

IDG Tech Report

데이터 속에서 가치를 창출하는 데이터 과학자

아는 만큼 보인다. 현재 통계 전문가나 컴퓨터 전문가들은 기업의 비즈니스를 잘 알지 못한다. 반면 해당 비즈니스를 잘 아는 전문가는 데이터를 분석할만한 수학이나 컴퓨팅 지식이 없다. 데이터 과학자는 대량의 데이터 속에서 비즈니스 가치를 창출해 내는 전문기를 일컫는다. 데이터 과학자는 말보다는 데이터 가치 창출 전문가라고 불리는 편이 이 직업의 의미 전달에 보탬이 될 듯하다. 빅 데이터 시대를 주도할 데이터 과학자에 대한 정의와 요건, 그리고 이 직업의 현황과 미래에 대해 살펴보자.

- ▣ 빅 데이터를 다루는 자, 데이터 과학자
- ▣ BI 문제의 근원적 극복 방안, 바로 '사람'
- ▣ 데이터마이닝+통계+분석+통찰력 = 가치 창출
- ▣ 부정적으로 바라본 데이터 과학자, 5D 직종 가능성

무단 전재 재배포 금지

본 PDF 문서는 IDG Korea의 프리미엄 회원에게 제공하는 문서로, 저작권법의 보호를 받습니다.
IDG Korea의 허락 없이 PDF 문서를 온라인 사이트 등에 무단 게재, 전재하거나 유포할 수 없습니다.

데이터 속에서 가치를 창출하는 데이터 과학자

편집부 | ITWorld

□ 국드라마 범죄과학수사대 CSI에서 길 그리섬 반장은 감식곤충학 전문으로, 시체에 붙어있는 파리에서도 단서를 찾아내 사건을 해결한다. 데이터 과학자 얘기하면서 생뚱맞게 미드 얘기를 꺼낸 이유는 과학수사관이 하는 일이 앞으로 설명할 데이터 과학자가 하는 일과 비슷하기 때문이다. 이 드라마에서 과학수사관들은 별로 도움이 안될 듯한 파편화된 사건 조각들을 모아 사건 해결에 결정적인 단서와 증거를 발견한다. 이처럼 데이터 과학자들이 하는 일이 바로 수없이 많은 데이터 속에서 새로운 비즈니스 가치를 만들어내는 것이다.

데이터 속에서 가치를 창출한다

우선 빅 데이터 관련 전문가들이 데이터 과학자에 대해 내린 정의를 살펴보자.

EMC 분석 담당 부사장 스티브 힐리온은 “컴퓨터 분석 기술과 모델링 기술을 사용해 방대한 양의 정보로부터 통찰력 있는 결론을 얻어내는 사람이 데이터 과학자”라고 말했다.

IBM 빅 데이터 제품 담당 부사장 앤줄 밤브리는 “데이터 과학자란 실제 비즈니스 결과를 이끌어내는 활동 간의 트렌드를 발견하기 위해 기업의 대용량 데이터를 분석하는 사람”이라고 정의했다.

아마존의 데이터 마이닝 전문가 존 라우저가 정의한 데이터 과학자는, 공학과 응용 수학, 특히 통계학 지식을 배경으로 경제학, 소프트웨어 공학 등의 학문적 지식을 더해 데이터 분석에서 유의미한 결과를 도출해 내는 사람이다.

이런 추상적인 정의로는 데이터 과학자를 제대로 이해하기가 힘들다. 앞서 설명한 대로 데이터 과학자

는 데이터를 분석해 새로운 가치를 창출해내는 사람으로 받아들이자. 이제 데이터 과학자가 어떻게 나오게 됐으며, 무엇을 하는 사람이며, 어떤 역량이 요구되는지, 그리고 어떤 곳에서 일하는지 알아가면서 데이터 과학자의 개념을 구체화해 보자.

빅 데이터를 다루는 자, 데이터 과학자

데이터 과학자의 정의에 맞춰 본다면 역사적으로 이런 업무를 하고, 업적을 남긴 이들은 상당히 많다. 존 라우저는 최초의 데이터 과학자로 18세기 수학자이자 천문학자인 토비어스 마이어를 꼽았다. 토비어스 마이어는 1750년에 9개의 상수를 3개의 방정식에 적용해 27개의 공식으로 지구의 위도 측정 오차를 최소화했다. 라우저는 “토비어스 마이어는 데이터가 많을수록 더 정확하게 측정할 수 있다는 것을 입증했기 때문에 최초의 데이터 과학자로 볼 수 있다”고 말했다.

이처럼 기준에도 이런 역량을 가진 이들은 존재했다. 또한 이전에도 산업계마다 해당 데이터 분석 전문가도, 비즈니스 전문가도 수없이 많이 종사하고 있다. 그런데 왜 데이터 과학자일까?

