

Development of Scoring Model on Customer Attrition Probability by Using Data Mining Techniques¹⁾

Sang Tae Han²⁾, Seong Keon Lee³⁾
Hyun Cheol Kang⁴⁾, Dong Kyun Ryu⁵⁾

Abstract

Recently, many companies have applied data mining techniques to promote competitive power in the field of their business market. In this study, we address how data mining, that is a technique to enable to discover knowledge from a deluge of data, is used in an executed project in order to support decision making of an enterprise. Also, we develop scoring model on customer attrition probability for automobile-insurance company using data mining techniques. The development of scoring model in domestic insurance is given as an example concretely.

Keywords : data mining, scoring model, customer attrition

1. 서론

최근 많은 기업들은 자사가 보유한 고객데이터를 이용하여 시장에서의 경쟁력을 갖출 수 있는 다양한 관점의 모델을 개발하는데 데이터마이닝을 적극 활용하고 있다. 즉, 고객과의 커뮤니케이션 및 관계유지로 대변되는 고객관계관리(CRM : Customer Relationship Management)에 데이터마이닝을 핵심정보기술로 활용하고 있는 것이다. 특히 은행, 카드사, 보험사 등 금융 관련 기업에서 가장 적극적으로 활용되고 있는 편인데, 보험사의 경우 국내 대부분의 기업이 데이터마이닝을 이용한 시스템을 운영하고 있다. 보험사 중 특히 손해보험 업계에서는 자동차보험 가입자에 대한 이탈 방지에 가장 큰 관심을 두고 있다.

손해보험 업계의 시장상황을 살펴보면, 최근 자동차의 수가 급격히 증가하고 있는 추세에 있음을 알 수 있는데, 통계청에서 2001년 7월에 발표한 자료에 의하면 한 가구당 자동차 보유대수가 0.88대를 넘어서고 있는 상황이다. 자동차 수가 늘어나면서 자동차 보험사들은 자사의 고객 유치

1) This research was supported by a Grant from Institute of Small and Medium Industry of Hoseo University, in 2002

2) Assistant Professor, Department of Mathematics, Hoseo University 29-1, Asan, 336-795, Korea.

3) Graduate Student, Department of Statistics, Korea University 5-1, Anam-Dong, Sungbuk-Gu, Seoul, 136-701, Korea.

4) Senior Lecturer, Department of Mathematics, Hoseo University 29-1, Asan, 336-795, Korea.

5) Graduate Student, Department of Mathematics, Hoseo University 29-1, Sechul-Ri Asan, 336-795, Korea.

를 위하여 보험상품을 다양화하고 이를 통해 타사와의 차별화를 이루려는 다양한 전략을 세우고 있다. 또한 현재 자동차보험 시장은 포화상태에 이르러 기존의 신규고객 유치전략에서 기존고객 관리전략으로 마케팅 패러다임이 바뀌고 있는 실정이다. 이러한 변화에 발 맞추어 국내의 각 보험사들은 새로운 보험료 체계를 마련하고 고객관리를 강화하기 위해 데이터베이스 마케팅(Database Marketing)의 일환으로 보험가입자의 이탈 방지, 새로운 보험상품 개발 및 자동차 보험 지원 시스템 개발에 심혈을 기울이고 있다. 이런 관점에서 본 연구의 목적은 고객 이탈 방지에 그 초점을 맞추고, 데이터마이닝을 활용하여 경쟁력 있는 자동차 보험 지원 시스템과 향후 캠페인 활동의 중요한 초석이 될 수 있는 자동차 보험 고객 이탈 스코어링 모델을 개발하고자 하는 것이다.

특히 본 연구는 국내 A 손해보험사에서 실제 진행되었던 데이터마이닝 프로젝트를 중심으로 구성하였는데, 프로젝트의 진행과정으로서 CRISP-DM 방법론(Pete and Julian, 1999)이 적용되었고, 데이터마이닝 소프트웨어로는 SAS사의 Enterprise Miner 4.0을 이용하였다(SAS, 1997).

