

An Empirical Study on Telemarketing Business (L Insurance Case)

Yon-Hyong Kim¹⁾ · Seok-Won Lee²⁾

Abstract

The purpose in this datamining modeling is to maximize the number of L insurance' new customer selected from the S corp.'s customers through the telemarketing. We demonstrated the superiority of this method by comparing the existing marketing method and campaign result. The used software in this analysis is SAS 9.1 and so on.

Keywords: 로지스틱회귀분석; 반응율; 생명보험.

I. 서론

보험시장의 변화는 크게 소비자측면과 공급자측면으로 나누어 생각할 수 있다. 소비자측면은 지속적인 정보통신 기술의 발달에 따라 웹, 휴대인터넷, 텔레마케팅, 은행 등 새로운 판매 제도가 등장하게 되었고 이에 따라 보험 상품 정보를 얻을 수 있는 채널이 다변화됨에 따라 고객 선택 폭이 확대된 것이며, 공급자 측면에서의 변화는 다수의 보험사가 시장 진입을 함으로써 공급자간 경쟁이 심화됨에 따라 마케팅 활동이 기존 가입자 증가의 양적 성장추구와 대량판매를 통한 비용절감에서 경쟁력 있는 분야를 선택하여 자원을 집중하고 이를 통해 고객에게 경쟁자와 차별화되는 서비스를 제공하여 가치 창출을 이루려는 것이다. 특히, 보험설계사의 지나친 연고주의 영업과 실적주의 영업은 모집인의 대량도입, 대량탈락을 유발하여 전체 모집조직의 전문성 부족을 가져오는 폐해가 발생됨에 따라 보험회사에서는 고객과 인적대면을 거치지 않고 우편(DM), 전화(TM), 인터넷(CM) 등의 매체를 이용하여 고객, 또는 예상고객에 대하여 상품 정보를 제공하고 직접 보험상품을 판매하는 다양한 마케팅 방법을 활용하고 있다(최인호, 2004).

1) Professor, Department of Public Survey and Applied Statistics, Jeonju University,
560-759, Korea
E-mail : yhkim@jj.ac.kr

2) Corresponding Author : Adjunct Professor, Department of Public Survey and Applied
Statistics, Jeonju University, 560-759, Korea
E-mail : leeseokwon@hanmail.net

그러나 신규 판매 채널의 효율성은 직접비, 광고비 등에서 전통적인 보험설계사 보다 우수함에도 불구하고 자필서명을 전자인증으로 대체하는 방법 등 제도적 미비점이 존재하며, 보험회사들의 낮은 동기 부여로 텔레마케팅을 제외한 나머지 채널에 대해서는 실적이 높지 않은 상태이다(생명보험협회, 2001).

이는 텔레마케팅에서 보험사들이 지금까지 경쟁력 있는 상품 개발, 우수한 텔레마케터 조직 운영, 양질의 DB 확보에 노력을 기울인 결과 일정 수준의 보험가입 수율을 유지하고 있기 때문이다. 그렇지만 수년간에 걸쳐 개선이 진행된 영역임에도 불구하고 업무 표준화가 쉽지 않고 단기간 내에 개선이 어려운 영역이므로 신규 진입을 원하는 보험사나 기존 보험사들도 비용대비 개선의 폭이 낮은 분야이다.

그러나 확보된 DB의 효율적 활용을 통한 보험가입 수율 제고 영역은 아직도 미개척 분야로 가망고객 선별 모델링 기법이 요구되며 단기간에 효과 검증이 가능하고, 보험에 가망고객만을 대상으로 전화를 진행함에 따라 고객 만족도를 증가시킬 수 있는 장점이 있다(Youg Jun, Cho., et al., 2006).

본 논문에서는 텔레마케팅에서 L생명보험사 가입 고객 극대화를 위하여 DB제휴사인 S주식회사 고객들을 대상으로 데이터마이닝 모델링을 진행하였으며, 기존 마케팅 방법과 캠페인 결과를 비교함으로써 그 우수성을 증명하고자 한다. 분석에 사용된 소프트웨어는 SAS 9.1 등이다.

II. 제휴사 사전분포

생명보험사 DB는 가입고객만을 관리함에 따라 텔레마케팅을 이용한 보험가입 유치를 위해서는 다량의 DB를 보유한 통신사, 쇼핑몰, 정유사, 카드사 등의 전략적 제휴를 통한 제휴사 확보로 이루어진다.

L생명보험사의 경우 6개 제휴사를 통해 신규 가입자를 늘려가고 있다. 제휴사 중 특히 S사의 경우 국내의 온·오프라인 제휴 가맹점을 두고 운영하고 있으며, 고객은 제휴사에서 사용한 일정한 금액에 따라 포인트를 적립받기도 하고, 포인트를 차감하기도 한다. 가맹점에서는 고객이 사용한 금액에 따라 포인트 점수를 부여하고 자사 사이트에 포인트를 통합하여 일정점수에 도달하면 해당 가맹점에서 점수만큼 상품을 무료로 구입하거나 다양한 제휴사에서 서비스를 이용할 수 있다. 그리고 5,000점 이상이면 현금처럼 사용이 가능하고 50,000점 이상이면 현금으로 돌려받는 서비스를 제공함으로써 마일리지 적립을 바탕으로 고객 사용자를 확보하는 데 성공하였다. 결국 L생명보험사는 S주식회사의 고객 중 통신판매의 동의를 얻어 텔레마케팅을 진행한 후 S주식회사에 일정한 수수료를 지불하는 수익구조를 담고 있다(이석원, 2008).

텔레마케팅을 통하여 판매되는 L생명보험사 상품은 연령을 기준으로 24세에서 34세까지의 가족사랑보험, 35세에서 39세까지의 상해보험, 40세에서 49세까지의 암보험, 50세 이상의 실버보험으로 나눌 수 있으며 보험설계사 자격이 있는 텔레마케터가 영업을 직접 진행하는 원콜과 1차에서 고객 접점을 확인하고 보험설계사 자격이 있는 텔레마케터에 넘긴 후 영업을 진행하는 듀얼콜로 세분화된다. 본 논문에서는 원콜에서 암보험에 대한 모델링을 진행하기로 하며 2005년 11월에서 2006년 2월까지의 모델 설정을 위한 분석용 데이터로 사용하고 2006년 3월에서 2006년 5월까지의 검증데이터로 이용하였다. 이 때 S주식회사의 전체 고객 분포는 <표 2-1>과 같다.