데이터 과학자라는 새로운 직업이 필요하게 된 이유 가운데 하나는 바로 빅 데이터(Big Data)라는 시대적 흐름 때문이다. 빅 데이터는 단순히 데이터 양의 증가가 아니라 데이터의 형식, 입출력 속도 등을 함께 아우르는 의미로, 기존 방법으로는 수집, 저장, 검색, 분석 등이 어려운 데이터를 총칭해서 일컫는 단어다.

지난 20년 간 인터넷과 컴퓨팅의 발전, 모바일 기기와 센서들의 진화로 인해 데이터, 텍스트 및 문서,

통화 기록, 웹 클릭스트림(Web Clickstream) 등이 데이터의 폭증을 불러왔다.

예를 들면, 웹 로그, RFID, 센서 네트워크, 소셜 네트워크, 인터넷 텍스트/ 문서, 인터넷 검색 인덱싱, 음성 통화 상세 기록, 천문학/ 대기과학/ 유전학/ 생화학/ 생물학 등 학문적 연구 기록, 군사 경계 기록, 의료 기록, 사진 목록, 동영상 목록, 전자 상거래의 행적 등이 빅 데이터에 해당한다.

이런 빅 데이터 속에 소비자의 행동 양식이나 보안 위협, 잠재적인 시스템 문제 등의 귀중한 정보가 묻혀있다는 것이다. 하지만 18개월마다 두 배로 증가하는 테라바이트 단위의 데이터를 어디서부터 어떻게 처리해야 할 것인가? 여기서부터 데이터 과학자의 필요성이 시작됐다.

BI 문제의 근원적 극복 방안, 바로 '사람'

한편 데이터를 이용해 고객의 행동 양식을 파악해 부가가치를 창출하려는 기업들의 노력은 이미 20년 이 훌쩍 넘었다. 또한 지난 십 수년간 기업들은 고객의 행동과 패턴에서 부가가치를 창출하기 위해 CRM, BI, DW(Data Warehouse) 등에 IT 예산을 쏟아부었지만, 원하는 결과를 얻지 못했다.

이에 대한 원인을 분석한 결과, 근원적인 문제는 데이터 분석 결과를 읽고 이를 해석하고 가치를 창출



할 수 있는 능력을 지닌 사람, 즉 데이터 과학자가 없었다는 것에 초점이 맞춰졌다.

포레스터 리서치 수석 애널리스트 브라이언 흉킨스는 “최근 기업들은 방대한 양의 정보를 잘 사용하기 위해서는 기존의 데이터 마이닝이나 BI와 같은 도구만으로는 부족하다는 점을 깨달았다”고 말했다. 흉킨스는 “기업은 항상 경쟁업체보다 더 많은 정보를 얻고자 한다. 이를 바탕으로 예측 기능이 있는 분석 도구를 사용하면 데이터에 대한 이해가 저절로 따라올 것이라고 생각하는 기업도 있는데, 이는 잘못된 생각이다. 데이터 과학자들이 있어야 통계 모델을 만들고 사용할 수 있으며, 그 결과를 경영진들에게 이해시킬 수도 있다”고 설명했다.

스티브 힐리온 역시 빅 데이터를 처리하기 위해서는 도구만으로는 부족하다는 점에 동의하면서 기존 BI 툴의 한계와 병렬 컴퓨팅의 발달, 그리고 정교한 데이터 모델링 도구가 합쳐져 데이터 과학자가 등장한 것이라고 주장했다.

일반적으로 BI는 기존 데이터의 합산과 분석을 거쳐 과거의 특정 지역에서, 특정 기간 동안, 특정 소비자에게 어떤 일이 있었는지를 알아낸다. 이런 데이터가 기업을 운영하는데 매우 중요한 정보다. 힐리온은 “BI가 기업들이 더 많은 정보를 원하도록 만들었다는 점에서는 성공한 것”이라며, “사업이 어떻게 돌아가는지를 알게 될수록, 어떻게 하면 더 잘 될 수 있는지가 궁금해지기 마련”이라고 말했다. 그래서 일부는 BI가 실패했다기보다는 한계에 다다른 것이라고 주장한다.

또한 DW 역시 나름의 한계점이 있었다. 현재 DW는 주로 숫자가 딱딱 들어 맞아야 하는 정형화된 리포팅에 사용되고 있다. 그런데 DW는 너무 크고, 느리며 규제가 심해 다루기가 힘들다. 그래서 빅 데이터를 탐색하기에는 적합하지 않다. 만약 예측성 질문을 하려 한다면 좀 더 융통성 있는 데이터 소스가 필요했다.