2. 프로젝트 일정 및 CRISP-DM 방법론

2.1 프로젝트 일정

프로젝트는 2000년 10월에서 2001년 1월까지 3개월 동안 진행되었으며, 프로젝트의 일정은 업무 파악, 분석용 마트 구축, 확증적 모형 개발, 이탈 스코어 산출 및 적용의 4단계로 구성되어 진행되었다. 전체 일정의 약 85%가 최종 분석용 마트의 구축에 활용되었고, 실제 모형개발 및 이탈 스코어 산출에 사용된 기간은 전체의 15%였다. 이는 데이터마이닝 프로젝트에서 일반적으로 나타나는 현상인데, 분석용 마트의 구축 단계가 전체 프로젝트에서 매우 중요한 위치를 차지하고 있다는 점과 일반적으로 데이터 추출에 너무 많은 시간이 소요되어 모델링 개발에 충분한 시간이 활용되지 못하고 있다는 점을 시사해 주는 것이다.

2.2 CRISP-DM (CRoss Industry Standard Process of Data Mining) 방법론

본 프로젝트의 단계별 주요 과제는 CRISP-DM 방법론을 바탕으로 진행되었다.

<표 2-1> CRISP-DM 방법

적용 단계 구분	단계별 과제	
비즈니스 이해	1. 업무목적결정 3. 데이터마이닝목표결정	2. 상황평가 4. 프로젝트계획수립
데이터 이해	1. 초기데이터 수집 3. 데이터탐색	2. 데이터기술 4. 데이터품질 검증
데이터 준비	1. 데이터설정 3. 데이터정제 5. 데이터통합	2. 데이터선택 4. 데이터생성
모델링	1. 모델링기법 3. 모델생성	2. 테스트설계 4. 모델평가
모델 평가	1. 결과평가 3. 향후단계결정	2. 프로세스검토
전개	1. 전개계획수립 3. 최종보고서작성	2. 유지관리계획수립

CRISP-DM 방법론은 유럽에서 SPSS Inc, NCR, DaimlerChrysler, OHRA가 함께 개발한 모델로써 초보자나 전문가가 비즈니스 전문가와 함께 모델을 만들어 내는 포괄적인 데이터마이닝 방법론 표준 프로세스이다. 프로젝트의 상황 및 일정상 CRISP-DM 방법론의 일부 과제를 생략하여 진행하였다. 생략된 과제는 비즈니스 이해 단계에서 상황평가, 데이터이해 단계에서 초기 데이터 수집, 모델평가 단계에서 프로세스 검토와 향후단계결정, 전개 단계에서 전개계획수립 등이다.

3. 모형정의 및 데이터 추출

3.1 목적 및 모형의 정의

자동차보험에 가입한 고객을 대상으로 보험만기가 도래했을 때 이탈하는 고객의 특성을 설명할 수 있는 자동차 보험 이탈방지 모형개발을 목적으로 하였다. 이에 따른 모형의 목표변수는 자동차 보험 만기 시점에서의 재계약 여부이며, 입력변수는 자동차 보험 계약건에 대한 만기월 기준의 고객 속성정보 및 거래정보와 과거 5년간의 거래특성을 이용하였다. 또한 마케팅 부서와의 협의를 통하여 자동차보험의 성격에 따라 이탈 패턴이 다르다고 판단되어 자동차보험을 “승용”, “승합용”, “화물용”으로 구분하여 3개의 모형을 개발하였다. 즉, 이질적인 집단들이 섞여 있을 때는 동질적인 집단들로 나누어 개별적으로 분석하는 것이 타당하다고 생각했기 때문이다.

