

미래 인재로의 도약, 스마트기술 기반 다지기	
05차시	빅데이터의 개념과 활용

## 1. 빅데이터의 이해

빅데이터(Big Data)란 대량의 정형 또는 비정형 데이터 세트 및 이러한 데이터라는 의미로 출발했지만, 이러한 대용량 데이터를 활용·분석하여 가치있는 정보를 추출하고 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술이라고 의미로 확대되었다. 즉 빅데이터란 기존의 관리 및 분석 체계로는 감당할 수 없을 정도의 거대한 데이터의 집합으로 대규모 데이터와 관계된 기술 및 도구(수집·저장·검색·공유·분석·시각화 등)를 모두 포함하는 개념인 것이다.

### 1) 빅데이터의 등장

#### 가. 빅데이터 등장 배경

인터넷이 일상화된 최근 10여년 사이 인류는 디지털 데이터가 폭증하는 데이터 홍수(data deluge) 현상에 직면하게 되었다. 2007년부터 전 세계적으로 생성된 디지털 정보량이 사용 가능한 저장 공간을 초과하기 시작하였으며 정보량이 기하급수적으로 증가하여 2030년에는 관리해야 할 정보량이 현재보다 30배 급증하고 10배 많은 서버가 필요할 것으로 전망되었다.

기술발전에 따른 데이터 저장 및 처리비용의 하락, 소셜 네트워크 서비스의 확대 등으로 막대한 데이터 폭발은 가속화되고 있으며 앞으로 도로, 건축물 등에 내장된 임베디드 시스템(embedded system)과 사물인터넷에서도 막대한 데이터가 생성될 것이다. 이와 같이 폭발적으로 증가하는 대규모 데이터가 중대 이슈로 부각되면서 '빅데이터(Big Data)'라는 용어가 등장하게 된 것이다.

디지털 기술 인프라의 고도화와 성숙에 따라 새로운 투자처에 대한 수요(needs)가 증가하면서 새로운 공급 체계와 시장을 창출할 빅데이터에 대한 관심 역시 증가하고 있다. 또한 기업은 그간 축적된 자산 데이터(asset data)와 소셜 네트워크에 기반한 대규모 외부 데이터의 활용으로 혁신과 기회 포착을 기대하기 위해 빅데이터에 관심을 기울이기 시작했다.

#### 나. 빅데이터의 개념

빅데이터(Big Data)는 새로운 개념이 아니다. 이는 1990년 이후 인터넷이 확산되면서 정형화된 정보와 비정형 형태의 정보가 무수히 발생하게 되면서 정보홍수(information overload)나 정보폭발(information explosion)이라는 개념으로 논의되었고, 오늘날 '빅데이터'라는 개념으로 이어지게 된 것이다. 그동안 인터넷에서 발생한 수많은 정보는 인터넷 서비스 기업이 보관하거나 일부 상업적으로 이용되기도 하였다. 더불어 모바일 스마트 기기의 확산으로 개인과 관련된 비정형 데이터가 축적되면서 데이터는 더욱 증가하게 된다. 특히 소셜 미디어의 활성화는 공적인 정보뿐만 아니라 사적인 정보까지 증가하게 함으로써 빅데이터의 서막

을 알리는 계기가 되었다.

국가전략위원회에서는 '대용량 데이터를 활용·분석하여 가치있는 정보를 추출하고 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화기술'이라고 정의하였다. 또한 삼성경제연구소는 '빅데이터란 기존의 관리 및 분석체계로는 감당할 수 없을 정도의 거대한 데이터의 집합으로 대규모 데이터와 관계된 기술 및 도구(수집·저장·검색·공유·분석·시각화 등)를 모두 포함하는 개념'으로 정의하였다. 이와 같은 정의를 살펴볼 때 빅데이터란 엄청나게 많은 데이터이지만 양적인 의미를 벗어나 데이터 분석과 활용을 포괄하는 개념으로 사용하고 있다. 빅데이터의 정의는 데이터 규모와 기술 측면에서 출발했으나 빅데이터의 가치와 활용 효과 측면으로 의미가 확대된 것이다.

