

이 책을 통해 일관적으로 통계적 모델의 중요성을 강조해왔다. 제3장: '빅데이터의 핵심은 분석이며 분석의 중심은 모델링'에서 여러 가지 고전적인 모델 적용방법을 구체적으로 소개한 바 있으나, 데이터가 넘쳐나서 고민인 빅데이터 시절에는 복잡하고 방대한 양의 데이터를 간단한 '점수'로 변환하는 기능이 모델링의 가장 중요한

기능이라고 감히 말하고 싶다. 그 다음으로 중요한 요소는 'unknown'을 'potentials'로, '모르는 부분'을 '가능성'으로 변환시키는 기능, 즉 데이터에 존재하는 수많은 구멍들을 추정치로 매워주는 역할이라고 할 수 있겠다. 아무리 데이터가 사방에 널려 있는 시절이라 해도, 그 누구도 모든 사람에 관한 모든 것을 알 수는 없는 일이다.



예를 들어 접촉이 가능한 십만 명 이상의 대상을 분석할 때 그 중 '확실히 아는 정보'로 골퍼로 분류될 사람들은 극소수인 경우가 많다. 모델을 사용하면 '골프를 즐길 가능성'에 관한 점수를 그 데이터베이스 안의 모든 사람에게 적용할 수 있다. 모델이란 늘 100% 정확한 것은 아니지만, 회사 आयुषी 날 '오후에 소나기가 올 확률이 70%'

라는 일기예보는 아예 모르고 당하는 것보다 훨씬 낫는데 이의가 없을 것이다.

‘모르는 부분’을
‘가능성’으로 변환

글을 여기까지 이끌어 오면서 모델링의 많은
혜택들에 관해 자세한 설명을 했지만,

여기서 그 리스트를 아주 짧게 요약한다면 모델은 ① 마케터들이 누구를 우선적으로 상대해야 할지 정하는데 도움을 주며 ② 일단 어떤 대상을 접촉하기로 정해진 경우, 자신에게 상관 없다고 여겨지면 아무리 멋지게 꾸며진 광고라도 가차 없이 무시해버리는 현대의 소비자들에게 도대체 어떤 오퍼와 메시지를 가지고 접근해야 할지를 정하는데도 가이드가 된다는 것이다.

이어지는 질문은 모델링이 그렇게 중요하고 유용한 것이라면 도대체 ‘누가 그런 수학적 작업을 수행할 것인가’다. 한국이야 미국에 비해 수학을 잘하는 사람들이 넘쳐나 보이지만, 과연 이런 일들이 통계적 지식만으로 가능한 일인가?(제12장: ‘훌륭한 데이터 사이언티스트란?’ 참조) 그리고 통계 전문가가 아닌 사람들은 무슨 기준으로 이런 일을 누구에게 맡길 것인가?

만약에 조직 내에서 분석 팀을 만들 것이라면 몇 가지 고려해야 할 점들이 있다. 너무나 많은 조직들이 한두 명의 분석가, 혹은 데이터 사이언티스트를 고용하고 그들에게 가공 되지도 않은 산더미 같은 데이터를 몰아준다. 또 제대로 된 분석의 진로도 제시하지 않은 채, 그저 그들이 마법의 힘으로 데이터로 돈을 벌어줄 방법을 찾아 주리라고 막대한 기대들을 한다(미국에서도 사정이 비슷한 경우가 많다). 사실 이런 경우 일이 제대로 돌아간다면 그게 더 놀라운 일이다.

팀의 목적을
분명히 해야

내부의 분석 팀을 만들 때 첫 번째 관문은
훌륭한 후보자를 찾는 것이다.

제12장: ‘훌륭한 데이터 사이언티스트란?’에서 이미 자세히 다룬 바 있으나, 일단 제대로 된 자격요건을 만드는 것부터가 쉬운 일이 아니고, 통계 전문가가 아닌 사람이 전문가를 면접 심사한다는 것도 간단한 일이 아니다.

게다가 요즘에는 모두가 자신을 데이터 사이언티스트라고 소개하고 다니니 그 중에서 누가 전문지식을 갖추고 또 일을 제대로 할지를 알기가 쉽지 않은 것이다. 더욱이 한국에서는 이 데이터 산업이 비교적 새로운 분야라서 경력이 있는 사람들이 흔하지도 않을 것이다. 이런 일을 해본 사람들도 모든 테크닉에 능통할 수 없는 일이고, 그 중 기획 전문가와 실제로 숫자를 만지는 사람들 간의 차이도 또한 크다. 그래서 팀의 목적부터 분명히 해야 하는 것이다.

팀이 만들어진 다음에도 그것을 유지하는 것이 쉬운 일이 아니다. 일단 수학이나 통계전문가들은 단지 금전적이나 직책에 관한 보상뿐 아니라 끊임없는 도전적 과제를 원한다. 구글에서 늘 강조하는 Smart-Creatives, 즉 창조적이고 똑똑한 사람들은 더 그렇다. 그들이 판단하기

첫 번째 관문은
훌륭한 후보자를
찾는 것

에 쉽고 지루하고 반복적인 작업만 계속 주어진다면 요즘같이 분석가들에 대한 수요가 많은 시절에 가만히 그 자리에 앉아있을 리가 없다.

