 보험 추가 가입 여부(1: 가입, 0: 미가입)이며, 설명변수는 보험 계약 고객의 인구통계학적 속성 정보 및 거래 행태, 모집자 관련 정보 등에 대한 데이터를 활용하였다.

2) 분석 대상 정의

일반적으로 추가가입은 1건 이상의 승인된 계약을 보유하고 있는 고객이 계약이 유지된 상태에서 1건 이상을 계약하는 경우이지만 본고에서는 “계약이 승인된 상품을 1건만 보유하고 있는 계약자가 계약이 유지된 상태에서 추가로 1건 이상의 계약을 체결했을 경우”로 분석 대상을 한정한다. 2007년 1월 31일을 유지 계약 판정 시점으로 정하고, 추가가입기간은 2007년 2월 1일에서 2007년 7월 31일까지 6개월로 정한다. 또한 기존 유지계약 만기일(해약 또는 해지 포함)이 추가 가입 기간 중인 경우는 분석대상에서 제외하고, 외국인 및 주민번호 오류건과 계약자가 직원이거나 보험관리사인 경우도 분석 대상에서 제외한다.

2. 자료탐색 및 변수선택

1) 자료 수집

본 연구에서 사용된 데이터는 우체국 보험 데이터베이스에서 추가가입모형에 유의하다고 생각되어지는 고객 데이터이다. <표 2>와 같은 29개의 변수로, 크게 인구통계학적인 변수와 계약과 관련된 변수들로 구성되어 있다.

〈표 2〉 분석 모형을 위한 필요 변수 항목

| 대분류 | 변수명 | 내 용 |
|------|-----------|-----------------------------|
| 인구통계 | 계약자 나이 | 계약자 연령 |
| | 계약자 성별 | 계약자 남자/여자 |
| | 계약자 직업 | 11개 범주의 직종대분류로 구성 |
| | 계약자 지역 | 16개 지역으로 분류 |
| | 피보험자 연령 | 피보험자 연령 |
| | 피보험자 성별 | 피보험자 남자/여자 |
| 계약관련 | 보유 상품군 | 건강/암/재해/중신/저축/연금/어린이/교육 |
| | 납입방법 | 월납/3개월납/6개월납/연납/일시납 |
| | 수급방법 | 방문/창구/우체국이체/은행이체 |
| | 납입기간 | 보험료 납입기간(월 단위) |
| | 보험기간 | 보험계약 유지기간(월 단위) |
| | 보장경과기간 | 계약일로부터 기간(월 단위) |
| | 월납환산보험료 | 납입방법에 따라 환산된 매월 납입 보험료 |
| | 과거 실�효 횟수 | 해당 증서 과거 실�효 횟수 |
| | 지급보험금 | 사고 보험금 |
| | 환급금 대출 잔액 | 환급금 대출 잔액 |
| | 가입기간 | 계약자의 첫 가입증서 계약일로부터 기간 |
| | 소멸건수 | 과거 증서 소멸 건수 |
| | 모집직군 | 모집직군 |
| | 모집자성별 | 모집자의 성별 |
| | 모집자 연령 | 모집자의 연령 |
| | 모집자 입사차월 | 모집자 입사차월 |
| | 계_피 동일여부 | 계약자와 피보험자가 동일이면 1, 아니면 0 |
| | 모_수 동일여부 | 모집자와 수급자가 동일이면 1, 아니면 0 |
| | 미아고객 여부 | 미아고객이면 1, 아니면 0 |
| | 보험료 완납 여부 | 보험료를 정상 및 면제완납이면 1, 미완납이면 0 |
| | 피보험자 보유건수 | 피보험자 기준 보유건수 |
| | 보험금 지급 여부 | 사고보험금 지급한 경험이 있으면 1, 아니면 0 |
| | 환급금 대출 여부 | 환급금 대출한 경험이 있으면 1, 아니면 0 |

인구통계학적인 변수 중 계약자의 직업은 전문직, 행정·사무 관리직, 농림수산업 및 광업, 섬유제조업, 인쇄·출판업, 식료품 제조업, 화학·비금속·금속산업, 전기 통신업, 건설업, 유

흥·서비스업, 기타의 11개 범주로 구성되어 있다.

보험계약과 관련된 변수들로는 보유 상품군, 납입·수금 방법, 납입기간, 보험기간, 지급 보험금, 월납환산보험료, 미야고객여부, 모집자 입사차월, 과거 실효횟수 등이 있다. 보험 테이블에 원시적으로 존재하는 변수를 이용하여 새롭게 구성된 파생변수에 대한 설명을 한다면 다음과 같다.

보장경과기간이란 보유 중인 1건 증서의 계약일로부터 기준 월까지의 기간을 월단위로 나타낸 것이며, 과거 실효횟수는 보유 중인 증서의 계약일로부터 기준 월까지의 기간 동안 실효 횟수를 의미한다. 지급보험금은 사고발생으로 인한 지급 보험금만을 한정하며, 가입기간은 보유하고 있는 증서뿐만 아니라 고객이 우체국에 최초 가입한 계약일로부터의 기간으로, 고객에 따라서 보장경과기간과 동일할 수도 있고, 보장경과기간보다 훨씬 더 길수도 있다.

소멸건수는 2007년 1월말 현재 고객별 소멸된 증서수를 의미하며, 모집자 입사차월은 모집자의 입사일로부터 기준 월까지의 월 단위 기간이고, 모집자가 퇴직이나 이직 등의 이유로 수급자가 관서로 바뀌게 되는 경우가 있는데, 고객별 모든 수급자가 관서로 되어 있는 경우 미야고객이라고 한다. 계_피 동일여부, 모_수 동일여부는 계약자와 피보험자, 모집자와 수급자가 같은 1, 다르면 0으로 정의하였고, 보험료 완납여부, 보험금 지급여부, 환급금 대출여부는 해당사항이면 1, 아니면 0으로 하였다.

