

# 데이터 마이닝의 금융권 활용방안

이번 호에는 금융업계에서의 데이터 마이닝의 적용사례에 대해서 살펴보겠다. 그동안 5회에 걸쳐 1, 2회에서는 데이터 마이닝의 전반적 소개를, 그리고 3~5회동안 활용사례를 집중적으로 소개했다. 데이터 웨어하우스의 활용에서 가장 각광받고 있는 분야가 데이터 마이닝이라는 것에는 많은 사람들이 공감을 표하고 있다. ROI라는 측면에서 보아도 가장 쉽게, 안심하고 다가갈 수 있는 분야 또한 데이터 마이닝이다. 그러나 마이닝이란 단순히 마이닝 툴 혹은 마이닝 기법이라는 요술 방망이를 통해서 나오는 금은보화는 결코 아니다.

글/한국싸소소프트웨어 DB 마케팅팀

연	재	순	서
1.	데이터마이닝이란?		
2.	마이닝의 접근방법론 및 기법		
3.	활용사례 : DB마케팅을 위한 접근 방법론		
4.	활용사례 : DB마케팅 사례		
5.	활용사례 : 금융권 활용방안(이번호)		

## 카드의 도용사고방지 (Fraud Detection)

금융개방화 및 자율화를 비롯한 기업의 내·외부적인 환경변화로 인해 각 업체들은 더욱 더 심화되어가는 경쟁 상황하에 놓이게 되었고 이에 보다 효과적이고 실질적인 생존전략이 필요하게 되었다.

이러한 상황은 신용카드업계에도 예외는 아니어서 카드발급에 있어서의 자격심사나 카드론 등 고객 중심의 관리(CRM : Customer Relationship Management)를 비롯하여, 신용카드의 분실도난, 위조 및 변조, 부실가맹점, 부실채권 등으로부터 발생하는 다양한 리스크를 관리하기 위한 체계적인 방안이 필요하게 되었다. 더군다나 최근 분실도난이나 위조 및 변조와 같은 각종 신용카드를 이용한 범죄의 규모와 발생률이 증가하면서 신용카드사가 입게 되는 피해손실액이 점차 증가하고 있는 실정이다.

그러나, 분실도난이나 위조 및 변조를 포함한 각종 부정사용에 대해 대부분의 기존 신용카드회사에

서 취하고 있는 대책은 다분히 사후 조치에 머무르고 있어서 실질적인 예방효과는 전무한 상태이다. 이러한 상황에서 BC카드는 우선 분실도난에 의한 부정사고를 효율적으로 예방·관리하기 위해 데이터 마이닝 기법을 적용하게 되었다. 즉, 분실도난에 의한 부정사고를 효과적으로 예방하기 위하여 카드로부터 승인요청이 들어온 시점에서 해당 카드의 이전 사용패턴과 비교하여 사용된 카드의 부정사용 여부를 자동적으로 판단해주는 일명 '분실도난 조기 검색 시스템'을 개발하게 되었다.

분실도난 조기검색 시스템은 승인요청이 들어오면 부정사용의 가능성 정도에 대한 점수를 산출하고 일정기준 이상의 점수를 받은 즉, 부정사용이 의심되는 승인전에 대해서는 승인을 보류하여 본인여부를 확인하는 절차를 거친 후 승인여부를 결정할 수 있도록 하는 것이다.

## 시스템 환경과 개발 과정

분실도난 조기검색 시스템 개발이 진행된 시스템 환경은 데이터베이스 서버로 디지털 Alpha 8000 시리즈와 상용 RDBMS가 사용되었고, 데이터 마이닝 서버로는 디지털 Alpha 4100이 사용되었다.

분실도난 조기검색 시스템 개발 프로젝트는 크게 3단계 즉, ▲과거의 부정사용 거래내역을 중심으로 부정사용여부의 예측을 위한 모형의 구축단계 ▲재



모형화 작업의 용이성을 위한 모델링 시스템의 개발 단계 ▲그리고 실제 승인시스템에서의 적용단계로 진행되었다.