승진 등 관리직에 대한 제안도 먹혀 들어가지 않을 때가 많은 것이, 그들은 워낙 사람들보다는 숫자나 컴퓨터를 다루는 것을 선호하는 부류이고, 다른 사람을 관리하지 않고도 비슷한 보수를 받을 수 있으니 더 그렇다. 그래서 팀에게 적당히 흥미로우면서도 어렵고, 또 조직에게도 도움을 주는 과제를 계속 만들어주는 것 자체가 일이 된다. 그게 유지가 안 되면 유능한 사람들부터 조직을 떠날 것이다(인재의 유통이 많은 미국에서는 더 그렇다).

아무리 똑똑한 분석가라도 지속적인 멘토링이 필요한 것이,
이 데이터 사이언스라는 것 자체가 단순한 수학적 작업이 아니라

과학과 비즈니스와의 접목과정이며 유능한 분석가가 되려면 마케팅과 소비자의 관계와 사업자체의 생태계도 잘 알아야 하는 것이다.

팀 전체를 개인별로 멘토링 해주는 것은 불가능할 터이니, 도대체 자신의 월급이 어떤 과정을 통해 만들어지고 지급되는지 관심이라도 보이는 통계전문가부터 시작할 일이다. 게다가 이 글을 통해 계속 강조해 왔듯이 분석에 최적화된 환경도 필요하다. 소프트웨어, 하드웨어는 물론이고 각종 툴 셋, 가끔은 아주 비싼 도구도 필요할 것이다.

그리고 그 Analytics-Ready Environment, 즉 분석에 최적화된 환경이란 단지 기술적 요소들만으로 이뤄지는 것이 아니고, 이미 ‘데이터베이스의 디자인 개념’, ‘데이터베이스는 분석을 위해 최적화돼야’, ‘랭킹이 관건이다’ 등에서 강조한 바와 같이 데이터 자체가 고등분석에 유용하도록 재구성돼야 하는 것이니, 통계전문가 말고도 데이터를 자유자재로 다루는 사람들도 필요하다.

그런 데이터 작업은 통계작업과 근본적으로 다른 일이고, 서로의 일에 대한 이해는 필요하겠지만 반드시 분업화돼 이뤄져야 한다. 그렇지 않으면 통계전문가가 대부분의 시간을 데이터를 고치고 재구성하는데 소비할 것이다.

‘아웃소싱이
각광받고 있다’

이쯤 되면 도대체 일 년에 모델 몇 개 만들자고
내부에 팀을 만드는 것이 필요한 것인가에 대한
질문이 나올 법하다.

결론은 일이 아주 많아지고 회사 전체의 의사결정 과정이 데이터를 통해 이뤄지는 날이 오기 전까지는 이 모든 일을 시험적으로 해 차츰 그 강도를 높이는 것이 낫다는 것이다.

처음부터 모든 일을 내부에서 해결하려고 들면 일단 과도한 투자를 하게 되고, 그것을 성공으로 보이게 만들려면 그야말로 기적적인 결과가 나와야 한다. 규모의 경제(Economy of Scale)를 무시하고 모든 것을 내부에서 해결하려고 하는 것은 마치 부엌에 간장이 필요하다고 집집마다 간장 공장을 차리는 격이다.

게다가 이미 여러 차례 설명했듯이 데이터에서 의미를 찾아내는 것 자체도 공정을 따라해야 하는 것이고, 아무리 날고 긴다는 조직도 그 모든 분야에 다 정통할 수는 없다. 그래서 툴 셋이 흔해 빠진 시절에도 각종 아웃소싱(outsourcing)이 각광을 받고 있는 것이고, 통계적 분석 및 그에 관련된 데이터와 해석 작업도 예외가 아니다.

그렇다면 도대체 어디에 전화를 해야 하는가? 아웃소싱이 체계화돼 있지 않고 모든 것이 대기업 중심으로 내부에서 해결되는 한국에서는 더 난감한 일일 수 있다. 하지만 이 아웃소싱이야말로 전 세계를 무대로 진행되고 있으며, 미리 밝히자면 필자도 글로벌 아웃소싱 회사에서 일하고 있다.

그런데 문제는 거의 모든 데이터나 마케팅 회사들이 묘사적 분석(Descriptive Analytics), 예측 분석(Predictive Analytics), 규범적 분석(Prescriptive Analytics) 등의 모호한 표현들을 마구잡이로 사용하고 있다는 점이다. 심지어 미국에선 통계 전문가가 한 사람도 없는 광고회사의 웹 사이트들도 그 애널리틱스(Analytics)라는 말이 빠져 있는 경우는 없다.

애널리틱스(analytics)도 빅데이터만큼 남용되고 있는 단어라서 그 정의를 확실히 해야 할 필요가 있고, 거기에 대해서는 이 책의 제3장: '빅데이터의 핵심은 분석이며, 분석의 중심은 모델링'에서 비즈니스 인텔리전스(BI) 리포팅, 묘사적 분석, 예측 분석, 최적화 모델(Optimization Model)로 나눠 설명한 바 있다. 여기서는 그 중 고등 분석에 속하는 통계적 분석, 즉 예측 분석을 어떻게 아웃소싱 할 것인가를 중점적으로 다뤄보겠다.