10여 년 전만 해도, 매우 기본적인 도구를 사용해 데이터를 분석했다. 엑셀을 사용할 때도 있었고 기능적인 측면에서 훌륭했지만 더 이상 확장이 불가능했

던 통계 도구들도 사용했다. 지금은 훨씬 더 진보된 통계 도구들이 많다. 병렬 처리가 가능한 컴퓨터와 정교한 모델링 도구, 빠른 데이터 처리 기술 등이 현재 데이터 과학자가 제대로 일을 할 수 있게 해준다. 힐리온은 “이런 조건들이 제대로 갖춰지기 시작한 것은 10년이 채 되지 않았다”며, “데이터 과학자의 등장은 현재의 기술이 데이터 과학자의 필요로 하는 시점에 다다랐다는 것을 증명해준다”고 말했다.

데이터마이닝+통계+분석+통찰력 = 가치 창출

지난 2년간, 기업들은 최적화를 매우 신중히 고려해 왔다. 직원 수나 예산이 적고 경쟁이 심할 때는 어떻게 성장할 것인가 뿐 아니라 보유한 자원을 어떻게 최적화할 것인지도 고민해야 하는데, 그것이 매우 복잡하고 어려웠다. 기업들은 스스로를 간소화시키는 데 있어 매우 혁명해져야 했고, 그러기 위해서는 비즈니스 프로세스에 대한 깊은 이해가 필요했다. 이런 필요성이야말로 데이터 분석에 꼭 필요한 것이다.

데이터 과학자는 데이터 마이닝과 통계, 그리고 모델링 기술을 대량의 데이터에 적용해, 기업에 도움이 될 만한 결론을 이끌어 내거나 최근의 트렌드나 패턴을 연구한다.

레볼루션 애널리틱스 CEO 노만 나이는 “현재 기업들은 내부에서 비즈니스 기회를 찾아내기 위해서는 데이터 과학자를 통해 데이터를 활용해야 한다는 것에 동의하고 있다”며, “데이터는 새로운 자원이라는 말이 있다. 이 원자재를 잘 정제하는 기업은 성장 잠재력이 클 것”이라고 예상했다.

기업 측면에서 봤을 때 데이터 과학자는 고객의 행동이나 시장 주기 같은 구조화되지 않은 데이터의 숨겨진 패턴을 찾아냄으로써 새로운 기회를 창출할 수 있다. 개발자 측면에서는 데이터 과학자가 발견한 심오한 데이터 동향을 이용해 고객 유지를 위해 웹사이트를 최적화할 수도 있다. IT 부서 내에서도 데이터 과학자는 초기에 잠재적인 스토리지 클러스터의 장애를 찾아내거나 포렌식 분석을 통해 보안 위협을 추적할 수도 있다.

스티브 힐리온은 “기업들은 이제 단순히 지난 분

기별, 월별, 연도별 단위의 실적을 알고 싶은 것이 아니라, 전체적인 소비 패턴과 왜, 어떻게, 이런 패턴들이 생겨났는지를 알고 싶어 한다”고 설명했다.

사용자와 데이터 간 가교 역할

IBM의 앤솔 밤브리는 데이터 과학자는 기업이 경쟁우위를 가지고 살아 남기 위해 무엇이 필요한지를 제시하는 사람이라고 말했다. 또한 데이터 과학자는 빅 데이터에서 어떻게 하면 최대 가치를 끌어내고 새로운 정보를 통합할 지에 대해 C레벨 임원들에게 직접 조언을 주고 함께 일한다.

다양한 방법으로 오늘날의 업무에서 변화 담당자로서 부서를 조정하고 정보를 통합하기도 한다. 그래서 데이터 과학자는 IT 부서의 일원인 경우가 있긴 하지만, 대부분의 경우 마케팅이나 기획, 재무팀과 같은 실용적인 부서에 있을 확률이 높다. 가끔은 IT와 비즈니스의 중간지점이라고 할 수 있는 연구 부서에 있는 경우도 있다.

빅 데이터 관련 최초의 모임인 오릴리 스트라타 컨퍼런스(O'Reilly Strata Conference) 창립자인 에드 텀빌은 데이터 과학자와 관련해 ‘데이터와 파이널 마일(final mile)’이라는 표현을 썼다. 네트워크 종사자에게는 이 파이널 마일이란 표현이 친숙하게 들릴 수 있다. 이는 최종 사용자와 인터넷 간의 최종 연결 구간을 의미하기도 한다.

텀빌은 “비즈니스 내부의 데이터 과학과 분석에 관련된 것으로, 데이터 측면에서 파이널 마일이란 데이터를 적절히 처리하고 진정으로 중요한 정보를 전달하는 능력을 의미한다”고 말했다. 또한 데이터 과학자는 데이터를 정보로 바꾸는, 그리고 이를 토대로 최종 사용자와 비즈니스 가치를 연결해주는 가교 역할이라고 설명했다.