분실도난 조기검색 시스템의 핵심은 부정사용여부를 모형이 얼마나 정확하게 가려낼 수 있는가 하는 것이다. 즉, 모형의 예측력이 관건이 된다.

예측력은 크게 두가지 요소를 가지고 있는데 모형으로부터 산출된 점수를 사용하여 부정사용여부를 예측했을 때 예측의 결과가 얼마나 정확한가를 의미하는 검출효율과 모형을 활용했을 때 전체 부정사용을 어느 정도 감소시킬 수 있는가를 의미하는 검출력이라는 부분으로 구분된다(이와 같은 검출효율과 검출력은 실제 모형구축단계 중 모형평가단계에서 다양한 후보 모형들 중에서 최종모형을 선택하는 평가 기준으로 이용된다).

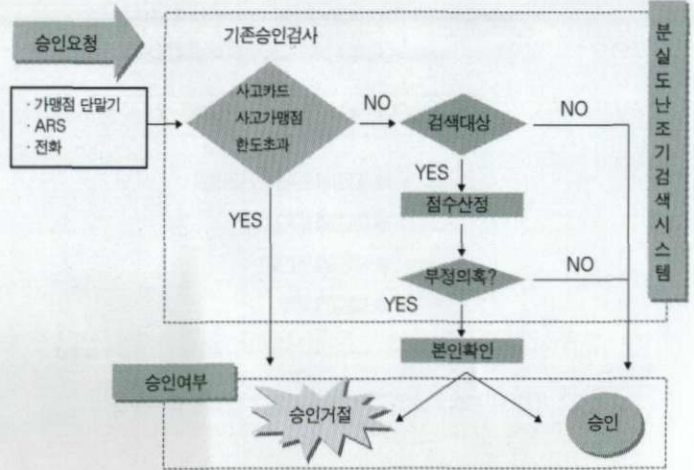
분실도난 조기검색 모형의 구축의 핵심은 카드의 사용에 있어서 정상 사용과 부정 사용이 구별되는 분기점에 대한 다각도의 탐색과 이에 대한 적절한 모형을 생성하는 작업으로서, SAS의 데이터 마이닝 프로세스인 SEMMA단계에 따라 진행되었고, 데이터 마이닝 기법으로는 신경망 기법(Neural Networks)이 이용되었다.

분실도난 조기검색 모형의 구축과정을 좀더 자세히 살펴보면 다음과 같다.

먼저 데이터 마이닝을 위한 데이터준비작업이 선행되었는데 과거 일정기간동안 카드거래 및 승인내역을 비롯한 카드/고객에 관한 속성 데이터 등 가용 데이터와, 실제 발생한 부정거래의 내역에 관련된 데이터를 적절한 형태로 준비하였다. 부정사용의 패턴을 찾아내기 위하여 각 개별카드의 과거 부정사용으로 판명된 거래와 정상적인 사용에 의한 거래데이터에 대해 카드단위로 샘플링 기법을 이용하여 충분한 샘플을 생성하였다.

정상사용과 부정사용의 특성을 잘 반영하는 변수들을 그래프나 통계적 방법을 이용하여 검색하는데, 이때 실제 운영시스템으로부터의 변수들을 근간으로 다양한 패생변수를 생성하였고 이러한 변수들과 부정사용여부의 관계를 다각도로 탐색하여 주요변수를 최종 선정하였다.

다양한 모형화 기법을 적용할 수 있는데, BC카드

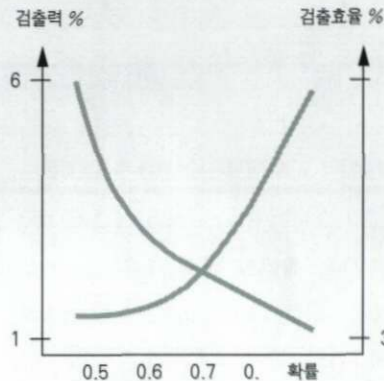


(그림 1) 분실도난 조기검색의 흐름

(표 1) 모형의 비교 기준

부정여부	모형	정상 934	부정 66
정상 990 (99%)	검출건수	92	6
	검출력	93.7	6.3
	검출효율	99.36	93.84
부정 10 (1%)	검출건수		4
	검출력	60	40%
	검출효율	0.64	6.06%

