빅데이터
승리의 과학

빅데이터는 당신이 무엇을 선택할지 알고 있다

빅데이터, 승리의 과학
The Science of Winning with Big Data

초판 1쇄 인쇄 · 2013년 4월 19일
초판 1쇄 발행 · 2013년 4월 25일
아이패드용 전자책 발행 2013년 5월 25일

지은이 | 고한석
발행인 | 이지연
펴낸곳 | 이지스퍼블리싱(주)
출판사 등록번호 | 제313-2010-123호
주소 | 서울시 마포구 동교로 22길 29(구 서교동 375-13번지) 성지빌딩 301호
대표전화 | 02-325-1722 팩스 | 02-326-1723
홈페이지 | www.easyspub.co.kr 이메일 | service@easyspub.co.kr

기획 및 책임 편집 · 이지연 | 교정교열 · 이명애 | 표지 디자인 · 디박스
내지 디자인및 편집 · 트인글터 | 인쇄 · 이쪽피앤피 | 영업 · 이주동

이 아이패드용 전자책은 '빅데이터 승리의 과학' 초판 1쇄를 아이패드용 애플리케이션으로 제작한 것입니다.

이 책에 실린 모든 내용, 디자인, 이미지, 편집 구성의 저작권은 이지스퍼블리싱(주)과 지은이에게 있습니다.
허락 없이 복제할 수 없습니다.

ISBN 978-89-97390-19-9 13320
종이책 가격 15,000원

빅데이터 승리의 과학

빅데이터는 당신이 무엇을 선택할지 알고 있다

고한석 지음(하버드대 케네디스쿨 IT정책 연구)

이지스퍼블리싱



저자 **고한석**

서울대 중문과를 졸업하고, 하버드대학 케네디스쿨에서 IT정책으로 석사학위를 받았다. 95년 솔빛조선미디어 멀티미디어 제직팀장으로 일하다 IT 기술과 정보화 사회에 대한 관심이 커지면서 직장을 그만두고 유학길에 올랐다. 그후 하버드대학 케네디스쿨에서 석사학위를 받은 후 SK에 입사하여 SK China IT/인터넷 사업개발팀장으로 중국에서 4년 동안 일했다. 이때 외국인으로는 최초로 중국 동창회 커뮤니티 포털인 ViaFriend.com을 오픈하여 가입자를 1,000만 명까지 늘리기도 했다. 이후 삼성네트웍스의 글로벌 사업추진팀장으로 5개 해외사무소를 총괄하는 일을 하던 중 젊은 시절 가졌던 꿈인 사회변화에 기여하는 삶을 살고자 당시 열린우리당 정책연구원에 정책기획 연구원으로 들어갔다. 그후 당의 정세분석국장으로 여론조사 데이터를 관리하였다. 정책분야에 대한 관심을 구체화하고 싶어 (주)사회디자인연구소를 만들어 초대 소장을 맡아 활동하기도 하였다. 현재는 IT에 기반을 둔 사회혁신과 데이터 마이닝에 관심을 가지고 집필활동과 강연을 하고 있다.

E-메일 ko.hanseok@gmail.com

차가운 숫자 너머로 따뜻한 사람을 보다

'데이터'의 어원은 라틴어 'datum'의 복수형이고, 'datum'은 '주다'라는 뜻을 가진 'dare'의 과거분사다. 즉 '주어진 (것)'을 말한다. 따라서 현실세계에서 우리 눈앞에 주어진 모든 것은 다 데이터이다. 세상은 그 자체로서 '빅 데이터'인 것이다. 그리고 데이터 더미 속에서 의미를 '캐내는'(mining) 일이 데이터 분석이다.

우리가 일상에서 가장 많이 하는 데이터 분석은 '평균'을 구하는 것이다. 평균을 통해서 한 집단의 특성을 파악한다. 하지만 세상에 평균은 실재하지 않는다. 평균적인 한국인이란 없으며 평균적인 진보도 평균적인 보수도 없다. 단 한 사람도 똑같은 사람은 없다. 모두 각자의 특징을 가지고 있고 조금씩 서로 다른 관심과 생각을 가지고 있다.