2) 자료 분석

(1) 인구통계학적 속성

분석 대상자에 대한 인구통계학적인 특성을 살펴보면 다음과 같다. 성별로는 전체 2,127,901명 중 계약자는 여성 58.4%, 남성 41.6%인 반면, 피보험자는 남성이 47.3%로 계약자로는 여성이 월등히 높은 비율을 보이고 있다. 지역별로는 서울이 약 16.7%로 가장 높은 비율을 차지하고 다음으로 경기, 부산, 대구 순이다. 연령대로는 계약자는 46세~55세가 가장 높은 비율을 차지한 반면, 피보험자는 전체 연령대별로 고른 분포를 볼 수 있다. 즉 40~50대 여성 계약자가 다양한 연령대의 피보험자를 대상으로 보험에 가입함을 알 수 있다.

(2) 보험계약과 관련된 변수의 특성

보험계약과 관련된 변수들의 특성을 살펴보면 다음과 같다. 보험 상품의 종류로는 재해보험이 44.7%로 가장 높은 비율을 차지하고, 건강·암 보험이 18.6%, 11.8%의 비율로 보장성 위주의 보험이 높은 비율을 차지하고 있다. 납입방법은 매월 보험료를 납입하는 월납이 97% 이상으로 가장 많으며, 수금방법으로는 우체국 및 은행 이체가 89.2%로 가장 높은 비율을 차지하고 있다. 보장경과기간(계약일로부터의 기간)은 4년~7년이 52.5%로 가장 많은 분포를 보이며, 미아고객은 전체 고객의 49.2%를 차지하고 있다. 보험료 납입기간은 49.1%가 5년에서 10년 사이에 분포함을 알 수 있다. 보험기간은 55년이 넘는 보험계약이 40.4%로 가장 많고 다음으로 10년에서 20년 사이의 보험계약이 약 36.4%를 차지하고 있다.

모집직군으로는 보험관리사 36.5%, 기능직 내근자 29%, 일반직공무원 15.8%의 순서이고, 모집자 성별로는 67.7%로 여성의 비율이 높으며 모집자 연령대는 40대가 가장 높은 비율을 차지한다. 모집자와 수급자가 동일한 경우가 48.7%이고 계약자와 피보험자가 동일한 경우가 78.8%로 계약자가 다른 피보험자로의 추가가입 여력이 있다. 보험료 완납인 경우는 16.1%에 불과하여 납입방법이 일시납에 비해 월납이 많은 부분과 같은 의미일 수 있다. 모집자 입사차월을 보면 10~20년이 62.0%, 5~10년이 30.8%로 장기근속의 특징을 알 수 있다. 보험금을 지급받은 경험이 있는 고객은 13.5%로 많지 않음을 알 수 있다.

대상기간을 판단하는 시점 이전에 실효한 적이 있는 횟수는 0건이 95%를 넘는 것으로 나타나며, 피보험자 기준으로 보유하고 있는 증서수가 1건이 90%를 넘는 것으로 보아 추가 가입 여력이 충분함을 알 수 있다. 계약자가 과거에 가입한 증서 중 소멸된 증서 수는 0건이 72.7%이고, 가입기간이 1년 초과~10년 이하가 80%를 넘는 것으로 보아 현재 보유하고 있는 증서가 우체국 보험에 첫 번째 가입한 증서이면서 가입한 지 1년~10년 정도인 고객이 70~80%를 차지함을 알 수 있다.

3) 모형 구축에 사용할 변수 선택

전체 분석 대상 중 분석하고자 하는 추가 가입률이 2.4%(2,127,901명 중 50,624명 추가 가입)로 추가가입자의 비율이 현저히 낮아서 모형 분석 시 추가 가입자의 속성 및 행태가 잘 드러

나지 않을 수 있어, 모형을 위한 마스터 테이블 생성 시 Over-sampling 방식을 적용하였다.³⁾ 즉, 추가 가입한 고객은 50,624명을 전체 추출하고 비가입한 고객은 추가 가입 고객의 3배를 표본으로 추출하여 분석 대상으로 선정한다.

모형 구축 시 모든 변수를 이용하여 모형을 구축하지는 않는다. 지나치게 많은 변수를 사용하면, 모형의 적합도는 올라가지만 모형의 구축에 사용되지 않은 새로운 데이터에 적용시켰을 경우에는 오류율이 오히려 높아지는 과잉맞춤(Over-fitting) 문제가 발생할 수 있기 때문이다.

모형 구축에 사용할 변수를 선정하기 위하여 입력 변수들을 줄이기 위한 다양한 통계적 방법들이 다양하게 사용되고 있다. 본 연구에서는 다양한 입력 변수 및 파생변수 중 추가가입 모형 구축에 사용될 변수를 예측하는데 유의하지 않은 변수를 일부 제거하기 위해 카이제곱(Chi-Square) 통계량과 Correlation 분석, F통계량, 그리고 로지스틱 회귀분석을 실시하였다. 먼저 16개의 범주형 변수들을 대상으로 목적변수(추가가입, 미가입)와의 교차분석을 실시한 결과, 유의수준 0.05에서 모든 변수 값이 유의하여 변수 목록에 모두 포함되었다.

다음으로 13개의 연속형 변수들을 대상으로 목적변수(추가가입, 미가입)와의 상관분석과 F-test를 실시하여, 유의수준 0.05에서 유의하지 않은 변수들을 제외시켰다. 그 결과 지급보험금과 환급금 대출 잔액의 상관분석 p-value가 0.5742, 0.5669로 유의하지 않게 나타나서 분석에 사용되는 변수목록에서 제외되었다.

마지막으로 29개의 변수들 모두를 목적변수(추가가입, 미가입)와의 로지스틱 회귀분석을 스텝와이즈(stepwise)⁴⁾ 방식으로 실시하여 유의수준 0.05에서 최종적으로 선택되는 변수들을 살펴보면, 모집자 성별의 유의수준이 0.1450으로 유의하지 않게 나타나서 변수목록에서 제외되었다. 이상의 분석결과를 바탕으로 네 가지 분석 방법 모두에서 선택된 변수들을 사용하여 고객 추가가입 예측 모형 구축에 사용하였다.

3) Over-sampling 방법: 모집단 내에 Event 발생 건은 샘플링 수행 시 발생 건을 모두 이용하며, 상대적으로 Non-Event 영역은 샘플링을 수행하여 Event 발생 정보의 손실을 줄여 특성을 밝혀내는 데 유리한 샘플링 기법.

4) Stepwise(증감법): 독립변수를 하나도 포함시키지 않은 회귀식에서 각 독립변수에 대하여 독립변수가 추가 혹은 제거함을 반복하여 유의한 변수들만 선택하는 통계적 기법.