성별에서는 남성이 62.7%로 여성의 36.3% 보다 높게 나타나고 있으며, 결혼여부에서는 기혼의 비중이 57.31%로 미혼 42.69% 보다 높다. 차량 보유 여부에서는 보유자가 55.16%로 구성되고 있다. 연령은 30대 38.43%, 20대 29.70%로 S주식회사의 전체 고객 중 68.13%로 비중이 상당히 높게 나타나고 있다. 지역은 서울 24.58%, 경기 24.11%로 나타나고 있어 서울, 경기 수도권 지역이 전체에서 약 50%를 차지하는 것으로 나타나고 있다.

<표 2-1> S주식회사 고객 분포

| 구분 | 구성비(%) | | | | | |
|--------|--------|-------|-------|-------|------|------|
| | 성별 | 남성 | | | 여성 | |
| 62.7 | | | 36.3 | | | |
| 결혼여부 | 미혼 | | | 기혼 | | |
| | 42.69 | | | 57.31 | | |
| 차량보유여부 | 보유 | | | 미보유 | | |
| | 55.16 | | | 44.84 | | |
| 연령대 | 10대 | 20대 | 30대 | 40대 | 50대 | 60이상 |
| | 0.89 | 29.70 | 38.43 | 19.98 | 7.87 | 2.13 |
| 지역 | 서울 | 부산 | 대구 | 인천 | 광주 | 대전 |
| | 24.58 | 7.54 | 6.02 | 4.79 | 3.01 | 2.90 |
| | 울산 | 경기 | 강원 | 충북 | 충남 | 전북 |
| | 3.29 | 24.11 | 2.55 | 2.22 | 2.88 | 2.78 |
| | 전남 | 경북 | 경남 | 제주 | | |
| | 2.52 | 5.00 | 4.77 | 1.04 | | |

암 상품의 연령대인 40세에서 49까지의 분포는 <표2-2>와 같다. 성별에서는 남성이 55.12%로 여성의 44.88% 보다 높게 나타나고 있으나 전체 분포와 비교하면 남성 비중이 약7%포인트 정도 줄어든 수치이다.

<표 2-2> S주식회사 40~49세 고객 분포

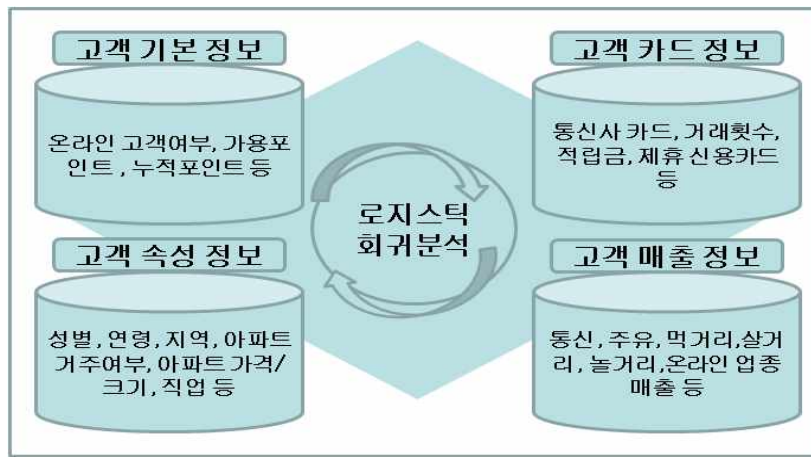
| 구분 | 구성비(%) | | | | | |
|--------|--------|-------|-------|-------|------|------|
| | 성별 | 남성 | | | 여성 | |
| 55.12 | | | 44.88 | | | |
| 결혼여부 | 미혼 | | | 기혼 | | |
| | 19.35 | | | 80.65 | | |
| 차량보유여부 | 보유 | | | 미보유 | | |
| | 84.89 | | | 15.11 | | |
| 지역 | 서울 | 부산 | 대구 | 인천 | 광주 | 대전 |
| | 25.56 | 5.99 | 5.65 | 6.51 | 3.26 | 3.44 |
| | 울산 | 경기 | 강원 | 충북 | 충남 | 전북 |
| | 2.27 | 20.34 | 2.57 | 2.65 | 3.55 | 3.13 |
| | 전남 | 경북 | 경남 | 제주 | | |
| | 2.69 | 5.02 | 5.6 | 1.77 | | |

결혼여부에서는 기혼의 비중이 80.65%로 미혼 19.35% 보다 상당히 높게 나타나고 있다. 이는 40대 경우 기혼의 비율이 높기 때문으로 사료된다. 차량보유여부에서는 보유자가 84.89%로 전체의 55.16%보다 약30%포인트 정도 더 높게 나타나고 있다. 이러한 결과는 연령대 특성과 무관하지 않다. 지역에서는 서울 25.56%, 경기 20.34%로 나타나고 있다. 이는 95% 신뢰수준내에서 카이제곱 검증결과 서울은 높고, 경기 지역은 약간 낮은 수치로 파악되고 있으며, 전체 고객 분포와 마찬가지로 서울, 경기의 분포가 집중되어 있다.

Ⅲ. 로지스틱회귀분석을 이용한 모델 설정

본 논문에서는 텔레마케팅에서 L생명보험사 보험수율을 증가시키기 위하여 데이터 마이닝 기법 중 로지스틱회귀모형을 사용하여 모델을 설정하였다. 로지스틱회귀모형은 모델 설정 후 캠페인 진행 시 선형식으로 계산되어 고객을 추출할 수 있기 때문에 기업의 데이터베이스에서 쉽게 사용이 가능하며, 의사결정나무, 신경망 방법 등과 대동소이하게 결과 값을 얻을 수 있기 때문이다.

S주식회사의 경우 방대한 자료에서 주로 사용되어지고 있는 100여개 변수를 1차로 선택하였다. 100여개 변수는 크게 4가지 형태로 구분할 수 있으며 <그림 3-1>과 같다.