#### 다. 빅데이터의 영향

빅데이터는 가치 창출 방식에 있어 경제 부문, 공공 부문, 사회 부문에 영향을 미치고 있으며, 그 영향은 다양하게 나타날 수 있다. 빅데이터를 통해 창출된 가치는 산업과 기업에게 있어서는 혁신의 원동력을 제공할 수 있고 이를 통해 경쟁력을 향상시킬 수가 있으며, 나아가 생산성의 향상을 통해 기업 전반의 효율성을 높여주는 역할을 한다. 정부 및 지자체에게는 빅데이터를 통해 국가의 전반적인 환경을 탐색하고 이를 통해 새로운 가치를 분석하여 기존의 방식으로는 해결을 하지 못했던 다양한 문제점들을 쉽고 빠르게 대응할 수 있도록 해준다.

첫째, 기업에서는 빅데이터를 활용하여 소비자의 행동 패턴을 분석하고 이를 통해 시장의 동향을 예측하여 기업이 기존에 가지고 있던 비즈니스 프로세스의 문제점을 파악하고 개선 및 수정을 통해 기업의 비즈니스에 도움을 줄 수 있다. 기존에 기업에서 활용하고 있는 ERP, SCM, MES 등으로부터 생성되는 비즈니스 전반의 데이터 활용으로 원가 절감, 제품 차별화가 가능해졌다. 과거에는 생산 과정에서 발생하는 데이터의 양이 적었을 뿐만 아니라 역량 부족으로 인해 제대로 활용되지 않고 버려지곤 했으나, 빅데이터 관련 기술의 발달에 힘입어 손실을 줄이고 생산성을 높이는 데 데이터가 활용될 수 있게 되었다.

세계 1위의 풍력 터빈 제조 및 풍력 단지 설비 업체인 베스타스는 빅데이터 분석을 통해 풍력 터빈(바람의 힘을 전기로 변환시키는 핵심 부품)의 에너지 효율성을 높였다. 베스타스는 날씨, 조수 간만의 차, 위성 이미지, 지리 데이터, 날씨 모델링 조사, 산림 및 해상 지도 등 페타바이트 규모의 정형 데이터와 비정형 데이터를 수집한다. 날씨 하나만 보더라도 4년 동안 20페타바이트 이상의 전 세계의 날씨 데이터를 집적하고 있다. 베스타스는 실시간으로 변하는 바람의 방향, 높이에 따른 변화, 기후 요소 데이터를 수집하고 분석하여 최적의 풍력 발전소 부지를 선정하고 있다.

둘째, 정부도 빅데이터를 활용하기 위해 많은 노력을 경주하고 있다. 그중에서 환경을 탐색하거나 상황에 대한 분석, 사회적인 현안 해결 및 국가의 미래에 대한 대응 등을 위해 빅데이터를 활용하고 있다. 이를 위해 정부는 공공데이터인 기상 데이터, 인구 데이터, 각종 통계지표 등을 수집하고 이를 토대로 하여 전반적인 각 상황에 대한 정보를 추출한다. 또한 수집된 데이터를 바탕으로 시스템 다이내믹스, 복잡계 이론 등과 같은 분석을 통해 미래 전

락을 수립하기도 있다.

셋째, 빅데이터가 사회 부문에 미치는 영향은 매우 크고 흥미롭다고 할 수가 있다. 빅데이터 분석을 통해 사회적 약자를 위한 활동들을 수행하는데 도움을 줄 수 있으며, 다양한 사회적 기회를 창출할 수 있도록 도움을 줄 수도 있다.

이탈리아의 저명 브랜드인 아날도바시니(Arnaldo Bassini)는 기브디렉틀리(GiveDirectly)라는 기관과 데이터카인드(DataKind) 프로젝트를 수행하고 있다. 데이터카인드는 모바일을 통해 아프리카의 가난한 가정들에게 무조건적으로 현금 기부를 이끌어내기 위한 프로젝트이다. 또한 케냐와 우간다 중에서 가장 가난한 마을을 찾아내기 위해 위성 이미지를 사용하고 있는 바시니는 기브디렉틀리가 체계적인 데이터를 모으는 데 빅데이터를 활용하여 비용과 시간을 절약하고 있다. 바시니는 서부 케냐에서 우간다까지, 집의 지붕 형태로 각 마을의 상대적 빈곤 상태를 파악할 수가 있으며, 또한 다양한 인터넷 웹사이트들의 스크랩을 통해 식량 가격과 관련된 새로운 데이터를 만들어 내기도 한다.

넷째, 아직까지 빅데이터를 활용하는 개인은 드물지만, 유명 정치인이나 연예인의 경우는 빅데이터를 활용하여 성공을 거둔 사례가 종종 있다. 사회적으로는 빅데이터 분석을 통해 다양한 기회를 창출할 수 있게 해주고 이를 통해 사회 전반에 긍정적인 영향을 미칠 수 있도록 해준다. 물론 아직까지는 다양한 활용 사례가 나오지 않았지만, 다양한 통찰을 통해 사회의 여러 가지 문제를 해결하는데 있어 새로운 돌파구를 마련하는 계기가 될 수 있을 것으로 보인다.

