

## 빅데이터의 분야별 활용사례

Big Data Use Cases of the Sector

---

저자 (Authors)	김동완 Dong Wan Kim
출처 (Source)	<a href="#">경영논총 34</a> , 2013.12, 39-52(14 pages)
발행처 (Publisher)	<a href="#">동아대학교 경영문제연구소</a> Dong-a University Business Research Center
URL	<a href="http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE02348298">http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE02348298</a>
APA Style	김동완 (2013). 빅데이터의 분야별 활용사례. 경영논총, 34, 39-52
이용정보 (Accessed)	이화여자대학교 211.48.46.*** 2020/04/29 15:33 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

## 빅데이터의 분야별 활용사례

### Big Data Use Cases of the Sector

김동완\*  
Dong Wan Kim

#### Abstract

---

스마트폰의 급격한 보급과 페이스북, 트위터, 카카오톡 등 다양한 정보채널의 등장으로 우리 주변에는 다양하고 수많은 비정형 데이터들이 만들어지고 있다.

이러한 대용량의 불확실한 데이터는 분석에 의해 가치 있는 정보로 창출되고 이를 바탕으로 능동적인 대응 및 변화를 예측하는 기술로 사용되고 있다. 그러나 빅데이터는 기존의 기술로는 처리하기 힘들 정도로 규모가 방대하고 형식이 다양하여 이의 활용에는 전문화된 IT기술을 필요로 한다. 따라서 IT강국임을 자부하는 우리나라에서 빅데이터의 광범위한 활용을 위해서는 기존의 정형화된 데이터 분석이 아닌 비정형 데이터를 분석하여 시각화 할 수 있는 창의적 사고를 가진 인력 양성이 필요하다.

본 연구는 아직 확산되지 않은 빅데이터 환경에서 각 분야별 빅데이터 선진 사례를 중심으로 빅데이터가 공공 및 기업의 이익을 위해 사용되고 있는 여러 가지 사례를 제시하였다. 또한 본 연구에서 제시된 사례를 바탕으로 우리 사회의 각 분야에서 빅데이터가 활용될 수 있을 것이다.

그러나 이 연구는 모든 자료 및 데이터를 인터넷에 공개된 자료와 보고서 및 학술정보를 이용했다는 한계점을 가지고 있다. 후속 연구에서는 빅데이터의 이용성과에 대한 실증적 자료를 바탕으로 보완한다면 빅데이터를 필요로 하는 공공 및 기업에 좋은 지침을 제공할 수 있을 것이다.

핵심주제어 : 빅데이터, 사례연구

---

---

\* 동아대학교 경영정보학과 박사수료

## I. 서론

스마트폰의 급격한 보급과 페이스북, 트위터, 카카오톡 등 다양한 정보채널의 등장으로 우리 주변에는 다양하고 수많은 비정형 데이터들이 만들어지고 있다. 트위터는 전세계 1억 명의 이용자들이 하루 평균 2억 개의 트윗을 발생시킨다. 오늘날 11억 인구가 SNS를 이용하고 있는데 2억 5000만 명이 매일 페이스북에 사진을 업로드하고 있다. 이와 같이 스마트 기기의 보편화와 무선 인터넷의 안정적인 사용이 데이터의 확산을 가속화 시키고 있다. 또한 이러한 데이터들로 인해 데이터 트래픽이 급증하고 있으며, 데이터의 크기와 형태가 다양하고 데이터가 기하급수적으로 증가하는 빅데이터 시대에 접어들고 있다. 글로벌 리서치 및 컨설팅 전문기업인 Gartner는 모바일 기기 전쟁, 하이브리드 IT 및 클라우드 컴퓨팅, 전략적 빅데이터 등 10대 기술이 향후 3년 동안 IT 업계에 상당한 영향을 미칠 것이라고 발표했다[1].

이러한 빅데이터는 기존의 데이터에 비해 너무 방대하여 일반적인 데이터베이스, 소프트웨어로는 관리하기 어려운 정형 및 비정형 데이터를 의미한다. 초기에는 수십~수천 테라바이트에 달하는 방대한 양의 데이터 집합 자체를 지칭하였으나 최근에는 기존의 방식으로 저장, 관리, 분석이 어려울 정도로 규모가 크고 순환 속도가 빠르며, 형식이 다양한 데이터 또는 이러한 데이터를 분석하는 방법을 포괄하는 용어로 통칭된다. 이러한 빅데이터는 분석에 의해 가치 있는 정보로 창출되고 이를 바탕으로 능동적인 대응 및 변화를 예측하는 기술로 사용된다[2].

빅데이터 기술의 발전은 다변화된 현대 사회를 보다 정확하게 예측하고 효율적으로 작동하도록 정보를 제공하며 개인화된 현대 사회 구성원들에게 있어서 맞춤형 정보를 제공, 관리, 분석 가능케 하며 과거에는 불가능 했던 기술을 진일보 시킨다.

최근 빅데이터는 정치, 사회, 경제, 문화, 의료와 같은 전 영역에 걸쳐 그 중요성이 부각되어 빅데이터 활용으로 주목할 만한 성과를 끌어내고 있으며, 많은 분야의 경영진들은 빅데이터의 이용 방법을 제대로 배워야 할 필요를 느끼고 있다. 그러나 빅데이터에 대한 많은 관심이 집중되어 있음에도 불구하고 실제로 빅데이터를 어떻게 활용하고 있는지 분석한 정보는 많지 않다.

이에 본 연구에서는 빅데이터의 개념을 기반으로 빅데이터가 공공 및 기업의 이익을 위해 어떠한 분야에서 어떻게 사용되고 있는가를 여러 가지 사례를 통해 제시하였다.