Ⅲ. 분석 결과

1. 추가가입예측모형 성능 비교

로지스틱 회귀분석과 의사결정나무모형(Decision Making Tree Analysis)을 이용한 추가 가입 예측모형의 성능에 어떠한 차이가 있는지 살펴본다. 로지스틱 회귀분석, 의사결정나무모형을 각각 이용한 추가가입 예측모형에서 정확도를 나타내는 지표인 오분류율⁵⁾과 실제 추가 가입고객들을 얼마나 잘 예측하였는지를 나타내는 검출력을 기준으로 살펴보기로 한다.

〈표 3〉 로지스틱 회귀분석에 의한 추가가입모형의 오분류표

| 구 분 예측 실제 | 모형 구축용 | | | 모형 검증용 | | | 모형 시험용 | | |
|-----------------|-------------|-------|--------|-------------|-------|--------|-------------|-------|--------|
| | 미가입 | 가입 | 합계 | 미가입 | 가입 | 합계 | 미가입 | 가입 | 합계 |
| 미가입 | 59,527 | 1,318 | 60,845 | 44,529 | 962 | 45,491 | 44,369 | 913 | 45,282 |
| 가입 | 18,217 | 1,835 | 20,052 | 13,847 | 1,335 | 15,182 | 14,047 | 1,343 | 15,390 |
| | 정분류율: 75.9% | | | 정분류율: 75.6% | | | 정분류율: 75.3% | | |

〈표 4〉 의사결정나무분석에 의한 추가가입모형의 오분류표

| 구 분 예측 실제 | 모형 구축용 | | | 모형 검증용 | | | 모형 시험용 | | |
|-----------------|-------------|-------|--------|-------------|-------|--------|-------------|-------|--------|
| | 미가입 | 가입 | 합계 | 미가입 | 가입 | 합계 | 미가입 | 가입 | 합계 |
| 미가입 | 59,231 | 1,614 | 60,845 | 44,211 | 1,280 | 45,491 | 44,083 | 1,199 | 45,282 |
| 가입 | 17,774 | 2,278 | 20,052 | 13,519 | 1,663 | 15,182 | 13,762 | 1,628 | 15,390 |
| | 정분류율: 76.0% | | | 정분류율: 75.6% | | | 정분류율: 75.3% | | |

로지스틱 회귀분석의 세부옵션은 가장 일반적인 형태인 디폴트로 하였다. 의사결정나무분석

5) 오분류율: 목표변수의 범주가 0, 1일 때, 실제 1인 관찰치의 빈도 중 (실제 0, 예측 1)의 빈도+(실제 1, 예측 0)의 빈도의 비율로서 정분류율(Correct Classification)은 1-오분류율과 같은 개념.

은 기본설정에 따르기로 하고, 분리기준으로는 Chi-square를 이용하여 유의수준이 0.2를 넘지 않도록 하였으며 뿌리마디로부터 끝마디까지의 깊이는 최대 10을 넘지 않도록 설정하였다. 기타 Advanced 옵션들은 디폴트로 하였다. 예측의 정확도를 나타내는 로지스틱 모형과 의사결정나무모형의 오분류표를 살펴보면 <표 3>, <표 4>와 같다.

로지스틱 회귀분석 모형에서는 전체 202,242명에서 모형 구축용(Training)으로 추출된 80,897명 중에서 가입을 가입으로 분류하고, 미가입을 미가입으로 분류한 고객의 수가 61,362명이고 정확도는 75.9%이다. 모형 검증용(Validation)으로 추출된 60,673명의 고객 중에서 가입을 가입으로 분류하고, 미가입을 미가입으로 분류한 고객의 수가 45,864명이고 정확도는 75.6%이다. 모형 시험용(Test)으로 추출된 고객 60,672명 중에서 가입을 가입으로 분류하고, 미가입을 미가입으로 분류한 고객의 수가 45,712명으로 정확도는 75.3%이다.

의사결정나무 모형에서는 전체 202,242명에서 모형 구축용(Training)⁶⁾으로 추출된 80,897명 중에서 가입을 가입으로 분류하고, 미가입을 미가입으로 분류한 고객의 수가 61,509명이고 정확도는 76.0%이다. 모형 검증용(Validation)⁷⁾으로 추출된 60,673명의 고객 중에서 가입을 가입으로 분류하고, 미가입을 미가입으로 분류한 고객의 수가 45,874명이고 정확도는 75.6%이다. 모형 시험용(Test)⁸⁾으로 추출된 고객 60,672명 중에서 가입을 가입으로 분류하고, 미가입을 미가입으로 분류한 고객의 수가 45,711명으로 정확도는 75.3%이다.

앞의 표와 같이 의사결정나무분석, 로지스틱 회귀분석 모두 정분류율은 75.9~76.0%이고, 검증 결과나 테스트 결과의 정분류율은 75.6%, 75.3%로 잘 분류되므로 안정적인 모형이며, 정확도가 유사함을 알 수 있다.

검출력은 <표 5>와 같이 모형 검증용(Validation) 데이터 60,673개를 대상으로 Threshold가 15, 20, 25, 30, 35일 때를 살펴보면, 로지스틱 회귀분석이 Threshold가 20일 때는 0.766으로 의사결정나무분석 0.803보다 낮지만, 나머지 Threshold에서는 검출력도 높고, 전

6) 모형 구축용(Training): 모형에 가장 적합한 가중치(weight)를 찾기 위해 사용하는 데이터 셋트.
 7) 모형 검증용(Validation): 분석용(Training) 데이터를 이용하여 찾은 모형의 적합도를 평가하는 데이터 셋트.
 8) 모형 시험용(Test): 모형의 일반오류의 최종적인 불편추정치(unbiased estimate)를 선택하는 데 사용하는 데이터 셋트.

반적으로 안정적임을 볼 수 있다.

