

# ISSUE REPORT

2021.

Vol. 4

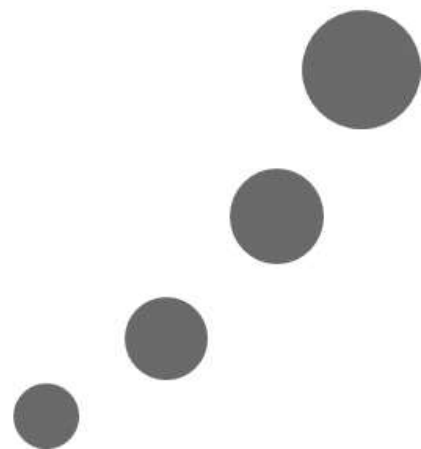
## 데이터 댐(Data Dam) 구축과 AI 기반 데이터 가치 고도화



# TIPA | 이슈 리포트

## 데이터 댐(Data Dam) 구축과 AI 기반 데이터 가치 고도화

- I. 데이터 댐의 개요
- II. 국내외 데이터 산업 정책 분석
- III. 데이터 댐 구축 관련 기술 동향
- IV. 데이터 댐 관련 주요 플레이어 동향
- V. 결론 및 시사점



# I. 데이터 댐의 개요

## 1 데이터 댐이란?

- 데이터 댐은 지난 2020년 7월 발표된 『디지털 뉴딜』의 대표 과제로 공공기관 및 민간기업의 데이터를 수집하고, 유용한 정보로 데이터가 가공될 수 있도록 지원하는 데이터의 집합 시스템을 의미
  - 세계 경제 대공황을 극복하기 위해 미국 대통령 플랭클린 루즈벨트가 펼쳤던 뉴딜(New Deal) 정책의 핵심이었던 후버 댐(Hoover Dam)과 같이, 차세대 산업의 원동력이자 코로나19로 인한 난관을 극복하기 위한 한국판 뉴딜의 핵심 사업
  - 후버댐 건설을 통해 주변에 물을 활용하는 산업이 생겨나고 일자리가 창출된 것과 같이, 데이터 댐을 구축하여 수집되는 데이터를 활용할 수 있는 데이터 기반 신산업이 활성화될 것으로 기대
- 데이터 댐의 핵심 목적은 저장된 데이터를 각종 신기술과 다양한 산업 분야에 연계시켜 데이터 중심의 새로운 비즈니스 모델을 발굴하고 이를 바탕으로 국내 데이터 경제 조성, 일자리 창출 등, 데이터 기반의 국가 성장 동력을 마련하는 것
  - 빅데이터 플랫폼 및 센터구축사업을 통해 공공데이터뿐만 아니라 민간데이터를 통합 구축하여 공공의 사회문제 해결과 공공데이터를 활용한 다양한 혁신 서비스의 창출을 도모

### [ 대한민국 정책브리핑, 데이터 댐 ]



\* 출처: 대한민국 정책브리핑(과학기술정보통신부 블로그), 2020.03.13.

- 데이터 댐은 크게 데이터 수집, 데이터 축적 및 가공·결합, 그리고 데이터 활용, 이 세 가지 단계로 구분이 가능하며 특히 본 이슈 리포트에서는 각 단계별 기술 동향과 핵심 플레이어 동향을 위주로 기술



## 2 데이터 댐의 필요성

- 전 세계적 디지털 전환(Digital Transformation) 추세로 인해 데이터량이 급격히 증가하여 빅데이터 수집·분석 기술에 대한 수요 대폭 증대 예상
  - 전 세계에서 연간 생성되는 데이터량은 2016년 96EB<sup>1)</sup>에서 2021년 278EB로 증가할 전망<sup>2)</sup>
  - 무선 네트워크의 고도화와 스마트폰의 보급 확대로 비정형 데이터 기반의 모바일 데이터는 2016년 7.2EB에서 2021년 48.2EB로 6.7배 폭증할 것으로 예상<sup>3)</sup>
  - 급격히 증가한 데이터량에 따라 빅데이터 수집·분석 기술에 대한 수요 대폭 증대되고 있으며, 데이터 베이스 구축을 위한 기초 인프라 투자의 중요성이 부각
- 빅데이터 기술은 초연결 사회, 4차 산업혁명 등 기존 사회에 획기적인 변화를 가져오는 기술의 진보를 위한 기반 기술로 관련 수요가 지속적으로 증가
  - 모바일 기기의 확산 및 사물인터넷의 발달로 데이터량은 기하급수적으로 증가할 것으로 예상되고 있어 데이터로부터 새로운 가치와 혁신을 창출하기 위한 관련 기술 중요성 부상
  - 빅데이터는 인공지능 분야의 학습에서 사용되거나, 적용되는 분석 기법에 따라 다양한 활용 분야에서 유의미한 지식을 창출 가능
  - 플랫폼 기반 기업이 향후 10년 간 글로벌 GDP 중 신규 부가가치의 70% 창출할 것으로 전망되며 이에 따라, 각국이 위기극복 및 국가경쟁력 제고를 위해 디지털 투자를 경쟁적으로 추진

### [ 데이터 경제 구조 ]



\* 출처: "혁신성장동력 시행계획", 관계부처 합동, 2018.05.28

- 이와 같이 디지털 경제로의 전환 속도가 국가의 산업·기업 경쟁력을 좌우할 것으로 전망되어, 데이터 댐 구축을 통한 데이터 생태계 기반 조성이 중요한 상황

1) Exabyte, EB:  $10^{18}$  bytes

2) Cisco, Visual Networking Index: Forecast and Methodology, 2016-2021

3) Cisco, Mobile Visual Networking Index: Global Mobile Data Traffic Forecast, 2016-2021

## II. 국내외 데이터 산업 정책 분석

### 1 주요국 데이터 산업 관련 정책

- 데이터라는 새로운 자원을 확보하기 위한 경쟁은 기업 차원을 넘어서 국가 간의 ‘데이터 주권’ 경쟁으로 확대되는 추세
  - 국경 간 데이터 이동과 글로벌 데이터 기업의 조세 형평성 등의 데이터를 매개로 한 국가 간 무역 분쟁 사례가 증가
  - 특히, 미국은 자국 내의 플랫폼 사업자를 전달자(Carrier)로 보고 이용자 불법행위에 대해 면책하거나, 망 중립성 규제를 대폭 완화하는 등의 정책 기조로 미국 이외의 국가와 마찰
- 최근에는 데이터 경제 체제에서 공정경쟁 규제가 국제적이면서도 내부적으로도 논란이 되는 핫이슈로 부상하면서 자국 내 기업에 대한 제재 사례가 발생하며 양상이 복합적으로 변화
  - 미국의 40개 주에서 구글이 온라인 광고 시장을 불법 독점하고 있다며 반독점법 위반 혐의 고소<sup>4)</sup>
  - 중국 금융 당국의 알리바바 관계사인 앤트 그룹 상장 중단 조치<sup>5)</sup>

#### ◎ 미국

- 2020년 1월 연방데이터 정책위원회(Federal Data Policy Committee: FDPC)를 설립하여 데이터 정책 및 거버넌스, 자원에 대한 조정을 강화
  - 연방기관들이 기존 또는 새로운 데이터 관련 법적 요건 충족은 물론 인공지능에 사용하기 위한 데이터 준비 등 준비사항을 관리
  - 데이터 관리의 비용, 중복 부담을 줄이는 노력과 함께 데이터 관리표준화 및 활용에 대한 정부 차원의 관리를 통해 거버넌스 자원의 우선순위 결정 등의 역할 수행
- 미국 연방 정부는 2019년 1월 정부의 정책 평가에 필요한 증거(evidence)를 ‘통계’로 규정하는 증거기반정책법(Foundations for Evidence-Based Policymaking Act of 2018) 제정
  - 증거기반정책법은 연방정부의 증거구축활동을 규정하고(제1장) 연방정부가 생산하는 공공데이터의 활용에 초점을 맞춘 OPEN(Open, Public, Electronic, and Necessary) 정부데이터법(제2장)과 비밀정보보호와 통계적 효율성(제3장)으로 구성하여 데이터와 통계의 유기적인 연계를 강조

4) 구글, 美 40개 주에 줄소송 직면...“온라인 광고·검색 독점”, 조선비즈(2020.12.17.)

5) 마윈 '中 정부 비판' 화 불렀나...앤트그룹 상장 전격 중단, 한국경제(2020.11.03.)

- 10대 원칙은 윤리적 거버넌스, 의식적 설계 및 학습 문화라는 세 범주로 구성되고, 40대 실천 전략은 10대 원칙 실현을 위한 5~10년간의 중장기 전략으로, 데이터의 가치를 육성하고 공공 부문에의 활용을 촉진하기 위한 실천 전략, 데이터의 관리와 보호를 위한 실천 전략의 효율적이고 적절한 데이터 활용 촉진을 위한 실천 전략으로 구성

□ 트럼프 행정부의 주요 정책 중 하나로 추진되던 ‘클린 네트워크’로 인하여 화웨이와 ZTE 등의 중국 기업을 배제하던 양상에서 불거진 미·중간의 무역 분쟁이 존재

- 클린 네트워크는 클린 패스(Clean Path), 클린 캐리어(Clean Carrier), 클린 스토어(Clean Store), 클린 앱(Clean Apps), 클린 클라우드(Clean Cloud), 클린 케이블(Clean Cable)로 구성
- 다만, 바이든 행정부가 당선되며 클린 네트워크 정책은 앞으로 둔화될 것으로 전망

## ◎ EU

□ 2020년 출범한 유럽연합 집행위원회는 향후 5년간 유럽연합 디지털 정책의 방향을 제시하는 정책 중 하나로 “유럽 데이터 전략(European Strategy for Data)”을 2월 19일 발표

- 유럽만의 인간 중심적 가치에 기반한 데이터 환경 구축을 위한 정책의 기본 틀로 크게 네 가지 축을 중심으로 정책을 발표할 예정
  - 네 가지 축은 1) 데이터 접근과 이용을 위한 포괄적 거버넌스 체계 구축, 2) 데이터에 대한 투자 확대와 유럽의 역량 및 데이터 인프라, 호환성 강화, 3) 개인의 권리 보장 강화, 기술 인력과 중소기업자들에 대한 투자 확대, 4) 산업 부문별 유럽의 공통 데이터 풀 조성
- 2020년 11월 25일 EU 집행위원회는 데이터 중개자에 대한 신뢰도를 높이고 EU 전역의 데이터 공유 체제를 강화하여 데이터 가용성 제고를 위한 데이터거버넌스법(Data Governance Act) 발표
  - 공공 부문 데이터의 재사용, 데이터 중개자의 데이터 공유 서비스 신뢰성 확보, 이타적 목적으로 제공하는 데이터의 수집 및 처리와 관련한 EU 전역의 거버넌스 프레임워크 구축이 주요 목적
- 2020년 12월에는 디지털 서비스 법안(Digital Services Act) 및 디지털 시장 법안(Digital Market Act)을 발표하며
  - 디지털 서비스 법안은 이용자의 게시물, 의견, 메시지 등에 대한 의사 결정 및 처리 방법에 대해서 인터넷 중개서비스 제공 사업자들이 준수해야 할 책임과 규칙을 제시
  - 디지털 시장 법안은 기존 반독점 제도로 부족한 억제 효과를 향상시키기 위해, 집행위원회에 거대 게이트키퍼(gatekeeper)<sup>6)</sup>들에 대한 조사권, 전 세계 매출액의 10%에 해당하는 과징금 부과, 체계적인 법 위반 플랫폼 사업자의 해체 등의 권한을 부여

□ 2015년 5월 유럽 집행위원회(장클로드 융커(Jean-Claude Juncker) 재임 기간)의 “디지털 단일 시장 전략(Digital Single Market Strategy) 발표

- 2018년 5월부터 일반 개인정보보호 규정(GDPR, General Data Protection Regulation) 시행
  - 유럽 의회에서 유럽 시민들의 개인정보 보호를 강화하기 위해 만든 통합 규정으로 국제적인 데이터 프라이버시 보호 규범으로 영향력을 확장 중

6) 핵심 플랫폼 서비스를 제공하는 사업자를 의미

- 2018년 3월 EU 집행위원회는 글로벌 디지털 기업의 유럽 내 매출에 대해 세금을 부과하는 “디지털세(Digital Tax)” 법안을 발표
  - 기존 법인세 체계에 디지털 비즈니스 모델이 적용될 수 있도록 ‘주요 디지털 사업장(significant digital presence)’ 개념을 추가
  - EU 회원국 중 영국, 프랑스 등 6개국이 매출의 2%에서 7.5%에 해당하는 디지털서비스세(DST: Digital Service Tax)를 부과

## ◎ 중국

- 2021년 6월 10일, 제 13회 인민대표대회상무위원회는 ‘중화인민공화국 데이터안전법(일명 데이터보안법)’의 3번째 개정 버전을 공개하였으며, 이번에 개정된 법률은 총 7장 55조로 구성되어 2021년 9월 1일부터 정식 시행
  - <데이터보안법>은 데이터 보안영역에서의 최상위법으로 2017년 6월 1일 시행된 <사이버보안법>과 함께 <국가보안법> 프레임워크 하에서 보안 거버넌스 법률 체계를 보완
  - 주요 내용은 중앙국가기구의 국가 데이터 수집·저장·전송 등 안전을 감독·관리하는 매커니즘 수립과 중국 영토 안의 데이터의 해외 반출을 강력하게 처벌하여 규제의 범위를 확대할 것
  - 데이터 관리에 관한 규정이 세분화되고 처벌이 강화되어, 중국과 관련된 소셜 네트워킹 서비스(SNS) 및 전자상거래 등 플랫폼 관련 기업들은 규제에 대비가 필요
- 2020년 9월 8일 데이터 안보에 관한 글로벌 이니셔티브(全球数据安全倡议)를 발표
  - 미국 정부가 중국산 소셜네트워크(SNS) 틱톡·위챗 사용을 금지한 조치에 대해서 대응하기 위해 발표한 글로벌 데이터 안보 표준안(8개 조항으로 구성)으로 데이터의 국적성을 강조한 것이 특징
  - (제4항) 기업은 사업하는 국가의 법률을 존중하고, 기업이 해외에서 취득한 데이터를 그 국가(해외국가)에만 저장하도록 강요해서는 안 된다
  - (제5항) 다른 국가의 주권, 관할권 및 데이터 관리 권한을 존중하고 타국 데이터에 접근해선 안 된다