3.2 데이터 추출

앞서 정의한 모형에 필요한 데이터를 A보험사의 기간계 데이터베이스(DB)와 데이터웨어하우스(DW)에서 자동차 보험 만기월 현재 기준으로 최근 5년까지의 고객 속성정보 및 거래정보와 만기월 이후 3개월 사이의 개인 고객의 갱신 정보를 추출하였다. 이 때 주의해야 할 점은, 고객의 이탈에 영향을 미친다고 예상되는 항목은 마케팅 실무자와 분석자가 충분히 논의하여 선정해야 하며 데이터를 추출하는 전산담당자와의 긴밀한 협조가 이루어져야 모형에 적합한 양질의 데이터를 추출할 수 있다는 것이다. 이 과정이 전체 프로젝트 진행과정 중에 가장 중요하고도 어려운 부분이라 할 수 있다.

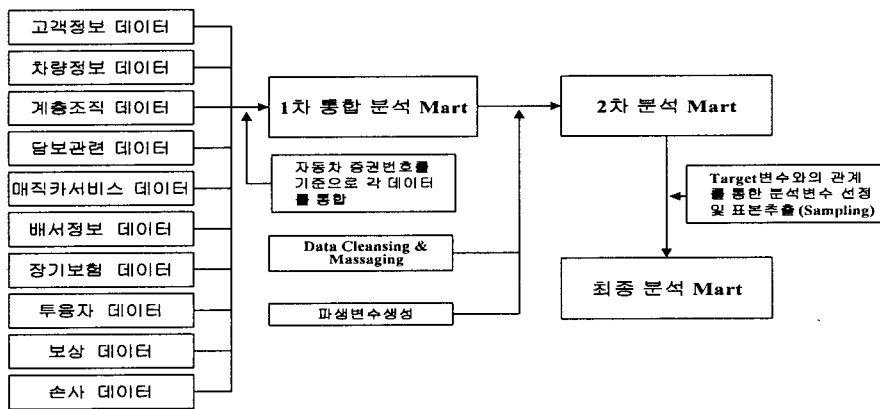
4. 분석용 마트 구성

개인고객의 이탈을 예측하는데 필요하다고 판단되는 각 데이터를 A사의 기간계 시스템에서 추출한 후, 보험 증권번호를 기준으로 통합하여 1차 통합 분석 마트를 구성하였다. 다음으로 데이터 정제와 파생변수를 생성하여 2차 분석 마트를 구성하였고, 마지막으로 목표변수와의 관계를 통한 분석변수 선정 및 표본추출(sampling)을 통하여 최종 분석 마트를 구성하였다. <표 4-1>은 분석용 마트에 대한 구성 흐름도이다. 데이터마이닝 프로젝트에서 대부분의 작업시간이 분석용 마트를 구성하는 데에 할당되므로 매우 중요한 단계라 할 수 있다.

4.1 데이터 탐색

추출된 데이터는 대부분 텍스트나 특정 데이터베이스 파일로 분석자에게 전달되는데, 데이터 탐색 또는 분석을 위해 분석자는 사용가능한 데이터베이스 또는 통계분석 소프트웨어에 데이터를 업로드(upload)해야 한다. 또한 이렇게 전달된 초기 데이터에는 입력오류, 결측값, 불필요한 정보 등이 포함되어 있는 경우가 많으므로, 좋은 모형을 개발하기 위해서는 이에 대한 탐색과 정제가 선결되어야 한다. 본 연구에서는 빈도분석, 분할표분석, 기초통계분석 등을 통하여 데이터의 충실도, 변수들간의 상충, 논리적 오류 등을 살펴보았다.

<표 4-1> 분석용 마트의 구성 흐름도



예를 들어 본 연구에서는 고객의 신상정보에 관련된 대부분의 변수들이 최종 분석용 마트에서 제외되었는데, 이들 변수들이 모형개발에 매우 중요한 역할을 할 수 있음에도 불구하고 <표 4-2>에서 볼 수 있듯이 기록율이 매우 떨어져 분석변수로 사용하기가 어려웠기 때문이다. 이는 고객들이 자신의 개인 정보가 유출됨을 걱정해 알려지기를 꺼려하기 때문으로, 따라서 개인의 프라이버시를 보호해 주면서 필요한 정보를 분석에 사용할 수 있도록 고객에게 신뢰를 심어 주어야 하며 또한 기업에서도 개인의 정보를 다른 용도로 사용하지 않는 도덕성이 담보되어야 할 것이다. 현재의 상황에서 고객의 신상정보를 획득하는 하나의 방법은 캠페인을 통해 신상정보를 제공하는 고객에게 인센티브를 제공하는 것이다. 이렇듯 고객 신상 정보를 획득함으로써 기록율이 향상되면 보다 안정적이고 적중률이 높은 스코어링 모형을 개발할 수 있을 것이라 여겨진다.