## 2. 빅데이터 유형과 특징

### 1) 빅데이터의 유형

빅데이터는 고객정보와 같은 정형화된 자산 데이터(내부)뿐만 아니라 센서 데이터나 비디오, 소셜, 실시간 데이터 등이 복합적으로 구성되어 있다. 빅데이터는 정형 데이터, 반정형 데이터, 비정형 데이터로 구분할 수 있다.

정형 데이터(structured data)는 일정한 규칙을 갖고 체계적으로 정리된 데이터를 의미한다. 표 안에서 행과 열에 의해 지정된 각 칸에 데이터를 저장하는 엑셀의 스프레드시트, 관계 데이터베이스의 테이블이 대표적인 정형 데이터이다. 예를 들어, 기업의 내부데이터, 통계청의 통계자료, 각종 과학적 데이터 등이 이에 해당한다. 정형 데이터는 그 자체로 의미 해석이 가능하며, 바로 활용할 수 있는 정보를 내포하고 있다.

반정형 데이터(semi-structured data)는 구조에 따라 저장된 데이터지만 정형 데이터와 달리 데이터 내용 안에 구조에 대한 설명이 함께 존재한다. 그렇기 때문에 데이터 내용에 대한 설명인 구조를 파악하는 파싱 과정이 필요하고, 보통 파일 형태로 저장된다. 반정형 데이터의 예시로 웹에서 데이터를 교환하기 위해 작성하는 HTML, XML, JSON 문서나 웹 로그, 서버 데이터 등이 있다.

비정형 데이터(unstructured data)는 정해진 구조가 없이 저장된 데이터다. 소셜 데이터의 텍스트, 이미지, 영상, 워드나 PDF 문서와 같은 멀티미디어 데이터가 대표적인 예다. 페이스북, 트위터, 카카오톡, 라인 등 SNS 이용률이 크게 증가하면서 많은 양의 비정형 데이터가 실시간 생산되고 있다. 오늘날 빅데이터는 비정형 데이터에 관심을 두고 있다.

이러한 빅데이터는 해당 데이터를 분석하고 처리함으로써 기존의 데이터에서 볼 수 없었던 새로운 의미를 창출하게 한다. 즉, 수많은 데이터를 분석하여 사용자에게 유용한 정보를 제공할 수 있어야 빅데이터는 효용성을 갖는다. 따라서 빅데이터에서 중요한 것은 형식적인 데이터 소스(source) 내에서 외부로 새로운 가치를 창출할 수 있느냐 하는 것이다. 결국 새로운 가치와 의미를 산출하기 위해서는 축적된 데이터를 갖고 무엇을 분석할 것인가에 대한 문제제기가 필요하다. 이러한 문제제기는 마이닝과 연결되는데, 빅데이터에서 마이닝(mining)은 데이터 분석을 통한 가치있는 고급 정보를 생산하는 것을 의미한다.

## 2) 빅데이터의 특징

가트너(2012)에 의하면, 빅데이터 특징은 3V로 지칭되는 규모(Volume), 다양성(Variety), 속도(Velocity)이다. 그러나 빅데이터가 본격적으로 위력을 발휘하기 시작하면서 빅데이터에 관한 개념이나 특징이 더욱 정교해지기 시작했다. 이로 인해 최근에는 빅데이터 특징을 5V로 부르게 되었다. 즉 5V로 대표되는 빅데이터 특징은 규모(Volume), 다양성(Variety), 속도(Velocity), 정확성(Veracity), 가치(Value) 등 5가지 구성요소이다.

첫째, 규모(Volume)는 그야말로 빅데이터의 양적 규모를 말한다. 각종 기술의 발전과 ICT의 일상화가 진행되면서 해마다 디지털 정보량이 기하급수적으로 폭증하고 있는 것이다. 과거 메가바이트(MB, 백만 바이트), 기가바이트(GB, 십억 바이트) 수준으로 축적되던 데이터량은 제타바이트(ZB, 1십해 바이트,  $1\text{ ZB} = 10^{21}\text{ bytes} = 1,000,000,000,000,000,000\text{ bytes}$ ) 수준으로 쌓이는 시대로 진입하게 된 것이다.

