

의사결정트리를 통한 자동차산업의 구매패턴분류

Buying Customer Classification in Automotive Corporation with Decision Tree

이병엽*, 박용훈**, 유재수**

배재대학교 전자상거래학과*, 충북대학교 전기전자컴퓨터공학부**

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

요약

일반적으로 데이터 마이닝은 다양한 예측기법이나 차이점의 분석을 통하여 유용한 정보 도출을 통해 매출의 증대나, 비용 절감 등의 효과를 가져 올수 있다. 데이터 마이닝 분석은 정보의 분류 또는 다양한 각도나 영역의 관점을 통하여 새로운 뷰를 분석할 수 있다. 기술적으로 데이터 마이닝의 연관규칙이나, 패턴의 분석은 대량의 데이터베이스에서 분석된다. 따라서 데이터 마이닝은 가장 빠르게 성장한 산업중의 하나이다. 그 이유는 정보기술이 발달하면서 수많은 자료들이 체계화된 데이터베이스에 저장이 되고, 기업의 데이터베이스의 규모는 폭발적으로 증가되고 있다. 본 논문에서는 자동차 마케팅에서 이용 가능한 데이터를 데이터 마이닝 분석 기법중의 하나인 Quinlan의 C4.5에 따라 분석 하였다. 본 논문에서는 기존 고객에 대한 고객 속성파악, 고객 분류 및 분석에 따른 고객의 구매패턴을 분석하여 해당 기업의 전략적 마케팅 수립을 통해 경영 과학적으로 접근할 수 있는 데이터 마이닝 분석에 관한 사례 연구이다.

■ 중심어 : | 데이터 마이닝 | 의사결정트리 | CRM마케팅 |

Abstract

Generally, data mining is the process of analyzing data from different perspectives and summarizing it into useful information that can be used to increase revenue, cuts costs, or both. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Data mining is one of the fastest growing field in the computer industry. Because of According to computer technology has been improving, Massive customer data has stored in database. Using this massive data, decision maker can extract the useful information to make a valuable plan with data mining. Data mining offers service providers great opportunities to get closer to customer. Data mining doesn't always require the latest technology, but it does require a magic eye that looks beyond the obvious to find and use the hidden knowledge to drive marketing strategies. Automotive market face an explosion of data arising from customer but a rate of increasing customer is getting lower. therefore, we need to determine which customer are profitable clients whom you wish to hold. This paper builds model of customer loyalty detection and analyzes customer buying patterns in automotive market with data mining using decision tree as a quinlan C4.5 and basic statics methods.

■ keyword : | Data Mining | Classification | Decision Tree |

1. 서론

데이터웨어하우징의 급속한 확산 및 이용 목적의 고급화에 따라 데이터웨어하우징과는 독립적인 특정 목적을 위해 데이터 소스의 형태에 관계없이 일반 데이터베이스나 텍스트를 대상으로 하는 지식 추출이 활성화 되고 있고, 또한 기업의 ERP시스템의 정착과 더불어 ERP로부터의 방대한 데이터에 대한 분석의 필요성이 증대되고 있다. 이에 데이터웨어하우징 기술만으로는 향후 고객의 욕구를 충족시키기 어려울 뿐만 아니라 기업들간의 경쟁 심화로 인해 데이터 마이닝의 중요성에 대한 인식이 확대되고 활용은 극대화 되고 있다. 또한 데이터 마이닝은 정보기술의 발달과 더불어 발전되어 왔다. 특히 데이터베이스기술의 발달과 데이터웨어하우징 기법, 인공지능의 전문가 시스템과 기계학습 등이 데이터 마이닝을 발달시키는데 주요한 요인을 제공하였다. 기업경영에서 마케팅 정보에 대한 요구와 기대가 달라진 것도 데이터 마이닝을 확산시키는데 중요한 요인이라고 볼 수 있다.

데이터 마이닝은 현재 광범위한 영역에서 활용되고 있으며, 또한 다양한 부문에 구축되어 있는 자료들을 통합해서 보다 상위수준의 전략적 지식을 도출하기 위해서도 많이 이용되고 있다. 현재는 비단 기업에서 뿐 아니라, 공공기관 등 보다 복잡한 정보분석이 요구되는 기관에서는 예외 없이 데이터 마이닝을 활용하고 있다. 더욱이 인터넷상의 트랜잭션이 늘어나면서, 인터넷상에서 사용자의 행동특성을 분석하는 웹마이닝과 자료검색 기능을 보다 확장한 텍스트 마이닝 기법 등을 통해, e-비즈니스를 위한 제반 마케팅 정보와 소비자 행동 모형들이 만들어지고 있다. 이런 추세라면 데이터 마이닝 기법과 그 분야는 앞으로도 계속 확장될 전망이다. 또한 최근 데이터 마이닝에 대한 많은 연구가 이루어졌는데, 그동안 제한된 다양한 데이터 마이닝 기법들은 어떤 형태의 지식을 발견하고자 하는가, 어떤 종류의 데이터베이스에 적용될 수 있는가, 어떤 분야의 기술에 바탕을 두고 있는가 등의 기준에 의하여 데이터 집합의 일반적 특성을 분석하는 특성화, 특정 인자에 의해서 구분되는 데이터베이스 상의 다른 클래스에 대한 차별적인 특성을 추출하는 분류화, 사전정보가 주어지지 않은 상태에서 데이터베이스