<표 4-2> 빈도분석을 통한 결측값 확인(일부)

필드명	형 태	비 고
녹색면허여부	결측값 99%	분석변수에서 제외
면허취득년도	결측값 99%	
운전자경력	결측값 99%	
자녀수	결측값 99%	
학력ID	결측값 99%	
직위	결측값 99%	
선납일	99.9%가 '0'	
고객결혼상태	99.8%가 '미혼'	

4.2 데이터 정제 및 파생변수 생성

고객의 갱신에 관련된 변수들에 포함되어 있는 결측값(missing value) 및 오류를 파악한 후 제거 또는 적절한 값으로 대체 하였는데, 먼저 1단계로 결측값 및 오류의 비율이 90% 이상이 되는 변수들을 제거하고, 2단계로는 필드간 상충, 업무적인 결측값 등은 변수간 교차검토(cross check)를 통하여 적절한 값으로 변환하였다. 3단계는 입력오류 및 값 범위 초과 레코드들을 삭제하였다. 또한, 인수된 데이터 이외에 목표변수에 유의한 영향을 준다고 생각하는 변수를 추가적으로 생성하였는데, 이는 실무자들과 충분한 협의를 통해 진행되었다.

4.3 표본추출

앞의 과정을 거쳐서 수집된 데이터는 약 100만건에 이르는 매우 큰 크기를 가지고 있었기 때문에 이를 모두 이용하여 분석하는 것은 시스템적으로 불가능하다고 할 수 있다. 따라서 승용, 승합용, 화물용 각각에 대해 다음 <표 4-3>과 같이 약 10만건~15만건을 표본추출하여 분석용 마트를 구성하였다.

<표 4-3> 최종 분석 마트

보종	필드수	전체건수	제갱신건	미갱신건	갱신율	이탈율
승용	163	150,000	88,042	61,958	58.69%	41.31%
승합용	167	90,802	41,979	48,823	46.23%	53.77%
화물용	165	150,000	74,878	75,122	49.92%	50.08%

4.4 변수선택

기업 데이터베이스에서 분석에 활용하는 변수는 일반적으로 매우 많을 뿐만 아니라 이들 변수들끼리는 상호 연관성이 크다. 또한 대부분의 경우 이렇게 많은 변수들이 목표변수를 예측하기 위해 모두 유용하게 사용되는 것은 아니므로 모든 변수를 이용하여 모형을 개발하는 것은 타당하지 않다. 따라서 사전에 모형개발에 유용한 변수들만을 선택할 필요가 있다. 본 연구에서는 실무자와의 협의 및 목표변수와의 관계분석을 통해 각 모형별로 입력변수를 선정하였는데, 목표변수를 예측하는데 연관성이 낮은 변수를 제거하기 위해 범주형 변수인 경우에는 카이제곱(Chi-Square)검정을, 연속형 변수인 경우에는 t-검정을 이용하였다.