<그림 3-1> 로지스틱회귀분석에 사용된 변수

고객 기본 정보에는 일반적인 S주식회사의 온라인 고객여부, 가용 포인트, 누적 포인트 등을 포함하고 고객 속성 정보에는 성별, 연령, 지역, 결혼여부 등 인구통계학적 정보와 아파트 거주여부, 아파트 가격, 아파트 크기, 차량 보유여부, 캐쉬백 등급 등이 포함된다. 그리고 고객 카드 정보에는 통신사 카드 매출, 거래 횟수, 적립금 등이 포함되며 제휴 카드 매출(EBC 카드), 거래 횟수, 업종별 상세 내용을 이용한다. 마지막으로 고객 매출 정보에서는 주유, 온라인 가입 개월 수, 온라인 구매 형태, S주식회사 기준에 의한 5가지 분류 형태의 업종 거래 횟수, 매출 등을 이용하였다.

S주식회사의 기준에 의한 5가지 분류 형태는 자사 온·오프라인 제휴사를 기준으로 나누는 것으로 <표 3-1>과 같다.

<표 3-1> S사 기준에 의한 온·오프라인 제휴사 분류

| 업종 분류 | 업종 |
|--------|--|
| 먹거리 | 주점, 다과점, 음식점 |
| 살거리 | 예술품/선물용품, 식료품, 신변잡화, 의료직물, 서적/문구/인쇄/복사, 의약품, 통신판매, 컴퓨터 사무용기기, 차량용연료/기타연료, 인테리어/건축자재, 자동차/오토바이, 오락/취미, 가전용품/통신기기, 유통점 등 |
| 생활/건강 | 교육기관/교습학원, 광고/이벤트, 개인간병/유사서비스, 의료시설, 여행사, 수리, 관광/숙박시설, 운송업(여객/화물), 법무/회계관리, 이미지/욕탕/몸매관리, 예식준비/예식관련 등 |
| 놀/유희거리 | 공연/전시, 오락 시설, 운동/여가시설, 영화/비디오, 후원지/놀이공원, 놀거리기타 등 |
| 온라인 | 전자상거래, 포인트판매 등 |

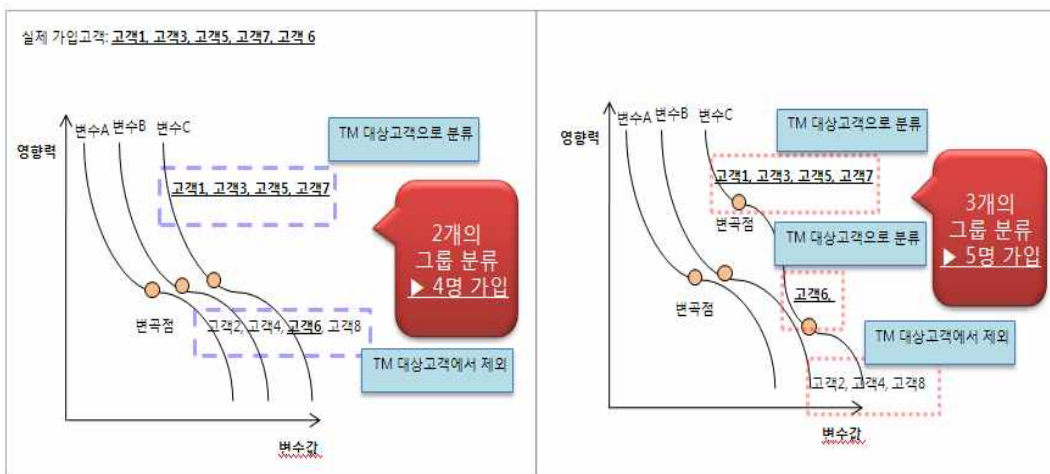
그러나 S주식회사에서 1차로 추출한 100여개의 변수들 중 모델 설정에 영향을 주는 변수를 찾기란 쉽지 않다. 그래서 최종 모델을 설정한 후 변수의 유의한 변수가 누락되거나 누락되었다 하더라도 이유를 해석하기 위한 보완 방법으로 먼저 요인분석과 군집분석을 통해 개략적인 변수를 파악하고자 한다.

<표 3-2> 요인분석, 군집분석 결과

| 변수명 | 요인분석(설명력) | 군집분석(설명력) |
|-------------------|-----------|-----------|
| 3개월평균주유매출금액 | 0.9455 | |
| 6개월평균주유매출금액 | | 0.9101 |
| 3개월평균이마트매출금액 | 0.95678 | 0.97 |
| 6개월평균이마트매출금액 | | 0.8276 |
| 6개월평균통신매출금액 | 0.895 | |
| 6개월평균통신매출횟수 | | 0.8426 |
| 6개월평균살거리업종매출횟수 | 0.939 | |
| 온라인이용개월수 | 0.6738 | |
| 6개월평균신용카드매출횟수 | 0.9023 | 0.798 |
| 3개월평균먹거리업종매출횟수 | 0.9253 | 0.8084 |
| 6개월평균놀/유희거리업종매출금액 | 0.8835 | 0.7501 |
| 6개월이용주유소개수 | 0.6532 | 0.8784 |
| 캐쉬백전체사용개월수 | 0.895 | 0.7562 |
| 누적포인트 | 0.6895 | |
| 3개월평균온라인업종매출금액 | 0.9319 | 0.9233 |
| 6개월평균생활/건강업종매출금액 | 0.8776 | 0.8666 |
| 3개월평균온라인업종매출횟수 | 0.9588 | |
| 6개월평균온라인업종매출횟수 | | 0.9389 |
| 쇼핑몰사용개월수 | | 0.6952 |

요인분석과 군집분석을 통해 공통으로 추출된 변수는 3개월 평균 이마트매출금액, 6개월평균신용카드매출횟수, 3개월평균먹거리업종매출횟수, 6개월평균놀/실거리업종매출금액, 6개월평균놀/실거리업종매출금액, 6개월이용주유소개수, 캐쉬백전체사용개월수, 3개월평균온라인업종매출금액, 6개월평균생활/건강업종매출금액이며 이들의 변수가 중요변수로 파악되고 있다.