본 논문은 2장에서 빅데이터의 개념과 특징을 기술하였고, 3장에서는 정치, 사회, 경제, 문화, 의료분야의 빅데이터 활용사례를 기술하였다. 그리고 4장에서는 시사점을 기술하였고 마지막 장에서는 결론으로 연구의 요약 및 공헌도와 한계점을 기술하였다.

## II . 빅데이터의 개념 및 특징

### 1. 빅데이터의 개념

기존 빅데이터의 개념은 단순히 데이터의 양이 많은 것을 의미했다면 최근의 일반적인 빅데이터의 개념은 기존 데이터에 비해 너무 방대해 일반적으로 사용하는 방법이나 도구로 수집, 저장, 검색, 분석, 시각화 등을 하기 어려운 정형 또는 비정형 데이터 집합을 의미한다. 빅데이터의 정의에 합의된 바는 없지만 Manyika 외(2011)는 데이터베이스의 규모에 초점을 맞춰 일반적인 데이터베이스 SW가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터[3], Gantz 외(2011)는 데이터베이스의 업무수행에 초점을 맞춰 다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고 데이터의 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처[4]라고 정의하였다. 최근 논의되는 빅데이터는 대용량의 데이터 외에 앱, SNS 등에서 생산되는 데이터를 포함하는 개념이다[5].

빅데이터란 우리가 알지 못하는 새로운 것을 의미하는 것은 아니다. 우리가 지금까지 데이터라고 인식하지 못했던 것들까지 모두 데이터화 시키는 것이다. 기존 SNS로 대표되는 소셜 미디어의 성장과 최근 스마트폰으로 대변되는 모바일 장치의 확산이 결합되어 일상 속에서 다양한 종류의 대규모 데이터가 급속히 생성, 유통, 저장되고 있다. 또한 RFID와 같이 정보를 감지하는 센서 장비의 이용 확대와 이러한 정보를 수집하는 클라우드 컴퓨팅 기술의 확산은 물류의 이동 및 재고의 변화뿐만이 아닌 개별 소비자들의 개인정보 및 소비행태와 같은 모든 일상에 대한 디지털 기록을 가능케 하고 있다.

이러한 광범위한 데이터는 기존 민간 기업이나 정부의 데이터 분석 범위를 넘어선 것이다. 따라서 빅데이터를 기존의 시스템, 서비스, 기업 등에서 주어진 비용이나 시간 내에 처리 가능한 데이터 범위를 넘어서는 데이터[6]라고 보는 시각이 대부분이다.

이러한 데이터는 정형화된 데이터, 반정형화된 데이터, 비정형 데이터로 구분할 수 있다. 정형화된 데이터는 일정한 규칙을 갖고 체계적으로 정리된 데이터를 의미한다. 예를 들면 매년 통계청에서 발표하는 통계자료, 방송통신 실태조사, 각종 과학적 데이터 등이 이에 해당된다. 정형화된 데이터는 그 자체로 의미 해석이 가능하며, 바로 활용할 수 있는 정보를 내포하고 있다. 반정형화된 데이터는 한글이나 MS Word 등으로 작성도니 데이터를 의미한다. 대표적인 예가 인쇄매체의 텍스트라 할 수 있다. 반정형화된 데이터는 표나 그림이 될 수도 있지만 일반적으로 문자로 서술된 정보를 담고 있다. 비정형화된 데이터는 스마트 기기 등을 통해서 생성되는 데이터로 페이스북, 트위터, 카카오톡 등으로 상호 교류되는 정보가 이에 해당한다. 비정형화된 데이터는 개인, 집단, 사회, 국가 등과 관련된 주제를 스마트 기기 이용자들이 상호 의견을 교류함으로써 생성되는 정보들이다. 오늘날 빅데이터는 비정형화된 데이터에 관심을 두고 있다. 기업 및 사회는 이러한 수집된 데이터를 기반으로 예측 분석을 하기 위하여 다양한 종류의 대규모 데이터 처리, 분석 및 활용 기술을 필요로 하고 있다. 대량의 다양한 데이터 생산 기술의 진보와 이에 필요한 데이터 저장, 관리, 분석 기술

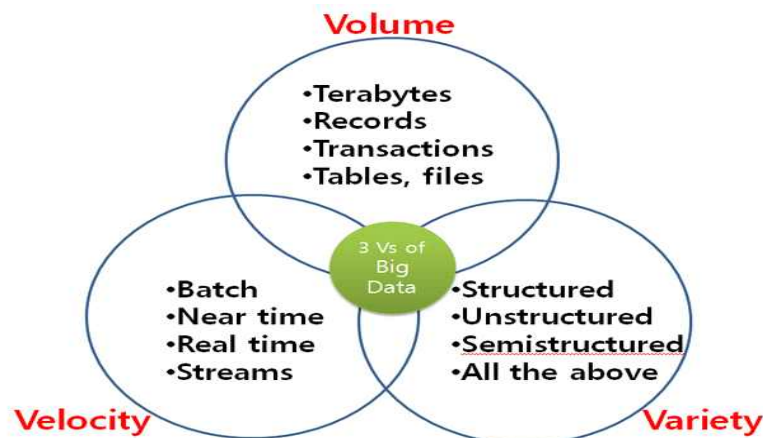
의 발전 속에서 빅 데이터가 출현하게 된다.

## 2. 빅데이터의 특징

빅데이터는 TB(테라바이트)단위의 데이터 크기로 정의되거나 데이터 수집 및 분석에 장기적인 시간을 요하므로 데이터 크기의 증가를 그 특징으로 한다. 그러나 단순한 데이터 크기의 증가를 넘어서서 빅데이터는 크게 데이터 크기(volume), 데이터 속도(velocity), 그리고 데이터 형태(variety) 등 세 가지 요소의 복합적인 변화를 그 특징으로 한다[7]. 또한 빅데이터의 활용을 위한 3대 요소에는 자원, 기술, 인력이 있다[8].