언제부터인가 우리는 정치를 우리의 일상과는 상관없는 '그들만의 리그'라고 생각해 왔다. 그것은 정치가 다양한 시민 개개인의 구체적인 관심사에 대해서 듣고자 노력하지 않았으며, 현실적 조건의 제약으로 인해서 들을 수도 없었기 때문이었다. 그러나 2012년의 오바마 캠프는 최첨단의 테크놀로지와 지식을 동원하여 시민들이 제각기 무엇을 궁금해하고 어디에 관심을 가지는지 듣고자 노력하였으며 그에 대해서 답변하고자 노력하였다. 시민들은 투표로써 이에 회답하였고 오바마는 선거에서 승리할 수 있었다.

작년부터 오바마의 빅데이터 선거전략을 조사, 연구하면서 숫자들이 단지 차가운 활자체가 아니라 그 뒤에 가슴 뜨거운 한 명 한 명의 사람들이 있다는 것을 느꼈다. 정치를 하든, 경제활동을 하든, 학문을 하든, 예술을 하든, 사람에 대한 관심과 애정이 없다면 무슨 의미가 있겠는가?

길고 추웠던 지난 겨울 내내 원고와 씨름하느라 지친 내게 웃음을 잃지 않게 해준 아내 은경과 아들 주몽에게 사랑과 감사를 전한다.

고한석

추천의 글

데이터로 판단하려는 합리적인 사람들에게 통찰력을 주는 책

어떤 주제를 '피상적'으로 아는 것과 '구체적' '실제적'으로 파악하는 것은 커다란 차이가 있다. 〈빅데이터, 승리의 과학〉은 이러한 차이가 어떤 것인지를 극명하게 알려주는 뛰어난 책이다. 빅데이터와 관련되어 회자되는 '이야기' 수준의 논의와, 그에 근거한 터무니없는 '무용론(無用論)'들은 이 책에서 펼쳐지는 빅데이터 활용과 그 효과에 대한 구체적인 논증 앞에 무색해진다.

흔히 정치는 기업에 한참 뒤져 있다고들 한다. 그러나 빅데이터의 활용이라는 측면에서 보면 미국의 정치는 기업을 한참 앞서가고 있는 것으로 보인다. 기업이 ROI 타령을 하며 투자를 주저할 때, 미국의 대통령 선거에서는 2억 명이 넘는 유권자에 대한 빅데이터를 가지고 가장 확실한 ROI인 '승리'를 엮어냈다는 것을 이 책을 통해 확인했다. 저자는 이러한 '승리'의 과정을 기업 등의 조직에서 재현할 수 있도록 단계적인 방법론도 제시하고 있다.

최근 포춘 500대 기업 중 90% 이상이 빅데이터를 활용하고 있거나, 활용할 예정이라는 발표를 보았다. 더 이상 빅데이터는 유행이 아니다. 이 책이 기업에서 데이터로 판단하려는 합리적인 분들과 빅데이터를 활용하려는 분들에게 확실한 가이드와 통찰력을 줄 것이다.

LG CNS 어드밴스드 애널리틱스(Advanced Analytics) 사업부문장 • **박용익**

역사상 가장 방대한 데이터를 사용한 미 대선을 통해 구체적인 활용을 배운다

미래를 준비하는 과정은 항상 유쾌한 일이다. 희망이라는 마약이 현실의 고통을 잊게 해주기 때문이다. 나는 이 책을 읽고 얼룩진 안경을 깨끗이 닦고 하늘을 바라볼 때의 느낌을 받았다. 저자는 국내 정치분야에서 활동하고 있는 많은 사람들, 특히 스핀닥터(spin doctor)와 폴스터(pollster)* 들이 앞으로 어떤 일에 집중해야 할지 명확하게 알려주었다.

---

* 스핀닥터(spin doctor)는 주로 정치권의 여론 담당 홍보전문가를, 폴스터(pollster)는 여론조사원을 말한다.

이 책을 통해 저자는 지난 미국 대선에서 승리한 오바마 캠프의 무대 뒤에는 '빅데이터와 그것을 운영한 분석팀'이 있었다는 것을 구체적으로 보여주고 있다. 그간 국내에서 출판된 빅데이터 관련 저서들은 거의 모두 알고리즘 측면에서 접근하거나 비즈니스의 비전을 조망하였다. 그러나 이 책은 그 영역을 정치사회 분야로 확장시킨 최초의 서적일 뿐 아니라, 역사상 가장 방대한 데이터를 다루어 성공에 이른, 미 대선의 빅데이터 전략을 구체적으로 밝혀 그 가치가 매우 높다고 할 수 있다. 이 책을 읽은 국내 많은 전문가들이 데이터를 모으고 연결하고 정제하고 분석하는 과정을 상상해 본다. 나는 다음 선거가 기다려진다.