통계, 모델링, 수학적 지식은 우선 조건

스티브 힐리온에 따르면, 데이터 과학자는 통계와 모델링, 그리고 수학적인 지식이 있어야 한다. 이것이 가장 우선 조건이라고 강조했다. 물리학자가 데이터 과학자로 두각을 드러내는 경우가 많은데, 그들은

실제 상황에 수학을 적용시켜 모델을 만들어 내는 데 익숙하기 때문이다. 생물 정보학을 연구하던 사람들도 데이터 과학자에 적합하다.

또한 일하려는 분야나 사업에 대한 깊은 이해도 요구된다. 데이터 과학자들은 가설을 세우는 것으로 일을 시작한다. 그 분야에 대한 충분한 경험 없이는 가설을 세울 수 없다.

한편 다른 사람들 말에 귀를 기울이는 것도 중요하다. 소프트웨어 엔지니어는 고객의 요구를 잘 알고 있어야 하는 것처럼, 데이터 과학자도 경영진이나 관리직 임원의 질문을 잘 듣고 답할 수 있어야 하기 때문이다.

좋은 데이터 과학자가 되기 위한 또 다른 요건은 ‘데이터를 얼마나 잘 가시화시켜 설명해 내느냐’다. 대부분의 경영진들은 데이터 과학자들이 내놓는 정보를 다 이해하기에는 역부족일 것이다. 데이터 과학자는 비즈니스에 가장 필요할 만한, 가장 큰 영향력을 가질 만한, 그리고 실제로 사업 구상이 가능한 정보를 추려내야 한다. 또한 결과를 시각적으로 설득력 있게 표현해내는 것도 중요하다.

핵심 역량은 정확한 질문을 이끌어내는 경험

IBM의 앤줄 밤브리는 성공적인 데이터 과학자는 데이터 관리에 대한 기능적인 통찰력보다는 데이터가 무엇을 말하는지를 알려줄 때, 비로소 최고의 가치를 제공할 수 있다고 전했다. 밤브리는 무엇보다 IT를 통해 정확한 질문을 적용하는 경험이야말로 데이터 과학자의 중요한 요건이라고 주장했다. 데이터를 찾아내는 방법이나 예측 분석을 활용하는 방법에 대한 깊은 지식은 부차적이라는 얘기다.

이에 따라 기업들은 데이터 기술에만 초점을 맞춘 전통적인 체크리스트를 사용한다면, 뛰어난 데이터 과학자 후보군을 놓칠 것이라고 충고했다.



아마존의 존 라우저는 자신의 경력을 통해 데이터 과학자의 요건을 설명했다. 라우저는 18년 전 우주공학과 컴퓨터공학 학위를 받았고, 10년 정도 소프트웨어 엔지니어로 근무했으며 2003년 아마존에 입사하면서 데이터 과학자의 길을 걷게 됐다고 말했다. 라우저는 “데이터 과학자를 고용하기란 매우 어려울 것이다. 또한 사내에서 양성하는 것도 쉽지 않다”고 전했다.

라우저는 데이터 과학자가 갖춰야 할 능력에 대해 여섯 가지를 제시했다. 첫째와 둘째는 수학과 공학이다. 셋째는 글쓰기다. 글쓰기는 커뮤니케이션으로 다른 사람에게 요점을 잘 전달하는 능력이다. 넷째는 비평적인 시각이다. 이는 가설을 검증하거나 가설에 어긋나는 데이터를 분석하는 데 필요하다. 다섯째는 호기심이다. 호기심은 생산성을 높여준다. 마지막으로 ‘행복’이다. 라우저는 “데이터 과학자가 행복을 느낀다는 것은 자신이 하는 일이 기업에 중요한 영향을 미치며 더 나아가 기업과 사회를 좀더 나은 세상으로 만드는 게 기여하는 것”이라고 말했다.

기술적 배경에 따라 추가로 배워야 할 것이 다르다

IT 채용 및 컨설팅 업체 모디스의 휴스턴 지부 부사장 로라 켈리는 데이터 과학자가 되기 위해 필요한 습득해야 할 기술은 그 사람이 어떤 기술적 배경을 지녔느냐에 따라 다르다고 설명했다.

MBA 졸업생이라면, 통계 소프트웨어 프로그램 자격증이 따야 한다. 일반 컴퓨터 과학 박사 학위를 취득하지 말고 주력 분야에 좀더 집중할 필요가 있다.

컴퓨터 과학 전공자라면, MBA 입학을 고려해야 한다. 비즈니스를 이해하고 커뮤니케이션을 잘 해야 하는 것은 데이터 과학의 필수 요소다. 또한, BI와 분석 소프트웨어 관련 자격증 또한 취득해야 한다.

통계나 정량 분석 전문가라면, 컴퓨터 과학과 비즈니스 커뮤니케이션 교육을 집중적으로 받아야 한다. 컴퓨터 과학에는 자연어 처리와 기계 학습(machine learning)이 포함된다.

