

데이타마이닝을 이용한 CRM 事例研究

- A 패션企業을 中心으로 -

李 裕 順

梨花女子大學校 大學院 經營學科 博士課程 修了, 第一毛織 三星패션연구소

A CRM Study on the Using of Data Mining

- Focusing on the "A" Fashion Company -

Lee, Yu-Soon

Completion of Ph. D., Dept. of Management, Ewha Womans University
Chief Researcher, Cheil Industries, Samsung Fashion Institute

Abstract

In this study, we proposed a method to be standing customers as the supporting system for the improvement of fashion garment industry which was the marginal growth getting into full maturity of market. As for the customer creation method of Fashion garment company is developing a marketing program to be standing customer as customer scoring to estimate a existing customer's buying power, and figure out minimum fixed sales of company to use a future purchasing predict. This study was a result of data from total sixty thousands data to be created for the 11 months from september. 2000 to July. 2001. The data is part of which the company leading the Korean fashion garment industry has a lot of a customer purchasing history data. But this study used only 48,845 refined purchased data to discriminate from sixty thousands data and 21,496 customer case with the exception of overlapping purchased data among of those. The software used to handle sixty thousands data was SAS e-miner. As the analysis process is put in to operation the analysis of the purchasing customer's profile firstly, and the second come into basket analysis to consider the buying associations for Association goods, the third estimate the customer grade of Customer loyalty by 3 ways of logit regression analysis, decision tree, Artificial Neural Network. The result suggested a method to be estimate the customer loyalty as 3 independent variables, 2 coefficients. The 3 independent variables are total purchasing amount, purchasing items per one purchase, payment amount by one purchasing item. The 2 coefficients are royal and normal for customer segmentation. The result was that this model use a logit regression analysis was valid as the method to be estimate the customer loyalty.

Key words: basket analysis(장바구니분석), buying power(구매력), CRM(고객관계마케팅), customer loyalty(고객충성도), logit regression(로지트회귀분석)

I. 연구목적과 필요성

70년대부터 고성장을 거듭하던 국내 패션산업이 최근에는 시장 완속기에 접어들면서 시장 성장이 점차 둔화되고 있다. IMF를 맞이한 1997년 이전에는 전반적으로 두 자리 수의 성장을 거듭하였으나, 2000년에 들어서서는 3%대의 성장에 머물고 있는 실정이다¹⁾. 이러한 시장 정체기에 패션기업들은 불황을 극복하기 위하여 다양한 미래전략을 추진하고 있다. 그 대표적인 전략이 핵심고객창출이다. 즉 한정된 시장 내에서 기업에 가치를 주는 우수 고객을 발굴하여 집중 마케팅을 통해 개별고객의 구매력을 향상시키고 더불어 안정된 수익성을 유지하는 것이다²⁾³⁾. 최근 핵심고객 발굴에 대한 방법론으로 CRM(Consumer Relation Management)이 부상하고 있다. 특히 일부 패션 대기업들이 CRM 솔루션을 도입하면서 고객정보 분석 및 고객발굴을 위한 마케팅 수단에 대한 연구가 활발히 진행되고 있다⁴⁾. 유통업계로는 현대백화점이 1997년 CRM을 위한 마스터 플랜에 착수하여 1999년에 고객 데이터를 정비하고 2001년에 3단계 구축을 완료하였다. 패션 제조업체로는 주식회사 세정이 1999년에 CRM 서비스를 도입하여 2002년 현재도 진행하고 있으며 대외적으로 구축 성공사례에 대한 발표를 하고 있다⁵⁾. 그러나 CRM을 도입한 기업은 물론, 도입하고자 하는 기업들이 우려하는 것은 패션산업의 상품과 유통형태가 매우 다양하여 구입 고객의 구입 이력을 상세히 획득할 수 있는 시스템 구축이 어렵다는 것이다. 더불어 구입고객 중에서 우리 기업의 핵심고객을 추출할 수 있는 필요 정보가 무엇인지에 대한 노하우도 부족하여 비싼 경비를 지불하고 구축한 데이터웨어하우스(Data warehouse)의 활용도가 취약할 것이라는 것⁶⁾이다. 즉 기업들은 이를 위해 유효한 데이터 및 의사결정 요인 추출이 가능한 데이터웨어하우스 구축을 지원하는 방법론이 필요한 것이다.

따라서 본 연구는 패션 산업이 사용할 수 있는 전문적인 CRM 시스템구축을 지원하기 위하여 실제 국내 최우수 패션기업이 보유한 고객 데이터 및 고객별 구매이력 정보를 통해서 패션기업이 보유한 데이터웨어하우스의 현황과 문제점을 파악하고자 한다. 또한

보유 데이터를 이용하여 다양한 CRM 분석 기법을 통해 고객 세분화 및 고객의 연관 구매력 등을 분석할 것이다. CRM관련한 선행연구들이 2000년을 기점으로 많이 있지만, 패션산업에 대한 연구사례는 아직 없었다. 따라서 본 연구는 실제 패션기업의 고객정보를 기반으로 한 데이터웨어하우스 실태조사로서 그 초석을 다진다는 데에 의의를 가진다.

본 연구는 문헌연구는 물론, 패션기업이 보유한 6만 여 개의 실제 고객정보에 대한 분석을 중심으로 이루어졌다. 본 연구의 분석도구로는 SAS(Statistical Analysis System)의 e-miner를 사용하였고, 주로 마이닝기법을 통해 다층적인 분석 단계를 거쳤다. 첫째는 보유 데이터의 문제점과 보완 방법을 파악하였고, 둘째는 데이터의 정제를 통해 기존 고객의 구매력을 평가하여 고객 등급화를 통해 핵심고객을 추출하는 방법론을 개발하였고, 셋째는 고객의 연관 구매력을 평가하여 기업의 경영성과를 높이는 기법을 도출하였다.

이에 본 연구는 패션기업의 경영관리에 중요한 대안으로 부상하는 핵심고객창출에 대한 방법론 연구에 대한 도움이 될 것으로 판단된다. 또한 CRM 시스템을 도입하는데 기초 기반이 되는 데이터웨어하우스의 최소 필요단위의 데이터베이스 정보를 제공함으로써 기업의 구축비용 및 도입 실패율을 개선시키는데 일익을 담당하는 것에 그 의의를 두었다.

II. CRM 개념 및 선행연구

CRM에 대한 이론적인 개념은 Peppers & Rogers의 저서 "The One to One Future"(1993)에 따르면, 장기적인 관점에서 기업에 가치를 주는 고객에 대하여 지속적인 관계를 형성함으로써 고객이 해당기업의 제품을 평생 소비하여 생애가치(LTV, Life Time Value)를 극대화한다는 개념이다⁷⁾. 고객의 생애가치가 기업의 장기적인 경영효율에 제공하는 높은 효과에 대한 인식이 확대됨에 따라, 점차 많은 기업이 고객을 기업의 중요한 자산으로 인식하기 시작했다. 과거의 고객만족 영역이나 데이터베이스 마케팅은 기업관점에서 고객

을 관리하고자 했다. 그러나 CRM은 고객과의 관계를 통해 축적된 정보를 분석하여 고객을 이해할 수 있는 지식을 얻고, 이를 통해 개별고객 수준에서 고객만족을 이끌어낼 수 있도록 마케팅, 판매, 서비스 활동을 효율적으로 전개함으로써 장기적으로 고객과 우호적인 관계를 형성하고 개별고객의 수익성을 극대화하기 위한 총체적인 경영활동이다⁸⁾. 총체적인 경영활동의 의미는 전략과 업무 프로세스, 조직, 정보시스템이 CRM의 목적을 위해 통합적으로 운영되어야 한다는 것이다. 그러므로 CRM의 구성요소 또한 최근에는 DB 마케팅의 핵심 구성요소인 고객 데이터베이스와 분석 기반(분석도구 및 분석능력)에 마케팅, 판매, 서비스 등 고객과의 직접적인 상호작용이 필요한 업무 프로세스 및 기업조직과 고객접점을 이루는 매체를 모두 포함하게 되었고, 이를 통합적으로 활용하기 위한 전략적 요소도 포함시켜야 한다⁹⁾¹⁰⁾.