요인분석과 군집분석을 통해 개략적인 변수를 파악하였다면 선택된 변수와 S사 마케팅 경험에서 얻어진 주요 변수를 포함하여 변수내 속성을 정의한다. 이 때 선택된 변수의 영향력을 함께 반영함으로써 캠페인 진행 시 고객의 누락방지 뿐만 아니라 변별력을 동반하여 보험수율을 향상시킬 수 있다.



<그림 3-2> 로지스틱회귀분석에 사용된 변수

<그림 3-2>에서 실제 캠페인 결과 고객1, 고객3, 고객5, 고객6, 고객7이 보험가입자로 확인 되었으나, 그림 좌측은 변수내 집단을 2개의 그룹으로 설정하여 상단의 고객1, 고객3, 고객5, 고객7만을 선택하여 캠페인을 진행한 결과 고객6 가입자를 포함시키지 못했다. 하지만 그림 우측의 경우 변수 내 집단을 3개의 그룹으로 설정함으로써 고객6을 포함한 5명의 보험가입자를 유도할 수 있었다.

그러므로 모델 설정 시 변수내 변곡점을 어떻게 설정하느냐에 따라 최종 캠페인 진행 후 보험가입자 결과 값이 달라질 수 있다.

<표 3-3> 변수내 변곡점 설정

| 변수명 | 구분 | Estimate | 변수명 | 구분 | Estimate |
|-----------|-------------------|----------|-----------|-------------------|----------|
| 아파트 가격 | Missing | 0.0169 | 가용 포인트 | 935.1미만 | -0.0188 |
| | 0 | 0 | | 935.1이상,5,194미만 | -0.0126 |
| | 0초과,7,118미만 | 0.465 | | 5,194이상,10,000이상 | -0.0086 |
| | 7,118이상,13,563미만 | 0.5329 | | 10,000이상,17,490미만 | -0.008 |
| | 13,563이상,35,224미만 | 0.0016 | | 17,490이상,35,066미만 | -0.0069 |
| | 35,224이상 | 0 | | 35,066이상 | 0 |

S주식회사의 경우 <표 3-3>과 같이 아파트 가격과 가용포인트를 변곡점에 따른 범주화를 실시하였으며 그 방법론은 지니계수를 사용하였다.

이를 바탕으로 로지스틱회귀모형을 구축하며 확률 산출식은 다음과 같이 표현할 수 있다.

$$P = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}$$

P : 텔레마케팅시 고객이 보험상품에 가입할 확률, X_k : 보험상품가입 확률에 대한 영향 변수, β_k : 영향 변수의 상관성 및 방향성 계수(값이 클수록 상관성이 크며, (+)이면 양의 방향으로, (-)이면 음의 방향의 상관성을 가짐)

로지스틱회귀모형에서 목적성 변수는 보험가입여부(Target)로 보험가입자는 1, 비가입자는 0의 값을 가지며, 설명변수 중 범주형 변수와 연속형 변수 중 구간을 나누는 변수들로 더미 변수를 생성하였다(김연형외, 2007, Fu, 2004).

더미 변수는 SAS프로그램에서 E-miner가 아닌 SAS 프로시저를 사용할 때 이용되며 더미를 취한 변수는 지역, S주식회사 등급, 직업, 연령대, 차종 등이며 그 구성은 <표 3-4>와 같다.

<표 3-4> 변수의 더미화

| 지역 | 더미화 | 등급 | 더미화 | 직업 | 더미화 |
|----|---------|---------|---------|------|-----------|
| 서울 | haddr01 | 최우수 | grade01 | 전문직 | jobt_cd01 |
| 부산 | haddr02 | 우수 | grade02 | 공무원 | jobt_cd02 |
| 대구 | haddr03 | 일반 | grade03 | 회사원 | jobt_cd03 |
| 인천 | haddr04 | 신규 | grade04 | 자영업 | jobt_cd04 |
| 광주 | haddr05 | 단기휴먼 | grade05 | 주부 | jobt_cd05 |
| 대전 | haddr06 | 장기휴먼 | grade06 | 학생 | jobt_cd06 |
| 울산 | haddr07 | 이탈 | grade07 | 직종기타 | jobt_cd07 |
| 경기 | haddr08 | 잠재 | grade08 | 미상 | jobt_cd08 |
| 강원 | haddr09 | | | | |
| 충북 | haddr10 | 연령대 | 더미화 | 차종 | 더미화 |
| 충남 | haddr11 | 40세~43세 | age01 | 대형 | Car01 |
| 전북 | haddr12 | 44세~45세 | age02 | 소형 | Car02 |
| 전남 | haddr13 | 45세~46세 | age03 | 경차 | Car03 |
| 경북 | haddr14 | 47세~49세 | age04 | RV | Car04 |
| 경남 | haddr15 | | | 생계형 | Car05 |
| 제주 | haddr16 | | | 기타 | Car06 |
| | | | | 외제차 | Car07 |

종속변수가 보험가입여부로 1과 0값만 취하는 이항 변수이므로 로지스틱회귀모형을 자료에 적합 시키기 위해서 분석용 데이터에 PROC LOGISTIC 문을 사용하였다. 로지스틱모형의 구성은 “종속변수(Target) = 독립변수들”이며 옵션은 내림차순으로 중

