

## 자동차 재구매 증진을 위한 데이터 마이닝 기반의 맞춤형 전략 개발

이동욱\* · 최근호\*\* · 유동희\*\*\*

I. 서론	IV. 분석 결과 및 맞춤형 전략 제안
II. 문헌연구	4.1 예측모형 구축과 예측률 비교
III. 연구방법	4.2 재구매 규칙에 따른 전략 제안
3.1 연구 진행 과정	V. 결론
3.2 분석 데이터	참고문헌
3.3 목표변수와 독립변수	<Abstract>
3.4 데이터 균형화	
3.5 변수 선택	

### I. 서론

국내 자동차 산업은 1955년 8월, 최초로 수공업형태로 제작된 지프형 자동차인 ‘시발’ 이후로 급격하게 발전하여 오늘날 연간 약 452만 대를 생산하는 세계 자동차 강국의 대열에 합류하였다. 우리나라는 2013년 기준으로 중국, 미국, 일본, 독일에 이어 세계 5위의 자동차 생산국으로 자리매김하고 있으며, 현재 자동차를 제작하는 국내 기업에는 현대자동차, 삼성자동차, 기아자동차, 쌍용자동차, 한국지엠이 있다

(국가지표체계, 2014). 여기에서 2016년 기준으로 국내에 등록된 자동차의 수는 약 2,180만 대에 달하며 이는 대략적으로 자동차 1대당 국민 2.37명에 해당하는 수치이다. 이 중에서 등록된 국내 자동차의 약 2,016만 대(92.5%)가 국산차이며 수입차는 약 164만 대(7.5%)를 차지한다(국가지표체계, 2017). 미성년자나 군인 등 자동차를 소유하기 힘든 인구를 제외하면 국내 자동차 시장은 거의 포화상태에 가깝다고 할 수 있다.

이처럼 오늘날 자동차 산업에서의 상품과 서비스에 대한 경쟁이 심화되면서 자동차에 대한

\* 경상대학교 경영정보학과 석사과정, ldw1391@hanmail.net(주저자)

\*\* 근로복지공단 근로복지연구원 책임연구원, ckh0515@hanmail.net

\*\*\* 경상대학교 경영정보학과 부교수, 경영경제연구소, dhyoo@gnu.ac.kr(교신저자)

재구매의 중요성이 함께 높아지고 있다. 전성은과 정기한(2006)은 경쟁적 환경에서 기업의 성장을 위해서는 기업이 많은 고객들을 유인하는 것보다 방문한 고객을 재방문하게 하는 것과 고객의 지속적인 방문을 이끌어내는 것이 중요하다고 언급하였다. 이와 관련하여 McDougall and Levesque(2000)은 재구매를 소비자가 이용한 서비스를 다시 이용하고 싶은 정도와 타인에게 서비스를 이용하도록 권유하는 정도로 정의하였으며, 한은경 등(2011)은 재구매 의도를 현재의 서비스 제공자를 지속적으로 반복하여 이용할 가능성이라고 언급하였다. 따라서 자동차를 판매하는 기업 입장에서는 신규 고객 유치와 함께 자사의 자동차를 구매한 경험이 있는 고객을 대상으로 자동차의 재구매를 증진시키는 전략을 마련해야 한다.

이를 위해서, 기업은 자동차 재구매에 영향을 주는 요인들을 분석한 후 전략 수립에 활용해야 한다. 이와 같은 요인들을 분석하기 위해서는 1) 자동차를 구매한 경험이 있는 고객을 대상으로 설문조사를 실시하여 재구매 의도에 영향을 주는 요인들을 분석하는 방법과 2) 실제 재구매가 이루어진 고객의 판매 데이터를 분석하여 재구매에 영향을 주는 요인들을 분석하는 방법을 활용할 수 있다. 전자의 경우, 설문조사 항목을 다양하게 구성하여 고객들로 하여금 재구매와 관련된 여러 영향 요인들에 대한 자료 수집이 가능하지만, 실제 재구매를 한 고객이 아닌 재구매 의도를 지닌 고객들로부터 자료가 수집되는 특징이 있다. 후자의 경우, 기업에서 관리하고 있는 재구매 관련 고객 데이터의 종류에 따라 분석할 수 있는 영향 요인들이 제한되지만 실제 재구매가 이루어진 고객의 데이터

를 분석할 수 있기 때문에 분석 결과를 좀 더 실용적으로 활용할 수 있다. 따라서 자동차 기업들은 앞서 언급된 방법들을 활용하여 자사의 상품을 구매한 고객의 성향과 자사의 자동차 특성들을 반영한 맞춤형 자동차 재구매 전략을 마련해야 한다.

본 연구에서는 자동차 재구매 증진을 위한 보다 실용적인 전략을 개발하기 위한 목적으로 실제 재구매가 이루어진 고객의 데이터를 데이터 마이닝 기법을 활용하여 분석하고자 하며, 분석 결과를 통해 자동차 재구매 증진에 도움을 주는 맞춤형 전략을 제안하고자 한다. 이를 위해 국내 자동차 제작사인 S사의 판매 데이터를 이용하여 자동차 재구매에 관한 예측모형을 데이터 마이닝 알고리즘인 의사결정나무(decision tree: DT), 나이브 베이지안(naive Bayesian: NB), 인공신경망(artificial neural network: ANN)을 활용하여 구축하였다. 또한 구축된 예측모형을 통해 자동차 재구매에 관한 10개의 주요 규칙들을 도출하였고, 도출된 규칙을 분석하여 자동차 재구매를 증진시킬 수 있는 맞춤형 전략을 제안하였다.

본 연구의 구성은 다음과 같다. 2장에서는 상품 재구매에 영향을 주는 요인들을 조사한 기존 문헌들에 대해 살펴본다. 3장에서는 본 연구에서 제안한 연구 방법과 데이터 마이닝 기법을 활용하여 자동차의 재구매 여부를 판단하는 예측모형이 구축되는 실험 과정에 대해 설명한다. 4장에서는 연구 결과를 해석하고, 자동차 재구매에 도움을 주는 전략을 제시하고자 한다. 마지막으로 5장에서는 본 연구의 요약과 함께 시사점과 한계점에 대해 다룬다.