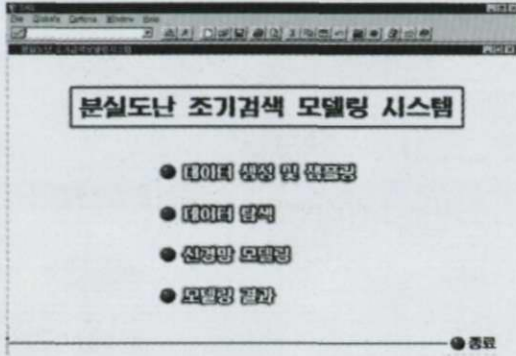
· 검출효율: 모형이 부정으로 추천한 거래 중에서 실제 부정거래의 비율  
· 검출력: 전체 부정거래 중에서 모형을 통해 검출된 부정거래의 비율



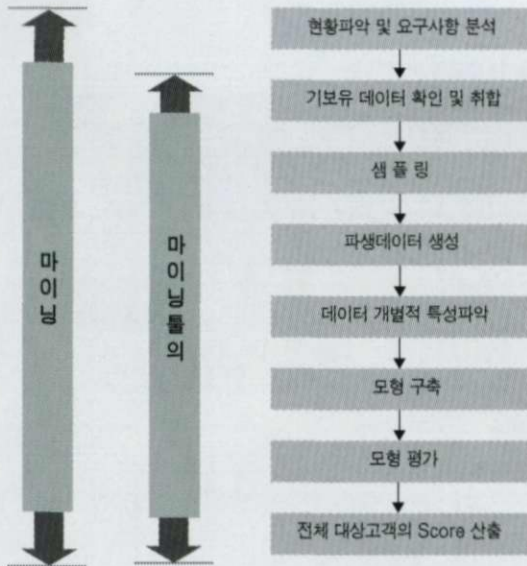
(그림 2) 모형의 비교 기준

의 경우에는 선정된 변수를 중심으로 SAS의 신경망(Neural network) 알고리즘을 이용하여 여러 가지 가능한 아키텍처로 모형화를 시도하였다.

모형의 평가 단계에서는 몇 가지 후보 모형의 성능을 검출효율과 검출력 측면에서 비교하여 최종적으로 예측모형을 선택하였다.



〈화면〉 분실도난 조기 검색 모델링 시스템의 초기화면



〈그림 3〉 A보험사의 고객이탈방지시스템의 추진방법론

또한 카드사용 패턴의 변화 혹은 일정시간의 경과 등에 의해 모형을 다시 조정해야 할 필요성이 생기게 되는데, 이런 경우 실제 담당자들이 직접 모형화 작업을 시도하여 모형을 갱신하고 이에따른 예상 ROI를 쉽게 확인할 수 있도록 담당자들의 직접 데이터 마이닝작업을 수행할 수 있게 해주는 분실도난 조기 검색 모델링 시스템 애플리케이션으로 개발하였다.

분실도난에 의한 부정사용여부의 예측을 위해 구축된 모형은 기존의 승인시스템에 이식되어서야 비로서 그 진가를 발휘하게 된다.

즉, 모형구축단계에서 얻어지는 산출물은 분실가

능성 정도에 대한 점수를 산출하는 모형식이고 이러한 모형의 수식을 C와 같은 적절한 언어를 이용하여 승인시스템 내에 이식하는 것이다. 실제 승인요청시, 여러 정보들이 이 모형식을 거쳐서 부정사용 가능성에 대한 점수를 계산하고 적절한 점수기준에 따라 부정사용 여부를 판단한다. 여기서 중요한 것은 점수의 계산이 전체적인 승인체계에 부하를 주어서는 안된다는 것이다. 따라서, 이를 위해 프로그램의 최적화, 사전 테스트 등을 통한 충분한 검토가 필요했다.

또한, 기준점수를 얼마로 할 것인가, 일정점수 이상의 승인요청건에 대한 실제적 조치는 어떻게 할 것인가 즉, 본인을 확인하는 방법과 이를 위한 인력의 수급방안 등에 대한 충분한 사전검토가 필요하다.