미국에서와 같이 외주를 주는 것이 일상화돼 있고 데이터와 분석의 생태계가 오랫동안 진화돼 온 곳에서는 선택의 여지가 많고, 그에 따른 장·단점이 구체화돼 있다. 간단하게 정리하자면

01 개인 컨설턴트(Individual Consultants)

계약기간 동안 고용인처럼 팀에 완전히 속한다는 장점이 있지만 한시직이라서 항상 다른 프로젝트를 찾고 있으며, 그들이 떠날 때 그들의 지식도 같이 떠난다는 단점이 있다. 계약 조건에 따라 시간이 많이 지나면 전체적 비용이 올라갈 수도 있다.

02 분석전문외주회사(Standalone Analytical Service Provider)

그야말로 분석전문회사이며, 다른 일은 하지 않는다. 축적된 지식과 경험이 많고 각종 산업별 노하우도 풍부할 수 있다. 다만 규모가 작은 경우가 많고 그래서 과도한 분량의 일을

말으려고 하는 경향이 있을 수 있다. 데이터에 관한 한 어떤 종류의 데이터도 가리지 않고 다루지만, 미국과 같이 외부데이터가 흔한 나라에서는 자신들만의 데이터가 없다는 것이 단점이 될 수도 있다.

03 데이터 서비스 제공회사(Database Service Providers)

미국에서는 데이터를 수집하고 거래하는 데이터 브로커(Data Broker)나 데이터 컴파일러(Data Compiler)들이 많은데, 예외 없이 그런 회사들은 분석팀을 운영하고 있다. 일단 그것은 그들의 데이터에 있는 빈 곳들을 통계적 모델로 채워 넣기 위해서인데, 그런 과정을 통해 축적된 지식도 상당하다.

그들과 상대하면 분석뿐 아니라 다른 데이터 가공 등 많은 공정을 한 곳에서 해결할 수 있다는 장점이 있지만, 회사 규모들이 커서 민첩하고 기민하지는 못할 수도 있다. 그리고 수익을 올리기 위해 자신들이 운영 관리하는 데이터를 중점적으로 사용하려는 경향이 있다.

04 광고회사/광고 에이전시(Direct Marketing Agencies)

광고회사들 중 특히 1-to-1 마케팅 전문으로 하는 회사들은 분석에 많은 투자를 하고 있고, 그들의 장점은 전략적 차원의 시각을 갖고 있다는 것이다. 특히 많은 분석회사들이 다루지 않는 Optimization Model, 즉 광고비 최적화 모델 등이 그들이 강점이며, 전략 수립과 함께 세분화(segmentation)를 통한 광고의 개인화와 최적화 등 모든 과정을 다 관리한다.

하지만 모든 광고회사들이 그런 것은 아니고, 구멍가게 수준의 분석팀을 갖고 그 능력을 과장하는 경우도 많다. Predictive Analytics는 다루지 않는다고 하는 경우도 많으며 그들도 그런 일은 아웃소싱을 한다(오히려 그게 나을 수도 있다).

이렇듯이 딱 누구와 어떻게 일해야 한다는 정답은 없다. 그래서 사업 목적에 따라 그 전문가나 아웃소싱회사를 찾아야 하는 것인데, 여기서 중요한 것은 그 고려대상이 단지 수학적 능력이 아니라는 점이다.

너무나 많은 경우 단지 가장 싼 값에 입찰하는 회사에 외주를 주는 경우가 많은데, 그러다가 낭패를 보는 수가 허다하다. 그래서 '고등 분석을 아웃소싱 할 때 고려해야 할 열 가지 요소들'을 여기서 소개하고자 한다.

1. 컨설팅 능력 (Consulting Capabilities)

‘훌륭한 데이터 사이언티스트란?’ 등 이전의 챕터에서 강조했듯이 데이터 분석을 하는 사람들에게 가장 요구되는 점은

마케팅 등 비즈니스의 세계와 테크놀로지 세계의 중간에서 통역을 할 수 있는 능력이다. 그들은 비즈니스의 목적과 마케팅에서 요구되는 점들 두루 이해하고, 적합한 처방을 내리며,

그러한 목적들은 수학적 표현으로 변환하고, 주어진 데이터를 최대한 이용해 타깃을 찾아내야 한다.

만약 분석을 맡은 사람들이 전략적인 시각을 갖추지 못하면 통계적 지식만으로는 목표를 달성할 수 없다. 그리고 그러한 비즈니스 전략과 목표는 속에 있는 산업, 마케팅 채널, 그와 관련된 성공지표에 따라 달라지기 때문에 훌륭한 컨설턴트라면 일단 사업에 관한 질문부터 하고 고객의 말에 귀 기울일 줄 알아야 한다. 반면 수준이 떨어지는 분석가들은 자신들의 톤 셋과 분석방법에 마케터들의 목적을 끼워 맞추려 든다.

마케팅의 목적을 자세한 스텝들로 변환하는 것도 기술이다. 제대로 된 분석 파트너라면 구체적인 모델이 비즈니스에 어떤 영향을 미치는지를 고려해 단지 분석의 단계를 넘어 전체적인 데이터 로드맵을 짤 수 있어야 한다.

그 후 큰 목표를 전체적 비용까지 고려하면서 작은 단계로 나눠 일을 진행해야 하며, 거기에는 어떤 종류의 모델이 몇 개가 필요인지, 모델단계에 들어가기 전의 데이터 변환과정에 관한 설계, 모델이 만들어진 다음에 알고리즘을 접근 가능한 모든 대상에 어떻게 적용할 것인지에 대한 구체적 방안, 외부 데이터는 어떤 경로로 들여올 것인지 등이 포함된다.