내에서 유사한 특성을 갖는 데이터들을 묶는 군집화, 여러 개의 트랜잭션들 중에서 동시에 발생하는 트랜잭션의 연관관계를 발견하는 연관규칙탐사, 시계열 데이터(주식, 물가, 판매량, 과학적 실험 데이터)들이 시간 축으로 변하는 전개과정을 특성화하여 동적으로 변화하는 데이터의 분석을 수행하는 경향 분석, 대용량 데이터베이스 내의 명시된 패턴을 찾는 패턴 분석 등으로 분류 될 수 있다. 데이터는 의사결정에 필요한 새롭고 가치 있는 정보와 지식을 획득할 수 있는 잠재적인 원천으로 인정되고 있다[1][2]. 기업들은 급변하는 경영환경에서 기업의 경쟁력을 강화하기 위해서 축적된 데이터를 분석하고 정보와 지식을 획득할 수 있는 능력과 정보기술을 보유해야 한다[3][4].

본 논문에서는 이러한 데이터 마이닝의 기술을 이용해 완성자동차 업체의 설문데이터를 토대로 자동차 구매 고객의 패턴을 분석하고자 한다. 2002년 이후 수출이 큰폭의 증가세를 유지, 이는 국산차의 제품경쟁력 향상과 브랜드 인지도의 상승, 업계의 적극적인 수출시장 개척에 따른 결과이며, 최근 3~4년간 내수 부분에서의 부진이 지속되고 있는 것은 고유가로 인한 차량 유지비의 상승, 경기부진에 따른 소비심리위축, 고용불안정 등의 요인을 반영으로 반영되고 있다.

최근 글로벌 경쟁 심화와 업계판도의 변화, 시장구조의 큰 변화가 예상되지만 향후 내수는 경기둔화, 고용불안정, 고유가 지속 등의 요인들이 크게 개선되기 어려울 것으로 보이나, 차량 노후화에 따른 대체 수요 발생, 신차 효과 등이 긍정적인 요인으로 작용할 것으로 예상된다. 하지만 자동차 산업은 한미 FTA 및 여러 가지 어려운 글로벌한 시장과 더불어 자동차 생산 및 내수의 보급이 급격하게 증가 되고 있는 추세이고, 현재 우리나라 자동차 보급률은 IMF의환위기 이전 150만대 수준이던 내수시장은 외환위기 이후 '02년 160만대 수준까지 증가 하였으나 최근 고유가 등으로 인한 내수 시장 침체로 인하여 부진한 상황을 면치 못하고 있다. 반면 자동차 업계는 기업간 보유고객의 수가 늘어남에 따라 고객 데이터와 정보의 양도 기하급수적인 증가를 보여 왔으며 이러한 데이터와 정보의 증가는 데이터 활용에 대한 방법과 그 영역의 증대를 가져왔다[5]. 따라서 자동차 산업의 경

쟁방식은 기존고객의 유지, 서비스의 차별화와 고도화를 통해 이루어져야 하며 신규고객 1명 유치에 드는 비용은 기존고객 1명을 유지하는데 드는 비용보다 4-6배 더 들게 되는 것을 감안한다면 양적 성장인 신규고객 유치보다는 보유고객에 대한 다각적인 분석에 따른 효율적인 마케팅 방법 적용으로 심화되는 자동차 완성 업체간의 경쟁에서 우위를 차지할 수 있을 것이다[6]. 이에 본 논문은 데이터 마이닝을 통한 고객 분석을 통하여 자동차 산업의 고객 마케팅의 좋은 분석 방법으로 활용하기를 기대 한다.

