
제조업분야의 고객 성향 및 추이 분석

Analysis of Customer Behavior and Trend of Manufacture

이병엽*, 임승빈*, 박용훈**, 유재수**
배재대학교 전자상거래학과*, 충북대학교 전기전자컴퓨터공학부**

Byoung-Yup Lee(bylee@pcu.ac.kr)*, Seung-Bin Yim(bin@pcu.ac.kr)*,
Yong-Hoon Park(yhpark@netdb.chungbuk.ac.kr)**, Jae-Soo Yoo(yjs@chungbuk.ac.kr)**

요약

최근 기업은 업무의 효율적 수행을 위해 데이터베이스를 사용하고, 저장된 데이터베이스의 데이터로부터 분석을 통해 행동 패턴을 추출해내어 그 결과를 마케팅과 생산의 효율성 증대를 위해 데이터마이닝을 많이 사용한다. 데이터마이닝을 통해 얻어진 지식의 활용은 기업 활동을 정비하고 활동방향을 제시하며 의사결정의 순간에 기반 자료로 활용될 수 있는 부가적 경쟁력이라 할 수 있다. 본 논문에서는 제조업체의 실제 데이터를 가지고 데이터마이닝 방법론을 이용하여 기존고객의 등급 및 소비행위 파악을 위한 예측모델을 설계한다. 이를 통해 고객의 등급 및 소비행위를 파악하여 이를 마케팅까지 연결, 수익을 창출하고, 기업의 브랜드 가치를 향상시키는데 목적이 있다.

■ 중심어 : | 데이터 마이닝 | 분류 규칙 | 의사결정나무 | 연관규칙 |

Abstract

Companies often use database for performing task more efficiently and data mining for marketing and production efficiency through analyzing of the stored database. The use of the knowledge through the data mining maintains and provides a direction of development for the company. It could be as an additional competitive power for the company when decision making is necessary. This study is designing a model that predicts a rating of existing customer and consumption pattern with using actual data of the manufacturer and data mining methodology. The objective of this model is to improve profits for the company and brand value through connecting the marketing with identifying the customer's rating and consumer behavior.

■ keyword : | Data Mining | Classification | Association Rule | Decision Rule | Sequence Rule |

I. 서론

기업이 그들의 고객과 상호 교류하는 방식은 지난 몇 년 사이 엄청나게 바뀌어 왔다. 이제 더 이상 고객의 지속적인 거래는 보장되지 않는다. 이 결과로 기업들은

고객을 더 잘 이해하고 그들이 원하고 필요로 하는 것에 더 빨리 대응해야 할 필요성을 느낀다. 고객의 불만족 사인이 명백해져 어떤 조치를 취할 때까지 기다리는 것은 이제 불가능하다. 성공하기 위해서는 고객이 원하는 것을 좀 더 앞서 분석하고 그것이 무엇인지 예측 해

* 이 논문은 2009년 지식경제부의 특구연구개발사업으로부터 지원받아 수행된 연구임.

야만 한다[1]. 이런 상황에서 경영자는 성공적으로 기업을 이끌어가기 위해 변화가 필요함을 인식하게 되었다. 기업은 다양한 고객들의 요구가 무엇인지를 알고 싶어 하고 이를 경영에 포함시켜 좀 더 내실 있는 경영으로 나아가기를 원한다. 이러한 요구를 충족시켜 줄 수 있는 방안이 고객관계관리이다[3].

고객관계관리는 경영학적인 개념이지만 기술적인 방법론이 접목되지 않고서는 결코 만들어 질 수 없는 개념이다. 고객의 중요성이 부각되는 시대적인 흐름과 발전된 컴퓨팅 및 통계학의 기술이 만나 고객관계 관리가 이제 경영의 필수조건이 되어가고 있다. 고객관계관리가 중요 할 수밖에 없는 시대적인 흐름을 이해하고 고객관계관리란 무엇이며 여기에 실천적으로 응용될 수 있는 기법을 탐색하는 것이 본 논문의 주된 목적이다.