### 2.1 빅데이터의 3대 요소

빅데이터의 요소에는 [그림 1]과 같이 크기, 속도, 형태가 있다.



[그림 1] 빅데이터의 요소 <출처: 구글이미지>

#### 2.1.1 데이터의 크기(Volume)

단순 저장되는 물리적 데이터 크기의 증가뿐만이 아닌 이를 분석 및 처리하는 데 어려움이 따르는 네트워크 데이터의 급속한 증가는 빅데이터의 가장 기본적인 특징이다[7].

#### 2.1.2 데이터의 속도(Velocity)

빅데이터는 데이터의 실시간 처리 및 장기적 접근을 요구한다. 데이터 생산 및 유통, 수집 및 분석 속도의 증가와 이에 대한 실시간 처리 및 장기간에 걸쳐 데이터를 수집, 분석하는 장기적 접근이 빅 데이터의 속도적 특성이다[7].

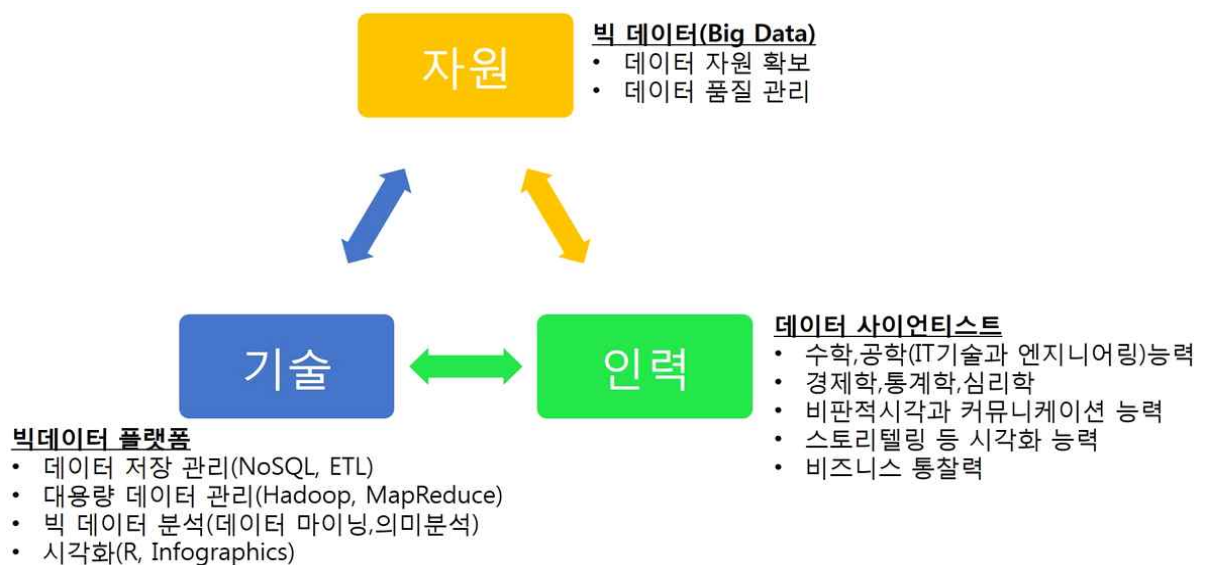
#### 2.1.3 데이터의 형태(Variety)

기존 예측 분석에서 데이터 분석은 기업 내부에서 발생하는 운영 데이터인 ERP, SCM, MES(Manufacturing execution system), CRM 등의 시스템에 저장되어 있으며 잘 정제되어 있고 의미도 명확한 RDBMS(관계형 데이터베이스) 기반의 정형 데이터를 통해 이뤄졌다 [8].

최근 빅데이터를 이용한 데이터 분석은 고정된 시스템에 저장되어 있지 않은 XML, HTML 등과 같이 데이터베이스 스키마를 포함하는 반정형 데이터를 이용한 분석뿐만이 아닌 사진, 오디오, 비디오 형식의 소셜 미디어 데이터나 로그파일(Database log) 같이 비정형 데이터도 처리할 수 있는 능력을 요구한다. 즉 빅데이터의 성장이란 단순히 데이터의 크기가 증가하는 것을 넘어서서 다양한 형태의 데이터 크기가 증가하는 것을 의미하는 것이다 [7].

## 2.2 빅데이터의 활용을 위한 3대 요소

빅데이터의 활용을 위한 요소에는 [그림 2]와 같이 자원, 기술, 인력이 있다.



[그림 2] 빅데이터 활용 요소 <출처: 구글이미지>

### 2.2.1 자원

미래에는 빅데이터를 핵심 자원으로 인식하고 필요한 정보를 뽑아낼 수 있도록 리소스를 키우는 것이 중요하다. 그러므로 주어진 빅데이터를 관리·처리하는 측면과 함께 활용할 수 있는 외부 빅데이터 자원을 발견하고 확보하는 전략이 필요하다. 또한 데이터의 품질은 데이터 활용 결과에 중대한 영향을 미치므로 데이터 관리체계 및 데이터의 신뢰성 확보가 매우 중요하다[9].