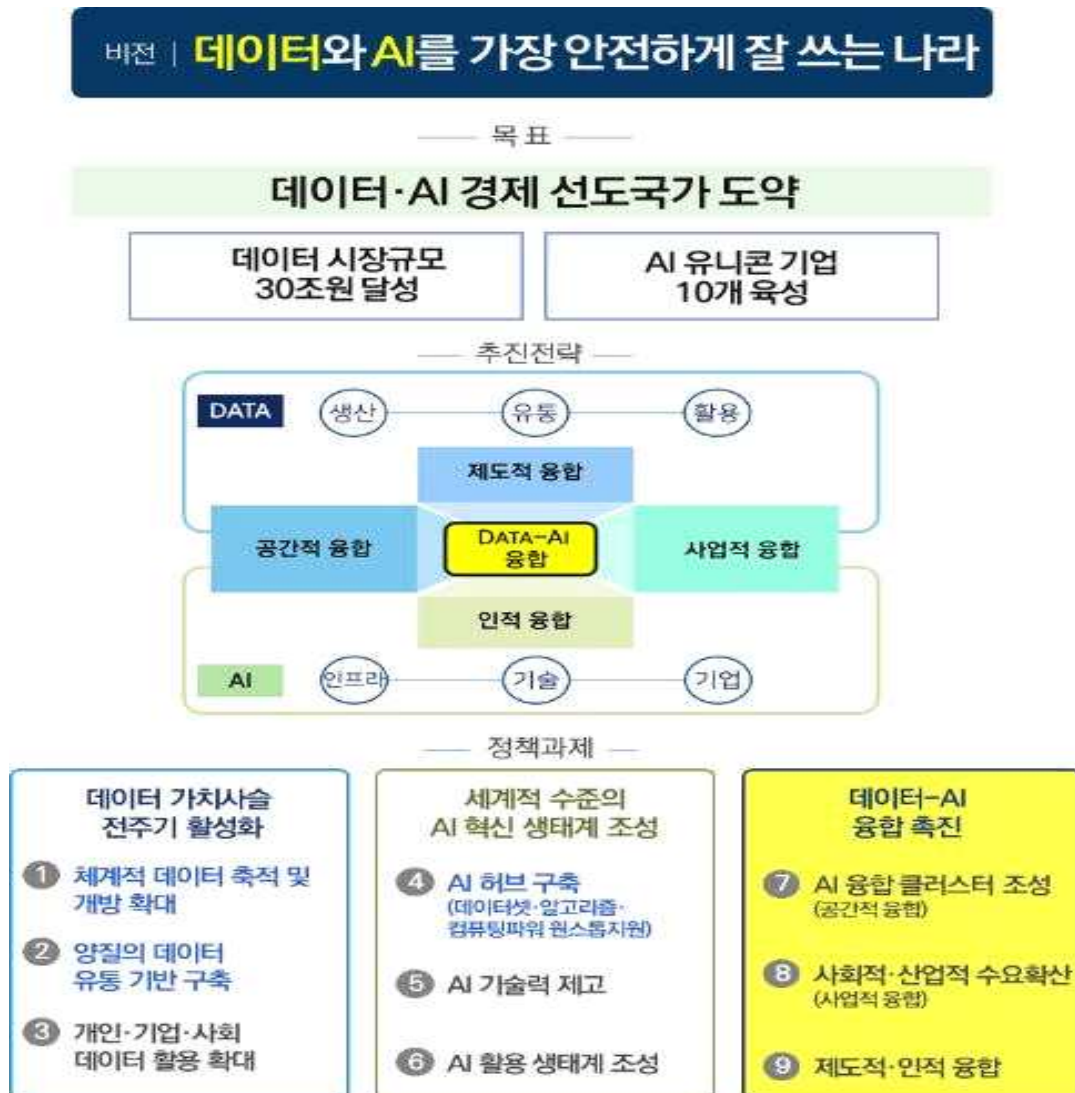
## ◎ 일본

- 2016년 관민 데이터 활용 추진 기본법을 제정하며 정부와 민간이 보유하고 있는 데이터를 공동 활용할 수 있도록 하는 법적 근거를 마련
  - (제11조) 민간 공공을 포괄한 관민 데이터에 대하여 공익 증진을 위하여 국민이 쉽게 이용할 수 있도록 필요한 조치를 강구하도록 노력할 것을 규정

## 2 국내 데이터 정책

- 정부는 10대 대표과제 중 하나로 “데이터 댐”을 선정하여 2020년 추경부터 2025년까지 26.6조원의 사업비를 투입하여 38만 9천개의 일자리를 확보할 것이라고 발표
  - 범국가적 데이터 정책 수립, 공공·민간데이터 통합관리·연계·활용 활성화, 데이터 산업 지원 등을 위한 민관합동 컨트롤타워 마련
  - 분야별 데이터 확충·연계·활용, 5G·AI 융합서비스를 통한 글로벌 선도로 2025년까지 데이터 시장 규모는 43조 원으로 성장 예상
    - 2022년까지 총사업비 8.5 조원 (국비 7.1조원) 투자 일자리 20.7만개 창출
    - 2025년까지 총사업비 18.1 조원 (국비 15.5 조원) 투자 일자리 38.9만개 창출

### [ 대한민국 정책브리핑 ]



\* 출처: 대한민국 정책브리핑.(2020.03.13.)



- 데이터 3법(개인정보 보호법, 정보통신망법, 신용정보법) 개정안은 2020년도 8월부터 시행
  - 추가 정보 없이는 특정 개인을 알아볼 수 없도록 처리한 정보인 ‘가명정보’ 개념의 도입으로 추후 데이터 활용 및 분석에 유리
  - 다수의 부처에 데이터 관련법들이 산재되어 생긴 중복 규제의 해결을 위한 법 제도 및 감독 기구 일원화하여 규제 완화
  - 데이터 보안 강화 및 개인정보처리 위반 처벌 강화를 통해 ‘활용 범위는 넓히되 처벌은 강하게’라는 방식 추진
  
- 2020년 7월 ‘국가빅데이터융합관리처신설(안)’의 법안 발의를 통해 데이터 댐 추진을 위한 국가차원의 통합데이터 관리에 대한 논의가 진행 중
  - 다양한 형태와 유형, 그리고 생산기관의 데이터를 수집·탐색하고 저장·관리하는 데이터 댐을 가능하게 하기 위한 국가차원에서의 통합데이터 관리 논의가 활발함
  - 부처별 기관별로 산재되어 있는 데이터 생산 및 관리 규정과 규칙 및 제도를 일원화하고 통합관리의 필요성이 증가
  - 범부처 데이터 기구의 필요성이 강하게 요구되고 있지만, 이해관계의 조정과 해결이 시급함
  
- 정보화진흥원(NIA)의 데이터 거래를 위한 10개 빅데이터 플랫폼 사업 (2019년)
  - “데이터를 가장 잘 다루는 나라”라는 데이터 경제선언(2018.8.31) 이후 데이터 고속도로 구축의 일환으로 민간과 공공에서 보유하고 있는 데이터를 오픈하고 유통을 통해 데이터 거래를 활성화시키기 위한 정책을 추진
  - 빅데이터 플랫폼과 센터를 통해 주요 분야별, 다양한 형태의 데이터 수집, 분석, 유통을 지원하고, 주요 기간별 데이터를 체계적으로 생산·관리 하도록 추진
  - 빅데이터 플랫폼 및 센터 구축사업에는 금융, 환경, 문화미디어, 교통, 헬스케어, 유통, 통신, 중소기업, 지역경제, 산림 분야에 대한 주요 10개 분야를 선정하여 플랫폼 주관 기관과 데이터를 제공하는 센터(7개~11개 데이터 기업)를 두고 있음
  - 각 센터는 공공 및 민간기관에서 보유하고 있는 데이터를 공개하고 유통하고 판매하는 등 데이터 거래를 통해 빅데이터 생태계를 조성하고, 신규 사업 창출, 창업 등 데이터 경제를 활성화 할 수 있도록 지원하는 데이터 개방과 혁신을 추진하고 있음
  - 과학기술정보통신부는 ‘빅데이터 플랫폼 및 센터 구축 사업’을 위해 2019년 6월 10개 분야 플랫폼 및 100개 전문센터를 선정. 향후 3년간 1,516억원을 투입할 계획을 발표하였으며, 2019년에는 총 640억원 규모로 사업 추진 중

☐ 데이터산업진흥원(Kdata)의 데이터 거래를 위한 바우처 사업 (2019년)

- 데이터산업진흥원에서는 데이터 바우처 제도를 통해 데이터 공급기업과 이에 대한 데이터 수요기업을 매칭하여 공급기업이 보유하고 있는 데이터를 필요에 따라 거래할 수 있도록 정책적으로 지원하고 있음
- 통신 대기업 등 199곳이 공급기업으로 참여하고 Kdata의 데이터 스토어에 등록하면, 상권분석, 산업별 특허정보 등을 제공하며, 데이터 수집, 분석, 가공 등의 중개서비스를 진행하고, 데이터산업진흥원을 통해 비용을 청구하는 방식
- 중소·벤처기업 등 200개사를 선정하여 데이터 구매 부문과 일반 가공부문, AI가공 부문에 대하여 원하는 데이터 공급자를 검색하여 각 데이터 구매 형태에 따라 1,800만원에서부터 최대 7,000만원까지 지원을 받아 바우처를 사용할 수 있음
- 이밖에도 데이터 거래 활성화에 필수인 개인정보보호법 개정 등 비식별조치, 가명 및 익명 정보 개념을 법제화하여 제도적인 활성화 지원이 기대되며, 이를 통해 중국 등 주변국의 빅데이터 시장과 경쟁우위를 확보하여 한국형 데이터거래소를 성장시키는 구상을 가지고 있음

### 3 데이터 댐의 활용을 돕는 통합 데이터지도

- 과학기술정보통신부와 한국지능정보사회진흥원은 데이터 댐에 축적된 데이터가 효과적으로 검색·활용될 수 있도록 기능을 확대·개편한 통합 데이터지도 서비스를 2020년 3월부터 제공
  - 통합 데이터지도 서비스(www.bigdata-map.kr)는 각 빅데이터 플랫폼에 축적된 데이터의 소재를 한 곳에서 파악할 수 있도록 구성

#### [ 통합 데이터지도 사용법 안내 ]



〈주화면 하단에서 플랫폼 조화선택〉



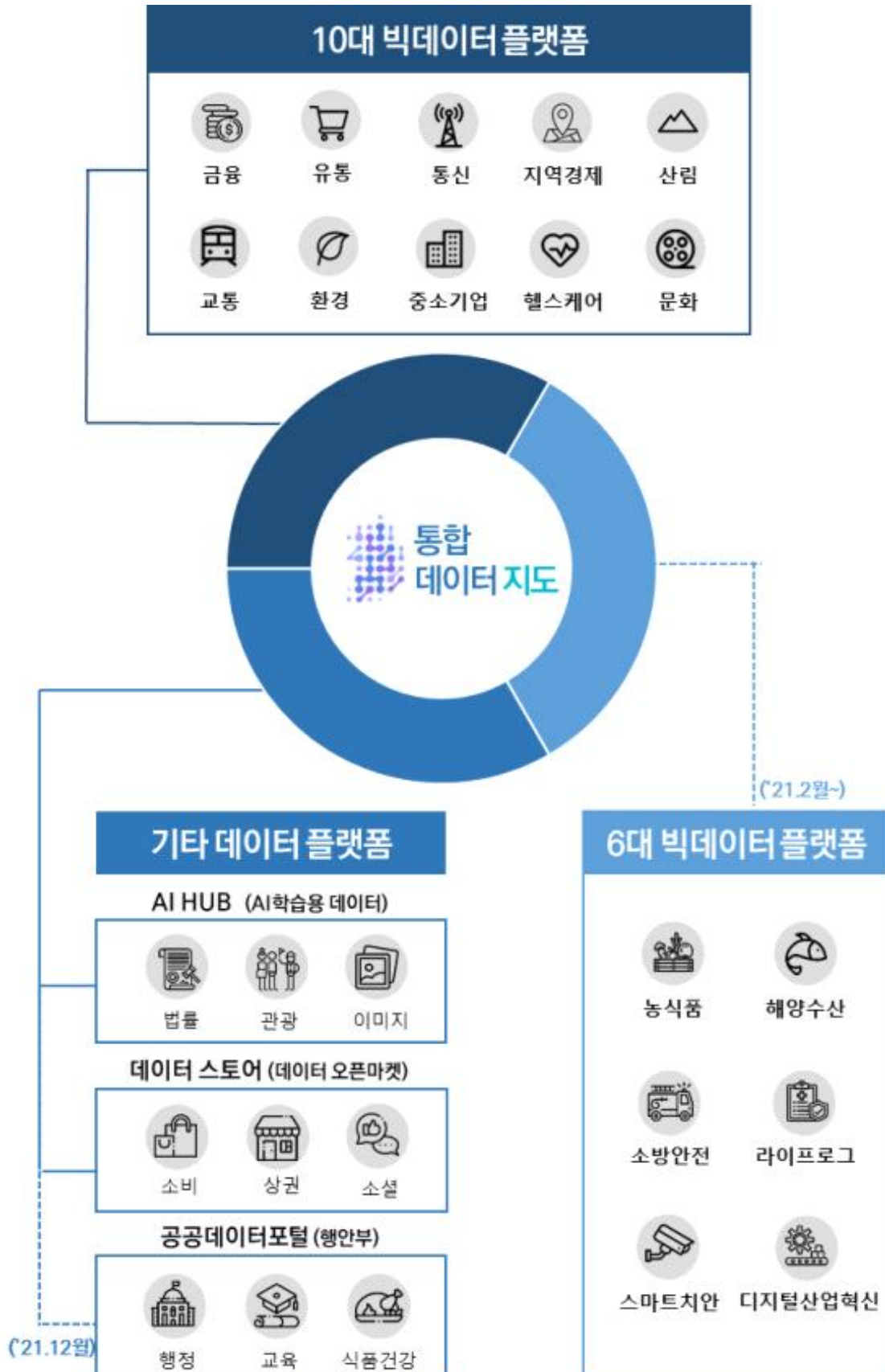
〈해당 플랫폼으로 이동〉

- 2019년에 구축한 10대 빅데이터 플랫폼과 2020년 추가 구축한 6개 분야 빅데이터 플랫폼을 포함한 16개 빅데이터 플랫폼과의 연계 서비스를 제공 중
  - 현재 서비스 중인 16개 빅데이터 플랫폼 : 문화, 통신, 유통, 헬스케어, 교통, 환경, 금융, 중소기업, 지역경제, 산림, 소방안전, 스마트치안, 해양수산, 농식품, 라이프로그 및 디지털 산업혁신
- 또한, 통합 데이터지도는 과기정통부가 구축한 16대 빅데이터 플랫폼 외에도 국내 다른 데이터 플랫폼과의 연계를 확대하는 중

#### [ 통합 데이터지도 서비스에서 연계 중인 데이터 플랫폼 개요 ]

플랫폼명	주소	개요
인공지능(AI) 허브	aihub.or.kr	<ul style="list-style-type: none"> <li>AI 학습용 데이터, AI SW/알고리즘, 컴퓨팅 자원 등 AI개발 인프라 제공 플랫폼</li> </ul>
데이터스토어	datastore.or.kr	<ul style="list-style-type: none"> <li>데이터를 온라인으로 판매하거나 구매할 수 있는 데이터 오픈마켓</li> </ul>
KOTRA 무역투자 빅데이터	kotra.or.kr/bigdata/dashboard	<ul style="list-style-type: none"> <li>해외시장 정보, 수출입 통계 등 국내외 무역 관련 정보 제공 플랫폼</li> </ul>
한국관광 데이터랩	datalab.visitkorea.or.kr	<ul style="list-style-type: none"> <li>이동통신, 신용카드, 내비게이션, 관광통계, 조사연구 등 다양한 관광 빅데이터 및 융합분석 서비스를 제공하는 관광특화 빅데이터 플랫폼</li> </ul>

[ 통합 데이터지도 네트워크 구성도 ]



\* 출처: 과학기술정보통신부 보도자료(2021. 2. 3.)