한국의 CRM 시장은 고객과의 접촉을 관리하고 고객 응대의 효율성을 확고하기 위한 운영시스템으로서의 CRM이 주도하여 왔다. Siebel 대표되는 CRM 솔루션이 바로 그런 것들이다. 기업은 이와 같은 CRM 시스템의 도입에 많은 비용의 지출을 감수하였는데 기업의 기대수준에 비해 그 효과가 부진한 상황이다. 따라서 기업들은 데이터마이닝을 통한 고객의 심도있는 분석에 기반하여 고객응대 및 마케팅을 효율화하는 데에 관심을 두기 시작하였다[2].

데이터마이닝은 정보기술의 발전을 바탕으로 대규모로 축적되고 있는 데이터를 정보화하기 위한 노력의 일환으로 볼 수 있다. 데이터마이닝을 통해 산더미처럼 쌓여있는 데이터로부터 의미있는 정보를 캐내고 이를 기업의 의사결정에 반영함으로써 보다 과학적인 기업 경영을 실현하게 한다. 데이터마이닝이 CRM의 핵심기술요소로서 활약하면서 현재 국내의 CRM은 제2의 성장기를 맞이하고 있다[2].

본 논문은 국내 제조업체인 A사의 자료를 입수하여 실제 데이터를 중심으로 연구하였다. 또한 의사결정나무, 연속규칙, 연관규칙을 이용하여 고객속성 및 구매추이를 분석하였다.

II. 분석 방법 및 구성

본 연구에서는 국내 제조업체인 A사의 실제 데이터 자료를 입수하여 연구하였고 수행하는 연구의 진행절차는 데이터마이닝 프로세스의 각 단계별 과정에 따라 진행을 하고자 하며 과정은 [그림 1]과 같다.

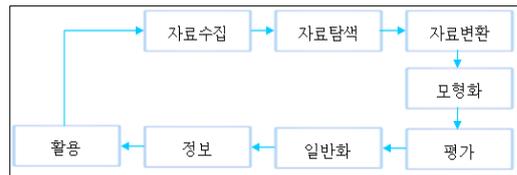


그림 1. 데이터마이닝의 과정

데이터마이닝의 효율적인 적용과정을 살펴보면 대용량의 자료를 수집, 탐색, 보완을 하는 과정 (Data cleansing) 정제된 자료를 적절한 기법을 이용하여 분석하는 모형화 과정(Modeling), 그리고 모형화에 대한 평가 (Assessment)를 하는 과정을 거쳐 일반화된 정보를 얻어 활용하게 하고 또 다른 분석에 대한 피드백 과정을 주게되어 보다 유용한 정보를 얻을 수 있게 한다고 볼 수 있다.

일반적인 데이터마이닝 과정은 8단계로 구성되어 있다. 각 단계는 여러 업무로 분할 가능하며 상이한 단계를 반복적으로 수행하여 최적의 결과를 만들어 낸다[4].

- Step 1 : 기초 통계 분석

대상문제에 대한 명세화 또는 데이터 마이닝의 목표에 대한 명확한 정의를 내린다. 이 단계의 산출물은 이후 단계들의 준비와 실행에 관한 전략적 계획이다.

- Step 2 : 도메인 분석

응용 도메인, 데이터, 환경적 특성에 관한 지식을 분석하여 초기 데이터마이닝 계획을 수립하다.

- Step 3 : 데이터 집합 정의

데이터 마이닝의 대상이 될 데이터 집합 또는 변수집합을 정의한다.

- Step 4 : 사전처리

기법 적용 전에 필요한 모든 태스크가 포함된다. 여기에는 데이터의 적재, 변환 및 클리닝이 포함된다.