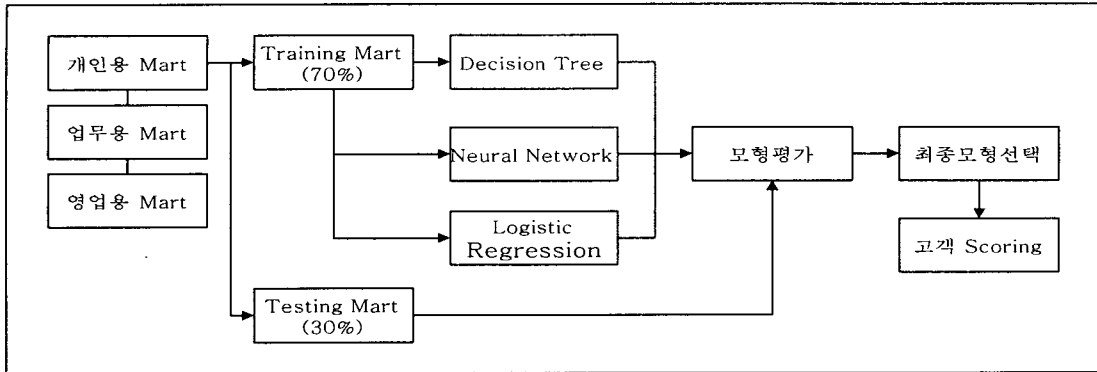
5. 이탈 스코어 모형 개발

5.1 분석흐름도

본 프로젝트의 모형개발을 위해 E-Miner에서 구성한 분석흐름도는 다음 <그림 5-1>과 같다.

E-Miner의 Data Partition 노드를 이용하여 승용, 승합용, 화물용 각각에 대해 모형개발을 위한 훈련용 데이터(training data)를 70%, 모형평가를 위한 평가용 데이터(test data)를 30%로 분할한 후 의사결정나무, 로지스틱 회귀모형, 신경망모형을 개발하고 각 모형을 비교 평가하여 최적의 모형을 선택하였다.

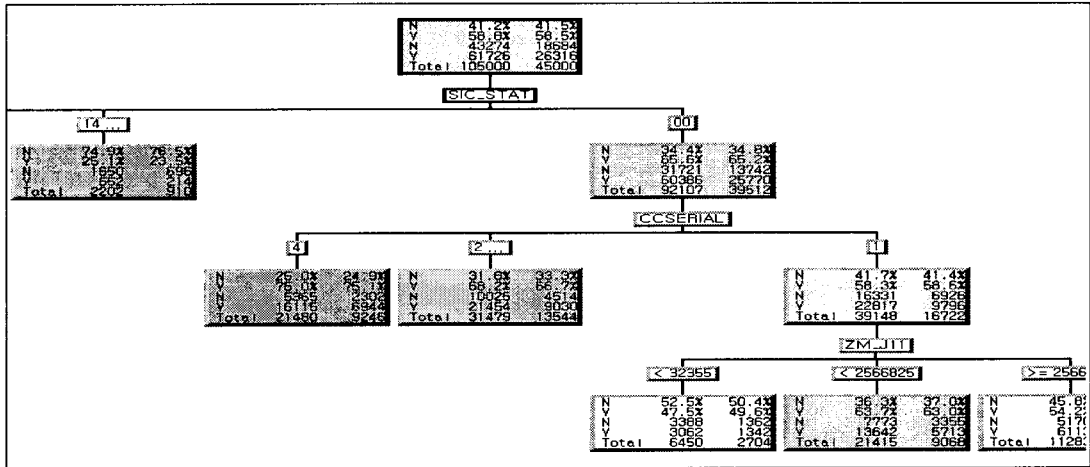
<그림 5-1> 모형개발 흐름도.



5.2 의사결정나무분석 결과

의사결정나무 분석은 E-Miner의 Decision Tree 노드를 이용하였고, 알고리즘은 CHAID를 이용하였다. 승합용 마트를 분석한 결과 중 의사결정나무 결과 일부는 <그림 5-2>와 같다.

<그림5-2>의사결정나무의나무그림(일부)



<표 5-1> 의사결정나무에 의한 승용모형의 오분류표

예측 실제	학습 결과			테스트 결과		
	미가입	가입	합계	미가입	가입	합계
미가입	15,876	27,398	43,274	6,678	12,006	18,684
가입	4,571	57,155	61,726	1,982	24,334	26,316
합계	20,447	84,553	105,000	8,660	36,340	45,000
	Prior Distribution : 58.79%			Prior Distribution : 58.48%		
	학습 정분류율 : 69.55%			테스트 정분류율 : 68.92%		

<표 5-1>은 학습 결과와 테스트 결과로 구분된 오분류표이다. 학습 결과의 정분류율은 69.55%이고, 테스트 결과의 정분류율은 68.92%로 안정적인 모형임을 볼 수 있다. 이 때 승용 모형의 재갱신 규칙의 일부를 살펴보면 다음과 같다.