미디어리서치 상무이사 • **김지연**

빅데이터로부터 가치를 창조하는 과정이 흥미진진하게 담겨…

이 책, 아주 재미있게 봤다. 2000년부터 웹 검색엔진 관련 일을 하면서 대용량 데이터를 다뤄본 사람으로서, 빅데이터란 대용량 데이터를 처리하여 이를 통해 새로운 의미나 가치를 창조하는 새로운 시스템이라 생각한다. 그런 면에서 〈빅데이터, 승리의 과학〉은 대용량 데이터로부터 부가가치를 창조하는 여러 가지 예들을 아주 흥미진진하게 보여주고 있다.

이 책에서 눈여겨봐야 할 것은 부가가치를 창조하는 사람의 역할이다. 빅데이터는 누가 어떻게 쓰느냐에 따라 결과가 크게 달라질 수 있다. 이 시스템은 대용량 데이터의 처리를 빠른 시간 내에 가능하게 해줄 뿐, 사람들의 문제를 알아서 해결해 주는 전지전능한 도구는 아니기 때문이다. 결국 빅데이터 시스템은 사용하는 사람에 따라 강력한 힘을 발휘할 수도 있고, 아무 의미 없는 일을 할 수도 있다.

빅데이터 시스템은 사람들이 문제를 발견하고 해답을 찾아내는 것을 과거보다 효과적으로 구현해 준다. 그리고 이 과정에 수많은 가설, 실험, 분석의 반복이 동반된다는 점도 이 책에서 주의 깊게 보기를 권한다. 빅데이터의 실체를 알고 싶은 사람들에게 이 책을 강력하게 추천한다.

《Do it! 직접 해보는 하둡 프로그래밍》 저자 • **한기용**

차 례

이 책을 시작하며 ■ 승리의 과학, 빅데이터 ····· 14

데이터 분석이 가져온 애슬레틱스 구단의 성공 ····· 15
'직관'의 시대에서 '데이터'의 시대로 ····· 16
미국의 대기업들, 오바마의 테크놀로지팀 모시기 열풍 뜨거워 ····· 17
빅데이터에 대한 가장 구체적인 사례 연구 ····· 18
오바마의 빅데이터 전략에서 배우는 미래의 경영과 마케팅 ····· 19
이제, 대중이 아니라 한 사람을 만나라 ····· 20

01 빅데이터는 오바마의 당선을 알고 있었다
변화와 혁신의 토대, 빅데이터

그들의 데이터 분석은 얼마나 정확했나? ····· 24
오바마 캠프의 선거운동본부장이 제일 먼저 했던 일 ····· 25
스티브 잡스와 스티븐 스필버그 등 각계 인사의 조언들 ····· 27
폭발적인 소셜 네트워크의 성장 ····· 28
스마트폰과 무선 인터넷 서비스 ····· 32
빗나간 언론의 호들갑 ····· 33
선거 관련 IT 활동은 모두 클라우드 컴퓨팅을 이용 ····· 35
빅데이터, 가장 주목해야 할 과학기술 키워드 1위 ····· 36
진실은 빅데이터 속에 묻혀 있다 ····· 41
선거운동과 빅데이터: 마이크로 타기팅 ····· 42
신이시여, 부동층에게 제발 초록색 코를 주소서 ····· 45
융단폭격의 시대를 너머 '드론 전쟁'의 시대로 ····· 47

전문가에겐 상식 ■ 클라우드 컴퓨팅 ····· 49

02 빅데이터, 합칠수록 가치가 빛난다
유권자 데이터베이스 구축 과정

공화당이 먼저 유권자 데이터베이스 구축을 시작 ····· 54
두 번 연속 패배한 민주당 진영의 와신상담 ····· 55
1억 6천6백만 명의 정보를 저장한 보트빌더 ····· 57
협력의 촉매가 될 회사 구축: 캐털리스트 ····· 59
NGP와 VAN의 합병: 민주당 협력 1위 IT업체 ····· 63
소비자 정보를 가지고 있는 데이터베이스 업체들 ····· 64
오바마 캠프는 어떻게 데이터를 모았을까? ····· 66
미네소타 주 '할머니 여단'의 데이터 자원봉사 ····· 67
빅데이터와 롱데이터 ····· 69
전문가에겐 상식 ■ 개인정보의 수집과 제3자 제공에 관한 규제 ····· 72
전문가에겐 상식 ■ 미국의 대통령 선거 제도, 한국과 무엇이 다른가 ····· 73