속변수가 1인 확률모형을 추정하기 위한 것이다. 설정된 모형에 포함된 많은 변수들 중에서 유의한 변수를 선택하기 위해서 단계별 선택법을 사용하였고 Slentry와 Slstay 옵션은 변수 추가 또는 제거 시 유의수준을 정하는 것으로 <그림 3-3>에서는 0.1로 정하였다. 단계별 선택법은 각 단계마다 변수를 모형에 추가적으로 포함시키거나 제외시키기 위해서 전진 선택법과 후진 선택법을 혼합하여 사용하는 방법으로 각 단계마다 모형에 추가 혹은 제외될 변수, 각 변수에 대한 카이제곱 검정통계량 값과 대응되는 P 값을 보여준다(이성우, 2005, Hosmer, 2000).

```
proc logistic data=Lee.association descending ;
model target =
/* target=1이 가입이고, target=0이 비 가입이면, descending 옵션을 달아준다.
*/
가용포인트, 3개월평균먹거리업종매출횟수, 연령, 아파트사이즈, 3개월평균이마트매출금액, 6개월평균신용카드매출횟수, /* 중간 생략 */ 6개월이용주유소개수,
캐쉬백전체사용개월수, 3개월평균온라인업종매출금액
/*범주형 변수의 더미화*/
지역01 지역02 직업01 직업 성별1 결혼여부1 연령01 연령02
/selection=stepwise slentry=0.1 slstay=0.1 STB
/*slentry=0.1 : 변수 추가시 유의수준을 0.1로 한다. */
/*slstay=0.1 : 변수 제거시 유의수준을 0.1로 한다. */
/* STB : 표준화 회귀 계수 */
/* STB 값의 절대값이 1보다 크면 strongest effect variable 이다. */
;
output out=result.Association_rslt p=phat; run;
proc sort data=result.Association_rslt; by descending phat; run;
/* 검정 Validation */
data d_fib.result_fib3_vali04; set d_fib.fib_validation04 ;
zz= -3.9088
+3개월주유소이용회수*(0.0693) +3개월평균통신사용금액*(0.000003679)
+신용카드 사용*(-0.1934)
/*중간 생략 */
+서울*(-0.3983)+아파트 거주*(-0.966) ;
zzz=exp(zz);phat=zzz/(1+zzz);run ;
proc sort data=d_fib.result_fib3_vali04;by descending phat;run ;
```

<그림 3-3> 로지스틱회귀분석을 위한 SAS프로그램

<표 3-5> 로지스틱회귀분석 단계별 선택법 변수 선택 요약

| Paramete | DF | Estimate | Standard Error | Wald Chi-Square | Pr>ChiSq |
|-------------------|----|-------------|----------------|-----------------|----------|
| Intercept | 1 | -3.9088 | 0.0909 | 1850.5004 | <.0001 |
| 3개월주유소이용회수 | 1 | 0.0693 | 0.0145 | 22.9507 | <.0001 |
| 3개월평균통신매출금액 | 1 | 0.000003679 | 1.55E-06 | 5.66 | 0.0174 |
| 6개월평균통신매출횟수 | 1 | -0.7527 | 0.2242 | 11.2712 | 0.0008 |
| 3개월평균생활/건강업종매출금액 | 1 | 0.000003143 | 1.18E-06 | 7.0839 | 0.0078 |
| 지역(서울) | 1 | -0.3983 | 0.1297 | 9.4298 | 0.0021 |
| 직업04(자영업) | 1 | -0.4854 | 0.2572 | 3.5603 | 0.0592 |
| 차종06(기타) | 1 | 0.3662 | 0.1518 | 5.8218 | 0.0158 |
| 등급03(일반) | 1 | -0.3166 | 0.135 | 5.5016 | 0.019 |
| 등급04(신규) | 1 | 0.6205 | 0.3173 | 3.8252 | 0.0505 |
| 아파트거주여부1 | 1 | -0.966 | 0.1397 | 47.7974 | <.0001 |
| 신용카드보유여부1 | 1 | -0.3983 | 0.1128 | 2.9401 | 0.0864 |
| 6개월통신사용여부1 | 1 | 0.3203 | 0.1673 | 3.6648 | 0.0556 |
| 6개월평균먹거리업종구매여부1 | 1 | -0.3933 | 0.1639 | 5.7561 | 0.0164 |
| 6개월평균살거리업종구매여부1 | 1 | -0.1764 | 0.0969 | 3.3106 | 0.0688 |
| 아파트사이즈03(18평이하) | 1 | 0.66 | 0.2568 | 6.6025 | 0.0102 |
| 아파트사이즈04(18~24평) | 1 | 0.2667 | 0.1567 | 2.8969 | 0.0888 |
| 아파트가격03(7,118미만) | 1 | 0.465 | 0.1978 | 5.528 | 0.0187 |
| 아파트가격04(13,563미만) | 1 | 0.5329 | 0.1667 | 10.2162 | 0.0014 |

<표 3-5>의 SAS 결과값을 기초로 추정된 회귀계수를 이용하여 구매 고객이 될 확률이 큰 고객을 다음과 같이 예측하였다.

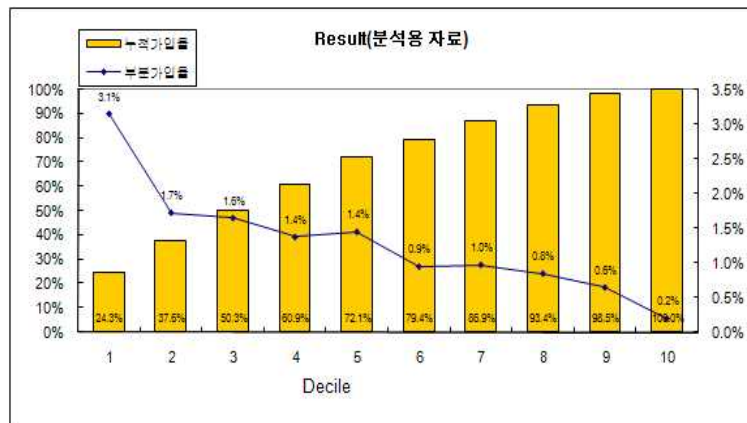
$$\hat{Y} = -3.9088 + (0.0693 * 3\text{개월주유소이용회수}) + (-0.3983 * \text{서울}) + (-0.0119 * \text{아파트사이즈}) + (0.3662 * \text{차종06}) + (-0.3933 * 6\text{개월평균먹거리업종구매여부1}), \dots, + (0.5329 * \text{아파트가격04})$$

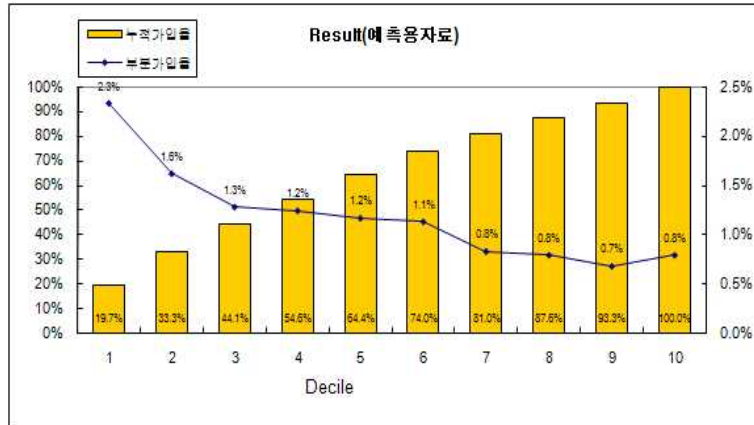
어떤 사건이 발생할 확률이 발생하지 않을 확률에서 차지하는 정도를 오즈라 하며 오즈비 추정 값과 그 한계 값을 표시하면 <표 3-6>과 같다.