모델링에서 타깃의 정의는 다년간의 경험과 산업지식이 필요한 작업이다. 만약에 그 정의가 틀리거나 적합하지 않으면 아무리 수학적으로 훌륭한 모델도 좋은 결과를 낼 수 없다. 타깃이 엉뚱한 곳에 걸려있으면 아무리 총을 잘 쏘도 소용이 없는 것과 같은 이치다.

또 다른 중요한 자질은 혼란스럽고 정리되지 않은(unstructured) 데이터 환경에서 유용한 정보를 만들어내는 능력이다. 모델링은 단지 완벽한 데이터베이스로만 가능한 것이 아니고 주어진 데이터를 최대한 활용하는 작업이다. 모델 경시대회 등에서 좋은 성적을 거두는 회사들은 주어진 데이터를 창조적으로 사용하는 참가자들이지, 단지 통계에만 도통한 사람들이 아니다.

컨설턴트로서 접근방식이 중요한 것은 모델이나 통계적 분석이 자체적으로 존재할 수 없는 것이고 그것이 전체적 마케팅 엔진의 일부여야 하기 때문이다. 그래서 모든 활동을 자신들이 창출한 모델 공식을 중심으로 하려는 사람들을 조심해야 한다. 더욱이 전체적 그림을 이해하고 있는 분석가라면 다음 단계가 무엇이어야 하는지도 알려줄 것이다. 마케팅은 연속적인 활동이지 한시적 작업이 아니고, 거기에 따른 분석도 마찬가지다.

그렇다면 어떻게 컨설팅 자질도 있는 파트너를 찾을 것인가? 아래의 질문부터 하기를 권한다.

- 당신의 말과 고민에 진정으로 귀 기울이는가?
- 당신의 목적을 그들만의 언어로 재해석해 반복할 수 있는가?
- 장·단기 목표를 모두 커버하는 데이터 로드맵을 갖추고 있는가?
- 당신의 말을 교정할 수 있을 정도로 자신감이 있는가?
- 비수학적인 마케팅 요소들을 이해하고 있는가?
- 단지 이론만이 아닌 실천경험이 있는가?

2. 데이터 처리 (Data Processing) 능력

데이터 프로세싱이란 말을 여태 쓰냐고 반문하는 사람들도 있지만, 데이터를 제대로 처리하고 변환하는 것은 모델링이 시작되기 전에 거쳐야 할 필수적인 단계이다.

영어 표현으로 “garbage-in, garbage-out”이란 말이 있듯이 제대로 처리되고 가공되지 않은 데이터로 고등 분석을 할 수 없는 일인데, 불행히도 현대의 많은 데이터베이스들은 ‘전혀’ 분석에 적합하지 않은 경우가 허다하다.

여하간 누군가가 그런 저저분한 데이터를 청소해야 하는 것이고, 워낙 이 데이터 비즈니스에서는 마지막으로 데이터를 만드는 사람이 그 이전의 저질러진 모든 실수도 다 고쳐야 하는 법이다. 물론 공평하지는 않은 일이지만, 그래서 분석을 맡을 회사나 사람들의 통계적 지식과 거창한 프레젠테이션뿐 아니라 크고 저저분한 데이터도 처리할 수 있는 능력도 고려해야 하는 것이다.

그 쓰레기 처리능력이란 데이터 변환, 수정/보정, 카테고리화/태깅(tagging), 데이터 집적(summarization)과 변수창출(variable creation) 등을 포함하며, 이전에 집중적으로 다룬 숫자적 변수와 자유형 데이터를 다루는 능력을 아우른다(제7장: ‘RFM 데이터를 넘어서’와 제8장: ‘자유형(freeform) 데이터는 결코 공짜가 아니다’ 참조).

이런 작업들은 겉으로 보기에 화려하지 않지만 일관적인 형태의 데이터는 고등분석의 제1 성공요소이다. 그래서 필자가 강조해온 ‘Analytics-Ready’ 환경이 이뤄져 있지 않다면 분석 파트너라도 주어진 데이터를 분석을 위해 최적화할 수 있어야 한다. 전체적인 프로젝트로 볼 때 가장 큰 실패요인들은 모델이 아니라 그 모델링의 ‘before’와 ‘after’과정에서 생기는 경우가 대부분이다. 그래서 후보자들에게 이런 질문들을 던져봐야 한다.

- 데이터 변환, 수정, 카테고리화, 집적화 등의 작업을 하기는 하는가?
- 얼마나 큰 데이터까지 다뤄봤으며, 얼마나 많은 변수들을 동시에 다뤄봤는가?
- 얼마만큼의 자유형(freeform) 데이터가 부담스러운가?
- 과거에 새로 창출한 변수들의 샘플을 보여줄 수 있는가?

3. 산업지식 (Industry Knowledge)

약간의 과장을 보태자면 산업지식이 통계적 지식과 모델링 노하우보다도 더 중요하다고 말할 수도 있는데,

그것은 특정 산업에서의 경험이 없는 경우 비즈니스적 니앙스가 제대로 전달이 되지 않는 경우가 생기기 때문이며, 심한 경우 아예 대화가 이어지지 않을 수도 있다. 농담이 아니라 필자는 (수확은 도사급인) 로켓 전문가가 신용카드 캠페인을 아주 망가뜨려 놓은 경우도 본 적이 있다.