03 실리콘밸리의 최고들이 모였다
빅데이터 선거를 위한 IT 인프라 개발과 운영

대통령 선거 캠프에 CTO를 영입하다 ····· 80
왜 하퍼 리드를 영입했나? ····· 82
CTO가 테크놀로지팀으로 불러모은 사람들 ····· 83

클라우드 컴퓨팅은 '아마존 웹 서비스' 이용 ····· 84
역사상 가장 큰 규모의 데이터를 다룬 테크놀로지팀 ····· 86
IT 통합 프로젝트의 코드명, '외뿔고래' ····· 88
클라우드 기능 100% 활용 ····· 92
진보적인 민주당의 보수적인 IT 사용 ····· 94
혹독한 훈련을 한 시스템 운영 그룹 ····· 96
허리케인 샌디라는 악천후와 장애 복구 훈련 ····· 97
테크놀로지팀과 다른 부서의 문화적 충돌: 해결법은? ····· 99
우리는 '화력을 증강시키는' 팀: Force Multiplier ····· 100
전문가에겐 상식 ■ 외뿔고래와 데이터베이스 ····· 103
전문가에겐 상식 ■ 아마존 웹 서비스, AWS ····· 104
전문가에겐 상식 ■ 쉽게 이해하는 빅데이터와 하둡 ····· 107

04 빅데이터는 당신이 누구인지 알고 있다
빅데이터 분석과 마이크로 타기팅

23살 사라를 대하는 두 진영의 접근법 ····· 112
보수적인 동네에서 오바마 지지자 찾아내기 ····· 114
미국을 18만 개 지역으로 나누어 점수를 매긴 NCEC 점수 ····· 117
숫자, 숫자, 숫자: 슬로건 중심에서 데이터 중심으로! ····· 121
오바마 진영의 데이터 분석팀 책임자는 29살의 댄 와그너 ····· 124
데이터 마이닝, 어떻게 했나? ····· 126
군집분석과 라이프 스타일 ····· 130
최초의 마이크로 타기팅 선거는 부시가 먼저 ····· 135

연관성 분석과 장바구니 분석 ····· 138
좋아하는 맥주를 보면 어느 후보를 찍을지 알 수 있다 ····· 141
회귀분석: 어떤 요인이 지지도에 얼마만큼 영향을 미치는가 ····· 144
당신이라면 누구를 공략할 것인가? ····· 150
설득할 대상 추려내기 ····· 153
2008년 대규모 전화조사와 마이크로 타기팅 ····· 154
선거 평가 태스크포스팀과 시스템 혁신 ····· 155
전문가에겐 상식 ■ 오바마 캠프가 사용한 데이터 분석 플랫폼, HP 버티카 ····· 157
전문가에겐 상식 ■ 분석 도구, R ····· 160

05 마음을 움직이는 메시지, 과학으로 만든다

오바마 진영이 사용한 행동과학과 실험주의

데이터 분석팀은 조직 내부의 컨설턴트 ····· 164
조지 클루니, 오바마 그리고 당신을 찾아내는 능력 ····· 165
여론조사와 빅데이터 분석, 어떤 차이가 있나? ····· 173
수많은 TV 신문 중 어디에 광고하는 게 효과적일까? ····· 175
새로운 지지자를 '발굴'하는 광고 ····· 177
옵티마이저: 광고 1달러당 가장 효율적인 TV 채널은? ····· 179
공화당은 추측할 수 없는 오바마의 타기팅 알고리즘 ····· 182
드림캐처: 온라인 텍스트에서 정치성향을 캐내다 ····· 184
장바구니를 분석하면 인간의 욕구가 보인다 ····· 187
A/B 테스트로 700억 원의 후원금을 더 모으다 ····· 190
행동과학과 동기부여: 분석가 연구소 ····· 196