본 논문에서는 데이터마이닝 방법론 중의 하나인 Quinlan의 의사결정 트리 C4.5을 이용하여 기존고객을 세분화한 다음 분류고객의 특성에 맞는 마케팅 프로모션을 하려고 하고 신규고객을 획득할 때는 신규고객의 특성을 미리 예측하여 분류함으로써 고객의 평생가치(Life Time Value)를 촉진하여 기업과 고객과의 관계성을 높여서 기업은 안정된 고객층으로부터 수익을 창출하고 고객들은 해당 기업으로부터 더 많은 혜택을 받게 하는데 목적이 있다. 그러므로 본 논문에서 제시하고자 하는 것은 기존 구매 고객에 대한 분류규칙, 및 고객속성 파악, 기존 고객 분류에 따른 신규고객의 분류 예측 두 가지를 수행하여 해당 기업의 마케팅 전략수립에서부터 경영 과학적으로 접근할 수 있는 데이터마이닝 분석 방법을 제시 하고자 한다[1].

II. 데이터마이닝(Data Mining)의 활용

국내 자동차 시장은 크게 4가지측면에서 예측가능하다. 첫째, 글로벌 경쟁 심화, 시장구조의 변화가 예상되고 이는 미국 빅3의 비중축소와 일본업체와 아시아 업체의 약진이 예상되며, 미국 유럽, 일본 등 선진시장의 비중축소와 중국, 인도, 러시아 등 신흥시장의 확대가 예상된다. 또한 최근 고유가의 지속으로 중대형차에서 소형차로 무게중심이 상당폭 이동 예측이 가능하며 향후 내수는 경기둔화, 고용불안정, 고유가의 지속 등의 요인들이 크게 개선되기 어려울 것으로 보이나 국산 자동차의 품질 및 브랜드 이미지 상승, 판매망 확대, 해외선진

업체의 판매망 활용 등의 요인이 긍정적으로 작용할 것으로 예측할 수 있다. 이에 마케팅의 측면에서 고객전략을 정의하고, 경영자원의 분배를 최적화하며, 고객과 관련된 모든 부분에서 최상의 서비스를 제공하기 위해, 고객의 욕구와 행동, 고객의 수익성을 규명해 가는, 지속적이고 자동화된 프로세스로서 CRM은 고객과 관련된 기업의 내외부 자료를 분석, 통합하여 고객특성에 기초한 마케팅활동을 계획하고, 지원하며, 평가하는 과정이다. 또한 CRM은 데이터 마이닝을 통해 고객데이터의 세분화를 실시하여 신규고객획득, 우수고객 유지, 고객가치증진, 잠재고객 활성화, 평생 고객화와 같은 사이클을 통하여 고객을 적극적으로 관리하고 유도하며 고객의 가치를 극대화 시킬 수 있는 전략을 통하여 마케팅을 실시한다. 이러한 CRM의 정의로 볼 때 CRM에서의 데이터 마이닝은 필수 불가결한 도구라고 볼 수 있다[7][8].

최근 비즈니스 세계에서, 고객중심의 시각에서 모든 것을 바라보며, 그것에 맞추어 경영 전략을 수립하는 일들이 이제는 거의 상식의 관점에서 이루어지고 있다. 이러한 관점에서 비즈니스의 당위 목적들이 제품 또는 서비스의 단순한 판매에만 있는 것이 아니라 고객과 관계를 맺어가는 과정임을 인식하는 것이라고 볼 수 있다. 이 CRM은 크게 2가지 영역의 형태로 나누어 볼 수 있다. 첫 번째는 기능적 CRM, 다음은 분석적 CRM으로 나누어 볼 수 있다[7][8]. 기능적 CRM을 요약하면 비즈니스에서 고객과의 관계를 기능적인 단위 측면에서 초점을 맞춘 경우를 의미한다. 예를 들면 모든 고객들의 접점에서 어떤 유용한 특정 고객의 상호 작용에 대한 정보만을 생성하여 이를 활용하는 경우를 말한다. 분석적 CRM은 개인적인 측면과 전체적인 측면에서 고객의 이해를 더욱 더 높여줄 수 있는 정보를 제공한다[9].

따라서 자동차산업에서 데이터 마이닝과 CRM의 활용을 요약 하면 다음과 같다. 첫째, 고객 세분화와 고객 가치의 이해증대로 더 많은 고객의 획득이 예상된다. 특정한 고객과 집단의 분류와 가치의 극대화는 타겟 마케팅의 캠페인의 목표가 될 수 있다. 둘째, 고객 세분화와 신규 자동차 구매간의 관계에 대한 더 나은 이해로 교차 판매의 효율이 증대 될 수 있다. 이러한 요약된 개념들은 데이터 마이닝 또는 분석적 CRM의 기법에서 얻어진

결과로서 운영 CRM에 의해 활용 될 수 있다[7].