## III. 데이터 댐 구축 관련 기술 동향

### 1 데이터 기술의 분류

- 빅데이터 산업은 데이터의 생산, 수집, 처리, 분석, 유통, 활용 등을 통해 가치를 창출하는 상품과 서비스를 생산·제공하는 산업으로 구성
  - (후방산업) 데이터 생산 및 수집에 제반이 되는 ICT 인프라 기술, IoT 분야 등이 존재
  - (전방산업) 생성된 데이터를 다루는 데이터솔루션, 데이터구축, 빅데이터 분석 컨설팅 및 서비스와 기술 발전의 중심인 인공지능 시스템이 존재

[ 빅데이터 가치사슬 ]

후방산업	빅데이터	전방산업
센서, 반도체, 컴퓨팅 시스템, 통신 등 ICT 인프라 기술	빅데이터	데이터솔루션, 데이터 구축, 빅데이터 분석서비스, 인공지능 시스템

- 데이터 댐은 데이터 수집 기술, 데이터 축적·가공 기술 및 데이터 활용 기술로 구성된 개념
  - 데이터 수집 기술은 도로, 교량, 상하수도, 기상, 건물 등 국가인프라(SOC)를 포함해 다양한 분야에서 수집되는 원시 데이터들을 적극적으로 수집하기 위한 기술
  - 데이터 축적·가공 기술은 수집된 데이터를 서로 결합하거나 분석, 표준화를 통해 활용 가능한 데이터의 형태로 저장하기 위한 기술
  - 데이터 활용 기술은 자율주행차, 스마트공장, 생활안전, 에너지 절감 등 다양한 분야에 제공, 활용토록 지원하여 AI 혁신서비스를 창출하기 위한 기술
  - 이에 맞춰, 데이터댐의 7개 사업 역시 각 단계에 맞춰 연계되는 형태로 구성되어 있음

[ 데이터댐 7개 사업의 연계 구성도 ]



\* 출처: 과기정통부, AI타임즈



## 2 데이터 수집 관련 기술 동향

### ◎ 다양한 데이터 형태의 빅데이터 수집을 위한 수집 기술, 데이터 분석 기술 개발

#### □ 다양한 분야의 공공 빅데이터의 유형과 형태를 수집하기 위한 특화된 수집기 개발

- 문서, 이미지, 영상, 통계 등 공공 빅데이터의 다양한 유형과 부처별 기관별로 상이한 데이터 개방 규칙에 따른 원활한 수집이 이루어지기 위한 검색엔진 및 Query 기술 개발
- 수집기는 데이터가 존재하는 위치의 오픈 및 수집 정책에 따라 조직적, 자동화된 방법으로 문서와 데이터를 추적하여 필요한 데이터를 수집하는 기술의 개발을 필요

#### □ 수집된 공공 빅데이터에 대한 데이터 보관과 연계통합 할 수 있는 기술요구

- 다양한 형태의 데이터를 통합할 수 있는 ETL(Extract,Transform,Load), CDC (ChangeDataCapture), EAI(EnterpriseApplicationIntegration), ESB(EnterpriseServiceBus)등의 기술의 요구
- 부처별 기관별 공공데이터로부터 수집된 데이터를 정제 및 가공하여 적재하는 데이터 통합연계 기술
- 부처간 공공 데이터를 전자문서 형식을 통해 상호 교환할 수 있는 프로토콜 변환, 전문변환, 거래추적 등을 제공하는 형태의 데이터 교환 및 개방을 데이터 솔루션 도구 형태로 개발이 진행 중

#### □ 정형과 비정형 등 공공 빅데이터를 가공하고 분석하는 기술의 요구

- 데이터의 유형이 다양한 공공 빅데이터를 가공하고 분석할 수 있는 방법론과 기술 개발이 진행 중
  - 수집된 공공 빅데이터에 대한 데이터 수준과 형태를 볼 수 있는 정형과 비정형 데이터의 현황 보고서를 생성하고 모니터링 할 수 있는 기능
  - 문서 등 비정형 데이터를 추출하고 의미있는 단어 조합과 담론을 분석하는 텍스트마이닝 기능
  - 신문기사, 논문 및 특허 등 비교적 표준화된 텍스트 정보를 정형화된 형태로 추출하여 분석에 활용하는 기능
  - 음성 및 영상 데이터를 추출하고 분석하는 비디오데이터 가공솔루션, 음성데이터 가공솔루션 기능
  - 행위자 기반의 공공 빅데이터에 대한 네트워크 처리 및 분석 기능
- 분석 시 활용되는 데이터 정제, 중복제거 등 데이터 프로파일링을 통해 공공 빅데이터에 대한 분석적 가치 향상해 주는 기술
  - 공공 빅데이터 분석에서 비정형데이터 처리 및 데이터 전처리 자동화 기능
  - 빅데이터의 특성이 고려된 분석용 데이터 셋을 구축하는데 필요한 데이터 보정 솔루션
- 통합연계 저장된 공공 빅데이터의 현황정보 및 분석정보를 제공하기 위한 데이터 시각화 분석 기능
  - 정형과 비정형 형태의 공공 빅데이터 유형에 따라 적합하게 보여주는 시각화 기능
  - 수집된 데이터에 대한 사용자의 선택적 데이터 시각화 기능

## ◎ 일상에 배치된 수많은 디지털 센서와 디지털 디바이스, 특히 헬스케어 웨어러블 디바이스에서 이뤄지는 데이터 수집

- 디지털 센서와 웨어러블 기기는 유헬스케어(Ubiquitous Healthcare)와 wBAN(wireless Body Area Network) 기술을 융합하여 신체 정보를 환자나 의사에게 전달하는 역할을 함
  - 디지털 센서는 목적과 제작자에 따라 여러 가지 형태로 제작 가능한 무선 소형 감지 및 데이터 수집 장치
  - 헬스케어 웨어러블 디바이스는 사용자가 자신의 건강 상태를 모니터링하기 위한 자가관리 웨어러블 디바이스와 만성질환자나 심장질환자들과 같은 특정 목적을 달성하기 위해 개발된 메디컬 웨어러블 디바이스, 소형 센서가 내장되어 생체 신호를 모니터링하는 피트니스 및 웰니스 웨어러블 디바이스로 구분. 특히, 피트니스 웨어러블 디바이스는 인포테인먼트와 결합하여 실시간 건강정보를 활용하는 형태로 발전 중
  - 디지털 센서와 웨어러블 디바이스가 의료기기에서 사용될 경우, 외래 및 입원환자의 건강 상태를 구체적으로 계속 모니터링함으로써 의사에게 더 많은 정보를 제공하며 당뇨, 간질, 부정맥, 심부전, 욕창 및 알츠하이머와 같은 다양하고 심각한 만성질환의 모니터링에 도움을 줌
- 최근에는 마이크로프로세서(중앙처리장치)를 포함하여 지능형 센서(Intelligent Sensor)라고도 불리는 스마트 센서의 활용이 증대되는 추세
  - 마이크로프로세서는 센서의 출력 신호를 유지하며, 신호를 통해 낮은 수준의(Low-level) 처리를 수행. 스마트 센서는 대부분 통신 인터페이스를 가지고 있어 디지털화된 신호를 제어 네트워크로 전송이 가능
- 헬스케어 데이터의 경우, 유전자 정보기술로 환자와 일반 소비자 개인의 유전체 정보 획득이 가능해지고 병원 의료 정보시스템, 의료영상기술과 의료정보 기술의 발전으로 환자의 진단, 처방과 진료 내역이 모두 저장되어 있어 과거 병력과 진행 상황의 파악이 가능
  - 개개인이 생성해내는 방대한 양의 비구조화 데이터를 분석하게 되면 기존의 정형화된 치료방식이 아닌 개인에게 맞춰진 정밀 치료를 할 수 있게 되고 효과는 극대화될 수 있음

### 3 데이터 축적·가공 관련 기술 동향

#### ◎ 다양한 형태와 유형의 대용량 빅데이터의 통합수집 저장관리를 위한 인프라 기술

- ☐ 빅데이터의 특성상 데이터 형태의 다양성 문제를 통합하고 표준화해야 하는 문제를 해결하기 위해 메타데이터를 관리하고 데이터의 변경이 일관된 규칙을 따르는 기능 요구
- ☐ 빅데이터의 통합수집과 연계시 필요한 데이터베이스 관리 기술
  - RDBMS, NoSQL DBMS 등 관계형 모델과 비 관계형 구조를 갖는 데이터 관리 기술
  - 빅데이터의 대용량 처리를 위한 데이터 메모리와 하드웨어 저장장치의 성능을 향상시킬 수 있는 인메모리 DBMS와 어플라이언스, 클라우드, 하이브리드DBMS 등 인프라 기술
- ☐ 수집된 빅데이터에 대한 데이터 보관과 연계통합 할 수 있는 기술요구
  - 다양한 형태의 데이터를 통합할 수 있는 ETL(Extract,Transform,Load), CDC (ChangeDataCapture), EAI(EnterpriseApplicationIntegration), ESB(EnterpriseServiceBus)등의 기술의 요구
  - 부처별 기관별 공공데이터로부터 수집된 데이터를 정제 및 가공하여 적재하는 데이터 통합연계 기술 요구
  - 부처간 공공 데이터를 전자문서 형식을 통해 상호 교환할 수 있는 프로토콜 변환, 전문변환, 거래추적 등을 제공하는 형태의 데이터 교환 및 개방을 데이터 솔루션 도구 형태로 개발이 진행 중
- ☐ 빅데이터 통합 데이터베이스에 관리되고 있는 데이터 품질을 측정하고 분석하여 개선데이터를 도출할 수 있는 기능
  - 실시간 데이터 품질 검측, 비정형 데이터 쿼리 매칭 등 데이터 품질 관리 솔루션에 대한 기술적 요구
- ☐ 클라우드 컴퓨팅 기술
  - 클라우드 컴퓨팅은 데이터 분석과 저장 능력의 향상에 있어 중요한 역할을 함
  - 데이터 활용 서비스의 질과 운영 효율성의 향상, 다른 지역과의 정보 공유촉진, 관리 비용 절감을 통해 데이터의 저장, 정보 공유, 손실 예방, 정보 기록 등 다방면에 적용될 수 있음
- ☐ 데이터의 유지 보존 및 폐기
  - 데이터는 장기간에 걸쳐 가치를 창출하거나, 장기간 유지함으로써 가치가 증가하는 데이터도 있는 반면, 시간이 지남에 따라 가치가 감소하거나 수명이 제한되는 데이터도 존재
  - 데이터의 효율적인 운영·이용에 있어 데이터 폐기의 필요성 및 폐기 기준 등에 대한 검토가 필요하며 특히 개인정보 처리의 경우는 복구 불가능한 확실한 폐기 방법 등 미리 검토 필요

## ◎ 다양한 분야에서의 생성하여 사용하는 데이터 유형의 분류 및 탐색 기술의 요구

- ☐ 수집된 데이터들에서 데이터 생성규칙과 분포를 찾아 학습하고 학습 데이터를 근거로 오류 데이터를 판정하는 기술
  - 빅데이터 기반의 분석을 통해 민간과 공공의 영역에서 다양한 분석을 시도할 때 발생하는 데이터 품질과 편향의 문제를 극복하기 위해 데이터로서의 가치를 높이기 위한 AI기반의 데이터 오류 및 보정 자동화 기술
- ☐ 분야마다 상이한 데이터의 유형과 특징을 탐지하고 분류하여 데이터의 보정 및 대체와 재생 방법 추천을 지원하는 기술
  - 데이터의 형태에 따라 데이터 가치에 오류가 발생하는 조건과 생성과정이 다르기 때문에 이를 인식하는 기술이 필요
  - 데이터의 추출, 변환 적재 등 ETL(Extract, Transform, Load)기술의 발달로 인해 영상, 음성, 텍스트 데이터의 추출과 변환 기술이 상당수준 개발되어 왔음
  - 데이터 유형 및 형태에 따른 오류수준을 탐색하는 기술 개발은 고도화되어야 할 부분임
- ☐ 데이터로서의 온전한 가치를 가지는지 판단할 수 있는 데이터 가치유형 분류기술의 요구
  - 장기간 축적되어 시계열적인 추세를 보이는 데이터에 대한 가치판단을 통해 분류되는 기술
  - 횡단/종단 자료 유형을 판단하고, 이에 따른 판단을 통해 데이터 가치를 분류하는 기술(예를들어 횡단데이터는 일정 시점에서 특정 표본으로 식별하는 반면, 종단데이터는 일정기간 변화하는 특정 표본을 식별하므로 데이터를 읽고 오류를 판단하는 방향과 방법이 상이)
  - 패널데이터의 경우 특정 표본을 장기간 동일한 조건에서 추적해야 하므로 동일성 조건 즉, 자료의 신뢰성을 판단하여 분류하는 기술
- ☐ 데이터의 형태와 유형을 탐지하고 분류하여 데이터 특징과 오류를 판단하는 기준에 따라 데이터의 오류형태를 탐지하는 기술

### [ 이상치 오류 판단의 예시 ]

구분	정의	예시
글로벌 이상치(점 이상치)	전체 집단에서 크게 벗어나는 개별 관찰치	이상탐지
상황적 이상치(조건적 이상치)	특정상황(조건)에서 크게 벗어나는 관찰치	상황에 따라 정상적 혹은 비정상적으로 나타나는 데이터
집단적 이상치	전체 집단에서 크게 벗어나는 집단적 관찰치	DDos 공격

\* 출처: 정하영(2019).이상탐지 활용 전자집단민원 추정 방법론에 관한 탐색적 연구. 정보화정책 제26권 제4호.

- ☐ 정형/비정형, 횡단/종단, 시계열 형태 이상치, 결측치, 특이치 유형별 데이터 가치 판단의 근거가 다른 형태를 인식할 수 있는 기술
  - 기존 ETL기술과 AI 및 머신러닝 기술의 결합을 통해 규칙 기반의 데이터 오류검출 기술 개발
  - 데이터에 대한 가치 인식을 가능하게 할 분야와 형태가 다양한 데이터 수집을 통한 학습셋 구축이 선행되어야 함
  - 데이터 가치판단의 일반적인 규칙을 고려하고, 다양한 분야와 유형에서 생성되는 데이터를 학습셋으로 구축하여 AI기반의 머신러닝 과정에 필요
- ☐ 데이터 유형과 형태 그리고 생성과정을 AI와 머신러닝으로 학습하여 유형별, 형태별 오류데이터를 검출할 수 있도록 학습데이터 셋 개발
  - 데이터 수집단계에서부터 데이터가치에 대한 오류 분류 및 탐지가 가능하도록 설계
  - 학습 셋과 수집단 데이터 셋의 비교를 통해 분야별, 유형별, 특징별 데이터의 오류를 식별할 수 있도록 설계
  - 대용량의 정형/비정형 데이터의 학습셋 구축을 위한 인프라 설계와 구축필요
  - 대용량의 계량, 수치 정보 뿐만 아니라, 영상, 음성, 텍스트 등 질적 데이터를 수집하고 학습셋을 구성하기 위해 대용량 데이터의 적재관리가 필요한 서버 및 소프트웨어 구축 필요
  - 데이터의 오류 검출이 가능하게 하기 위한 원문의 정제 작업 및 DB 설계
  - 사용자가 쉽게 학습데이터를 구성하고 오류데이터를 검출하며, 보정과 재생 등의 가치 고도화 작업을 수행할 수 있도록 플랫폼 개발

