

碩士學位論文

데이터마이닝을 이용한
쇼핑몰고객 이탈방지 향상기법



濟州大學校 大學院

컴퓨터工學科

任 晶 弘

2006年 6月

데이터마이닝을 이용한 쇼핑몰고객 이탈방지 향상기법

指導教授 金 壯 亨

任 晶 弘

이 論文을 工學 碩士學位 論文으로 提出함.

2006年 6月



任晶弘의 工學 碩士學位 論文을 認准함.

審査委員長 安 基 中 印

委 員 郭 鎬 榮 印

委 員 金 壯 亨 印

濟州大學校 大學院

2006年 6月

**A Techniques to enhance the Prevention of
Defection by Shopping Mall
Customers using Data Mining**

Chung-Hong Lim

(Supervised by professor Jang-Hyung Kim)



**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENT FOR THE DEGREE OF
MASTER OF ENGINEERING**

**DEPARTMENT OF COMPUTER ENGINEERING
GRADUATE SCHOOL
CHEJU NATIONAL UNIVERSITY**

2006. 6.

목 차

SUMMARY	v
I. 서 론	1
II. 데이터마이닝	5
1. 데이터마이닝 개요	5
2. 데이터마이닝의 활용	7
3. 데이터웨어하우스 & OLAP	8
4. 데이터마이닝과 CRM	10
5. 고객 세분화 및 차별화 전략	15
6. 마케팅 CRM에서 데이터마이닝 기법	18
III. 데이터마이닝을 이용한 고객 이탈방지 알고리즘 설계	25
1. 효율적인 고객 이탈 방지를 위한 방법	26
1-1 기존고객 이탈 방지 방법의 문제점	28
1-2 사례를 통한 작업 흐름도	30
1-3 고객 산출 및 세분화	35
1-4 고객 패턴분석 및 이탈 원인분석	38
1-5 고객세분화를 통한 이탈방지 마케팅 전략	40
2. 향상된 알고리즘을 이용한 고객 세분화 설계	41

IV. 구현 및 실험 결과	45
1. 시스템 환경	45
2. 제안 알고리즘 평가	46
3. 실험 결과 분석	46
V. 결 론	49
참고문헌	50



[그림 목차]

Fig. 1	데이터마이닝의 활용분야	8
Fig. 2	데이터웨어하우스와 데이터마트.....	9
Fig. 3	분석 및 운영 CRM	14
Fig. 4	고객 세분화의 개념.....	16
Fig. 5	고객 세분화에 따른 마케팅 전략 수립.....	17
Fig. 6	ECMiner사의 데이터마이닝 기법의 분류.....	18
Fig. 7	의사결정나무의 지식표현.....	20
Fig. 8	인공신경망의 원리.....	22
Fig. 9	인공신경망의 구조.....	23
Fig. 10	고객 세분화의 목적.....	27
Fig. 11	본 연구 흐름도.....	28
Fig. 12	군집분석 결과	32
Fig. 13	신용카드업체의 고객이탈 정의사례.....	33
Fig. 14	분석 흐름도에 따라 수행결과 구성된 흐름도.....	34
Fig. 15	구축된 두가지 모형을 이익도표 이용 향상된 모형 구축.....	34
Fig. 16	신경망 분석에 의한 이익도표.....	35
Fig. 17	고객 세분화의 흐름도.....	36
Fig. 18	RFM 구매.....	37
Fig. 19	RFM Score.....	37
Fig. 20	고객패턴 및 이탈분석.....	38

Fig. 21	고객세분화를 통한 이탈방지 마케팅 전략 흐름도	41
Fig. 22	구매패턴과 구매이력의 흐름도	41
Fig. 23	고객 이탈분석 프로세스의 흐름도	42
Fig. 24	쇼핑몰에 적용한 상품재배열 흐름도	43
Fig. 25	접속통계 대한 내용	44
Fig. 26	시스템 구현 화면	46
Fig. 27	성별 및 성향 분포	46
Fig. 28	재구매 건수 / 이탈방지 통한 구매 분포	47
Fig. 29	연 구매횟수 / 연 재구매 횟수.....	48
Fig. 30	월 구매횟수 비교 / 월 재구매 건수비교.....	48



[표 목차]

Table. 1 고객관리의 시대적 변천.....	11
Table. 2 국내/외 연구기관들의 CRM 정의.....	12
Table. 3 정량조사를 위한 조사 설계.....	31
Table. 4 RFM Table.....	36
Table. 5 RFM 정의 및 세분화 기준.....	38
Table. 6 Experiment Environments.	45



Summary

The popularization of the Internet and the development of information communication technology today caused explosive increase of customer related data. Although these data have potentials to be used very valuably for forecasting change of external environment or special change of customer behavior, the potentials are not being used sufficiently. There are many experts in technical areas of management, but almost of them do not have in-depth penetration for management process, understanding of information technology and data analyzing ability comprehensively. In particular, each data are not being integratively managed as data with the identical viewpoint by being connected with each other. Also understanding and use of statistical tools to induce various knowledge using these data are insufficient. Accordingly, the importance of marketing activities focusing on various customer managements under a perception that accurate understanding of customers and proper service are the essences of company's competitiveness is emerging. From this viewpoint, the needs for datamining that enables to analyze large-size data and provide various services by characteristics of customer will be continuously expanded and continuously highlighted as a functional tool supporting various decision-making.

The current study aims to improve an algorism in which customers' connection record, purchasing record and personal information are statistically analyzed by datamining method, the customers are subdivided, and the customers can automatically organize the display of goods according to popularity level in order to design and realize an intelligent internet shopping mall for strategical marketing.

Through this system, the existing shopping mall management handled manually by subjective judgement by a shopping mall manager can be automatized. In addition, new marketing methods that can reinforce competitiveness in rapidly increasing electronic market can be suggested.

I. 서 론

오늘날 인터넷의 확산과 정보 통신 기술의 발달은 고객 관련 데이터의 폭발적인 증가를 가져왔다. 이러한 데이터는 외부환경변화의 추세나 특별한 고객 행동의 변화를 예측하는데 매우 유용하게 쓰일 수 있는 잠재력이 있으나 이러한 잠재력이 충분히 활용된다고 볼 수 없다. 이러한 이유로는 크게 두 가지로 나누어 볼 수 있는데 첫째, 각 데이터가 서로 연관되어 단일 관점의 데이터로 통합 관리되지 못하고 있다는 점이다. 최근에 데이터 퓨전(Data Fusion)이라는 용어로 소수의 기업들이 각 분야에서 이러한 단일 고객 관점으로 본 데이터통합 노력은 하고 있지만 여전히 총체적인 데이터 통합관리는 아직 많이 진전되지 못하고 있는 상황이다. 둘째, 이러한 데이터를 이용하여 다양한 지식을 끌어낼 수 있는 통계적 도구에 대한 이해나 활용이 미흡하다는 것이다. 경영의 기능적 분야에 대한 전문가가 많이 있어도 조직에서 경영 프로세스에 대한 깊은 통찰력, 정보기술에 대한 이해, 데이터 분석 능력을 종합적으로 갖고 있지 못한 경우가 대다수라고 할 수 있다.

이러한 문제점들이 최근에는 많이 극복되어지고 있는데, 그 현상으로 첫째, 하드웨어와 소프트웨어 성능은 향상되고 가격은 점점 낮아지면서 기업이 고객의 데이터에 점점 쉽게 접근하고 획득이 가능해지고 있다는 것이다. 특히 유비쿼터스 사회로 고객에 대한 데이터는 폭발적으로 늘어날 것으로 보인다. 둘째, 기업의 대용량 데이터를 취급하고 분석하는 문제에 대하여 컴퓨팅과 통계학의 분야를 필두로 다양한 방법론적 연구가 계속 진행되고 발전하고 있다는 것이다. 특히 고객에 대한 정확한 이해와 적절한 서비스가 기업경쟁력의 핵심이라는 인식하에 다양한 고객관리 중심의 마케팅 활동의 중요성이 요구되는 것도 이러한 연구의 확산에 일조하고 있다. 이러한 관점에서 대용량 데이터를 분석하고 고객특성별로 다양한 서비스를 가능하게 하는 데이터 마이닝에 대한 요구가 계속 확산될 것이며 여러 가지 의사결정을 지원하는 지능형 도구로서 계속 각광을 받을 것이다.

시장의 규제완화로 인하여 새로운 시장으로의 진입기회가 늘어남에 따라 동일업종에서의 경쟁사가 많아지기 시작하였다. 그래서 고객세분화를 통하여 서로 다른 고객

집단의 필요와 요구를 정확히 파악하고 이들이 원하는 상품과 서비스를 차별적으로 제공하지 않는 기업은 경쟁력을 잃고 도태될 수밖에 없다. 이러한 시장 상황 하에서 고객에 대한 정보를 바탕으로 전략적인 고객세분화를 하여 목표 고객군에 더 나아가서 개별화된 목표 고객을 설정하고, 그에 적절한 마케팅 믹스를 적용하고 실행하여 기존 고객과의 관계를 강화해야할 필요성이 대두되었다.

CRM(고객관계관리)이 등장하게 된 환경적 요인으로 컴퓨터 및 그에 관련된 정보기술의 급격한 발전을 들지 않을 수 없다. 우선 컴퓨터 하드웨어의 저장용량이나 데이터 처리속도 면에서 과거와는 비교가 되지 않을 정도로 컴퓨터의 성능이 향상 되었다. 이를 통해 기업들은 낱알이 늘어나는 방대한 양의 고객 거래 데이터를 데이터 웨어하우스라 불리는 시스템에 특정체계를 갖추어 자동적으로 내장하고 관리할 수 있게 되었다.

최근 들어 어느 시장을 막론하고 고객들이 갖고 있는 기대와 요구가 다양화되고 있다. 그럼에도 불구하고 고객의 요구는 고객조차 모를 정도로 끊임없이 변화하여 고객 만족은 더욱 더 복잡해져 가고 있는 실정이다. 기업이 이러한 고객의 기대에 부응치 못하면 고객은 다른 업체로 옮겨간다. 뿐만 아니라 고객의 가치 및 라이프스타일도 시간의 흐름에 따라 매우 달라지고 있는데, 그 대표적인 것이 편리성의 추구이다. 이런 욕구를 충족시키기 위해 기업들이 고객 중심적인 경영방식인 CRM의 도입을 서두르고 있는 것이다.

기업은 끊임없이 고객의 니즈를 파악하려 애를 쓸 것이고 서비스를 제공하는데 있어 가장 효율적이면서 효과적인 방법을 찾으려고 노력 할 것이다. RFID와 같은 태그를 인식하는 다양한 형태의 센서와 위치 추적장치를 이용하여 고객에게 즉각적이면서 상호적인 응답을 하게 될 것이다. 따라서 이제는 데이터를 모아 분석한 후에 그 특성에 맞추는 서비스를 제공하는 개념을 벗어나 즉각적인 대응이 가능하게 실시간 데이터마이닝 분석이 필요할 것이고 그런 면에서 각 분야별로 많은 데이터마이닝의 특성이 이미 가미된 지능형 칩의 개발이 계속 확산 될 것이다.

이렇게 변하는 경영환경 하에서 기업들은 다음과 같은 문제를 해결하기 위하여 많은 노력을 할 것이다.

첫째, 어떻게 하면 자사의 상품과 서비스가 제공되는 지역을 넓힐 것인가?

둘째, 어떻게 고객의 요구를 빨리 파악할 것인가?

셋째, 어떻게 각 지역별로 자사의 자원을 최대한 효율적으로 이용할 것인가?

넷째, 어떻게 고객의 개인 프라이버시를 침해하지 않으면서 고객과의 관계를 좋게 유지하여 갈 것인가?

물론 이러한 문제는 아직 닥쳐오지 않은 사회에 대한 예측이지만 기본적으로 고객 관계관리에 있어 데이터마이닝의 역할은 더욱 더 증대될 것이라고 예측된다. 실제 과거에 비하여 사물에 컴퓨터가 내장되고 사물과 사물, 사람과 사물 간에 네트워크가 이루어지면 현재보다 수십배, 수백배의 고객 데이터가 수집될 것이고 특히 이러한 데이터를 실시간으로 처리하여 고객 특성별로 시간, 장소, 채널에 관계없이 정확한 서비스를 전달하려면 현재보다 보다 정확하고 빠른 데이터 분석 능력이 필요할 것이며, 훨씬 더 복잡한 여러 가지 환경인자들을 고려하여 데이터마이닝을 하여야 할 것이다.

다시 한번 “언제 어디서나 사용자가 환경에 구애받지 않고 원하는 정보를 주고받을 수 있는 미래의 네트워크 환경”이라는 유비쿼터스 환경하의 고객관계관리를 고려하면 다음과 같은 변화를 예상할 수 있다.

- ① 기업 고객의 위치가 곧바로 비즈니스의 활동지역이 될 것이다. 즉 모바일과 위치추적 장치의 발달로 기업 활동의 지역적 경계는 더 이상 의미가 없을 것이다.
- ② 언제 어디에서든지 기업의 상품과 서비스가 고객에게 전달 가능하도록 할 수 있는 것은 기업경쟁력의 중요한 요소가 될 것이다.
- ③ 모바일 장치와 장비들은 원격으로 비즈니스를 하는 서비스 제공자의 중요한 도구가 될 것이다.

- ④ 서비스 제공자는 24시간 끊임없이 고객의 니즈가 무엇인지를 파악하려고 할 것이다.
- ⑤ 서비스 제공자는 고객과 접촉할 때 매우 신중하고 선택적으로 타겟화하여 시도할 것이다.
- ⑥ 고객에 대한 정보는 그 어느 때보다 강력한 자원이 될 것이며, 고객데이터의 가격은 매우 높게 책정될 것이다.

요약하여 정리하면 향후 기업의 고객관계관리에 있어 중요한 인자로 고객에 대한 정확한 이해, 고객에 대한 접근, 고객에 대한 즉각적인 응대 등이 유비쿼터스 하의 U-상거래 환경에서 특히 중요한 것으로 재조명되며, 이러한 관리를 어떻게 할 것인가는 고객데이터의 수집, 분석, 고객세분화 등 데이터마이닝 방법의 발전과 더불어 해결 가능 할 것으로 보인다. 따라서 고객이탈방지 알고리즘 향상을 위한 데이터마이닝이 필요하며, 여기에 덧붙여 다양한 채널에 대한 고객과 기계의 인터페이스 디자인에도 데이터마이닝 기법의 적용이 많이 예상되고 있다.[1][2][3][4][13][19]



II. 데이터마이닝

1. 데이터마이닝 개요

데이터마이닝(Data Mining)이란 대규모의 데이터베이스로부터 과거에는 알지 못했던, 그리고 데이터 속에서 유도된 새로운 데이터 모델로 미래에 실행 가능한 정보를 추출해내어 중요한 의사결정에 이용하는 과정이다. 외국의 경우에는 은행이나 신용카드사, 보험회사, 유통회사 등 많은 데이터를 보유하고 있는 회사에서 이미 데이터마이닝이 구현되어 쓰이고 있다. 제품과 부서간의 특수 영업으로 판매량을 늘릴 수 있을까? 상품판매 캠페인이 얼마나 효과가 있을까? 특정 고객과의 마케팅 상담을 위한 전략을 어떻게 공식화 할 수 있을까? 등 수 많은 기업들의 고민들을 데이터마이닝 혹은 다른 경험적인 방법으로 이를 해결해 가고 있다. 좀 더 쉽게 생각해보면, 매달 신용카드 결제일이 되면 신용카드 사용내역과 함께 여러 상품의 선전물이 함께 동봉돼 날아온다. 그 선전물들 중에 관심이 가는 제품은 무엇인가? 20대냐, 30대냐, 여성이냐, 남성이냐, 직장여성이냐, 가정주부냐에 따라 관심이 가는 제품이 다 다를 것이다. 바로 데이터마이닝을 통해 이것들을 구분할 수가 있다. 회원의 나이와 성별, 직업에 따라 관심을 가질 제품을 구분하여 필요한 사람에게 필요한 정보만을 전달한다면 경제적, 시간적 그 외 여러 측면에서 훨씬 효율적인 비즈니스가 이루어 질 수 있는 것이다. 광고학 등에서 나오는 개별마케팅의 개념과도 관계되는 것이다. 이런 것이 데이터마이닝이다.