데이터 마이닝은 각각의 산업의 특성에 따라 표준화된 분석방법과 사례들이 있다. 본 논문에서 제시하는 데이터 마이닝은 분석자들이 새로운 모형을 구축하는데 소요되는 노력을 최소화 할 수 있고, 더 나아가 기존 고객 및 신규 고객관리에 측면에서 마케팅 자료로 활용 될 수 있다.

III. 분석의 절차 및 변수 선정

3.1 연구 절차와 방법

국내 완성자동차 업체인 A사의 자료를 입수하여 연구하였고 수행하는 연구의 진행절차는 데이터 마이닝 프로세스의 각 단계별 과정에 따라 진행을 하고자 하며 과정은 첫째 고객자료 수집 및 통합, 둘째 수집 자료에 대한 추출 및 정제 작업, 셋째 변수의 선정, 넷째, 분류에 의한 고객 그룹속성 분석 및 고객그룹에 대한 분류규칙의 분석 순서로 진행 되었다.

3.2 자료의 구성 및 기초 통계

본 연구에서 사용한 데이터는 국내 완성형자동차 업체의 샘플자료로서 소비자구매 행동에 대한 일반적인 데이터 즉 구매시 자동차 비교 여부, 구매시 비교차종, 비교사유, 현소유차 구매 사유 등이며, 자동차에 대한 만족도를 위한 항목은 전반적인 만족도, 소유차 제조회사 이미지 개선도, 이미지 적합도, 개인 속성별 만족도, 연비/유지비 평가 등이고 재구매 측정을 위한 항목은 소유차 재구매 의도, 소유차 제조회사 재구매 의도로 구성되었으며 과거 신차 구입자를 대상으로 이루어진 설문은 대상으로 하였다. 고객에서 샘플링을 하여 선정하였고, 조사 결과중 null 값은 평균값으로 일괄 처리, 또는 데이터 마이닝 알고리즘에 적합한 값으로 변환 하였다. 따라서 대체 평균값이나, 기타 값들에 의해서 추출된 데이터 마이닝 룰은 분석/해석 대상에서 제외를 하였다. 반면 무응답의 경우도 일정한 패턴의 규칙성을 가질 수 있고, 의미 있는 결과를 가질 수도 있지만 분석에서는 일괄적으로 무응답인 경우는 룰 생성에서 제외하였다. 최초 원시 데이터로부터 고객의 기본정보의 데이터 셋, 고객 성향들의 유의한 값을 가진 변수들의 집합, 그리고 이 두 데이터를 합한 데이터를 최종 분류규칙 데이터 마이닝 모델에 적용하였다.

데이터 마이닝 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. 잠정적으로 선택된 많은 변수 중에서 원하는 모형에 가장 적합한 변수를 선정해야 하는데 선정방법에는 대표적으로 입력변수와 목표변수간의 결정계수를 이용하는 방법과 카이제곱 통계량을 이용하는 방법이 있다[10]. 또한 기초통계 분석을 함으로써 원시데이터 변수들 중 유의한 변수들을 정의하고 효과적이고 이상적인 마이닝 분석 결과를 얻기 위한 과정으로서 주로 분석한 통계분석은 평균값, 최빈값, 최소값, 최대값 등을 분석하였다. 또한 차종을 차량 급별로 분류 하였다. 예를 들면 경소형, 소형, 준준형, 중형, RV, 스포츠, 기타 등으로 재분류 하였다.

가계소득에 관한 평균값을 살펴보면 평균 328만원 가량의 총 가계소득을 살펴볼 수 있고, 평균연령은 30대, 20대의 순으로 조사되었고, 가족수는 4명의 가족구성원이 가능 높은 평균값의 분포를 보였으며 직업으로는 전문직, 공무원, 자영업/회사원의 순으로 조사되었다.

가계소득에 관한 평균값을 살펴보면 평균 328만원 가량의 총 가계소득을 살펴볼 수 있고, 평균연령은 30대, 20대의 순으로 조사되었고, 가족수는 4명의 가족구성원이 가능 높은 평균값의 분포를 보였으며 직업으로는 전문직, 공무원, 자영업/회사원의 순으로 조사되었다.

3.2.1 유의변수 도출

유의한 변수를 도출하기 위하여 선택된 변수들의 대환 상관관계를 알아보기 위하여 분석한 결과 [표 1]의 결과를 볼 수 있다.

표 1. 변수의 Correlation

변수	Correlation Coeff	변수의수
차명 VS 마지막고려차종	0.7252	2385
고려차종 VS 마지막고려차종	0.92082	2385
소유차종 VS 고려차종	0.75543	2385
소유차종 VS 마지막고려차종	0.72874	2385

마지막고려차종 VS 소유차종	0.72874	2385
조작용이성 VS 조작편의성	0.79233	3159
가족총소득 VS 생활비	0.70127	3078
연령 VS 가족형태	0.77786	3159

※ Prob > 0.0001, Correlation Coeff < 0.7

[표 1]의 평균값과 상관관계(Correlation) 분석을 통하여 간단히 정리되기 어려운 변수들이 있었지만 중요 변수들과 통계분석 결과 얻어낸 유의한 변수들을 정리하면 [표 2]와 같다.