- Step 5 : 데이터 탐구

감독/무감독(Directed/Undirected) 데이터 마이닝과 가설검정에 적용할 데이터에 대한 통찰과 흥미있는 데이터 또는 특성 부분집합을 파악한다.

- Step 6 : 데이터마이닝 기법의 적용

다양하고 상이한 기능별 데이터 마이닝 기법을 선택 적용한다. 데이터 마이닝 기법의 기능은 분류, 추정, 예측, 군집화, 유사 집산화, 서술/설명 등 6가지로 분류할 수 있다.

- Step 7 : 해석과 평가

데이터 마이닝의 결과는 사용자가 해석 가능한 용어 또는 의사결정에 이용할 수 있는 지식으로 표현되어야 하고, 단계 1에서 정의된 평가기준에 의해서 평가된다. 평가가 성공적으로 이루어지지 못한 경우에는 이전 단계로의 회귀와 단계의 반복이 필요하다. 적용된 모델을 실제상황에 적용해 보고 새로운 상황에 맞게 피드백 하는 단계이다.

- Step 8 : 데이터 마이닝 결과의 통합

성공적인 데이터 마이닝 태스크의 결과는 의사결정 문제의 해결을 위해서 사용된다. 결과는 사후처리를 위해 다른 시스템에 통합되거나 필요로 하는 사용자를 위해서 문서화된다.

III. 분석의 절차 및 변수 선정

3.1 자료의 구성

본 연구에서 사용한 데이터는 국내 제조업체의 자료를 대상으로 한다. 데이터를 이루고 있는 고객의 수는 955명, 사원 1359명으로 총 고객 및 사원에서 샘플링을 하여 선정하였고, 고객 및 상품관련 데이터로 나누어진다. 다음 [표 1]은 보유 데이터를 이용하여 산출한 변수를 설명한다. 그룹에 대한 연관규칙의 분석의 순서로 진행 되었다.

표 1. 산출된 생성 변수

범주	변수	수
고객인적정보	연령, 주소, 결혼여부, E-Mail 유무, 직업, 직위	6
사원인적정보	연령, 주소, 결혼여부	3
상품정보	2번이상 구매고객, 2개이상 구매고객	2

3.2 변수의 선정

데이터마이닝 알고리즘을 적용하기 위해서는 데이터에 대한 사전처리가 필요하다. 표본의 각 사례의 특정 변수에 누락된 데이터가 있는지 확인하여 누락된 데이터에 대해서는 변수의 평균값, 중위수 혹은 중간값 등으로 결측치를 대체시킨다. 만약 결측치를 지닌 변수가 이산형이라면 해당 변수의 최빈값으로 대체하는 것이 일반적이다. 또한 중복된 레코드를 제거하고 변수의 값이 일관성이 없으면 이를 바로 잡아야 한다.

3.2.1 연령별

구매고객의 연령을 나타내는 변수로 연령그룹별 분류에서 40대 이용자가 49.65%를 차지함을 보여주며, 30대 40대 순으로 주요 고객임을 확인 할 수 있다. [표 2]는 이용고객의 연령별 고객현황을 나타낸다.

표 2. 연령별 고객 현황

고객그룹연령	인원수	백분율
10대	10	1.42%
20대	80	11.35%
30대	130	18.44%
40대	350	49.65%
50대	100	14.18%
60대	30	4.26%
60대 이상	5	0.71%

3.2.2 주소별

구매고객의 지역 분포도를 나타내는 변수로 대도시 위주인 서울, 인천, 부산 순으로 분포되어 있으며, 서울 근교 도시의 고객이 많은 것을 확인 할 수 있다. [표 3]은 이용고객의 연령별 고객현황을 나타낸다.