〈표 1〉에서와 같이 데이터베이스 마케팅 전략은 데이터에 근거하여 실행 가능한 마케팅에 초점이 맞춰진 프로모션 전략과 Customer Reward Program, Customer Service Program등을 실행하기 위한 부분 전략에 가깝다¹¹⁾.

이러한 CRM 전략에서 기초가 되는 고객 정보 분석에는 데이터 마이닝 기법(Data Mining Method)이 대표적이다. 데이터 마이닝에 대한 정의를 보면 '대용량의 데이터에 숨겨져 있는 데이터 간의 관계, 패턴을 탐색하고 이를 모형화하여 업무에 적용할 수 있는 의미있는 정보로 변환함으로써, 기업의 의사 결정에 적

용하는 일련의 과정', '대규모 데이터 베이스 내에 존재하는, 그러나 데이터 사이에 숨겨져 있는 상호관련성과 일반화할 수 있는 특성에 대한 탐색', '대량의 데이터로부터 패턴 인식기술과 통계기법, 수학적 기법 등을 이용하여 의미있는 새로운 상관관계, 패턴 그리고 추세를 발견하기 위한 기법'이라고 볼 수 있다. OLAP(On Line Analytical Processing)은 이미 정의된 데이터 모델에 근거하여 프로그래밍 없이 사용자가 여러 형태의 비정형 분석을 하는 것으로, 데이터 베이스에 있는 자료를 여러 각도로 조회하는 것이다. 이에 비해 마이닝은 방대한 양의 정보로부터 쉽게 드러나지 않는 유용한 정보의 패턴을 추출하는 과정으로 감추어진 지식, 기대하지 못했던 패턴, 새로운 법칙을 발견하여 경영의사 결정에 활용하는 것이다¹²⁾.

데이터 마이닝의 활용범위는 유통 산업에서 장바구니 분석(Market Basket Analysis), 표적 마케팅, 교차판매, 금융산업에서는 신용평가(Credit Scoring), 카드 부정사용 탐지(Fraud Detection), 고객 세분화 및 프로파일링, 고객수익성 분석, Claim 분석, 통신 산업에서는 고객이탈 방지(Churn Management), 표적 마케팅 등으로 매우 다양하다. 그러나 데이터 마이닝에는 활용하는데 몇 가지 제약이 있다¹³⁾. 첫째는 데이터 베이스의 준비이다. 데이터가 체계적으로 잘 정비되어 있어야 하며 그렇지 않으면 많은 시간을 데이터 검증에 사용하게 된다. 둘째는 데이터 마이닝의 결과를 해석하고 전략적 의미와 연결시키기 위해서는 통계와 비즈니스 전문가가 혼합된 팀이 구성되어야 한

〈표 1〉 데이터베이스 마케팅 전략(박찬욱, 1997)

구분	영역	전 략	효과 Customer Reward Program	프로그램 Customer Service Program
기존고객	기존상품	고객 활성화 전략	구매빈도 증가	우량고객 우대 프로그램, 마일리지서비스
		고객애호도 재고 전략	브랜드 스위칭 방지	서비스 차별화, 정보제공
		고객 유지 전략	반복 구매	고객서비스 센터, 할인쿠폰 DM
	신상품	교차판매 전략	범위의 경제	제휴서비스
잠재고객	신규고객	신규고객 확보전략	지속적 성장	직접반응광고, 정품응모행사
	과거고객	재유치 전략	Win-back	할인쿠폰 DM

다. 통계 전문가 없이 단순히 도구만 사용할 경우 효과를 얻을 수 없다. 셋째, 마이닝 모델의 선택이 중요한 변수가 된다. 데이터 마이닝은 복잡한 알고리즘을 구현한 강력한 도구이다. 그러나 어떤 모델을 완성하는 데에는 수개월이 들어갈 정도로 많은 시간이 소요된다. 업계에서도 보통 시스템 가동 이후 최소 1년은 경과해야 시스템의 구축효과를 조금씩 느낄 수 있다고 한다¹⁴⁾.

그러나 CRM을 위한 데이터 마이닝 기법에 대해서는 현재까지 다양한 분야에서의 연구는 부족한 편이다. 특히 패션산업에서의 리서치 연구사례는 전무한 것으로 판단된다. 반면 계량적인 데이터로 주로 구축된 금융권에서는 다양한 연구사례가 발표되고 있다. 마이닝 기법이 국내에 소개되어 많은 연구가 일어난 것은 90년대 초반으로 특히 연속형 데이터 분석이 주로 있었다. 고객 신용도 평가나 주식평가, 부모평가 등으로 대표적인 것이 Data Mining을 위한 속성변수 추출개발을 연구한 이승민(1999)¹⁵⁾이 있다. 그리고 권오준(1997)¹⁶⁾은 통계적 기법과 인공지능망을 이용한 신용카드 고객 신용도 평가모형에 관한 연구를 하였다. 그러나 90년대 후반에 일반 제조 및 서비스 기업에서 장기간 축적된 고객정보 분석에 대한 욕구가 증대되면서 고객성향 및 고객구매 분석 등에 대한 연구가 활발하면서, 서은호(2000)¹⁷⁾는 E-Marketplace에서 고객 성향분석을 위한 웹마이닝기법의 적합성 측정, 김기운(2000)¹⁸⁾은 데이터 웨어하우스 기반의 고객 관계관리 모델링을, 김성수(2000)¹⁹⁾는 국내 기업의 성공적인 고객관계관리 구현방안에 관한 연구 등이 있었다.

III. 연구 방법 및 절차

1. 분석 데이터

본 연구의 분석데이터를 수집하기 위하여 국내 패션 산업에서 가장 ERP system 구축이 잘 된 기업인 A사를 선정하였다. A사는 130개 매장 내에 POS 시스템(Point of Sales System)과 Lan network이 설치되어 구매 즉시 판매결과 혹은 구매고객 데이터가 중앙 컴퓨터시스템에 전송되어 축적되고 있다. 본 연구에서 사용된 데이터는 A사가 고객서비스 차원에서 발행한 '고객 마일리지 카드'의 사용기준이다. 그러므로 고객 마일리지 카드에 축적되지 않은 고객정보는 제외된 것이다.²⁰⁾

본 연구에 사용된 data의 record는 총 6만 여 개로 지난 2000년 9월부터 2001년 7월까지 11개월간 생성된 것이다. Record별 설명력(field)은 <표 2>와 같이 총 9개이나, 품번으로 표시되는 제품의 스타일 세분구분²¹⁾으로 결국 총 14개가 형성되었다.

2. 데이터(Original Data)의 정제(整除)

2000년 6월부터 11개월 동안 고객 구매력(Customer Purchase History)에서 Original Data의 획득 고객(Record)는 총 60,251개이나, 첫째, 성별, 주민번호, 지역 등의 정보(Field) 중에서 하나라도 정보가 없는 경우(null로 표시되는 부분), 둘째, A기업이 타겟으로 하지 않는 제외 정보(Data Outliner)인 17세 이하, 70세 이상의 고객, 셋째, 반품된 것으로 추정되는 것으로 구매금액에서 마이너스인 경우를 제외하면 48,945개이다.