## ◎ 다양한 분야의 데이터서비스 가치 극대화를 위한 데이터 관점에서의 기초 원천기술

- ☐ 수집된 데이터들에서 오류를 가진 데이터를 자동으로 검출, 정제, 전처리하여 활용에 최적화된 데이터를 만들어 데이터 가치를 향상시키는 기술
  - 단순한 데이터 규칙과 패턴의 오류를 수정하는 것을 넘어 데이터 유형과 특징을 학습하여 특정 유형과 특징을 갖는 데이터의 결측치, 이상치 등을 재생 하여 데이터 가치를 고도화
- ☐ 다양한 분야에서 생성되고 활용되는 데이터의 특성과 유형에 따라 데이터분석을 통한 서비스 품질 향상을 위해 원천데이터의 가치와 유용성을 고도화하는 기술 및 방법론이 요구됨
  - 정형데이터의 자료생성시 결측치가 발생되어 데이터의 연속성을 방해하여 분석시 요구되는 신뢰성을 보장하기 위해 결측값을 처리하는 기존의 방법론 활용 및 개발
  - 시계열 및 종단자료 등 추세를 확인하는 데이터에서 자주발생하는 이상치 및 특이치에 대한 다양한 방법론을 활용한 탐지기술 개발(예를들어 SVD(Singular Value Decomposition: 특이치분해)
  - 특이치분해 방법의 경우 영상 및 이미지데이터의 손실을 복원하기 위해 개발된 방법론으로 네트워크 데이터 및 자연어 처리 등의 분야로 응용확장 되고 있음



- 일반적인 통계방법(모수적방법)을 포함한 비모수적 방법을 다각도로 활용하고 연계하는 기술 및 방법론 개발
- 한가지 방법론 및 알고리즘을 활용하는 것보다는 다각적인 EDA방법을 통해 기존의 데이터 가치를 탐지하는 방법을 응용하거나 연계하여 새로운 형태의 객관적이고 높은 성능의 탐지기술을 개발하는 것이 필요

### [ 이상치 탐지기법 ]

방법	모델	알고리즘	측정기준
모수적방법	일반모델	Grubbs	평균
		Cochran	분산
	회귀모델	Robust regression	오차
		ARIMA	오차
	혼합모수모델	EM	잠재 변수의 분포 추정
비모수적방법	지도학습	KNN	거리 기반 군집화
		SVM	이진 분류
		LOF	밀도 기반
		의사결정나무, 아이솔레이션 포레스트, 랜덤 포레스트	독립 변수의 차원 축소
		나이브 베이지안	이진 및 다중 분류
	비지도학습	히스토그램	밀도 추정
		커널 밀도 추정	밀도 추정
		박스 플롯, 바이올린 플롯	사분위 범위 추정
		다층 퍼셉트론	차원 축소
		K-평균 클러스터링	거리 기반 군집화

\* 출처: 정하영(2019).이상탐지 활용 전자집단민원 추정 방법론에 관한 탐색적 연구. 정보화정책 제26권 제4호. 재구성.

### ☐ 비정형 데이터의 자료 특성에 따른 데이터 가치 품질 측정지표 개발

- 상대적으로 비정형 질적데이터에 대한 결측 및 이상치, 특이치 등의 데이터 가치 측정과 탐지에 대한 기술 및 방법론 개발이 미흡
- 텍스트 데이터의 경우 형태소분석 및 감성 분석 등 원천데이터의 정제가공을 거치기 때문에 데이터 가치의 손실률이 높아 데이터의 분해와 복원 등 고려할 사항들이 많음
- 특히 질적자료는 축적의 시기와 자료의 형태등 정형과 비정형의 특징이 혼합되어 있는 경우가 많아 이를 처리할 새로운 방법론 및 기술개발이 필요

### [ 이상탐지 방법론 ]

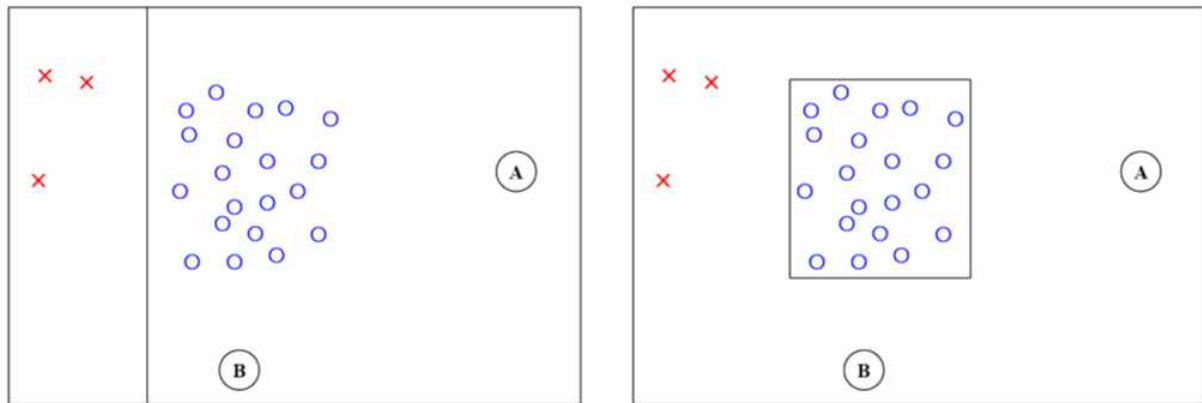
구분	부분	상세	
추정프로세스	시간데이터	시계열 분석	ADF 단위근 검정
		이상치 탐지	박스 플롯
	텍스트데이터	n-gram 벡터 생성	형태소 분석(MPH)
		유사성 분석	레벤슈타인 거리(LD)
	임베딩	시각화	다차원척도법(MDS)

\* 출처: 정하영(2019).이상탐지 활용 전자집단민원 추정 방법론에 관한 탐색적 연구. 정보화정책 제26권 제4호. 재구성.

□ AI 및 머신러닝을 활용한 특이치, 이상치, 결측치에 대한 보정 및 재생 추천기술개발 필요

- 데이터 유형과 특징에 따른 학습데이터 셋을 통해 분류된 데이터 가치 규칙기반으로 특이치, 이상치, 결측치 등 오류데이터 분류를 위한 시기반의 머신러닝 응용 및 연계 기술 개발
- AI 및 머신러닝의 분류 및 추천 알고리즘을 통해 오류 및 정답데이터 학습으로 데이터 대체, 패턴추정, 복원, 제거, 변환, 증폭 등 데이터로서 사용가능한 가치 재생 기술 개발

[ 머신러닝을 활용한 특이치 분석 기술 접근도 ]



Binary classification

Novelty Detection

\* 출처: Novelty Detection - github, 2017

◎ 빅데이터에 대한 보안 기술

□ 무단 액세스에 의한 파손을 방지하기 위해, 접속자의 엄격한 인증(본인 확인) 및 인증(권한 확인)의 구조와 신속하게 복구할 수 있는 백업 등의 대책이 필요함. 백업 데이터 보호는 일반적으로 암호화 기술 및 비밀 균형 조정 기술이 이용

- 블록체인 기반의 빅데이터를 활용한 물류공급망/항만운영 효율화 기술을 통해 화물조달, 화물 운송, 화물 추적 및 감시, 물류 관리, 물류 계약, 물류 금융 등 다양한 물류 영역에 적용 시도 중
- 빅데이터 변조는 무단 액세스, 조작 실수나 시스템 문제이며, 랜섬웨어(Ransomware) 등의 사이버 공격의 피해 등이 예상

□ 은닉 암호화는 암호화 상태 처리에서 얻은 결과를 이용자가 소유하는 개인 키로만 해독한 후 암호화된 결과로 변환하여 이용자에 제공함으로써 빅데이터 시스템 운영자가 데이터 유출 방지

□ 또한, 은닉 검색은 공개키 암호화 또는 대칭 키 알고리즘을 이용하여 AND 검색, OR 검색, 범위 검색, 유사 검색 등을 구현할 수 있는 기술 적용

### 3 데이터 활용 관련 기술 동향

#### ◎ 빅데이터 기반 마케팅 활용

- ☐ 마케팅 인텔리전스 플랫폼(Marketing Intelligence Platform: MIP)이란 빅데이터와 정보 분석 도구와 방법론을 활용하여 마케팅의 의사결정을 지원하는 서비스로 데이터 기반의 시장 및 정책정보를 모니터링하고 분석하여 다양한 통찰과 예측을 제공
- ☐ 마케팅 인텔리전스 플랫폼은 비즈니스 인텔리전스(Business Intelligence: BI)와도 유사하며, 다양한 빅데이터와 데이터 수집도구, 분석도구를 포함한 분석방법론이 결합한 빅데이터 분야 종합기술의 총체
- ☐ 급격한 사회변화와 고객들이 생산하는 다양한 실시간 정보로 기업 입장에서 제품의 모니터링 신제품 기획 시기, 재고량 조절 등 적절하고 신속하게 대응해야 생존할 수 있는 환경이 조성됨에 따라 이를 모니터링하고 예측하는 지능형 플랫폼의 요구가 지속해서 증가
  - 산업 분야별로 기업들의 마케팅 대응방식은 평준화되어 있으며 이에 따른 치열한 경쟁상황에서 시장을 선점하고 유지할 수 있는 지속 가능한 경영을 위해서는 신속한 시장정보 모니터링은 물론 모니터링 결과를 분석하고 예측하여 경쟁사보다 빠르게 시장의 움직임에 대응하는 것을 의미
  - 기존의 마케팅 분석과 대응은 단편적인 정보와 지식, 부분적인 분석으로만 대응하기 때문에 불확실성이 높고 미래를 예측하는 데 한계가 있었으나, 마케팅 인텔리전스 플랫폼의 등장은 복잡하고 다양성 높은 데이터를 결합하여 분석함으로써 예측력과 대응력을 동시에 향상시킬 가능성 증대
  - 다양한 마케팅 정보를 부분이 아닌 전체를 직관적으로 판단할 수 있게 하여 마케팅 담당자가 시장의 이슈를 파악하고 대응하는 시간을 줄여 예산 절감 등 효율적인 마케팅 업무를 수행할 수 있도록 지원
- ☐ 마케팅 인텔리전스 플랫폼의 사용을 통해 마케팅 과정 전반에 걸쳐 효율성을 달성 가능
  - 제품 및 서비스에 대한 고객 이슈를 모니터링할 수 있으며 이를 통해 신제품의 기획과 기존 제품의 개선 사항을 파악하고 조사하는 데 활용
  - 기업이나 제품에 대한 인식이나 감정 등을 모니터링하여 기업 평판과 기업위기 관리에 활용
  - 제품과 경쟁사에 대한 시장 상황을 파악하고 경쟁 관계 설정과 대응 요인 지원
  - 제품 및 서비스의 이슈와 트렌드를 예측하여 향후 시장선점을 위한 제품 및 서비스 기획, 재고량 조절, 공격적 마케팅 시점 계획 등에 활용

□ 마케팅 인텔리전스 플랫폼을 구성하는 요소들은 다음의 세 가지 기술이 적용

- 첫째, 빅데이터 기술로 규모와 다양성, 실시간성, 정확성, 가치추구 등이 그것이며, 기존의 정형 데이터와 함께 영상, 텍스트, 이미지나 음성 등을 포함한 비정형 데이터와 온라인, 1, 2차 자료, 센싱 디바이스(사물인터넷)를 통해 전송되는 데이터 모두를 사용하며, 특히, 비정형 데이터를 정형 데이터와 융합분석이 가능하도록 비정형 지수를 도출하는 기술이 중요
- 둘째, 데이터 인프라기술로 빅데이터를 수집하고, 저장하며 가공처리를 자동화하여 빠르게 데이터를 처리할 수 있는 인프라기술을 사용하는 하둡(Hadoop) 에코 시스템이 중심
- 셋째, 빅데이터 분석기술로 정형과 비정형, 대규모 데이터의 융합적 분석이 가능한 분석 툴과 분석방법론이 사용되며, 통계를 비롯해 머신러닝, 텍스트마이닝(감성 분석 포함), 네트워크 분석 등에 이용

□ 인공지능, 머신러닝, 텍스트 마이닝, 통계 분석 등 정착된 분석 기법과 공공부문과 민간부문에 존재하는 내/외부 빅데이터를 융합해 분석하는 것이 주요 역할인 마케팅 인텔리전스 플랫폼은 비정형 데이터의 사용자 사전 및 감성사전 자동화와 비정형 지수 개발이 이슈

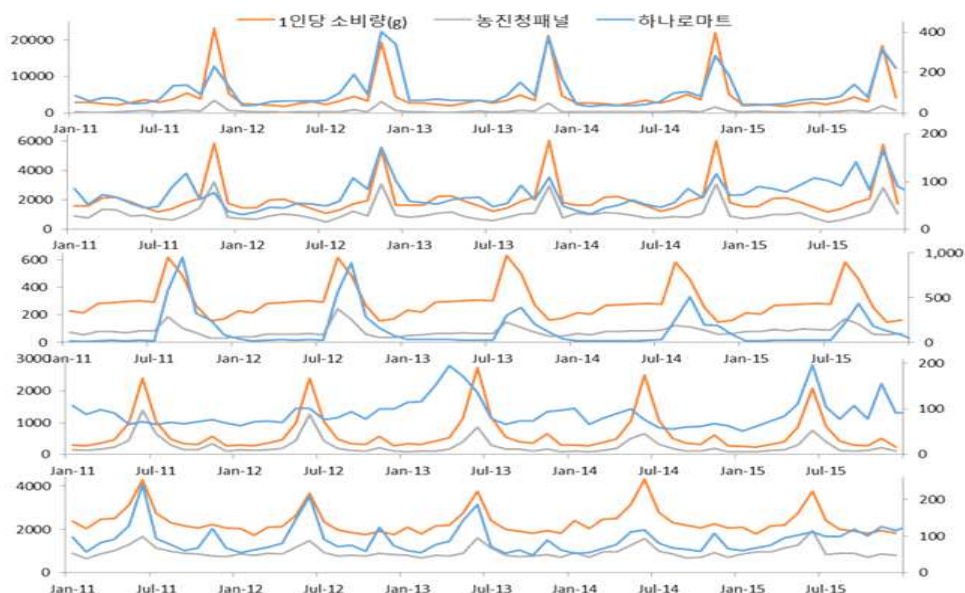
- 텍스트 기반의 내/외부 빅데이터를 수집하여 분석 목적과 특성에 맞게 정제 가공처리 한 후 사용자사전과 감성사전을 도출하고, 이를 통계 및 제3자 데이터와 통합하기 위한 지수화 과정을 거쳐 최종적인 분석결과를 도출하기 때문에 자연어 처리와 사전분류모델 기술, 사용자 데이터 특성에 맞는 크롤링 개발 등의 이슈가 존재

□ 기존의 비즈니스 인텔리전스 플랫폼에서는 비정형 데이터와 정형 데이터의 분석결과가 독립적으로 표출되었으며 이에 대한 시각화 정보의 해석은 오롯이 사용자의 주관에 맡겨야 했으나 비정형 데이터의 지수화를 통해 정형 데이터와 통합된 분석결과 도출은 예측 가능한 패턴과 추세를 시각화해주는 방향으로 변화

- 정형 데이터는 통계 분석, 비정형 데이터는 소셜네트워크 분석 및 텍스트마이닝에 국한되어 있었다면, 비정형 지수를 산출하여 패턴을 도출하여 통찰력을 직관적으로 제공해주는 방법으로 변화
- 텍스트마이닝을 통해 정제된 단어의 빈도, TF-TFIDF, N-GRAM, 연결 중심성 등 비정형 지표와 표준화 수식을 통해 정형화하고, 경제지표, 각종 통계지표, 또는 조사자료를 활용한 지표 등을 통합해 상관성, 인과관계, 교차분석 등 다양한 분석을 시도할 수 있으며, 기존의 단편적인 분석보다 풍부한 예측 가능
  - 단어의 정제를 통해 사용자 사전을 정의하고 구축하는 단계가 무엇보다 중요해지며 이 과정을 통해 특정 이슈나 키워드에 대한 추적과 국가통계체계 간의 비교가 가능해져 비정형 지표의 정량적 객관화가 가능해지며 감성 분석을 활용하는데 인공지능의 역할이 증대
  - 비정형 데이터인 텍스트 데이터를 수집하고 전처리하고 가공하여 분석하는 전 과정에 인공지능의 자연어처리 기술의 결합으로 사용자 사전 구축과 감성사전 구축 등 비정형 데이터의 자동화 기술이 이슈가 될 전망