둘째, 다양성(Variety)은 텍스트 이외의 멀티미디어 등 비정형화된 데이터 종류와 유형의 다양화를 말한다. 과거에는 기업이나 공공 부문의 정형 데이터 중심으로 데이터가 축적되었으나, 이제는 로그기록, 소셜, 위치, 소비, 현실데이터 등 비정형 데이터 등 데이터 종류의 상상을 초월할 정도로 증가하고 있는 것이다.

셋째, 속도(Velocity)는 데이터가 모이고 축적되고 전송되는 시간적 의미를 말한다. 과거에는 데이터가 일정한 시간적 간격을 두고 축적되거나 전송되었다. 그러나 이제 사물(센서, 모니터링) 데이터, 스트리밍 데이터 등 실시간성 데이터가 증가하고 있다. 그리고 실시간성으로 인한 데이터 생성, 이동(유통) 속도도 증가하고 있다. 대규모 데이터 처리 및 가치 있는 현재(실시간) 데이터 활용을 위해 데이터를 처리 및 분석 속도가 중요해지고 있는 것이다.

넷째, 정확성(Veracity)은 데이터의 저장 및 처리과정에서의 정확하고 안전한 상태를 의미한다. 빅데이터의 특성상 방대한 데이터들을 기반으로 분석을 수행하게 되므로 정확성은 매우 중요한 것이다. 특히 분석의 정확도와 이로 인한 의사결정의 질은 데이터 분석에서 질이 높

은 데이터를 활용하는 것에 영향을 받기 때문에 데이터의 정확성은 매우 중요한 빅데이터의 특징인 것이다.

다섯째, 가치(Value)는 빅데이터로부터 얻을 수 있는 경제적 가치를 말한다. 결국 빅데이터가 추구하는 것은 가치 창출인 것이다. 빅데이터 분석을 통해 도출된 최종 결과물은 기업이 현재 당면하고 있는 문제를 해결하는 데에 통찰력 있는 유용한 정보를 제공하고 결국에는 가치 창출의 토대가 되는 것이다.

이제 빅데이터는 특정 규모(big volume) 이상을 넘어서 원하는 가치(big value)를 얻을 수 있는 정도로 상대적인 해석을 해야 하는 것이 된 것이다. 예를 들어, 통계 패키지로 유명한 SAS는 데이터 가치에 중점을 두어 가치를 창출하기 위한 비즈니스 예측 및 최적화 주제를 선정하여 빅데이터에서 어떤 가치가 있는 정보를 얻을 것인가에 분석 관점을 가지고 있다.

### 3) 빅데이터 추진 3대 요소

#### 가. 자원 : 활용할 수 있는 빅데이터 발견하기

미래에는 빅데이터를 핵심 자원으로 인식하고, 필요한 정보를 뽑아낼 수 있도록 자원을 축적해 가는 것이 무엇보다 중요하다. 주어진 빅데이터를 관리하고 처리하는 측면과 함께 활용할 수 있는 외부 빅데이터 자원을 발견하고 확보하는 전략이 필요하다. 데이터의 품질은 데이터 활용 결과에 중대한 영향을 주기 때문에 데이터 관리 체계 및 데이터의 신뢰성 확보가 매우 중요하다.

빅데이터 자원 확보 방안은 생성, 수집(검색), 공유와 연계, 협력, 참여, 오픈 등에 따라 단계적으로 확장이 가능하다. 최근에는 소셜미디어, 포털 검색 데이터 등을 분석하여 여론 확인, 기업의 마케팅 서비스 등에 적용하는 사례가 증가하고 있다.

#### 나. 기술 : 빅데이터 프로세스와 신기술

빅데이터는 데이터 자체뿐만 아니라 관련 도구나 플랫폼, 분석 기법까지 포괄하는 용어로 확장되어 디지털 기술 패러다임의 변화를 견인하고 있다. 성장을 촉진하는 신기술에 대한 이해가 부족하면, 미래 경쟁력 강화를 위한 기회 포착도 어려워진다. 조직과 기업의 혁신 전략으로 적용할 수 있도록 빅데이터 플랫폼, 빅데이터 분석 기술 및 기법에 대한 이해가 필요하다.

기업의 의사결정에 유용한 정보를 추출한 비즈니스 인텔리전스(Business Intelligence : BI)는 분석 기능이 강화된 비즈니스 애널리틱스(Business Analytics : BA)로 진화 중이다. BI는 신속하고 정확한 비즈니스 의사결정을 위해 사용하는 데이터의 접근, 수집, 보관, 분석 등 애플리케이션과 기술의 집합이다. 최근 BI에서 BA로 진화하며, 데이터의 생성부터 폐기까지 전사적인 범위에서 기업의 미래를 예측한다.