### 2.2.2 기술

빅데이터는 데이터 자체뿐만 아니라 관련 도구나 플랫폼, 분석기법까지 포괄하는 용어로 의미가 확장되며 IT패러다임의 변화를 견인하고 있다. 성장을 촉진하는 신기술에 대한 이해가 부족하면 미래 경쟁력 강화를 위한 기회 포착에도 어려움을 겪을 수 있다. 그러므로 조직과 기업의 혁신 전략으로 적용할 수 있도록 빅데이터 플랫폼, 빅데이터 분석기술 및 데이터 분석기법에 대한 이해가 필요하다[9].

### 2.2.3 인력

신기술과 툴이 아무리 뛰어나도 실제 성과를 낼 수 있는 것은 이를 활용하여 적용하는 사람의 역량에 좌우된다. 데이터 처리와 분석능력을 갖춘 인력은 IT뿐만 아니라 대부분의 기업과 조직에서 필수적으로 확보해야 할 핵심인력으로 조직의 차원에서 인재를 확보하기 위해 내부 역량 강화 및 외부 협력이 중요하다[9].

## III. 빅데이터 활용 사례연구

빅데이터는 기술 분야뿐 아니라 우리 사회의 다양한 분야에서 여러 가지 방법으로 사용되고 있다. 특히 빅데이터는 기업들이 오랫동안 고민해 온 난제에 대한 해결책을 제공해 줄 뿐 아니라, 프로세스와 조직, 산업 전반, 심지어 사회 자체를 변화시킬 수 있는 새로운 방법까지 제시하고 있다.

본 연구에서는 각 분야에서 광범위하게 사용되어 성과를 내고 있는 빅데이터의 선진사례 연구를 바탕으로 빅데이터 활용사례를 연구한다. 사례연구는 다양한 동기에 의해 활용될 수 있는데, 특히 어떠한 상황이나 환경에 대한 구체적인 과정을 이해하는데 효과적으로 활용할 수 있는 연구방법이다[10]. 하지만 사례연구는 일반화부분에서 약하다는 단점이 있지만 현상에 대한 이해를 바탕으로 이론을 이해하고 정립할 수 있다는 장점이 있기도 하다[11]. 따라서 본 연구에서는 빅데이터 활용의 초기 단계인 정치, 사회, 경제, 문화, 의료분야에서 빅데이터를 활용하여 새로운 가치가 창출된 분야에 대한 탐색적 연구이기 때문에 사례연구방법이 적절하다고 판단하였다.

빅데이터는 우리가 살아가는 모든 영역에서 발생하는 데이터들을 기반으로 한다. 이러한 방대한 디지털 흔적인 데이터를 분석하여 사회의 움직임을 예측하여 새로운 가치를 창출하는 빅데이터 기술이 성과를 내고 있다. 본 장에서는 정치, 사회, 경제, 문화, 의료 분야의 빅데이터 활용사례를 기술한다.

### 1. 정치 분야 활용사례

2008년 미국 대통령 선거에서 버락 오바마 미국 대통령 후보는 다양한 형태의 유권자 데이터베이스를 확보하여 이를 분석, 활용한 유권자 맞춤형 선거 전략을 전개했다. 당시 오바마 캠프는 인종, 종교, 나이, 가구형태, 소비수준과 같은 기본 인적 사항으로 유권자를 분류하는 것을 넘어서서 과거 투표 여부, 구독하는 잡지, 마시는 음료 등 유권자 성향까지 전화나 개별 방문을 또는 소셜 미디어를 통해 유권자 정보를 수집하였다. 수집된 데이터는 오바마 캠프 본부로 전송되어 유권자 데이터베이스를 온라인으로 통합 관리하는 보우트빌더(VoteBuilder.com)시스템의 도움으로 유권자 성향 분석, 미결정 유권자 선별, 유권자에 대한 예측을 해나갔다. 이를 바탕으로 유권자 지도를 작성한 뒤 유권자 맞춤형 선거 전략을 전개하는 등 오바마 캠프는 비용 대비 효과적인 선거를 치를 수 있었다[12].

대한민국 중앙선거관리위원회는 제19대 총선부터 소셜 네트워크 등 인터넷 상의 선거 운동을 상시 허용하였다[13]. 이에 소셜 미디어 상에서 선거 관련 데이터는 증폭되었으며 2010년 대한민국 제5회 지방 선거 및 2011년 대한민국 재보궐선거에서 SNS의 중요성을 확인한 정당들 또한 SNS 역량 지수를 공천 심사에 반영하는 등[14] 소셜 네트워크 활용에 주목했다. 이 가운데 여론 조사 기관들은 기존 여론조사 방식으로 예측한 2010년 제5회 지방 선거 및 2011년 재보궐선거의 여론조사 결과와 실제 투표 결과와의 큰 차이를 보완하고자 빅 데이터 기술을 활용한 SNS 여론 분석을 시행했다. 그러나 SNS 이용자의 대다수가 수도권 20~30대에 쏠려 있기에[15], 빅 데이터를 이용한 대한민국 제19대 총선에 대한 SNS 분석은 수도권으로 한정되어 일치하는 한계를 드러내기도 했다.