40명의 생각이 1천만 명을 대표하지는 않는다 ····· 197
마음을 움직인 메시지를 찾아라 ····· 200
더 잘게 사람들을 나누어라. 그리고 마음을 얻어라 ····· 201
에밀리의 리스트와 함께한 실험: 어떤 사람들이 흔들렸나? ····· 202
2012년 대선, 실험주의 꽃피우다 ····· 204
'먹히는' 메시지 만들기 ····· 205
넛지와 행동과학자 컨소시엄(COBS) ····· 207
투표를 하게 만드는 미묘한 동기부여 방법 ····· 210

06 오바마 캠프의 스마트한 무기들

빅데이터로 무장한 현장 자원봉사자들

오바마 캠프의 IT 전략: 활동가를 위해 복무하라 ····· 216
조직 내 소셜 네트워크 서비스: 내셔널필드 ····· 217
모바일 선거운동사무소: 대시보드 ····· 223
누워서도 가능한 전화홍보: 가상 콜센터 ····· 229
투표참여 독려를 위한 스마트폰 앱, 고든 ····· 233
젊은 유권자를 공략하는 페이스북 맞춤식 공유 ····· 233
선거운동 자원봉사자를 위한 여행 플래너 사이트 ····· 237
디지털팀의 사람들 ····· 238
데이터 과학을 사용하면 후원금을 더 낸다 ····· 241
조직활동가의 사관학교: 신 조직화 연구소(NOI) ····· 242
민주당의 또 다른 전당대회, 뿌리 캠프 ····· 248

07 빅데이터 전략을 사용하지 못한 공화당
오바마와 롬니의 IT 운영전략 비교

출발부터 뒤처진 공화당 롬니의 테크놀로지팀 ····· 252
친구에 의한, 친구를 위한 IT 아웃소싱 ····· 253
롬니 스타일의 클라우드 컴퓨팅 ····· 256
공화당, '범고래' 프로그램의 굴욕 ····· 260
외뿔고래와 범고래: 더 근본적인 차이 ····· 264

08 빅데이터 전략에서 배워야 할 것들

사람의 마음을 얻어야 하는 모든 조직에 필요 ····· 268
데이터로 리드하라 ····· 268
데이터 생태계를 조성하라 ····· 271
빅데이터는 '발' 데이터이다 ····· 272
마이크로 타기팅의 6단계 프로세스 ····· 275
선거 후 오바마팀은 업계 스카웃 1순위 ····· 279

이 책을 끝내며 ■ **빅데이터냐 빅브라더냐** ····· 282
덧붙이기 01 ■ 오바마 선거캠프의 IT 관련 조직도 ····· 286
덧붙이기 02 ■ 참고문헌과 인용된 이미지 안내 ····· 290
찾아보기 ····· 301

승리의 과학, 빅데이터

영화 〈머니볼〉의 한 장면.
새로운 선수를 뽑기 위해 오클랜드 애슬레틱스 구단의 노장 스카우터들이 모여서 여러 선수들을 평가하고 있었다.

"내 생각에는 제로니모가 딱이야. 진짜 운동선수지. 덩치도 크고 빠르고 재능도 있고…."

"딱 부러지지. 인물도 좋아. 턱도 멋지고 공수투타 모두 능하지."

"타격은요?"

"스윙이 근사하잖아. 공이 방망이 끝에서 폭발해 버려."

"방망이 끝이 공에 맞는 순간 그 경쾌한 '탕!' 소리가 온 구장에 울려퍼진다구."

"난 페레즈가 좋아. 걘 아주 클래식한 스윙을 하지. 아주 깨끗한 스윙이야."

"하지만 커브볼을 못 치잖아."

"훈련이 더 필요하긴 해. 하지만 눈여겨볼 만하다구."

"하지만 여자친구가 못생겼어."

"그게 뭐 어때서?"

"못생긴 여자친구를 사귄다는 건 걔가 자신감이 없다는 증거지."

"헛소리하지 마. 아티 말이 맞아. 걘 자신감 충분해. 이 친구는 태도가 좋다고. 태도는 중요한 거야."

데이터 분석이 가져온 애슬레틱스 구단의 성공

모두 경험에 의존하거나 눈썰미로 파악한 평가들이었다.