<표 3-6> 오즈비 추정값

| Effect | Point Estimate | 95% Wald Confidence Limits | |
|-------------------|----------------|----------------------------|-------|
| 3개월주유소이용회수 | 1.072 | 1.042 | 1.103 |
| 3개월평균통신매출금액 | 1 | 1 | 1 |
| 6개월평균통신매출횟수 | 0.471 | 0.304 | 0.731 |
| 3개월평균생활/건강업종매출금액 | 1 | 1 | 1 |
| 지역(서울) | 0.671 | 0.521 | 0.866 |
| 직업04(자영업) | 0.615 | 0.372 | 1.019 |
| 차종06(기타) | 1.442 | 1.071 | 1.942 |
| 등급03(일반) | 0.729 | 0.559 | 0.949 |
| 등급04(신규) | 1.86 | 0.999 | 3.464 |
| 아파트거주여부1 | 0.381 | 0.289 | 0.501 |
| 신용카드보유여부1 | 0.824 | 0.661 | 1.028 |
| 6개월통신사용여부1 | 1.378 | 0.992 | 1.912 |
| 6개월평균먹거리업종구매여부1 | 0.675 | 0.489 | 0.931 |
| 6개월평균살거리업종구매여부1 | 0.838 | 0.693 | 1.014 |
| 아파트사이즈03(18평이하) | 1.935 | 1.169 | 3.201 |
| 아파트사이즈04(18~24평) | 1.306 | 0.96 | 1.775 |
| 아파트가격03(7,118미만) | 1.592 | 1.08 | 2.346 |
| 아파트가격04(13,563미만) | 1.704 | 1.229 | 2.363 |

단계별 선택법을 이용한 변수 선택 시 모형 검증 및 평가는 분석용 자료에서 시행한 결과에서 선택된 독립변수들과 추정된 회귀계수를 예측용 자료에 적용하여 모형 성과를 평가하게 되는데, 추정된 회귀모형을 예측용 자료에 적용하여 \hat{P} 값을 기준으로 내림차순으로 정렬한다. \hat{P} 값이 큰 순서대로 상위 10%씩 구간을 나누어 구간별 구매 고객의 비율을 비교하고 추정된 회귀모형의 성과를 평가한다. 여기서 분석용 데이터를 분석한 결과 상위 10%의 구간에서 구매율은 3.1%를 나타내고 있으며 이 때 전체 구매자 중 24.3%를 차지하는 것으로 파악된다.





<그림 3-4> 로지스틱회귀분석 - 분석용, 예측용 자료 적합

분석용 자료와 예측용 자료에서의 모형의 성과 차이는 거의 나타나지 않으며 추정된 로지스틱회귀모형에 따르면 상위 50%에서 구매고객의 누적비율이 전체 대비 약 64%~72%를 차지하고 있음을 알 수 있다.

IV. 캠페인 실행을 통한 모델효율성 검증

텔레마케팅에서 보험 가입을 위한 캠페인 진행시 상담원들은 표준화된 스크립트를 기본으로 영업에 임하게 된다. 그래서 고객입장에서 전화를 받아보면 동일한 영업흐름이 느껴지게 된다. S사의 경우 고객과 처음 통화가 이루어 진후 <그림 4-1>과 같은 방법으로营业을 진행하고 있다. 이 때 보험영업에서 고객과 가장 친근하게 다가갈 수 있는 방법이 S사의 가용포인트다. 포인트의 유무와 현재 가지고 있는 포인트를 효과적으로 사용할 수 있다는 메시지를 담아 일면식이 없는 고객과 대화에서 거부감을 최소화 시키고 있다.

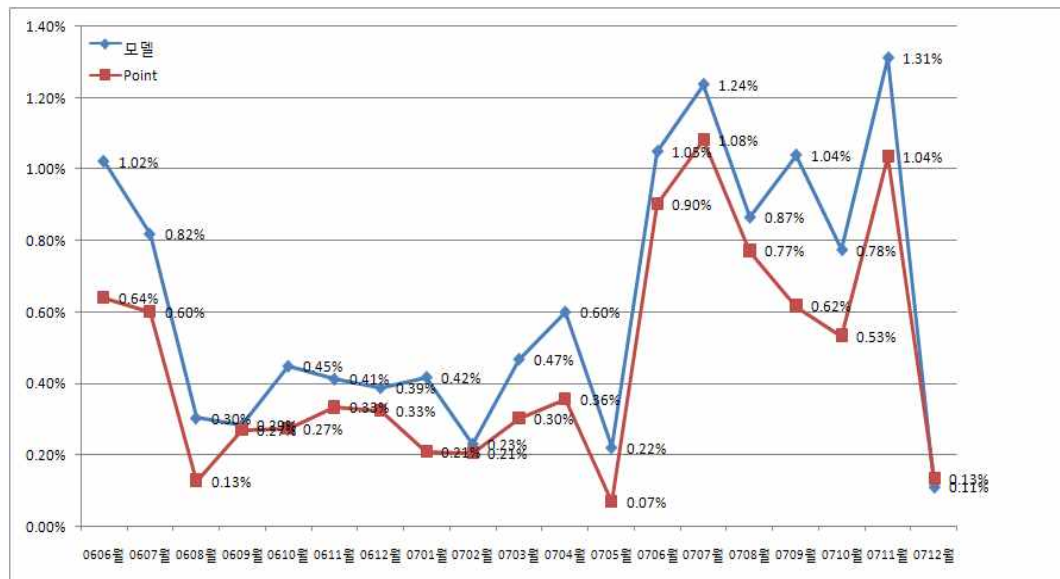
[첫인사] 안녕하세요? 000 상담원 000입니다.
 [수신자 확인] 000고객님 맞으십니까? (000 이용해주셔서 감사합니다.^^)
 [통화가능여부확인 및 감사표현] 캐쉬백포인트와 서비스 안내차 연락드렸는데 잠시 통화 가능하십니까?
 (고객:NO) 바쁘시면 용건만 간단히 말씀드리는 건 괜찮으시죠?
 (고객:YES) 소중한 시간 내주셔서 감사합니다.
 [포인트 안내 및 목적안내]
 현재 적립되어 있는 포인트는 00점인데 5천점이상부터는 가맹점에서 현금처럼 이용 가능하고 5만점 이상부터 현금으로 환급받을 수 있으니 많은 이용 부탁드립니다
 요...앞으로도 저희 000 꾸준히 이용해 달라는 의미로 000까지 준비 했습니다
 (000->상품명, 상해, 암 등)