## II. 문헌연구

기업이 제품 판매를 통해 얻게 되는 수익을 지속적으로 유지하거나 증진시키기 위해서는 구매 경험이 있는 소비자들을 대상으로 제품의 재구매를 높이는 전략을 실행해야 하며, 제품 재구매에 영향을 주는 요인들에 대한 분석이 이루어져야 한다. 이를 위해, 본 연구에서는 재구매 영향 요인을 분석한 기존 연구들을 자동차 분야와 그 외의 분야로 구분하여 조사하였다.

먼저 자동차 분야에서 재구매 요인을 분석한 연구는 다음과 같다. Oliver and Swan(1989)은 자동차 재구매에 영향을 주는 요인으로 제품 만족도를 언급하였다. 권기대 등(2004)은 자동차의 품질 보증이나 안정성 요인이 자동차 재구매에 영향을 준다고 하였으며, 고객이 소형차를 선택할 때는 자동차의 유지비와 연비를, 중형차를 선택할 때는 자동차의 성능을 주로 고려한다고 언급하였다. 이장로와 김미옥(2010)은 중국 소비자를 대상으로 자동차의 구매 만족도에 영향을 미치는 요인을 분석하였는데, 그 결과 경제성, 기능성, 외형감각성 순으로 영향력이 있음을 파악하였다. 여기에서 중국의 자동차 소비자들은 자동차에 대한 전반적인 만족도가 높아 질 경우 재구매 의도와 함께 추천 의도도 높아지는 것을 알 수 있었다. Ranjbarian et al.(2010)의 연구에서는 합리적 행동이론에 따라 자동차 재구매 의도에 영향을 주는 요인들을 분석했는데 그 결과 주관적 규범이 가장 큰 영향을 주는 것으로 확인되었다. Nistorescu et al.(2013)은 자동차에 관한 고객 충성도를 만족감, 추천, 재구매 영역으로 구분하면서 각 영역

들간의 상관관계를 분석하였는데, 재구매와 추천이 높은 상관관계가 있음을 밝혔다. Balla et al.(2015)은 자동차에 관한 고객의 신뢰, 고객의 몰입 그리고 만족도가 자동차 재구매 의도에 긍정적인 영향을 준다고 하였다. Koay and Derek(2016)은 자동차 품질, 합리적 가격, 고객 만족이 자동차 재구매에 영향을 준다고 설명하였으며, 그 중, 고객 만족은 자동차 품질과 재구매 간의 관계 뿐 아니라, 합리적 가격과 재구매 간의 관계에도 매개변수의 역할을 하는 것으로 분석되었다. 또한 석경하 등(2011)은 렌터카 재구매에 관한 예측 모형을 개발하기 위해 데이터 마이닝 기술을 활용하였으며, 고객성별, 나이, 총 구매횟수, 평균 구매간격, 마지막 거래일자와의 차이를 변수로 활용하였다. 그 결과 신경망 기반의 예측 모형이 재구매 여부 판단에 우수한 성능을 보인 것을 파악하였다.

다음으로 자동차 이외의 분야에서 재구매 요인을 분석한 연구는 다음과 같다. 이낙귀(2003)는 여행 분야에 있어서 평판, 물리적 특성, 커뮤니케이션과 같은 도매업체의 관계적 특성과 전문성, 호감, 유사성, 접촉빈도와 같은 담당영업직원의 관계적 특성이 재구매 의도에 영향을 준다고 언급하였다. 김효경과 신홍철(2010)은 객실, 레스토랑, 부대시설 요인이 호텔의 재방문에 영향을 준다고 하였다. 김은정과 김종원(2012)의 연구에서는 사회적 기업 제품의 지각된 가치가 제품 재구매 의도에 유의한 영향을 준다고 분석하였다. 정태린 등(2012)은 청소년들이 아웃도어 의류를 구매하는데 있어서 구매장소, 광고, 브랜드 순으로 재구매 의사 형성에 영향을 주며, 정태영 등(2015)은 브랜드에 대한 만족이 의료기기의 재구매 의도에 영향을 준다

고 설명하였다. 한국과 중국 소비자를 대상으로 스마트폰 재구매 의도를 조사한 장서흠 등(2015)의 연구에서는 한국 소비자는 사용 의도와 만족도가 중국 소비자는 만족도만이 재구매 의도에 영향을 준다고 하였다. 또한, 조철호(2006)는 전자상거래에서 제품 다양성과 정보 제공성은 초기 신뢰에 영향을 주고, 정보 제공성과 검색 편리성은 고객 만족에 영향을 준다고 언급하였고, 이러한 초기 신뢰와 고객 만족은 전자상거래에서의 재구매 의도에 긍정적인 영향을 미친다고 하였다. 소셜커머스에서 고객의 재구매 의도에 영향을 주는 요인으로 손달호(2015)는 경제적 이익과 사회적 영향을, 김상현과 박현선(2013)은 구전효과, 집단주의, 다양성추구와 같은 소셜집단의 특성을 언급하였다.

지금까지 살펴본 문헌들을 통해 자동차 분야와 자동차 이외 분야 모두 구매한 상품에 대한 만족도가 높을수록 재구매에 긍정적인 영향을 준다고 생각해볼 수 있다. 이 중, 자동차 분야의 경우 고객의 성별, 나이 등과 같은 인구통계학적 변수들과 자동차 브랜드, 품질, 기능, 안정성, 유지비, 연비, 외형의 모양 등과 같은 자동차 특성에 관한 변수들이 자동차 재구매에 영향을 주는 요인들로 분석되었다.