표 3. 주소별 고객 현황

고객지역	인원수	백분율
서울	4000	39.6%
부산	1000	9.9%
광주	800	7.92%
대구	500	4.95%
인천	1200	11.88%
대전	600	5.94%
경기도	500	4.95%
충청도	300	2.97%
강원도	100	0.99%
경상도	700	6.93%
전라도	400	3.96%

3.2.3 결혼여부별 및 E-Mail유무별

구매고객의 결혼여부 및 E-Mail 존재 유무를 나타내는 변수로 기혼인 구매고객이 대부분이며, E-Mail을 보유하고 있는 고객은 크게 차이가 없는 것을 확인 할 수 있다. 고객 증가를 위해 미혼을 상대로 활성화에 관한 대책이 필요하다는 것을 알 수 있다. [표 4]는 결혼여부 및 E-Mail유무별 고객현황을 나타낸다.

표 4. 결혼여부별 및 E-MAIL 유무별 현황

고객결혼 여부	인원수	백분율	E-MAIL 존재유무		
			인원수	백분율	
기혼	900	94.24%	있다	500	53.59%
미혼	55	5.76%	없다	433	46.41%

3.2.4 직업 및 직위별

구매고객의 분포는 전문직 및 주부가 대다수를 차지 하며, 직위는 대리부터 차장까지 주요 고객층으로 나타난다. [표 5]는 직업 및 직위별 고객현황을 나타낸다.

표 5. 직업 및 직위별 현황

고객결혼 여부	인원수	백분율	E-MAIL 존재유무		
			인원수	백분율	
전문직	200	25.87%	사원	100	6.06%
회사원	55	7.12%	대리	200	12.12%
공무원	100	12.94%	과장	300	18.18%
자영업	80	10.35%	차장	500	30.30%
기술직	88	11.38%	부장	400	24.24%
주부	150	19.40%	이사	100	6.06%
기타	100	12.94%	대표이사	50	3.03%

3.2.5 사원의 연령별 및 주소별

직원 구매고객의 분포는 고객과 동일하게 40대에 많이 분포되어있으며, 서울근교 지역에 많이 거주하는 것으로 나타난다. [표 6]는 연령 및 주소별 고객현황을 나타낸다.

표 6. 사원의 연령 및 주소별 현황

고객결혼 여부	인원수	백분율	E-MAIL 존재유무		
			인원수	백분율	
10대	10	1.42%	서울	4000	39.60%
20대	80	11.35%	부산	1000	9.90%
30대	130	18.44%	광주	800	7.92%
40대	350	49.65%	대구	500	4.95%
50대	100	14.18%	인천	1200	11.88%
60대	30	4.26%	대전	600	5.94%
60대이상	5	0.71%	경기도	500	4.95%
			충청도	300	2.97%
			강원도	100	0.99%
			경상도	700	6.93%
			전라도	400	3.96%

3.2.6 상품별 2번 이상 구매고객 및 2개 이상 구매고객

2번 이상 및 2개 이상 구매한 고객의 분포는 VIP 회원이 대부분이며, 잠재고객에 대한 구매 횟수는 낮게 나타난다. [표 7]는 연령 및 주소별 고객현황을 나타낸다.

표 7. 상품별 2번 이상 구매고객 현황

2번이상 구매	인원수	백분율	2개이상 구매		
			인원수	백분율	
신규	5	2.56%	신규	50	19.23%
잠재	10	5.13%	잠재	40	15.38%
고정	80	41.03%	고정	70	26.92%
VIP	100	51.28%	VIP	100	38.46%

3.3 모형의 구성

모형의 구성 과정에서 데이터의 특성을 나열하고 데이터 마이닝의 문제정의 제약 조건을 정의 하면 다음과 같다. 첫째, 데이터 마이닝의 문제를 해결하는 가장 일반적인 5가지(Regression, Association, C4.5, Episode, Sequence) 알고리즘이 있다. 이러한 각각의 알고리즘 별로 데이터의 일정한 형식의 타입과 변수들이 필요한데, 현재 분석을 하기위해 보유하고 있는 로우 데이터의 특성을 살펴보면 고객의 의견이나 생각을 묻는 설문 조사 형식의 자료들로서, 고객의 거래 데이터 변수들이