지난 4월부터 분실도난조기검색시스템이 실제 승인시스템에서 운영되고 있는데, 이러한 시스템을 적용함으로써 부정사용으로 인한 손실액의 연 40% 정도의 감소를 가져올 것으로 예상되고 있다.

향후 부실가맹점 관리나 위조 및 변조에 의한 부정사용의 방지를 위한 데이터 마이닝의 활용을 고려하고 있으며, 신용평점시스템과 행동평점시스템의 신용관리시스템(Credit Scoring System)에서의 데이터 마이닝의 활용 또한 검토중에 있다.

### 보험사의 고객이탈방지시스템

A보험사가 고객이탈방지 시스템을 구축하게 된 배경 및 목적은 다음과 같다.

1. 신규고객 가입감소(시장의 포화상태로 인한 시장 성장률의 둔화) 및 치열한 업체간의 경쟁(보험상품 및 다양한 서비스)으로 인하여 고객이탈을 증가, 이에 따른 마케팅 비용증가 등 급변하는 경영환경변화에 능동적으로 대처하기 위하여
2. 이탈에 영향을 주는 주요 변수를 식별 및 분석할 수 있는 모형 개발과 이러한 변수를 이용하여 계약자 이탈을 사전에 예측할 수 있는 모델을 개발함으로써
3. 각각 계약자(고객)에 대해 이탈가능성을 점수화(Customer behavior Scoring)하여, 매출 및 손실에 영향을 미치는 이탈 예상 고객을 사전에 파악함으로써
4. 이탈 방지를 위한 고객과의 관계를 개선 및 지속적인 관



계강화를 통하여 고객 이탈을 최소화함으로써 기존의 시장 점유율을 유지하고

5. 매출 뿐만 아니라 수익 부문에서도 경쟁우위를 가질 수 있는 합리적인 경영 환경을 조성하고자 하는 데 있다.

A보험사의 고객이탈방지시스템의 추진방법론은 (그림 3)과 같다.

## 단계별 프로젝트 진행과정

### 1. 현황파악 단계

프로젝트에서 무엇을 할 것인지, 어떠한 방향으로 진행을 해갈 것인지 정의하는 단계로서 A보험사의 업무 이해와 비즈니스 목적의 명확한 이해를 통하여 프로젝트의 정의 및 방향성을 설정하는 단계이다.

A보험사의 매출액 및 기회손실액에 영향을 미치는 이탈예상계약(자)은 무엇이며, 분석을 하고자 하는 보험상품에서 이러한 이탈을 결정하는 기준은 무엇을 어떻게 선택할 것이며, 이 기준하에서 정의된 이탈예상자의 특성은 무엇이며, 이러한 이탈에 영향을 미치는 요인들은 무엇인지를 파악하여, A보험사는 고객의 이탈을 사전에 예측할 수 있는 모형을 구축하는 것을 프로젝트 목적으로 하여 진행하였다.

A보험사는 이탈에 가장 큰 영향을 주는 요인으로 영업사원의 특성, 예를 들자면, 근속기간, 근태상황, 계속하여 자신이 계약한 고객을 관리하고 있는 지의 관리지속성 여부 등으로 예상하고 접근하였다.

### 2. 기보유 데이터 확인 및 취합 단계

목적에 부합되는 필요한 데이터를 정의하고 데이터를 추출하는 단계이다. 필요한 데이터를 크게 주제별로 나누어보면, 고객의 성별, 연령, 현주소 등과 같은 고객관련 데이터, 영업사원(모집설계사)관련 데이터, 고객이 가입한 보험상품에 관련된 데이터인 가입상품관련 데이터, 대출여부와 대출종류에 관련된 대출관련 데이터, 탈락관련 데이터, 거래형태별 데이터를 사용하였다.