하지만 단장인 빌리는 예일대학 경제학과를 갓 졸업한 야구통계 전문가인 피터를 데려온다. 피터는 "좋은 선수를 뽑는 게 목적이 아니에요. 이기는 것이 목적입니다. 이기려면 출루를 해야 해요"라고 말했다. 그러나 가난한 오클랜드 애슬레틱스 구단은 출루율이 좋은 선수를 마음대로 데려올 형편이 못 됐다. 그래서 피터는 미국의 모든 야구선수들의 데이터를 모아서 통계분석을 하여 출루할 확률이 높은 선수들을 골라낸 후 그들의 몸값을 비교하였다. 그런 다음 1달러당 출루율이 좋은 선수들 순으로 점수를 매겨 그들을 데려왔다. 이 선수들은 비록 톱클래스는 아니었지만 승리에 기여하는 선수들이었기에 결국 2002

년 시즌에서 20연속 우승을 하고 아메리칸 리그에 진출하는 쾌거를 올린다는 게 이 영화의 줄거리다. 이 영화는 사실에 근거한 논픽션 책을 영화로 만든 것으로, 책제목은 〈머니볼: 불공평한 게임에서 승리하는 예술〉(Moneyball: The Art of Winning an Unfair Game)이다.

야구에 통계분석을 접목한 이 방법은 스포츠 업계에 일대 혁신을 불러와 요즘은 거의 모든 팀에 '스포츠 과학' 부서 또는 데이터 분석팀을 두게 되었으며, 한국에서도 SK와이번스의 김성근 전 감독이 데이터 야구의 신봉자로 이름을 날렸다. 최근에는 MIT대학 주최로 매년 '스포츠 데이터 분석 대회'(Sports Analytics Conference)가 개최되어 첼시나 맨체스터 유나이티드 같은 유명 스포츠 구단의 데이터 분석가는 물론 학자, 엔지니어들이 모여서 토론을 벌인다.

'직관'의 시대에서 '데이터'의 시대로

2012년 5월, 미국 온라인 시사뉴스 사이트인 슬레이트닷컴(Slate.com)은 오바마 대통령의 선거운동을 보도하면서 기사 제목을 '직관의 죽음'(The Death of the Hunch)이라고 붙였다.

미국 대통령 선거는 무려 2조 원이 넘는 비용을 들여 향후 4년 간 나라의 운명을 좌우할 인물을 뽑는 중요한 국가적 행사이니만큼 치열한 선거운동 경쟁이 뒤따른다. 그럼에도 불구하고 선거운동 과정에서 정작 중요한 의사결정은 정치권의 경험 많은 컨설턴트들의 '직관'에 따랐다. 하지만 2012년의 오바마 캠프는 달랐다. 무엇보다도 조직의 리더가 '데이터'를 중시하였다. 오바마 캠프의 총책임자였던 짐 메시나는 선거운동본부장 직책을 맡으

면서 "선거운동의 모든 것을 숫자로 측정할 것"이라고 선언하였다. 데이터 분석팀의 멤버들 또한 단순히 골방에서 연구 프로젝트를 진행하는 통계학자들이 아니었다. 그들은 선거캠프 내 모든 부서들이 올바른 의사결정을 하는 데 꼭 필요한 데이터 분석을 책임지는 내부 컨설턴트들이었다. 이들이 개발한 시스템은 골목골목에서 활동 중인 현장의 자원봉사자들의 스마트폰에서도 실행되고 있었다. 스마트폰의 애플리케이션은 자원봉사자들이 어디를 방문할지, 또 누구와 무슨 이야기를 할지 알려주었다.

테크놀로지팀은 이 모든 것을 실현할 수 있게 시스템을 설계하고 애플리케이션을 개발하였다. 매일매일 어떤 상황이 바뀌면, 비록 사소한 변화일지라도 그에 따라 입력 변수를 조금씩 조정하고 컴퓨터 시뮬레이션을 이용해 매일 밤 6만 6,000번의 모의선거를 치르면서 전국 각지의 변화를 점검하였다. 그리고 매일 아침 이 결과에 기반하여 자금 및 인력을 어느 지역, 어떤 분야에 할당할지를 결정하였다.