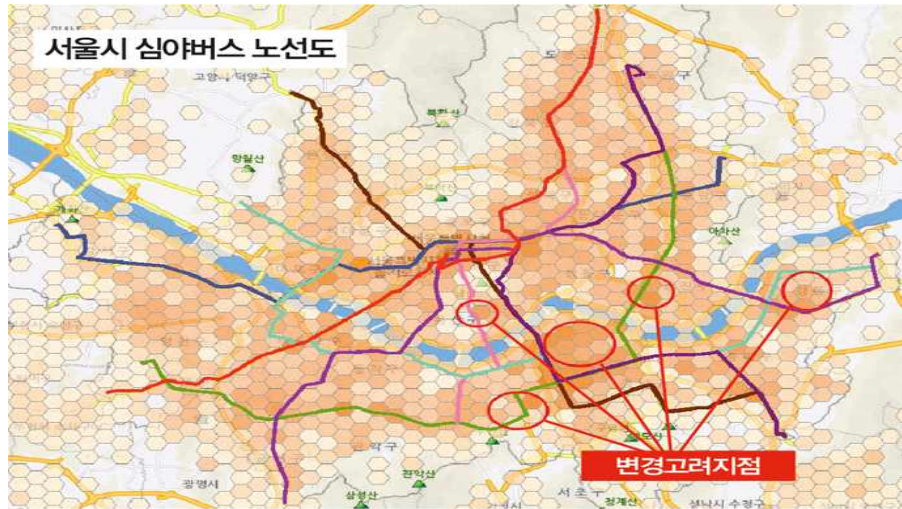


[그림 3] 오바마 대선용 모바일 앱 <출처: fastcompany.com>

## 2. 사회 분야 활용사례

서울시는 심야버스 노선을 확정하는데 빅데이터를 활용하였다. 먼저 KT와 협조하여 심야에 전화를 이용하는 사람들의 정보를 분석하였다. 발신자 위치(출발지)와 번호 소유자의 주소지(도착지)를 연결하여 심야에 이동해야 할 길을 찾아 낸 후, 이 데이터를 바탕으로 서울시는 새로운 노선을 확정하였다[16].





[그림 4] 서울시 심야버스 노선개발 상세화면

보스턴시는 수많은 도로에서 발생하는 도로파손은 예측하기도 힘들고 이를 직접 눈으로 확인하고 정보를 수집하는데 소요되는 인력과 예산을 절감하기 위해 스마트폰 어플리케이션을 통해 GPS와 센서 등을 활용하여 자동으로 도로파손을 감지하는 시스템을 활용하였다. 미국의 벤처회사에서 개발된 스트리트 범프(Street Bump)라는 모바일 앱을 시민들에게 무료로 제공하여 운전자가 직접 주변 환경의 데이터를 수집하고 전송하는 인간 센서 역할을 하게 하였다. 스트리트 범프는 차가 도로가 파손된 곳을 지나가게 되면 스마트폰의 GPS와 센서 등을 활용해 진동을 감지하게 되며, 진동이 감지될 경우 이를 도로관리국 도로정보 수집서버에 전송하고, 수집된 정보는 시청의 인터랙티브 지도상에 파손 위치 등이 기록되어 시는 곧바로 도로파손 보수를 실시하게 된다. 실시간으로 수집되는 막대한 양의 도로노면 정보를 기반으로 한 빅데이터 분석은 신속한 도로 유지보수 및 인력과 예산을 절감하였다 [17].



[그림 5] Street Bump 모바일 앱 버전 <출처: streetbump 홈페이지>

브라질에서 2011년 초 집중호우 때 발생한 산사태로 인해 발생한 대형 인명피해는 세계적인 뉴스가 되었다. 이 사건을 계기로 2014년 월드컵과 2016년 올림픽을 준비하는 리우데

자네이루시는 새로운 도시환경을 구현하기 위한 데이터 분석을 위해 도시 내 30여 개 기관의 정보와 프로세스를 단일체제로 통합해 자연재해, 교통, 전력공급 등을 24시간 실시간 모니터링하는 시스템을 도입하였다. 이후 새로운 경보시스템은 비상사태 발생 시 통보를 수작업으로 전달하던 과거 시스템과는 달리 자동 이메일 통보나 문자메시지 등을 즉각적으로 알리게 되었다. 이 시스템을 통해 응급상황에 대한 대응시간은 약 30% 정도 개선된 것으로 나타났다[18].



[그림 6] 리우데자네이루 지능형통합운영센터

### 3. 경제 및 경영 분야 활용사례

아마존닷컴은 모든 고객들의 구매 내역을 데이터베이스에 기록하고 이 기록을 분석해 소비자의 소비 취향과 관심사를 파악한다[19]. 이런 빅데이터의 활용을 통해 아마존은 고객별로 추천 상품(레코멘데이션)을 표시한다. 고객 한 사람 한 사람의 취미나 독서 경향을 찾아 그와 일치한다고 생각되는 상품을 메일, 홈 페이지상에서 중점적으로 고객 한 사람 한 사람에게 자동적으로 제시하는 것이다[20]. 아마존닷컴의 추천 상품 표시와 같은 방식으로 구글 및 페이스북도 이용자의 검색 조건, 나아가 사진과 동영상 같은 비정형 데이터 사용을 즉각 처리하여 이용자에게 맞춤형 광고를 제공하는 등 빅데이터의 활용을 증대시키고 있다.

인터넷으로 시작해서 인터넷으로 마감하는 생활, 스마트폰을 이용해 정보를 검색하고 쇼핑도하고 SNS를 이용해서 실시간으로 글을 남기는 등의 다양하게 인터넷을 이용하는 동안 남는 흔적같은 모인 데이터들을 분석하면 개인의 생활 패턴, 소비성향 등을 예측할 수 있고 기업들은 이런 데이터를 통해서 소비자가 원하는 것들을 미리 예측할 수 있다. 빅 데이터가 마케팅 자료로 활용되는 사례이다[21].

영국의 보험회사 아비바(AVIVA)는 주행거리 분석을 통한 합리적인 자동차 보험료를 제시하여 고객으로부터 좋은 반응을 얻고 있다. 혼잡 시간대와 사고 다발 지역의 운행 빈도가 낮은 운전자에게 보험료를 할인해 주는 주행거리연동보험은 기존 보험업의 관행을 깬 접근으로 고객으로부터 좋은 반응을 얻고 있다. 이는 차량 내 부착된 운행 기록 장치를 통해 실

제 운전 행태가 보험사에 전송되어 이러한 자료 분석을 통해 체계적인 보험료가 산정되는 방식이다. 또한 대부분의 보험사가 꺼려하는 고위험 운전자의 세분화 및 위험 수준에 근거한 자동차 운전자의 등급 산정을 통해 가입자별로 정확한 리스크 관리 수행이 용이하다 [22].