- 빅데이터 분석에 정착한 분석모델 및 기술과 달리 상대적으로 자연어처리 기반의 텍스트데이터의 사용자사전 및 감성사전의 자동화 방법은 많은 기술개발 노력이 필요한 분야
  - 자연어처리는 분야와 부문별 목적과 특징에 따라 단어정제의 배제와 포함 여부가 주관적인 경우가 대부분
  - 영문텍스트와 달리 음성어인 한글텍스트의 경우 의미가 맥락에 따라 다르기 때문에 많은 경우의 수를 계산한 머신러닝 기법이 적용 필요
  
- 지금까지 빅데이터 분석은 단위요소별로 독립적으로 수행되어 왔으나 데이터의 형태가 다양해지고 분석기술 및 방법론의 발달로 단위요소별로 이루어지던 빅데이터 분석이 융합되어 활용되기 시작
  - 복수의 데이터를 통해 통찰력 있는 분석결과를 제공
  - 플랫폼 서비스의 발달과 시장의 요구에 따라 다각적인 분석을 시각화하여 직관적으로 화면에 노출시킴으로써 사용자의 데이터 활용과 분석을 용이하게 할 수 있도록 텍스트 데이터를 정형화 시키려는 노력 지속
  - 통계자료나 조사자료 등 정형데이터와 텍스트, 영상, 음성, 이미지 등 비정형 데이터의 복수 사용과 융합, 응용분석기술들을 플랫폼에서 구현할 수 있도록 발전
  - 단위요소 기술 간, 이종 간 데이터의 결합을 가능하게 하기 위해서 정형 데이터와 비정형 데이터의 결합이 중요해졌고, 이에 따라 텍스트 데이터 등 비정형 데이터의 지수화 과정이 중요
    - 머신러닝, 베이지안 네트워크, 의사결정모형 등 분류모델 방식과 통계모델의 분석방법론은 결과적으로는 수치로 이루어짐
    - 비정형 데이터를 지수화하여 융합분석이 가능한 지수화 과정이 필수

[ 비정형지수산출을 통해 통계자료와 패턴을 비교한 시각화 화면 ]



\* 출처: 더아이엠씨, 2019



## ◎ 빅데이터 기반 스마트 헬스케어

- ☐ 헬스케어 분야의 데이터가 폭발적으로 증가하고 있는 가운데, 치료는 물론 사전예방 분야에서의 빅데이터 분석 및 활용의 중요성 부상
  - 헬스케어 분야 데이터는 보건, 의료, 건강 관련 디지털 데이터의 양적 증가와 더불어, 데이터의 생산·수집 채널도 다양화
  - 헬스케어 산업의 트렌드도 질병 치료 및 의료기관 중심에서 예방 및 소비자 중심으로 변화함에 따라 빅데이터 분석·활용의 중요성 증대
- ☐ 비교적 엄격한 규제가 적용되는 분야로, 규제샌드박스를 통한 규제완화 및 규제개선 필요성 증대
- ☐ IT 발달로 건강관리 및 의료정보의 접근성이 쉬워지고, 소비자 주권의식이 높아지면서 가능하게 되었으며, 환자는 의료 서비스 결과뿐 아니라 진료 과정의 만족도도 고려하기 시작
- ☐ 스마트 헬스케어 분야에서의 인공지능(Artificial Intelligence, AI) 활용 기술은 기존 의료데이터와 신규 의료데이터, 유전자 데이터, 환자 상태 정보로 인해 방대해진 의료데이터들을 인공지능 기술로 활용하는 것을 말함
- ☐ 인공지능 기술은 다양한 형태를 가진 방대한 규모의 의료 빅데이터를 분석 및 활용할 수 있는 기술 중 가장 주목받고 있는 기술

### [ 스마트헬스케어의 진화방향성 ]

인공지능 기술	헬스케어 적용
머신러닝, 딥러닝	대규모 의료 빅데이터를 기반으로 스스로 학습하고 데이터를 분석함으로써 질병 예측, 신약 개발 촉진 및 의료진에 대한 의사결정 지원
영상 인식 기술	컴퓨터가 환자의 MRI(Magnetic Resonance Imaging), PACS(Picture Archiving and Communication System) 등 의료영상데이터의 의료영상 이미지를 학습 및 분석하여 질환에 대한 진단정보를 제공하여 의사의 진단과 처방을 지원
음성 인식 기술	진료 시 의사와 환자 간 대화를 음성 인식 시스템을 통해 자동으로 컴퓨터에 입력, 저장하는 의료녹취 서비스 제공으로 의료기록 작성에 소요되는 시간 단축
자연어 처리 기술	임상시험 적합 환자 선별과 같이 방대한 자료를 이해하고 검토 및 분석하는 경우, 자연어 처리 기술을 적용하여 의료진의 업무 부담 경감 및 의료 업무 효율성 극대화

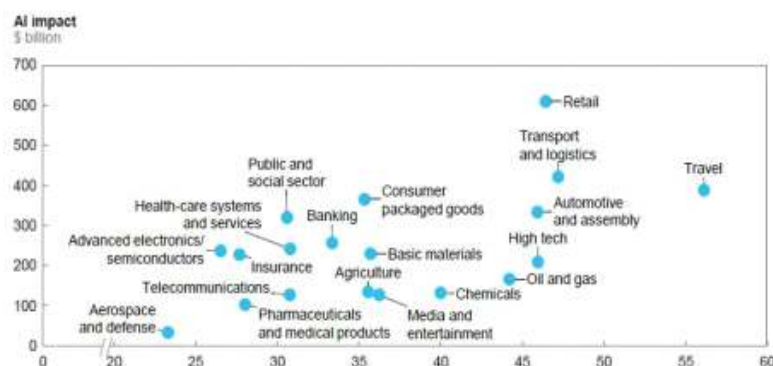
\* 출처: ETRI, 인공지능 헬스케어의 산업생태계와 발전 방향

- ☐ 헬스케어 산업의 패러다임이 질병이 발생한 후에 치료를 받는 치료·병원 중심에서 스스로 건강을 관리하는 예방·소비자 중심으로 변화하면서 헬스케어 산업 내 빅데이터 분석이 더 중요해지고 있음

## ◎ 빅데이터 기반 유통/물류 시스템

- ☐ 집하부터 배송까지 다양한 이해관계자들이 얹혀 있으며 톱니바퀴처럼 일사불란하게 움직이는 물류는 다른 산업에 비해 많은 트랜잭션이 발생
  - 인프라를 가지고 있는 사업자를 연결하는 포워딩(국제물류 주선사업)의 중요성 증대
  - 수송 규모의 확대와 소량·다품종 처리 시설 도입 등에 따라 물류 창고의 대형화 및 고도화
- ☐ 물류공급망의 효율화를 위한 IT 시스템 접목
  - 데이터를 기반으로 미래를 예측하여 과잉 재고와 재고 부족을 막는 데 활용
  - 복잡한 프로세스와 불확실성을 보완하기 위해 디지털 포워딩 업체들이 주목
- ☐ 택배 송장 유통 인덱스를 활용하여 온라인 소비 트렌드를 분석하고, 소셜데이터 버즈를 분석하여 온라인 소비자 니즈를 파악
  - 온라인 상품 가격정보를 수집하여 가격전략 수립에 이용하며 상권분석을 통해 매출 예측 기반으로 입지 전략을 수립하거나, 와이파이를 이용한 방문자 위치정보를 수집하여 유동인구를 상세히 파악
- ☐ 물류 유통산업 빅데이터 적용을 위한 중요 요소
  - (판매 이력 데이터) POS 데이터로 대표되는 판매 이력 데이터의 경우, POS를 통해 정산된 구매일시 구매품목 개수, 금액, 구매점포 등의 데이터가 기록
  - SCM(Supply Chain Management)은 일반적으로 최종 소비자에게 전달되는 제품과 서비스의 형태로 가치를 창출하는 다양한 프로세스와 활동하는 조직의 네트워크로 정의
  - (핵심 알고리즘과 분석) 통합재고관리, 제품 탐색 및 매치, 택배 매치, 배송가격 결정, 택배 선택, 다이내믹 라우팅, 고객 커뮤니케이션, 수요 예측 등
  - (물류 산업의 디지털 요소) 물류 통제 타워, 서비스 분석론, 디지털 크로스 보더 플랫폼, 도시 물류, 당일 배송, 드론, 자율주행 트럭, 클라우드 소싱, 순환경제, 공유운송능력, 공유보관능력 등

### [ 물류산업에서의 빅데이터/인공지능 영향력 ]



\* 출처: McKinsey&Company, 2019

## IV. 데이터 댐 관련 주요 플레이어 동향

### 1 해외 플레이어

#### ◎ 빅데이터 수집/분석

☐ (Clickworker) 기계 학습 및 인공지능 AI 교육을 위한 데이터 세트 수집 및 분석을 제공

- 특정 학습 목표에 대한 요구 사항을 충족하는 AI 훈련 데이터를 텍스트뿐만 아니라 사진, 오디오 및 비디오 녹화와 같은 고유하고 새로 생성된 교육 데이터 제공
- 이미지 주석, 텍스트 마이닝을 통한 텍스트 분석, 학습 기반 알고리즘의 출력 평가 등 데이터 라벨링 및 유효성 검사 제공

☐ (Qburst) 데이터를 비즈니스 과제 해결에 도입 및 실제 사용

- 기업이 페타바이트 단위의 데이터를 처리하고 그 데이터로부터 가치를 창출할 수 있도록 함. 기존의 데이터 분석 방법이 불충분하다고 판명될 경우, 실행 가능한 정보를 제공
- 기계 학습, 데이터 분석, 데이터 시각화 데이터 저장고, 데이터 처리, 데이터 관리 및 BI 제공
- 데이터 기반으로 인공 지능 솔루션, 소셜 미디어 분석, 로봇 프로세스 자동화 등의 솔루션 제공

☐ (ScienceSoft) 분석, 컨설팅, 구현, 지원 및 빅데이터 서비스를 포함한 모든 범위의 빅데이터 서비스 제공

- 빅데이터 구현 및 개선 컨설팅
- 데이터 품질 관리 및 데이터 보안을 설정하고, 실행하여 기업이 예측 높은 정확도를 얻을 수 있도록 빅데이터 솔루션 구현
- 데이터 추출 및 정리, 머신러닝 모델 트레이닝, 조정 등 빅데이터 관리 분석 서비스 제공

#### [ ScienceSoft의 데이터 분석 서비스 기술 구성 요소 ]



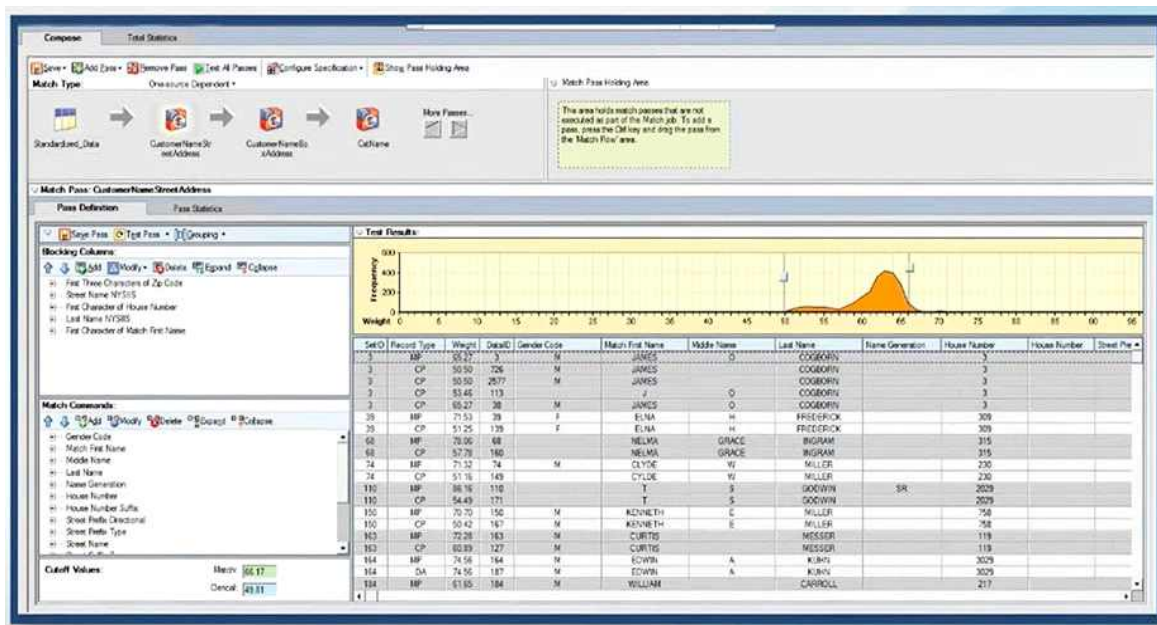
\* 출처: ScienceSoft

## ◎ 데이터 품질 관리

□ (IBM) 데이터를 신뢰할 수 있는 정보로 전환하기 위한 품질관리 솔루션을 제공

- 데이터와 그 관계를 이해하는데 도움이되는 중단 간 데이터 품질도구를 제공하여 데이터 품질을 모니터링하고 데이터 정리, 표준화, 데이터 계층 정보를 제공
- 소스데이터를 조사하고, 정보표 표준화 및 레코드 일치를 자동화시키고 있으며, 부정확하거나 일관되지 않은 데이터의 확산을 줄이는데 필요한 기술을 제공하고 있음