산업지식이 중요한 첫째 이유는 성공을 가능하는 잣대부터 다르기 때문이다. 몇 가지 예를 들자면 금융, 신용카드, 보험, 투자, 항공, 숙박, 엔터테인먼트, 포장상품(packaged goods), 온라인과 오프라인 유통, 카탈로그, 출판, 텔레콤(telecommunication), 자선단체, 정치단체 등이 다른 형태의 분석과 모델을 요구하는 것이고, 그것은 그들의 비즈니스 모델과 그들이 고객을 상대하고 거래하는 수단과 채널이 전부 다르기 때문이다.

모델이나 혹은 데이터베이스 자체를 구축할 때 ‘돈 받기 전에 물건을 건네주는 사업’과 ‘지불과 동시에 물건이 전해지는 사업’을 위한 그 구조의 차이는 엄청나다. 그 두 개의 경우 간단한 ‘날짜’도 전혀 다르게 기록되고 해석돼야 하는 것이 유통에서는 ‘최근의 거래’가 높은 점수를 받는 경우가 많은 반면, 구독이나 장기계약이 필요한 사업에서는 꼭 그렇지 않기 때문이다. 이 예는 단순한 날짜에 불과하고, 액수나 기타 변수로 파고들어가자면 차이점은 한두 가지가 아니다.

더욱이 이런 사업들은 그 사업 대상에 따라 B2B와 B2C로 나뉘는데, 그 둘의 차이는 다뤄야 하는 변수의 차이로부터 시작해 엄청나다. 거기에 회사 내 부서에 따른 ROI Metrics의 차이도 고려돼야 하며, 심지어는 쓰는 단어들과 기업문화의 차이도 분석에 영향을 미친다.

하지만 자신을 컨설턴트로 부르는 사람들은 미팅이 시작되는 순간부터 그 분야의 전문가로 활동해야 하는 법이다. 그래서 후보자를 고를 때 (계약상 제약이 있는 경우가 많으니) 허용되는 범위 안에서 과거에 상대해본 기업과 부서에 관해 알아봐야 한다. 그게 여의치 않다면 산업에 상관없이 이미 꾸며진 방법만을 고집하는 분석가를 조심할 일이다.

4. 다루는 모델 종류 (Types of Models Supported)

특정방법에 관한 말이 나온 김에 덧붙이자면, 아웃소싱 파트너가 다루는 모델의 종류도 알아봐야 한다.

물론 모든 방법을 다 쓰는 분석가는 없지만(그들의 프레젠테이션은 어떨지 모르지만), 한 가지 방법만 고집하면 곤란하다(미국식 표현으로 ‘One-trick Pony’라고 한다).

이것은 약간은 전문적 분야로 파고들어가야 하므로 어려울 수도 있는 부분이다. 동시에 마케터가 목적을 분명히 제시하는 대신 특정 방법을 멋대로 지정하는 것도 바람직하지 않다(제15장: ‘IT와 마케팅의 다소 결끄러운 관계’ 참조). 그래서 구체적인 목적의 예를 나열해보자면

- 마케팅 대상에 랭킹을 매겨 고르기(Prospect ranking and selection)
- 영업대상의 스코어(Lead scoring)
- 부가 판매(Cross-sell/up-sell)
- 메시징(Messaging)을 위한 세그멘테이션(Segmentation)
- 계약 및 구독 중지(Attrition) 시점 예측
- 고객가치(Lifetime-value) 예측
- 미디어와 채널 최적화(Media and channel optimization)

- 신상품/패키지 개발(New product and package development)
- 사기예방(Fraud detection)
- 기타 등등

과거에 성공적으로 아웃소싱 파트너와 일해 본 경험이 없다면 신경망(Neural-Net), 체이드(C HAD), 클러스터 분석(Cluster Analysis), 다중회귀(Multiple Regression), 판별함수 분석(Di scriminant Function Analysis) 등의 기술적 용어를 나열하는 것은 곤란하다. 그것은 마치 의사가 검진하기도 전에 환자가 나서서 특정 약을 요구하는 것과 비슷한 일이다.

이는 유행어를 반복하는 대신 비즈니스의 목적을 분명히 설명하라는 것이다. 목적에 관한 토론을 마친 이후에 파트너가 방법론을 제시하는 것이 올바른 순서다. 다만 그 후보자가 한두 가지 특정 방법만을 고집하는지 살필 일이다.

5. 일처리속도 (Speed of Execution)

현대 마케팅에서는 속도가 왕이다.
스피드가 승리하고 스피드가 존경을 얻는 법이다.

하지만 모델링과 분석에 들어서면 아웃소싱 회사마다 작업속도에 엄청난 차이가 있다는 것을 발견하게 된다. 물론 그 주된 이유는 그들이 어느 정도로 망가진 데이터를 다루야 할 지 몰라서 미리 많은 시간을 요구하기 때문이기도 하다.

언급한 바와 같이 모델링의 전, 후가 가장 시간과 노력을 많이 필요로 하는 부분이고, 데이터가 아주 지저분하다면 시간을 하염없이 잡아먹을 수도 있다. 모델이 만들어진 후 적용단계도 많은 주의를 필요로 하고, 거기에서 시간이 많이 흘러가는 경우도 흔하다. 그래서 어느 정도 시간이 걸릴 지 예측할 때 모델링의 전과 후 단계를 나눠서 견적해달라고 해야 한다.