[그림 7] 자동차 주행에 따른 보험료 계산 <출처: BBC News, 2004>

#### 4. 문화 분야 활용사례

머니볼 이론이란 경기 데이터를 철저하게 분석해 오직 데이터를 기반으로 적재적소에 선수들을 배치해 승률을 높인다는 게임 이론이다[23]. 이는 미국 메이저 리그 베이스볼 오클랜드 어슬레틱스의 구단장 빌리 빈이 리그 전체 25위에 해당하는 낮은 구단 지원금 속에서도 최소비용으로 최대효과를 거둔 상황에서 유래되었다. 빌리 빈은 하버드대에서 경제학을 전공한 폴 데포데스터를 영입하여 타율, 타점, 홈런 등 흥행 요소만을 중시하던 야구계에서 출루율, 장타율, 사사구 비율이 승부와 관련되어 있음을 간파하고 데이터를 수집, 분석, 활용하였다. 이를 통해 빈은 좋은 선수를 발굴하고 이들을 적재적소에 배치해 최하위에 그치던 팀을 4년 연속 포스트시즌에 진출시키고 메이저리그 최초로 20연승이라는 신기록을 세우도록 탈바꿈 시켰다. 미국 월스트리트 저널은 미국 경제에 큰 영향을 끼치는 파워 엘리트 30인에 워렌 버핏, 앨런 그린스펀과 함께 빌리 빈을 선정[24]하는 등 머니볼 이론은 경영, 금융 분야에서도 주목받았다. 최근 들어 과학기술 및 카메라 기술의 발달로 더욱 정교한 데이터의 수집이 가능해졌으며 투구의 궤적 및 투수의 그림, 타구 방향, 야수의 움직임까지 잡아낼 수 있게 되었다. 류현진(LA 다저스)이 숙적 샌프란시스코 자이언츠를 상대로 시즌 14승을 따냈던 2013년 10월 25일 샌프란시스코 AT&T 파크 경기장. 5회말 원아웃에 스코어는 1대1, 원볼 투스트라이크 상황에서 류현진이 힘차게 71번째 공을 뿌린다. 91마일짜리 패스트볼이 뱀처럼 꿈틀거리며 홈플레이트 바깥을 타고 쏜살같이 지나간다. 불인지 스트라이

크인지 쉽게 분간을 할 수 없다. 그 순간 주심의 팔이 한 치 망설임 없이 힘차게 올라간다. 결과는 삼진 아웃.

이 극적인 순간 이면에는 빅데이터 분석이 있었다. 경기 전 포수 AJ 엘리스는 더그아웃에 붙은 구심 스트라이크존 분석 자료를 외우다시피 들여다봤다. 좌타자인지 우타자인지, 투스트라이크 이전인지 이후인지 등 모든 상황을 가정해 분석한 스트라이크존이 확률과 통계의 힘을 빌려 붉은색과 푸른색으로 표시를 하고 있다. 붉은색은 스트라이크 가능성이 높은 영역, 푸른색은 그 반대다. 오늘의 구심이 판정한 수백 수천 번의 경기 자료를 토대로 나온 자료다[25]. 이처럼 기존의 정형 데이터뿐만 아닌 비정형 데이터의 수집과 분석, 활용을 통해 최근 야구경기에서 빅 데이터의 중요성은 더욱 커지고 있다.

## 5. 의료 분야 활용사례

빅 데이터를 활용하면 미국 의료부문은 연간 3,300억 달러(미 정부 의료 예산의 약 8%에 해당하는 규모)의 직간접적인 비용 절감 효과를 보일 것으로 전망했다[3]. 특히 임상분야에서는 의료기관 별 진료방법, 효능, 비용 데이터를 분석하여 보다 효과적인 진료방법을 파악하고 환자 데이터의 온라인 플랫폼화로 의료협회 간 데이터 공유로 치료 효과를 제고하며 공중보건 영역에선 전국의 의료 데이터를 연계하여 전염병 발생과 같은 긴박한 순간에 빠른 의사결정을 가능케 할 전망이다[26].

미국 국립암연구소는 암 환자에 대한 빅데이터를 분석해 암 발생의 원인을 밝혔다고 한다. 또한 환자의 상태와 가족력 등을 분석하여 이를 고려한 식단, 운동, 치료가 가능한 이른바 맞춤형 의료가 가능할 것이라고 한다. 빅데이터와 같은 ICT가 의료 분야에 응용되면서 의료의 패러다임이 바뀌고 있다. 그 동안 의료는 병에 걸리면 치료를 한다는 치료의 개념이 강했지만 빅데이터를 활용하면 질병을 원천적으로 차단하는 예방 의료가 가능해진다.

## IV. 시사점

빅데이터 기술의 사용으로 성과를 내고 있는 정치, 사회, 경제, 문화, 의료 분야 빅데이터 활용사례의 시사점은 다음과 같이 정리할 수 있다.