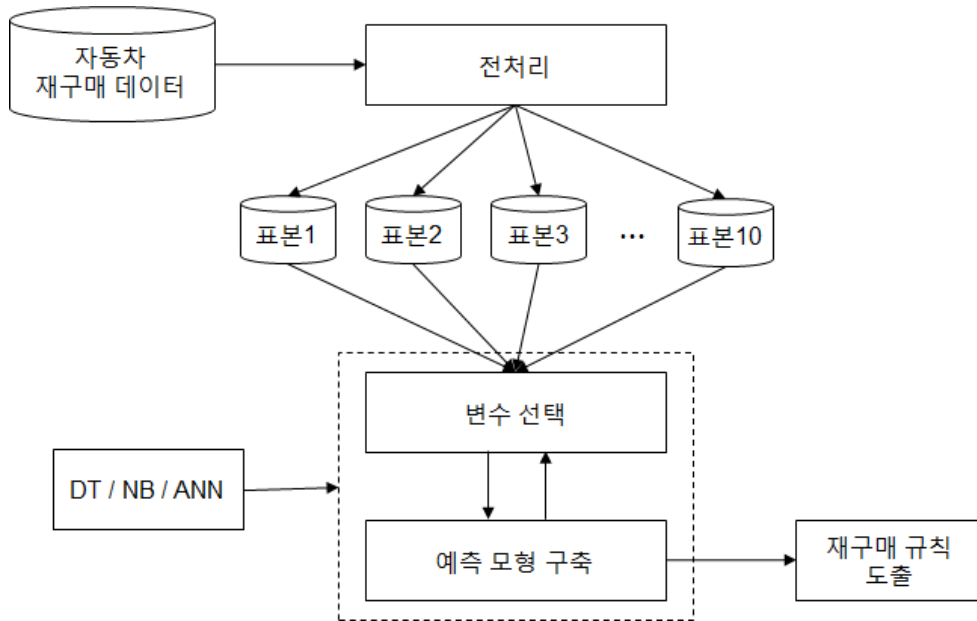
기존 연구와 본 연구와의 차이점을 데이터 수집 및 분석 방법의 관점에서 비교해 보면 다음과 같다. 본 연구에서는 설문조사를 통하여 데이터를 수집하지 않고 실제 자동차의 재구매가 이루어진 고객의 판매 데이터를 수집하여 재구매에 영향을 주는 요인들을 분석하였기 때문에 분석 결과를 좀 더 실용적으로 활용할 수 있다. 그리고 기존 연구에서는 주로 통계적 기법을 활용하여 상품 재구매에 영향을 주는 요

인들의 영향력을 분석하였지만, 본 연구에서는 데이터 마이닝 기법인 의사결정나무, 나이브 베이저언, 인공신경망 알고리즘을 활용하여 자동차 재구매에 관한 예측모형들을 구축하였고, 예측모형을 통해 도출된 10개의 규칙을 이용하여 자동차 재구매에 관한 맞춤형 전략을 제안하였다. 현재 데이터 마이닝 기법을 활용하여 자동차 재구매 여부를 분석한 연구가 많지 않고, 분석 결과를 토대로 재구매 증진 전략을 도출한 연구가 없는 점은 기존 연구와 비교해서 가질 수 있는 본 연구의 새로운 점이라 판단된다.

### Ⅲ. 연구방법

#### 3.1 연구 진행 과정

자동차 재구매에 관한 규칙이 도출되는 전반적인 과정을 요약하면 <그림 1>과 같다. 먼저 자동차 재구매에 관한 고객의 판매 데이터를 수집하였다. 전처리 과정(preprocessing)을 통해 수집된 데이터를 분석 가능한 형태로 변환하고 분석에 사용될 초기 독립변수들을 선별 및 추가하였다. 데이터 균형화(data balancing)를 실시하기 위하여 샘플링을 통한 10개의 표본을 생성하였다. 변수 선택(feature selection) 과정을 통해 목표변수에 영향을 주는 유용한 독립변수들을 최종 선택하였고, 의사결정나무, 인공신경망, 나이브 베이저언 알고리즘을 활용하여 예측모형들을 표본별로 구축하였다. 본 연구에서 사용한 3가지의 데이터 마이닝 알고리즘은 일반적으로 많은 데이터 마이닝 연구들에서 적용되어 왔으며, 그 우수성이 입증되어 왔



<그림 1> 연구 진행 과정

다. 의사결정나무는 일반적으로 높은 예측률을 보이고 설명력이 높아 결과에 대한 해석이 용이한 반면, 인공신경망은 결과에 대한 해석이 어렵지만 스스로 학습을 통해 예측률을 높일 수 있다는 장점이 있다. 또한, 나이브 베이지언은 통계이론에 기반한 알고리즘이기에 본 연구에서는 성격이 다른 위 3가지의 알고리즘을 사용하여 예측률을 비교하고자 하였다. 이렇게 구축된 예측모형들 중 가장 높은 예측률을 보인 예측모형 하나를 선택하였고, 선택된 예측모형을 통해 자동차 재구매에 관한 규칙을 도출하였다.

### 3.2 분석 데이터

본 연구에서는 국내 자동차 제작사인 S사의 2014년 판매 데이터 11,000건을 분석 데이터로 수집하였다. 수집된 데이터는 고객 정보, 자동

차 정보 및 재구매 여부에 관한 정보 등을 포함하고 있다. 수집된 11,000명의 판매 데이터 중 1,837명이 재구매를 하였고 나머지 9,163명은 신규로 구매를 하였다. 이 수치를 통해 수집된 데이터에서의 재구매 비율이 16.7%임을 알 수 있었다.

### 3.3 목표변수와 독립변수

고객의 자동차 재구매 여부에 대한 예측모형을 구축하기 위해 재구매 여부를 나타내는 'Repurchase' 속성을 목표변수로 선택하였다. 목표변수의 값은 재구매를 의미하는 'Yes'와 신규 구매를 의미하는 'No'로 구분된다.

독립변수 선정을 위해 앞서 문헌연구에서 조사된 인구통계학적 변수들과 자동차 특성에 관한 변수들이 고려되었고, 그 중 전문가 회의를 통해 재구매에 영향을 줄 것으로 판단된 14개

의 초기 독립변수들을 선정하였다. 이들 중 의 절한 ‘Time’ 속성을 제외하였고, 기존의 ‘Area’ 미가 없거나 널 값이 존재해 사용하기에 부적 와 ‘CarModel’로 부터 ‘AreaType’, ‘Price’,