### [ IBM InforSphere ]



\* 출처: IBM 제공

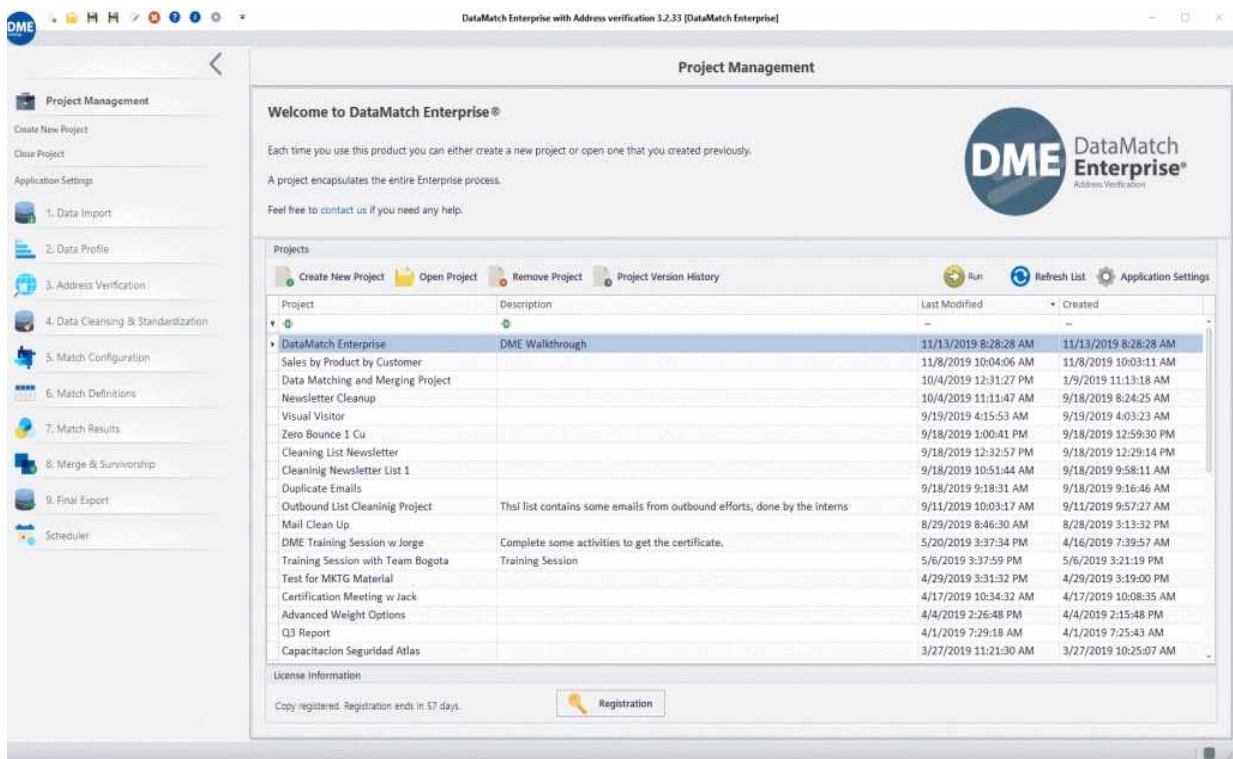
□ (Oracle) 폐쇄적인 데이터를 가지고 있는 기업 내의 품질관리를 구현하기 위한 소프트웨어를 개발하고 전사적 차원에서의 품질관리가 가능하도록 구현

- 조직내 연구개발, 제조, 고객센터, 규제, 공급업체에 대한 각기 다른 데이터를 취급하는 조직들의 맞춤형 품질관리 도구를 개발
- 조직내 연구개발, 제조, 고객센터, 규제, 공급업체에 대한 각기 다른 데이터를 취급하는 조직내 생산된 데이터의 중복방지과 중복제거 및 표준화, 중복레코드 식별, 이상치 탐지 및 유사한 유형의 데이터로 여러 클라이언트를 처리할 수 있는 시스템 통합자 및 컨설턴트가 사용할 수 있는 데이터 인터페이스 사용

□ (Data Ladder)데이터관련 솔루션은 광범위한 데이터 품질관리를 다루고 있음

- 데이터 강화 솔루션은 기존 데이터베이스에 누락된 정보를 추가되는 매칭을 통해 원시 데이터가 사용되기 전에 정리 및 최적화 되었는지 확인해줌
- 데이터 중복제거 소프트웨어로는 독자적인 알고리즘을 통해 데이터 중복문제 해결의 정확도를 향상시킴
- 데이터의 위치와 형식에 관계없이 퍼지매칭 및 머신러닝 알고리즘을 사용하여 일치하는 레코드를 찾고 데이터를 병합하여, 중복항목을 제거할 수 있게 함

[ DataMatch Enterprise ]



\* 출처: Data Ladder 제공



## ◎ 빅데이터 분석 및 시각화

- ☐ (Google) 클라우드 플랫폼의 빅데이터 분석 기술인 ‘클라우드 데이터플로우(Cloud Dataflow)’ 서비스와, 페타 바이트의 데이터 장 분석용 클라우드 서비스 개발하여 제공 중
  - GFS(Google File System)는 빅데이터의 저장을 위한 분산 파일 시스템으로 일반 상용 하드웨어를 이용하여 대량의 서버를 연결하여 데이터에 대한 접근이 효율적이고 안정적
  - Map Reduce는 빅데이터의 분산 병렬 컴퓨팅으로 처리하기 위한 분산 프로그래밍 기술
  - 압축된, 고성능 사유 데이터 스토리지 시스템인 Big Table 기술은 관계형 DB의 테이블 형태로 빅데이터를 저장 가능
  - Google은 그래프 알고리즘의 처리를 지원하기 위한 기술인 Pregel, 대용량 데이터를 분산 처리로 빠르게 분석할 수 있는 기술인 Dremel, 검색 인덱스를 작성하기 위한 기술인 Percolator 등의 빅데이터 관련 기술 및 도구들을 직접 개발하여 사용하고 있으며, 대용량 데이터의 실시간 처리 기술인 Dremel을 Big Query로 상용화하여 서비스 제공
  - Google 브레인팀이 개발한 딥러닝 플랫폼 TensorFlow는 다양한 작업에 대해 데이터 흐름 프로그래밍을 위한 오픈소스 소프트웨어 라이브러리로 딥러닝 관련 제반 도메인의 추론을 위한 기반 데이터를 Google의 클라우드로 수집하는 역할 수행
- ☐ (Oracle) 빅데이터 환경에 요구되는 시스템의 개념 아키텍처와 이와 함께 빅데이터 분석 참조 아키텍처를 발표
- ☐ (Spotfire) 시각화 도구 Spotfire, 자연어 검색 및 AI 기반 인사이트를 제공하는 Spotfire X, 다양한 데이터 소스에 대한 스트리밍 분석을 제공하는 Spotfire Data Streams 등의 다양한 솔루션 제공
  - 데이터 시각화를 위한 사용자 친화적인 인터페이스와 대화형 대시 보드 및 포인트-앤드-클릭 방식의 데이터 탐색 지원
  - 스마트한 데이터 발견, 데이터 랭글링, 지리 공간 분석, 원 클릭 예측 분석을 통해 의사결정을 개선할 수 있도록 지원
- ☐ (Data Planet) 개인과 정부, 다양한 기업, 기관이 보유한 공공데이터를 라이브 에디터를 활용해서 인터랙티브 시각화 라이브러리 서비스. 무료 차트 및 3D 입체 지도 제작의 데이터 시각화 서비스 제공
  - 소셜, 개인 블로그, 원하는 사이트로 차트를 공유해 다양한 목적과 용도로 활용

[ 주요 기업 동향 ]

주요 기업	주요 내용
Google	<ul style="list-style-type: none"> <li>빅데이터 분석 기술 ‘클라우드 데이터플로우’ 서비스 중</li> <li>페타 바이트의 데이터 장 분석용 클라우드 서비스 개발</li> </ul>
Oracle	<ul style="list-style-type: none"> <li>빅데이터 환경에 요구되는 시스템 개념 아키텍처 및 분석 참조 아키텍처 발표</li> </ul>
Spotfire	<ul style="list-style-type: none"> <li>시각화도구 ‘Spotfire’, 자연어 검색 및 AI 기반 인사이트 제공 ‘Spotfire X’, 다양한 데이터 소스에 대한 스트리밍 분석을 제공 ‘Spotfire Data Streams’ 등 다양한 솔루션 제공</li> </ul>
Data Planet	<ul style="list-style-type: none"> <li>개인, 정부, 기업, 기관이 보유한 공공데이터를 라이브 에디터를 활용, 인터랙티브 시각화 라이브러리 서비스 및 무료 차트, 3D 입체 지도 제작의 데이터 시각화 서비스 제공</li> </ul>

\* 출처: 빅데이터 분석 솔루션(빅데이터를 비즈니스에 활용하려는 수요가 시장 성장 견인) 요약 보고서, 2019.11, 한국IR협의회

[ 빅데이터 분석 서비스 분야별 제공업체 ]



\* 출처: Learnbigdatatools

## ◎ 물류/유통

- ☐ (DHL) Resilience 360은 데이터 취합과 평가를 통해 공급 사슬망 전반에 산재한 위험(risk)을 모니터링하고 미리 예측, 대응할 수 있도록 정보를 제공하는 시스템
  - 공급사슬 가시화, 리스크 평가, 사고 모니터링, 리스크 대응, 포탈시스템 등으로 구성되어 있으며, 리스크 평가는 국가 위험도와 자체 공급사슬 위험 노출도, 열(온도) 지도 등을 통해 평가·구현
  - 실시간 모니터링 기능은 사고 실시간 추적, 현장 피드백을 통한 경보 등을 수행하고, 리스크 대응 기능을 통해 응급 배송계획 수립 등을 수행
- ☐ (Amazon) 12개 이상의 Amazon 데이터 센터는 실시간으로 자사 데이터를 받아들이고 반영하여 시스템과 앱을 확장하며, 분권형 유통망 시스템(Decentralized Distribution System)을 통해 중앙의 물류 시스템이 지역별 물류센터를 빈틈없이 연결하여 자율주행자 동차와 로봇, 드론을 이용해 라스트 마일 시스템을 실제로 적용
  - 예측 물류 기법을 통해 소비자 데이터를 분석해 구매할 가능성이 높은 물품을 예측해 배송을 미리 준비하는 시스템으로, 자사 AI 플랫폼 알렉사를 글로벌 전자 기업들이 주요 스마트 가전에 탑재하여 그 정확도를 한층 높이고 있음. 이는 Amazon 물류 시스템 전체와 연결되어 소비자의 구매 예상 물품을 미리 준비할 수 있도록 전달하는 역할
- ☐ (Amazon) 자체 유통물류 서비스 사업 풀필먼트 바이 아마존(Fulfillment By Amazon) 시행
  - 아마존은 빅데이터 분석에 기반한 수요와 구매를 예측하여 배송시간을 단축. 또 물류 인프라와 설비 통합을 위한 투자를 확대해 물류 효율성을 높이고 있다는 보고

### [ 4차산업혁명의 기술과 Amazon의 대응 ]

활용 기술	서비스명	내용
빅데이터, 인공지능	Amazon 쇼핑(기존 상거래)	• 추천, 상품배열 등의 최적화, 배송 예측 등을 통해 소비자 만족 극대화
IoT, 인공지능	알렉사(Alexa)	• 인공지능 쇼핑 보조 OS로 최초에는 스피커인 에코(Echo)에 탑재
	Amazon 고(Amazon Go)	• 결제과정이 없는 쇼핑으로, 등록된 이메일로 청구서를 받는 쇼핑 시스템
클라우드 컴퓨팅	Amazon 웹서비스(AWS)	• 클라우드 컴퓨팅 사업을 직접 운영하여 Amazon의 새로운 캐시카우로 성장
로봇	키바 로봇(Kiva robot & systems)	• 물류 시스템에 도입하여 비용 절감, 배송시간 단축 등을 계획
드론	프라임 에어	• 배송 시스템에 드론을 도입

\* 출처: KIET 산업경제

- (트라센) 프랑스 대형 해운회사인 CMA CGM은 대형 컨테이너 선박에 실시간으로 화물을 모니터링할 수 있는 기술인 트라센(TRAXENS)을 적용
  - 트라센 기술은 스마트 컨테이너에 장착되어 컨테이너 간, 관련 인프라들 간에 커뮤니케이션을 가능하게 해주며, 육상 또는 해상에 있는 컨테이너로부터 실시간으로 데이터를 수집하여 컨테이너선, 고객, 보험회사, 세관 등에 유용한 정보(위치, 온도, 습도, 진동, 침입 여부, 통관 수속 상태 등)를 제공
- (UPS) ORION 솔루션으로 약 46,000대의 트럭에서 나오는 텔레메틱스 센서들로부터 데이터가 수집되며 차량의 속도, 방향, 제동, 차량의 성능 등의 정보가 포함
  - 이 데이터는 배송 기사의 평가 지표로도 활용되지만, UPS 중앙 서버로 수집되어 빅데이터를 구성하여 궁극적으로는 UPS 기사들이 이용하게 되는 주요 경로들을 재설정하는 데 사용
  - UPS는 이러한 목적을 위해서 On-Road Integrated Optimization and Navigation(ORION)이라는 시스템을 구축하고 그간 축적된 데이터, 2억 5천만 개의 주소 데이터, 그리고 지도 데이터를 기반으로 배송 기사들의 배달 및 집하 업무를 실시간으로 조절
- (히타치제작소) 축적되는 업무데이터를 자동으로 저장하고 이를 분석하는 인공지능(AI) 창고관리 시스템을 개발하여 효율적인 업무를 달성
  - 인공지능에 과거 작업지시와 관련된 빅데이터와 실시간으로 발생하는 상품 출고 실적 데이터를 입력한 후 출고에 걸린 시간과 공간 혼잡도 상관관계를 분석해 출고 작업의 효율을 분석하고, 이 작업효율 모델을 출고 카트 시스템에 반영해 다양한 상품을 카트에 집어넣는 순서를 최적화
  - 인공지능에 의해 최적화한 작업지시서를 발행하여 직원은 작업지시서에 나온 순서대로 창고를 돌면서 지시된 상품을 꺼내 카트로 선적
  - 물류 창고에서는 공간이 비좁아 한 선반에 한 명씩밖에 작업할 수 없기 때문에 다음 직원은 대기할 수밖에 없으며, 이때 인공지능을 통하여 같은 시간에 직원이 겹치지 않도록 순서를 조정하는 빅데이터 활용으로 평균 작업 시간을 8% 단축
- (야마토 예측 시스템) 과거 화물 수취시간의 데이터를 축적, 분석해 고객이 집에 있을 가능성이 높은 시간대를 예측하는 시스템으로, 20%에 달하는 첫 번째 배달 부재율을 낮출 수 있어 배송 기사는 남은 시간을 활용해 배달 시간을 세분화하는 등 보다 효율적인 서비스 제공이 가능
- (자라) 빅데이터 분석을 활용해 전 세계 매장의 판매 현황을 실시간으로 분석한 뒤 고객 수요가 높은 의류를 실시간으로 공급할 수 있는 물류망을 구축함으로써 재고 부담은 줄이고 매출은 극대화하는 성과

## ◎ 스마트 헬스케어

### ☐ (Apple) 헬스케어 플랫폼 'HealthKit' 출시로 의료 데이터와의 연계 시도

- iOS를 지원하는 다양한 개인건강기기(PHD: Personal Health Device)와 개인건강응용(PHA: Personal Health Application)로부터 활동량, 심박 수, 몸무게, 혈압, 혈당 등 여러 가지 개인건강정보(PHI: Personal Health Information)를 수집하고, 여러 PHA가 데이터를 공유할 수 있는 모래시계 구조(Hour-glass Architecture)를 채택하여 개발
- HealthKit의 출시로 미국 내 주요 22개 병원과의 연계가 가능

### ☐ (Google) 독보적 인공지능 분야 기술력, 10억 명 이상의 고객, 자금력, 스타트업 투자 및 자회사를 바탕으로 스마트 헬스케어 시장을 선도중

- 자회사 : Calico - 노화 관련 질병, DeepMind - AI 적용 헬스케어, Verily - 의료데이터 헬스케어

### ☐ (Epic Systems) 8백만 병의 개개인으로부터 EMR(Electronic Medical Record)을 수집하는 제품 'Cosmos'를 출시

- 에픽시스템즈는 미국 내 EMR 시장 점유율 54% 이상과 전 세계 환자들 중 2.5%에 대한 데이터를 관리
- 이를 통해, 에픽 시스템즈는 2억 건 이상의 환자 기록 즉, 빅데이터 분석을 통해 약물의 부작용과 연관성을 검색하고 도출 가능