<그림 4-1> 상담원 표준 스크립트

본 논문에서는 로지스틱회귀모형을 통하여 구축한 모델과 상담원들이 선호하고 실질적으로 S사의 영업방식인 가용포인트가 우수한 고객을 뽑아 두 집단을 상호 비교하였으며 대상 기간은 2006년 6월부터 2007년 12월까지 19개월간이다.

이 때 반응은 보험가입자수와 가입율을 기준으로 적용하였다. 매월 추출하는 모수 및 추출비율은 달랐으나 각 그룹별 효과를 통하여 모델의 효과성을 입증하고자 하였다.

데이터 추출은 가용포인트를 우선으로 고려하는 집단에서 해당월의 총 모수 중 10%를 추출하고 로지스틱회귀모형에서 도출된 모델을 통해 90%를 추출하여 평가가 이루어졌으며, 가용포인트를 우선으로 고려하는 집단에서 10% 추출한 이유는 비교대상 점점과 모델의 노후 여부를 판단하기 위해서다.



<그림 4-2> 로지스틱회귀모형과 가용포인트를 우선으로 고려하는 집단과의 비교분석

<그림 4-2> 결과처럼 2006년 6월의 경우 모델이 1.02%로 가용포인트를 우선으로 고려하는 집단의 0.64%보다 약 0.38%포인트 더 높았으며 19개월 평균으로는 0.17%포인트 더 효과가 있는 것으로 나타났다.

<표 4-1> 로지스틱회귀모형 결과

| 추출일자 | 총대상고객 | Random sample | 추출고객수 | 상위% | 배정 | 가입자 | 가입자 비율 | 평가 | |
|------------|---------|---------------|---------|---------|--------|---------|--------|-------|------|
| 2006-06-22 | 201,754 | 모델(9) | 181,578 | 36,315 | 20.0% | 24,046 | 246 | 1.02% | 60% |
| | | 포인트(1) | 20,175 | 4,035 | 20.0% | 2,965 | 19 | 0.64% | |
| 2006-07-25 | 220,188 | 모델(9) | 198,169 | 45,000 | 22.7% | 44,928 | 368 | 0.82% | 36% |
| | | 포인트(1) | 22,019 | 5,000 | 22.7% | 4,989 | 30 | 0.60% | |
| 2006-08-22 | 222,259 | 모델(9) | 200,033 | 56,700 | 28.3% | 26,640 | 81 | 0.30% | 137% |
| | | 포인트(1) | 22,226 | 6,300 | 28.3% | 6,232 | 8 | 0.13% | |
| 2006-09-20 | 246,115 | 모델(9) | 221,504 | 66,451 | 30.0% | 66,306 | 189 | 0.29% | 5% |
| | | 포인트(1) | 24,611 | 7,383 | 30.0% | 7,353 | 20 | 0.27% | |
| 2006-10-26 | 292,558 | 모델(9) | 263,302 | 52,661 | 20.0% | 52,517 | 236 | 0.45% | 64% |
| | | 포인트(1) | 29,256 | 5,851 | 20.0% | 5,835 | 16 | 0.27% | |
| 2006-11-21 | 285,026 | 모델(9) | 256,523 | 102,609 | 40.0% | 102,382 | 423 | 0.41% | 24% |
| | | 포인트(1) | 28,503 | 11,401 | 40.0% | 11,371 | 38 | 0.33% | |
| 2006-12-20 | 254,288 | 모델(9) | 228,859 | 91,544 | 40.0% | 91,302 | 355 | 0.39% | 19% |
| | | 포인트(1) | 25,429 | 10,171 | 40.0% | 10,129 | 33 | 0.33% | |
| 2007-01-17 | 264,808 | 모델(9) | 238,327 | 42,873 | 18.0% | 42,784 | 179 | 0.42% | 99% |
| | | 포인트(1) | 26,481 | 34,764 | 131.3% | 34,671 | 73 | 0.21% | |
| 2007-02-15 | 258,938 | 모델(9) | 233,044 | 57,648 | 24.7% | 55,701 | 129 | 0.23% | 12% |
| | | 포인트(1) | 25,894 | 6,405 | 24.7% | 6,283 | 13 | 0.21% | |
| 2007-03-15 | 274,939 | 모델(9) | 247,445 | 51,770 | 20.9% | 49,466 | 232 | 0.47% | 55% |
| | | 포인트(1) | 27,494 | 5,753 | 20.9% | 5,601 | 17 | 0.30% | |
| 2007-04-18 | 285,806 | 모델(9) | 257,225 | 25,722 | 10.0% | 24,956 | 150 | 0.60% | 68% |
| | | 포인트(1) | 28,581 | 2,858 | 10.0% | 2,794 | 10 | 0.36% | |
| 2007-05-17 | 312,848 | 모델(9) | 281,563 | 26,354 | 9.4% | 25,736 | 57 | 0.22% | 216% |
| | | 포인트(1) | 31,285 | 2,928 | 9.4% | 2,857 | 2 | 0.07% | |
| 2007-06-19 | 329,219 | 모델(9) | 296,297 | 29,638 | 10.0% | 28,074 | 295 | 1.05% | 16% |
| | | 포인트(1) | 32,922 | 3,293 | 10.0% | 3,207 | 29 | 0.90% | |
| 2007-07-18 | 373,390 | 모델(9) | 336,051 | 22,500 | 6.7% | 22,448 | 278 | 1.24% | 14% |
| | | 포인트(1) | 37,339 | 2,500 | 6.7% | 2,493 | 27 | 1.08% | |
| 2007-08-17 | 428,829 | 모델(9) | 385,946 | 35,100 | 9.1% | 34,973 | 303 | 0.87% | 12% |
| | | 포인트(1) | 42,883 | 3,900 | 9.1% | 3,885 | 30 | 0.77% | |
| 2007-09-19 | 455,387 | 모델(9) | 409,848 | 40,984 | 10.0% | 34,973 | 364 | 1.04% | 68% |
| | | 포인트(1) | 45,539 | 4,554 | 10.0% | 4,533 | 28 | 0.62% | |
| 2007-10-23 | 488,196 | 모델(9) | 439,376 | 43,937 | 10.0% | 43,724 | 339 | 0.78% | 45% |
| | | 포인트(1) | 48,820 | 4,882 | 10.0% | 4,861 | 26 | 0.53% | |
| 2007-11-20 | 486,009 | 모델(9) | 437,408 | 9,000 | 2.1% | 4,114 | 54 | 1.31% | 27% |
| | | 포인트(1) | 48,601 | 1,000 | 2.1% | 965 | 10 | 1.04% | |
| 2007-12-14 | 486,009 | 모델(9) | 514,581 | 13,500 | 2.6% | 13,500 | 15 | 0.11% | -17% |
| | | 포인트(1) | 57,176 | 1,500 | 2.6% | 1,500 | 2 | 0.13% | |