<표 1> 고객 속성의 전처리

종류	변수명	변수 설명	변수값	전처리 전 사용변수	전처리 후 사용변수
독립변수	Age	고객의 나이	20대, 30대, 40대, 50대, 60대, 70대	○	○
독립변수	Area	고객의 거주지역	강원, 경기, 경남, 경북, 광주, 대구, 대전, 부산, 서울, 울산, 인천, 전남, 전북, 제주, 충남, 충북	○	○
독립변수	CarModel	차량 배기량에 따른 모델	QM3, QM5, SM3, SM5, SM7	○	○
독립변수	ContractDay	구매한 요일	일, 월, 화, 수, 목, 금, 토	○	○
독립변수	ContractMonth	구매한 달	1월, 2월, 3월, 4월, 5월, 6월, 7월, 8월, 9월, 10월, 11월, 12월	○	○
제거변수	Time	자동차 판매 시간	10시, 11시, 12시, 13시, 14시, 15시, 16시, 17시, 18시, 19시	○	
독립변수	Color	차량의 색상	검정, 흰색, 노르딕블루, 레드바디, 마르스그레이, 백진주, 블랙바디, 아이보리바디, 오렌지바디, 울트라실버, 크림슨레드, 펄그레이	○	○
독립변수	Income/Year	고객의 연수입	없음, 3000만원미만, 3000만원~ 5000만원, 5000만원~7000만원, 7000만원~9000만원, 9000만원이상	○	○
독립변수	Job	고객의 직업	공무원, 무직, 법조인, 의료인, 자영업, 회사원	○	○
독립변수	CarEdition	차량 모델에 따른 등급	LE, LE25, PE, RE, RE25, RE35, SE	○	○
독립변수	Navigation	네비게이션 설치 유무	Yes, No	○	○
독립변수	Airbag	에어백 설치 유무	Yes, No	○	○
독립변수	Sex	고객의 성별	남, 여	○	○
독립변수	Sunroof	선루프 설치 유무	Yes, No	○	○
독립변수	AreaType	고객 거주지역의 특성	광역시, 특별시, 도		○
독립변수	Price	차량의 가격	숫자값		○
독립변수	Oil	차량의 유류	가솔린, 디젤, 가솔린/디젤		○
독립변수	CarType	차량의 종류	소형, 중형, 대형, 소형SUV, 중형SUV		○
목표변수	Repurchase	재구매 여부	Yes, No	○	○

‘Oil’, ‘CarType’이라는 4개의 새로운 변수를 파생시켜 초기 독립변수에 추가하였다. <표 1>은 전처리를 통해 선택된 초기 독립변수 17개와 목표변수 1개에 관한 정보를 보여준다.

### 3.4 데이터 균형화

데이터 마이닝 기법을 이용하여 예측모형을 구축할 때 목표변수의 클래스 분포를 파악해야 한다. 목표변수 내에 특정 클래스에 속하는 관측대상이 다른 클래스에 속하는 관측대상 보다 많을 경우, 관측대상이 많은 클래스를 중심으로 학습이 많이 이루어기 때문에 특정 클래스만을 잘 분류하는 편향된 예측모형이 구축된다. 따라서 예측모형을 구축하기 전 목표변수 내에 존재하는 클래스들의 비율을 맞추어 주는 데이터 균형화 작업이 필요하다.

본 연구의 목표변수인 ‘Repurchase’의 값은 2개 클래스인 ‘Yes’와 ‘No’로 구성되어 있다. 앞서 언급한바와 같이, 각 클래스별 관측대상의 수는 ‘Yes’가 1,837명이고 ‘No’가 9,163명이기 때문에 목표변수의 클래스 분포가 ‘No’ 방향으로 많이 편향되어 있는 것을 알 수 있다. 이와 같은 문제점을 해결하기 위해, 본 연구에서는 10개의 표본을 추출할 때 언더샘플링(under-sampling)을 통해 클래스의 비율을 맞추는 작업을 추가로 진행하였다. 그 결과 표본별 3,674명(Yes: 1,837명, No: 1,837명)으로 이루어진 데이터 셋을 최종 데이터 셋으로 사용하였다.

### 3.5 변수 선택

초기에 선정된 독립변수들 중에는 예측모형

을 구축하는데 아무런 도움을 주지 못하는 것들이 존재할 수 있다. 목표변수의 예측에 큰 도움이 되지 않음에도 불구하고 예측모형이 구축될 때 이용된다면 분류의 정확성을 저하시킬 수도 있기 때문에 이러한 변수들을 제거한 후 예측모형을 구축하는 것이 바람직하다(Dash & Liu, 1997). 이와 같이, 여러 초기 독립변수들 중 목표변수 분류에 중요한 역할을 하는 변수들을 선택하는 과정을 변수 선정이라고 한다. 본 연구에서는 변수 선정 알고리즘으로 gain ratio를 사용하여 초기 독립변수들의 중요도를 평가하였고, wrapper 방법 중 역방향 제거 방식을 사용하여 예측모형에 포함될 독립변수들을 최종 선택하였다. 여기에서 역방향 제거란 매 회 가장 중요성이 떨어지는 변수를 하나씩 제거한 후 예측모형을 구축하는 방식으로, 중요도가 가장 높은 변수가 1개 남을 때까지 진행되며, 새롭게 독립변수의 중요도 순위를 재 산정하기 때문에 전 단계에서 탈락된 변수의 영향을 제거한 상태에서 현재의 변수들 사이의 중요도가 계산된다(Witten & Frank, 2005). 이 과정을 통해 변수의 개수만큼 예측모형이 구축되고, 구축된 여러 개의 예측모형 중 정확도가 가장 높게 나왔을 때의 변수 조합을 최종 변수로 선택하게 된다.

## IV. 분석 결과 및 맞춤형 전략 제안

### 4.1 예측모형 구축과 예측률 비교

본 연구에서는 오픈소스 데이터 마이닝 툴인

Weka ver.3.8을 이용하여 자동차 재구매에 관한 예측모형을 구축하였다. 예측모형 구축에는 C4.5 기반의 의사결정나무, 나이브 베이지안, 인공신경망 알고리즘이 사용되었고, 언더샘플링 방식으로 추출된 10개의 표본들을 각각 66%의 학습데이터와 34%의 검증데이터 비율로 분할하여 실험을 진행하였다. 여기에서 학습데이터는 예측모형을 학습시키기 위해 사용되는 데이터이며, 검증데이터는 학습데이터를 통해 구축된 예측모형의 성능을 평가할 때 사용되는 데이터이다. 초기 17개의 독립변수들 중 중요도가 낮은 변수를 하나씩 제거하면서 예측모형의 예측률을 계산하였다. 그 결과 하나의 표본마다 서로 다른 3개의 알고리즘(DT, NB, ANN)을 적용하고, 각 알고리즘마다 역방향 제거를 통한 17개의 예측모형을 구축하여 최종적

으로 510개의 예측모형이 구축되었다.

<표 2>는 10개의 표본에 대한 예측모형의 평균 예측률과 예측모형이 각 표본에서 가장 높은 예측률을 기록한 횟수를 보여주고 있다. 그 결과 의사결정나무로 구축된 예측모형이 다른 예측모형들 보다 평균 예측률(84.7798%)이 높게 나타났고, 10개의 표본들 중에서도 8개의 표본에서 가장 높은 예측률을 기록하였다.