결여되어 있어 적용할 수 있는 알고리즘이 극히 제한적이다. 앞서 언급한 이유 이외에도 CRM 또는 마이닝에 대한 해석은 고객의 거래 데이터에 대한 분석이 주를 이루고 있고, 또한 고객의 성향을 알 수 있는 데이터들이 부족해서 알고리즘 적용시 특별한 툴들을 얻어내기가 어렵다. 따라서 제한 조건을 벗어난 데이터 마이닝의 문제점들을 도출한 결과 C4.5, Association Rule, Sequence Rule을 통한 고객의 성향 분석, 전체적인 데이터의 성격들의 결과를 도출할 수 있었다.

3.3.1 의사결정나무 분석(C4.5)모형 구성

본 논문에서 고객의 등급별 성향 분석 및 사원의 재직별 성향분석을 위해 C4.5 기법을 사용했다. 즉 어떤 고객이 어떤 등급에 해당하는지와 어떤 사원들이 어떤 재직에 해당하는가를 분석하였다. [그림 2]는 C4.5의 시나리오를 구성한 그림이다.

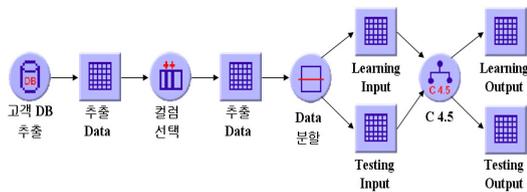


그림 2. C4.5의 시나리오 구성 모형

3.3.2 연속규칙(Sequence Rule)모형구성

본 논문에서 고객 등급별 순차 구매상품의 어떤 성향이 있는가를 분석하였다. Sequence Rule은 문제 정의가 불가능하고 변수들간의 시차 연관성을 통한 규칙을 발견하는데 이용한다. [그림 3]은 Sequence Rule의 시나리오를 구성한 그림이다.



그림 3. Sequence Rule의 시나리오 구성모형

3.3.3 연관규칙(Association Rule)모형구성

본 연구에서 고객의 등급별 속성에는 어떤 성향이 존

재하는가와 고객등급별 동시 구매상품에 어떤 성향이 있는가, 사원재직별 판매상품에 어떤 성향이 있는가를 분석하였다. 또한 사원재직별 판매설계 및 상품에 어떤 성향이 있는가를 함께 분석하였다. [그림 4]는 Association Rule의 시나리오를 구성한 그림이다.



그림 4. Association Rule의 시나리오 구성 모형

IV. 모델링 분석

4.1 의사결정나무 분석(C4.5)결과

C4.5에서 VIP 등급은 직업이 서울에 거주하며 E-MAIL을 하는 고객으로 신뢰도 64,317%로 나타났다. 고정 등급은 주부이고 연령이 40세 이하, 서울에 거주하는 고객으로 61.82%의 신뢰도를 나타냈으며, 이탈고객은 직업이 기타이고 사원이며 경기에 거주하는 고객으로 74.03%의 신뢰도로 나타났다. 또한 잠재고객은 전문직에 종사하며 직위는 대리이고 연령이 37세 이상, 40세 이하로 신뢰도 88.16%로 나타났다. [표 8]은 Decision Tree(Quinlan C4.5)의 수행결과를 나타낸다.