직면한 현상을 기술하기 위해서는 모델이 필요하다. 전통적으로 모델의 구성은 현상을 관찰하여 가설을 설정하고, 예측치와 실제결과를 비교함으로써 가설을 검증하였다. 모델이 부적절한 경우, 즉 예측치와 관찰치가 다른 경우 모델은 수정되거나 재구축되었다. 이러한 모델을 구축하기 위해서는 모델링 하고자 하는 현상을 관찰할 필요가 있다. 그러나 기업은 이미 자신들의 데이터베이스에 이것을 가지고 있으며, 이러한 데이터는 환경을 모델링하는 데 사용될 수 있다. 비록 데이터베이스 내에 모든 환

경이 다 포함되어 있지는 않지만, 데이터베이스 내에는 많은 객체들이 존재하며 객체들 사이에는 많은 상호관계가 존재한다.

모델을 구성하기 위해 다양한 기법들이 사용될 수 있다. 전통적으로 가장 대표적인 기법이 바로 통계분석이다. 그러나 이러한 기법은 모델에 관한 기본적인 아이디어를 이미 가지고 있다는 전제에서 이루어진다. 즉 어떠한 변수가 가장 중요하며, 이러한 변수들 사이의 관계를 설명하기 위해서는 어떠한 함수가 사용되어야 하는지를 알고 있을 때 최상의 결과를 이끌어 낼 수 있다.

반면, 데이터마이닝은 지금까지 질문되지 않았던 질문에 대한 답을 제공해 준다. 우선 데이터마이닝이 무엇인지 이해하기 위해 대표적인 정의들을 살펴보면 다음과 같다.

첫째, 대량의 실제 데이터로부터 묵시적이고 전에는 알려지지 않았지만 잠재적으로 유용한 정보를 추출하는 것(Frawley et al, 1994)

둘째, 대규모 데이터베이스 내에 존재하는, 그러나 대량의 데이터 사이에 숨겨져 있는 상호관련성(relationship)과 글로벌 패턴(pattern)에 대한 탐색(Holshemier & Siebes, 1994)

셋째, 대량의 데이터로부터 패턴 인식기술과 통계기법, 수학적 기법을 이용하여 의미 있는 새로운 상관관계(correlations), 패턴 그리고 추세(trends)를 발견하는 과정(Gartner Group)이다.

데이터마이닝이란 대량의 데이터 사이에 묻혀 있는 패턴을 발견하고 규칙을 추론함으로써, 의사결정을 지원하고 그 효과를 예측하기 위한 기법이다. 예를 들어 마이닝의 결과로 어떤 고객은 다른 고객에 비해 특정 상품을 더 잘 구매하는 경향이 있다는 사실을 알아 낼 수 있다. 기업은 이 두 유형의 고객에 대한 차이를 알게 됨으로써 불특정 대중이 아닌 목표 고객에 집중된 마케팅을 수행할 수 있을 것이다. 마이닝은 사용자의 질의나 보고서가 효과적으로 밝혀낼 수 없었던 데이터 웨어하우스 내의 정보를 밝혀낸다. 데이터마이닝은 데이터베이스 안으로 깊숙이 침투하여, 데이터 내의 패턴을 발견하고 규칙을 추론한다. 이러한 패턴과 규칙은 의사결정을 지원하고 기업 환경의 변화를 예측하는데 사용될 수 있다.

또한 다양한 통신기술을 활용한 매체들, 모바일이나 PDA등을 통한 여러 가지 형태

의 고객 데이터가 축적되고, 고객서비스의 활동채널이 늘어나면서 이제는 진정으로 고객을 이해하고 고객만족, 더 나아가 고객감동을 줄 수 있는 인프라가 점점 갖추어지고 있다고 할 수 있다. 여기에는 유비쿼터스 환경이 도래하면 컴퓨터가 사물에 내장되면서 거의 고객의 모든 일상데이터가 정보로 전환될 것으로 보여 어쩌면 고객자신도 모르는 것을 기계가 먼저 파악하여 서비스를 할 것으로 예상되고 있다.

이런 기업환경의 변화에서 당연히 기업은 경쟁력 강화를 위하여 고객에게 무엇을, 언제, 어디에서, 어떻게 서비스나 상품을 전달할 것이라는 필연적인 문제를 풀려고 노력 할 수밖에 없으며, 이러한 문제를 가장 효율적이고 효과적으로 대처하는 기업이 시장에서 최종 승리자가 될 것이다. 따라서 기업은 자연스럽게 고객의 관계를 재정립하여 잘 관리할 수 있는 노하우를 찾을 수밖에 없으며, 그런 관점에서 고객관계관리(CRM: Customer Relationship Management)가 주목 받고 있는 것이다.

2. 데이터마이닝의 활용

기업경영에서의 데이터마이닝의 활용은 데이터를 취급하는 거의 모든 분야에서 활용된다고 할 수 있다. 굳이 나누어 설명하자면 본 논문의 주된 연구 분야인 CRM 등의 마케팅 분야, 그리고 과거부터 고객자료뿐만 아니라 많은 기업의 재무, 비재무 자료를 잘 정리하였던 금융분야 등은 경영분야의 데이터마이닝 활용의 주된 연구 분야였다.

최근에는 생산현장의 많은 데이터를 취급하는데 있어 6-Sigma 등 품질관리 분야에서도 그 이용이 많이 확산되고 있다. SPC(Statistical Process Control)를 기반으로 한 품질관리에서 프로세스 모니터링, 생산계획 및 스케줄링 까지 그 이용이 계속 확산되고 있다. 또한 인터넷을 통한 많은 고객서비스가 이루어지면서 고객의 웹상에서의 행동 데이터를 웹로그 형태의 데이터(Click Stream Data)로 모아 분석하는 웹마이닝(Web Mining)도 eCRM 분야의 한 요소로 중요한 위치를 차지하고 있다고 할 수 있다. 그 외 Visual Data Mining이나 Multimedia Data Mining 등 그 활용분야가 계속 다양한 분야로 확산되면서 각각 고유한 데이터마이닝의 방법론을 정리하여 가고있다.



Fig.1 데이터마이닝의 활용분야

3. 데이터웨어하우스 & OLAP  제주대학교 중앙도서관
 JEJU NATIONAL UNIVERSITY LIBRARY

데이터마이닝을 설명할 때 당연히 중요한 것 중의 하나가 데이터의 획득 및 저장에 관한 것이다. 어쨌든 데이터마이닝을 하고자 한다면 데이터가 어느 한곳에 저장하여야 하며 그것도 다양한 형태의 데이터가 다양한 채널을 통하여 매우 큰 용량의 데이터 저장소에 모여져서 통합적으로 해석이 가능하고 분석이 용이한 형태로 저장되어야 할 것이다. 그러나 일반적으로 기업이 가지고 있는 데이터들은 다양한 운영체시스템으로부터 다양한 형태로 모아지므로 각기 다른 형식과 코드를 가질 수 있고, 따라서 서로 일치하지 않거나 부정확한 값을 가지고 있는 경우가 많다. 그러므로 먼저 이러한 문제들이 해결되어 양질의 데이터가 보장된 후에야 비로소 효과적인 데이터 마이닝을 수행할 수 있을 것이다. 이러한 점에서 거의 모든 국내 대기업들은 2000년도 전후에 데이터웨어하우스의 중요성을 인식하고 데이터웨어하우스 구축에 많은 노력을 기울였고 이제 데이터 구축에 대한 인프라가 어느 정도 완성되었으므로 향후 데이터마이닝의 활용이 점점 증가하게 될 것이라 예상된다.

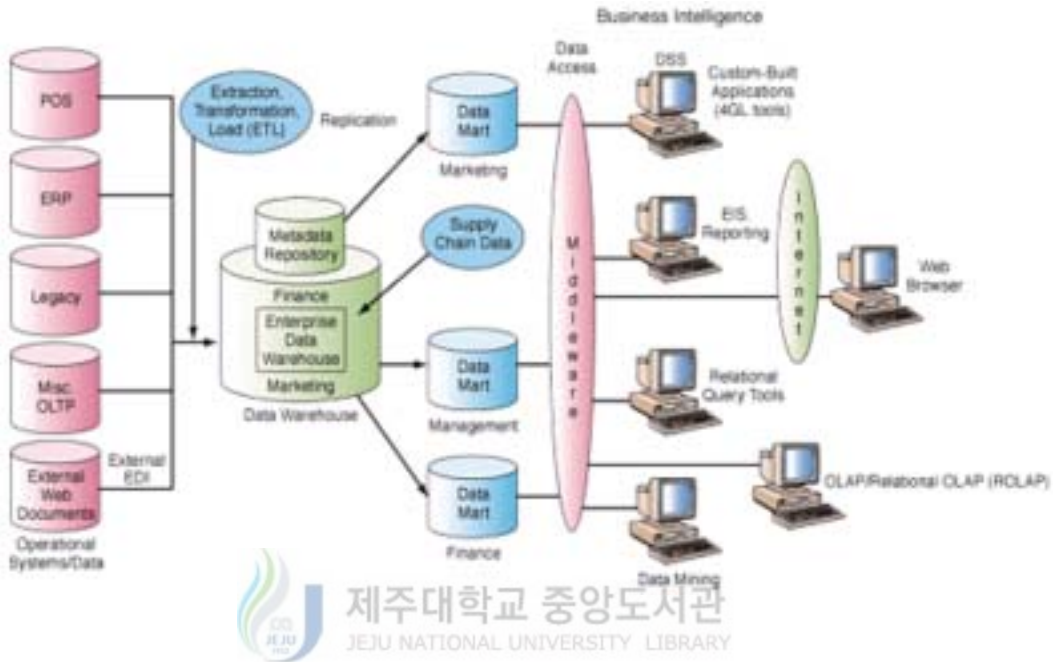


Fig.2 데이터웨어하우스와 데이터마트

[Fig 2]에서 보듯이 데이터웨어하우스는 POS(Point of Sale)나 ERP(전사적자원관리) 등으로부터 추출된 기업의 데이터를 저장하는 것이다. 이러한 데이터웨어하우스가 일반적인 데이터베이스와 다른 것은 물론 엄청나게 큰 대용량 데이터를 저장하는 것과 최소 데이터가 5년에서 10년 이상씩 저장되어 트렌드 비교가 가능한 시변성 (Time Variant), 한번 입력되면 지우지 않는 비휘발성(Nonvolatile), 관계적 구조를 사용하는 관계성(Relational) 등이다.

데이터웨어하우스의 높은 구축비용과 대용량 데이터를 취급하는데 드는 시간과 효율성의 부족으로 인하여 많은 기업에서는 데이터마트를 사용한다. 데이터마트는 데이터웨어하우스에 있는 데이터 중 각 사업단위에 필요한 데이터만 뽑아서 저장한 작은 웨어하우스라고 할 수 있다. 이러한 데이터마트를 사용함으로써 각 사업단위별로 필

요한 데이터마이닝 작업이 가능하며 훨씬 더 효율적이고 빠른 데이터 분석과 의사결정이 가능하여진다. 특히 정보시스템부서에 의존하지 않고 각 사업단위별로 데이터를 취급하여 보안과 유지보수에도 도움이 된다.

[Fig 2]를 다시 보면 이러한 데이터마트를 이용하여 당연히 데이터마이닝 분석이 가능한 데이터가 생성되고 그 이외에도 의사결정시스템(DSS: Decision Support System), 중역정보시스템(EIS: Executive Information System), 그리고 단순질의(Query) 기능을 가진 데이터 탐색 기능 등을 제공하게 된다. 이러한 단순 질의 및 분석을 통하여 얻을 수 있는 정보의 한계를 극복하고 대용량의 데이터베이스나 데이터 웨어하우스를 좀 더 면밀히 탐색할 수 있는 방법으로 등장한 것이 바로 OLAP(On-Line Analytical Processing)이다. 사용자들은 정보의 탐색과정에서 가질 수 있는 의문점들에 대해 가설을 세우고 여러 차원의 즉각적 질의를 통해 가설을 검증하고 확인하면서 풀려나가게 된다. 그러나 이러한 질의로부터 얻어내는 정보는 데이터에 대한 간략한 요약인 것에 반하여 데이터마이닝은 가설 자체를 만들어내고, 데이터를 탐색하는 그 자체를 수행하면서 생각하지 않았던 정보를 발견하기도 한다.

다시 말해 OLAP는 “어느 지역의 영업실적이 우수했는가?” 라는 의문에는 즉각적인 답을 줄 수 있으나, “그렇다면 부진한 지역에서의 원인은 무엇이며 그 해결책은 무엇인가?” 라는 본질적인 의문에는 답을 주지는 못한다. 이러한 의문은 바로 데이터마이닝을 통해 해결할 수가 있다. 즉, OLAP가 사용자가 보고자 하는 관점에서의 자료의 요약이라면 데이터마이닝은 원인과 결과를 연결하는 지식의 발견이라고 할 수 있다. 대용량 데이터를 기반으로 유용한 정보를 얻고자 하는 목적에 있어서 OLAP을 이용하여 얻을 수 있는 정보를 기반으로 데이터마이닝에 접근한다면, 보다 효과적으로 고급정보를 얻을 수가 있다. 또한 데이터마이닝의 결과는 데이터베이스에 적용되어 OLAP를 통해 그 효과가 극대화 될 수 있다.

4. 데이터마이닝과 CRM

1) CRM 등장배경

과거 상품의 차별성이 거의 없던 매스마켓이 지배하던 상황에서는 기업이 생산만

하면 소비가 되었으므로 규모의 경제(Economy of scale)의 원리하에서 기업은 대규모 생산을 하고자 노력하였다. 즉 기업과 고객과의 관계는 생산자가 제공하는 상품이나 서비스를 획일적으로 공급하고 고객은 일방적으로 살 수 밖에 없는 단순한 판매의 관계를 가지고 있었다.

Table.1 고객 관리의 시대적 변천

	단순판매 (1970년대)	고객만족 (1980년대)	데이터베이스마케팅 (1990년대)	CRM (2000년대)
고객	수동적구매자	선택적구매자	개성화, 다양화된 구매자	능동적 파트너
고객과의 관계	전체시장에 일방적 공급(배급)	고객만족도 측정 일방적관계	그룹화된 고객과의 일방적인 관계	개별고객과 쌍방향 의사소통
고객관리	단순 영업위주	영업과 판매위주	IT기술팀 위주	전사적 관리

그러나 80년대 들어 시장경쟁이 치열해 지면서 기업은 고객의 중요성을 인식하고 고객만족에 관한 관심을 두기 시작하였다. 즉 공급이 수요를 초과하고 소비자들의 파워가 증대하면서 기업들은 품질관리(Quality Control)에 대한 관심을 보이다가, 품질 관리 경쟁으로 품질차별화가 어려워지자 고객서비스와 고객 만족에 대한 중요성을 인식한 것으로 보인다.

90년대 말부터 고객이 모든 기업 경쟁력의 원천이라고 인식되어 고객관리가 기업의 성과와 생존에 직접적인 영향을 미치는 시대로 바뀌기 시작하였다. 인터넷과 무선 네트워크 기술의 발달로 더욱 다양해진 채널을 통하여 폭발적으로 늘어난 고객정보를 획득하게 되었고, 또한 이러한 다양한 채널들을 통하여 서비스가 가능해졌다. 특히 가치사슬의 역류화로 과거 공급자에서 소비자의 방향으로 형성되었던 가치사슬이 소비자로부터 기업경영활동이라는 방향으로 전환되기 시작하면서 이제는 기업과 고객이 쌍방향적 동반자로 인식되고, 특히 이러한 고객관계관리가 조직경영활동의 전체를 포괄하여 전사적으로 관리되어야 할 필요성이 생겼다. 다시 말해 CRM은 고객의 요구와 정보기술의 발전으로 인한 기업의 혁신적 노력에서 자연스럽게 탄생되었다고 할 수 있다.

2) CRM의 개념 및 구성요소

CRM의 개념

CRM에 대한 개념은 다양하게 여러 연구 문헌에서 정의 되고 있다. 아래 [표 2]는 여러 연구기관에서 제시하고 있는 CRM에 관한 정의를 정리하여 놓은 것이다.

Table.2 국내/외 연구기관들의 CRM 정의

연구기관	CRM 정의
Gartner	기업이 현재의 고객 및 잠재 고객관련 정보를 정확하게 파악하여 고객 관계 관리를 효과적으로 지원하기 위한 전략
Pricewaterhouse Coopers	조직의 전사적인 참여를 통하여 고객에 대한 정보를 획득하고 이러한 고객지식을 다양한 경로를 통하여 이용함으로써 기업의 수익을 늘리고 운영의 효율성을 도모하는 것
Carson Marketing Group	기업이 모든 구성원과 고객에게 조직에 대한 긍정적 선호도를 형성하여 우수고객 유지와 경영성과를 향상시키는 것
한국 소프트웨어 산업협회	기업이 보유하고 있는 고객데이터를 수집, 통합, 가공, 분석하여 고객 개개인의 특성에 맞게 마케팅 활동을 계획, 수행, 평가, 수정하는 일련의 과정

이러한 여러 정의들을 포괄하여 정리하면 "CRM은 고객 관련 정보를 획득하고 분석하여, 고객 개개인의 특성에 기초한 서비스를 제공함으로써, 고객의 평생가치(Life Time Value)를 높여, 기업의 수익성이 증가하도록 관련 활동을 계획, 지원, 평가하는 관리체계" 라고 할 수 있다.