표 2. 산출된 유의 생성 변수

구분	변수
고객배경정보	성별, 연령, 가족내에서의 위치, 가족형태, 직업, 회사 규모, 집소유, 가족수, 결혼, 맞벌이, 신문구독, 월소득, 가족총소득, 생활비
구매 경향	차명, 배기량, 차량타입, 변속기, 새 차 구입이유, 차량 구입시 가장 중요하게 생각하는점, 차량 선택시 의사결정, 구매비교차종, 현재소유차종, 고려차종, 마지막고려차종, 마지막고려 이유, 만족도, 회사이미지, 안전성, 성능, 소음, 외부 디자인, 내부디자인, 조작편의성, 용도(출퇴근, 등하교, 쇼핑용, 데이트, 업무/집대, 여행용, 종교활동, 친구/친척방문, 기타)

IV. 문제정의 및 모델링 분석

문제정의의 과정에서 데이터의 특성을 나열하고 데이터 마이닝의 문제정의의 제약 조건을 정의 하면 다음과 같다. 첫째, 데이터 마이닝의 문제를 해결하는 가장 일반적인 5가지(Regression, Association, C45, Episode, Sequence) 알고리즘이 있다. 이러한 각각의 알고리즘별로 데이터의 일정한 형식의 타입과 변수들이 필요한데, 현재 분석을 하기위해 보유하고 있는 로우 데이터의 특성을 살펴보면 고객의 의견이나 생각을 묻는 설문조사 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 물들을 얻어내기가 어렵다. 따라서 제한 조건을 벗어난 데이터 마이닝의 문제점들을 도출한 결과 분류규칙을 통한 고객의 성향 분석, 전체적

인 데이터의 성격들의 결과를 도출할 수 있었다.

4.1 의사결정트리 문제 정의

의사결정 트리는 매우 복잡한 구조를 가진 데이터들 사이의 관계나 패턴을 찾아내는 유연한 비선형 모형의 하나로, 신경생리학과 유사성 때문에 일반적으로 다른 (통계적) 예측모형에 비해 또 다른 물들을 예측할 수 있다. 이러한 신경망 모형은 고객의 신용평가, 불량거래의 색출, 의료진단예측, 우량고객의 선정 타겟 마케팅의 여러 주제 등을 비롯한 다양한 분야에 적용될 수가 있는데, 주로 교사학습에 적용되어 목적변수(target)에 대한 예측이나 분류를 목적으로 감춰진 패턴을 찾고 이를 일반화하는데 이용된다. 따라서 본 논문의 데이터를 분석한 결과 목적변수를 기준으로 우량고객의 선정, 타겟 마케팅 등에 적용할 수 있다. 결과의 해석은 [그림 1]에 제시된 바와 같이 유의한 변수의 규칙에 의한 의사결정 모형의 트리구조로 해석될 수 있다.

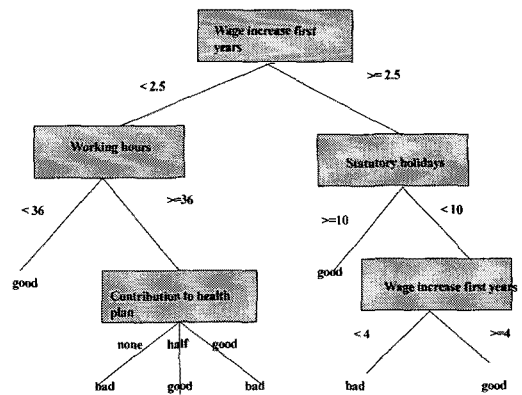


그림 1 C45 의사결정 모델 해석

따라서 본 논문에서 분석된 문제정의는 첫째, 어떤 고객들이 어떤 차량을 구입하는가? 둘째, 경소형차와 소형차 구매고객은 어떤 유의한 패턴의 차이를 보이는가?의 문제를 정의 하였고, 분석의 과정에서 특이하게 도출된 타겟변수들을 재선정하여 세부적인 분류 규칙을 분석 하였다.

4.2 의사결정트리분석 및 결과

정제된 데이터를 가지고 Quinlan의 C4.5 의사결정트리를 적용해본 결과를 정리하면 다음과 같다. 데이터 마이닝 분석 과정은 먼저 앞서 정의한 데이터들을 각각 분석하고 유의한 변수들과 타겟 변수들과의 마이닝을 통해 얻어진 결과이다.