그러나 이렇게 정제된 데이터도 구매고객과 실제 제품 사용고객과의 불일치성에 의해 분석오류가 상당

<표 2> Customer purchase history data의 예 (2000년 6월~2001년 4월 (N=6만))

이름	성별	연령대	지역	구매일자	제품 출시시점	품목(복종)	품번(스타일)	구매금액
김현호	남	45	전남	20010313	01S	66	EYC669262	59,400
김현호	남	45	전남	20010412	01S	35	HYC35B174	103,820
박부자	여	34	인천	20010412	01F	24	GY043521	97,230
.....

할 것으로 판단되나, 현재 집적된 정보로는 오류에 대한 판별이 불가능한 것이 사실이다. 이러한 축적 데이터의 오류는 향후 DSS(Decision Support System)구축을 통한 CRM 시스템에서의 전략정보도 결국 오류가 발생될 것으로 전망된다. 이러한 관점에서 성공적인 CRM구축에는 데이터의 관리가 가장 중요한 과제가 될 것이다.

3. 분석 절차

본 연구에서 사용되는 분석 데이터는 4만에서 6만개의 매우 큰 데이터베이스이다. 그러므로 SPSS나 SAS 등 일반 통계 패키지에서는 분석이 불가능하여 SAS의 CRM 분석도구인 e-miner를 분석도구로 사용하였다.

본 분석의 목적은 첫째, 다양한 분석을 통해 필요한 최소 Field 정보를 파악하고자 한다. 둘째, 패션산업에서는 제품간의 연계판매를 통한 고객효율화가 중요한 경쟁역량이므로 고객의 구매제품 이력으로 연계분석(basket analysis)을 통해 연결구매제품의 특성을 분석하고자 한다. 셋째, 고객의 구매강도, 즉 구매액 평가를 통해 고객 등급을 분류할 것이다. 이러한 분류를 통해 고객등급별 profile 분석을 하고자 한다. 이러한 결과는 고객 대응 공격적 마케팅전략을 수립하기 위한 기본 자료를 제공²⁹⁾할 것이다.

분석절차는 다음과 같이 총 6단계를 거쳐서 진행되었고, 시장 세분화, 목표 고객계층 선정, 고객 포트폴리오(Customer Portfolio)도출 및 핵심가치의 제공(Value Proposition)²⁹⁾을 거쳐 마케팅 믹스의 설계를 위한 최종 목적으로 분석되었다.

- 1차 : 통계분석을 통한 구매고객 profile
- 2차 : Data mining인 e-miner를 활용한 구매이력 분석
- 3차 : Market Basket analysis를 활용한 장바구니분석(관련상품구매 속성 평가)
- 4차 : 1차에서 3차에 이르기까지의 data outline을 기반으로 고객등급 분류(정성적 분석 및 정량적 분석 모두 적용)
- 5차 : 등급분류에 적용가능한 모형 선정(로지회귀분석, 의사결정나무, 인공신경망 등 3가지로 평가)
- 6차 : 등급별 고객특성 및 구매 제품 스타일 분석

IV. 연구 결과

1. 분석 정보의 프로파일(Record Profile)

60,251개의 Original Data Record가 1차 정제된 48,945개의 data를 빈도분석(Frequency Analysis)으로 분석 후, Record별로 동일인의 중복 구매 데이터를 제거한 결과, 연구에 적용되는 종속 변수는 총 21,496개 <표 3>이다.

<표 3>의 분석결과를 살펴보면, A사를 방문한 고객은 지난 2000년 9월부터 2001년 7월까지 11개월 동안 총 6,457,802,140원을 지불(Sum Observations)²⁹⁾하였다.

<표 3> 인구 통계적 특성에 따른 구매고객별 구매특성

구분	record 수	1인당 총구매가	1인당 구매별수	1회 구입금액	
전체	21,496	300,419	2.98	107,932	
성별	남자	13,253	305,740	2.98	109,422
	여자	9,243	299,863	2.99	105,537
연령별	20대	2,586	296,694	2.97	108,450
	30대	6,877	290,906	2.94	105,796
	40대	7,699	303,029	3.04	107,058
	50대	3,374	348,275	3.00	112,708
	60대	960	317,571	3.08	112,054
지역별	강원	1,575	293,451	2.59	115,157
	경기	2,514	305,664	3.21	104,393
	경남	1,262	321,798	3.21	105,592
	경북	3,079	364,658	3.63	108,018
	광주	176	318,877	3.33	102,592
	대구	1,257	277,812	3.12	91,442
	대전	169	147,087	2.17	66,522
	부산	1,422	306,944	3.40	102,195
	서울	1,815	291,939	2.83	122,416
	울산	10	203,240	2.50	88,728
	인천	751	268,882	2.54	105,267
	전남	2,709	308,660	2.93	108,844
	전북	2,310	280,676	2.65	110,145
	제주	204	259,234	2.46	110,561
	충남	1,611	234,847	2.31	105,007
	충북	632	312,965	2.71	123,001

이중 남자고객은 4,051,974,140원을, 여자고객은 2,405,828,000원을 지불하였다. 고객별 구매력을 살펴보면, 고객 1인당 평균 2.98개의 제품을 300,419원에 지불 (Mean)하였다. 제품 한 개당 구입금액은 107,932원으로 산출되었다.

성별, 연령별, 지역별로 고객의 구매특성을 살펴보았다. 남녀성의 구매특성의 차이는 크게 없는 것으로 나타났으나, 1회 구입금액은 남성이 우세하게 나타났다. 연령별로는 높은 연령대 일수록 구매력이 좋은 것으로 나타났으며, 특히 50대 고객이 heavy buyer였다.

총 16개 지역 중, 일인당 총 구매가가 높은 지역은 (경북> 경남> 광주> 충북> 전남> 부산> 경기 순 인 반면, 경북> 부산> 광주> 경기=경남> 대구로, 1회 구입금액은 강원> 충북> 제주> 전북> 전남> 경북 등으로 나타나서, 결과적으로 경북, 강원, 충북 고객이 heavy buyer일 것으로 추정되었다.

이와 같이 간단한 통계기법인 빈도분석(Frequency Analysis)으로도 Data의 Profile을 이해하고 전략적 마케팅 프로그램을 작성하는 데에도 유용하다. 예를 들어 일인당 구매액 평가를 통해 우리 고객의 경쟁력 정

도를 분석하고 이에 따른 시장경쟁전략을 도입할 수 있다. 또한 성별, 연령별, 지역별로의 시장 세분화 마케팅 전략을 전개할 수도 있다. 실제로 A社は 지역별 고객 수와 1인당 총 구매액 등을 기반으로 지역 집중화 전략을 실시하였다. 경북, 전남, 경기 3개 지역에 대한 '고객 사은 행사'를 통해 단골고객을 대량 확보하게 되었다.