이러한 CRM에 대한 개념을 중심으로 다양한 관련 속성들을 살펴보면 다음과 같다.

- ① 마케팅 패러다임은 제품판매 중심에서 신규고객의 획득 및 기존 우수고객을 유지하고 이탈을 방지하는 관계마케팅으로 전환이 필요하다
- ② 고객 관련 정보의 획득, 분석, 관리를 위하여 전산시스템만의 변화가 있어서 되

는 것이 아니라 조직 전체의 참여와 프로세스의 혁신이 필요하다

- ③ 20%의 상위 고객이 80%의 수익을 낸다는 파레토의 법칙에 따라 우수 고객을 어떻게 파악하고, 획득하고, 유지시켜 고객의 평생가치를 높일 수 있는가에 대한 분석이 필요하다
- ④ 고객 관련 데이터를 어떻게 획득하고, 축적하며, 분석하고, 궁극적으로 서비스를 할 것인가에 관한 고객 전략 수립과 인프라 구축에 대한 이해가 필요하다

CRM의 성공적 운영을 위하여 어떻게 데이터 마이닝이 활용되어야 하는지를 살펴보기 위하여 CRM의 구성은 어떤 요소로 형성되었는지를 정리하고, 관련 정보인프라는 어떤 것이 있는지 살펴본다.

CRM의 구성요소

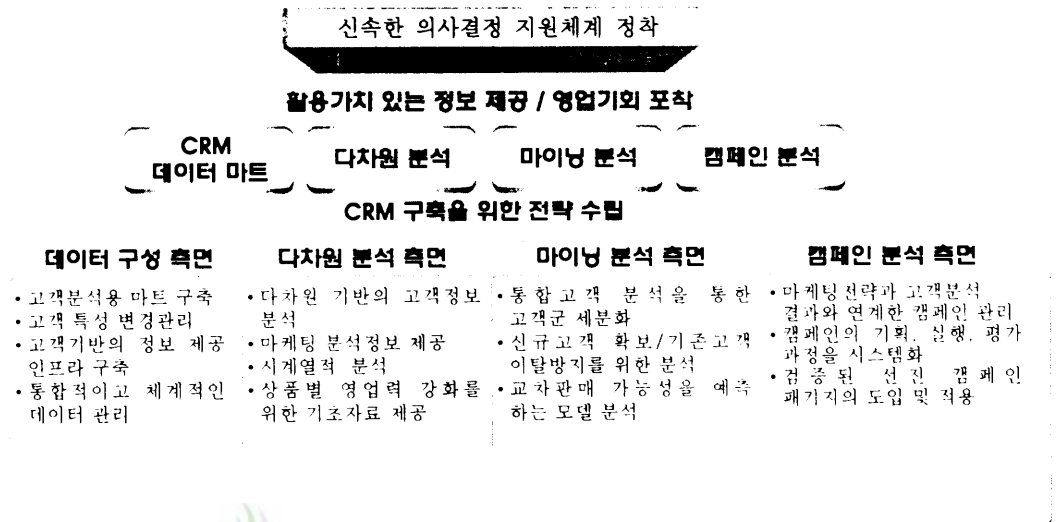
CRM의 정의와 마찬가지로 CRM의 구성요소를 설명하는 것도 매우 다양하다. 그러나 표현을 달리하여도 보는 관점에 의한 차이일 뿐 근본에 대한 이해는 동일하므로 본 논문에서는 가트너가 정리한 CRM 시스템 구성을 기반으로 한 실제 사례중심의 구성요소를 살펴보면


① 분석CRM

먼저 CRM의 구성요소 중 분석 CRM은 고객의 다양한 데이터를 통합하여 저장, 관리 하는 데이터웨어하우스(DW: Data Warehouse)와 분석시 필요한 데이터를 DW로부터 별도로 추출하여 저장한 데이터 마트(Data Mart)를 시작으로 기본 인프라가 구축 된다고 볼 수 있다. 물론 이러한 데이터웨어하우스나 데이터 마트에 있는 데이터를 분석하여 고객을 세분화하고 각 고객 특성별로 개인화된 서비스를 할 수 있게 하는 것을 데이터마이닝이라고 한다.

[Fig 3]에서 보듯이 데이터웨어하우스를 통하여 고객데이터를 통합 저장, 관리하게 되면 다차원분석(OLAP)이라는 분석을 하게 되는데, 다차원적으로 옵션을 주어 현재 고객의 특성을 이해하고자 할 때 가장 간편하면서도 유용한 도구라 할 수 있다. 또한

다양한 이벤트를 중심으로 캠페인 모델을 구축하고 활동을 벌인 후에 그 효과에 대한 분석을 하는 것도 분석 CRM의 한 분야라고 할 수 있다.




제주대학교 중앙도서관
 JEJU UNIVERSITY CENTRAL LIBRARY
 Fig.3 분석 및 운영 CRM

② 운영 CRM

운영 CRM은 기업의 운영과정에서 발생하는 모든 데이터를 관리 운영하는 것으로서 마케팅, 판매, 서비스 활동 등에 필요한 통합고객 정보, 서비스 정보, 지원자료 정보 등을 통합 관리하고, 후방조직(Back Office)의 ERP, SCM, 재무, 회계, 구매, 지불 등의 시스템과 연계하여 고객데이터의 관리 및 운영을 원활하게 하는 것이다.

③ 협업 CRM (채널관리)

협업 CRM은 고객 정보를 획득하거나 고객에게 서비스를 제공하는 채널에 관한 관리를 말하는 것으로 영업점, 콜센터, 인터넷 관리 등을 들 수 있다. 최근에 인터넷 채널의 폭발적 사용증가로 인터넷 이용자를 대상으로 한 CRM을 e-CRM이라 명칭하고

고객의 웹행동(Web Mining)이 등장하기도 하였다. 그러나 향후 유비쿼터스 사회가 도래하면 다양한 채널로부터 대량의 데이터가 획득되고 정리 될 것이므로 e-CRM도 결국 CRM으로 통합되어 사용될 것이다. 여기서 중요한 것은 고객은 필요에 의하여 다양한 채널을 혼용하여 사용하게 되므로 각 채널을 통합하여 고객을 중심으로 단일고객관점(Customer Single View)의 데이터를 정리하고 분석하는 것이 중요하다.

이상과 같이 CRM의 시스템 구성을 분석, 운영, 협업으로 나누어 살펴보았는데 중요한 것은 이것이 각각 따로 활용되는 것이 아니라 항상 유기적으로 서로 통합되어 CRM이 완성되는 것이다. 다시 말해 협업 CRM을 통하여 고객의 접점이 확보되고 운영 CRM을 통하여 기존의 기업 데이터베이스와 통합 정리되어 궁극적으로 모든 고객의 정보는 분석 CRM의 데이터웨어하우스에 저장되는 것이다. 분석 CRM의 OLAP 나 데이터마이닝 분석을 통하여 세분화된 고객을 대상으로 운영 CRM의 솔루션이나 아니면 곧바로 협업 CRM의 다양한 채널을 통하여 세분화된 고객에게 서비스가 제공된다. 고객의 반응 결과는 다시 데이터웨어하우스에 저장되어 좀더 개선된 방법으로 세분화된 고객에게 다시 서비스가 제공되고 그 피드백이 저장된다. 이러한 순화과정을 거쳐 기업의 고객관계관리가 완성되어진다고 할 수 있다.

5. 고객세분화 및 차별화 전략

CRM과 데이터마이닝을 논의 할 때 가장 많이 인용되거나 토의되는 것은 고객에 대한 다양한 데이터를 분석하여 고객을 세분화하여 분류하고, 각 고객별로 개인화된 서비스 전략을 세우는 고객 세분화에 관한 것이다. 고객세분화란 [Fig 4]에서 보듯이 고객을 잠재, 가망, 신규, 우량, 휴먼, 이탈 고객 등으로 분류하여 신규고객을 확보하고, 잠재고객을 활성화하며, 고객의 충성도를 높여 고객을 유지하거나 궁극적으로 교차판매 등의 판매 확대에 이용하는 것이다. 이렇게 고객을 세분화 하여 세분화된 고객별로 차별화된 서비스를 제공하게 되면 신규고객의 증가, 우량고객의 증가, 이탈고객의 감소 등의 효과를 통하여 매출을 증대시키고, 광고비, 채널이용비용, 인건비등의 절감을 통한 비용 절감 효과도 볼 수 있다.

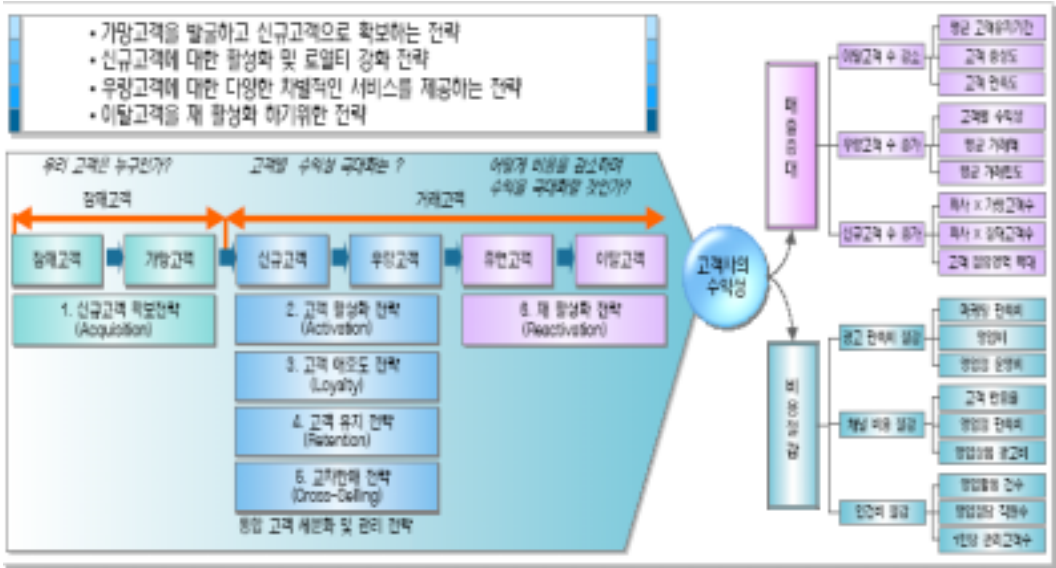


Fig.4 고객세분화의 개념

고객 세분화는 다양한 방법이 있으며 그중 가장 많이 이용되는 RFMT 방법을 중심으로 간략하게 정리한 것이다. RFMT 방법은 Recency(최근 거래 기간), Frequency(거래빈도), Monetary(거래금액), Type(거래상품유형)으로 나누어 각 고객들을 주어진 문제의 상황에 맞게끔 먼저 5등급이나 10등급 등으로 분류한다. 그리고 각 RFMT의 중요성을 점검하여 가중치를 산출하여 최종적으로 각 개인 개인별로 RFMT 지수를 만들어 최종 등급을 정하게 된다. 여기에서 CRM에서 많이 강조되고 있는 고객생애가치(LTV: Life Time Value)는 RFMT에 의한 수익성, 잔여거래기간의 예측, 이탈가능성을 종합적으로 고려하여 한 고객이 기업의 수익성에 미치는 영향을 계산하는 것으로 궁극적으로 기업의 수익성 향상에 진정으로 도움을 줄 수 있는 우량 고객을 파악하는 것이다. 이러한 데이터마이닝 기법을 이용하여 다양한 고객별로 다양한 마케팅 전략이 CRM의 목표 및 정의 이벤트 등에 따라 분석이 되어 정리된다. 예를 들어 우수고객 확인 및 잠재고객 활성화, 신용카드 개설 등 판매확대 전략, 우수고객유지를 위한 특별 서비스 제공전략 등이 설정되고 그러한 서비스를 받을 특정 타겟고객이 데이터마이닝을 통하여 확인된다. 이렇게 확인된 특정고객에 대한 다양한 캠페인 전략이 다양한 채널(예를 들어 영업점, 웹, 이메일, 콜센터, 우편)등을 통하여 고

고객에게 서비스되고 그 결과는 다시 데이터웨어하우스에 저장된다. 이러한 순화과정을 통하여 끊임없이 개선될 수 있는 CRM 시스템을 갖추게 되는 것이다. 어떤 산업분야 이든지에 관계없이 CRM에서 데이터마이닝이 가장 많이 활용되는 고객 세분화의 개념을 살펴보면 [Fig 5]과 같다. 일단은 다양한 캠페인을 통하여 최대한 양질의 신규 고객을 확보한 뒤, 고객의 충성도(Loyalty)를 높이기 위하여 확보된 신규고객을 어떻게 하면 기업에 수익성에 도움을 줄 수 있는 우수 고객 -> 최우수 고객 -> VIP 고객으로 만들 것인가가 CRM이고 이러한 분석과 전략제시에 데이터마이닝이 활용되고 있는 것이다. 결국 CRM을 성공적으로 완수하기 위하여서는 끊임없는 고객데이터의 확보와 분석이 있어야 하며 데이터마이닝이 이러한 고객데이터분석에 반드시 필요한 수단과 방법 이라고 할 수 있다.

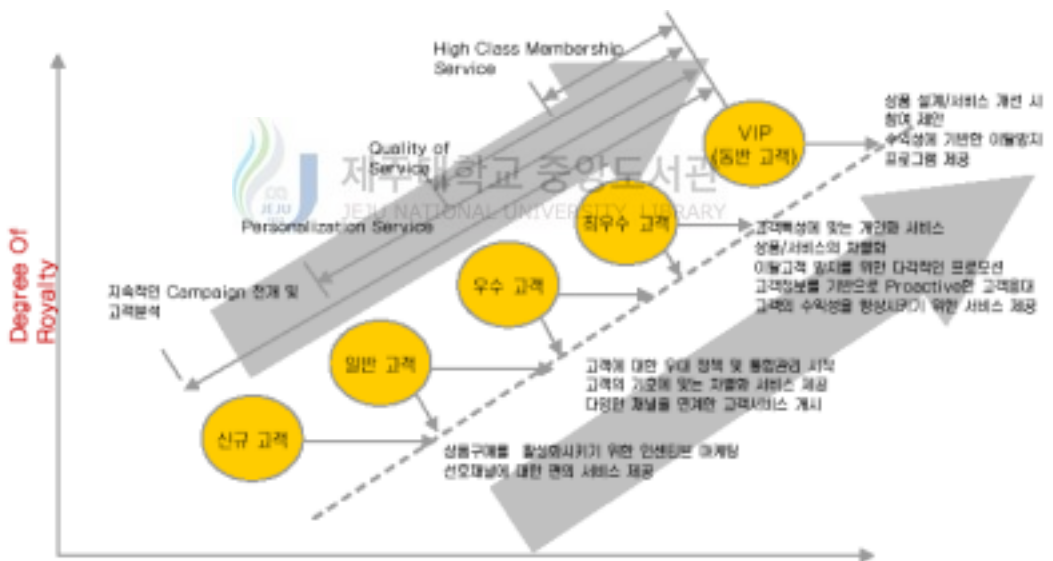


Fig.5 고객세분화에 따른 마케팅 전략 수립

6. 마케팅 CRM에서 데이터마이닝 기법

데이터마이닝 기법의 분류

[Fig 6]에서 보듯이 ECMiner사에서 제공하는 데이터마이닝 기법의 분류는 크게 계량적 데이터마이닝(Quantative Data Mining)과 비주얼 데이터마이닝(Visual Data Mining)의 두가지 분류로 시작된다. 일반적으로 많은 데이터마이닝소프트회사가 비주얼 데이터마이닝을 취급하지 않아 데이터마이닝 기법에 포함시키지 않았지만 ECMiner사는 관련 소프트웨어를 취급하여 이런 분류방법을 제공한 것으로 보인다. 계량적 데이터마이닝은 다시 크게 종속변수(보험사기의 예를 들면 사기의 유무, 기업 신용평가의 경우 도산/건전 등)의 유무에 따라 Supervised 와 Unsupervised 두 가지 방법으로 나누어진다. 즉 예측하고자하는 결과변수(Target)가 있는 데이터 구조를 Supervised 데이터라 하는 반면, 이런 예측의 대상이 되는 결과변수가 없으면 이를 Unsupervised데이터라 한다. Unsupervised의 경우는 자기조직화 지도(Self Organizing Map), 군집분석(Clustering), 장바구니 분석(Market Basket Analysis) 등이 포함되고, Supervised의 경우는 인공신경망(Neural Network), LOGIT이나 판별분석(Discriminant Analysis), 의사결정나무(Decision Tree), 사례기반분석(Case Based Reasoning)등이 포함된다.