<표 3>은 10개의 표본 중에서 예측률이 가장

<표 2> 표본별 예측모형의 평균 예측률 및 상위 예측률 보유 횟수

예측모형	평균 예측률	상위 예측률 보유 횟수
DT	84.7798	8
NB	73.3307	0
ANN	84.0592	2

<표 3> 표본별 예측모형들의 예측률 비교

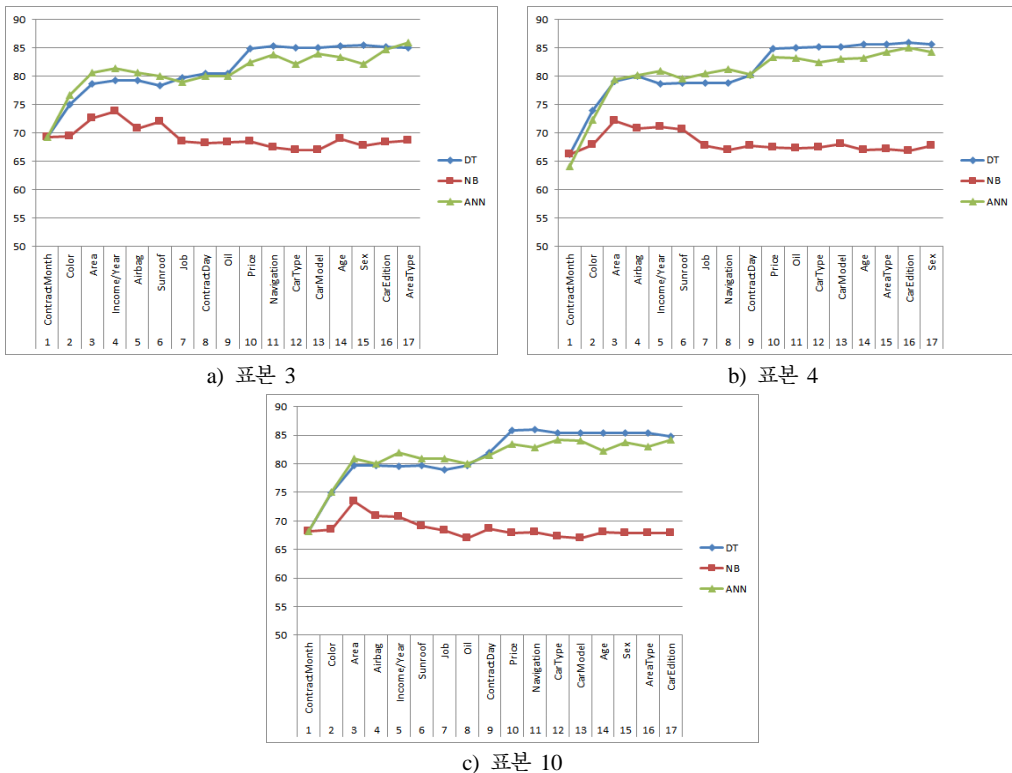
변수 순서	표본 3				표본 4				표본 10			
	변수	DT	NB	ANN	변수	DT	NB	ANN	변수	DT	NB	ANN
1	ContractMonth	69.34	69.34	69.34	ContractMonth	66.29	66.29	64.13	ContractMonth	68.21	68.21	68.21
2	Color	75.02	69.42	76.70	Color	74.06	67.89	72.38	Color	74.94	68.45	75.10
3	Area	78.62	72.62	80.62	Area	79.10	72.14	79.50	Area	79.66	73.50	80.94
4	Income/Year	79.26	73.82	81.35	Airbag	79.98	70.78	80.14	Airbag	79.74	70.94	80.06
5	Airbag	79.26	70.86	80.62	Income/Year	78.70	71.10	80.94	Income/Year	79.50	70.70	81.91
6	Sunroof	78.38	72.06	80.06	Sunroof	78.78	70.62	79.66	Sunroof	79.66	69.10	80.86
7	Job	79.82	68.53	78.94	Job	78.78	67.73	80.54	Job	78.94	68.37	80.86
8	ContractDay	80.46	68.29	79.98	Navigation	78.78	67.09	81.27	Oil	79.74	66.93	79.98
9	Oil	80.54	68.45	80.06	ContractDay	80.14	67.81	80.30	ContractDay	81.99	68.70	81.51
10	Price	84.87	68.53	82.55	Price	84.95	67.41	83.35	Price	85.83	67.89	83.43
11	Navigation	85.35	67.41	83.83	Oil	85.03	67.33	83.19	Navigation	85.99	68.05	82.79
12	CarType	85.11	67.01	82.23	CarType	85.27	67.41	82.47	CarType	85.43	67.25	84.23
13	CarModel	85.11	67.09	83.99	CarModel	85.27	68.05	83.11	CarModel	85.43	66.93	83.99
14	Age	85.35	69.02	83.35	Age	85.67	67.09	83.27	Age	85.43	68.05	82.23
15	Sex	85.43	67.73	82.15	AreaType	85.67	67.17	84.23	Sex	85.35	67.81	83.75
16	CarEdition	85.19	68.37	84.79	CarEdition	85.91	66.93	85.11	AreaType	85.35	67.81	82.95
17	AreaType	85.11	68.70	85.99	Sex	85.67	67.73	84.31	CarEdition	84.87	67.89	84.15



높은 예측모형을 지닌 표본 3개에 관한 정보를 보여주고 있다. 여기에서 변수 순서가 낮을수록 중요도가 높은 변수를 의미하며, 17번 변수부터 1번 변수 순으로 변수가 제거되면서 기록된 예측모형의 예측률을 보여주고 있다. 예를 들어, 표본 3의 경우 인공지능 기반의 예측모형이 17개의 변수 모두를 사용했을 때 가장 높은 예측률인 85.99%를 기록하였다. 표본 4와 표본 10에서는 의사결정나무 기반의 예측모형이 가장 높은 예측률을 기록하였는데, 표본 4에서는 16개의 변수를 사용했을 때 85.91%, 표본 10에서는 11개의 변수를 사용했을 때 85.99%의 예측률을 기록하였다. <표 3>의 결과를 그래프로 나타내면 <그림 2>와 같다.