- ① 계약상태가 전담보해지이면 이탈율이 92.7%이다.
- ② 계약상태가 정상이고 당사연속가입경력이 4년이면 이탈율은 24.9%이다.
- ③ 계약상태가 정상이고 당사연속가입경력이 1년이고 장기수당이 32,355원 이상 2,566,825원 이하이면 이탈율이 37.0%이다.

5.3 로지스틱 회귀분석 결과

로지스틱 회귀분석에서 변수선택법은 Stepwise 방법, 기준통계량은 AIC(Akaike's Information Criterion)를 선택하였고 이때 선택되는 변수와 제거되는 변수의 유의수준으로는 각각 15%, 5%를 지정하였다. 승합용 모형에 대해 로지스틱 회귀분석을 실시한 결과는 <표 5-2>와 같다.

<표 5-2> 로지스틱 회귀분석에 의한 승합용 모형의 오분류표

예측 실제	학습 결과			테스트 결과		
	미가입	가입	합계	미가입	가입	합계
미가입	21,947	12,116	34,063	9,508	5,252	14,760
가입	8,242	21,256	29,498	3,531	8,950	12,481
합계	30,189	33,372	63,561	13,039	14,202	27,241
	Prior Distribution : 46.41%			Prior Distribution : 45.82%		
	학습 정분류율 : 67.97%			테스트 정분류율 : 67.76%		

5.4 신경망 분석 결과

신경망 분석에서는 Neural Network 노드의 MLP(multi-layer perceptron) 알고리즘을 사용하였으며 네트워크의 구조는 은닉층 1개 및 은닉노드 3개로 구성하였고, 나머지 옵션은 디폴트로 하여 수행하였다. 화물용 모형에 대한 신경망 분석의 결과는 <표 5-3>과 같다.

<표 5-3> 신경망 분석에 의한 화물용 모형의 오분류표

예측 실제	학습 결과			테스트 결과		
	미가입	가입	합계	미가입	가입	합계
미가입	25,231	27,358	52,589	10,745	11,788	22,533
가입	8,824	43,587	52,411	3,721	18,746	22,467
합계	34,055	70,945	105,000	14,466	30,534	45,000
	Prior Distribution : 49.92%			Prior Distribution : 49.93%		
	학습 정분류율 : 65.54%			테스트 정분류율 : 65.54%		

5.5 모형평가 및 최종모형 선택

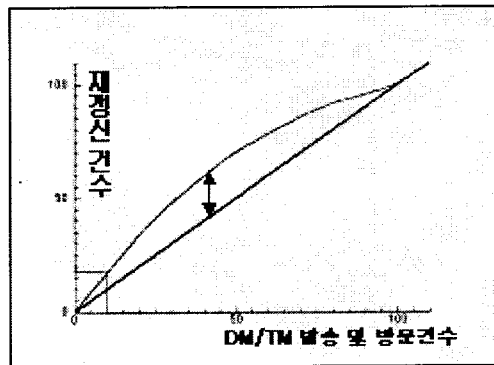
승용, 승합용, 화물용 모형에 대해 의사결정나무분석, 로지스틱 회귀분석, 신경망분석을 실시한 학습결과와 테스트 결과의 정분류율은 <표 5-4>와 같다. 결과를 살펴보면 승용, 승합용, 화물용