2. 장바구니 분석(Basket analysis Result)

본 연구에서는 패션제품 구매에서 제품연결판매를 유도하기 위하여 SAS의 e-miner, Basket analysis를 통해 <표 4>와 <표 5>를 도출하였다. 장바구니 분석(Basket Analysis)를 연관성 분석(Associations Analysis)이라고도 한다. 장바구니 분석은 하나의 거래나 사건에 포함되어 있는 항목들의 관련성을 파악하여 둘 이상의 항목들로 구성된 연관성 규칙을 찾아내는 탐색적 자료분석 방법이다. 즉 특정 항목(들)의 존재가 다른 항목(들)의 존재를 암시하는 것을 의미한다. 이 방법은 둘 또는 그 이상의 품목간의 상호 관련성을 파악하는

<표 4> Relation을 2개로 하여 Basket analysis 한 결과 (SAS e-miner)

	Lift(확율)	Support(%)	Confidence(%)	Transaction(Count)	Rule
1	2.19	222	1067	478	반팔셔츠 => 반바지
2	2.19	222	4561	478	반바지 => 반팔셔츠
3	2.02	282	5792	607	반바지 => 반팔니트
4	1.79	311	2347	669	긴팔니트 => 긴팔셔츠
5	1.79	311	2374	669	긴팔셔츠 => 긴팔니트
⋮	⋮	⋮	⋮	⋮	⋮
21	1.17	1041	4992	2237	반팔셔츠 => 바지
22	1.17	1041	2443	2237	바지 => 반팔셔츠
23	1.15	642	4901	1381	긴팔셔츠 => 바지
24	1.15	642	1508	1381	바지 => 긴팔셔츠
⋮	⋮	⋮	⋮	⋮	⋮
31	1.05	761	2563	1635	스웨터 => 점퍼
32	1.05	761	31.12	1635	점퍼 => 스웨터
33	1.04	426	4422	915	사파리 => 바지
34	1.00	1043	4267	2242	점퍼 => 바지
35	1.00	1043	2448	2242	바지 => 점퍼

것이나, 품목간의 직접적인 인과관계가 있는 것은 아니다. 분석을 위한 자료의 구조는 거래(transaction)와 항목(item)으로 구성된다. 거래는 항목들이 거래로 일어난 경우를 일컫고, 항목들은 거래를 위한 품목을 일컫는다. 연관성분석의 규칙을 평가하는 것에는 지지도(Support), 신뢰도(Confidence), 향상도(Lift) 3가지의 개념이 있다. 향상도에서는 결과 데이터가 1인 경우는 두 품목이 서로 독립적인 관계에 있는 것이고, 결과 데이터가 1보다 큰 것은 두 품목이 서로 양의 상관관계에 있다는 것이고, 결과 데이터가 1보다 작다는 것은 두 품목이 서로 음의 상관관계에 있다는 것을 나타낸다. Basket Analysis는 교차 판매(Cross Selling), 묶음 판매(Bundling) 등의 상품 기획에 활용되며, 기타 상품 진열(Inventory Display), 첨부 우편물, 쿠폰 배부 등 고객 마케팅 전술에도 사용된다. 최근 카탈로그 판매나 인터넷 쇼핑이 활성화되면서 카탈로그 디자인(Catalog

Design), 화면 디자인 등에도 사용된다²⁵⁾.

본 연구에서는 연관 항목(Relation)을 2개와 4개로 입력하여 분석하였다. 연관 항목 2개로 분석한 결과는 <표 4>와 같이, 향상도가 1 이상인 규칙은 총 35개의 규칙이 발견되었다.

반팔 셔츠를 구매한 규칙 1과 규칙 21을 비교 평가하면, 규칙 1은 반 팔 셔츠와 반바지를 함께 구입한 거래가 총 21,496 레코드에서 478개 거래인 반면, 규칙 21의 반팔 셔츠와 바지를 함께 구입한 거래는 총 21,496 레코드에서 2,238개 거래로 나타났다. 또한 규칙 1의 신뢰도를 보면 반팔 셔츠를 구입한 고객이 반바지를 꼭 구입하는 거래는 총 4,524 레코드에서 483개 경우이며, 규칙 21은 반팔 셔츠를 구입한 고객이 바지를 구입하는 거래는 총 4,524 레코드에서 2,237개 거래로 나타났다. 이 결과로 볼 때는 규칙 21의 상호 관련성이 높게 나타났다고 할 수 있다. 그러나 향상도

<표 5> Relation을 4개로 하여 Basket analysis 한 결과(SAS e-miner)

	Lift(확률)	Support(%)	Confidence(%)	Transaction(Count)	Rule
1	2.20	2.22	16.74	477.00	긴팔니트 => 점퍼 & 스웨터
2	2.20	2.22	29.17	477.00	점퍼 & 스웨터 => 긴팔니트
3	2.19	2.22	10.67	478.00	반팔셔츠 => 반바지
4	2.19	2.22	45.61	478.00	반바지 => 반팔셔츠
5	2.02	2.82	57.92	607.00	반바지 => 반팔니트
∴	∴	∴	∴	∴	∴
61	1.38	4.48	10.52	963.00	바지 => 점퍼 & 스웨터
62	1.38	4.48	58.90	963.00	점퍼 & 스웨터 => 바지
63	1.37	5.37	58.55	1154.00	반팔셔츠 & 반팔니트 => 바지
64	1.37	5.37	12.60	1154.00	바지 => 반팔셔츠 & 반팔니트
65	1.36	4.06	38.89	872.00	점퍼 & 바지 => 반팔니트
∴	∴	∴	∴	∴	∴
101	1.06	5.98	14.03	1285.00	바지 => 긴팔니트
102	1.05	7.61	25.63	1635.00	스웨터 => 점퍼
103	1.05	7.61	31.12	1635.00	점퍼 => 스웨터
104	1.04	4.26	44.22	915.00	사파리 => 바지
105	1.02	4.06	30.21	873.00	반팔니트 & 바지 => 스웨터
106	1.02	4.06	13.68	873.00	스웨터 => 반팔니트 & 바지
107	1.00	10.43	42.67	2242.00	점퍼 => 바지
108	1.00	10.43	24.48	2242.00	바지 => 점퍼

로 검정하면 규칙 1의 경우 반팔 셔츠 구매 시에 반바지를 구입하게 될 가능성은 반팔 셔츠 구매가 전체되지 않았을 경우보다 2.19배나 높아지는 반면, 반팔 셔츠 구매 시에 바지를 구입하게 될 가능성은 반팔 셔츠 구매가 전체되지 않았을 경우보다 1.17배 높아지는 것으로 나타났다. 이와 같이 지지도와 신뢰도, 향상도 사이에는 상호 연관관계는 있으나, 기업의 제품 전략에 따라 3개의 연관성 규칙 분석모델을 활용하는 것이 좋을 것이다. 이 결과를 활용해 A社は 상품전략을 수립할 때, 점퍼와 바지의 수량은 동일하게 세웠으며, 반팔 셔츠 디자인은 반바지를 함께 착용하였을 때를 예상하여 코디네이션 되도록 진행하였다.

또한 연관성 분석에서 연관 항목을 4개로 입력하여 분석한 전체 결과, 2개로 분석한 경우보다 더 많은 108개의 규칙이 <표 5>와 같이 도출되었다. 장바구니 분석에서는 분석대상이 되는 연관 항목의 수를 증가할 수록 규칙 사례도 증가하는 것을 살펴볼 수 있다. 그러나 너무 많은 규칙사례의 도출은 오히려 상품전략의 방향을 흐리게 할 수도 있으므로 적절한 조정이 필요하다.

3. 정보의 세분화 및 등급화 (Record Segmentation & scoring)

고객의 구매 충성도를 평가하여 향후 고객의 총 구매액을 추정하고 분류고객별 특성에 따른 미래고객의 고객평가 및 고객별 마케팅 실행을 용이하게 할 수가 있다면, 기업의 의사결정에 높은 적중률을 가질 수 있을 것이다²⁶⁾. 데이터베이스 마케팅에서 유용하게 이용

되는 전통적인 분석방법으로는 RFM 점수분석(Recency, Frequency, Monetary Score analysis)이 있으나, 상품 종류가 한가지 범주에 들어가는 경우에는 RFM 공식이 소비자 행동을 잘 반영하지만, 여러 종류의 품목을 판매하는 기업의 경우에는 잘 적용되지 않는다. 특정 품목의 구입자들의 특성들이 다른 품목 구입자들의 특성과 다른 경우가 많기 때문이다. 그러므로 본 연구 분석에서는 기존의 RFM 공식의 특성을 활용하여 패션상품에 대한 고객의 기여도를 측정하기 위해 새로운 모형을 제시하고자 한다.