〈표 5〉 의사결정나무, 로지스틱 회귀분석 모형 검출력

| Threshold | 의사결정나무분석 | 로지스틱 회귀분석 |
|-----------|----------|-----------|
| 15 | 0.803 | 0.893 |
| 20 | 0.803 | 0.766 |
| 25 | 0.375 | 0.620 |
| 30 | 0.321 | 0.474 |
| 35 | 0.313 | 0.347 |

추가로 [그림 2], [그림 3]의 ROC 곡선⁹⁾과 Lift 도표¹⁰⁾를 통해서도 모형을 평가하고 비교할 수 있는데, 2개 모형의 ROC 곡선 패턴이 유사하며, 도표의 왼쪽 상단에 더 가까운 모형일수록 성능 면에서 우수한 모형으로 판단한다고 할 때, 두 모형 모두 민감도가 비교적 높은 편임을 알 수 있다. Lift 도표를 보면, 상위 10%에서 Lift값이 모두 2를 넘는 것으로 볼 수 있다. 즉, 모형을 고려하지 않은 상태에서 10%를 무작위 표본 추출하여 캠페인 등의 마케팅을 할 때의 성공률보다 모형의 추가가입점수 상위 10% 집단에 대해서 마케팅 활동을 시행할 때의 성공률이 2배 이상이 됨을 알 수 있다. 결국, 모형 구축의 효과를 2배 이상 볼 수 있다는 점에서 두 모형 모두 우수한 모형임을 알 수 있다.

앞에서 본 바와 같이, 두 가지 모형의 정분류율이 75.9~76.0%이고, 검증결과와 테스트 결과 모두 안정적이므로, 정확도에 큰 차이가 없는 것으로 판단되며, 검출력 부분에서는 로지스틱 회귀분석이 더 안정적임을 볼 수 있으나, ROC 곡선 및 Lift 도표를 통해 2가지 모형이 유사한 성능을 나타내는 모형임을 알 수 있다.

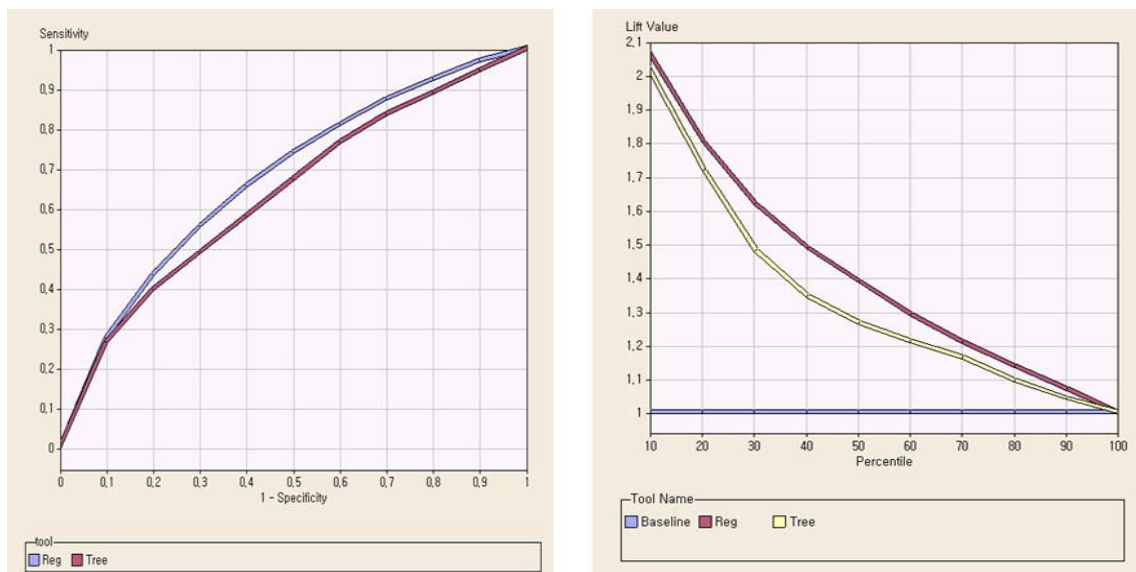
9) ROC 곡선: 이진형의 목표변수를 가지는 모형들의 성능을 민감도와 특이도를 이용하여 비교, 평가하는 도표.

10) Lift 도표: 사후확률을 정렬한 뒤 각 10% 집단별 향상도를 나타냄.

2. 추가가입 요인 분석

본 절에서는 로지스틱 회귀분석, 의사결정나무분석 각 모형에 의해 추가가입에 유의한 설명 변수를 찾아내고, 선정된 유의 변수에 의한 유망 노드 발굴 및 고객 군집별 profiling을 시도한다.

(그림 2) ROC 곡선 및 Lift 도표



1) 로지스틱 회귀분석

로지스틱 회귀분석 결과를 살펴보면 다음과 같다. <표 6>은 유의한 변수들을 정리한 것이다. 추가가입에 유의한 변수들을 살펴보면, 계약자 피보험자 동일여부, 모집자 수급자 동일여부, 보험기간, 모집자 연령 및 입사차월, 소멸건수, 계약자 성별 및 연령, 보장경과기간, 환급금대출 여부, 지급보험금 여부가 있다.

이들 변수들 중 소멸건수, 보험기간, 계약자_피보험자 동일 여부는 추가가입에 긍정적인 영향을 미치는 것으로 나타났다. 즉, 소멸건수가 많을수록, 보험기간이 길수록, 계약자와 피보험자가 다른 경우 추가가입률이 향상되는 것으로 판단된다. 보험기간은 회귀계수 값이 0.003 이하로 그 영향력이 비교적 미비한 것으로 나타난다.¹¹⁾

모집자_수금자 동일 여부, 모집자 연령 및 입사차월, 계약자 성별 및 연령, 보장경과기간, 환급금 대출여부, 지급보험금 여부는 추가가입에 부정적인 영향을 미치는 것으로 나타났다. 즉, 모집자나 계약자 연령이 낮을수록, 보장경과기간이 짧을수록, 계약자 성별이 여성인 경우 추가 가입 확률이 높은 것으로 나타났고, 환급금 대출이 없는 경우, 모집자와 수금자가 다른 경우, 지급 보험금을 받지 않은 경우는 추가가입 확률이 낮아진다. 입사차월은 회귀계수 값이 -0.000188 로 영향력이 다소 미약한 것으로 판단된다.