표 8. Decision Tree(Quinlan C4.5)의 수행결과

Decision Tree		
TreeNode/Leaf	Target	신뢰도(%)
주소 = 서울	고정	44.06
주소 = 경기	-	-
직업 = 공무원	VIP	39.62
직업 = 기술숙련직	고정	36.52
직업 = 자영업	고정	52.15
직업 = 8	VIP	53.87
직업 = 9	VIP	49.37
직업 = 주부	고정	54.18
직업 = 경영/회사원	-	-
E-Mail = O	VIP	45.56
E-Mail = X	-	-
주소 = 부산	고정	37.27
주소 = 서울	VIP	64.31
직업 = 전문직	-	-
결혼 = O	고정	43.83
결혼 = X	VIP	50.82
직업 = 기타	-	-
연령 <= 38	고정	42.87
연령 > 38	VIP	53.71

Rule(등급 = VIP)						
직업	직위	나이	주소	결혼	E-MAIL	신뢰도(%)
경영/회사원	-	-	서울	-	O	64.3%
공무원	대리	-	서울	-	-	63.8%
기타	-	>44	경기	-	-	61.6%
전문직	-	>37	서울	O	-	57.4%
기타	-	>38	광주	-	-	53.1%
전문직	-	-	대구	X	-	50.8%

Rule(등급 = 고청)						
직업	직위	나이	주소	E-MAIL	신뢰도(%)	
주부	-	<=40	서울	-	61.82%	
경영/회사원	과장	>31	경기	-	61.23%	
-	사원	>31	대구	X	59.01%	
기술 숙련직	-	>31	경기	-	58.40%	
기타	대리	>31	-	-	56.71%	
주부	-	>31	-	-	56.45%	
자영업	-	-	-	X	55.86%	
기타	-	>31	광주	-	55.36%	
주부	-	>31	-	-	56.44%	

Decision Tree		
TreeNode/Leaf	Target	신뢰도(%)
직위 = 사원	잠재	56.39
직위 = 부장	이탈	67.78
직위 = 사장	이탈	55.78
직위 = 사원	-	-
직업 = 기타	잠재	62.20
직업 = 기술 숙련직	이탈	45.27
직업 = 자영업	이탈	50.72
직업 = 주부	잠재	66.25
직업 = 공무원	-	-
연령 <= 26.0	이탈	44.38
연령 > 26.0	잠재	67.35
직업 = 전문직	-	-
연령 <= 23.0	이탈	67.49
연령 > 23.0	잠재	56.27
직업 = 경영/회사원	-	-
결혼 = O	이탈	64.52
결혼 = X	잠재	47.05
직위 = 부장	-	-
연령 > 48.0	이탈	61.29
연령 <= 48.0	-	-
연령 <= 44.0	이탈	44.24
연령 > 44.0	잠재	64.52
직위 = 대리	-	-
직업 = 공무원	이탈	51.46
직업 = 전문직	이탈	62.58
직업 = 기타	잠재	59.40
직업 = 기술 숙련직	이탈	61.23
직업 = 자영업	잠재	54.09
직업 = 주부	이탈	51.20

Rule(등급 = 이탈)						
직업	직위	성별	나이	주소	결혼	신뢰도 (%)
기타	사원	-	-	경기	-	74.03%
-	사원	-	-	-	X	73.10%
전문직	사원	남	>26	-	-	70.01%
공무원	사원	-	>26	-	-	67.35%
주부	-	-	-	-	-	66.25%
-	사원	-	-	경기	O	64.52%
-	-	-	<=48	-	-	64.52%
경영/회사원	대리	-	>44	-	-	63.85%
기타	사원	-	-	-	-	62.20%
전문직	대리	남	>23	-	-	62.05%

Rule(등급 = 잠재)						
직업	직위	성별	나이	주소	결혼	신뢰도 (%)
전문직	대리	-	<=40	-	-	88.16%
전문직	과장	-	>37	-	-	73.77%
-	과장	-	<=35	-	-	71.85%
전문직	과장	-	>30	-	-	70.43%
공무원	과장	-	-	-	-	70.21%
전문직	과장	-	-	-	-	69.62%
-	과장	-	-	-	-	67.78%
전문직	과장	-	<=23	-	-	67.49%
회사원	사원	-	-	-	X	64.52%
전문직	대리	-	-	-	-	62.58%