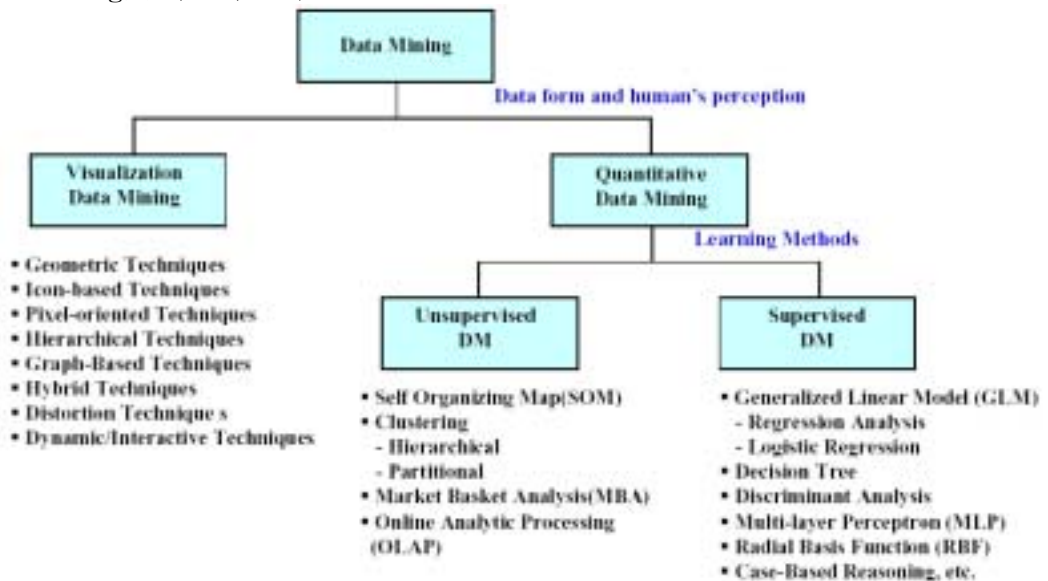


Fig.6 ECMiner사의 데이터마이닝 기법의 분류

비주얼 데이터마이닝기법은 데이터의 관계를 도표화시켜 데이터와 관련된 패턴이나 추세를 한눈에 파악하여, 전문가의 의사결정을 하는데 도움을 주는 방법으로, 과거부터 자료를 요약하여 그래프로 나타내는 방법을 지칭하였으나, 최근에는 이러한 도표화기술을 발전시켜 대용량 데이터의 상호연관성을 그림으로 보여주고, 그림의 관계를 수치로 분석가능하게 함으로써, 사기적발 등 수학적 계량화된 모델을 적용하기 힘든 데이터마이닝 분야에서 각광을 받기 시작하였다. 비주얼 데이터마이닝을 이용하여 사기를 적발하는 비주얼 데이터마이닝은, 그림으로 표시된 데이터간의 상호 연관관계를 보다가, 만일 의심스러운 부분이 있으면 그 부분을 클릭하여 관련 데이터의 수치를 이용한 실시간 분석을 하도록 설계되어있다.

의사결정나무

의사결정나무(Decision Tree)는 분류 및 예측에 있어서 자주 쓰이는 기법으로, DM(direct mail)의 응답여부 등에 영향을 미치는 변수들과 변수들의 상호작용을 누구나 쉽게 이해할 수 있도록, 굳이 통계학적인 용어를 쓰지 않더라도 설명이 가능하다는 점이 큰 장점이며 특징이다. 또한 이러한 특징과 장점 때문에 데이터마이닝을 언급할 때마다 빠지지 않고 소개되고 있으며, 실제 기업에서도 가장 많이 활용되는 데이터마이닝 분석기법 중의 하나이다.

어떤 경우에 있어서는, 예측력을 높이는 것이 분석의 최종적인 목표일 때가 있다. 이유를 알아내는 것은 부차적이면서 DM에 응답할 가능성이 높은 사람, 프로모션 상품을 구매할 가능성이 높은 사람들을 우선적으로 예측하면 되는 경우가 있는 반면, 대출거절, 카드사용승인 거절과 같이 고객에게 그 이유를 반드시 설명해야 하는 경우가 있다. 후자의 경우, 의사결정나무를 사용하게 된다면 ‘당신에게 대출을 거절하는 이유는 수입이 얼마 이하이며 담보가치가 얼마이하이기 때문입니다’ 라는 식의 근거 이유를 제시 할 수 있게 되나, 신경망모형(Neural Networks)을 사용하게 되면 채무를 상환하지 않을 가능성이 높은 고객을 찾아낼 수 있을지언정, 대출을 거절하는 이유를 설명하지는 못하게 된다. 오히려 전자의 경우처럼, DM에 응답할 가능성이 높은 고객을 추출하여 발송하기만 하면 되는 경우에 신경망모형(Neural Networks)이 더 적합할 것이다.

의사결정나무는 흔히 [Fig 7]와 같은 형태로 그 결과가 출력된다. 더불어 의사결정수를 이해하는데 있어서 필요한 용어들도 같이 그림에 나타내어져 있다. [Fig 7]에 그려진 네모들은 모두 마디(Node)로 불리어지는데, 이는 각각 DB에서 사용되는 용어로는 레코드(Record), 통계에서 사용되는 용어로는 관측치(observation)들의 집합체라 생각하면 된다. 맨 위에 그려지는 마디를 뿌리마디(Root Node)를 어떤 특정한 법칙에 의하여 나누게 되면 바로 밑에 생기는 마디들은 자식마디(Child Node)라 한다. 결국 자식마디는 뿌리마디의 한 부분이 되는 것이다. 이런 식으로 어떤 법칙을 가지고 위의 마디를 나누는 것을 가지치기라 하고, 더 이상 가지가 쳐질 마디가 없으면 이 마디는 종단마디(Leaf Node)가 된다. 다시 말하면 종단 마디(Leaf Node)는 의사결정수의 맨 하단에 생기는 마디가 된다.

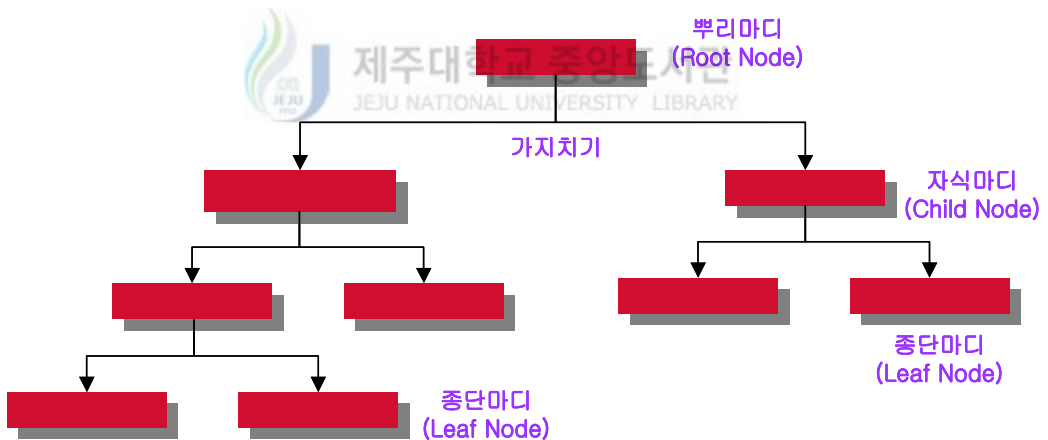


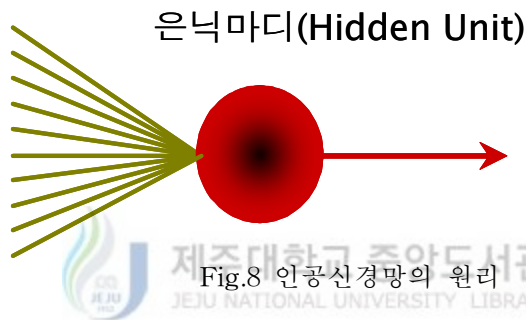
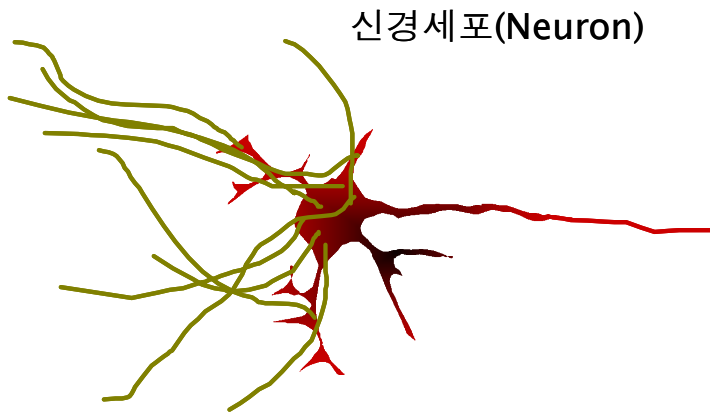
Fig.7 의사결정나무의 지식표현

인공신경망(Neural Networks)

인공신경망모형은 데이터마이닝에 대한 관심이 모아지면서 가장 일반적으로 언급되어지고 또한 다양한 응용분야를 가지고 있는 기법일 것이다. 인공신경망모형은 신경생리학분야에서 두뇌의 활용을 이해하고자하는 하는 목적 하에, 신경의 작업을 설명하려는 시도에서 출발하여 생물학적인 프로세스를 컴퓨터를 이용하여 모형화 하려는 노력에서 비롯된 것으로, 80년대 이후 생물학적 활동의 모형발전과 더불어 컴퓨터 성능의 진보, 신경망이론에 대한 통계학적인 접목으로 인해 빠르게 진보되면서 최근에는 데이터마이닝에 있어서 유용한 기법이 되고 있다.

데이터마이닝에 이용되는 한 기법으로서의 신경망모형은, 인간이 경험으로부터 학습해 가는 두뇌의 신경망 활동을 모방하여 자신이 가진 데이터로부터의 반복적인 학습과정을 거쳐 패턴을 찾아내고 이를 일반화함으로써, 특히 향후를 예측(Prediction)하고자 하는 문제에 있어서 유용하게 이용되는 기법이다. 매우 복잡한 구조를 가진 데이터들 사이의 관계나 패턴을 찾아내는 유연한 비선형모형(Flexible nonlinear Model)의 하나로, 신경생리학과 유사성 때문에 일반적으로 다른 (통계적) 예측모형에 비해 보다 흥미롭게 여겨지고 있다.

이러한 인공신경망모형은 고객의 신용평가, 불량거래의 색출, 의료진단 예측, 우량고객의 선정, 타겟 마케팅의 여러 주제(DM, TM)등을 비롯한 여러 분야에서 적용될 수가 있는데, 주로 Supervised 데이터에 적용되어 결과변수(target)에 대한 예측(Prediction)이나 분류(Classification)를 목적으로 감춰진 패턴을 찾고 이를 일반화 하는데 이용된다. 혹은 Unsupervised 데이터에서 코흐넨 맵(Kohonen maps)을 이용하여 데이터의 클러스터링 작업을 수행하는데 쓰이기도 한다. 인공신경망모형은 상당히 다양한 산업분야의 다양한 문제에 적용될 수 있고, 입력변수와 결과변수의 관계를 그리기가 어려운 복잡한 데이터에 대해서도 좋은 결과를 주는 것으로 알려져 있다. 또한 입력변수와 결과변수의 속성이 연속형이나 이산형인 경우를 모두 다룰 수가 있다. 그러나 신경망모형은 매우 유연하기는 하지만 설명력이 부족하여 종종 블랙박스(Black box)로 불리고 있다.



의사결정수(Decision Trees)나 연관성측정(Associations)은 명쾌하고 쉽게 이해할 수 있는 결과물(일종의 Rule)을 제공하는데 반해, 신경망모형은 인간이 어떠한 현상을 인지하게 되는 것처럼 쉽게 설명되지 않는 내부적인 작업을 수행하고 이를 통해 얻어진 결과물을 제공할 뿐, 어떠한 변수가 중요한지, 어떻게 상호작용이 이루어져 그러한 결과물을 주게 되는지에 대한 설명은 하지 않는다. 따라서 설명력(Comprehensibility)보다는 더욱 정확한 예측을 주는 것이 더 중요하게 고려되는 경우에 이용될 수 있을 것이다.

인공신경망에는 다양한 여러 모형이 있는데, Supervised 데이터에 적용되는 가장 일반적인 모형은 [Fig 9]에서 나타나는 다계층인식인자(MLP: Multi Layer Perceptron)이다.

다계층인식인자 신경망은 입력계층, 출력계층, 그리고 은닉계층(Hidden Layer)으로 이루어져 있다. 입력계층은 결과변수를 설명하는데 이용하고자 하는 입력변수들(예를 들어, 카드의 부정거래를 색출하려는 문제에 있어서 승인시간간격, 승인금액, 누적승

인건수등)이고 출력계층은 예측치를 얻고자 하는 결과변수(부정 또는 정상)이다. 두 개 이상 놓여질 수 있는 은닉계층은 인간의 신경망을 모형화한 몇 개의 은닉마디 (Hidden unit)로 이루어져 있는데, 각 은닉마디는 연결함수(Combination function)를

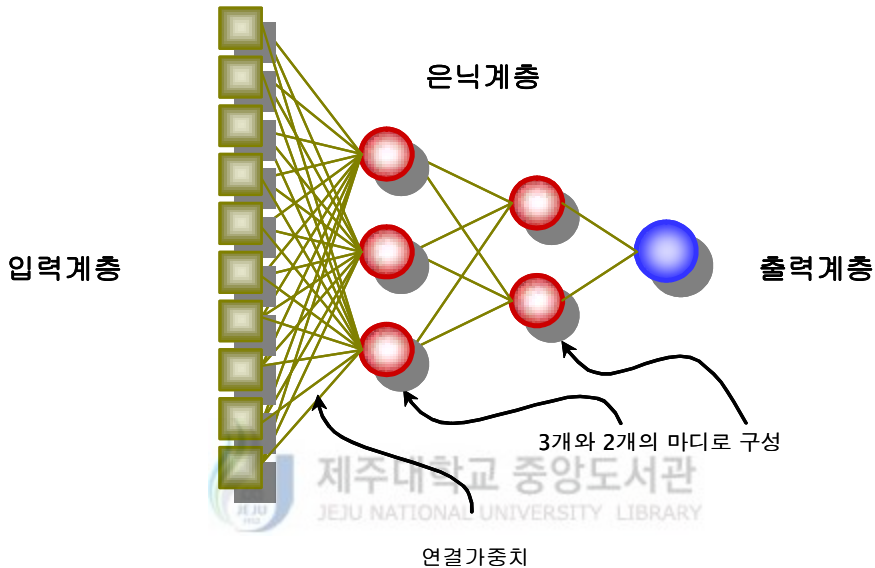


Fig.9 인공신경망의 구조

통해 입력변수들과 연결되어 있고 이러한 각 연결에 사용되는 계수는 연결가중치 (Synaptic Weights)라고 불린다. 각 은닉마디와 은닉계층에는 대개 비선형함수인 활성화함수(Activation Function)를 지정하게 되는데, 입력으로 들어오는 데이터는 이 함수에 의해 변환되어져 다음 마디나 계층으로 전달된다.

데이터를 가지고 신경망을 학습(Training)시키는 것은 여러 가지 입력변수의 정보로부터 은닉계층 내에서의 다소 복잡한 내부작업을 통해, 가장 정확한 결과를 주도록 연결가중치의 값을 찾아가는 것이다. 가장 일반적인 훈련방법은 후진전파(Back Propagation)이다. 데이터의 각 레코드별로 그 값을 읽어 신경망을 훈련하고 이에 의

한 출력 결과와 이미 알고 있는 실제값을 비교하여 그 차이가 충분히 작아져 오차가 일정한 수준에 수렴할 때까지 반복적으로 가중치를 조정해간다. 일련의 훈련과정이 끝나고 연결가중치의 값이 정해지면 이 가중치들을 아직 결과변수의 값을 알지 못하는 새로운 데이터에 적용하여 예측치를 얻게 된다. 신경망모형을 적용하여 좋은 결과를 얻기 위해서는 데이터의 준비와 신경망의 아키텍처 구성이 중요한 요소가 된다.