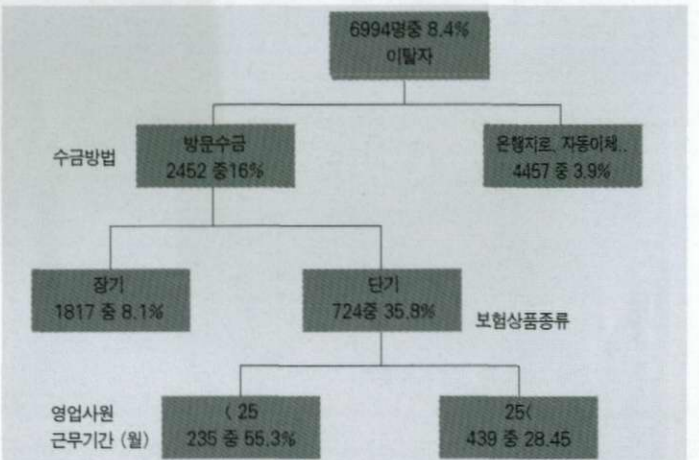
### 3. 샘플링 단계

분석의 편이성을 위하여 데이터의 크기를 조절할

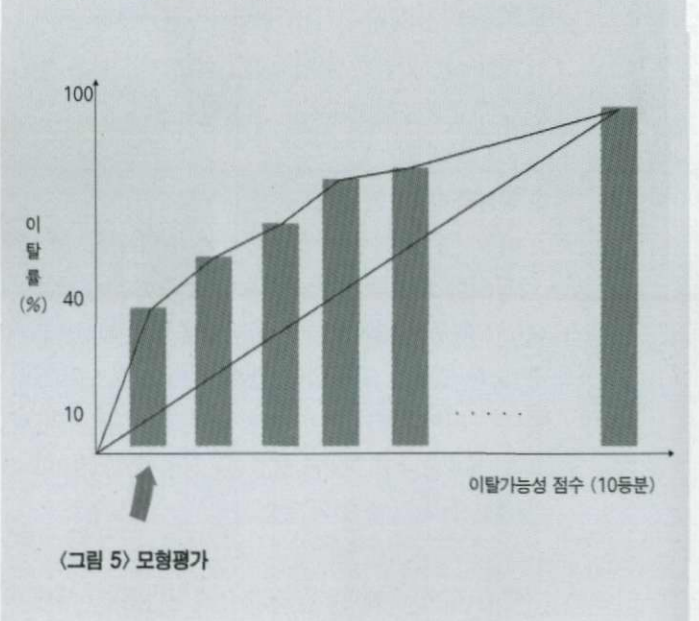
〈표 2〉 생성된 파생데이터의 예

요인	기본데이터	파생데이터	생성규칙
이탈자	탈락일자	이탈여부	이탈 : 1 정상 : 0
성별/연령과 이탈	주민등록번호	성별, 연령	
가입기간과 이탈	가입일, 이탈일	유지개월수	탈락월-가입월
계약액에 대해 현재까지 납입한 금액의 비에 따른 이탈 여부	계약년월일 월보험료 납입기간	납입비	월보험료*유지개월수 계약액

(\*위의 표의 예는 이해하기 쉽고, 손쉽게 만들 수 있는 몇 개의 데이터를 예로 든 것이며 실제 프로젝트와는 무관하다.)



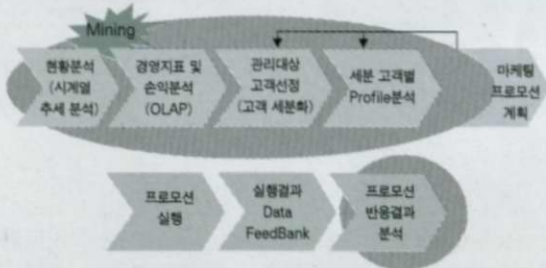
〈그림 4〉 의사결정나무의 예



〈그림 5〉 모형평가

필요가 있었다. 이에 샘플링기법을 사용하였는데 분석할 샘플이 전체 모집단의 성격을 충분히 반영할 수 있도록 적합한 샘플링기법을 사용하는 것이 중요하다. 예를 들면, 모집단(관측치수 : 1만건)에서 남녀의 구성비가 7 : 3이라고 하면, 샘플링한 집단(관

〈그림 6〉 고객세분화 및 고객특성파악



〈그림 7〉 프로젝트 절차

측치수 : 1000건)에서도 남녀의 구성비가 7 : 3이어야 한다는 것이다. 즉, 분석을 용이하게 하기 위해 관측치의 수를 줄이기는 하지만, 모집단의 특성을 간과해서는 안된다는 것이다. A보험사에서는 단순 임의 추출방법을 이용하여 샘플링하였다.