〈표 6〉 E-miner의 로지스틱 회귀분석 결과

| 구 분 | Estimate | pr.)Chisq | 구 분 | Estimate | pr.)Chisq |
|-------------|----------|-----------|------------|-----------|-----------|
| 계_피 동일여부(0) | 0.0726 | <.0001 | 입사차월 | -0.000188 | <.0001 |
| 소멸건수 | 0.1775 | <.0001 | 계약자 성별 | -0.3186 | <.0001 |
| 환급금대출여부(0) | -0.1883 | <.0001 | 계약자 연령 | -0.00885 | <.0001 |
| 모집자 연령 | -0.00472 | <.0001 | 보장경과기간 | -0.0111 | <.0001 |
| 모_수 동일여부(0) | -0.1261 | <.0001 | 지급보험금여부(0) | -0.0499 | <.0001 |
| 보험기간 | 0.00226 | <.0001 | | | |

로지스틱 회귀분석의 결과 중 입력변수가 분류결정에 얼마나 영향을 미치는지의 정도를 오즈비(Odds Ratio)¹²⁾를 통해서도 살펴본다.

〈표 7〉을 살펴보면 다음과 같다. 계약자와 피보험자가 다른 경우는 동일한 경우에 비해 추가 가입 가능성이 1.156배 높고, 계약자 성별이 여성인 경우 남성에 비해 1.890(1/0.529)배 높다. 또한 소멸건수가 적은 고객이 많은 고객에 비해 추가 가입할 가능성이 1.194배 높으며, 모집자와 수금자가 동일한 조건의 고객이 동일하지 않은 고객에 비해 1.287(1/0.777)배 높음을

11) 회귀분석 결과, 각 변수의 Estimate(회귀계수)가 양의 값이면 값이 커질수록 영향력이 높고, 음의 값이면 값이 작을수록 영향력이 높아짐.

12) 오즈비(odds ratio): 다른 모든 입력변수가 일정한 상태에서 x_i 가 1단위 증가할 때의 영향력. 여기서 오즈비가 1보다 작다는 것은 입력변수가 x_i 가 감소방향으로 영향을 미침을 의미하고, 반대로 오즈비가 1보다 크다는 것은 증가방향으로 영향을 미침을 의미한다.

알 수 있고, 환급금 대출을 받은 고객이 받지 않은 고객에 비해 1.458(1/0.686)배 높음을 알 수 있다. 보험금 지급받은 경험이 있는 고객이 받지 않은 고객보다 1.105(1/0.905)배 높다. 입사차월은 0.999, 보험기간은 1.002로 영향력이 약함을 볼 수 있다.

〈표 7〉 E-miner의 로지스틱 회귀분석 오즈비 결과

| Input | Odds Ratio |
|----------------------|------------|
| 계약자 피보험자 동일여부 0 vs 1 | 1.156 |
| 보장경과기간 | 0.989 |
| 계약자 성별 1 vs 2 | 0.529 |
| 소멸 건수 | 1.194 |
| 모집자 수급자 동일여부 0 vs 1 | 0.777 |
| 환급금 대출여부 0 vs 1 | 0.686 |
| 계약자 연령 | 0.991 |
| 입사차월 | 0.999 |
| 보험기간 | 1.002 |
| 보험금지급여부 0 vs 1 | 0.905 |

앞에서와 같이 로지스틱 회귀분석 결과를 회귀계수와 p-value를 통한 유의성 검증, 오즈비를 통한 목표변수예의 영향정도의 두 가지 기준으로 살펴보았다. 공통된 점으로는 계약자와 피보험자가 다른, 낮은 연령의 여성 계약자, 보장경과기간이 짧고, 모집자와 수급자가 동일한 경우에 추가가입 확률이 높아지는 것으로 나타난다.

2) 의사결정나무분석

SAS의 Enterprise-Miner의 Tree를 이용한 의사결정나무모형 결과를 살펴보면 [그림 4]와 같다. 노드의 depth를 7로 하였을 때 총 19개의 터미널 노드¹³⁾가 생성되며, 분석대상의 추가가입률이 25%에 비하여 최고 72.2%의 추가가입률을 나타내는 유망고객노드도 나타남을 볼 수 있다. 유망고객을 선정하기 위한 분류규칙 변수로는 보장경과기간, 계약자 성별, 소멸건수,

13) 터미널 노드: 가장 마지막에 있는 Node이며 세분화 집단의 수를 나타냄.

계약자 연령, 모집자 입사차월, 계약자_피보험자 동일여부가 선정됨을 알 수 있다.