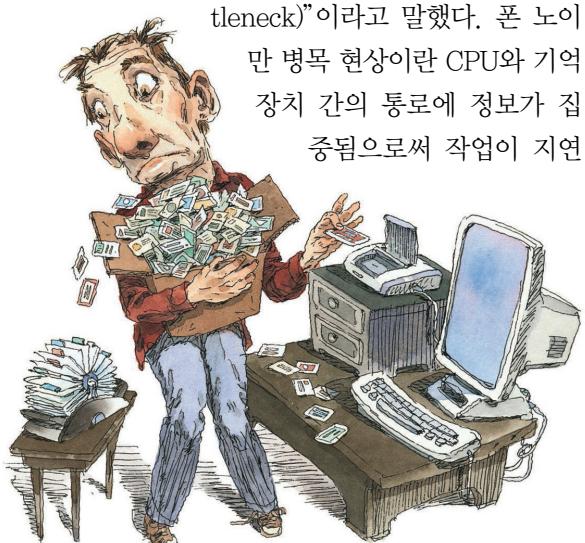
켈리는 미래에 데이터 과학자가 되려는 이는 주변의 동료나 데이터 과학자라고 하는 다른 전문가에게 직접 혹은 사이트를 통해 질문을 던지는 것도 좋다고 권유했다.

데이터 과학자가 데이터 분석가 등의 다른 데이터 전문가와 구별되는 점은 비즈니스적 결정을 이끌어내는 데이터 이면의 로직(Logic)을 생성할 수 있는 능력이다. 켈리는 “데이터 과학자는 데이터를 추출하고 모델을 구성하며 주도적으로 양적 분석을 적용할 수 있어야 한다”라고 말했다. 켈리에 따르면, 이런 막중한 책임 때문에 데이터 과학자의 평균 급여는 11만~14만 달러 정도로 책정되고 있다.

노만 나이는 데이터 과학 분야에는 초보 수준의 데이터 클리너(Data Cleaner)부터 고급 수준의 통계학자에 이르기까지 다양한 수준의 데이터 과학자를 필요로 하기 때문에 초보자에게도 이 분야에 진입할 수 있는 기회가 부여된다고 설명했다.

에드 덤빌은 “핵심은 데이터 과학자가 거론된 지 1년이 지난 현재에도 매우 적은 수의 데이터 과학자만이 활동하고 있다는 사실이다. 데이터 과학자의 부족은 새로운 폰 노이만 병목현상(Von Neumann bottleneck)”이라고 말했다. 폰 노이

만 병목 현상이란 CPU와 기억장치 간의 통로에 정보가 집중됨으로써 작업이 지연



되는 현상을 말한다.

덤빌은 데이터 과학자가 모든 작업을 한번에 처리하는 천재일 필요는 없다고 강조했다. 대신 데이터 과학과 관련한 여러 능력을 갖춘 구성원들과 하나의 팀을 꾸려 효율성을 증대시킬 수 있을 것이라고 말했다.

데이터 과학자, 수요 급증

대규모 데이터를 분석하고 이를 통해 가치를 창출하려는 노력은 인터넷이 없던 시절부터 시도됐다. 힐리온은 “코카콜라나 P&G와 같은 소비자 업종의 경우 예전부터 리서치 전문가나 통계 전문가를 고용해 계량 경제학적 기술을 적용, 고객들의 데이터를 분석해왔다”고 말했다.

그러나 인터넷이 발전함에 따라 상황은 달라졌다. 기업마다 더 많은 데이터가 생겨나고 있으며 분석 기술들도 더 다양해지고 있다. 예전에는 데이터와 관련된 산업에서만 필요로 하던 이 기술들이, 점차 확산됐다. 현재 거의 모든 산업군에서 기업은 대규모 데이터 분석을 통해 수많은 데이터에 묻힌 전략적 통찰이 가진 엄청난 비즈니스 가치를 창출할 수 있는 전문가를 찾고 있다.

데이터 전문가들은 주로 대량의 소비자 정보를 관리하는 대기업에서 데이터 과학자의 수요가 클 것으로 예상했다. 또한 웹 기반의 업체나 광고 업체, 통신 및 휴대폰 업체, 또는 영업이나 마케팅 데이터를 추적하는 기업들도 주요 수요처다. 실제로 데이터 과학자들은 에너지, 전자 상거래, 건강관리, 재무 서비스 등 다양한 분야로 진출하고 있다. 전문가들은 이런 움직임은 시작에 불과하다는 데 모두 동의했다.

최근 구인 공고를 통해 데이터 과학자를 채용하고 있는 IT 기업으로는 페이스북, 구글, 스텁블어핀, 그루폰, ADP, 페이팔 등이 있다. 이 기업들의 구인 광고를 살펴 보면, 각 기업들이 원하는 데이터 과학자의 역량이 조금씩 다르다는 것을 알 수 있으며, 또한 데이터 과학자가 갖춰야 할 현실적인 능력이 어떤 것인지 파악할 수 있다.