4. 파생데이터 생성 단계

이탈여부를 차별화 할 수 있는 요인들을 정의한 후 정의한 요인들을 실제로 수치로 구현하는 단계이다. 이 과정이 실제로 마이닝작업을 하는데 있어 가장 많은 노력이 요구되는 단계이다. 그러므로 이러한 파생데이터를 용이하게 활용할 수 있는 툴의 확보가 중요하다고 할 수 있겠다. 생성된 파생데이터의 예를 들면, 〈표 2〉와 같다.

만일 데이터 웨어하우스가 구축된다면, 파생데이터를 만들기 위한 기본데이터의 질에 대한 문제는 해결될 것이다.

5. 데이터의 개별적 특성 파악

분석에 사용할 데이터의 특성을 파악하여, 데이터가 많을 경우 모형화하는 과정에서 포함시킬 데이

터를 선정하는 단계이다. 부연하면, 생성된 데이터 중에서 실제로 이탈에 영향을 미치는 변수는 무엇이며 해당변수들이 영향을 미치는 정도는 어떠한가를 판단하는 단계이다. 이 단계에서는 의사결정나무와 그밖의 통계분석기법을 사용하였다. 〈그림 4〉는 의사결정나무의 예이다.

의사결정나무(Decesion Tree)는 특정한 경우(이탈)이 일어날 가능성이 높은 집단과 낮은 집단으로 나눈다.

6. 모형구축 단계

여러 후보변수들이 또는 의사결정나무의 결과로 나온 변수들이 동시에 이탈여부에 어떠한 영향을 미치는지를 파악하는 모델을 만들고, 이 모델식이 만들어지면, 이를 근거로 하여 이탈가능성에 대한 점수(Scoring)를 산출한다.

A보험사에서는 신경망기법과 로지스틱회귀분석(Logistic regression)를 모델링 방법으로 사용하였는데, 다음 단계인 모형평가 단계에서 로지스틱회귀분석이 더 좋은 모델로 평가되었기에 로지스틱회귀분석을 사용하였다.

7. 모형평가 단계

여러 후보변수들의 다양한 조합과 다양한 모델링 기법들을 이용하여 모형을 구축하였을 때 그 모형들 중 이번 목적에 가장 적합한 모델을 찾아야 하는데, 이렇듯 가장 효율적인 모델을 찾기 위해, 각 모형들을 평가하는 단계이다. 〈그림 5〉는 모형을 평가하는 방법의 예이다.

이탈점수가 높은 전체 고객의 10%중에 실제 이탈자의 40%가 포함됨을 의미한다. 이러한 방식으로 구한 여러가지의 모형을 대상으로 평가가 가능하다.

고객행동점수를 마케팅 또는 캠페인 계획측면에서의 활용

각 목적에 따라 구한 다양한 점수들을 이용하여 정교하게 고객들을 세분화하고, 세분화된 특수고객 군들에게 적합한 마케팅 계획이나 캠페인 계획을 수립을 세울 수 있도록 지원하는 데 사용할 수 있다.



이러한 점수들을 사용하여 마케팅 부서 등 현업에서 활용할 수 있는 이탈점수 활용의 예를 들어보자.

1. 이번달 DM발송을 하고자 하는데 이탈점수가 70점 이상인 고객에 대해서 하고자 한다.

2. 이번달에는 이탈로 인하여 회사에 손실을 주는 고객을 관리고객으로 삼고자 하는데, 그 기준을 어떻게 선정할 것인가?

3. 이번달 관리고객 100명에게 행사를 하고자 한다. 이때 100명은 누구인가?

위에서 예로 든 것과 같이 좀 더 세밀한 계획을 수립할 수 있다. 그러므로 최적의 캠페인 및 마케팅 계획을 수립할 수 있으며, 이에 따라 캠페인 반응률 및 성공률을 극대화할 수 있고, 고객의 선정이 최적화 되었으므로 유치비용 대비 수익을 증가시킬 수 있다.