연관 규칙(Association Rule): 장바구니 분석

한 슈퍼마켓에서 과거 판매 데이터를 조사해보니 이상하게 맥주와 기저귀와의 관계가 있어 이상하다 싶어 조사하여 보니, 구매자가 기저귀를 사러오면서 맥주를 함께 사 간다는 것을 발견하고, 데이터분석을 통하여 과거에 전혀 예상하지 못하였던 맥주와 기저귀간의 상호연관률 찾아냈다는 예는, 데이터마이닝을 소개할 때 가장 흔하게 언급되지는 것이었다. 이와 같이 상품 혹은 서비스간의 관계를 살펴보고 이로부터 유용한 규칙을 찾아내고자 할 때 이용될 수 있는 기법이 연관 규칙(Association Rule)이다. 연관성 규칙은 어떤 특정 문제에 대해 아직은 일어나지 않는 답을 얻고자 하는 예측(Prediction)의 문제나, 고객들을 특정목적에 따라 분류(Segmentation)하는 문제가 아니라, 상품 혹은 서비스(이하 상품)의 거래기록 데이터로부터 상품간의 연관성 정도를 측정하여 연관성이 많은 상품들을 그룹화 하는 군집방법의 일종으로서, 동시에 구매될 가능성이 큰 상품들을 찾아냄으로써 시장바구니분석(Market Basket Analysis)을 다루는 문제들에 적용될 수 있다.[1][2][3][4][13][19]

Ⅲ. 데이터마이닝을 이용한 고객 이탈방지 알고리즘 설계

데이터마이닝(data mining)방법론 중의 하나인 분류 규칙(classification)을 이용하여 기존고객을 세분화한 다음 고객 개개인의 특성에 맞는 마케팅 프로모션을 하게하고 신규고객을 획득할 때는 신규고객의 특성을 미리 예측하여 세분화함으로써 고객의 평생가치(LTV : Life Time Value)를 촉진하여 기업과 고객과의 관계성을 높여서 기업은 안정된 고객층으로부터의 수익을 창출하고 고객들은 해당 기업으로부터 더 많은 혜택을 받게 하는 것이다[5]. 그러므로 본 논문에서 제시하고자 하는 것은 기존 고객에 대한 분류 및 고객속성 파악, 기존 고객 분류에 따른 신규고객의 분류 예측 두 가지를 수행하여 해당 기업의 마케팅 전략수립에서부터 경영 과학적으로 접근할 수 있는 데이터베이스 마케팅을 하려고 하는 것이다.

기존의 데이터베이스마케팅에서는 20 : 80 법칙이라는 방법론을 사용하여 우수고객과 비 우수고객을 나누어 관리하여 왔다. 20 : 80법칙이라고 하는 것은 전체 매출에 있어서 매출이 많은 상위 20%의 고객층이 전체매출의 80%를 차지한다고 하는 것으로 상위20%의 고객들을 중점 관리하여 안정된 매출을 유지하면서 더 좋은 서비스를 제공하는 방법으로 백화점에서 매출 비례 마케팅을 실시할 때 많이 이용되어 왔다 [11].

본 연구에서 제안하려고 하는 것은 신규고객획득 비용이 크다는 점을 감안하여 기존의 20 : 80법칙에 의한 고객세분화 방법에서 탈피하여 나머지 20%매출을 차지하는 하위 80%의 고객들에 대한 적극적인 프로모션을 하고 상위 20%고객들에 대한 고객 세분화도 하여서 적극적인 마케팅을 실시하려고 하는 것이다. 이러한 방법을 사용하게 되면 신규고객획득 비용도 줄일 수 있으며 하위20%의 매출에 대한 세밀한 조사를 통한 잠재 우수고객층을 발굴할 수 있을 것이다. 또한 현재 사용되는 고객 이탈방지 보다 세분화하여 고객이탈방지 예방을 함으로써 고객의 유지율과 신규고객을 새롭게 창출하는 밑거름이 될 것이다. 향상된 알고리즘은 고객의 이탈 징후나 예상되는 변수를 추출하여 그에 맞는 적절한 예방책을 적용하거나 고객의 Needs 및 Wants를 달성할 수 있는 Loyalty 프로그램 개발 하여 보다 능동적인 방지책을 적용할 것이라 본다.

1. 효율적인 고객 이탈방지를 위한 방법

기업 CRM 역량과 현재의 CRM 수준을 파악하기 위해 사용하는 방법론 중 일반적으로 많이 사용되는 것은 PwC의 성숙도 모델 6가지 영역의 CRM역량으로 구분된다.

- ① 고객정보관리: 고객의 정보를 수집하고 활용하는 역량은?
- ② 고객 가치평가: 지속적으로 고객의 가치를 평가하고 활용할 수 있는가?
- ③ 고객 기대 측정: 고객의 기대와 니즈를 이해하고 충족시키는 정도는?
- ④ 고객 정보 활용: 고객을 중심의 정보 전략과 사업계획이 체계적으로 실행되고 있는가?
- ⑤ 고객 접점 통합: 고객 중심의 기업 활동을 통하여 가치를 극대화할 수 있도록 고객 접점이 유기적으로 통합되어 있는가?
- ⑥ 지속적인 고객 만족: 고객과의 모든 상호작용에서 고객 경험을 최적화하고 고객만족을 유도하는가?

무엇보다 고객의 CRM 전략은 크게 고객의 세분화 전략과 고객 차별화 전략으로 구분할 수 있다. CRM전략 Framework를 정착시키기 위한 중요한 부분은 다양한 고객 세분화를 시도해 보고 전략적으로 연관성이 있다고 판단되는 Frame을 Master 세분화로 결정하고, 다른 요소들은 Master 세분화를 설명하기 위한 Sub-세분화 변수로 사용한다.

- ① 인구통계(Demographics), 상품(Product): 가장 기본적인 정보이지만 구체적인 고객 세분화 기준을 제시함. 대체로 DW가 구축되어있는 경우는 이러한 고객정보는 어느 정도 축적되어 있는 것이 보통이며 주기적인 update를 통해 정보의 정확도 제고가 필요
- ② 행동양식(Behavior): 고객의 행동양식 및 이력 관련 정보로 세분화를 함 OLAP성 분석 혹은 고객 Survey 등으로 가능하며, Data Mining으로도 다양한 고객 행동 예측 모델 구축 후 세분화 가능
- ③ 심리적 특성: 고객의 심리적 특성을 이용한 세분화로 고객의 잠재적, 심리적

Needs에 깊숙이 접근할 수 있는 장점은 있으나 현업에 실용적이게 나오지 않는 경우가 발생

- ④ 가치기반: 가치기반 고객 세분화로 고객의 가치 즉 LTV, Profitability, Loyalty, 구매패턴 점수 등을 활용 등이 있다.

그래서 본 논문에서는 기존의 매스마케팅에서 고객중심의 마케팅 CRM활동을 효과적이고 차별적으로 추진하기위해, 유사한 특성을 이해하는 기반을 마련하는데 있다.

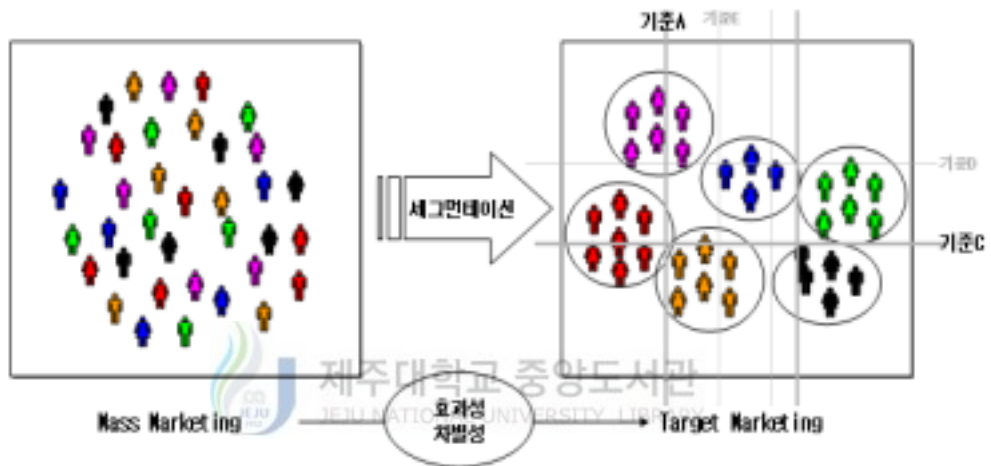


Fig.10 고객 세분화의 목적

전자상거래에서 고객마케팅에서 CRM전략으로 고객세분화 하여 보다 효과적인 고객 이탈방지 알고리즘을 적용하는 것이다.

그러기 위해서 본 연구에서는 기존의 CRM전략을 이용하여 연구대상으로 선정한 신생 인터넷 쇼핑몰의 경우 국내에 벤처 열기를 타고 급속도로 늘어나고 있는 추세이지만 아직까지는 제대로 된 고객관리가 이루어지지 않고 있다. 왜냐하면 신생기업이기 때문에 아직까지 구축된 자료가 적거나 미비한 수준인 경우가 많고, 또한 고객 세분화 마케팅에 대한 인식의 부족으로 준비가 되지 않는 기업이 많기 때문이다. 그래서 현재 널리 사용되는 전자상거래에 접목하여 ASP로 제작한 인터넷 쇼핑몰을 데이터마이닝의 군집분석을 활용한 고객수입, 고객 선호상품 등을 예측하여 고객이 로그인시 웹 서버에 기록된 로그 파일을 이용하여, 상품구매력을 높이기 위한 구매욕구

순서를 동적으로 구성하며, 보다 나은 경쟁력강화를 위해 고객이탈방지 알고리즘을 연구하는 것이 주목적이다.

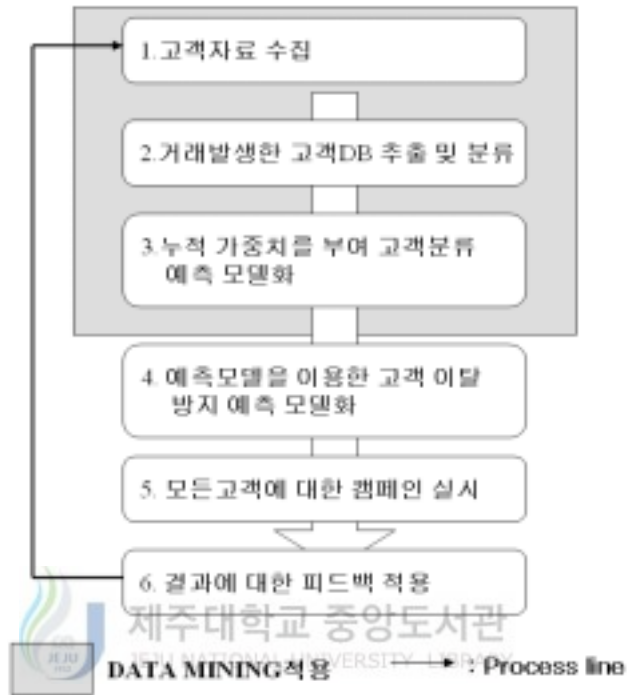


Fig.11 본 연구 흐름도

1-1 기존고객 이탈방지 방법의 문제점

기존고객 이탈방지 방법은 가공되어 있지 않는 데이터나 단순 통계나 정렬 방식으로 가공처리가 되어 있지 않은 상태에서 처리하므로 정확한 분석이 어려웠으며 정보 품질 또한 기대하기 어려웠다. 그래서 다음과 같은 문제점을 안고 있었다.

① 정보화시대의 도래

18세기 중엽부터 시작된 산업화의 물결은 지난 200여 년 동안 인류 문명을 지배하였고, 이는 다량 생산과 소비를 가능케 하여 이전에는 기대할 수 없는 물질적인 풍요

를 가져왔다. 이 시기의 기업은 정부의 보호와 규제 속에서 독과점의 형태를 띠고 있었으며, 전통적인 자원인 인력, 자금, 그리고 물자의 효율적 운용을 통해 다량의 제품이나 서비스를 받으면 성공적이라 할 수 있었다.

이러한 환경에서 기업이 생존하고 발전하기 위해서는 지속적으로 소비자의 동향을 파악하고, 자사는 물론 경쟁사의 경영전략을 효과적으로 분석하고 대처할 수 있는 능력을 절실하게 요구되는데, 이를 가능케 하는 것이 바로 정보다. 따라서 정보화 시대의 기업경영에서 정보란 전통적인 자원들을 효과적으로 운영, 관리하며 새로운 제품이나 서비스를 창출하는 역할을 담당하는 또 다른 자원으로서, 이것을 제대로 수집하고 활용할 수 있는 기업만이 경쟁에서 살아남고 우위를 확보하게 되는 것이다.

② 정보기술의 가속적인 발전

70년대만 하더라도 정보기술은 일부연구소나 학계 등 특정직에 종사하는 사람들의 전유물로 여겨졌다. 그러나 80년대 중반 이후 개인용 컴퓨터가 시장을 주도하기 시작하면서 전문지식이 별로 없는 일반 사용자들도 컴퓨터를 쉽게 사용할 수 있는 환경이 조성되었다. 그러나 오늘날 학습능력을 지닌 프로그램의 등장으로 컴퓨터의 가공할만한 정보처리 능력뿐만 아니라 정보통신 기술의 정보 유통기능과 연계하여 대량의 데이터를 실시간에 분석하고 이를 통해 경영전략을 수립할 수 있는 길을 열었다.

③ 전통적인 전문가 시스템의 한계

전문가시스템은 인공지능의 한 분야로서 특정분야에서 전문가의 축적된 지식과 경험을 시스템화하여, 필요할 때 사용하도록 하는 소프트웨어이다. 전문가시스템의 구성요소 중 핵심이 되는 지식베이스(knowledge-base)에는 전문가의 지식을 저장하는데, 과거에는 주로 해당분야 전문가와의 인터뷰나 전문서적 등을 토대로 구축하였으나 기업환경의 급속한 변화와 시스템을 이용해 해결하고자 하는 문제영역이 점차 확대되고 복잡해짐에 따라 해당 분야에 통달한 전문가를 찾기가 어려워지면서 지식획득에 장기간에 걸쳐 많은 비용을 투자해야 하는 경제적인 문제에 봉착하였다.

이것을 소위 “지식획득의 병목현상(knowledge acquisition bottle-neck)”이라 표현되는데, 이로 인해 데이터를 분석하여 자동적으로 지식을 추출해 내는 방안에 대한 요구가 급속히 증가하고 있는 것이다.

④ 데이터의 홍수, 정보의 빈곤

가공되지 않은 데이터 자체는 아무런 의미가 없다. 데이터의 양이 오늘날과 같이 방대하지 않았던 과거에는 소수의 전문가들이 통계기법이나 질의를 통해 데이터를 분석하고 요약된 결과를 보고서 형식으로 제공하곤 했다. 이러한 접근법은 데이터의 양이 증가하고 차원의 수가 증가할수록 효율성이 떨어지고, 분석을 통해 얻을 수 있는 정보의 품질도 기대하기 어렵다. 수백 기가 또는 테라바이트 상당의 데이터를 제대로 이해하고 분석할 수 있는 사람은 아무도 없다. 따라서 데이터의 홍수와 정보의 빈곤이라는 환경에 처한 지금, 기업들은 정보기술을 이용하여 데이터를 여과하고, 분석하며, 결과를 해석하는 자동화 된 데이터 분석 방안에 높은 관심을 갖게 되었다.

이렇듯 4가지 사항처럼 지금 쇼핑몰에서 사용되고 있는 고객 이탈방지 기법은 데이터를 단순 비교하여 고객의 정확한 요구사항을 파악하지 못하고 통계처리 된 결과값만 이용하여 이벤트나 각종 행사 혜택을 고객에게 적용한 마케팅을 하고 있는 것이라 볼 수 있습니다.



인구통계 기반의 세분화는 성별, 연령, 거주지역 등의 고객정보를 이용하여 고객을 세분화하는 방법으로 고객의 일반적인 특성을 파악할 수 있으나 고객이 상품/서비스 등에 대한 성향 등을 파악하기가 어려움이 존재한다. 그리고 대부분의 업체에서 인구통계정보가 정확하지 않거나 부족한 경우가 존재하고 있다.

고객 Needs/Wants기반의 세분화는 고객의 Needs/Wants를 직접 알 수 있어 전략 수립에 용이하지만 대부분의 경우 Survey를 통해 획득되므로 전 고객에 대해 확대 적용하기 어려움이 존재하고 있다. 최근 이러한 문제점을 해결하기 위해 프로젝트를 진행한 사례를 다루고자 한다.