미국의 대기업들, 오바마의 테크놀로지팀 모시기 열풍 뜨거워

선거 열기가 한창 달아오르던 2012년 상반기에 뉴욕, 샌프란시스코, 텍사스 오스틴, 브루클린 등지에서 후원금 모금 행사가 개최되었다. 예전 같았으면 이런 자리의 메인 게스트는 으레 유명 정치인을 비롯해 캠프 본부장들이나 전략가들이었지만 이날은 테크놀로지팀과 디지털팀의 엔지니어들이 메인 게스트로 각광을 받았다. 사람들은 이들의 경험과 전략을 고작 한두 시간 남짓 듣기 위해서 일반 입장권 50~75달러, 스페셜 입장권 500~1,500달러를 내고 후원금 모금 행사에 참석하였다.

11월 6일 대통령 선거에서 오바마 대통령이 당선된 다음날부터 이들 엔지니어와 데이터 분석가들의 전화통에는 불이 나기 시작했다. 미국의 수많은 대기업들이 이들을 모셔가기 위해서 전화를 걸어댄 것이다. 아직 어떤 대기업도 1인당 1,000여 개꼴 되는 2억 명의 데이터를 그토록 정교하게 분석하여 경쟁에서 이겨본 경험이나, 이를 실현하기 위한 테크놀로지 운영의 노하우를 가진 곳이 없기 때문이다.

〈워싱턴 포스트〉지는 이렇게 철저하게 데이터에 기반하여 진행된 오바마의 선거운동을 '머니볼 선거운동'이라고 이름 붙였다.

빅데이터에 대한 가장 구체적인 사례 연구

최근 빅데이터에 대한 관심이 높아지면서 이를 소개하는 책들이 연일 출판되고 있다. 대부분의 경우 빅데이터란 무엇이며 왜 그것이 중요한지에 대한 배경을 설명해 주고 그것이 경영에 미치는 영향에 대한 설명이 주류를 이룬다. 또는 비교적 전문적으로, 빅데이터를 가능하게 만든 테크놀로지에 대해서 일반인은 잘 알아듣지 못하는 프로그래밍 용어를 써가며 설명하는 책들도 나왔다. 하지만 정작 빅데이터를 이용해 성공한 사례에 대한 깊이 있는 분석은 거의 없는 실정이다. 일부 사례가 소개되고 있지만 기껏해야 사례별로 한두 페이지에 그치고 있다. 그런 까닭에 현재 빅데이터는 주로 IT업계 전문가들 사이에서만 논의되고 있으며, 기업이나 조직 경영에 관련된 부서들에서는 빅데이터가 중요하다는 말은 많이 듣고 있지만 정작 이것으로 어떤 일을 어떻게 할 수 있는지에 대해서는 제대로 감을 잡고 있지 못하다.

50명의 데이터 분석가와 50명의 소프트웨어 엔지니어들, 그리고 200명의 디지털 미디어팀 등 총 300명의 전문가들이 18개월 동안 2억 명의 유권자 개개인의 정치적 프로필을 파악하기 위해서 다양한 베이터베이스를 통합하고 분석하고 활용한 예는 지금까지 어디에도 없었다. 이 책은 이들이 경험하고 시도한 모든 것을 마치 눈앞에서 벌어지듯 자세하게 보여줄 것이다.

오바마의 빅데이터 전략에서 배우는 미래의 경영과 마케팅

선거운동을 영어로 캠페인이라고 부르고 마케팅도 종종 캠페인이라는 용어를 쓴다. 같은 맥락에서 기업들도 오바마의 빅데이터 전략에서 작은 노하우뿐만 아니라 21세기적 경영이 어떠해야 하는지를 배울 수 있다. 다만 기업들이 빅데이터를 다루는 데 있어서 간과하지 말아야 할 점은, 기업 경영에 빅데이터가 왜 필요한지 그 이유를 명확히 하는 것이다. 오바마 캠프의 최고기술책임자(CTO: Chief Technology Officer)인 하퍼 리드가 겸손하게 말하였듯이 테크놀로지는 '화력 증강자'(force multiplier)일 뿐이다. 승리하기 위한 전략과 이를 구체적으로 표현한 핵심성과지표(KPI: Key Performance Index)가 명확할 때만이 빅데이터는 의미를 가진다. 전략과 목표가 없는 빅데이터는 '빅 쓰레기더미'(Big Garbage)일 뿐이다. '무엇'을 '왜' 하는지 확실할 때 '어떻게' 할지도 알 수 있다.