페이스북, SW엔지니어+정량조사 전문가

데이터 과학자는 소프트웨어 엔지니어와 정량 조사에 익숙한 전문가다. 온라인 소셜 네트워크의 연구에 강한 관심과 자사가 최고의 제품을 만들도록 도울 질문을 확인하고 답을 줄 수 있는 열정을 가진 데이터 과학자를 찾고 있다.

***자격 조건 :** 관련 분야의 이공계 석사 또는 박사 학위 소지자거나 관련 업무 경력 4년 이상인 사람, 정성적인 접근을 사용한 분석 문제 해결하는 데 경험이 풍부한 사람, 다양한 정보에서 데이터를 가져와 대용량의 복잡한 고차원 분석에 익숙한 사람, 데이터와 관련한 고난이도의 질문에 답을 찾고 경험적 연구에 대해 열정적으로 도전할 사람, R, 맷랩(Matlab), SAS 등의 분석 툴 전문가

구글, 의사결정 지원 분석가

의사 결정 지원 엔지니어링 애널리스트로서, 데이터 과학자는 구글의 광고를 측정하고 개선하는 것을 돋는 일을 할 것이다. 또한 포괄적인 문제에 대해 엔지니어들과 애널리스트로 이뤄진, 다양한 배경을 가진 팀과 협력할 것이다. 데이터 과학자는 사후 분석



및 통계 방법을 활용해 광고 품질과 매출을 어떻게 최적화하고 사용자 행동 모델링을 어떻게 설계할 지와 같은 문제를 들여다 보면서 광고 품질을 측정해야 하는 과제를 해결할 것이다.

***자격 조건 :** 관련 분야의 이공계 석사 또는 박사 학위 소지자, 관련 업무 경력 4년 이상인 사람, 특히 데이터 관련 분야의 실제 경력자, 데이터와 행동 패턴에서 결론을 끌어낼 수 있는 사람, 다른 사람에게 기술을 가르치고 새로운 기술을 습득하는 것을 즐기는 사람, 검증된 리더십 및 자기 주도적인 사람



스텀블어펀, 대량의 데이터 처리 전문가

스텀블어펀은 우연한 발견, 점수 매기기, 평가 기록 150억 건 이상의 데이터뿐 아니라 5,000만 개 이상의 웹사이트, 1,500만 사용자에 대한 마크업 및 선호도 데이터에 접근할 수 있는 전문가를 구한다. 데이터 과학자는 여기에서 흥미로운 패턴을 발견하고 데이터 변경 내용을 추적하기 위해 데이터마이닝 알고리즘을 쓰거나 데이터마이닝 툴을 사용할 것이다. 또한 연구팀에서 조사 결과를 발표한다.

***자격 조건 :** 분석이나 기술 분야에서 박사 또는 석사 학위 소지자(예 : 응용 수학, 통계, 물리학, 컴퓨터 과학, 운영 연구 또는 관련 분야), 대용량 데이터 세트를 다뤄본 경력자, 데이터 마이닝 툴 사용 경험이 있는 사람, 수학 및 통계학 지식이 풍부한 사람, 자바/PHP 또는 다른 언어로 프로그래밍 기술 보유자

그루폰, 비즈니스 가치 창출가

대규모 데이터 분석 및 모델링에 대한 열정이 있는



가? 인터넷 업계에서 가장 흥미롭고 도전적인 과제에 대한 아이디어를 실용적인 솔루션으로 바꿀 수 있는 기회를 갖고 싶은가? 그리고 가장 중요한 부분인, 높은 야망을 가지고 재미있는데도 빠르게 성장하는 팀의 일원이 되고 싶은가?

* 자격 조건 : 데이터 마이닝,

기계 학습, 통계 분석, 응용 수학 박사 학위 또는 이와 동급의 자격 소지자, 대규모 데이터 분석에서 3년 이상 실무 경험이 있는 사람, SAS, R 등 분석 툴을 완벽하게 다룰 수 있는 사람

페이팔, 빅 데이터 관리 전문가

페이팔은 자사의 빅 데이터 플랫폼 팀과 함께할 사람을 찾고 있다. 이 팀은 주로 리스크 관리를 담당하는데, 여기에 지원하면 기계 학습 엔지니어와 과학자의 핵심 부서의 일원이 될 수 있으며, 빅 데이터를 관리하고 마이닝 솔루션을 구축할 수 있는 기회이기도 하다. 페이팔은 사기 탐지 비즈니스 문제를 해결하는데 고도의 분석 기법을 적용하기 위해 통계 전문가와 분석 전문가들과 매우 긴밀하게 협력한다.