#### 다. 인력 : 데이터 사이언티스트 역량

기술과 분석 도구가 뛰어나도 실제 성과를 낼 수 있는 것은 이를 활용하여 적용하는 사람

의 역량이다. 데이터 처리와 분석 능력을 갖춘 인력은 디지털 기술 분야뿐만 아니라 대부분의 기업과 조직에서 필수적으로 확보해야 할 핵심 인력이다. 조직 차원에서 인재를 확보하기 위해서는 내부 역량 강화 및 외부 협력이 중요하다.

데이터 사이언티스트의 역량과 조건은 다양한 문헌에서 제시되고 있다. 대표적인 데이터 사이언티스트인 존 라우저, 아마존 수석 엔지니어는 데이터 사이언티스트가 갖추어야 할 6가지 자질을 제시하였다.

데이터 사이언티스트는 기본적으로 ① '수학'과 ② '공학' 능력을 갖추어야 한다. 데이터를 분석하는 데 있어 필수인 가설을 세우거나 검증하는데 필요한 ③ '비판적 시각'과 이를 작성할 수 있는 ④ '글쓰기 능력'을 보유하여야 한다. 또한 다른 사람에게 구두로 내용을 적절히 전달하기 위한 ⑤ '대화 능력'도 필요하다. 나아가 라우저는 ⑥ '호기심과 개인의 행복'도 중요한 소양으로 제시하였다.

데이터 사이언티스트는 대규모 데이터를 분석한 결과를 생생하고 차별화되게 시각화하여 이해하기 쉽게 전달하는 역량도 갖추어야 한다. 시각화는 데이터 분석 결과를 전달하는 마지막 단계로써 가시적 데이터의 문맥화를 통한 해석 작업이라고 할 수 있다. 정교한 모형 및 시각화 도구를 활용하면 비즈니스 가치와 통찰력을 제공할 수 있다.

### 3. 빅데이터 비즈니스 모델과 활용

#### 1) 빅데이터 비즈니스 모델의 이해

일반적으로 빅데이터 비즈니스 모델이란 빅데이터 산업에 관련된 모든 비즈니스를 의미한다.

빅데이터의 산업 구조는 크게 인프라 부분과 서비스 부분의 두 부분으로 나눌 수 있다. 우선 인프라 부분은 ①빅데이터를 구성하기 위한 데이터의 수집, 저장, 분석, 관리 등의 기능을 담당하는 컴퓨터, 단말, 네트워크, 서버, 스토리지 등의 하드웨어와 ②관리, 분석틀 등의 소프트웨어로 구분할 수 있다. 또한 서비스 부분은 ③교육, ④컨설팅, ⑤솔루션(분석, 저장, 관리, 검색, 통합 등), 데이터 및 정보 제공, ⑥데이터 처리 등 다양한 서비스 종류가 있다.

#### 2) 빅데이터 비즈니스 모델 유형

##### 가. 교육 관련 비즈니스 모델

빅데이터의 발전으로 관련 인력에 대한 수요가 폭발적으로 증가하고 있다. 즉 빅데이터 활용이 증대되면서 관련되는 다양한 직종이 탄생하면서 이에 대한 수요가 증가하고 있다. 빅데이터 관련 분야의 전문 직종은 아래와 같이 6가지 분야로 분류하여 정리해 볼 수 있다.

직업	주요 업무 내용
데이터 분석가	- 제조, 유통, 금융, 의료 등 각 산업 분야에 대한 도메인 지식과 해당 산업에 특화된 데이터 분석 수행

데이터 사이언티스트	- 다양한 비즈니스 지식을 바탕으로 빅데이터를 활용하여 비즈니스에 새로운 가치 부여
데이터 시각화 전문가	- 데이터, 분석 결과 및 관련 지식을 효과적이고 이해하기 쉽도록 시각적으로 표현
데이터 엔지니어	- 데이터를 안전하고 효과적으로 수집 및 저장하고 필요할 때 빠르게 분석 및 출력할 수 있는 시스템 구축
빅데이터 기반 사기방지 기술자	- 전자금융 거래로 인해 발생한 데이터를 분석하고 이상 거래를 미리 탐지하고 사기를 미연에 방지
빅데이터 운영 관리자	- 데이터를 수집, 저장 및 처리하는 시스템을 체계적으로 관리
빅데이터 큐레이팅 전문가	- 방대한 양의 데이터 속에서 숨어 있는 가치 발굴