첫 번째 사례 : 설문조사를 통한 쇼핑스타일 유형 세분화

본 고객 니즈 기반의 세분화는 설문조사를 통해 백화점 고객의 쇼핑 스타일과 Life-Stage 유형을 확인하고, 모집단의 니즈 기반의 세분화를 위해 판별 규칙을 개발하였다. 고객심리확인 은 백화점 내부 전문가 인터뷰를 통해 현업에서 근무하는 직원

들이 이미 알고 있는 고객의 심리 특성을 확인하였다. 다음으로 조사 대상 중 연령대, 직업 유무, 결혼 유무 등의 구분의 구분에 따라 각 군별로 3~5명의 고객을 초청하여 정성조사(F.G.I)를 실시하여 백화점에서 제품구매에 영향을 주는 요인을 파악하였고 이를 기반으로 설문지를 개발하였다. 정량조사는 300명 대상으로 개별면접조사를 통해 쇼핑스타일 및 Life-Stage 유형을 파악하였다.

정량조사를 위한 [표2]과 같이 연령별, 고객등급별, 구매상품별, 거주지역별 층화추출을 통해 모집단을 대표할 수 있는 정보를 획득하고자 하였고, 정확한 설문 응답을 확보하기 위해 일대일 개별면접조사를 실시하였다.

Table.3 정량조사를 위한 조사 설계

구분	내용
모집단	00년 본점 숙녀의류 구매고객
조사기간	00년 0월 00 일 - 0월 00일(3주)
표본구성	- 25~29세, 30~34세, 35세~39세 = 1:1:1 - 00년 본점 현재가치 고객등급 A/B/C등급 =1:1:1 - 숙녀 5개 상품군 할당 - 지역: 강남/강북/일산/분당 DB비중에 따른 할당 추출
표본크기	300명
표본추출 방법	연령별 / 구매숙녀 브랜드별 / 고객등급별 균등 분포를 감안한 유의할당 추출 (Purposive Quota Sample)
조사방법	구조화된 질문지(Structured Questionnaire)를 이용한 일대일 개별면접조사

설문조사를 통해 308명의 데이터를 수집하였으나, 응답의 신뢰성이 낮은 8명을 제외한 300명을 데이터를 분석에 사용하였다. 설문 문항은 총 27문항으로 구성되었으며 5점 척도로 측정되었다. 설문조사 고객의 쇼핑스타일을 정의하기 위해 사용된 기법은 주요인분석 -> 군집분석을 적용하였고, 분석에 사용한 툴은 spss를 이용하였다.

주요인분석은 spss의 메뉴에서 “analysis -> Data reduction -> Factor”를 클릭하여 프로세로 분석을 수행하게 된다.

이렇게 하여 얻어진 값을 바탕으로 군집분석을 하면 다음과 같이 결과를 얻을 수 있다.

요인	군집(Cluster)				
	1	2	3	4	5
충동 구매	-0.406	0.076	-0.229	0.563	-0.318
가격 충성	-0.508	-0.589	0.880	0.341	-1.064
조기 수용	-0.306	10102	-0.061	-0.423	-0.906
유행 추구	-0.339	-0.008	0.615	-1.025	0.402
품질 지향	-1.911	0.242	0.019	0.197	0.629

↓

저관여형

↓

조기수용형

↓

실속구매형

↓

충동구매형

↓

신중구매형

제주대학교 중앙도서관
 JEJU UNIVERSITY LIBRARY

Fig.12 군집분석 결과

두 번째 사례 : 모 은행 거래를 통한 고객이탈방지 유형 세분화

고객 이탈모델은 우선 고객 이탈에 대한 정의에서 시작한다. 고객 이탈에 대한 정의는 산업에 따라 차이가 존재하며, 마케팅 활용 관점에서 여러 기준을 도입할 수 있다. 업체에서 주로 사용하는 이탈고객 정의를 살펴보면 첫째, 학습지 구독 중지, 보험 해지, 카드 해지 등 고객이 명확하게 이탈 의사를 표현한 경우 이탈로 정의한다. 둘째, 유통업체의 경우 최종 구매 후 경과 기간에 따라 이탈 고객을 정의한다. 예를 들어 최근 1년 미거래시 완전 이탈고객으로 정의할 수 있다. 셋째 다음과 같이 고객등급이 이전 기간에 비해 하락하는 경우를 이탈 고객으로 정의한다.

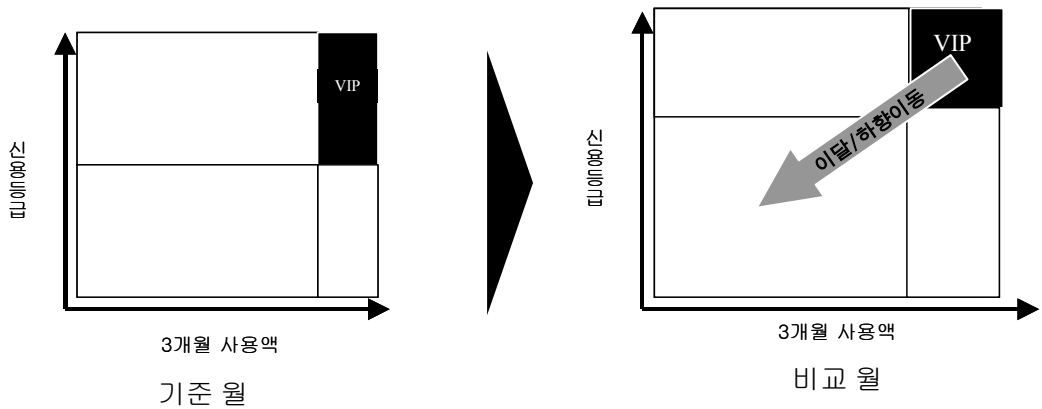


Fig.13 신용카드업체의 고객 이탈 정의 사례

이탈고객 분석 순서

- ① 데이터 설정
- ② 신경망 분석과 의사결정나무분석에 대한 모형화
- ③ 만들어진 두 가지 모형을 이익도표 등을 이용 더 나은 모형으로 만든다.
- ④ 위에서 선택된 집단의 고객 정향을 파악
- ⑤ 결론 도출

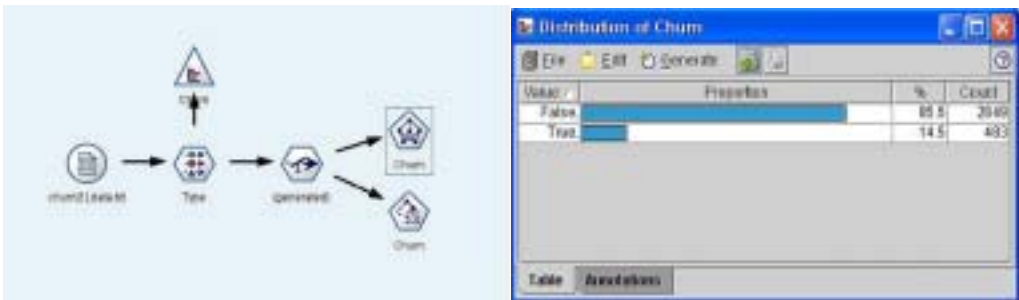
1) 데이터설정

:데이터 분석을 위해 사용된 데이터 : 5,000명(20개 변수)

* 검증을 위한 Test Data : 1,667Case

Training Data : 3,333Case

2) 신경망 분석과 의사결정나무분석에 대한 모형화



클레멘타인 모형구축 분석 흐름도

- ① 데이터 입력을 위한 노드
- ② Type 노드로서 변수유형과 방향 지정

- ③ Distribution 노드 : 범주별 구성 비율 확인
- ④ Balancing 노드 : 균형화 작업
- ⑤ 신경망 노드와 C5.0알고리즘 노드를 통해 분석 수행

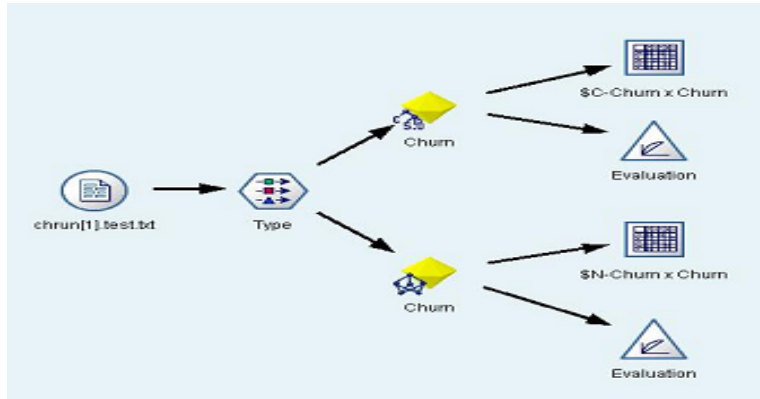


Fig.14 분석 흐름도에 따라 수행결과 구성된 흐름도

3) 만들어진 두 가지 모형을 이익도표 등을 이용 더 나은 모형으로 만든다.

		False	True	Total
False	Count	1322	46	1368
	Total %	79.352	2.761	82.113
True	Count	120	178	298
	Total %	7.203	10.684	17.887
Total	Count	1442	224	1666
	Total %	86.655	13.445	100

Fig.15 구축된 두가지 모형을 이익도표 이용 향상된 모형으로 구축

- ① 의사결정나무 분석에 의한 오류분포 결과치
 - * 전체 Test Data 1,666개 관찰치 중 "True" 또는 "False"로 분류한 경우 1,500개(90.14% 정상 분류)
- ② 신경망 분석에 의한 오류분포 결과치
 - * 전체 Test Data 1,666개 관찰치 중 "True" 또는 "False"로 분류한 경우 1,502개(90.16% 정상 분류)

4) 위에서 선택된 집단의 고객 성향을 파악

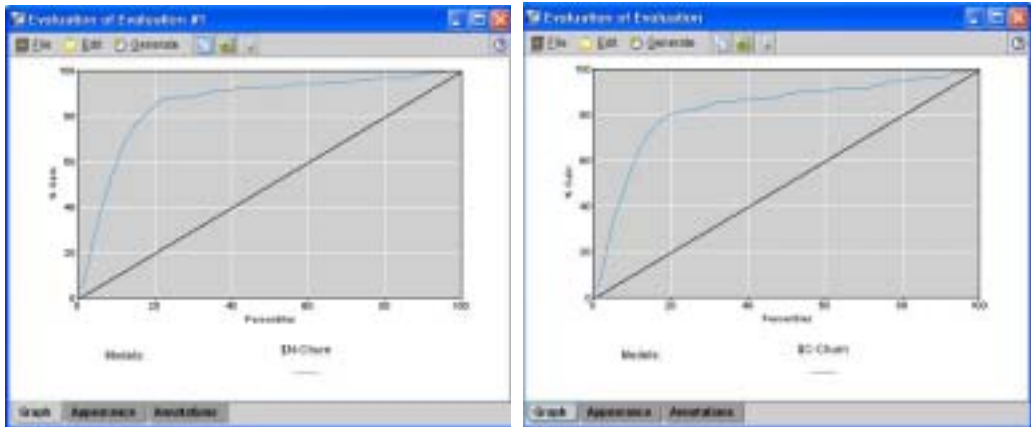


Fig.16 신경망 분석에 의한 이익도표

두 모형 분석결과 결과의 상이성이 뚜렷이 나타나지 않음 그러나 이익도표에서 신경망분석에서 약간 유효성이 높게 나타남



5) 결론 도출

이 분석을 통하여 데이터마이닝에 있어서 여러 모형을 적용하여 분석하는 경우 분석결과에 대한 적용은 각 분야의 이용목적에 부합되게 모형을 결정하는 것이 바람직하다

[신용카드사의 고객이탈방지를 위한 방법]

- ① 적립 포인트 리워드 (3만점 이상이면 현금처럼 사용)
- ② 제휴 할인 서비스 (주유소, 패밀리 레스토랑, 대형 할인매장, 의류점 등)
- ③ 사은품 행사 (백화점, 면세점, 할인매장에서 00카드로 00만원이상 구매시 사은품 증정)
- ④ 레저, 스포츠 입장 / 사용권 할인
- ⑤ 여행관련(호텔, 콘도, 비행기) 제휴할인 및 적립
- ⑥ 현금서비스 수수료 일시 인하
- ⑦ 통신이용료 카드 결제시 00% 할인

1-3 이탈고객 산출 및 세분화

고객들로부터 생성된 구매패턴과 접속패턴의 기반으로 테이블에 상품별로 가중치를 부여하여, 신규고객과 기존고객의 가중치변화를 고려하여 적절하게 화면을 구성하고, 가장 가중치가 높은 데이터를 근거로 우선순위로 출력하며, 기존고객의 패턴을 기반으로 고객이탈방지를 모델화 하여,신규상품 이나 그와 관련된 부가서비스 정보

를, 모바일이나 그 외 인터넷 서비스방법으로 고객에게 제공함으로써 재 구매요구를 증가 시킨다. 그러기 위해서는 첫 번째로 고객의 구매패턴과 접속패턴을 파악 하는게 가장 급선무인 것 같다.

전자상거래에서 구매이력을 통한 구매패턴을 데이터마이닝 기법을 이용하여 적용하여 고객 세분화를 하여 이탈고객을 산출한다.



Fig.17 고객세분화의 흐름도
JEJU NATIONAL UNIVERSITY LIBRARY

보다 세부적으로 전자상거래에서 고객으로 세분화 하려면 다음과 같이 입력한 데이터에 추가적으로 필드를 더 삽입하여 데이터를 분석해야 할 것이다.

Table.4 RFM Table

파일명	필드	유형	설명
CRO	고객번호	문자형	고객식별키
	구매여부	날짜형	종속변수, 구매: 1
	연령	숫자형	
	가입경과일	숫자형	
	최근성(전체)	숫자형	전체 상품 기준 구매경과일
	최근성(상품별)	숫자형	각 상품군(10개) 기준 구매경과일
	구매빈도(전체)	숫자형	전체상품 기준 구매 빈도
	구매빈도(상품별/지역별)	숫자형	각 상품군(10개) 기준 구매빈도
	구매금액(전체)	숫자형	전체 상품 기준 구매 금액
	구매금액(상품별)	숫자형	각 상품군(10개) 기준 구매금액
	가중치	숫자형	고객에 대한 가중치

각 데이터를 쇼핑몰을 구축 시 필드를 추가하여 만든다. 그리고 다음과 같이 분석을 RFM 이용하여 구매 예측 모델을 삼는다.

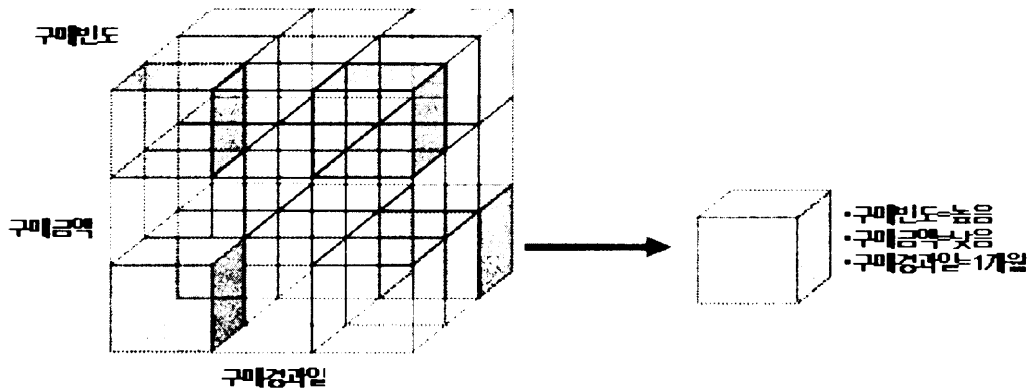


Fig.18 RFM 구매

- ① 고객의 가치를 평가하는 방법 중의 한 가지
- ② 고객을 서열화하여 우수고객에게 차별화된 서비스 제공
- ③ R(Recency), F(Frequency), M(Monetary) 등 3요소 적용

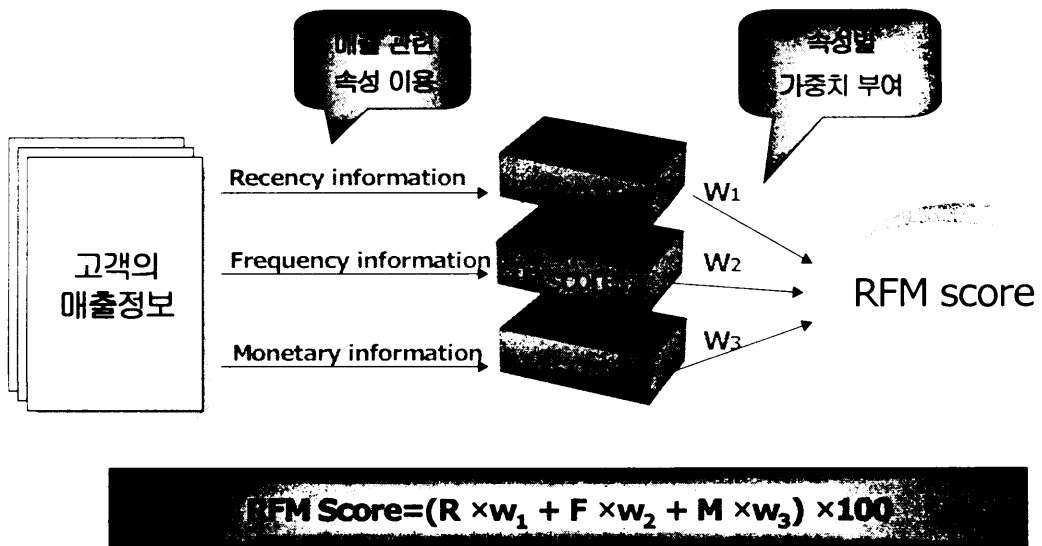


Fig.19 RFM Score

Table.5 RFM 정의 및 세분화 기준



RFM Score 점수가 높은 상품에 대한 세분화를 한다.