#### 4.2 재구매 규칙에 따른 전략 제안

일반적으로 의사결정나무 알고리즘과 인공신경망 알고리즘을 비교해 보면, 결과의 안정성 측면은 인공신경망이 우세하고, 결과 해석의 측면은 의사결정나무로부터 도출된 분류 규칙이 이해하기 쉬운 형태로 만들어지기 때문에 의사결정나무가 우세하다고 평가되고 있다(Witten and Frank, 2005). <표 3>의 결과를 살펴보면, 표본 3의 인공지능 기반 예측모형과 표본 10의 의사결정나무 기반 예측모형에서 가장 높은 예측률인 85.99%를 동일하게 보여주었다. 본 연구에서는 두 예측모형 중 의사결정나무 기반 예측모형의 결과를 활용하여 재구매 전략을 도



<그림 2> 자동차 재구매에 관한 예측모형들의 예측률 도식화

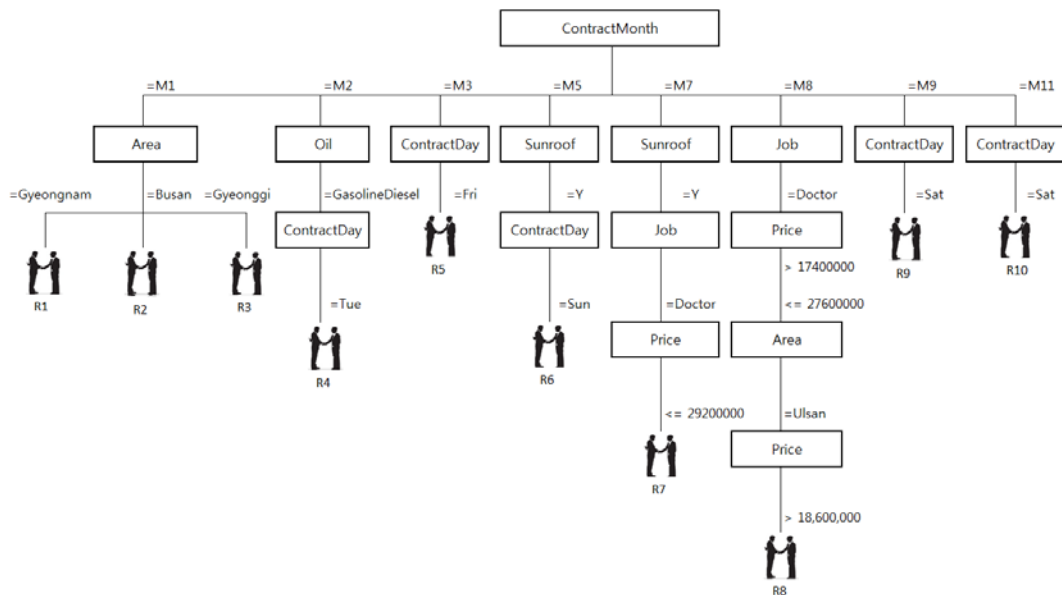
출하고자 한다. 그 이유로는 본 연구에서는 의사결정나무 기반 예측모형이 전반적으로 우수한 예측 성능을 보여주었고(<표 2> 참고), 해당 예측모형의 결과를 활용할 경우 좀 더 설명력이 높은 재구매 전략을 개발할 수 있다고 판단하였기 때문이다.

<표 3>의 표본 10과 같이, 자동차 재구매에 영향을 주는 11개의 독립변수인 ‘ContractMonth’, ‘Color’, ‘Area’, ‘Airbag’, ‘Income/Year’, ‘Sunroof’, ‘Job’, ‘Oil’, ‘ContractDay’, ‘Price’, ‘Navigation’이 최종적으로 선정되었으며, 최종 선정된 독립변수들을 이용하여 의사결정나무를 구축할 경우 가장 높은 예측률을 얻을 수 있었다. <그림 3>은 의사결정나무 기반 예측모형에서 도출된 분류 규칙들을 보여주며, 의사결정나무 표현의 간결화를 위해 예측률이 100%인 규칙들 중 재구매자의 수를 많이 예측한 상위 10개의 ‘Yes(재구매)’ 규칙들만 그림에 표시하

였다. 이러한 상위 10개의 규칙은 총 860명 고객의 재구매 패턴을 설명해 주며, 이는 전체 재구매자 수 1,837명의 46.8%에 해당되는 수이다.

<표 4>는 <그림 3>에서 보여준 10개의 규칙에 관한 세부 내용들을 설명하고 있으며, 본 연구에서는 각 규칙들의 주요 특성들을 분석하여 다음과 같은 맞춤형 전략을 제안하고자 한다.

첫 번째, 최종 독립변수에 포함되지 못한 6개의 초기 독립변수에는 ‘CarType’, ‘CarModel’, ‘Age’, ‘Sex’, ‘AreaType’, ‘CarEdition’이 있다. 이를 통해 고객들은 기존에 소유했던 동일한 회사의 자동차를 재구매할 때에는 차량의 종류나 차량 배기량에 따른 모델은 크게 참고하지 않은 것으로 나타났다. 그리고 고객의 나이나 성별, 거주지역의 특성에 상관없이 자동차 재구매가 이루어졌고, 자동차 모델에 따른 등급 역시 고객들이 재구매하는 것에는 특별한 영향을



<그림 3> 의사결정나무로 도출된 자동차 재구매에 관한 규칙

<표 4> 자동차 재구매에 관한 규칙 명세

순서	규칙	클래스	재구매자 수
R1	ContractMonth = 'M1' & Area = 'Gyeongnam'	'Yes'	157
R2	ContractMonth = 'M1' & Area = 'Busan' & Price <= 32200000	'Yes'	13
R3	ContractMonth = 'M1' & Area = 'Gyeonggi'	'Yes'	139
R4	ContractMonth = 'M2' & Oil = 'GasolineDiesel' & ContractDay = 'Tue'	'Yes'	61
R5	ContractMonth = 'M3' & ContractDay = 'Fri'	'Yes'	12
R6	ContractMonth = 'M5' & Sunroof = 'Y' & ContractDay = 'Sun'	'Yes'	133
R7	ContractMonth = 'M7' & Sunroof = 'Y' & Job = 'Doctor' & Price <= 29200000	'Yes'	91
R8	ContractMonth = 'M8' & Job = 'Doctor' & Price > 17400000 & Price <= 27600000 & Area = 'Ulsan' & Price > 18600000	'Yes'	11
R9	ContractMonth = 'M9' & ContractDay = 'Sat'	'Yes'	24
R10	ContractMonth = 'M11' & ContractDay = 'Sat'	'Yes'	219

주지 않았다. 따라서 기존 고객들로부터 자동차의 재구매를 유도하기 위해서는 최종 선택된 11개의 독립변수의 특성을 활용하여 재구매 전략을 수립해야 한다.