#### 나. 컨설팅 관련 비즈니스 모델

컨설팅 관련 비즈니스 모델은 크게 두 가지로 기업을 상대로 한 빅데이터 활용 컨설팅과 개인을 상대로 한 컨설팅으로 나누어 생각할 수 있다. 개인을 상대로 하는 컨설팅 비즈니스의 예로는 개인의 신용카드 사용 데이터와 각종 금융(은행, 보험 등) 데이터를 이용한 개인별 종합 금융 컨설팅 또는 건강 및 진료 데이터를 이용한 건강관리 컨설팅 등이다.

기업 또는 개인 사업자를 대상으로 하는 컨설팅 비즈니스 모델은 카드사들의 고객 결제 정보 및 SNS 데이터를 활용한 상권 분석 컨설팅, 고객의 구매 패턴 데이터를 활용한 제품 개발 컨설팅 등 다양한 유형으로 가능하다.

#### 다. 솔루션 및 도구 공급 관련 비즈니스 모델

솔루션 및 도구 공급 관련 비즈니스 모델은 크게 H/W 부분과 S/W 부분으로 분류할 수 있다.

첫째, H/W(Hardware) 부분은 빅데이터의 수집, 저장, 분석을 위한 하드웨어 장비로서 네트워크 장비, 서버, 스토리지 등이 있다. 기존 ICT 분야의 비즈니스 모델과 중복되기 때문에 대부분 기존 사업자들이 빅데이터 처리를 위해 기존 하드웨어보다 성능이 우수한 고용량, 고속의 장비들로 대체하는 시장을 개척하고 있다.

둘째, S/W(Software) 부분의 비즈니스 모델 역시 기존 솔루션 사업자와 스타트업들이 새로운 비즈니스 창출을 위해 빅데이터 수집, 저장, 분석 관련 솔루션이나 툴을 개발하여 공급하고 있다. 솔루션은 오픈소스(open source) 진영과 상용 솔루션 진영으로 나뉘어 공급하고 있으며, 점차 오픈소스에 기능성과 관리 도구를 갖추어 상용화 서비스를 하는 비즈니스 모델을 개발하며 시장을 개척하고 있는 상황이다.

#### 라. 빅데이터(원시, 분석, 가공) 판매 비즈니스 모델

빅데이터 관련 데이터의 판매 비즈니스 모델은 빅데이터의 특성인 다양성, 대용량, 빠른 데이터 생성에 따라 다량의 다양한 데이터에 기반한다. 데이터들은 서비스의 종류에 따라 필요한 데이터와 불필요한 데이터로 분류할 필요가 있다. 통상 데이터는 원시 데이터 또는 가공한 형태의 데이터를 필요로 하는 사업자에게 판매하는데, 불필요한 데이터를 제외하고 필

요 부분 데이터만을 별도로 판매하게 된다. 따라서 이들 데이터를 필요로 하는 사업자를 찾아 데이터를 직접 판매하거나 데이터를 필요로 하는 사업자를 연계시켜주는 비즈니스 모델이 각광을 받을 것이다.

#### **마. 빅데이터 기반 서비스 비즈니스 모델**

빅데이터 서비스 비즈니스 모델이 실제 가장 큰 빅데이터 시장을 형성할 수 있는 비즈니스 모델이다. 따라서 빅데이터 비즈니스 모델을 말하면 이 서비스 비즈니스 모델을 의미한다고 볼 수 있다. 즉 서비스 비즈니스 모델은 빅데이터를 수집, 저장하여 이를 필요한 서비스 요건에 맞추어 분석하고 분석된 데이터를 이용하여 필요한 서비스를 제공하는 것이다. 빅데이터 기반 서비스 비즈니스 모델은 데이터의 활용 시나리오 측면에서 ①이상 현상 감지 시나리오, ②가까운 미래 예측 시나리오, ③현 상황 분석 시나리오, ④고객 맞춤 시나리오로 나누어 볼 수 있다.