1-4 고객 패턴분석 및 이탈 원인분석

- ① 고객의 연령대별로 구분하여 RFM를 적용하여 CRO파일에 저장한다.
- ② 상품별로 고객 구매빈도를 분석하여 CRO파일에 저장한다.
- ③ 지역별로 고객 구매빈도를 분석하여 CRO파일에 저장한다.
- ④ 저장되어 있는 파일을 고객, 상품, 지역별로 분석하여 가중치를 누적시킨다
- ⑤ 누적된 가중치중 구매빈도가 급격히 떨어지는 고객에 대한 정보를 세분화시킨다.
- ⑥ 세분화된 정보를 연령별, 상품별, 지역별로 구분한다.

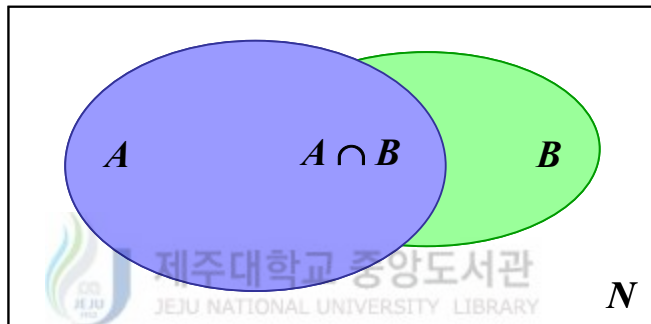


Fig.20 고객패턴 및 이탈분석

고객패턴을 쇼핑몰에서 구매한 경력의 DB를 바탕으로 연령, 결혼유무, 지업유무, 지역별로 구분하고 자주 방문하여 여러분 구매하는 고객의 쇼핑스타일을 경력DB와 연계하여 구축한다. 크게 지지도, 신뢰도, 향상도를 바탕으로 쇼핑스타일을 구축한다
 고객패턴을 쇼핑몰에서 구매한 경력의 DB를 바탕으로 연령, 결혼유무, 지업유무, 지역별로 구분하고 자주 방문하여 여러분 구매하는 고객의 쇼핑스타일을 경력DB와 연계하여 구축한다. 크게 지지도, 신뢰도, 향상도를 바탕으로 쇼핑스타일을 구축한다

(1) 지지도

전체 거래 중 품목 A와 B를 포함하는 거래는 어느 정도인가?

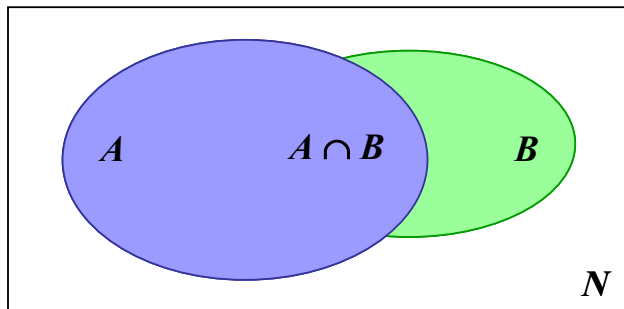


$$\text{Support}(A \Rightarrow B) = \Pr(A \cap B)$$

$$= \frac{\text{항목 } A \text{와 항목 } B \text{가 동시에 포함된 거래수}}{\text{전체 거래수 } (N)}$$

(2) 신뢰도 (Confidence)

품목 A를 포함하는 거래 중에서 품목 B가 포함된 거래는 어느 정도인가?



$$\text{Confidence } (A \Rightarrow B) = P(B|A) = \frac{\text{Pr}(A \cap B)}{\text{Pr}(A)}$$

$$= \frac{\text{항목 } A \text{ 와 항목 } B \text{ 가 동시에 포함된 거래수}}{\text{항목 } A \text{ 가 포함된 거래수}}$$

(3) 향상도 (Lift / Improvement)

임의로 B가 구매되는 경우에 비해 A와의 관계가 고려되어 구매되는 경우의 비율은?

$$\text{Lift } (A \Rightarrow B) = P(B|A) / P(B) = \frac{\text{Pr}(A \cap B)}{\text{Pr}(A) \cdot \text{Pr}(B)}$$

$$= \frac{\text{항목 } A \text{ 와 항목 } B \text{ 가 동시에 포함된 거래수}}{\text{항목 } A \text{ 가 포함된 거래수} \cdot \text{항목 } B \text{ 가 포함된 거래수}}$$

Lift	의미	예
1	두 품목이 서로 독립적인 관계	과자와 후추
>1	두 품목이 서로 양의 상관 관계	빵과 버터
<1	두 품목이 서로 음의 상관 관계	지사제, 변비약

이러한 비율을 통한 고객패턴을 DB에 누적 시킨다 또한 구매고객 중 이탈 고객을 찾은 후 고객스타일 즉 구매 내용과 지역별, 연령별, 상품별로 DB화 시킨다.

1-5 고객세분화를 통한 이탈방지 마케팅 전략

- ① 세분화된 DB를 바탕으로 고객 분석된 현황을 새롭게 정렬한다
- ② 정렬된DB를 바탕으로 고객의 라이프스타일에 추가내용을 첨가한다
- ③ 고객의 쇼핑물에 접속 시 고객의 라이프스타일에 맞게 쇼핑물 진열을 바꾼다
- ④ 바뀐 진열된 쇼핑물에 대한 접속통계를 만든다.
- ⑤ 접속통계를 통한 이탈방지의 효과를 비교분석 한다.
- ⑥ 새롭게 접속된 쇼핑물에서 재 구매를 유도한다..

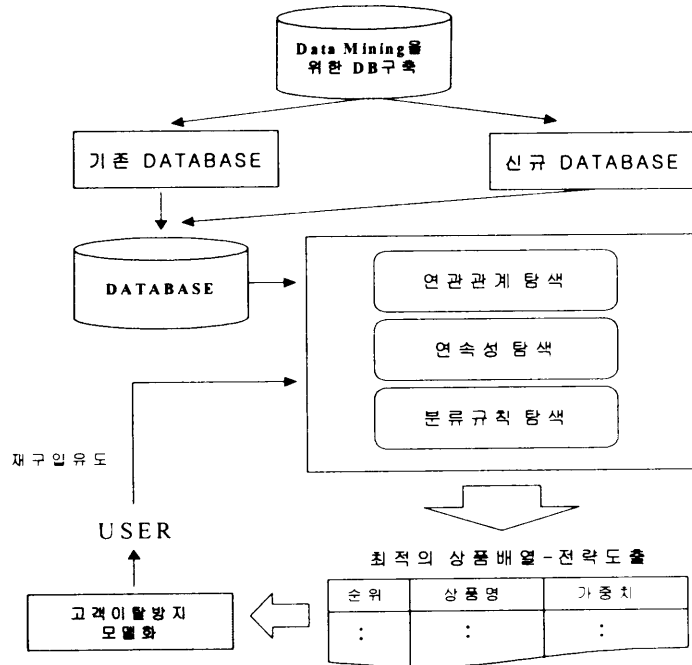
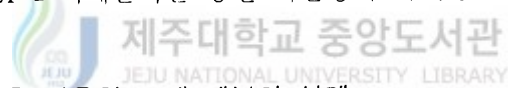


Fig.21 고객세분화를 통한 이탈방지 마케팅 전략 흐름도



2. 향상된 알고리즘을 이용한 고객 세분화 설계

- ① 기존 고객구매 패턴과 이력 관계를 DB화 시킨 후 연관규칙과, 연속규칙, 분류 규칙을 적용한다.

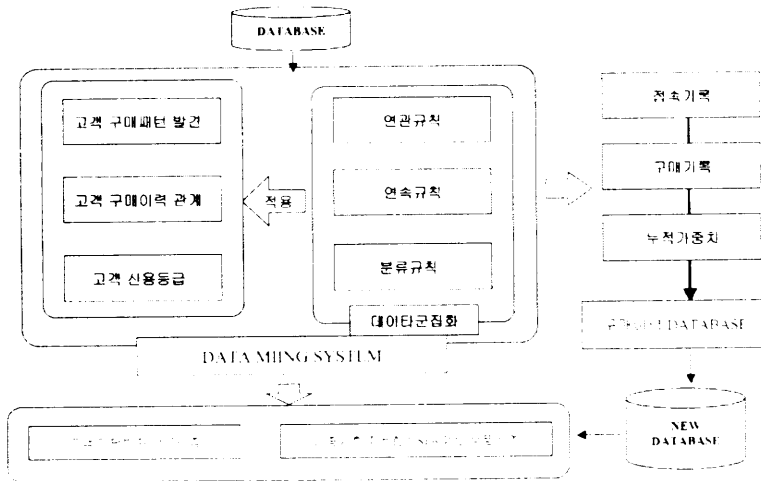


Fig.22 구매패턴과 구매이력의 흐름도

② 실제 예로 사용했던 고객 이탈분석 프로세스를 적용한다.

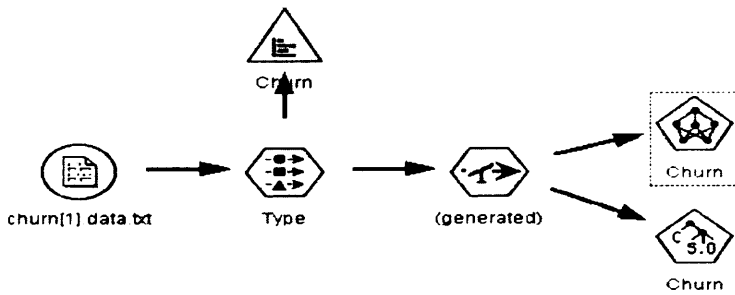


Fig.23 고객 이탈분석 프로세스의 흐름도

③ 이렇게 입력된 노드 중 지정한 데이터를 RFM테이타가 저장한 CRO를 이용하여 쇼핑몰에서 로그인한 회원에 따라 상품배열을 다르게 한다. 구체적인 소스는 다음과 같이 개발하였다. 쇼핑몰에서 로그인하면 저장한 상품 파일 코드가 있고 그 코드가 상품별로 지정이 되어있다 예를 들어 농산물은 코드가 A1이고 수산물 B1이고 축산물은 C1이라고 가정하자 그 회원은 상품구매패턴은 A1이 5개이고 B1이 2개이다 이렇게 구매패턴이 저장 되었다고 한다면 A1이 먼저 쇼핑몰 화면에 진열하는 방식으로 변환이 되면 그 전보다 클릭수가 많은 지 비교 분석하고 또한 구매액수 및 구매횟수를 그 전보다 훨씬 많았는지 적었는지 알 수가 있다.

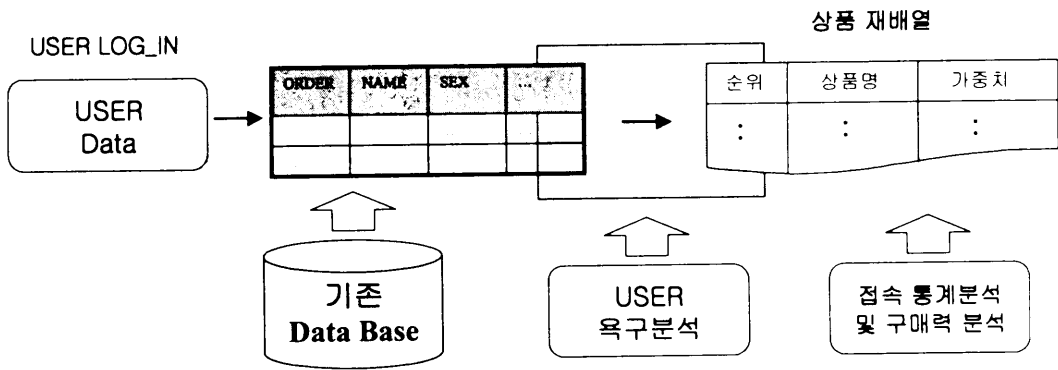
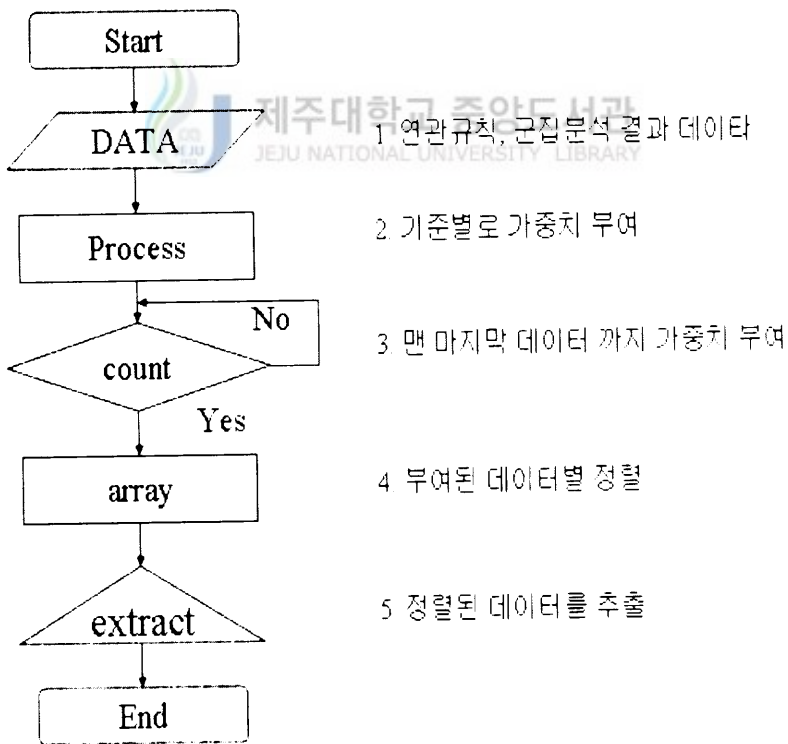


Fig.24 쇼핑몰에 적용한 상품재배열 흐름도

4. 적용한 소스



```
<%imagesql="select * from main_image" where a1 and b1 order by cross  
set imagers=dbcon.Execute(imagesql)
```

```
if imagers.eof or imagers.bof then
```

회원별로 진열을 다르게 함

```
gap="blank"
```

```
else
```

```
if imagers("cur_status")="no" then
```

```
gap="no"
```

```
elseif imagers("cur_status")="image" then
```

```
image=imagers("image")%>
```

```
<table width=619" border="0" cellspacing="0" cellpadding="0">
```

```
<tr>
```

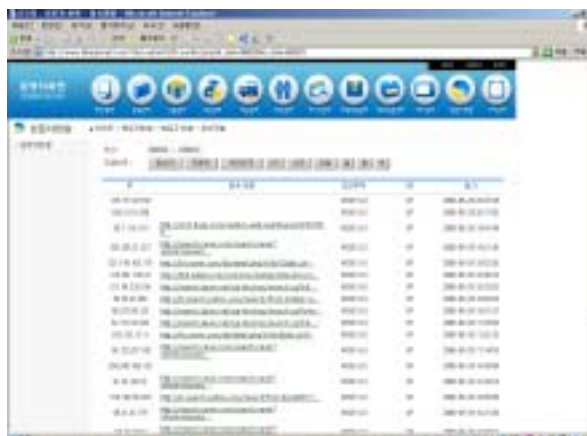
```
<td valign="top">
```

```
<%if right(image,3)="swf" then%>
```

```
<embed src="/main_images/<%=image%>" width="619"><% else %>
```

```
<%end if%>
```

⑤ 접속통계를 분석한다.