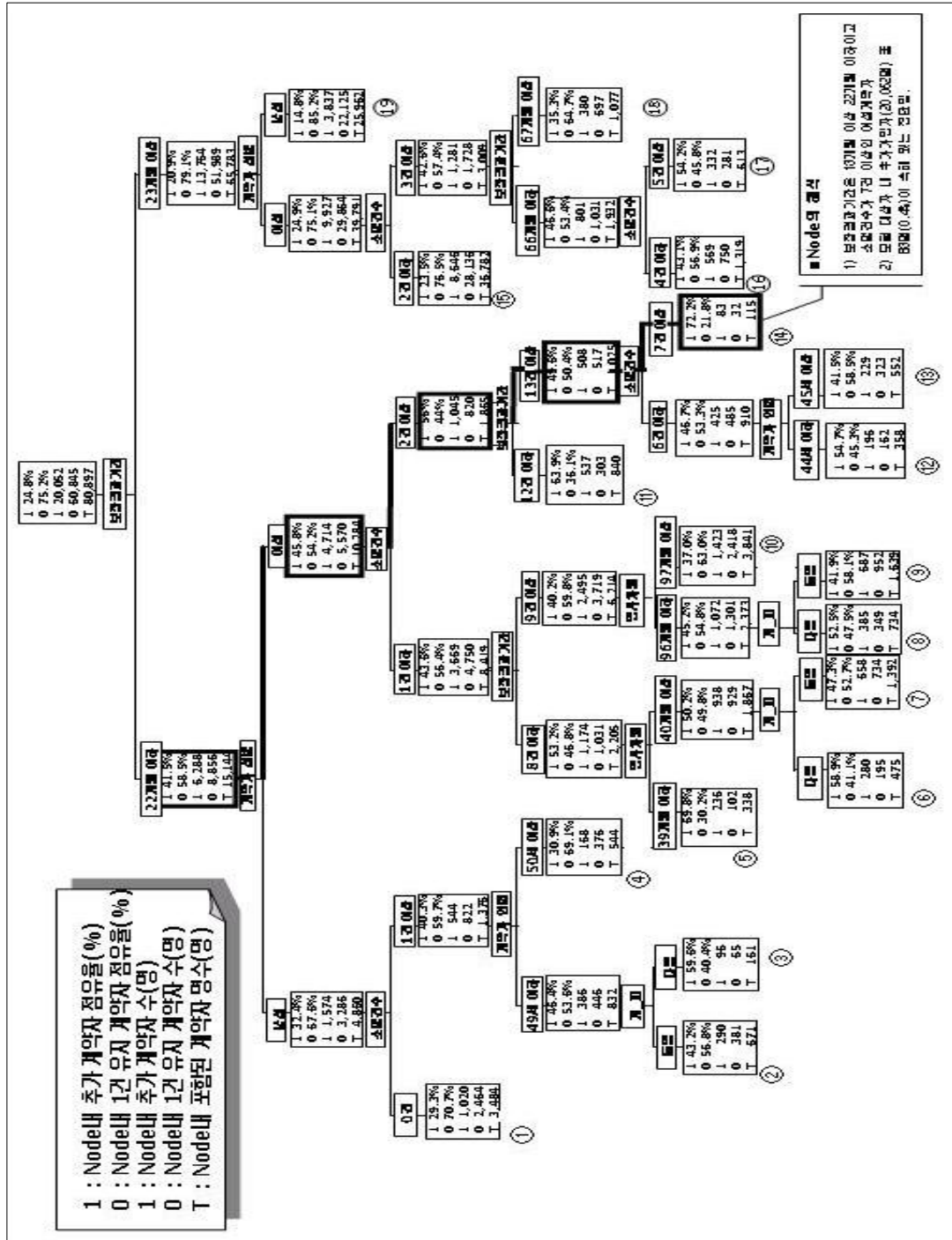
첫 번째 분류기준은 보장경과기간 22.5개월 이상, 22.5개월 미만으로 분할되고, 두 번째는 계약자 성별, 세 번째는 소멸건수에 의해 분할되었고, 네 번째 이하로는 직전 노드의 성향에 따라 달라지는데, 새로운 변수로는 계약자 연령, 계약자 피보험자 동일여부, 모집자 입사차월이 분류규칙변수로 나타나고, 보장경과기간, 소멸건수가 또다시 분류변수로 나타나 분류된다.

〈표 8〉 의사결정나무분석 결과 추가가입률 상위 6개 노드 특성

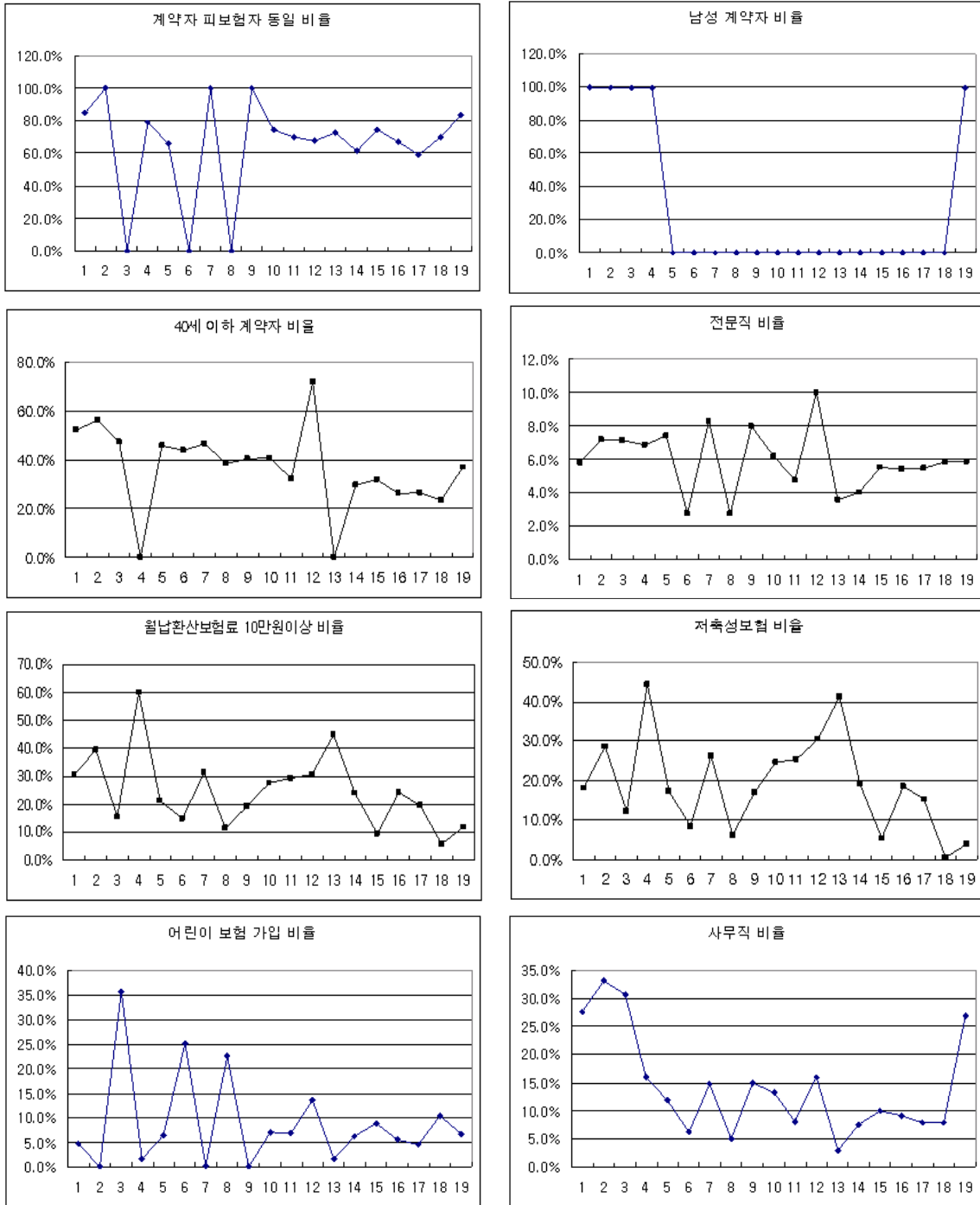
| 순위 | 노드 | 추가가입 고객 특성 | 추가가입률 | | 비율 |
|----|----|---|-------|-------|---------------------------------------|
| | | | 구축용 | 검증용 | |
| 1 | 14 | 보장경과기간 13개월 이상 22개월 이하, 여성 계약자, 소멸건수 7건 이상 | 72.2% | 79.6% | 전체 80,897명 중 115명이며 83명(72.2%)이 추가가입 |
| 2 | 5 | 보장경과기간 8개월 이하, 여성 계약자, 소멸건수 1건 이하, 모집자 입사차월 3년 3개월 이하 | 69.8% | 57.2% | 전체 80,897명 중 338명이며 236명(69.8%)이 추가가입 |
| 3 | 11 | 보장경과기간 12개월 이하, 여성계약자, 소멸 건수 2건 이상 | 63.9% | 63.8% | 전체 80,897명 중 840명이며 537명(63.9%)이 추가가입 |
| 4 | 3 | 보장경과기간 22개월 이하, 49세 이하 남성계약자, 소멸건수 1건 이상, 계약자_피보험자 다름 | 59.6% | 53.3% | 전체 80,897명 중 161명이며 96명(59.6%)이 추가가입 |
| 5 | 6 | 보장경과기간 8개월 이하, 여성계약자, 소멸건수 1건 이하, 모집자 입사차월 3년 4개월 이상, 계약자_피보험자 다름 | 58.9% | 56.1% | 전체 80,897명 중 475명이며 280명(58.9%)이 추가가입 |
| 6 | 12 | 보장경과기간 13개월 이상 22개월 이하, 44세 이하 여성계약자, 소멸건수 2건 이상 6건 이하 | 54.7% | 49.6% | 전체 80,897명 중 358명이며 196명(54.7%)이 추가가입 |