1) 적용 모형의 설계

먼저 개별 고객에 대한 충성도를 측정하기 위해 개별고객을 표기하여야 하므로, 다음과 같이 몇 가지 점자들을 사용하기로 한다.

구매 고객들을 일련번호로 분류하기 위한 점자는 i 이고, 같은 고객이 서로 다른 기업 제품을 구매한 경우를 구분하기 위한 점자는 j 이다. 또한 같은 고객이 구매한 어느 일정 기간을 구분하기 위한 점자는 t 이다.

이상과 같은 구매고객 분류방법에 따라 각각의 구매고객을 다음과 같이 3개의 점자를 이용하여 표시할 수 있다.

$$\begin{aligned} \text{고객}(i, j, t) \quad & i=1, 2, \dots, I \quad I = \text{고객수}, I = 21,496 \\ & j=1 \quad \quad \quad 1 = \text{A社 제품} \\ & t=1 \quad \quad \quad 1 = 2000\text{년 } 6\text{월} \sim 2001\text{년 } 4\text{월} \end{aligned}$$

각 구매고객(ijt)의 구매충성도를 평가하는 매장 방문 건 수, 제품 구매 횟수, 제품 구매금액을 나타내는 계수들을 다음과 같이 정의한다.

X_{ijt} = t 기간동안 j 社の 제품을 구매한 i 번째 고객

<표 6> A社の 21,496명의 구매고객 정보 중에서 필요한 계수값을 정리한 결과

구매고객(ijt)	1인당 매장 방문수	1인당 구매별수	1인당 1회 매장 방문시 구매별수	1인당 총구매가	1인당 1회 제품 구매시 구입금액
	C_{ij}	P_{ij}	EP_{ij}	Q_{ij}	EQ_{ijt}
X00001_01_01	4	8	2	1,850,000	231,250
X00002_01_01	2	6	3	2,450,340	408,390
⋮	⋮	⋮	⋮	⋮	⋮
X21496_01_01	1	1	1	230,000	230,000

C_{jt} = t기간동안 j사의 제품을 구매한 i번째 고객의 매장 방문 수

P_{jt} = t기간동안 j사의 제품을 구매한 i번째 고객이 구매한 제품의 총개수

Q_{jt} = t기간동안 j사의 제품을 구매한 i번째 고객이 구매한 제품의 총구입금액

위의 계수들을 이용하여 구매고객(X_{jt})이 매장에 한번 방문하여 구입한 제품 개수와 구매고객이 하나의 제품을 구입하기 위해 지출한 비용을 나타내는 계수는 다음과 같이 정의할 수 있다.

EP_{jt} = t기간동안 j사의 제품을 구매한 i번째 고객이 매장에 한번 방문하여 구입한 제품 개수

EQ_{jt} = t기간동안 j사의 제품을 구매한 i번째 고객이 하나의 제품을 구입하기 위해 지출한 비용

이와같이 모형에 적용하려는 변수와 계수들의 정의를 기반으로 본 모형의 적용대상인 A사의 21,496명의 구매고객 정보를 <표 6>과 같이 수집하였다.

2) 적용 모형의 구축

위에서 정의된 변수 및 계수들을 이용하여 A사의 21,496명의 고객에 대한 구매충성도를 측정할 수 있는 모형을 다음과 같이 제시하였다.

$$Y_{ijt} = Q_{jt} \cdot EP_{jt} \cdot EQ_{jt}$$

Y_{jt} = t기간동안 j사의 제품을 구매한 i번째 고객의 구매 충성도

Q_{jt} = t기간동안 j사의 제품을 구매한 i번째 고객이 구매한 제품의 총구입금액

EP_{jt} = t기간동안 j사의 제품을 구매한 i번째 고객이 매장에 한번 방문하여 구입한 제품 개수

EQ_{jt} = t기간동안 j사의 제품을 구매한 i번째 고객이 하나의 제품을 구입하기 위해 지출한 비용

위 모형에서 구매고객 충성도를 평점식에 의해 고객의 등급을 다양하게 세분화할 수 있으나, 너무 많은 등급으로 나눌 경우에 향후 마케팅 대응의 복잡성을 가져올 우려가 있어서, 우선 세 개의 등급과 두 개의 등급에 대한 결과를 유추하여 정확도를 측정해 보았다. 우선 세 개의 등급을 가지는 변수(level)로 normal, good, royal에 대하여 e-miner의 나무구조분석(Decision

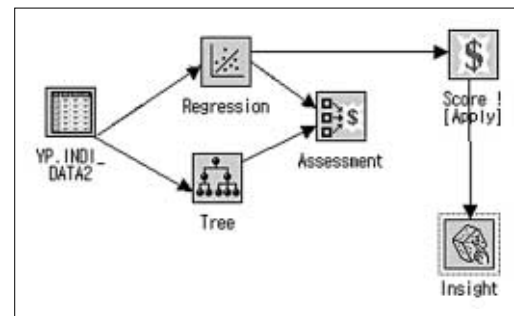
tree Analysis)를 이용하여 분석해 보았다. Level이라는 변수는 개인지출금액, 이용빈도, 1회당 지출비용에 대하여 분포를 기반으로 세 개의 수준으로 나눈 다음, 이를 곱하고 다시 1점부터 27점까지의 고객점수에 대하여 normal은 3점 이하, good은 4점부터 10점, royal은 11점 이상으로 분류하였다. 두 개의 등급을 가지는 변수로 normal, royal로 명명하였고, 10점 이하를 normal로, 11점 이상을 royal로 분류하였다.

3) 적용 모형의 평가

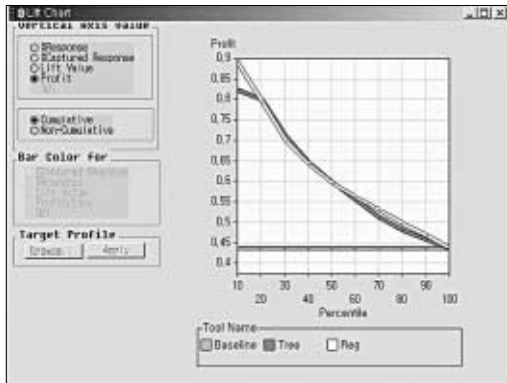
본 모형의 적합도를 측정하기 위하여 나무구조 분석을 이용하였다. 그 결과 세 개의 고객분류모델은 (563+5,197+5,909)/21,496의 값이 54.3%의 정확도를 나타낸다. 반면, normal, royal의 두 개의 등급을 가지는 변수(level)의 고객분류모델은 그 정확도가 (10,228+4482)/21,496=68.4%로 나타났다. 그러므로 본 분석에서는 2등급 한 추정모델이 3등급 한 추정모델보다 높은 정확도를 나타냈으므로, 향후 모든 연구결과는 고객을 2등급 분류한 모델을 사용하였다.

본 연구에서는 이 두 개의 등급을 로열고객(royal)과 일반고객(normal)으로 나누었다. 로열고객으로 분류된 고객은 총 지출액이 270,000만원 이상이거나 한번 방문 시 아이템을 구매한 아이템의 개수가 3개를 넘거나 한번 방문 시 아이템 구입을 위해 지출한 비용이 203,700원 고객이었다. 이에 대한 마이닝의 분석 흐름도는 다음 <그림 1>과 같다.