* 자격 조건 : 데이터 마이닝에 사용되는 다양한 기술을 이해하는 숙련된 엔지니어 및 과학자, 하둡, 맵리듀스, PIG 경력자, 자바나 스크립트 언어(필/파이썬/셀 스크립트 등)같은 프로그래밍 언어를 자유자재로 다룰 수 있는 경력자, 석사 학위 또는 동급의 자격 소지자



피트비트, 해커+과학자 형태의 연구원

피트비트는 프로토 타입을 만들고 선구적 제품을 제공하는, 세계적 수준의 해커-과학자 형태의 연구 팀을 만들고 있다. 피트비트의 데이터 과학자로 입사

하면, 정형/비정형 데이터를 이해하는 방향으로 기계 학습과 통계를 적용하는 경험하게 될 것이다. 연구 범위는 제품 개발과 의사 결정을 이끌어낼 프로토 타입과 매트릭스 개발, 데이터 분석, 데이터 시각화다.

* 자격 조건 : 경력 3~5년 이

상인 사람 또는 통계, 응용 수학이나 컴퓨터 과학 석/박사, 기계 학습 및 데이터 마이닝(분류, 클러스터링, 시차 분석, 추천 시스템) 전문가, 통계 또는 정량 분석 툴(r, 맷랩) 전문가, 자바, 스트라이프, 스프링, MySQL, Solr, 분산 eh캐시(ehcache) 분야 경력자



페이팔, 팀으로 데이터 과학자의 부족 해결

이베이의 지불결제를 담당하는 페이팔의 수석 과학자 목 오는 환상의 데이터 과학자 팀을 이끌고 있다. 목 오는 앞으로도 계속 노스 캐롤라이나 대학원

이나 일리노이 대학의 데이터 과학 여름학교와 같은 데이터 과학자 교육 기관에서 프로그램을 이수한 사람들을 발굴한다는 계획이다.

페이팔은 고객과 협력업체들이 생성하는 수십 테라바이트의 데이터를 연구해 구매 추이를 예측하기 위한 계획을 세우고 있다. 목 오는 지출 및 행동 패턴 데이터를 세심하게 혼합해 새로운 고객을 페이팔과 협력업체 생태계로 끌어들이는 프로파일을 개발하고 추이를 파악할 수 있기를 기대하고 있다.

오가 원하는 이상적인 지원자는 비즈니스, 데이터 분석, 컴퓨터 공학 등 모든 기술을 갖춘 인물이다. 그러나 충분한 자질을 갖춘 사람이 드물다는 것이 문제다. 오는 “이 세 가지를 모두 갖춘 사람을 만나는 것은 쉽지 않다”면서, 따라서 세 가지 훈련을 조합해 데이터 공학 팀을 꾸리고 있다”고 전했다.

- 다수(80%)는 기계 학습, 자연어 처리, 데이터 마이닝에 초점을 맞춘 박사 학위 소지자들이다.
- 10%는 핵심 성과 지표를 개발하기 위한 데이터 모델링과 분석기술을 갖춘 통계학자들이다.
- 나머지 10%는 ‘왜 사람들이 더 이상 페이팔 사용하지 않는가’ 등 적절한 질문을 던질 수 있는 MBA 학위 소지자들이다.

오는 자신의 팀이 페이팔을 다음 세대로 발전시키고 고객에게 더 나은 서비스를 제공할 것이라 확신했다.

데이터 과학자, 실패를 동반한 황금거위

BI 업체인 클릭 뷰의 제품 관리 부사장 도널드 파머는 대부분의 기업이 데이터 공학을 활용해 프로세스를 향상시키고 새로운 비즈니스 기회를 발견할 수 있다고 말했다. 예를 들어, 투자정보 서비스 기관에서 데이터 과학자들은 거래 및 위험 관리를 위한 알고리즘을 개발할 수 있으며, 제약 분야에서는 약제 시험 결과를 연구할 수 있다.

하지만 파머는 데이터 과학자를 영입하는 기업이 감안해야 할 점은 실패에 관대할 수 있어야 한다는 것이라고 말했다.

파머는 “데이터 공학은 전적으로 경험적인 학문이다. 기업들은 실패할 가능성이 있을 뿐 아니라 실패할 것으로 예상되는 구조를 세워야만 한다. 그렇지 않으면 데이터 과학자들은 충분히 노력하지 않을 것”이라고 말했다.

이는 자금이 부족한 기업에 있어서는 쉽지 않은 결정이다. 데이터 과학자가 되기까지 자신의 일상을 계시한 블로그 ‘데이터 과학 101(Data Science 101)’ 운영자인 라이언 스완스트롬은 “이따금 새로운 모델에서 적절한 아이디어를 얻기 위해서는 실패해야 한다. 물론 위험이 따르며 비용도 발생한다”고 말했다.