19개의 노드 중 추가가입률 상위 6개의 노드에 대한 고객 특성 및 노드 내 1건 유지 계약자 및 추가 계약자 비율을 살펴보면 〈표 8〉과 같다. 구축용과 검증용의 추가가입률이 큰 차이를 보이지 않으므로, 안정적인 모형이라고 볼 수 있다. 49세 이하 남성계약자가 포함된 추가가입률이 네 번째로 높은 노드를 제외하고는 여성 계약자가 공통 유의변수로 포함되고, 보장경과기

(그림 3) E-miner에 의한 의사결정나무분석 결과



(그림 4) 의사결정나무분석 결과 19개 노드별 고객 속성



간은 22개월 이하로 압축된다. 모집자 입사차월을 살펴보면, 5번 노드는 모집자 입사차월이 3년 3개월 이하, 6번 노드는 3년 4개월 이상으로 상이하게 나타나지만, 5번 노드의 추가가입률이 6번 노드의 추가가입률보다 높고, [그림 4]에서 5번 노드가, 6, 7번 노드보다 추가가입률이 높고, 8, 9번 노드의 추가가입률이 10번 노드보다 높음을 볼 때, 모집자 입사차월이 낮을수록 추가가입률이 높음으로 판단할 수 있다. 소멸건수도 입사차월과 마찬가지로, <표 8>은 소멸건수가 낮은 경우, 높은 경우 모두 나타나지만, [그림 4]에서 보면, 소멸 건수가 1.5 미만인 1번 노드가 소멸건수가 0.5 이상인 2, 3, 4번 노드에 비해 추가가입률이 낮고, 소멸건수가 2.5 미만인 15번 노드가 소멸건수가 2.5 이상인 16, 17, 18노드에 비해 추가가입률이 낮음으로 소멸건수가 많은 경우가 추가가입률이 높음으로 예측할 수 있다.

<표 9> 유망 고객군 분석

| 노드 | 분석 결과 및 평가 |
|----------|---|
| 4, 13 | 추가가입가능성이 높다. 보험료 10만 원 이상 비율이 높고, 저축성보험 비율이 높은 노년층 집단. 특히 일시납 비율이 높은 것으로 보아 경제적으로 여유가 있는 고객 층 |
| 2 | 추가가능성도 높고 사무직 비율이 높은 40세 이하 남성 계약자 집단. 모두 계약자와 피보험자가 동일하고, 10만 원 이상의 보험료 비율이 높은 편이므로, 추가가입 여력이 있음으로 판단되며, 추가계약을 통해 관리강화를 할 경우 회사입장에서 유리한 고객임 |
| 7, 9, 12 | 추가가능성이 높고 전문직 및 사무직 비율이 높은 40세 이하 여성 계약자 집단. 10만 원 이상의 보험료 비율이 높으며 저축성 보험 비율이 높으므로, 향후 밀착된 고객관계관리를 통한 안정 고객에서 우수고객으로의 가능성이 높은 고객층. 특히 어린이보험 가입비율이 12노드는 13.4%, 7, 9노드는 0%로 대조적인 결과를 보이는데, 두 가지 모두 추가가입기회가 많을 것으로 보이는 고객 층 |
| 3 | 추가가능성이 높고 40세 이하 비율이 47.4%로 높고 사무직 비율이 30%를 넘는 남성 집단. 보험료는 낮은 편이고 타인을 피보험자로 한 계약이 100%로 본인을 담보로 한 추가계약이 가능한 고객층 |
| 5 | 추가가능성이 높음. 40세 이하 그룹과 40세 이상 그룹이 50%씩 점유하고 있는 여성 고객층이며, 전문직 비율이 높은 편이며 본인을 담보로 한 계약이 60%를 넘으며 건강보험 가입비율이 높은 고객층으로 본인을 담보로 한 기타 상품 추가가입이 가능해보임 |
| 6, 8 | 추가가입가능성이 높으며, 10만 원 이하의 보험료 비율이 높고, 타인을 피보험자로 한 계약이 100%로 어린이 보험 비율이 높은 여성계약자 집단. 본인을 담보로 한 추가계약이 가능한 고객층 |