Quinlan의 C4.5의 분석절차는 적절한 Tree Parameters들과 C4.5의 Parameters들의 적절한 조건이 주어지지 않으면 적게는 몇 개에서 많게는 몇 십개 이상의 과생변수들이 생성되어 의사결정트리의 적절한 파라미터의 조건과 이에 맞는 변수들의 선정이 가장 중요한 문제이다. 따라서 본 알고리즘을 적용 하기전 선행된 분석에 맞는 변수들의 통계적인 수준을 분석하고 이에 맞는 변수들을 선정하여 의사결정트리에 적용하였다.

(1) 어떤고객이 어떤 차량을 구입하였는가의 분류

위의 분류규칙을 생성하기 위해서 의사결정트리 규칙을 적용하기 위해서 앞서 분석된 통계적인 유의한 변수들중에 독립변수로는 성별, 연령, 가족형태, 직업, 가족수, 결혼여부, 맞벌이 여부, 신문구독, 월소득이 사용되었고, 종속변수로는 문제정의의 분류 타겟인 차종에 대한 데이터로 정의 하였고, MINOBS의 수는 30, CF 0.25로 설정하였다.

[표 3]의 분석결과를 타겟마케팅 측면에서 분류된 규칙을 살펴보면 월소득 260만원 이상인 가정에서 중형차의 구입 가능성이 높음을 알 수 있다. 또한 맞벌이를 하지 않는 월소득 450만원이상인 고객 분류층에서는 대형차를 선호하는 것으로 분석되었고, 38세 이상의 연령대에서 대형차를 주로 구매하는 패턴을 분류할 수 있었다.

따라서 좀더 세부적인 분석을 위해서 분류타겟을 좀더 세분화 하여 종속변수를 대형, 중형으로 구분하여 분석을 하였다. [표 4]은 종속변수를 “대형”으로 설정하여 분석하였다.

[표 5]은 종속변수를 중형으로 선정하여 분석한 결과 중형차를 구매하는 몇가지 주요한 분류 규칙을 발견한수 있었다. 월소득 260만원 이상의 40대 이하의 주부그룹에서 신뢰도 61.82%의 신뢰도를 가지고 중형차 구입의 분류 규칙이 도출되었다. 또한 31세 이상의 신혼부부중

260만원보다 적거나 같은 급여수준을 가지고 있는 집단에서 중형차를 구입하는 패턴을 발견할 수 있었다. 분석결과를 요약하여 보면, 주부 집단이나, 신혼부부집단은 우리가 알고 있는 일반적인 소형차 위주의 구매패턴에서 벗어나 삶의 질이 고급화 되어가고 젊은 층의 의식변화에 따른 소형차 보다는 중형차의 구매 패턴을 반영한 결과라고 예측되고, 이러한 분석의 패턴을 기반으로 중형차 판매의 마케팅을 보다 새로운 계층, 신혼부부와 주부들을 대상으로한 마케팅 활동도 보다 좋은 판매 전략이 될 것이라고 예측할 수 있다.

표 3. 고객 분류 구매 패턴

Tree Node/Leaf	Target	신뢰도(%)
월소득 <= 260	중형	44.06
월소득 > 260	-	-
직업 = 공무원	대형	39.62
직업 = 기술숙련직	중형	36.52
직업 = 자영업	중형	52.15
직업 = 주부	중형	54.18
직업 = 무응답	기타	25.00
직업 = 경영/회사원	-	-
맞벌이 = O	대형	45.56
맞벌이 = 무응답	중형	32.30
맞벌이 = X	-	-
월소득 <= 450	중형	37.27
월소득 > 450	대형	64.31
직업 = 전문직	-	-
가족수 <= 3	중형	43.83
가족수 > 3	대형	50.82
직업 = 기타	-	-
연령 <= 38	중형	42.87
연령 > 38	대형	53.71

(2) 고객 자동차 구입시 제조사 선택에 있어 어떤 유의한 패턴의 차이를 보이는가?

위의 분류규칙을 생성하기 위해서 의사결정트리 규칙을 적용하기 위해서 앞서 분석된 통계적인 유의한 변수들중에 독립변수로는 성별, 연령, 가족형태, 직업, 가족수, 결혼여부, 맞벌이 여부, 신문구독, 월소득이 사용되었고, 종속변수로는 문제정의의 분류 타겟인 제조사에 대한 데이터로 정의 하였고, MINOBS의 수는 50, CF 0.25로 설정하였다.