순전히 수학적인 부분도 그 과제의 복잡함에 따라 공정기간에 많은 편차가 있을 수 있다. 물론 기본적인 변수를 이용한 간단한 클로닝 모델(Cloning Model)은 다양한 채널에서 비롯된 많은 양의 거래와 이벤트 데이터를 사용해야 하는 모델에 비해 시간과 노력이 훨씬 덜 들기 마련이다.

만약 시간차 변수(time-series variables)들이 더해지면 일이 더 복잡해지며, 클러스터링(c lustering) 같은 작업은 타깃이 분명히 정해진 회귀(regression) 모델보다 일반적으로 많은 시간이 소요된다. 게다가 한 가지 사업목적 위해 여러 가지 모델이 필요하다면 당연히 더 많은 시간이 필요할 것이다.

그런데 재미있는 것은 마케터들이 프레젠테이션을 만들 때처럼 많은 분석가들은 모델링을 원하는 대로 끝마치는 것이 아니라 단지 주어진 시간을 다 소모해버리는 경우가 많다는 점이다. 그 공

통점은 모델이나 프레젠테이션이나 시간이 한없이 주어진다면 끊임없이 수정을 할 수 있는 것이지만, 언젠가는 마감일에 쫓겨 일을 마쳐야만 하게 된다는 것이다.

하지만 지난 수십 년간 이뤄진 많은 자동화와 툴 셋의 발전으로 - 근본적인 수학적 개념에는 큰 변화가 없지만 - 모델 작업의 시간도 비약적으로 줄어들었다. 1960~1970년대에 모델링이 마케팅에 처음으로 적용된 이후로 정말로 많은 변화가 있었으니, 아직도 1980년대식의 공정기간을 요구한다면 그건 관련한 일이다. 물론 '을'의 입장에서는 지키지 못할 약속을 미리 피하고 싶은 것이지만 말이다.

여기서 고려해야 할 점은 그 소요시간도 약간의 정확성을 포기하고 더 간단한 모델을 요구하는 식으로 조정과 타협이 가능하다는 것이다. 만약 98%의 정확도를 성취하는 데 일주일만 걸리고 95%의 정확도는 하루 만에 가능하다면 과연 어떤 선택을 할 것인가? 그것이야말로 비즈니스적 결정인 것이다.

그렇다면 일반적 가이드라인은 무엇인가? 물론 많은 요소들이 영향을 미치겠지만 아웃소싱이 일 변화돼 있는 미국의 예를 들어보는 것이 도움이 될 수 있겠다.

- Pre-modeling Processing
 - Data Conversions: 반나절 - 수주일
 - Data Append/Enhancement: 하룻밤 - 이틀
 - Data Edit & Summarization: 주어진 데이터의 상태에 따라 크게 다름
- Modeling: 반나절 - 수주일
 - 모델의 종류, 방법, 모델 수, 복잡성에 따라 크게 다름
- Scoring: 반나절 - 일주일
 - 적용돼야 할 데이터베이스의 크기와 개체 수, 그리고 상태에 따라 다름

물론 이것은 일반적인 추정이라 그 편차가 크지만, 간단한 모델을 만드는데 한 달씩 요구하는 파트너를 주의할 일이다. 그런 경우 그들이 실전경험이 없어서 일을 제대로 할 줄 모르거나, 더 나쁜 경우 비즈니스의 속성을 이해하지 못하고 순수한 수학적 완성도를 고집하는 경우일 수가 있다.

6. 가격 (Pricing Structure)

많은 마케터들은 오로지 가격만을 보고
아웃소싱 파트너를 정하는 경우가 많은데,
그건 정말 바람직하지 않다.

물론 필자는 늘 '을'의 입장에서 경력을 쌓아온 사람이지만, 모든 사업에서의 수익성의 중요성을 이해 못하는 것은 아니다. 하지만 싼 값만 보고 후보자를 고르는 것은 무모한 일이라서

이 긴 리스트를 소개하는 것이고, 예산이 아무리 중요해도 그런 식으로 대화를 시작해서는 안 된다.

비즈니스의 목적과 분석의 방법에 대한 토론 이후에 가격을 따져야 하는 것인데, 중요한 것은 그 가격이 무엇을 포함하고 있는가이다. 이 애널리틱스란 끊임없는 순환과정이고, 각 단계가 다음 단계로 가는 디딤돌이기 때문이다. 게다가 요즘은 ‘공짜’로 모델을 만들어준다는 회사들도 있는데, 세상에 일거저 얻는 것이란 없는 것이고 반드시 가격이 붙어있는 단계가 나중에 등장하게 마련이다. 가격에 관한 한 중요 요소들을 나열해 보자면

- 여러 개의 모델이 필요한 경우 첫 모델 이후 가격할인이 있는가?
- 맞춤형 모델이 비싸다면 이미 만들어진 모델 중 카테고리별로 사용 가능한 것을 싸게 제공할 수 있는가?
- 특정 기업이나 부서별 데이터가 사용되지 않는다면 할인이 가능한가?
- 방법론에 기초한 가격차별이 있다면 어떤 옵션들이 있는가?
- 클러스터링이나 세그멘테이션이 일반 모델에 비해 얼마나 높게 가격이 책정되었나?