첫째, 이상 현상 감지 시나리오는 업무에서 발생한 이벤트 기록을 수집·분석하여 정상 상태, 비정상 상태의 패턴을 파악, 새로운 현상이 발생할 경우 이상 여부를 판단하고 대응하도록 하는 시나리오이다. 그러므로 부정행위 검출, 시스템 사고 예방, 마케팅 전략 수립 등의 다양한 분야에서 활용될 수 있다. 부정행위 검출의 활용 사례로서 비자(VISA)사는 카드 부정검지모델의 설계 처리에 하둡(Hadoop)을 도입하여 부정검지 이용 패턴의 갱신을 매일, 수회로 실시할 수 있도록 정밀도를 향상시킬 수 있었다.

둘째, 가까운 미래 예측 시나리오는 빅데이터를 통해 수 분 또는 수 시간 후를 예측하는 시스템을 실현함으로써 현상에 대한 조치를 신속하게 실시할 수 있다. 주행 시스템에 활용한 사례로서 미국 포드사는 개발 중인 커넥티드 자동차의 주행 시스템에 빅데이터를 이용하여 내비게이션이 운전자의 주행 이력과 패턴을 분석하여 운전자에게 앞으로 도착해야 할 목적지에 이르는 최단 또는 최적 경로와 연료 배분을 제안하도록 구현하였다.

셋째, 현 상황 분석 시나리오는 빅데이터를 이용하여 지금까지 살펴보지 못했던 사업 측면의 분석을 가능하게 하여 자사의 현황을 보다 명확하게 이해하도록 지원한다. 일본의 니시테츠 스토어(Nishitetsu Store)는 빅데이터를 이용하여 매일 단일 상품별 원가율과 원가 변동 추이를 분석하여 이익률이 높은 상품에 대한 일자별 주력 마케팅 정책을 수립하여 수익성을 극대화하고 있다.

### **3) 빅데이터의 활용**

#### **가. 미래사회의 빅데이터 역할**

사회가 빠르게 발전하고, 리스크(risk) 요인과 복잡성이 증가할수록 시스템을 통해 환경 변화를 신속하게 감지할 필요성이 증가한다. 따라서 빅데이터는 미래 사회에서 나타날 수 있는 리스크를 줄이고 새로운 기회를 창출하는 엔진의 역할을 수행할 것으로 기대된다. 즉 미래 사회는 불확실성, 리스크, 스마트, 융합 등의 특징을 보이면서 기회와 위협 요소를 출연케 할 것이다. 이에 대해 빅데이터는 여러 가지 가능성에 대한 시나리오 시뮬레이션을 통해 통찰력, 대응력, 경쟁력, 창조력 등으로 불확실한 각각의 상황 변화에 유연하게 대처할 수 있



게 지원할 것이다.

정리하면, 빅데이터에 기반을 둔 정보 패턴 분석을 통해 리스크를 사전에 파악하고 실시간으로 대응할 수 있다. 또한 빅데이터 분석을 통해 개인화, 지능화된 서비스 제공을 확대하여 스마트 사회에서 삶의 질을 향상시키고, 트렌드 변화 분석을 통해 각 기업의 제품 경쟁력을 확보할 수 있다. 빅데이터를 활용하여 다양한 산업 분야의 결합을 유도하여 새로운 가치를 창출하고, 나아가 새로운 융합 시장을 창출할 것으로 기대된다.

#### **나. 산업별 빅데이터 활용**

산업 분야에서의 빅데이터 분석이란 의료, 자동차, 금융, 통신 등 각 산업내의 해당 기업들이 의사결정, 사업 및 업무 수행, 대 고객 서비스 제공 등의 비즈니스 프로세스에서 보다 과학적이고 합리적으로 수행하기 위해 빅데이터를 분석 및 활용하는 것을 의미한다. 따라서 각 산업내의 기업들은 자사가 처한 상황에서 문제 해결이나 신규 비즈니스 발굴 등 상황에 적합하고 필요한 데이터를 분석하여 적극적으로 활용하고자 할 것이다.

주요 산업(업종)별 빅데이터 활용에 대해 정리하면 다음과 같다.

첫째, 제조분야의 빅데이터 활용이 지속적으로 증가하고 있다. 디지털 기술의 급속한 발전으로 인해 생산 공정에도 뛰어난 효율성을 제공하는 기술 및 지식이 제공되고 있다. 이에 따라 기존 제조 산업은 미래형 제조 산업으로 빠르게 진화하고 있다. 사람, 기계, 자원, 생산라인, 물류 등 다양한 분야에서 발생하는 풍부한 데이터는 빅데이터 분석 소프트웨어와 연결되어 분석이 가능하게 되었다. 또한 이렇게 모아진 데이터는 제조 공정에서 비용을 감소시키고 생산 계획을 자동으로 만들어내는 알고리즘으로 변환되어 전체적인 생산성을 크게 향상시키는 데에 기여하고 있다. 많은 기업들이 단순한 제조 공정을 스마트한 제조 공정으로 혁신하는 데에 빅데이터를 활용하고 있다.