적절한 환경에서 훌륭한 데이터 과학자는 기업에게 새로운 비즈니스 기회를 가져다 주는 황금거위 역할을 할 것이다. 특히 페이팔 등 새로운 고객을 찾고 기존의 고객에 대한 서비스 향상에 초점을 맞춘 기업

에게는 그 효과가 더욱 클 것이다. 스완스트롬은 컴퓨터 공학, 비즈니스, 분석 등을 물리학 또는 심리학 등 주변 학문들과 연계함으로써 결과를 향상시킬 수 있을 것으로 판단했다.

모디스의 켈리는 데이터 과학자의 역할을 ‘진행 중인 작업’으로 명명했다. 켈리는 “1년 전과 지금 기업들이 생각하는 데이터 과학자는 완전히 다르다”라고 평가했다.

부정적으로 바라본 데이터 과학자, 5D 직종 가능성

화려하고 멋드러지게 사건을 해결하는 CSI 드라마와는 달리 실제 과학수사관은 시체 조각과 배설물, 정액, 바퀴벌레, 쓰레기들을 분석하며, 매일 밤 야근은 당연한 일과다. 특히 형사나 경찰에게는 CSI는 자신의 수사에 증거를 가져다 주는 도우미 역할로 취급받는다.

앞서 말한 데이터 과학자의 지대한 역할이나 화려한 분석 스킬, 뛰어난 통찰력, 높은 보수 등은 이 직업에 대한 환상을 심어주기에 충분하지만, 현실은 CSI의 그것과 동일하다.

우선 데이터 과학자는 데이터 쓰레기 속에서 헤매고 다닐 수밖에 없는 운명이다. 뿐만 아니라 새로운 비즈니스 기회를 창출한다는 것은 기존에 없던 것을 만들어내는 마치 예술 작품을 창작하는 스트레스를 받을 것이다. 특히 심혈을 기울여 만든 비즈니스 기회를 경영진들이 알기 쉽게 설명하고 설득해 이것이 실제 적용 단계에 이르기까지는 무엇보다 지난한 업무가 될 것이다. 그래서 일부 전문가들은 데이터 과학자가 가져야 할 최고의 덕목은 통찰력이 아닌 인내라고 주장했다.

특히 국내 비즈니스의 현실을 본다면 데이터 과학자는 4D 직종인 개발자에 이어 5D 직종의 충분 조건을 갖추고 있다고 해도 과언이 아니다.

그럼에도 불구하고 전 세계는 데이터 과학자를 구하고 있으며, 이에 대한 관심은 더욱 커져갈 것은 자명한 일이다. [ITWORLD](#)

IDG Korea Contacts

The collage consists of several screenshots from different IDG Korea websites:

- CIO Korea:** Shows sections like "Office 365 vs 구글 문서도구 대결: 성능 비교" (Comparison between Office 365 and Google Docs: Performance), "Digital Marketing 2011" forum, and "IDG News".
- IT World Korea:** Shows sections like "IT World News", "IT World Events", "IT World Features", and "IT World Specials".
- IDG News Korea:** Shows sections like "IT News", "Business News", "Technology News", and "Opinion".

Each screenshot includes a header with the IDG logo and a footer with copyright information.

Sales Contact

김성일, Eddy Kim

Tel: + 82-2-558-6939 Mobile: + 82-10-2702-0360

Email: sungil_kim@idg.co.kr

이포원, Tony (Powon) Lee

Tel: + 82-2-558-6924 Mobile : + 82-10-2053-2413

Email : tony_lee@idg.co.kr

양지은, Michelle Yang

Tel: + 82-2-558-6980 Mobile: + 82-10-9160-5454

Email: michelle_yang@idg.co.kr

CIO Editorial Contact

천신웅, Brian Cheon

Tel: + 82-2-558-9882 Mobile: + 82-10-3221-7824

Email: psy_cheon@idg.co.kr

박해정, Jenny Park

Tel: + 82-70-7725-9955 Mobile: + 82-10-2810-4518

Email: jenny_park@idg.co.kr

IT World Editorial Contact

박재곤, Jay Park

Tel: + 82-2-558-9887 Mobile: + 82-10-2782-6722

Email: jay_park@idg.co.kr

이대영, Dy Lee

Tel: +82-70-7725-9951 Mobile: +82-10-6511-8982

Email: dy_leed@idg.co.kr

김현아, Hyuna Kim

Tel: + 82-2-558-6956 Mobile: + 82-10-8300-8511

Email: hyuna_kim@idg.co.kr

PR & Marketing Contact

김종민, Matt Kim

Tel: + 82-2-558-6079 Mobile: + 82-10-2408-4121

Email: matt_kim@idg.co.kr

박용학, Yong Park

Tel: + 82-2-558-6965 Mobile: + 82-10-4744-2823

Email: yongpark@idg.co.kr



IDG KOREA

4F ChangHwa Bldg., Bongnae-dong 1Ga, Jung-Gu, Seoul, Korea 100-161