4. 예측한 이탈자로 인하여 경영손실은 어떻게 되며, 경영손실을 줄이기 위하여 어떤 고객을 관리해야 하는가?

등과 같이 경영지표 및 손익분석을 할 수 있다. 그리고 고객세분화와 그에 따른 고객특성을 분석함으로써 보다 합리적인 마케팅 계획을 수립할 수 있다. 물론 이러한 경영지표 및 손익분석 그리고 고객세분화와 프로파일분석은 OLAP을 통하여 구현될 수 있다.

<그림 6>은 이탈점수를 이용하여 고객세분화 및 세분화된 고객들의 특성파악이 가능하도록 구현한 OLAP의 한 예이다.

마지막으로, A보험사의 프로젝트절차를 나타내면, <그림 7>과 같다.

즉, 현재 비즈니스 이슈가 무엇인지, 즉, 회사의 문제가 무엇인지를 파악한 후 마이닝을 통하여 분석을 수행하고 그 결과를 마케팅, 캠페인 계획을 수립하는데 근간이 되는 정보로 제공하는 절차이다.

## 연재를 마치며

그동안 6회에 걸쳐 데이터 마이닝에 대하여 소개해 보았다. 1, 2회에서는 데이터 마이닝의 전반적 소개를, 그리고 3~6회동안 활용사례를 집중적으로

소개하였다. 데이터 웨어하우스의 활용에서 가장 각광받고 있는 분야가 데이터 마이닝이라는 것에는 많은 사람들이 공감을 표하고 있다. ROI라는 측면에서 보아도 가장 쉽게, 안심하고 다가갈 수 있는 분야 또한 데이터 마이닝이다. 그러나 마이닝이란 단순히 마이닝 툴 혹은 마이닝 기법이라는 요술 방망이를 통해서 나오는 금은보화는 결코 아니다.

데이터 마이닝은 잘 준비된 데이터, 비즈니스와 고객에 대한 깊은 통찰력, 다양한 분석기법을 활용하여 정확하게 문제를 풀어낼 수 있는 분석능력, 구현 단계에서 발생할 수 있는 기술적 문제들을 해결할 수 있는 능력, 모든 과정을 조화롭게 이끌어 갈 수 있는 리더쉽 등 여러가지 요인이 함께 어우러졌을 때만 가능한 어쩌보면 고통스런 작업이다.

서양의 우수한 성공사례를 일일이 들추어보지 않더라도 데이터 마이닝이 가지고 있는 거대한 잠재력을 우리는 쉽게 감지할 수가 있다. 그 구현과정은 비록 복잡하고 힘들지라도 그 개념은 매우 명확하고 쉽게 우리에게 다가오기 때문이다. 특히 데이터 마이닝이 약속하는 장미빛 ROI를 그 누구도 외면할 수는 없는 것이다.

마지막으로 성공적인 마이닝을 위한 간단한 제언으로 글을 맺을까 한다. 성공적 데이터 마이닝을 위해서는 무엇보다 철저한 사전 준비 단계가 매우 중요하다. 특히 분명한 목적의식이 있어야 한다. 목적이 없거나 불분명한 채로 데이터 마이닝을 시도하는 것은 광부가 목적없이 곡괭이질을 하는 것과 같다. 정확한 목적을 가지고 정확한 곳을 파고들어가야만 목표표한 것을 발견할 수 있다는 것은 자명한 사실이다.

이밖에도 최고경영 층의 확고한 의지와 지원, 프로젝트 관리자의 리더쉽, 구성원들의 추진력과 팀웍, 프로젝트 과정에서 발생할 수 있는 현업부서와 IT간에서 발생할 수 있는 어려움 등 보이지 않는 변수들도 프로젝트의 성패에 중요한 요인으로 작용할 수 있다.

특히 데이터의 준비과정에서 발생할 수 있는 여러가지 기술적 문제점들과 데이터의 신뢰성의 검증이 사전에 반드시 행해져야 한다는 점을 강조하고 싶다.