### ☐ (IBM) 데이터 관련 기업의 인수에 40억 달러 이상을 투자하고, 미국 내 주요 병원과 협력 관계를 맺으며, 잠재력이 큰 스타트업에 투자하는 등 IBM을 중심 생태계 형성

### ☐ (Philips) 환자의 생명 정보를 모바일 기기를 통해 실시간으로 확인하는 커넥티드 모니터링 솔루션 서비스를 제공하며, 영상진단장비/초음파/마취기 및 신생아 중환자 관리 등 스마트 헬스케어 사업으로 영역을 확대

### ☐ (Quell) 뇌파와 신경을 자극해 천연 마취제인 엔케팔린을 생성하고, 통증 신호가 뇌로 전달되는 것을 방지하는 웨어러블 신경자극 기기를 개발

### ☐ (Zephyr) 웨어러블 센서가 내장된 셔츠를 이용하여 운동선수의 성과를 심장박동수, 신진대사, 신체 포지션, 폐활량을 통해 측정

## ◎ 마케팅

- (Amazon) 1억 2,000만명의 고객정보를 보유한 장점을 활용해 구매 내역과 검색어, 상품탐색 정보 등을 활용하여 자체적인 플랫폼을 운영
  - 대표적으로 날마다 사이트에 접속하는 사용자의 선택이나 결정 데이터를 수치화해 목표상품에 접근하는 시간을 고려하여 상품탐색 경로를 제시하고 고객 구매전환을 촉진하는데 내부의 빅데이터와 플랫폼을 결합
  - 고객의 과거 구매정보를 이용하여 특정 고객의 과거 구매품목, 시기, 상품검색 기록, 위시리스트 등을 활용하여 특정 시기에 예상되는 구매목록 리스트를 만들어 소비자와 가까운 물류센터로 예측 배송하는 서비스를 구현
- (IBM) 현업부서에서 생성되는 모든 정형 및 비정형 데이터를 활용하여 잠재적 리스크를 확인하였고, 사기방지 등에 활용
  - 매출 하락을 확인하고 스마트데이터 디스커버리를 사용하여 고객만족에 따라 구매량이 달라짐을 확인, 고객이탈성향에 대한 예측모델을 개발
  - 이탈 가능성이 가장 높은 고객을 찾아내어 새로운 주문 및 조정된 가격을 제공하여 고객이탈을 줄일 수 있었음

## [ IBM의 BI(Business Intelligence) ]



\* 출처: IBM 홈페이지, 2019



- (Anexinet) ListenLogic Omni-Channel Analytics Platform을 개발하여 고객경험에 대한 심층적인 이해와 행동 패턴을 분석하여 마케팅에 활용할 수 있도록 함
  - 기업이 운영하는 콜센터기록, 이메일, 채팅, 텍스트 및 CRM정보와 같은 다양한 데이터소스를 통해 고객의 패턴을 분석할 수 있도록 함
  - 비정형 데이터로부터 통찰력을 얻을 수 있도록 직관적인 시각화 정보를 세부적으로 제공하며, 기계학습을 사용하여 고객과의 관계를 이해하고 예측할 수 있도록 함으로써 향후 구매 가능성을 예측하는데 기여
  - 이를 통해 콜센터 운영비용의 20% 이상을 절감하고 고객경험을 15~20% 향상시킴
- (REI) 90TB나 되는 데이터를 통해 협동조합 멤버십을 분석하였음. 이를 통해 아웃도어 유통과정에서 추가적인 신규 고객을 다수 확보하는데 성공
  - 90TB 이상의 구매정보, 고객보유현황 정보들을 분석하여 고객만족도를 향상시키고 배송방법을 개선
  - 고객의 수명주기관리와 제품 범주 분류와 같은 의사결정에 도움이 되는 정보들을 제공하였고 이를 통해 아웃도어 소매업에 대한 오프라인 혹은 온라인 투자처를 결정하는데 도움
- (HelloFresh) 디지털 마케팅보고가 시간이 많이 걸리고 수작업으로 비효율적인 것을 중앙집중식 비즈니스 인텔리전스 솔루션을 도입하여 전환율을 높임
  - 자동화 구축을 통해 마케팅 분석팀의 작업시간을 하루 10~20시간을 절약하여 더 큰 규모의 마케팅 팀과 개별화된 마케팅 캠페인을 기획할 수 있는 역량을 갖게 됨
  - HelloFresh는 고객행동분석 집계를 토대로 세 가지 구매자 유형을 만들었으며 실시간으로 데이터를 보고 추적할 수 있도록 하였음 그 결과 전환율이 증가하고 고객 유지율을 향상
- (Tableau) 2019년 비즈니스 인텔리전스를 기반으로 한 인텔리전스 플랫폼에 대해 설명 가능한 AI를 추가하여, 음성 데이터를 자연어 처리하고, 데이터 설명기능으로 데이터 모델링 또는 전문지식 없이도 누구나 데이터에 대한 인사이트가 가능
  - 사회공헌을 확대하는 데이터 협업, 데이터와 발맞추는 윤리강령, 최신 BI 플랫폼으로 수렴되는 데이터 관리, 기업의 새로운 언어가 된 데이터 스토리텔링, 더 스마트하게 분석을 채택하는 엔터프라이즈, 데이터 과학자의 역량을 강화하는 데이터 민주주의, 최신 BI 채택에 동력이 되는, 가속화된 클라우드 데이터 마이그레이션에 대한 10개를 선정

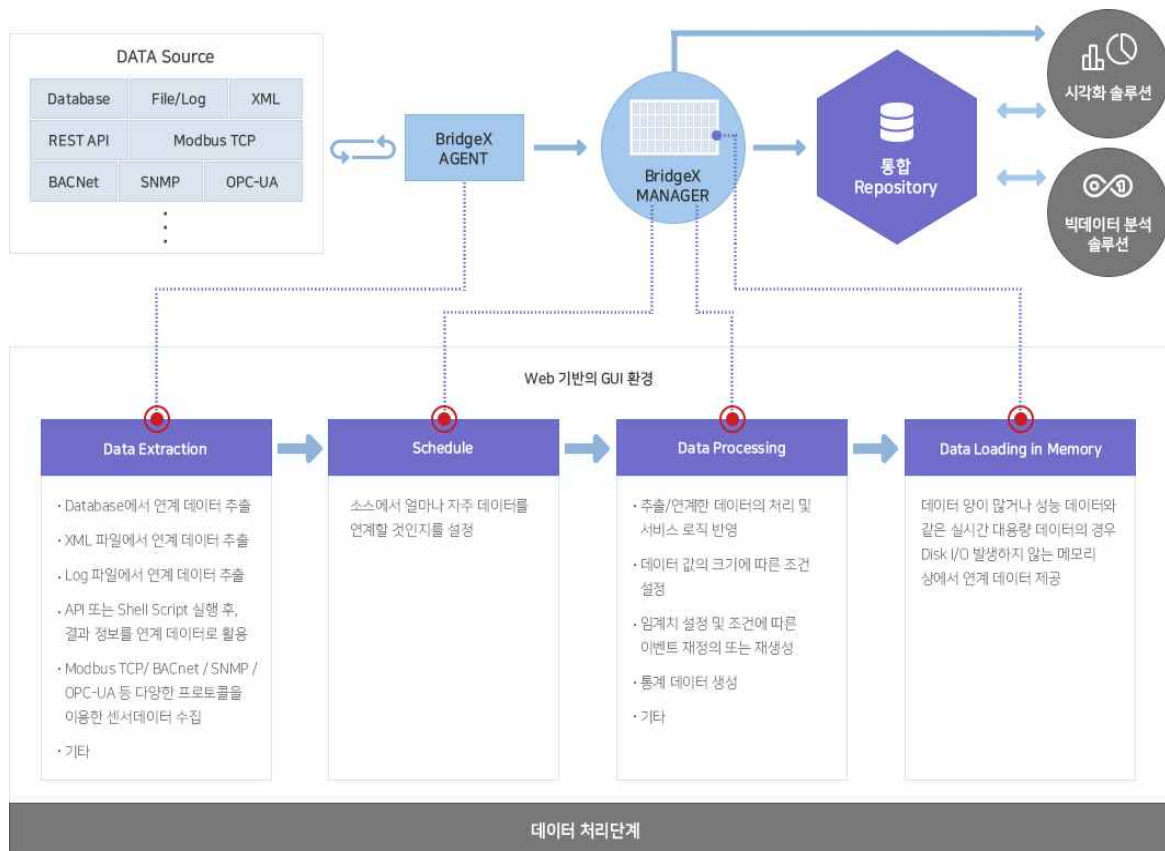
## 2 국내 플레이어

### ◎ 빅데이터 수집/분석

#### □ (위엠비) BridgeX는 빅데이터 수집 활용 솔루션

- 다양한 데이터 소스로부터 데이터를 수집하고 저장하는 빅데이터 수집 솔루션으로 연계 대상 정보로부터 주기적으로 정보 연계를 자동화함으로써 자동화의 신뢰도와 안정성을 보장
- 추출/연계된 데이터는 통합 레포지터리로 저장되거나 서비스 로직이 반영된 의미있는 데이터로 재해석 되어 사용자에게 실시간으로 제공

[ BridgeX의 시스템 구성도 ]



\* 출처: 위엠비

□ (빅인사이트) SI기반 데이터가공, 데이터분석 솔루션, 고객사 맞춤 컨설팅 및 개발 제공

- 기존에 보유 또는 수집하고 있지만, 활용되지 않고 있는 데이터 또는 분석을 위해 추가로 필요하지만 수집이 되고 있지 않은 데이터 공급, 수집/설계, 데이터를 통한 인사이트 도출 등 전반적인 데이터 SI 수행
- 머신러닝 데이터 분석도구, 빅데이터 처리 기술의 설계, 구축, 분석, 관리 뿐만 아니라 기업에서 활용중인 기존의 BI 및 웹 분석 도구의 비교/평가를 통해 비즈니스 목표 실현을 위한 문제 해결 지원
- 기술적 지원을 통해 지속적인 UX의 탐구와 실험을 적용하고 최적화할 수 있도록 설계, UI/UX, 개발 및 운영/유지보수에 걸친 전 과정 제공

□ (와이즈넷) 빅데이터 정보수집 솔루션 'WISE BICrawler', 빅데이터 의미분석 솔루션 'WISE BIC Analyzer'

- WISE BICrawler는 빅데이터 수집과 언어분석 기술 기반, 소셜 데이터 분석에 최적화된 빅데이터 정보수집 솔루션. 대규모 데이터를 정확하고 빠르게 수집, 클렌징 기술로 수집 데이터의 스토리지 공간을 최소화함으로써 불필요한 스토리지 운영으로 인해 발생하는 데이터저장 비용 및 관리 업무의 시간을 절감
- WISE BIC Analyzer는 비정형 빅데이터를 의미 분석함으로써 필요 정보의 획득 시간 및 비용을 절감, 빠른 의사결정을 수립

[ WISE BICrawler의 시스템 구성도 ]



## ◎ 빅데이터 품질관리

- (데이터스트림즈) QualityStream은 데이터의 품질을 모니터링 하고 정제하며, 프로파일링을 통해 사용자들이 데이터를 검증할 수 있도록 데이터 품질정보를 제공
  - 사용자들이 일치하지 않는 데이터를 발견하고 속성을 이해할 수 있도록 표준 통계측정값과 비교를 통해 데이터를 검증
  - 중복된 데이터 또는 비표준 및 알 수 없는 데이터 유형을 표준화하여 데이터 셋으로부터 개선된 인사이트를 도출할 수 있도록 지원
  - 거버넌스 규칙 및 절차를 통해 데이터의 일관성 및 정확성을 확인해 데이터 품질 개선유지
- (위세아이텍) 머신러닝기반의 데이터품질관리와 데이터정비를 지원하는 웹기반의 데이터 품질 및 정비 솔루션 WiseDQ를 제공
  - 데이터 탐색을 통한 패턴을 분석하고 특질을 파악하여 기초통계를 제공
  - 데이터 값에 근거한 도메인의 자동 추천과 정확한 데이터 범위나 업무규칙이 없어도 인공지능기반의 비정상적인 값을 탐지

[ WiseDQ의 시스템 구성도 ]



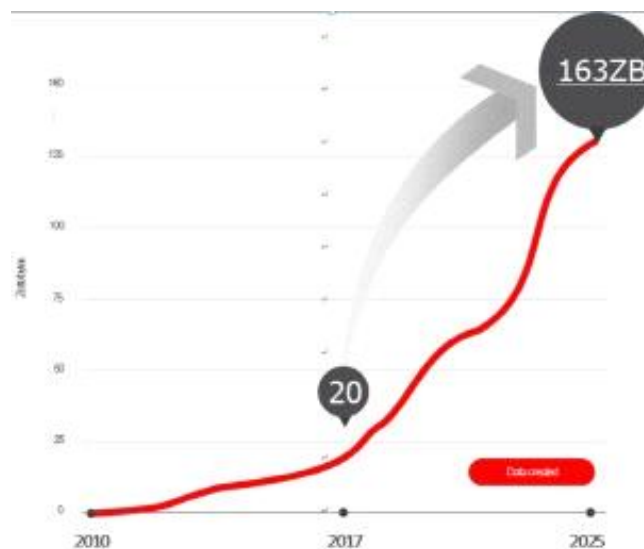
\* 출처: 위세아이텍 제공

- (GTONE) DQ시리즈는 데이터 품질관리 솔루션은 다양한 데이터 유형에 대한 품질관리를 지원
  - DQMiner는 기업데이터에 대한 자동 데이터 프로파일링, 데이터 오디팅, 데이터 품질분석과 결과보고 제공
  - DQXpress는 정부의 공공데이터의 관리지침준수여부에 따라 품질관리의 수준평가와 인증을 지원하는 진단도구
  - DQIoT는 사물인터넷환경에서 생성되는 실시간 스트림 데이터를 품질관점에서 분석하고 모니터링

## ◎ 빅데이터 분석 및 시각화

- (엑셈) 머신러닝 자동화(AutoML) 솔루션 기업인 데이터로봇(Data Robot), 빅데이터 분석 솔루션 기업인 나임(KNIME)과 파트너 계약을 체결하여, 데이터에서 최적의 알고리즘 조합을 찾아내 결과를 예측함과 동시에, 분석된 데이터 시각화까지 제공하는 서비스인 '아울아이(Owleye)' 사이트를 '19년 오픈
  - 다양한 고객사들을 대상으로 데이터 분석에 대한 전문적 지식을 공유할 예정
  - 경쟁력 있는 글로벌 솔루션 기업인 나임과 데이터 로봇과 파트너십을 맺음으로써, 고객사에 더욱 차별화되고 발전된 데이터 분석 서비스 제공 목적
  - 생산성 최적화 및 Seamless 설비 운영을 위한 인공지능기반 제조상황 진단/예측 시스템 개발
- (KT 넥스알) 클라우드 환경에서 분석하는 신규 빅데이터 플랫폼 '콘스탄틴'을 출시
  - 제주도에 구축된 스마트 관광 플랫폼을 통해 공공 와이파이로 얻은 내·외국인 관광객의 위치 정보를 통신, 카드, GIS 같은 지리 데이터와 융합해 유동인구와 상권 경쟁력, 관광지 추천과 관광 수요 예측 등 다차원 분석 정보를 도출
  - 초대용량 데이터를 클라우드 환경에서 실시간으로 저장하고 분석하는 기능

[ KT 넥스알 빅데이터 증가량 ]



\* 출처: KT NexR(<https://www.ktnexr.com>), 2019.