<표 4-1> 결과를 살펴보면 2006년 6월의 경우 추출일자는 22일이었으며, 총 대상 고객은 201,754명이었다. 캠페인의 공정성을 기하기 위하여 가용포인트를 우선으로 고려하는 집단 10%와 모델집단 90%를 임의추출(Random Sampling)방식으로 분배하였으며 각각의 방식으로 추출된 모수에서 상위20%를 뽑아 캠페인을 진행하였다. 모델에서 최종 추출한 고객수는 36,315명이었고, 가용포인트를 우선으로 고려하는 집단에서 추출한 인원은 4,035명이었다. 이 인원을 캠페인에 적용하였으며, CTI(Computer

Telephony Integration)에 배정된 인원은 모델이 24,046명, 가용포인트를 우선으로 고려하는 집단인원은 2,965명이었다.

텔레마케팅을 통한 보험가입자 수는 모델은 246명, 가용포인트를 우선으로 고려하는 집단은 19명으로 모델이 약60%의 개선효과를 보이는 것으로 나타났다.

V. 결론 및 향후 과제

텔레마케팅에서 데이터마이닝의 로지스틱회귀모형을 이용하여 캠페인을 진행한 결과 정량적인 효과에서는 직접적인 보험가입자 증대와 이에 따른 거수보험료도 함께 증가되는 것으로 나타났다. 또한 정성적 효과로는 가망고객 대상 추출로 인한 불필요한 고객 접촉이 줄어들어 따라 고객 클레임 감소, 텔레마케팅에서 보험 수율 증가에 따른 텔레마케터 수익 증대로 이직 감소, 마지막으로 가용 DB의 효율적 활용을 들 수 있다. 그러므로 DB를 통한 수율제고만으로도 텔레마케팅 센터의 안정을 도모하고 이를 순환할 수 있는 매체로써 충분한 가치가 있다고 판단된다.

하지만 데이터마이닝을 통한 모델을 구축하고 실행함에 있어서 한번의 모델 구축만으로는 그 성과가 오래가지 못하고 있다. 모델 구축 후 일정한 시간이 흐르면 모델의 노후화로 인하여 주기적으로 모델개선작업이 필요하기 때문이다. 본 논문에서 구축한 모델도 약2년 동안 1회에 걸쳐서 개선작업을 하였으나 그 주기를 통계적으로 파악하지 못한 상태였다. 향후 모델 주기를 정확하게 찾을 수 있다면 지속적인 텔레마케팅 수율 향상을 이루어 낼 수 있을 것이다.

[참고 문헌]

1. 김연형, 김재훈, 이석원(2007), 『고객관계관리와 데이터마이닝』 교우사, 서울.
2. 이성우, 민성희, 박지영, 윤성도(2005), 『로짓·프라빗모형 응용』, 박영사, 서울.
2. 이석원(2008), 『데이터마이닝의 Association Rule 기법에 관한 연구』, 박사학위논문, 전주대학교.
3. 최인호(2004), 『보험회사 다이렉트 마케팅 활성화 방안 사례연구』, 석사학위논문, 성균관대학교.
4. 21세기 보험산업 발전방안 연구(2001), 생명보험협회, 생명보험협회보, Vol22, No2.
5. Fu, C. Y., Hung, J. H., Liu, S. H., and Chien, Y. L.(2004). 『A new algorithm for solving binary discrimination in conditional logistic regression, with two choices of strata』, *Computational Statistics & Data Analysis*, Vol 49, 85-97.
6. Guadagni, P. M. and J. D. C. Little(1983). 『A logit model of brand choice calibrated on scanner data』, *Marketing Science*, Vol. 2, No3, pp203-238.

7. Hosmer, D. W. and Lemeshow, S.(2000). 『*Applied Logistic Regression*』, 2nd edition, Wiley, New York.
8. Yong Jun, Cho, Joon, Hur, Myoung Joon, Kim(2006). 『A Strategic Plan for Improving Customer Satisfaction in Auto Insurance』, *Journal of the Korean Data & Information Science Society*, Vol. 17, No. 2, pp 355 ~366.

[접수일(2008년 8월 6일), 수정일(2008년 8월 17일), 게재확정일(2008년 8월 23일)]