모형에서 의사결정나무분석 및 신경망 분석보다 로지스틱 회귀모형의 정분류율이 약간 더 우수한 것을 볼 수 있다. 하지만 뛰어난 정.도가 매우 미약하고 모형의 안정성과 시스템 이식성 및 향후 현장에서의 활용 용이성을 고려하여 의사결정나무분석모형을 고객 스코어링에 사용할 최종 모형으로 선택하였다. <그림 5-3>의 리프트 도표(lift chart)는 데이터마이닝을 통해 얻을 수 있는 이익을 보여주고 있는데, 대각선은 모형을 개발하기전 무작위로 고객과 접촉했을 때의 반응율을 나타내고 위쪽의 곡선은 모형을 이용한 반응율을 나타낸다(정분류율 및 리프트 도표 등 데이터마이닝 모형평가에 관한 자세한 사항은 강현철·한상태 외(2001), 최종후·한상태 외(2001) 등을 참조하기 바란다).

<표 5-4> 각 모형의 정분류율 비교

모형의 정분류율 비교		학습 결과	테스트 결과
의사결정나무	승 용	69.55%	68.92%
	승합용	66.80%	66.31%
	화물용	67.28%	67.18%
로지스틱 회귀분석	승 용	70.35%	69.98%
	승합용	67.97%	67.76%
	화물용	68.49%	68.30%
신경망 분석	승 용	69.80%	69.50%
	승합용	61.08%	61.12%
	화물용	65.54%	65.54%

<그림 5-3> Lift Chart(decision tree)



5.6 개발된 모형을 통한 캠페인 활동

각 보증별로 개발된 모형을 토대로 하여 상대적으로 예상 갱신율이 낮은 고객을 선정하고, 그 고객을 상대로 관촉활동을 하는 캠페인을 고려하자. 향후 특정 월에 보험만기가 도래하는 고객을 기준으로 하는 적용데이터(applying data)를 최종 선택된 모형인 의사결정나무에 적용시켜 이탈 및 재갱신 스코어를 산출하고 산출된 고객별 스코어를 영업소 화면에 출력시켜 보험설계사의 활동 효율성을 높일 수 있다.

다음 <그림 5-4>의 영업소 현장화면을 보면 고객별 재갱신 스코어와 고객 세그먼트(segment)

별 특성들이 나타나게 되는데, 영업사원은 이 화면을 참조하여 방문할 고객을 선정할 수 있고 또한 대상 고객에 대한 특성과약을 함으로써 영업활동에 많은 도움을 받을 수 있다. 화면에 나타나는 고객 세그먼트의 특성은 의사결정나무 모형에서의 끝마디(terminal node)를 표현하는 의사결정 규칙에 사용된 변수들의 조합을 나타낸다.

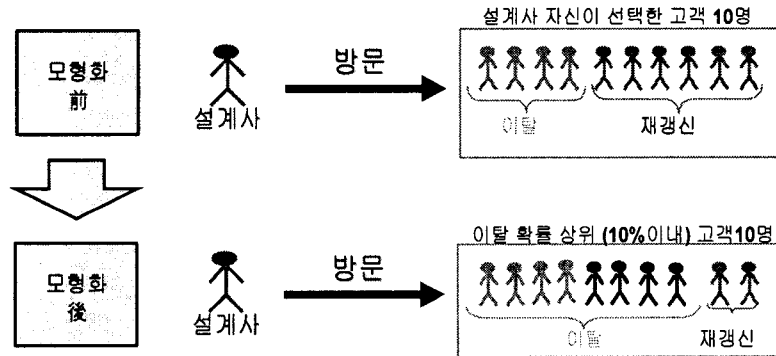
모형을 통한 캠페인 활동의 효율성 측면을 고려할 때 설계사가 기존의 무작위 선택 방법으로 이탈할 고객을 접촉하는 경우 10번 방문해서 4번의 성공기회를 갖지만, 모형에 의한 이탈확률 상위 10% 고객을 접촉하면 8번의 성공기회를 갖게 되므로 설계사의 활동 효율성을 제고할 수 있다 (<그림 5-5 참조>). 이러한 결과는 <그림 5-3>의 리프트 도표를 통해서도 확인할 수 있다.