첫째, 모든 분야에서 빅데이터 활용에 성공하기 위해서는 양질의 데이터 획득이 중요함을 시사한다. 미국은 전화나 개별방문 또는 소셜 미디어를 통해 수집된 유권자 정보를 보우트 빌더(VoteBuilder.com) 시스템의 도움으로 유권자 성향 분석, 미결정 유권자 선별, 유권자에 대한 예측으로 비용 대비 효과적인 선거를 치를 수 있었다. 대한민국은 중앙선거관리위원회가 제19대 총선인 2012년 4월 11일부터 SNS 등 인터넷 상의 선거운동을 상시 허용함에 따라 소셜 미디어 상에서 선거 관련 데이터가 증가되었다. 이후 각 정당들은 SNS의 중요성을 인식하고 SNS 역량지수를 공천 심사에 반영하는 등 SNS 활용에 주목했으나 당시의 20~30

대(90.3%)와 60대 이상(9.5%) 연령층간 스마트폰 보유율 격차가 80.8%p로[27] 빅데이터를 이용한 SNS 분석은 20~30대로 한정되어 일치하는 한계를 드러내었다. 데이터 양이 많고 분석 기술이 뛰어나더라도 데이터의 품질이 낮으면 좋은 정보를 얻기가 어려움을 의미한다.

둘째, 확보된 데이터의 필요성과 가치를 중시하는 공감대 형성과, 사업 진행 중에 생산되는 데이터를 단순하게 생각하지 않고 적극적으로 활용하려는 자세가 중요함을 시사한다. 서울시는 KT의 협조로 심야 전화 이용자의 정보를 분석하여 가장 효율성이 높은 버스 노선을 확정할 수 있었다. 보스턴시는 실시간으로 수집되는 도로노면 정보를 기반으로 도로 유지보수 및 인력과 예산을 절감하였다.

셋째, 각 사업부처별로 흩어져 있는 데이터들을 전사적 차원에서 관리하기 위한 데이터 통합이 필요함을 시사한다. 브라질 리우데자네이루시는 도시 내 30여 개 정보와 프로세스를 단일체제로 통합하는 지능형통합운영센터의 운영으로 응급상황에 대한 대응시간을 약 30% 정도 개선시켰다.

넷째, 개인정보보호 환경이 마련되어야 함을 시사한다. 빅데이터의 많은 부분을 차지하는 비정형데이터는 스마트폰이나 소셜미디어를 기반으로 한 개인의 다양한 정보를 포함하고 있다. 그러므로 고객정보 수집 및 활용에 앞서 정보 제공의 권한 및 책임 문제와 소유권, 개인정보보호, 내부정보유출, 보안 등에 관한 내부 가이드라인 등이 필요하다. 또한 개인의 프라이버시는 충분히 보호하면서도 빅데이터 활용을 통해 생활의 편의를 증진시킬 수 있다는 공감대 형성이 필요하다.

## V. 결론

스마트폰의 급격한 보급과 페이스북, 트위터, 카카오톡 등 다양한 정보채널의 등장으로 우리 주변에는 다양하고 수많은 비정형 데이터들이 만들어지고 있다. 이른바 데이터의 크기와 형태가 다양하고 증가 속도가 가파른 빅데이터 시대에 접어든 것이다. 이러한 대용량의 불확실한 데이터는 분석에 의해 가치 있는 정보로 창출되고 이를 바탕으로 능동적인 대응 및 변화를 예측하는 기술로 사용되고 있다. 그러나 빅데이터는 기존의 기술로는 처리하기 힘들 정도로 규모가 방대하고 형식이 다양하여 이의 활용에는 전문화된 IT기술을 필요로 한다. 따라서 IT강국임을 자부하는 우리나라에서 빅데이터의 광범위한 활용을 위해서는 기존의 정형화된 데이터 분석이 아닌 비정형 데이터를 분석하여 시각화 할 수 있는 창의적 사고를 가진 인력 양성이 필요하다.

본 연구는 아직 확산되지 않은 빅데이터 환경에서 각 분야별 빅데이터 선진 사례를 중심으로 빅데이터가 공공 및 기업의 이익을 위해 사용되고 있는 여러 가지 사례를 제시한 연구로 이와 관련된 후속연구를 유발할 것으로 사료된다. 또한 본 연구에서 제시된 사례를 바탕으로 우리 사회의 각 분야에서 빅데이터가 활용될 수 있을 것이다.

그러나 이 연구는 모든 자료 및 데이터를 인터넷에 공개된 자료와 보고서 및 학술정보를 이용했다는 한계점을 가지고 있다. 후속 연구에서는 빅데이터의 이용성과에 대한 실증적

자료를 바탕으로 보완한다면 빅데이터를 필요로 하는 공공 및 기업에 좋은 지침을 제공할 수 있을 것이다.