본 분석에서는 회귀분석(Regression, speciality Logit



<그림 1> Level 2의 추정모델에 대한 분석 흐름도(e-miner)



<그림 2> 모형적합도 평가표(Lift Chart)

Regression), 의사결정나무(Decision Tree), 인공신경망 (Artificial Neural Network) 등 3가지 기법을 적용해 보았다. 그러나 인공신경망 기법의 경우 시스템이 지속 다운되면서 분석이 불가능하였다. 이러한 결과는 아마도 6만 여 개의 많은 분석 데이터로 인해 시스템 부하 때문인 것으로 추정되지만 정확하지는 않다. 모델의 적합도를 Lift Chart를 통해 평가한 결과, 로지스틱 회귀분석 모형이 의사결정나무보다 높은 모델 적합도가 나타났다. <그림 2>와 같이 회귀분석 모델은 상위 20% 우수고객과 60~90% 수준의 우수고객에서 의사결정나무보다 높은 모델 적합도를 나타내고 있다.]

로열고객과 일반고객을 분류하는 Binomial Model에서 로지스틱 회귀분석결과, 로열고객은 9,313명, 일반고객은 12,183명이다. 이 분석모형의 적합성을 평가하

<표 7> Logit Regression 모형에 의한 변수별 유의성 검증 결과 I

변 수	DF	Wald Chi-Square	Pr>Chi-Square
연령	1	17.6302	0.0001*
지역	15	247.6016	0.0001*
성별	1	1.2214	0.2691
제품(복종)	8	180.4090	0.0001*
제품(스타일)	17	5973.1251	0.0001*
구매일자	10	238.5379	0.0001*

* p < 0.005, 사후 검정결과 p < 0.005 수준에서 유의한 차이가 있는 경우 **로 표시함

기 위해 Dual Quasi-Newton Optimization(최적의 값을 찾아내는 알고리즘)을 사용한 결과, 최대 우도비(-2 Log L, 범주형 자료를 분석할 때 사용하는 검증통계량)가 p<.0001로 분석되어 본 모형이 유의한 것으로 나타났다으며, AIC와 SBC의 값이 별점에서 첨부한 Number of Response levels이 3개인 경우보다 작게 나와 타 모형보다 좋은 모형으로 평가되었다.

<표 7>은 로지스틱 회귀분석을 통해 2개 수준의 고객분석 모형을 활용하여 각 변수들에 대한 유의성을 검증한 결과이다.

<표 7>에 의하면, 연령, 지역, 스타일별, 판매 월별에 대해서는 0.05보다 작은 p-value로 본 모형에서는 의미가 있는 것으로 분석되었으나, 성별에 대해서는 p-value가 0.2691로 나타나 무의미한 것으로 분석되었다. 이러한 기준을 기반으로 변수에 대한 상세 분석을 한 결과, <표 8>와 같이 나타났다. A사의 경우 연령이 많

Predicted: segment=royal	Predicted: segment=norma
0.2501910625	0.7496089375
0.2849868813	0.7150131187
0.450659759	0.549340241
0.1626164254	0.8373835746
0.69909615	0.30090385
0.4843403679	0.5156596321
0.2482882669	0.7517117131
0.3211530086	0.6788469914
0.2083088976	0.7916911024
0.1589634083	0.8410365917
0.9826060668	0.0173919332
0.2433792193	0.7566207807
0.367375094	0.632624906
0.320072293	0.679927707
0.726611972	0.271388028
0.2754047268	0.7245952732
0.3470890967	0.6529109033
0.5724606362	0.4275393638
0.2403067343	0.7596932657
0.4339394112	0.5660605888
0.2601036253	0.7398963747
0.3577839558	0.6422160442
0.2088298762	0.7911701238
0.8565426958	0.1434573042
0.2058685612	0.7941314388
0.4970418932	0.5029581008
0.5305266431	0.4694733569
0.2374597198	0.7625402802
0.5434413534	0.4565586466

<그림 3> Logit Regression 모형에 의한 개별 고객 스코어링 결과

<표 8> Logit Regression 모형에 의한 인구 통계적 변수별 유의성 검증 결과 II

구분	Estimate(확률)	Expect(기대치)	Pr>Chi-square	구분	Estimate(확률)	Expect(기대치)	Pr>Chi-square		
전체	3.5208	33.813	0.8381	전체	3.5208	33.813	0.8381		
연령	0.00667	1.007	0.0001*	성별	-0.0183	0.982	0.2691		
지역별	강원	-0.0278	0.973	0.6988	제품	자켓	1.6031	4.968	0.0001*
	경기	-0.2310	0.794	0.0004*		면바지	1.1422	3.134	0.2866
	경남	0.0728	1.075	0.3424		울바지	0.2210	1.247	0.0010*
	경북	0.1960	1.217	0.0016*		반바지	-10.5165	0.000	0.7806
	광주	0.6493	1.914	0.0001*		코트	3.6270	37.599	0.9277
	대구	0.2139	1.238	0.0063		점프	1.0273	2.793	0.0001*
	대전	-0.4937	0.610	0.0235		사파리	1.1231	3.074	0.0001*
	부산	0.5404	1.717	0.0001*		가죽	11.0171	999.000	0.7446
	서울	-0.1454	0.865	0.0417		다운	1.5207	4.576	0.9696
	울산	0.4398	1.552	0.4963		무스탕	10.9778	999.000	0.8070
	인천	-0.3962	0.673	0.0001*		면니트	-8.8251	0.000	-
	전남	-0.0053	0.995	0.9336		울니트	-10.2663	0.000	0.7857
	전북	-0.1887	0.828	0.0043*		기타니트	0.1408	1.151	0.0691
	제주	-0.1588	0.853	0.2931		반팔니트	-10.0925	0.000	0.7892
충남	-0.3697	0.691	0.0001*	면스웨트	-0.1053	0.900	0.1409		
구매월	09월	0.9238	2.519	0.0001*	울스웨트	-11.1714	0.000	-	
	10월	0.7906	2.205	0.0001*	셔츠긴팔	0.3027	1.354	0.0109	
	11월	0.8717	2.391	0.0001*	셔츠반팔	0.7211	2.057	0.0004*	
	12월	0.8278	2.288	0.0001*					
	01월	0.4701	1.600	0.0001*					
	02월	0.1064	1.112	0.2134					
	03월	-0.3394	0.712	0.0001*					
	04월	-0.6157	0.540	0.0001*					
05월	-0.8488	0.428	0.0001*						
06월	-0.9700	0.379	0.0001*						

* p< 0.005, 사후 검정결과 p< 0.005 수준에서 유의한 차이가 있는 경우 *로 표시함

을수록 로열고객이 될 확률이 높아지고, 광주와 부산, 울산 고객이 충성고객이 될 확률이 높아진다는 것이다. 반면, 대전> 인천> 충남> 경기> 전북> 서울> 강원 순으로 로열고객이 될 확률이 낮아지는 것으로 나타났다. 가죽제품이나 무스탕 제품 등 고가의 제품을 구매한 고객이 로열고객이 될 확률이 높고, 반면에 울스웨트나 반바지 등을 구매하는 고객은 로열고객이 될 확률이 적다. 이상의 정보만을 보더라도 A사의 고객은 연령대가 높고, 높은 연령대의 고객이 선호하는 제품

군에서의 구매가 로열고객으로 확보할 수 있는 마케팅 제품으로 판명되었다. 또한 구입일자에 따른 로열고객이 될 확률을 살펴보면, 가을 및 겨울제품을 구입한 고객이 봄·여름제품을 구입한 고객보다는 로열고객이 될 확률이 높은 것으로 나타났다. 특히 9월> 11월> 12월> 10월> 1월> 2월 순으로 로열고객 될 확률이 높게 나타났다. A사는 향후 가을에 집중적인 판촉 마케팅을 통해 핵심고객을 확보하는 것이 효율적일 것이다.

