

시작하며...

- 데이터 마이닝이 기업에서 실제 활용되기 위해서는 적용하려는 분야의 **특성**이나 **프로세스**를 잘 이해하여야 합니다.
- 답변의 대부분은 아마 여러분들이 아직 신용카드를 사용해 본 적이 없어 카드 사용자에게 대한 패턴이나 신용카드사들이 어떤 프로세스를 거치는지 잘 알지 못해 나오는 답변들이 대부분이었던 것 같습니다.
- 혹은 아직 경제활동이 적어 사용하려는 금액이 낮거나 (2~3만원) 카드를 잃어버린 카드 소지자에 대한 예외로 유흥주점 같은 곳에서는 사용하지 않겠다는 답변도 있었습니다.
- 그렇다면 우리가 신용카드 부정사용을 탐색할 수 있는 모형을 만든다는 전제 하에 문제 접근 방법을 (정답은 아니지만, 아니 정답이 있을 수도 없지만) 나 나름대로 정리해 보겠습니다.

문제의 정의

● 범위 설정 (scoping)

- 부정사용에도 여러 종류가 있습니다. 우리는 현재 타인의 카드를 불법으로 사용하는 것에 국한하고 있습니다.
- 그리고 국내 부정사용에 국한할 것인지, 해외 부정사용도 고려할 것인지 등등 데이터 마이닝 적용범위를 확실히 하여야 합니다.

● 목표 설정

- 데이터 마이닝을 통해 어떤 효과를 얻을 것인지가 명확하여야 합니다.
- 즉, 데이터 마이닝을 통해 부정사용하는 결제 정보를 실시간으로 감지(detection)하는 것인지, 그래서 어느 정도로 부정사용을 잡아낼 수 있는 것인지 등이 여기에 해당됩니다.
- 이 기대효과를 명확히 하기란 쉽지 않지만 (마치 어머니가 이번 중간고사에는 몇등할거나 묻는 것과 똑같기 때문에...) 기업에서는 적지 않은 돈을 투자하는 것이므로 기대효과를 산출/언급하여야 하고 이 기대효과가 자연스럽게 목표가 되는 것이지요.

● 역할과 책임 정의 (Role & Repososnibility)

- 범위와 목표가 설정되면 어떤 일들이 몇명의 사람이 투입되어 진행되어야 하는지 윤곽이 잡히겠지요.
- 그리고 어떤 데이터들이 필요한지, 그 데이터의 제공 등 데이터 마이닝을 진행하는데 필요한 사람들이 정의되고 협력체가 구성이 되어야 할 것 입니다.
- 그때 그 협력체나 TFT(Task Force Team)에 참여하는 사람들의 역할과 책임의 명확이 정의되어야 합니다.

● 일정

- 결국 데이터 마이닝 작업은 한 사람이 하는 일이 아니라 여러 사람이 협력하며 진행하는 것이므로 R&R뿐 아니라 일정에 따라 누가 어떤 결과물을 산출하고 그 결과물은 누가 받아 작업을 하는지에 대한 부분도 명확히 정의되어야 합니다.

- 결국 신용카드 부정사용 탐지를 위한 데이터 마이닝 작업을 성공적으로 완수할 수 있는지는 일정의 정밀성이나 타당성을 보면 알 수 있겠지요.

비즈니스의 이해

비즈니스 이해란 데이터 마이닝을 적용하려는 업무의 이해 정도로 해석하면 됩니다.

● 프로세스의 이해

- 프로세스는 신용카드가 (식당이나 커피숍에서) 결제되면 그 데이터가 전송되어 어떻게 승인이 나는지의 과정과 같이 비즈니스의 흐름을 이야기 합니다.
- 우리 문제와 관련된 프로세스는 아무래도 “승인 단말기 -> VAN사 -> 카드사의 승인 시스템 -> 승인 결과 전송 -> 승인”의 흐름이 될 것 입니다.
- 그러나 우리가 더 파악하여야 할 것은 우리가 개발할 데이터 마이닝 결과를 언제 적용할 것인가에 대한 것이겠지요.
- 그리고 승인결과를 판단하기 전에 수상한 거래라고 판단되면 수상한 거래만 처리하는 프로세스로 넘어가야 하는 그 프로세스는 어떻게 되고 있는지, 혹은 어떻게 되어야 하는지를 이해하거나 디자인하여야 합니다.

● 조직의 이해

- “문제의 정의” 중 R&R을 정의하면서 조직에 대한 파악이 일부 수행될 것 입니다.
- 그러나 이 과정을 다시 언급하는 이유는 기업은 기본적으로 조직체이므로 결국 데이터 마이닝 결과는 프로세스에 적용되거나 변화를 요구할 것이고, 복잡하게 얽힌 프로세스의 유기체의 기업 입장에서는 관련된 조직을 더욱 정밀하게 이해하는 것이 매우 중요한 일입니다.
- 즉, 조직의 협조를 받는 일뿐 아니라, 해당 조직의 프로세스를 변경하거나 이해하여야 하는 일들이 발생할 수 있습니다.
- 예를 들어, VIP카드 소지자가 어제까지 잘 사용하였는데 오늘 갑자기 수상한 거래라 잡힌다면 부정탐지 목적에서 거래를 일시 중지하여야하지만 False-alarm인 경우 VIP 고객에게 상당한 불편을 끼칠 수 있습니다. 이럴때 사후 처리를 VIP고객관리팀과 공조한다던지의 프로세스가 필요하겠지요.

● 데이터의 이해

- 데이터 마이닝에 있어 가장 중요한 일 입니다. 어떠한 관련된 데이터들이 축적되어 있는지 파악되어야 합니다.
- 아울러 신용카드 부정사용 탐색은 실시간으로 처리되어야 하므로, 데이터 생성 및 축적주기 등도 세밀하게 파악되어야 합니다.
- 그리고 고객 관점, 가맹점 관점, 그리고 부정사용 감지를 위해 집계되거나 계산되어 축적되어야할 데이터 관점 등 다양한 각도에서 이해되고 정의되어야 합니다.
- 물론 최종적인 데이터 정의는 데이터 마이닝 작업을 통해 이루어질 것 입니다.

● 케이스의 이해

- 그간 축적된 사기 유형(케이스)를 이해하여야 합니다.

- 본 과정에서 어떤 데이터가 필요하며 데이터 마이닝 작업을 어떻게 진행할지가 구체화될 수 있습니다.
- 어떻게 보면 대부분의 해답을 이 과정에서 얻어낼 수 있습니다.
- 아무래도 관련된 업무를 계속하신 분이 가장 전무가이니 케이스 이해 과정 중에 담당자와 많은 대화를 나누고 도움을 얻는 것이 매우 중요합니다.
- 컴플라이언스 혹은 민원의 이해
 - 금융사들의 특징은 정부로부터 각종 규제를 받고 있다는 점입니다.
 - 카드사 부정탐지 시스템은 우리나라 금감원에서 반드시 갖춰야하는 시스템으로 정의한 만큼 시스템에 대한 요건들이 정의가 되어있을 것 입니다.
 - 또 하나 중요한 고려사항은 민원 입니다. 손님이 왕인 시대에 카드사와 같이 이미 포화상태인 시장에서는 고객 민원이 발생하지 않도록 하여야 합니다.
 - 관련된 민원 케이스의 파악이나 민원 소지가 많이 발생하는 false-alarm의 원인을 잘 파악하여야 합니다.

이상 2013-09-20