두 번째, R1, R2, R3 규칙을 비교하여 살펴보면, 1월일 때 고객들의 거주지역이 경남, 울산, 경기이면 자동차의 재구매율이 높은 것을 알 수 있다. 실제로 S사에서는 2014년 1월 경남, 울산, 경기지역을 대상으로 한시적인 프로모션을 실시한 것으로 분석되었다. 따라서 지역을 고려한 맞춤형 마케팅 전략을 실시하여야 한다.

세 번째, R5, R6, R9, R10 규칙을 비교하여 살펴보면, 봄이나 가을에는 주로 주말에 고객들의 재구매가 많이 일어난 것을 확인할 수 있다.

이 시기의 고객들은 평일보다는 시간적인 여유가 많은 주말을 택하여 자동차를 재구매하는데 시간을 소비한다는 것으로 이해할 수 있다. 따라서 봄과 가을에는 주말 프로모션 전략을 실시하여 고객의 자동차 재구매를 높여야 한다.

네 번째, R6, R7 규칙을 살펴보면, 5월, 7월 일 때 자동차를 재구매하는 중요한 요인으로 자동차의 선루프 장착이 공통적으로 나타났다. 이 결과를 볼 때 고객은 날씨가 더워질 것을 예상하거나 날씨가 더울 때 선루프가 장착된 차량을 특히 선호하는 것을 알 수 있다. 이와 같이, 계절적인 특성이 자동차의 재구매에 많은 영향을 주기 때문에 여름의 계절적 요인을 자동차 상품에 반영한 판매 전략을 수립해야 한다.

끝으로, R7, R8 규칙을 통해 직종이 의사인 고객이 자동차를 재구매한 것으로 나타났다. 실제로 S사에서는 2014년 의료인들을 대상으로 7월과 8월에 프로모션을 시행한 바 있다. 이를 통해, 특정 직업에 종사하는 사람들을 대상으로 프로모션을 진행할 때 자동차 재구매에 대한 유도를 이끌어 낼 수 있다는 것을 알 수 있었다. 따라서, 평균 자동차 교체주기가 5년인 점을 고려하여 자동차를 구매한지 5년이 지난 특정 직업 고객들을 대상으로 신차 홍보 전략을 실시해야 한다.

## V. 결론

본 연구에서는 날로 경쟁이 치열해지는 국내 자동차 시장에서 기업이 경쟁우위를 점하기 위한 방안의 일환으로 기존 고객들을 대상으로 자동차의 재구매를 높일 수 있는 전략을 제안하고자 하였다. 이를 위해 데이터 마이닝 기법을 활용하여 재구매에 관한 예측모형들을 구축하였고, 구축된 예측모형들 중 가장 높은 예측률을 보인 예측모형으로부터 재구매에 관한 규칙을 도출하여 기업에 맞춤형된 재구매 전략을 제시하였다.

그 결과, 고객들이 자동차를 재구매할 때 영향을 주는 요인으로 1) 구매한 달, 2) 차량의 색상, 3) 고객의 거주 지역, 4) 에어백 설치 유무, 5) 고객의 연수입, 6) 선루프 설치 유무, 7) 고객의 직업, 8) 차량의 유류, 9) 구매한 요일, 10) 차량의 가격, 11) 네비게이션 설치 유무가 분석되었다. 또한 도출된 10개의 자동차 재구매 규칙을 통해, 1월 재구매자의 경우 특정 지역에

관한 요인이, 봄이나 가을의 경우에는 주말 여부가, 여름의 경우에는 선루프 유무와 같은 계절과 관련된 차량 요인이, 그리고 특정 직종을 맞춤형한 프로모션의 진행여부 등이 반영된 자동차 재구매 증진 전략을 제안하였다.

본 연구에서는 자동차 재구매에 영향을 주는 의미 있는 요인들을 파악한 것과 재구매 규칙을 도출하였다는 점, 그리고 도출된 규칙을 활용하여 기업에 맞춤형된 재구매 증진 전략을 제안하였다는 점에서 연구의 의의를 찾을 수 있다. 향후, 제안된 전략들을 해당 기업에서 활용할 경우 기존 고객들의 재구매율을 높이는 데 도움을 줄 것이라 생각한다.

데이터 마이닝 분석의 특성상 분석 결과는 수집된 데이터에 영향을 받게 된다. 본 연구의 경우 1년이라는 제한된 기간의 고객 자료를 분석하였기 때문에 그 당시의 경제적 흐름, 국내외 상황, 국가 정책 등과 같은 외적 요소들을 분석에 반영하지 못한 점이 본 연구의 한계점으로 인식된다.

## 참고문헌

- 국가지표체계, 자동차 산업 동향, 2014, [http://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx\\_cd=1150](http://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1150)
- 국가지표체계, 자동차 등록 현황, 2017, [http://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx\\_cd=1257](http://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1257)
- 권기대, 이상환, 허원현, “자동차 평가 속성에 대한 중요도-성과 분석과 전반적인 만족도, 재구매 의도 및 추천의도간의 구