<그림 3>는 본 논문이 제시한 고객 등급 구분 모형

<표 9> 상위 20% 고객의 인구통계 특성

구 분		고객 수	(%)	구 분		고객 수	(%)
총고객		4215	100.00	총고객		4215	100.00
성별	남자	2654	62.67	지역별	강원	326	7.70
	여자	1581	37.33		경기	413	9.75
연령별	19세이하	5	0.12		경남	260	6.14
	20세~24세	71	1.68		경부	692	16.34
	25세~29세	407	9.66		광주	74	1.75
	30세~34세	661	15.68		대구	178	4.20
	35세~39세	633	15.02		대전	2	0.05
	40세~44세	770	18.27		부산	414	9.78
	45세~49세	669	15.87		서울	275	6.49
	50세~54세	459	10.89		울산	1	0.02
	55세 이상	540	12.81		인천	113	2.67
					전남	536	12.66
					전북	492	11.62
					제주	45	1.06
					충남	277	6.54
					충북	137	3.23

을 기준으로 SAS의 Logit Regression한 결과 중의 일부이다. 본 모형을 기준으로 21,496명에게 개별적으로 로열고객이 되는 확률과 일반고객이 될 확률(양 확률을 합하면 1이 됨)을 산정 한 결과에서 <그림 3>에서 12번째의 고객이 로열고객이 될 확률(0.9826080668)이 가장 높게 나타났고, <그림 3>의 11번째 고객이 로열고객이 되지 않을 확률(0.1589634083)이 가장 높게 나타났다. 이렇게 산정된 개별고객의 고객등급을 통해 로열고객의 특성을 분석하여 향후 신규고객을 창출할 때 로열고객이 될 확률을 미리 예측할 수 있고, 이 예측을 통해 일대일 마케팅을 지원함으로써 효과적인 고객관리가 될 수 있는 것이다.

예로써 A사의 경우 <그림 2>의 Lift 결과에 따라 기회비용이 동일할 때 상위 20%에 대해서 집중 고객관리를 한다면 가장 높은 효율을 산출할 수 있다는 것을 알 수 있다. 다음 <표 9>은 선택된 로지스틱 회귀모형에서 고객 각각에 대한 점수를 구한 후 상위 20%의 고객 프로파일을 살펴본 것이다. 로열고객의 특성은 30대 후반 및 40대 초반의 경북, 전남, 전북지역에서 9

월부터 12월 사이에 구매로 나타났다.

V. 결론 및 제언

1. 결론

국내 최대의 고객 데이터 베이스를 보유한 A사의 지난 2000년 9월부터 2001년 7월까지 11개월간 '고객 마일리지 카드'로 구입한 고객 21,496명중, 43.3%인 9,313명이 로열고객이 될 확률이 높은 것으로 나타났다. 이러한 결과는 총 지출액이 270,000만원 이상이거나 한번 방문 시에 아이টে를 구매한 아이টে의 개수가 3개를 넘거나 한번 방문 시에 아이টে 구입을 위해 지출한 비용이 203,700인 고객을 로열고객으로 구분하고 그렇지 않은 경우를 일반고객을 분류하는 Binomial Model에서 로지스틱 회귀분석모형을 적용한 결과이다. 본 모형은 의사결정나무와 인공신경망 모형보다 상대적으로 모형 적합도가 높고, 유의미한 것으로 나타났다.

본 연구결과 A社は 다음과 같은 유형의 데이터웨어 하우스와 CRM 분석 시스템을 도입하는 것이 효율적인 CRM 전략을 전개할 수 있을 것이다.

첫째, 고객 데이터를 수집할 때는 적어도 주민등록증에 기재된 내용을 빠짐없이 확보하여야 할 것이다. 주민등록증을 통해 정확한 성별과 연령, 거주 지역 등은 확실하게 확보될 수 있다.

둘째, 구매고객이 구매 제품에 대한 사용도를 기록하여야 한다. 구매고객이 본인이 착용할 수도 있지만 가족 및 주변인에게 선물할 경우도 많다. 그러므로 사용도에 대한 구분이 없을 경우, 장바구니 분석을 통한 연관구매 마케팅이나 고객 구매력에 따른 판매원의 상품제시에 혼란이 있을 수 있다.

셋째, 미래 창출고객에 대한 고객등급 우선 구분을 위해서는 적어도 체형특성, 예로써 키와 몸무게정도의 정보를 확보하여야 하고, 필요하다면 간단한 패션의식관련 질문지를 통해 선호 패션감도별 고객 특성정보를 확보한다면 매우 효율적일 수 있다.

넷째, 고객 등급 구분을 위한 모형으로 기존 RFM 공식보다는 본 논문에서 제시한 다음의 모형이 패션 상품에 대한 고객의 기여도를 측정하기 위해 유효한 것으로 판정되었다.

$$Y_{ij} = Q_{ij} \cdot EP_{ij} \cdot EQ_{ij}$$

Y_{ij} = t기간동안 j社の 제품을 구매한 i번째 고객의 구매 충성도

Q_{ij} = t기간동안 j社の 제품을 구매한 i번째 고객이 구매한 제품의 총구입금액

EP_{ij} = t기간동안 j社の 제품을 구매한 i번째 고객이 매장에 한 번 방문하여 구입한 제품 개수

EQ_{ij} = t기간동안 j社の 제품을 구매한 i번째 고객이 하나의 제품을 구입하기 위해 지출한 비용

본 연구는 문헌연구는 물론, 패션기업이 보유한 6만여 개의 실제 고객정보에 대한 분석을 중심으로 이루어졌다. 본 연구의 분석도구로는 SAS(Statistical Analysis System)의 e-miner를 사용하였고, 주로 마이닝 기법을 통해 다층적인 분석 단계를 거쳤다. 첫째는 보유 데이터의 문제점과 보완 방법을 파악하였고, 둘째는 데이터의 정제를 통해 기존 고객의 구매력을 평가하여 고객 등급화를 통해 핵심고객을 추출하는 방법

론을 개발하였고, 셋째는 고객의 연관 구매력을 평가하여 기업의 경영성과를 높이는 기법을 도출하였다.

이에 본 연구는 패션기업의 경영관리에 중요한 대안으로 부상하는 핵심고객창출에 대한 방법론 연구에 대한 도움이 될 것으로 판단된다. 또한 CRM 시스템을 도입하는데 기초 기반이 되는 데이터웨어하우스의 최소 필요단위의 데이터베이스 정보를 제공함으로써 기업의 구축비용 및 도입 실패율을 개선시키는데 일익을 담당하는 것에 그 의의를 두었다.

2. 분석결과 활용 및 제언

본 연구는 A社가 향후 CRM 및 e-CRM을 위한 Data-warehouse 구축에 따른 최적 분석모형을 제언코져 설계되었다. 향후 A社를 위한 고객 Data-warehouse에는 다음과 같은 준비가 있어야 될 것이다.

첫째, 고객에 대한 마케팅실행전략을 최우선 순위로 수립하여야 할 것이다. 즉 marketing action planning에 따라 분석 모형 혹은 분석을 위한 변수의 유형이 정해 질 수 있기 때문이다.

둘째, 고객을 설명하는 filed 정보에 대해서는 정확한 정보입력이 필수적이다. 특히 구매고객이 구입고객인지 혹은 실제 사용고객인지가 불분명함에 따라 결과의 오류가 많아지기 때문이다.