[표 6]의 결과를 요약하여 특이한 분류를 정리하면 가족의형태로는 자녀성장기에 있는 가정이나, 자녀독립기에 있는 가족의 형태에서 경소형 자동차를 구매하는 패턴을 발견할 수 있다. 이는 자녀의 성장, 독립과 발마추

어 1가구 2차량의 가정이라고 분석되어진다. 표의 분석된 결과의 분류는 대부분 소형, 경소형 차량으로 분류구획이 형성되어 경소형과 소형 차량을 독립변수로 설정하여 세부분석을 실시하였다.

표 4. “대형” 분류 규칙

직업	가족형태	성별	나이	월소득	가족수	맞벌이	신뢰도(%)
경영/회사원	-	남	>450	-	X		64.3
공무원	자녀독립기	남	>260	-	-		63.8
기타	-	남	>44	>260	-	-	61.6
전문직	-	남	>37	>260	>3	-	57.4
기타	-	남	>38	>260	-	-	53.1
전문직	-	남	>260	>3	-		50.8

표 5. “중형” 분류 규칙

직업	가족형태	나이	월소득	맞벌이	신뢰도
주부	-	(=40)	>260	-	61.82
회사원	신혼부부	>31	(=260	-	61.23
-	자녀성장기	>31	(=260	X	59.01
기술숙련직	-	>31	(=260	-	58.40
기타	가족형성기	>31	-	-	56.45
주부	-	>31	-	-	56.71
-	신혼부부	>31	(=260	-	56.27
자영업	-	>31	>175 (=260	-	56.36
기타	-	>31	-	-	56.45
주부	-	>31	-	-	56.27

[표 7]는 경소형 자동차를 구매하는 집단 분류의 패턴을 살펴보면 가족의 형태로는 미혼/독신, 신혼부부 등의 형태로 경소형 자동차의 구매패턴을 볼 수 있고, 연령별 분류를 살펴보면 26세 이상, 44세 이상의 상이한 패턴분류가 보이지만 26세 이상의 집단은 신규 차량의 구입이라고 예측되고 44세 이상의 집단은 자녀성장기 및 가족형성기의 특징에서 1가구 2차량의 구매 패턴으로 예측할 수 있다.

표 6. 제조사 선택 구매 유의 패턴

Tree Node/Leaf	Target	신뢰도(%)
가족형태 = 신혼부부	소형	56.39
가족형태 = 자녀성장기	경소형	67.78
가족형태 = 자녀독립기	경소형	55.78
가족형태 = 미혼/독신	-	-
직업 = 기타	소형	62.20
직업 = 기술숙련직	경소형	45.27
직업 = 자영업	경소형	50.72
직업 = 공무원	-	-
연령 (= 26.0	경소형	44.38
연령 > 26.0	소형	67.35
직업 = 전문직	-	-
연령 (= 23.0	경소형	67.49
연령 > 23.0	소형	56.27
직업 = 경영/회사원	-	-
가족수 (= 2.0	경소형	64.52
가족수 > 2.0	소형	47.05
가족형태 = 자녀성인기	-	-
연령 > 48.0	경소형	61.29
연령 (= 48.0	-	-
연령 (<= 44.0	경소형	44.24
연령 > 44.0	소형	64.52
가족형태 = 가족형성기	-	-
직업 = 공무원	경소형	51.46
직업 = 전문직	경소형	62.58
직업 = 기타	소형	59.40
직업 = 기술숙련직	경소형	61.23
직업 = 자영업	경소형	35.21
직업 = 주부	경소형	51.20

[표 8]는 소형 자동차를 중속변수로 설정하여 분석된 결과이다. 분석된 결과를 정리하면 주로 소형 자동차를 구매하는 패턴은 직업은 전문직에 대부분 자녀성장기에 있는 가정에서 소형자동차를 구매하는 분류 규칙을 볼 수 있다. 따라서 소형 자동차의 마케팅 패턴에서 젊은 층을 대상으로 한 마케팅의 범위를 넓혀서 전문직에 종사하는 자녀성장기에 있는 고객들을 중심으로 소형 자동차의 판매 전략을 수립할 수 있다.