오바마는 자신을 지지할 가능성이 있지만 투표장까지는 가지 않을 숨어 있는 사람들과, 상대방 진영에 속해 있지만 지지 정도가 약해 특정 이슈를 통해 자기편으로 만들 수 있는 사람들을 찾아내어 이들을 투표장까지 이끌어 내기를 원했다. 그래서 2억 명의 유권자 한 사람 한 사람에 대해 맞춤형 선거

전략을 세우기 위해 방대한 규모의 데이터베이스를 분석하였던 것이다.

기업이나 비영리단체, 심지어는 종교단체도 공통적으로 원하는 것이 있다. 그것은 자신의 주장을 더 많은 사람들이 받아들이길 원한다는 것이다. 그러기 위해서는 불특정 다수 속에서 설득 가능한 사람이 누구이며, 그 사람들은 어떤 사람들이며, 한 사람 한 사람은 어떻게 다른가 등을 알고 있어야 효율적으로 접근할 수 있다.

빅데이터는 이를 가능하게 해준다. 바로 그러하기에 기술이 아니라 조직문화가 되어야 한다. 최고경영자의 책상 위에 놓여지는 보고서로는 아무런 의미가 없다. 조직의 리더들이 경험과 주관적 판단이 아니라 철저하게 데이터에 기반하여 의사결정을 내리고, 조직의 최전방에 있는 말단사원들까지 데이터에 기반하여 그날그날의 영업활동을 해나갈 때 빅데이터는 의미를 가지게 된다. 그렇지 않으면 빅데이터는 기술쟁이와 숫자광들만 떠들어대는 또 하나의 IT 유행어가 될 뿐이다.

이제, 대중이 아니라 한 사람을 만나라

빅데이터가 단순한 개념이나 기술이 아니라 승리의 도구이기를 원하는 사람은 이 책을 통해서 다음과 같은 것들을 얻을 수 있다.

- **마케팅에 관심이 있는 사람** : 불특정 다수의 소비자 그룹을 대상으로 하는 저효율 마케팅의 시대는 갔다. 이제 소비자는 익명의 존재가 아니라 자신의 고유한 이름과 취향을 가진 존재이다. 오바마는 유권자 한 사람 한 사람에게 다가갔다. 이 책을 통해서 당신은 빅데이터를 어떻게 마케팅에 활용할 수 있

는지 그 방법을 배우게 될 것이다. 즉 한 사람의 소비자가 당신의 제품을 구매할 가능성이 몇 퍼센트인지, 당신의 홍보와 권유가 그의 마음에 어느 정도까지 변화를 일으킬 수 있는지, 어떤 홍보내용과 방식이 더 많은 변화를 일으킬 수 있는지 그 노하우를 배울 수 있을 것이다.

• **정치에 관심이 있는 사람** : '당신의 한 표가 중요하다'는 말이 있듯이, 겉에서 보기에는 똑같은 표도 그것을 행사하는 사람은 모두 다르다. 한 사람 한 사람 모두 특별하고 소중하며 오바마가 표현하였듯이 '단 한 명의 낙오자도 없어야 한다'(No one left behind). 오바마의 빅데이터 선거전략을 다 읽고 나면 정치란 한 덩어리의 얼굴 없는 대중을 상대로 하는 것이 아니라 한 사람 한 사람의 시민에게 말을 거는 것이라는 것을 이해할 수 있게 된다. 빅데이터는 민주주의에 새로운 실험과 도전을 던지고 있다.

• **빅데이터의 기술적 운용에 관심이 있는 사람** : 빅데이터를 다루는 것은 단순히 하둡(Hadoop) 프로그래밍을 잘한다고 되는 것은 아니다. 거기에는 사람과 조직, 시스템과 경험이 필요하다. 이 책을 통해서 2억 명이 넘는 사람들에 대한 정보를 다루기 위해서 오바마 캠프는 어떻게 클라우드 컴퓨팅을 활용하였고, 어떻게 다양한 데이터베이스를 통합하였으며, 수백만 명의 현장 활동가들의 스마트폰으로 무엇을 실시간으로 지원하였는지, 그리고 조직의 목적과 그 목적을 실행하고 있는 사람들을 위해서 테크놀로지는 어떻게 복무해야 하는지 생생한 현장 경험담을 접할 수 있을 것이다.