셋째, 본 논고에서 사용한 data는 DB구조를 기반으로 한 시스템으로 DBMS 분석을 위해서는 통계구조로의 변환이 필요하다. 그러므로 상당한 통계적 지식 및 결과를 분석할 수 있는 전문가에 의한 시스템 구축이 필요하다.

이러한 분석시스템을 위해서는 본 논고가 상당히 부족한 점이 많다.

첫째, 수집된 데이터의 불안정성이다. 즉 데이터가 본 논고가 출시되기 2년 전의 것으로 조금 오래되었으며, 1년도 되지 않은 단기간의 데이터라는 것이다. 둘째, 데이터에서 고객에 대한 설명 field를 좀 더 찾아야 할 것이다. 본 논고에서 분석한 3개의 설명력으로 고객을 구분한다는 것은 분석의 오류가 많아지는 것은 물론, 미래 잠재고객이 등록할 경우 구매 시기가 11개월 이상이 되어 구매이력이 있을 경우 고객

scoring이 가능하다는 점이다. 만일 고객의 특성, 예로써 패션에 대한 의식, 소득정도, 가족이력 등이 같이 평가된다면 미래고객의 구매신용도 평가가 용이 할 것이다. 셋째, 본고는 과거 구매경력만을 평가한 것으로 모델의 미래 적중도를 평가하지는 않았다. 즉 2001년 8월 이후의 고객 구매데이터로 본 모형의 적중성을 평가하지 못했다. 네째, 분석 사례 기업으로 A社 한 개만을 보았기 때문에 도출된 결과의 일반화하기에는 한계가 있었다.

향후 본 논고의 문제점을 심도있게 파악하여 좀 더 깊고 세밀한 분석을 하여야만 패션 기업이 구축할 고객 Data-warehouse의 효과성 및 효율성이 제고될 것이다.

오늘날 많은 기업이 CRM이라는 새로운 경영기법에 화두를 모으고 있다. 그러나 이러한 CRM을 하기 위해서는 과거 장기적인 고객 데이터가 축적이 되고 또한 정확한 데이터의 축적이 있어야만 가능하다. 특히 패션 산업의 전반적인 산업 영세성으로 대다수의 기업이 CRM의 중요성을 강조하고 있으나, 장기적인 정확한 데이터 축적에는 많은 투자와 인내심이 기반이 될 것이다. 패션산업이 모든 국가의 산업기반이듯이 A社와 같은 대기업이 아닌 중소기업 기반을 확고히 하기 위해 국가적인 차원에서의 고객 Data-warehouse 구축 및 CRM 모델 개발 지원이 필요하다. 21세기 국가 경쟁력은 지식축적을 통한 고객중심형 경영이 될 것이다.

참고문헌

- 1) 2002년 패션시장전망보고서, 삼성패션연구소. 2002. 1, pp. 4-6.
- 2) 유필화 외. 판매원의 고객지향지수 개발 및 판매원 성과와의 상관관계 분석. 소비자학연구, 7, 2(96.11), pp. 59-85.
- 3) 윤지영. 고객을 자본화하는 관계마케팅. 廣告情報, 205(984), pp. 77-81.
- 4) 이도현, 데이터 마이닝을 이용한 CRM. 情報科學誌, 138(2000.11), pp. 4-11.
- 5) 오세조 외, 고객 관계지향성 형성에 관한 연구. 1999 유통연구. 4, 2(99.12), pp. 41-58.
- 6) 성공적인 CRM 구축방법. 2001. 9, SAS 세미나.
- 7) 김성수. (2001). 국내기업의 성공적인 고객관계 관리(CRM) 구현 방안에 관한 연구. 고려대 경영대학원 석사학위논문, pp. 16-18.
- 8) 김광수, 관계마케팅을 통한 고객단골화, 마케팅. 345(97.10), pp. 80-85.
- 9) 이재문. 고객관계관리(CRM) 솔루션의 현황 및 구축. MARKETING COMMUNICATION REVIEW, 11(2001.6) pp. 3-22.
- 10) 이재문. 고객관계관리(CRM)의 이해 및 활용: 실행 Process 및 Tool을 중심으로. MARKETING COMMUNICATION REVIEW, 10(2000.12), pp. 3-33.
- 11) 데이터베이스월드, 98년 8월호.
- 12) 이도현, 데이터 마이닝을 이용한 CRM. 情報科學誌, 138(2000.11), pp. 4-11.
- 13) 이승민. (1999). Data mining을 위한 속성변수 추출방법 개발/亞洲大 大學院 석사학위논문.
- 14) 성공적인 CRM 구축방법 : 현대백화점 CRM 사례, 2001. 9, SAS 세미나.
- 15) 이승민. (1999). Data mining을 위한 속성변수 추출방법 개발/ 亞洲大 大學院 석사학위논문.
- 16) 권오준. (1997). 통계적 기법과 인공지능망을 이용한 신용카드 고객 신용도 평가 모형에 관한 연구. 한국과학기술원 테크노경영대학원 석사학위논문.
- 17) 서은호. (2000). E-Marketplace에서 고객 성향분석을 위한 웹마이닝기법의 적합성 측정. 포항공대 대학원 석사학위논문.
- 18) 김기운 외. 데이터 웨어하우스 기반의 고객관계 관리 모델링: 프로세스 및 데이터 관점. Information Systems Review. 2, 2(2000. 12), pp.

- 283-299.
- 19) 김성수. (2001). 국내기업의 성공적인 고객관계 관리(CRM) 구현 방안에 관한 연구. 고려대 경영대학원 석사학위논문.
- 20) A社の 지난 2000년 9월부터 2001년 7월까지의 마감 매출은 총 780억이었으나, 본 데이터에서 합한 매출은 60억정도로 A社 총매출의 30%내 외임. 매출과 고객수가 일치할 수는 없지만 추정해 본다면, 본 데이터 고객이 21,500명으로 나타났으므로 결국 A社 총고객수는 6만명 내외로 추정됨.
- 21) 의류제품은 남성복, 여성복, 캐주얼 등 복종구분이 있으며, 각 복종에는 셔츠, 바지, 자켓 등의 스타일로 세분된다. A社は 품목구분을 위한 POS 번호에는 스타일구분, 원단구분, 색상구분, 상품 출고기간 구분 등으로 6가지로 세분되어 있다.
- 22) 이경전·이가은. 인터넷 기반 고객 관계 관리의 비즈니스 모델과 프로세스 모델에 관한 연구, 情報科學會誌, 138(2000. 11), pp. 12-21.
- 23) 핵심가치의 제공은 세분화된 각 고객집단별로 포지셔닝한 가치를 모두 포함하는 기업의 상품 및 서비스에 대한 핵심가치에 대한 제안을 구체화하는 것이다.
- 24) 본 연구에 사용된 고객 데이터는 A社가 직접 운영하는 점포의 POS시스템을 통해 수거한 것으로, 백화점 및 전문점 등 제 3의 중간 유통업체가 중개상으로 있는 점포에서는 고객 데이터가 수거되지 않았음. 그러므로 2000년 9월부터 2001년 7월까지 11개월 동안 발생된 매출 6,457,802,140원은 A社の 전체 매출에서 일부분으로 이해해야 된다.
- 25) 김형주(2001). 고객관계 관리를 위한 연관규칙 발견 및 확장. 아주대 정보통신대학원 석사학위논문.
- 26) 노경호. 인터넷 기반 고객관계관리의 전략적 도입에 관한 연구. 경영정보연구 5(2000. 12), pp. 61-79.
-
- (2002년 11월 18일 접수, 2002년 12월 9일 채택)