둘째, 금융분야의 빅데이터 활용 역시 급격히 증가하고 있다. 금융 서비스는 대다수의 국민들과 기업들이 항상 이용하고 있고 우리 생활의 일부분일 뿐만 아니라 금융 상품의 종류도 많기 때문에 데이터의 유입량, 보유량, 집적량이 광대하며 그 증가 속도가 실시간으로 이루어지고 있다. 이에 따라 다른 산업군에 비해 빅데이터의 활용 범위가 다양하고 활용 가치 또한 높을 것으로 기대하고 있다.

셋째, 유통·물류 분야에서는 주로 방문 고객의 구매 이력 데이터와 소비자들의 소셜미디어 데이터를 분석하여 주력 판매 제품, 매장 구도 계획 등 상품별 판매 전략의 수립에 활용하고 있다. 예를 들면, 추석, 구정 등의 명절이나 기념일(발렌타인 데이, 스승의 날 등) 전에 미리 고객의 SNS 활동과 과거의 구매 이력 데이터 등 기구축된 데이터를 분석하여 고객이 선물할 만한 대상을 추정하고 이에 적합한 선물을 추천할 수가 있다. 또한 글로벌 유통업체들은 네트워크를 활용하여 유통 데이터를 생산하고 영업 전략에 활용을 하고 있다. 중국의 우렌왕은 네트워크로 연결된 센서에서 사물의 움직임을 인식하고 정보를 분석하여 활용하는 유통 인프라를 추구하는 등 많은 부분에서 빅데이터를 활용하고 있다.

넷째, 보건 의료 산업의 경우 의료 데이터의 80% 이상이 비정형화된 데이터이기 때문에 기존 인프라로는 분석이 어려워 빅데이터의 도입 필요성이 증가하고 있다. 건강관리자료인 검진자료, 질병자료, EMR(Electronic Medical Record), 유전자 분석 등의 데이터와 바이오 데이터, 의료 영상 관련 데이터 등이 폭발적으로 증가하고 있는 추세이다. 이런 가운데 기존의 치료 중심에서 예방 및 건강관리 중심으로 트렌드가 바뀜에 따라 질병 및 전염병 발병 예측과 수요자 중심의 맞춤형 서비스의 중요성이 점차 증가를 하고 있다. 이에 따라 의료 분야에서 데이터 분석 활용 필요성과 이로 인한 가치 창출이 기대되고 있다. 의료 분야의 빅데이터 활용 효과는 미국의 경우 연간 3,000억 달러, 유럽의 경우 연간 2,500달러 규모에 이를 것으로 분석되었다.

#### **나. 공공 부문의 빅데이터 활용**

정부 등 공공 분야에서 업무 수행 과정상 축적한 데이터(이를 통상 공공데이터라 한다) 역시 무궁무진한데, 공공 분야 역시 이 데이터와 다른 데이터를 분석하여 해당 분야의 정책 결정, 업무 수행 및 대국민 서비스를 보다 과학적으로 수행하고자 할 것이다. 즉 공공 분야에서의 빅데이터 분석이란 정부, 지방자치단체 및 공공기관 등의 정책, 행정, 사업 등을 수행하는 과정에서 보다 과학적인 의사결정이나 합리적인 업무 수행을 위해 빅데이터를 분석하는 것을 의미한다.

#### **다. 기업의 가치활동별 빅데이터 활용**

오늘날 경쟁의 가속화와 더불어 소셜미디어의 보편화로 인해 소비자의 반응이 실시간으로 데이터로 노출되고 이러한 데이터를 경쟁전략과 연계하여 무한 경쟁에서 경쟁 우위를 확보하고자 하는 노력이 절대 필요하다. 고객의 마음(needs)을 정확하게 분석하여 고객이 움직이는 대로 기업의 비즈니스 모델을 집중시키는 전략이 필요한 것이다.

기업의 경영 분야에서는 전통적으로 계량적 분석과 밀접한 마케팅, 생산, 회계, 재무 분야에서뿐만 아니라 인사관리, 연구개발, 인수합병(M&A) 등 기업 내 거의 전 분야에서 운영의 효율을 높이고 현명한 의사결정을 하는 데에 애널리틱스를 활용하고 있다.