<그림 5-4> 영업소의 현장화면

점권번호 차량번호	피보험자명 보유기간	경신 확률 표준(L/R) 제결
4001200	은병	81.9 **
서울 3	2000	7/02 79.7 제결
4001200	한자	74.5 **
경기 9	2000	7/21 76.7 제결
4001200	지순	74.5 **
서울 5	2000	7/13 66.7 제결
4001200	김환	74.5 **
서울 3	2000	7/18 94.1 제결
4001200	정순	74.5 **
서울 5	2000	7/21 65.0 제결
4001200	김철	65.2 **
경기 3	2000	7/25 43.9 제결
4002200	홍길	71.0 **
경기 7	2000	7/12 0.0 미제결

✓ 집금자의 불 차보수액 <= 5% 041원 ✓ 차보(권)가입자 ✓ 자순 미가입 또는 가입금액이 1억원 ✓ 서명방법이 자필서명아님 또는 FAX 기타	1인2물자주 지보 이 23 인호 1인2물자주차 이 75 인호
---	---

<그림 5-5> 모형을 통한 캠페인 활동의 예시



6. 결론 및 토의

본 연구는 실제 데이터마이닝 프로젝트에서 진행된 과정을 모형의 정의와 데이터의 추출에서부터 스코어링 모형개발까지 일련의 흐름을 통해 소개하였다. 또한 의사결정나무, 로지스틱 회귀분석, 신경망분석을 이용하여 모형을 개발하고, 모형의 안정성과 향후 적용 등을 고려한 모형선택

과정과 실제 데이터마이닝 캠페인을 통해 얻을 수 있는 회사의 이익에 대해서도 살펴보았다.

본 연구를 진행하면서 데이터마이닝 프로젝트 수행 시 중요하게 고려되어야 할 것으로 생각되는 몇 가지 사항을 정리해 보면 다음과 같다.

첫째, 프로젝트를 수행하는데 있어 현장의 마케팅 실무자와 분석팀 그리고 전산팀의 상호 이해 및 긴밀한 협조체제가 유지되어야 한다는 것이다. 즉, 현장의 요구와 감각이 모형에 반영되어야 하며, 데이터마이닝 프로젝트의 대부분의 작업이 데이터를 추출하고 조작(manipulation)하는 데에 소요되므로 전산팀의 역할 또한 매우 중요하다.

둘째, 분석 데이터의 질이다. 앞서서도 살펴보았듯이 고객의 이탈에 영향을 줄 수 있는 고객의 신상정보는 거의 획득되어지지 않고 있었고 또한 획득되어진 자료도 오류가 많아 분석에 사용할 수 없었다. 따라서 양질의 데이터를 신속히 제공할 수 있는 기업환경이 데이터마이닝 프로젝트의 성공적인 진행과 안정적인 모형을 개발하는 중요한 요소로 깊이 인식해야 한다는 점이다.

향후 고객의 프라이버시를 침해하지 않는 범위 안에서 분석에 필요한 고객 신상정보를 획득하고 활용하는 방법론과 이를 위한 보안정책에 대한 연구가 활성화되어야만 하며, 자동차 보험의 차량용도와 장기보험 상품을 다각도로 고려하여 고객유형을 선정하고, 장기보험 상품과 연계된 연계 판매(cross-sell)모형 개발 등을 추후 연구 과제로 제안한다.

참고문헌

- [1] 강현철 · 한상태 · 최중후 · 김은석 · 김미경 (2001). 「SAS Enterprise Miner를 이용한 데이터마이닝-방법론 및 활용-」, 자유아카데미. 서울.
- [2] 최중후 · 한상태 · 강현철 · 김은석 · 김미경 · 이성건 (2001). 「SAS Enterprise Miner를 이용한 데이터마이닝-기능과 사용법」, 자유아카데미. 서울.
- [3] Chapman, P. and Clinton, J (1999). Crisp-DM Process Model , *Discussion paper*.
- [4] SAS Institute (1997). *Data Mining Using SAS Enterprise Miner Software*. SAS Institute Inc.

[2001년 10월 접수, 2002년 4월 채택]