## 참 고 문 헌

- [1] NIA(2013), “2013년 IT기술 트렌드 방향과 쟁점”, 동향분석 I-6, pp. 4-6.
- [2] 강만모, 김상락, 박상무(2012), “빅데이터의 분석과 활용”, 정보과학회지, 제 30권, 제 6호, pp. 25-32.
- [3] James M. & Michael C.(2011), “Big data: The next frontier for innovation, competition, and productivity,” McKinsey Global Institute, pp. 1-36.
- [4] John G. & David R.(2011), “Extracting Value from Chaos,” IDC IVIEW June, pp. 6.
- [5] 채승병(2011), “정보홍수 속에서 금맥 찾기: 빅데이터 분석과 활용”, SERI경영노트, pp. 28-47.
- [6] 김형준(2012), “빅데이터란 무엇이고, 어떻게 해야 할까?” <http://www.jaso.co.kr/451>.
- [7] Philip R.(2011), “Big Data Analytics”, TDWI Research Fourth Quarter, pp. 6.
- [8] 정병권 외 2명(2012), “미래사회와 빅데이터 기술”, 정보통신산업진흥원, pp. 13-14.
- [9] 정지선(2012), “성공적인 빅데이터 활용을 위한 3대 요소”, 한국정보화진흥원 연구보고서.
- [10] Yin, R. K.(2009), “Case Study Research-Design and Methods”, Fourth Edition, Applied Social Research Method Series, Vol. 5.
- [11] Cutler, A.(2004), “Methodical Failure: The Use of Case Method by Public Relations Researchers”, Public Relations Review, Vol. 30, pp. 364-365.
- [12] 차재필(2012), “빅데이터 시대의 국민공감 선거전략-미 대선사례를 중심으로”, 한국정보화진흥원.
- [13] 연합뉴스(2012), “선관위 인터넷 선거운동 상시 허용 결정(종합2보)”, <http://www.yonhapnews.co.kr/politics/2012/01/13/0502000000AKR20120113116600001.HTML>
- [14] 연합뉴스(2012), “새누리당 SNS 역량지수 주중 공천위 전달”, <http://www.yonhapnews.co.kr/bulletin/2012/02/21/0200000000AKR20120221198700001.HTML?sns>
- [15] 경향신문(2012), “소셜 여론의 총선 예측 실패는 이용자 수도권 집중 때문”, [http://news.khan.co.kr/kh\\_news/khan\\_art\\_view.html?artid=201204242003055&code=210100](http://news.khan.co.kr/kh_news/khan_art_view.html?artid=201204242003055&code=210100).
- [16] 서울뉴스1(2013), “서울시 빅데이터로 심야버스 노선 구축”, <http://news1.kr/articles/1219217>.
- [17] 세계도시동향(2012), “스마트폰 앱을 활용한 도로포장 파손위치 자동인식시스템 운영 및 신호장제 기술개발 지원”, 서울연구원.

- [18] 윤미영(2013), “더 나은 미래를 위한 데이터 분석”, 한국정보화진흥원, pp. 32-34.
- [19] 장영재(2012), “아마존닷컴, 현대의 서점 아저씨”, 비즈니스북스, pp. 119.
- [20] 데이코산업연구소(2010), “글로벌 전자책시장 실태와 관련 산업 동향(급성장하는 MID 단말 시장)”, 데이코산업연구소출판부.
- [21] 연합뉴스 기획취재팀(2013), “빅데이터 사회”, 연합뉴스(월간마이더스)
- [22] 김철원, 박기찬, 김승윤, 이선미(2012), “불확실한 경영환경 속 빛나는 글로벌 기업의 분석기반 혁신사례”, KT경제경영연구소.
- [23] 머니볼(Moneyball) 이론(2012), “에듀윌 정보통신/오늘의 일반상식”, 평생교육 No.1 에듀윌 [http://blog.eduwill.net/1079#comment\\_area](http://blog.eduwill.net/1079#comment_area)
- [24] The Wall Street Journal(2003), “Those Who Influence The Markets Most”, <http://online.wsj.com/news/articles/SB106843378249003800>
- [25] MK뉴스(2013), “류현진 돌풍은 빅데이터의 힘”, <http://news.mk.co.kr/newsRead.php?year=2013&no=1031744>
- [26] 이성춘, 임양수(2011), “Big Data, 미래를 여는 열쇠”, KT경제경영연구소, pp. 12.
- [27] NIA(2013), “2012 신 디지털 격차 현황 분석 및 제언”, 정보격차지수 및 실태조사 요약 보고서, pp. 31.
- [28] 빅데이터의 요소, 구글이미지 <2013.11.26. 검색>  
[https://www.google.co.kr/search?q=%EB%B9%85%EB%8D%B0%EC%9D%B4%ED%84%B0%EC%9D%98+%EC%9A%94%EC%86%8C&newwindow=1&rlz=1C2WPDB\\_enKR502KR518&source=lnms&tbm=isch&sa=X&ei=88GWUoniDsmyiAej-4CwCA&ved=0CAcQ\\_AUoAQ&biw=1920&bih=975](https://www.google.co.kr/search?q=%EB%B9%85%EB%8D%B0%EC%9D%B4%ED%84%B0%EC%9D%98+%EC%9A%94%EC%86%8C&newwindow=1&rlz=1C2WPDB_enKR502KR518&source=lnms&tbm=isch&sa=X&ei=88GWUoniDsmyiAej-4CwCA&ved=0CAcQ_AUoAQ&biw=1920&bih=975)
- [29] 빅데이터 활용 요소, 구글이미지 <2013.11.26. 검색>  
[https://www.google.co.kr/search?q=%EB%B9%85%EB%8D%B0%EC%9D%B4%ED%84%B0%EC%9D%98+%EC%9A%94%EC%86%8C&newwindow=1&rlz=1C2WPDB\\_enKR502KR518&source=lnms&tbm=isch&sa=X&ei=88GWUoniDsmyiAej-4CwCA&ved=0CAcQ\\_AUoAQ&biw=1920&bih=975](https://www.google.co.kr/search?q=%EB%B9%85%EB%8D%B0%EC%9D%B4%ED%84%B0%EC%9D%98+%EC%9A%94%EC%86%8C&newwindow=1&rlz=1C2WPDB_enKR502KR518&source=lnms&tbm=isch&sa=X&ei=88GWUoniDsmyiAej-4CwCA&ved=0CAcQ_AUoAQ&biw=1920&bih=975)
- [30] 오바마 대선용 모바일 앱, <http://www.fastcompany.com/>
- [31] Street Bump 모바일 앱 버전 [www.streetbump.org/](http://www.streetbump.org/)
- [32] BBC News(2004), “Could pay-as you-drive insurance work?”  
[http://news.bbc.co.uk/2/hi/uk\\_news/magazine/3574010.stm](http://news.bbc.co.uk/2/hi/uk_news/magazine/3574010.stm)