[그림 4]와 같이 19개 각각의 노드별로, 40세 이하 계약자 비율, 전문직 및 사무 관리직 비율, 월납환산보험료 10만원 이상자 비율, 판매_숙박_유흥업 종사자 비율, 납입 방법이 일시납 비율, 계약자_피보험자 동일 비율, 모집자_수급자 동일 비율, 저축성 보험 및 어린이 보험 등 각 상품 비율 등을 비교분석하였다. 3, 6, 8번 노드는 대상 고객 전체가 계약자와 피보험자가 다른 것으로 나타났고 다른 노드는 계약자와 피보험자가 동일한 비율이 60% 이상이 됨을 알 수 있다. 1, 2, 3, 4, 19 노드는 대상 전체가 남성 계약자이고, 나머지 노드는 여성 계약자이다. 4, 13 노드는 40세 이하 계약자 비율이 0% 임으로 보아 노년층 비율이 높음으로 판단할 수 있다. 노드별 월납환산보험료 10만 원 이상 비율과 저축성 보험 비율 패턴이 유사한 것으로 보아, 보험료가 10만 원 이상이면 저축성 보험 가입 비율이 높은 것으로 판단되며, 2, 4, 13 노드는 월납환산보험료 10만 원 이상 비율과 저축성 보험 비율이 둘 다 높고, 18, 19 노드는 월납환산보험료와 저축성보험 비율 모두 낮은 것으로 나타난다. 또한, 어린이보험 가입비율이 높은 3, 6, 8번 노드가 계약자와 피보험자가 모두 다른 것으로 보아, 3, 6, 8번 노드는 어린이 보험 혹은 기타 보험을 타인을 담보로만 계약하는 계약자의 속성임을 파악할 수 있다.

각각의 노드별 고객을 파악하여, 19개의 노드 중 추가가능성이 높고, 차별화된 특성을 가진 유망노드 10개를 <표 9>와 같이 특징이 유사한 6개의 군집으로 하였다. 각 군집의 특성을 파악할 수 있고, 이를 통해 고객집단의 특성에 맞는 추가 상품 제시가 가능하다.

앞에서와 같이 의사결정나무 결과를 통해서도 추가가입률을 높이기 위해 고객세분화를 위한 유의 변수를 찾아내고, 유망 고객들을 군집화하여 고객 특성을 분석하고 평가하였다. 즉 노드별 추가가입률을 고려한다면, 보장경과기간이 짧고, 소멸건수가 적고 모집자 입사차월이 낮고 계약자와 피보험자가 다르며 낮은 연령대의 여성계약자가 유망한 것으로 나타나며, 유망 고객군별 특성을 기반으로 추가 상품 판매 전략이 가능하다.

IV. 맺음말

연구결과, 로지스틱 회귀분석의 검출력이 더 안정적인 것으로 나타났으나, ROC 곡선이나 Lift 도표로 판단해 볼 때 두 모형의 차이는 크지 않으며, 모두 우수한 모형으로 나타났다.

두 가지 분석 모형을 통해서 도출된 결과를 바탕으로 고객추가가입에 영향을 주는 변수들을 살펴보면 다음과 같다. 로지스틱 회귀분석의 경우 계약자와 피보험자가 다른 낮은 연령의 여성 계약자, 보장경과기간이 짧고 모집자와 수급자가 동일한 경우에 추가가입 확률이 높아지는 것으로 나타났다. 의사결정나무모형은 보장경과기간이 짧고 소멸건수가 적고 모집자 입사차월이 낮으며 계약자와 피보험자가 다르고 낮은 연령의 여성계약자인 경우에 추가가입 확률이 높아지는 것으로 나타났다.

본 연구의 시사점으로는 첫째, 고객 관리 측면에서는, 보장경과기간이 짧고, 소멸건수가 적고, 계약자와 피보험자가 다른, 낮은 연령의 여성 계약자가 추가가입 유망고객임을 알 수 있었고, 신규 고객 관리의 중요성을 실감할 수 있다.

둘째, 모집자 입사차월이 낮을수록 추가가입 가능성이 높아지는 것으로 보아 모집자 코드를 부여받은 지 3년 이하가 되는 모집자의 추가가입확률이 높음을 알 수 있다. 즉 신규 모집자가 우체국 보험 가치 증대에 중요한 역할을 할 것임을 알 수 있다.

본 연구의 한계점으로는 분석에 활용된 변수를 들 수 있다. 특히, 고객 속성 정보 중 직업 변수의 충실도가 낮은 편이어서 변수의 유의성을 판단하기 힘들고, 군집별 속성 파악도 충분치 않다. 최근 2~3년 내에 연락정보 현행화율은 많이 향상된 반면, 분석에 필요한 정보는 아직 부족함을 알 수 있다. 앞으로 시행되는 각종 캠페인 시 고객정보 현행화를 통한 고객 정보 충실도율을 향상시키는 것이 중요하다고 할 수 있다. 고객관계 관리를 위한 모형의 구축은 단순히 일회로 끝나는 작업이 아니라, 지속적으로 이루어져야만 실효성을 거둘 수 있다. 그러나 본 연구에서는 시간상의 제약으로 인해서 예측결과를 실무에 직접 적용하여 실질적인 마케팅 성과나 수익성에 대한 영향을 조사·분석하지 못한 한계점도 있다. 향후 모형의 정확도와 예측력, 활용도 모두를 제고할 수 있는 연구가 계속되어야 할 것이다.

참 고 문 헌

- 박찬욱, 「금융기관의 데이터베이스 마케팅」, (주)시그마인사이트컴, 2002.
- 안철경·조혜원, 「보험회사의 CRM에 관한 연구」, 보험개발원, 2001.
- 이호영, 「생명보험회사 고객관계관리를 위한 고객이탈 예측모형에 관한 연구」, 한국외국어대학교, 2004.
- 정요천, 「교차판매 스코어링 모델 개발에 관한 연구」, 호서대학교, 2002.
- 정홍주, 「보험마케팅의 이해」, 보험일보, 2005.
- 최정환·이유재, 「죽은 CRM 살아있는 CRM」, 한연, 2001.
- 최종후·한상태 외, 「SAS Enterprise Miner를 이용한 데이터마이닝-기능과 사용법」, 자유아카데미, 1999.
- _____, 「SAS Enterprise Miner를 이용한 데이터마이닝-방법론 및 활용」, 자유아카데미, 1999.
- 한국마케팅연구원, 「CRM New Marketing」, 2003.
- SAS Institute, “Getting Started with the Enterprise Miner Software”, 1997.
- 질 디체, 「The CRM Handbook」, 야스미디어, 2003.
- 필립 코틀러, 「마케팅 A to Z」, 세종연구원, 2003.