다시 강조하지만 너무나 많은 모델 외적 요소들이 가격과 소요시간에 영향을 미치므로, 모델링의 전과 후 작업을 따로 떼어 견적을 요구하는 것이 바람직하다. 모델 적용도 반복적으로 하다 보면 그 비용이 만만치 않게 되는 경우가 있으니, 그런 반복 작업에 할인을 미리 요구하는 것도 고려해야 한다. 때로는 이 모든 과정을 패키지로 만드는 사업체들도 있으니 가격을 비교할 때 주의해야 한다.

7. 문서화 (Documentation)

미리 만들어진 모델을 할인가에 구매하는 것이 아니라 맞춤형 모델을 주문하는 경우 그 주문자가 그 모델 공식(algorithm)을 소유하게 되는 법이다.

그래서 그 공식과 그 모델을 만드는 과정, 가정들, 사용된 변수들의 변환과정 등이 문서들 통해 전해져야 한다. 바람직한 모델 문서는 최소한 다음과 같은 요소들을 갖춰야 한다.

- 타깃과 비교대상에 대한 정의(Target and Comparison Universe Definition)
무엇이 타깃 변수, 즉 Dependent Variable이었으며 어떻게 정의됐는가? 타깃 못지않게 중요한 비교대상은 또 어떻게 정의됐나? 그 두 집단에 적용된 선택 룰은 무엇인가? 이것은 모델의 구조보다도 더 중요할 수 있는 요소들이다.
- 변수 리스트(List of Variables)
모델을 묘사하는 변수들, 즉 Independent Variables는 무엇인가? 그것들은 어떻게 선택되고(selected), 변환되고(transformed), 또 구분 지어졌나(binned)? 그 변수들의 출처는 무엇인가? 이 모델 변수들은 모델의 성격을 대변하는 것이고, 상식적으로 이해가 가능해야 한다.
- 모델 공식(Model Algorithm)
모델에 사용된 변수들의 가중치를 포함한 공식 자체가 공유돼야 한다.

- 증가치 도표(Gains Chart)

모델이 점수 그룹에 따라 얼마만큼의 증가(gain)를 가져오는지 보여주며(예를 들어 ‘최상위 그룹의 반응률이 전체에 비해 320%가 높다’), 모델점수가 내려가면서 얼마나 그 증가치가 줄어드는지도 도식적으로 설명해야 한다. 동시에 개발용 샘플(Development Sample)과 검증용 샘플(Validation Sample)도 비교하는 것이 바람직하다.

맞춤형 모델일 경우 정식으로 모델 프레젠테이션을 하는 것도 관행이며, 주문자가 직접 모델을 데이터베이스에 적용할 경우 모델 공식을 실제 사용 가능한 프로그램 언어로 전달해야 한다. 그리고 만약에 여러 종류의 거래기록 등 주문자가 파일을 제공한 경우, 어떤 과정으로 그 파일이 가공됐고 각 스텝마다 얼마만큼의 기록들이 각종 이유로 제거됐는지도 보여주는 리포트(Waterfall Report)도 필요하다.

모델이 아웃소싱 회사에서 적용됐을 경우, 실제로 모델점수별로 분포도도 제공돼야 한다. 많은 회사들이 요구하지 않은 리포트를 알아서 만들어주지 않는 경우가 있으므로 계약 당시 이런 요구들을 명시하면 도움이 된다.

8. 실제 스코어 검증 (Scoring Validation)

모델의 개발이 끝나고 프레젠테이션까지 마쳤어도 업무가 끝난 것이 아니다.

모델은 나중에 사용될 데이터베이스에 적용돼야 진가를 발휘하는 것이고, 그 스코어링(scoring)이 끝날 때까지 긴장을 늦춰선 안 된다. 실제로 많은 오류는 이 단계에서 발생하는데, 그 이유는 모델 개발용 샘플은 아무리 커봐야 몇 십만 줄을 넘지 않지만, 메인 데이터베이스에는 수백만의 대상이 있을 수도 있기 때문이다.

일례로 미국 전체의 가구 수는 1억 4천만 정도인데, 그 모든 가구에 점수를 매기는 것은 간단한 일이 아니며, 성공적으로 프로그램이 돌아갔더라도 그 모델 점수의 분포도가 개발 샘플이나 검증 샘플의 그것들과 다르게 나오는 경우도 허다하다.

여기서 그 점수의 적용이 꼭 개발자의 일이 아닐 수도 있지만, 오류가 나올 경우를 대비해 그들이 질문이 대답할 수 있도록 대기상태에 두는 것이 바람직하다. 그리고 그런 식의 일관되지 않은 분포는 개발 샘플에 비해서만 아니라 시간이 지나면서 발생할 수도 있다. 예방차원에서 모델을 과거의 데이터베이스에 적용해 미리 살펴볼 수도 있다.

이는 많은 오류가 모델 개발이 끝난 후에 생길 수 있으므로 주의를 기울여야 한다는 것이다. 실제로 모델이 잘못됐다고 불만이 들어와 살펴보면 모델에는 오류가 없고 적용과정에서 문제가 생긴 경우가 대부분이다.