The screenshot shows a web analytics interface with a table of access logs. The table has columns for '시각' (Time), 'IP주소' (IP Address), '시도' (Country), and '시도' (Country). The data rows show various IP addresses and their corresponding countries, such as '192.168.1.1' from '대한민국' (South Korea) and '192.168.1.2' from '미국' (USA).

Fig.25 접속통계 대한 내용

IV. 구현 및 실험 결과

본 연구에서는 전략적 마케팅을 위한 지능형 인터넷 쇼핑몰을 설계하고 구현하기 위해, 고객들의 접속 기록과 상품 구매 기록 및 신상정보를 데이터마이닝 기법에 의해 통계적으로 분석하고 고객을 세분화하여, 고객이 상품에 대한 인기도에 따라 상품 진열을 자동적으로 구성할 수 있는 알고리즘 향상에 관한 연구이다.

1. 시스템 환경

본 논문에서 기존에 웹프로그램 언어인 ASP로 이용한 쇼핑몰에 데이터베이스에서 불러와서 SQL를 이용하여 고객의 구매패턴을 저장하여 비교분석하였다.

본 논문에서의 실험 환경은 CPU2.4Ghz, 메모리 512MB를 가지는 펜티엄 4 PC와 데이터베이스를 OFFICE에 기본적인 ACCESS를 이용했고 Windows 2000SERVER OS 환경에서 실험을 하였다. 실험을 위한 시스템 구현 프로그래밍 언어로는 ASP(Active server page)를 사용하여 구현 하였다.

Table. 6 Experiment Environments

시스템 사양	Pentium IV processor, 512MB RAM
운영체제	Windows 2000SERVER
프로그램 언어	ASP
사용된 데이터베이스	ACCESS

2. 제안 알고리즘 평가

데이터마이닝 기법을 이용하여 인터넷 쇼핑몰 사용자의 전공, 직업, 종교, 취미 등과 같은 신생정보나 구매 패턴에 따라 개인별로, 동적으로 화면을 구성하여 시스템을 구현할 수가 있었다.



Fig.26 시스템 구현 화면

3. 실험 결과분석

구매패턴 비교

대표적인 통계분석 시스템인 SPSS 9.0을 사용하여 모집군단인 1320명을 대상으로 적용했다.

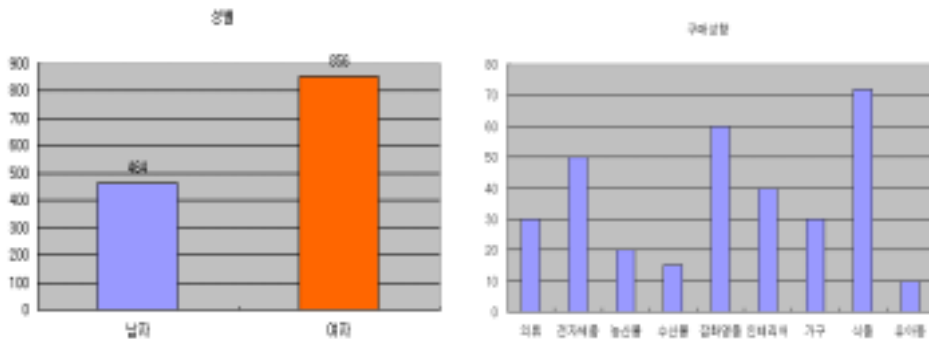


Fig.27 성별 및 성향 분포

어떤 경향이나 어떤 능력으로 구매하는지 비교 조사를 했으며 예측한 모델을 통한 구매건수와 이탈방지 모델을 통한 구매건수를 조사함.

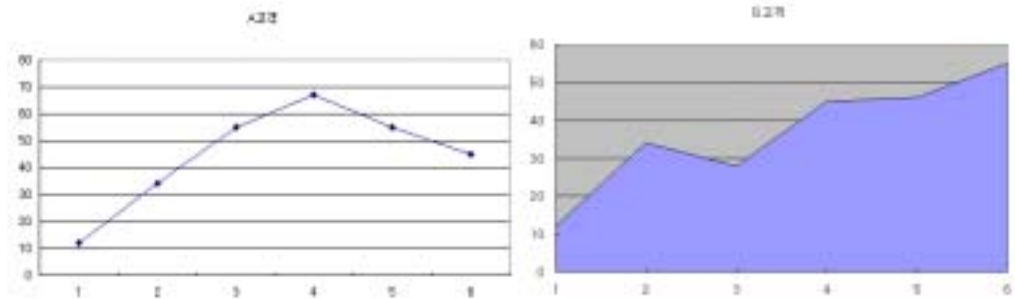


Fig.28 재구매 건수 / 이탈방지 통한 구매분포

개인별 차이도 있지만 평균적으로 긍정적인 효과가 있었으며, 보다 내실 있는 데이터를 바탕으로 보다 효과적인 모델을 제시 하려는, 재구매 욕구를 증대 시킬 수 가 있을 것이다.

현재의 신생 인터넷 쇼핑몰의 경우 데이터베이스가, 많은 문제점을 가지고 있어서 다채로운 군집분석을 할 수가 없었다. 따라서 기존의 알고리즘과 달리, 고객세분화 문제를 데이터마이닝 기법을 통해 고객집단을 분류하여, 고객속성에 대한 분석과 향후 신규고객에 대한 군집화를 실시하였으며, 여기서 얻어진 고객들의 속성을 참고하여 상품진열, 쇼핑몰제작 등에 개선 알고리즘을 적용하는데 있어서, 고객이탈 방지 알고리즘을 예측모델로 제공함으로써 기존 비구동방식의 쇼핑몰과 비교하여 보다 효과적인 마케팅을 할 수가 있다.

① 연간 구매횟수에 대한 고객그룹

고객그룹을 1-3회와 4회이상 그룹으로 나누어 구매자의 패턴을 조사했으며, 4회 이상의 고객의 재구매 했던 요인을 분석 예측 모델로 삼았다.

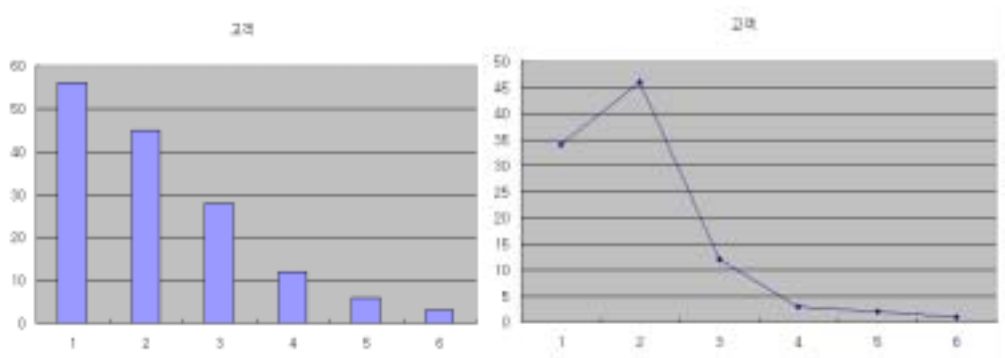


Fig.29 연 구매횟수 / 연 재구매 횟수

② 출력노드와 은닉노드을 이용한 분석

기존 고객의 테이블에 추가적으로 예측 모델로 쓰일 수 있는 월구매횟수, 연구매횟수, 연간구매횟수, 재구매횟수를 은닉노드로 적용하여 적용률을 극대화 시킬 수 있는 예측모델로 만든다.

기존방법과 제안방법의 비교

비구동방식의 쇼핑물과 유동방식의 쇼핑물을 접속패턴과 월,연구매횟수와 재구매건수의 수치를 비교하여 월 매출액과 조회건수의 차이점을 객관적으로 조사했다.

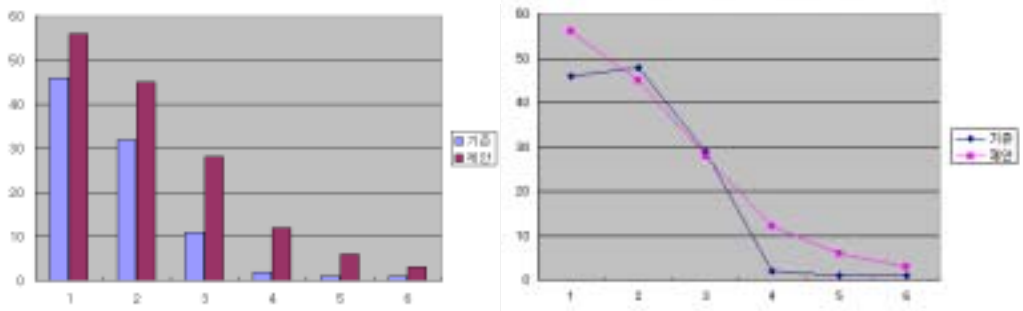


Fig.30 월 구매횟수 비교 / 월 재구매건수비교

유동방식의 쇼핑물은 구매욕구가 기존방식보다 재 구매 욕구를 자극시켰다는 증거가 되었으며 전체적으로는 큰 차이는 없으나, 예측 모델이 고객의 구매력과 이탈방지 시스템의 적용률이 높아 조금이나마 구매횟수를 늘렸으며, 고객의 세분화가 그만큼 전략적 마케팅에서 우수한 결과로 이어진다는 증거가 되겠다.

V. 결 론

본 연구에서는 기존 쇼핑몰 관리자의 주관적인 판단에 의해 수작업으로 이루어지는 쇼핑몰 관리업무의 일부를 자동화할 수 있었으며, 고객들의 접속기록과 구매 기록 및 개인 신상 정보로부터 상품들 간의 연관성을 분석하여, 관련 상품들끼리 그룹화 하여 제시함으로써, 고객이탈을 최소화하고 보다 지능적인 쇼핑몰의 개발과 쇼핑몰 관리를 자동화할 수 있는 알고리즘을 제시하였으며, 향후 신규고객의 예측모델과 기존고객에 대한 데이터마이닝 기법을 통한 분류 예측으로 보다 적극적인 마케팅을 할 수 있게 되었고, 최근 급격하게 증가하고 있는 쇼핑몰 시장에서 경쟁력을 강화할 수 있었다.

본 연구대상 기업의 쇼핑몰의 경우 데이터베이스가 많은 문제점을 지니고 있어서 다채로운 분석을 실시 할 수 없었기 때문에, 추후 연구과제로서 연구대상 쇼핑몰에서의 효율적인 매출분석과 고객관리차원에서 구매고객DB의 설계가 좀 더 효율성 있게 구성되어야하며, 구매고객들의 접속분석을 위해 로그(log)자료가 추가되어 동적인 분석을 실시해야 할 필요성이 있다. 왜냐하면, 인터넷 쇼핑몰에서 고객의 로열티 산정 문제는 오프라인 형태의 쇼핑몰과 달리 접속횟수(방문횟수)가 가장 중요한 변수가 되기 때문이다. 그리고 앞으로 이러한 자료가 계속 누적되고 웹사이트의 Hit수가 효율적으로 관리될 수 있도록 설계가 필요하다. 또한 신생 인터넷 쇼핑몰의 환경을 고려하여 기존의 데이터베이스 마케팅과 달리, 고객세분화 문제를 데이터마이닝 기법을 통해 고객집단을 분류하여, 고객속성에 대한 분석과 향후 신규고객에 대한 분류예측을 실시하였으며, 여기서 얻어진 고객들의 속성을 참고하여 웹 카탈로그 제작, 상품진열 등에 적용한다면 더 많은 매출을 올릴 것이라 기대된다.

[참고문헌]

- [1] 김진철, 황보윤, “A Design and Implementation of Intelligent Internet Shopping Mall using datamining technology”, 1999
- [2] 신승준 · 김순곤, “A study on Comparison of Analysis for Marketing Strategy using Datamining”
- [3] 김시환 · 권영식, “Customer Segmentation of the internet shopping mall using datamining”
- [4] 박성용, “데이터마이닝을 이용한 효과적인 데이터베이스 마케팅에 관한 연구”1998
- [5] 김정수, “통합 데이터베이스 마케팅 시스템”, 1997
- [6] 권오준, “통계적 기법과 인공지능망을 이용한 신용카드 고객 신용도평가 모형에 관한 연구”, 한국과학기술원, 1997
- [7] 정재호, “인공 신경망을 이용한 신용카드 부정사용 색출”, 한국과학기술원, 1998
- [8] Michael J. A .Berry Gordon Linoff, “Data Mining Techniques for Marketing, Sales, and Customer Support”, WILEY COMPUTER PUBLISHING
- [9] Joel Piphagen Alaina Kanfer, Ph.D. “In Search of the Elusive User: Gathering Information on Web Server Access”
- [10] Ravi Kalakota, Andrew B. Whinston, “Electronic Commerce A Manager’s Guide”, ADDISON-WESLEY
- [11] 차영준, “SPSS/PC+를 이용한 비모수 통계학”, 자유아카데미, 1993
- [12] Pieter Adriaans and Dolf Zantinge, “Data Mining”, 그린, 1998
- [13] 장남식, 홍성환, 장재호, “데이터마이닝”, 대청, 1999
- [14] 조재희, 박성진, “OLAP 테크놀로지”, 시그마 컨설팅 그룹, 1999
- [15] 조재희, “데이터웨어하우스와 OLAP”, 대청, 1996
- [16] 송현수, “통합 DBMS”, 새로운 제안, 1998
- [17] 김대수, “신경망 이론과 응용”, 1992
- [18] Han . Kamber, “Data Mining-Concepts and Techniques”, 자유아카데미, 2004
- [19] 김광용, 김명섭, “고객관계관리(CRM)를 위한 데이터마이닝의 활용과 실습”, 숭실대학교 출판부, 2005
- [20] 강현철, 한상태, 최종후, 김은석, 김미경, “SAS Enterprise Miner를 이용한 데이터마이닝-방법론 및 활용”, 자유아카데미, 2000
- [21] 강현철, 한상태, 최종후, 김은석, 김미경, “SAS Enterprise Miner를 이용한 데이터마이닝-기능과 사용법”, 자유아카데미, 2000

- [22] 강현철, 한상태, 최종후, 김은석, 김미경, “AnswerTree를 이용한 데이터마
이닝 의사결정나무분석”, SPSS 아카데미, 1998
- [23] Michael J. A. Berry and Gordon Linoff, “Data Mining Techniques
for Marketing, Sales, and Customer Support, New York: John Wiley &
Sons, Inc, 1997



감사의 글

힘차게 시작했던 2년간의 대학원 생활이 벌써 하루하루 흘러 어느덧 아쉬운 끝으로 다가 왔습니다. 2년 전만 해도 대학원이라는 생활이 참으로 힘들 거라는 두려움으로 시작했지만, 생활해 가면서 교수님들과 선배님들 그리고 같이 공부하고 연구하던 동기들이 있어 참으로 힘든 줄 모르고, 열심히 대학원 생활을 할 수 있었습니다.

이제 대학원 생활의 마지막 자리에 서면서, 언제나 관심을 가지고 끝까지 성심 성의껏 지도해 주신 김장형 지도 교수님과, 이렇게 좋은 결과를 맺을 수 있도록 조언해 주신 안기중 교수님, 곽호영 교수님, 변상용 교수님, 이상준 교수님, 송왕철 교수님, 변영철 교수님, 김도현 교수님께 먼저 감사의 마음을 전합니다.

또한 연구실에서 공부하는 동안 언제나 대학원 생활과 연구에 대해 조언을 아끼시지 않으시고, 관심을 가져주신 박사 강진석 선생님, 박사과정인 강영도 선생님, 김정효 선생님, 강길봉 선생님, 변태보 선생님, 고봉수 선생님, 양동호 선생님께도 진심으로 감사드립니다.

그리고 같이 연구하고 생활하던 석사 김제석 선생님과 멀티미디어 연구실의 여러 식구들과 대학원에 같이 들어와서 함께 열심히 공부했던 저의 동기들 및 학과 사무실에서 저희들을 위해 애써주신 이정하, 정은경 조교 선생님께도 감사의 마음을 전하는 바입니다.

2년이란 짧은 대학원 생활동안 이처럼 감사의 마음을 전할 수 있는 훌륭하신 교수님과 좋은 선배님들 그리고 함께 했던 동기들이 있어 대학원 생활이 저희 기억 속에 평생 좋은 추억으로 남을 것입니다.

끝으로 언제나 가슴 졸이면서 대학원 생활을 지켜봐 주시고, 힘차게 격려하여 주신 모든 분들에게 진심 어린 감사의 말씀을 드립니다. 이 모든 분들께 살아가면서 언제나 좋은 일만 있으시길 빌겠습니다.

감사합니다.