4.2 연속규칙(Sequence Rule) 분석 및 결과

Sequence Rule에서의 결과는 잠재고객은 상품A를 구매하고 C를 구매할 확률이 높은 것으로 나타났으며, 고정고객은 상품A를 구매하고 B,E를 구매할 확률이 높은 것으로 나타났다. 또한 VIP 고객은 상품A를 구매하고 C를 구매할 확률이 높은 것으로 추정됐다. 그러므로 고객 등급에 따라 상황판매(Up-selling)하여 구매 고객에 대한 활성화를 증진시키는 마케팅이 필요하다는 것을 알 수 있다. [그림 5]은 Sequence Rule에 의한 수행 결과를 나타낸다.

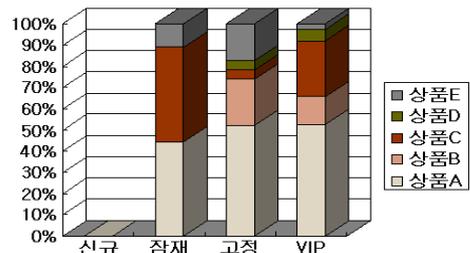


그림 5. Sequence Rule의 수행 결과

4.3 연관규칙(Association Rule) 분석 및 결과

Association Rule에서의 결과는 이탈, 잠재 고객은 상품A와 D를 구매할 확률이 높은 것으로 나타났으며, 고정고객과 VIP고객은 상품A와 B를 동시에 구매할 확률이 높은 것으로 나타났다. 그러므로 고객 등급에 따라 교차판매(Cross-selling)하여 구매 고객에 대한 활성화를 증진시키는 마케팅이 필요하다는 것을 알 수 있다. [그림 10]은 Association Rule에 의한 수행 결과를 나타낸다.

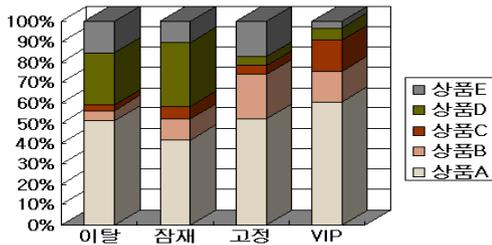


그림 10. Association Rule의 수행 결과

V. 결론

본 논문을 통하여 분석한 결과 고객 반복 구매유도를 위해 상황판매 및 교차판매를 이용하여 구매고객을 활성화하고, 고객 등급의 특성을 분석하여 적절한 마케팅 활용을 통한 우수고객 유치 및 영업력을 극대화 할 수 있다. 또한 고객 및 사원 이탈방지를 하여 고객가치 증진을 실현하고 평생 고객화를 할 수 있다. 본 논문에서는 국내의 제조업체의 고객을 대상으로 데이터마이닝을 이용하여 고객의 성향을 분석하였고 그 결과는 다음과 같다.

첫째, VIP 등급은 직업이 서울에 거주하며 E-MAIL을 하는 고객이며, 고정 등급은 주부이고 연령이 40세 이하, 서울에 거주하는 고객으로 61.82%의 신뢰도를 나타냈다. 또한 이탈고객은 직업이 기타이고 사원이며 경기도 거주하는 고객이며, 잠재고객은 전문직에 종사하며 직위는 대리이고 연령이 37세 이상, 40세 이하로 나타났다.

둘째, 잠재고객은 상품A를 구매하고 C를 구매할 확

률이 높은 것으로 나타났으며, 고정고객은 상품A를 구매하고 B,E를 구매할 확률이 높은 것으로 나타났다.

셋째, 이탈, 잠재 고객은 상품A와 D를 구매할 확률이 높은 것으로 나타났으며, 고정고객과 VIP고객은 상품A와 B를 동시에 구매할 확률이 높은 것으로 나타났다.