- 조연구,” 대한경영학회지, 제45호, 2004, pp. 1795-1821.
- 김상현, 박현선, “소셜집단특성이 소셜커머스 재구매의도에 미치는 영향과 실용적 가치의 조절효과,” 정보시스템연구, 제22권, 제2호, 2013, pp. 1-24.
- 김은정, 김종원, “사회적 기업의 제품 재구매 의도에 미치는 영향에 관한 연구,” 한국산업정보학회논문지, 제17권, 제1호, 2012, pp. 105-115.
- 김효경, 신흥철, “비즈니스 고객이 재구매에 영향을 미치는 선택속성과 영향요인의 차이,” 관광·레저연구, 제22권, 제3호, 2010, pp. 83-99.
- 석경하, 조대현, 이종언, 김재길, 전유중, 이영배, “렌터카 자료를 이용한 재구매 예측 모형,” 한국자료분석학회, 제13권, 제3호, 2011, pp. 1423-1432.
- 손달호, “소셜 커머스에서 재구매 의도의 결정 요인,” 정보시스템연구, 제24권, 제2호, 2015, pp. 1-22.
- 이낙귀, “여행도매업자의 관계적 특성이 신뢰와 재구매의도에 미치는 영향,” 관광학연구, 제26권, 제3호, 2003, pp. 321-339.
- 이장로, 김미옥, “자동차의 제품속성이 만족도, 재구매 의도, 추천 의도에 미치는 영향에 관한 연구,” 국제경영리뷰, 제14권, 제1호, 2010, pp. 19-54.
- 장서흠, 임규홍, 위오기, “스마트폰 소비자의 재구매의도 영향요인에 관한 한중 비교 연구,” 상업교육연구, 제29권, 제1호, 2015, pp. 287-308.
- 전성은, 정기한, “고객의 e-satisfaction, e-loyalty, 재구매의도 형성에 영향을 미치는 인터넷 쇼핑 사이트 특성에 관한 연구,” 인터넷전자상거래연구, 제6권, 제3호, 2006, pp. 133-158.
- 정태린, 황호영, 임범규, 이동영, “청소년의 아웃도어 의류 브랜드 구매 결정 요인이 브랜드 이미지와 구전 의도 및 재구매 의도에 미치는 영향,” 한국체육학회지, 제51권, 제3호, 2012, pp. 183-192.
- 정태영, 서건석, 김수범, “의료기기의 구매결정 요인과 만족 및 재구매 의도의 구조적 관계에 관한 연구: 초음파영상진단장치를 중심으로,” 한국산학기술학회논문지, 제16권, 제5호, 2015, pp. 3308-3314.
- 조철호, “전자상거래에서 초기신뢰와 재구매의도에 영향을 미치는 소비자 구매결정요인에 관한 연구-중소 인터넷쇼핑몰을 중심으로,” 중소기업연구, 제2권, 제1호, 2006, pp. 173-202.
- 한은경, 송석재, 임한나, “소셜 커머스의 이용동기와 만족, 재구매 의도에 관한 연구,” 한국광고홍보학회, 제13권, 제3호, 2011, pp. 298-325.
- Balla, B. E., Ibrahim, S. B., and Ali, A. H., “The impact of relationship quality on repurchase intention towards the customers of automotive companies in sudan,” *British Journal of Marketing Studies*, Vol. 3, No. 4, 2015, pp. 1-15.
- Dash, M. and Liu, H., “Feature selection for classification,” *Intelligent Data Analysis*,

Vol. 1, No. 3, 1997, pp. 131- 156.

Koay, K. Y. and Derek, O. L. T., “The mediating role of customer satisfaction in customer retention model: A case of local automobile brands in Malaysia,” *Pertanika Journal of Social Sciences and Humanities*, Vol. 24, 2016, pp. 27-40.

McDougall, G. H. G. and Levesque, T., “Customer satisfaction with services: putting perceived value into the equation,” *Journal of Services Marketing*, Vol. 14, No. 5, 2000, pp. 392-410.

Nistorescu, T., Barbu, M., and Dumitriu, R. I., “Study concerning customers’ loyalty in the auto market,” *Management and Marketing*, Vol. 11, No. 2, 2013, pp. 332-339.

Oliver, R. L. and Swan, J. E., “Consumer perceptions of interpersonal equity and satisfaction in transaction: A field survey approach,” *Journal of Marketing*, Vol. 53, No. 2, 1989, pp. 21-35.

Ranjbarian, B., Fathi, S., and Kamali, A., “The factors affecting on consumers’ repurchase intention toward an automobile brand: An application of reasoned action theory,” *European Journal of Social Sciences*, Vol. 16, No. 3, 2010, pp. 352-362.

Witten, I. H. and Frank, E., *Data mining: Practical machine learning tools and*

*techniques*. San Francisco: Morgan Kaufmann Publishers, 2005.

**이 동 욱 (Lee, Dong-Wook)**



현재 경상대학교 경영정보학과에서 석사학위 과정 중이다. 경성대학교 e-비즈니스학과에서 경영학사 학위를 취득하였다. 주요 관심분야는 데이터마이닝과 빅데이터 분석이다.

**최 근 호 (Choi, Keun-Ho)**



현재 근로복지공단 근로복지연구원의 책임연구원으로 재직 중이다. 인하대학교에서 금속공학과와 테크노경영학사 학위를 취득하고, 동대학 대학원에서 경영학 석사학위를 취득하였다. 고려대학교 일반대학원 경영학과 MIS전공에서 경영학 박사학위를 취득하였다. 주요 관심분야는 데이터마이닝, 추천시스템, 산재보험 빅데이터 분석 등이다.

**유 동 희 (Yoo, Dong-Hee)**



현재 경상대학교 경영정보학과에서 부교수로 재직 중이다. 고려대학교 경영정보학과를 졸업하고, 고려대학교 일반대학원 경영학과에서 경영학 박사학위를 취득하였다. 주요 관심분야는 시맨틱웹, 온톨로지, 데이터마이닝, 빅데이터 분석 등이다.

<Abstract>

## **Development of Customized Strategy for Enhancing Automobile Repurchase Using Data Mining Techniques**

Lee, Dong-Wook · Choi, Keun-Ho · Yoo, Dong-Hee

### **Purpose**

Although automobile production has increased since the development of the Korean automobile industry, the number of customers who can purchase automobiles decreases relatively. Therefore, automobile companies need to develop strategies to attract customers and promote their repurchase behaviors. To this end, this paper analyzed customer data from a Korean automobile company using data mining techniques to derive repurchase strategies.

### **Design/methodology/approach**

We conducted under-sampling to balance the collected data and generated 10 datasets. We then implemented prediction models by applying a decision tree, naive Bayesian, and artificial neural network algorithms to each of the datasets. As a result, we derived 10 patterns consisting of 11 variables affecting customers' decisions about repurchases from the decision tree algorithm, which yielded the best accuracy. Using the derived patterns, we proposed helpful strategies for improving repurchase rates.

### **Findings**

From the top 10 repurchase patterns, we found that 1) repurchases in January are associated with a specific residential region, 2) repurchases in spring or autumn are associated with whether it is a weekend or not, 3) repurchases in summer are associated with whether the automobile is equipped with a sunroof or not, and 4) a customized promotion for a specific occupation increases the number of repurchases.

**Keyword:** Automobile repurchase, Data mining, Prediction model, Decision tree, Customized strategy

\* 이 논문은 2017년 4월 25일 접수, 2017년 5월 24일 1차 심사, 2017년 7월 25일 게재 확정되었습니다.