9. 사용 후 분석 (Backend Analysis)

제다가 플랫폼과 사용 프로그램 언어의 차이로 기계끼리도 오해를 하는 경우도 있다. 혹은 언어적 오류는 없었지만 정보의 부재로 일관되지 않은 결과가 발생할 수도 있다(제9장: '정보의 부재에도 의미가 있다' 참조). 이런 경우 모델을 직접 개발한 사람들이 오류의 원인을 찾아내는 데 가장 유리한 통찰력을 갖고 있을 수 있으므로 질문에 즉각 대답할 수 있도록 계속적인 교류가 있어야 한다.

마케팅에서 훌륭한 분석이란 결과가 좋았든지 나빴든지 과거의 결과를 통해 새로운 것을 배워나가는 과정이다.

그것을 흔히 'Closed-loop Marketing', 즉 '연결고리가 끊어지지 않은 마케팅'이라고 부르는데, 불행히도 많은 마케터들은 이런 과정을 따르지 않는다.

제대로 된 분석전문회사라면 당연히 모델 사용 후의 일까지도 염려해줘야 마땅하다. 물론 그런 일은 모델 작업과는 분리돼 고려되겠지만, 요는 그러한 서비스의 제공여부도 중요한 고려 대상이라는 것이다.

사실 요즘처럼 많은 채널이 동시에 사용되는 옴니채널(Omni-Channel) 시대에는 어떤 요소와 채널이 좋은 결과를 가져왔는지를 알아내기가 쉬운 일이 아니고, 그 과정 자체가 여러 가지들과 더 나아가 전용 모델을 요구하게 될 수도 있다.

그리고 맞춤형 사용 후 분석은 그 자체가 고비용 프로젝트가 될 수도 있으므로 이미 개발된 리포팅 틀이 있는지도 살펴보는 것이 좋다. 그런 도구들은 사용자가 요구하는 모든 ROI 지수들이 포함돼있지 않을 수도 있지만, 기본적인 변수들, 즉 이메일 캠페인 분석이라면 개봉률(Open Rate), 클릭률(Click-Through Rate), 전환율(Conversion Rate)이나 거기에 따른 액수나 거래상황 지표들이 포함돼있으면 바람직하다.

그러한 지표들을 데이터의 출처, 캠페인, 시간진행, 모델그룹, 마케팅 메시지 버전, 타깃 정의, 채널, 배너 애드 파트너, 출판사, 키워드 등으로 전부 따지기 시작하면 리포트 자체가 너무 커져버리는 수가 있으니 거기에 대한 조언도 해줄 수 있는 파트너가 우선돼야 한다. 그리고 ROI(Return On Investment)가 기업 내에서 중요한 지표라면 마케팅 비용에 관한 자세한 정보도 공유할 수 있을만한 믿을 수 있는 파트너가 필요하다.

10. 계속적 지원 (Ongoing Support)

세상은 늘 변하고 데이터의 상관관계도 고정된 것이 아니기 때문에 모델의 수명도 한시적일 수밖에 없다.

사업이 계절을 타는 경우, 혹은 비즈니스 모델이나 전략 자체가 바뀌는 경우에 더 그렇다. 데이터에 대한 접근성의 변화나 데이터 질의 비일관성도 더욱 모델의 수명을 단축시킨다. ‘모든 변수가 일정하면’이란 가정은 교과서에서나 나오는 말이니 마케터들은 모델과 룰의 정기적 리뷰를 계획해야 한다.

뭔가가 잘못되고 있다는 신호는 모델의 효과가 떨어지는 것에서부터 보이기 시작한다. 의심이 들면 지체 없이 개발자와 상담해 모델을 재구성하든지 과감히 새로 개발하는 것이 필요하다. 보통 한 분기에 한 번쯤 리뷰를 하는 것이 바람직하지만 여의치 않다면 6개월이나 1년에 한 번은 하되, 1년은 넘기지 말아야 한다. 보통 기존 모델의 재개발은 가격이 더 저렴할 수 있으니 계약단계에서 확인해두면 좋다.

이 리스트는 의심의 여지없이 길지만, 아웃소싱이란 장기적 안목으로 봐야 하는 것이기 때문에 애초에 훌륭한하고 적합한 파트너를 찾는 것이 매우 중요하다. 그리고 필자가 여기에서 기술적인 면에 대해 전혀 언급을 하지 않아 의아해할 독자들도 있겠다.

그것은 첫째, 많은 마케터나 사용자들에게 기술적인 질문을 할 지식이 부족하기 때문이며, 둘째, 수학적이고 기술적인 요소들에서 발생하는 결과의 차이는 이 칼럼에서 나열한 요소들에서 비롯된 차이보다 일반적으로 훨씬 더 적기 때문이다. 즉, 모델 경시대회에서의 일등과 꼴찌와의 차이는 다른 비즈니스적 요소들에 비해 덜 중요하다는 말이고, 적절한 절차를 거친 모델은 아무리 수학적으로 완벽하지 않더라도 모델을 아예 사용하지 않는 것에 비해 더 좋은 결과를 가져온다는 것이다.

한국에서는 아직 아웃소싱이 활성화되지 않았을지 몰라도, 계약의 모양새를 떠나 좋은 파트너를 찾는 데 이 글이 도움이 됐으면 한다. 데이터가 넘쳐나는 시대에 분석의 활용은 더 이상 옵션이 아니고 필수사항이다. 고로 자체 내에서 해결이 어렵다면 지체 없이 도움을 청해야 한다.