추후 연구과제로는 다수의 변수 중에서 고객의 특성 속성을 인공지능경망에 적용함으로써 다양한 신규 고객층의 추출을 통한 보다 효율적인 마케팅 전략의 수립이 가능할 수 있도록 세부적인 분석이 필요하다.

참고 문헌

- [1] 알렉스버슨, *CRM을 위한 데이터마이닝*, 대청미디어, 2000.
- [2] 진서훈, *CRM과 데이터마이닝*, 교우사, 2005.
- [3] 김연형, *고객관리와 데이터마이닝*, 교우사, 2006.
- [4] M. Rob, *Data Warehousing and Data Mining for Telecommunications*, Artech House, 1987.
- [5] D. Pyle, "Data Preparation for Data Mining", Morgan Kaufmann Publishers, 1999.
- [6] H. Lan, "Data Mining, Morgan Kaufmann Publishers," 1999.
- [7] 최종후, *AnswerTreefmf 이용한 데이터 마이닝 의사결정나무분석*, SPSS 아카데미, 1998.
- [8] 강현철, "Enterprise Miner의 의사결정나무분석 알고리즘", SAS 사용자 컨퍼런스 발표자료집, SAS Korea, pp169-186, 1998.
- [9] J. Ross Quinlan, "Simplifying Decision Trees," Int. J.Man-Machine Studies, pp221-234, 1987.
- [10] J. Ross Quinlan, "Induction of Decision Trees," Machine Learning, pp.81-106, 1986.
- [11] J. Ross Quinlan, "C4.5:Programs for machine Learning," Morgan Kaufman, 1993.
- [12] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," In Proceedings of the 20th VLDB Conference, 1994(9).

저 자 소 개

이 병 엽(Byoung-Yup Lee)

정회원



- 1991년 2월 : 한국과학기술원 전산학과(공학사)
- 1993년 2월 : 한국과학기술원 전산학과(공학석사)
- 1997년 2월 : 한국과학기술원 경영정보공학(공학박사)

- 1993년 1월 ~ 2003년 2월 : 대우정보시스템 차장
- 2003년 3월 ~ 현재 : 배재대학교 전자상거래학과 부교수

<관심분야> : XML, 지능정보시스템, 데이터베이스 시스템, 전자상거래학

임 승 빈(Seung-Bin Yim)

정회원



- 1998년 2월 : 중앙대학교 행정학과(행정학사)
- 2000년 2월 : 중앙대학교 행정학과(정책학 석사)
- 2002년 4월 ~ 2007년 8월 : (주) 테이솔브 영업기획실 실장

- 2007년 8월 ~ 2008년 5월 : (주)네오플러스 SI 전략 기획팀장
- 2008년 6월 ~ 현재 : (주)하나아이엔에스 경영기획팀 신사업추진 담당 과장
- 2009년 3월 ~ 현재 : 배재대학교 전자상거래학과 박사과정

<관심분야> : 데이터베이스 시스템, 시스템다이내믹스, 지식생태계

박 용 훈(Yong-Hoon Park)

정회원



- 2005년 : 호원대학교 정보통신공학과 및 건축공학과(공학사)
- 2007년 : 충북대학교 정보통신공학과(공학석사)
- 2007년 ~ 현재 : 충북대학교 정보통신공학과 박사과정

<관심분야> : 데이터베이스 시스템, 정보검색, 시공간 데이터베이스, 센서 네트워크 및 RFID

유 재 수(Jae-Soo Yoo)

중신회원



- 1989년 : 전북대학교 컴퓨터공학과(공학사)
- 1991년 : 한국과학기술원 전산학과(공학석사)
- 1995년 : 한국과학기술원 전산학과(공학박사)

- 1995년 ~ 1996년 8월 : 목포대학교 전산통계학과 전임강사
- 1996년 8월 ~ 현재 : 충북대학교 전기전자컴퓨터공학부 부교수

<관심분야> : 데이터베이스 시스템, XML, 멀티미디어 데이터베이스, 분산 객체 컴퓨팅