표 7. “경소형” 자동차를 타겟으로 분류

직업	가족형태	성별	나이	월소득	가족수	신뢰도(%)
기타	미혼/독신	-	-	>135	-	74.03
-	신혼부부	-	-	-	>2	73.10
전문직	미혼/독신	남	>26	-	-	70.01
공무원	미혼/독신	-	>26	-	-	67.35
주부	미혼/독신	-	-	-	-	66.25
-	신혼부부	-	-	(= 125	(=2	64.52
-	자녀성인기	-	>44	-	-	64.52
경영/회사원	가족형성기	-	>44	-	-	63.85
기타	미혼/독신	-	-	-	-	62.20
전문직	미혼/독신	남	>23	-	-	62.05
기타	-	-	-	-	-	60.76

표 8. “소형” 자동차를 타겟으로 분류

직업	가족형태	성별	나이	월소득	가족수	신뢰도
전문직	자녀성장기	-	>37 <=40	-	-	88.16
전문직	자녀성장기	-	>37	-	-	73.77
	자녀성장기	-	<=35	-	-	71.85
전문직	가족형성기	-	>30	-	-	70.43
공무원	자녀성장기	-	-	-	-	70.21
전문직	자녀성장기	-	-	-	-	69.62
-	-	-	>48	-	-	68.65
-	자녀성장기	-	-	-	-	67.78
전문직	미혼/독신	-	<=23	-	-	67.49
경영/회사원	미혼/독신	-	-	-	<=2	64.52
전문직	가족형성기	-	-	-	-	62.58
전문직	-	여	-	-	-	62.31
-	-	-	>30	-	-	61.24
기술숙련직	가족형성기	-	-	-	-	61.23

V. 결론

완성 자동차 시장에서 기업간 경쟁이 더욱 심화되고 최근 글로벌 경쟁 환경에서 급격한 내수 감소 추세를 나타냄에 따라 기업이 마케팅 활동을 수행함에 있어 기존 고객의 유지가 중요한 과제가 되고 있다. 본 논문에서는 국내의 완성자동차 구매 고객을 대상으로 데이터 마이닝을 위한 기법 중 의사결정트리를 이용하여 기존 고객을 분류하였고 그 결과는 다음과 같다.

첫째, 분류규칙에 의한 구매패턴 분석 결과 타겟마케팅 측면에서의 고객 성향 분석을 통하여 몇가지 유의한 결과를 살펴볼 수 있었다. 첫째, 일반적인 매스미디어를 통한 마케팅의 규칙에서 조금 벗어난 몇가지 구매 패턴을 정의할 수 있었다. 예를 들면 중형차 구매의 패턴이 신혼부부나, 주부집단에서도 매우 유의한 구매 패턴을 도출할 수 있었고, 이는 새로운 마케팅에 대한 고객층을 제시할 수 있었다. 둘째, 소형차, 경소형차의 구매패턴을 살펴보면 전문직의 자녀성장기에 있는 그룹에서는 소형자동차의 구매 신뢰도가 상당히 유의하게 도출되었고, 경소형차의 구매 패턴에서도 44세 이상의 집단층에서 경소형차의 구입패턴이 유의하게 도출됨에 따라 소형자동

차 시장에서 이러한 분류 규칙을 적용하여 판매를 증가할 수 있는 유의한 규칙이라고 사료된다.

최근 미국의 빅3 자동차 상의 부도설과 더불어 내수경기의 침체, 유가의 급등으로 인한 여러 가지 사회적인 외면할수 있는 요인들이 발생함에 따라 추후 연구과제로는 사회적인 경제 지표 요인, 즉 유가, 환율 등 대외적인 경제 지표를 반영하여 보다 넓은 다수의 변수 중에서 고객의 특정 속성과의 상관관계 모형을 구축함으로써 구매고객분류 및 패턴과 사회요인의 역학적인 모형을 수립하여 좀더 효율적이고 효과적인 마케팅 전략의 수립이 가능할 수 있도록 세부적인 연구 모델의 수립이 필요하다.

참고문헌

- [1] M. Rob, Data Warehousing and Data Mining for Telecommunications, Artech House, 1987.
- [2] E. Carl and E. Katz, "Customer spending behavior and telecommunications services," Telecommunications Policy, Vol.22, pp.23-46, 1998.
- [3] W. John, "Data Mining Solution: Methods and Tools for Solving Real-World Problems," pp.501-517, 2003.
- [4] 장남식, 홍성완, 장재호, 데이터 마이닝, 대청, 1999.
- [5] <http://www.hri.co.kr/m3/vip/9928/ex2.htm>
- [6] J. P. Bigus, Data Mining with Neural Networks, McGraw-Hill, 1996.
- [7] 허준, 통신시장에서의 데이터 마이닝 Telco-CAT, SPSS아카데미, 2000.
- [8] H. Lan, Data Mining, Morgan Kaufmann Publishers, 1999.
- [9] U. M. Fayyad and G. P. Shapiro, Advanced in Knowledge Discovery and Data Mining, AAAAI Press/MIT Press, 1996.
- [10] 강현철, 한상태, 최종후, 데이터 마이닝 방법론 및 활용, 자유아카데미, 1999.

저자 소개

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, 데이터베이스시스템, 전자상거래학

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)

증신회원



- 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