- ☐ (솔트룩스) 빅데이터의 수집, 변환, 분석, 시각화, 의사결정 지원에 이르는 빅데이터 분석 가치사슬 전체를 커버하는 최고의 플랫폼과 시스템을 구축
  - 빅데이터 분석 플랫폼인 BigO는 자연언어처리(NLP)와 기계학습, 분산병렬 처리와 같은 핵심 기반 기술과 80억 건 이상의 소셜 및 공공 빅데이터에 기반
  - 소셜 빅데이터 분석뿐 아니라 기업 빅데이터 분석, 통신 및 금융 빅데이터 분석 등 다양한 도메인에 활용 가능
- ☐ (효성인포메이션시스템) 100% GUI 기반의 간편한 환경의 데이터 수집 및 통합부터 머신러닝 모델 구축, 모델기반 고급분석, 모델 업그레이드, 시각화 및 리포팅을 모두 제공하는 히타치 벤틀라의 원스톱 빅데이터 플랫폼인 '펜타호' 개발
  - 데이터통합, 머신러닝 기반 분석, 대시보드 및 시각화를 통해 GUI 기반 데이터 전처리, 분석 및 예측 모형, 고객 맞춤형 대시보드로 데이터 엔지니어, 데이터 분석가와 과학자, 비즈니스 분석가와 고객 등 다양한 사용자의 업무 자동화를 지원
- ☐ (에스씨플랫폼) 기업의 내부데이터와 외부데이터를 비교 분석하여 가공한 맞춤형 정보 기반의 플랫폼 및 데이터 서비스를 목표로 중소기업의 생산성을 높일 수 있는 재무, 작업 및 사업관리와 관련한 관리용 솔루션의 개발 및 보급과 기업에 필요한 시장정보를 맞춤형으로 제공
- ☐ (투그램시스템즈) 선박 엔진에서 추출한 각종 정보를 빅데이터 기술로 분석하여 현재 상태를 진단, 고장을 비롯한 미래의 변동 상황을 예측하여 고장이 나기 전에 정비할 수 있도록 지원하는 빅데이터 솔루션인 '선박엔진고장예측시스템(e-CBM)'을 제공
- ☐ (그루터) 오픈소스 빅데이터 분석엔진 개발업체로 '아파치 타조'에 관해 국제적인 명성을 얻을 정도로 중요한 공헌자로 평가받고 있으며, AWS의 빅데이터 분석 플랫폼 서비스인 EMR에서 동작하는 타조와, 자체 클라우드 분석 플랫폼 서비스인 TaaS(Tajo as a Service)를 선보임
- ☐ (데이터스퀘어) 소상공인의 상표권 침해 예방을 위한 빅데이터 분석 플랫폼을 개발하였다. 국내 소상공인과 자영업자가 기존에 상표권을 보유하거나 신규로 등록할 경우 해외 및 국내에서 자사 상표권의 침해 여부 및 관련 상품의 동향 모니터링 및 경쟁사 기업의 상표권 동향 모니터링과 해당 결과의 리포팅 서비스를 지원할 예정



## ◎ 물류/유통

- (삼성SDS) 스마트 물류 솔루션인 ‘Cello’를 통해 기존의 WMS 방식에서 탈피한 데이터 관리 및 최적 의사결정 제공을 지원하고 있으며, 이는 3PL을 넘어 4PL이라고 칭할 만큼 다양한 물류 효율화 방안을 제공
  - ‘Cello Virtual Warehouse’는 WMS 데이터를 분석해 다차원분석(Multi-Dimensional Analysis)을 지원하는데, 이를 통해 관리자는 자사의 창고운영 과정에서 현재 우수한 부분은 어떤 곳이고, 어느 부분을 개선해야 할지 알 수 있음 또한 창고 내 작업자의 동선, 재고 배치 등에 대해 다양한 대안을 시나리오 형태로 제공하며, 시뮬레이션 엔진을 통해 각 시나리오의 효과를 분석해 이를 알기 쉬운 형태로 시각화하여 제시
  - VR을 통해 창고의 현재 운영 상태를 보여주는 기능까지 지원하는데, 이는 WMS의 데이터를 바탕으로 배치 현황과 개선안을 시나리오별로 가상현실로 구성해 보여줌으로써 효과적인 의사결정을 지원하는 방식
  - Cello Loading Optimizer는 전 세계적으로 사용되고 있는 삼성SDS의 대표적인 적재 시뮬레이션 도구로 박스는 물론 팔레트, 컨테이너 등 다양한 적재 작업을 지원하며, 사전 설정을 통해 사용자가 원하는 방식으로 적재가 이루어질 수 있도록 도움
- (CJ대한통운) 최첨단 ‘3D Visibility(가시성) 시스템’을 개발하여 현장에 적용
  - 창고 내에 각 랙 공간마다 RFID(무선주파수인식) 칩을 부착해 특정 랙 공간에 보관되어 있는 제품의 정보를 중앙시스템에서 실시간으로 파악, 그 정보를 터치스크린 화면에 3D 영상으로 보여주는 시스템
  - 1단부터 층층이 보관되어 있는 제품의 정보를 한눈에 파악할 수 있는 장점이 있으며, 터치스크린 방식이기 때문에 보고 싶은 랙이나 셀을 찾아 이동하면서 상세정보를 확인 가능
- (한국네트웍스) ‘한국테크놀로지’ 그룹 소속으로 매년 매출의 3% 이상을 R&D에 투자하여 통합물류솔루션(WMS, TMS, OMS, FMS, VMS 등) Suit 완성
  - 물류솔루션과 자동화솔루션을 연계하는 설비제어 소프트웨어(WCS, ACS, MFS 등) 영역과 IoT 활성화를 위한 IoT 플랫폼 및 솔루션을 강화
- (위비즈테크) 빅데이터 전문기업으로 데이터 관리에서부터 데이터의 시각화, 빅데이터 분석과 클라우드 인공지능을 활용한 예측 등 각종 통합 솔루션 및 플랫폼 서비스 제공
  - 축적된 빅데이터 개발 사업수행을 바탕으로 Google Map 서비스와 빅데이터 등을 가공, 모델링해 시각화하여 전달하는 솔루션 Wiver를 출시
  - 클라우드 기반의 머신러닝 자동화 플랫폼인 Smart AEO Platform을 개발하여 공인인증 프로세스 자동화, 위험요소 모니터링, 감사일 자동 알람 등의 지원. 빅데이터로부터 정보를 추출하고 결과를 분석, 예측하는 서비스를 제공

- ☐ (광원엘에스) 물류창고자동화설비 제조 기업으로 AS/RS 자동화 창고는 수직 공간을 활용하여 보관물량을 늘리고, 비용, 시간을 최소화시킴. 물류자동화 시스템은(KLAS) 단순 입출고에 한정된 소규모 보관의 자동화 창고부터 복잡한 WIP(Work In Process) 라인까지 적용 가능
- ☐ (밸류체인씨엔티) Global 솔루션(JDA, Mahantan, SAP, IBM)과 자체 개발한 솔루션(VC Suite)을 가지고 기업의 여건에 따라 선별적으로 구축
  - 특히 JDA Logistics Solution의 경우 ABI, 삼성전자, LG전자, CJ대한통운, 판토스 등의 글로벌 법인을 대상으로 'Implementation Service'를 제공
  - OM, WM, TM, FIS, BMS, VMS 각 모듈별로 비즈니스 진행하고 있으며 19년 매출 100억 내외
- ☐ (네오시스템즈) 물류 솔루션 '인트라로지스(IntraLogis)'를 개발했으며, 통합물류관리 클라우드 서비스인 '로지스허브(LogisHub)'를 론칭
- ☐ (택트레이서) 자율주행 기반 무인 재고관리 시스템(드래곤플라이)를 개발하고 있으며 기존의 바코드와 RFID 기술의 단점인 비효율성을 자율주행 방식을 활용한 창고관리 시스템을 안산시, 경기테크노파크, 과기부의 지원을 받아 개발 중
- ☐ (에이아이엠) 화장품, 전자부품처럼 정밀화를 요하는 제작 시스템의 조립 및 품종별 공급 시스템을 개발, 상용화에 성공했으며, '플렉시블 피딩 시스템'으로 불리는 자동화 생산 설비로 부품 공급부터 조립까지 완성
- ☐ (랩투마켓) 자동화 장비 및 시스템을 개발하는 기업으로 첨단 물류 장비·시스템 개발하여 화주기업의 물류시스템에 대한 최첨단 기술적용 및 지속적인 최신성 유지가 가능하고 IoT-빅데이터-생산·물류 자동화가 연계 및 융합 종합시스템 구축
- ☐ (진코퍼레이션) 공급망관리 솔루션을 개발하는 기업으로 WMS 창고관리시스템은 입고, 적치, 재고, 피킹, 출고 등 물류센터 프로세스 전체를 통합 관리하여 고객기업의 물류관리 및 운영능력을 향상시킴으로써 경영자원의 유용한 활용과 고객서비스 향상을 지원하는 시스템
- ☐ (더블유앤이케이) 스마트제조자동화를 구현하는 기업으로 WMS 창고관리시스템은 창고 내의 재고 흐름을 추적 및 통제하는 시스템으로 재고관리 및 창고운영을 용이하게 하며, 이노비즈 기업 인증을 보유

## ◎ 스마트 헬스케어

- 디지털 헬스케어 스타트업을 전문 육성하는 엑셀러레이터 ‘디지털 헬스케어 파트너스’가 출범했고, 세계적으로 기술력을 인정받은 ‘루닛’과 같은 국내 헬스케어 스타트업 등장
  - 데이터의 측정, 분석, 연계, 통합까지 전 데이터 사이클 주기를 둘러싼 제품 및 서비스가 활발하게 출시되고 있음
  - 주로 병원에 축적된 의료 빅데이터를 중심으로 비즈니스 기회가 열리고 있으며, 복지부의 규제 완화로 제한적이지만 소비자 의뢰(DTC) 유전자 분석 시장도 열림
  - 헬스케어 데이터 통합·연계 서비스도 등장했으며, 정부의 정책적 노력도 이어지고 있음

### [ 국내 데이터 기반 헬스케어 비즈니스 유형과 가치 창출 ]

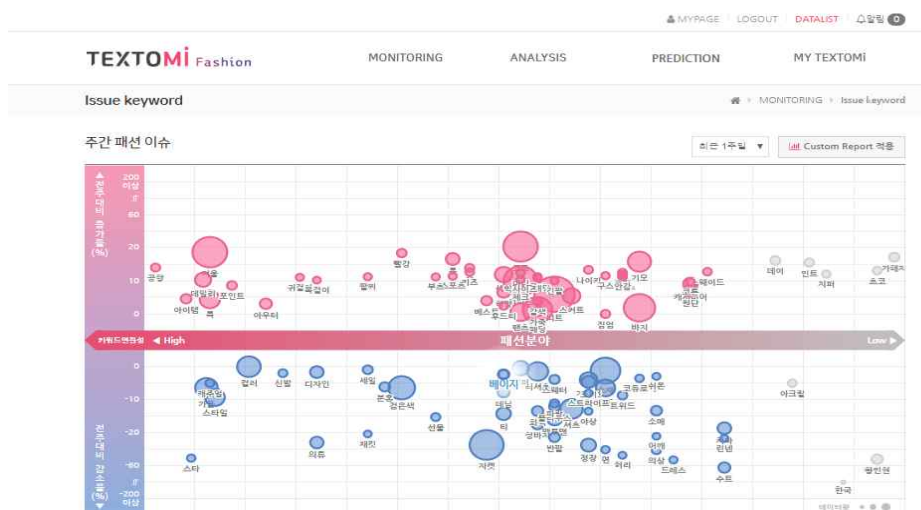
비즈니스 유형		주요 기능 및 기업	새로운 가치 창출
데이터 측정	건강데이터 측정 기반 건강관리 서비스	<ul style="list-style-type: none"> <li>• 일반인의 자가건강관리 (활동량, 심전도, 자세 등)</li> <li>- 사례 : 엠트리케어 등</li> </ul>	<ul style="list-style-type: none"> <li>• 치료에서 예방으로 의학 패러다임 전환</li> <li>- 일반인의 적극적 건강관리로 의료비 절감</li> </ul>
		<ul style="list-style-type: none"> <li>• 혈당, 혈압 측정을 위한 만성질환관리</li> <li>- 사례 : 핑거앤, 닥터다이어리 등</li> </ul>	<ul style="list-style-type: none"> <li>• 상급종합병원의 업무 효율화에 기여</li> <li>- 만성질환 자가관리로 상급종합병원의 업무 효율화에 기여</li> </ul>
데이터 분석	병원 내 축적된 의료 빅데이터 분석	<ul style="list-style-type: none"> <li>• 주요 상급종합병원의 빅데이터 센터</li> <li>- 사례 : 서울 아산병원, 서울 성모병원, 고대 구로병원 등 의료영상 데이터 및 빅데이터 센터 개소</li> </ul>	<ul style="list-style-type: none"> <li>• 개방형 혁신 플랫폼으로서 병원 혁신</li> <li>- 병원에 축적된 임상 빅데이터 활용으로 연구개발을 통한 수익 경로 확보</li> <li>- 새로운 의학적 통찰력 발견의 가능성 확대</li> </ul>
		<ul style="list-style-type: none"> <li>• 임상 데이터 분석 스타트업</li> <li>- 사례 : 뷰노, 루닛</li> </ul>	<ul style="list-style-type: none"> <li>• 의사 역할 및 1, 2차 의료 전문성 보완</li> <li>- 1, 2차 의료의 전문성 보완</li> <li>취약계층의 의료접근성 확대</li> </ul>
	소비자의뢰 유전자 분석 서비스	<ul style="list-style-type: none"> <li>• 규제 완화로 소비자의뢰 시장 형성</li> <li>- 사례 : 제노플랜, 디엔에이링크, 테라젠이텍스</li> </ul>	<ul style="list-style-type: none"> <li>• 정밀의학 구현</li> <li>- 유전체 분석을 바탕으로, 증상이 아닌 질병 전에 근거한 진단과 치료 가능</li> </ul>
데이터 연계·통합	헬스케어 데이터 연계	<ul style="list-style-type: none"> <li>• 기 구축된 병원 빅데이터에 건강측정데이터 통합 및 연계</li> <li>- 사례 : 라이프시맨틱스</li> </ul>	<ul style="list-style-type: none"> <li>• 맞춤 의료/수요자 중심 헬스케어 구현</li> <li>- 분절된 여러 종류의 데이터를 연계, 통합하여 수요자 중심/맞춤 의료 구현에 기여</li> </ul>

\* 출처: 디지털헬스케어혁신동향과 시사점

## ◎ 마케팅

- (다음소프트) 자체 보유하고 있는 막대한 빅데이터를 활용하여 통신 및 카드데이터와 융복합 교차분석을 시도하여 사회이슈를 분석
  - 비정형 데이터뿐만 아니라 정형 데이터와 결합하여 AI를 활용한 다양한 분석으로 패턴을 모니터링하고 예측
  - 특히, 데이터 검색 및 분석 인사이트를 발굴하는 솔루션과 자연어처리 기반을 위한 사용자 인터페이스 솔루션을 최신 이슈로 발굴하고 사회현상을 분석하고 인사이트를 제공하는 리포트를 제작
  - 다음소프트는 다양한 분야에서 축적되고 있는 비정형 데이터를 강점으로 AI와 결합시켜 내재된 패턴을 찾는 데 주력하고 있음. 이를 적용하여 관광동향, 여론, 주식시장 분석 등에 활용
- (더아이엠씨) 온라인 포털 및 뉴스, 소셜미디어의 텍스트 데이터 기반의 패션 인텔리전스 플랫폼인 패션 텍스트ومي 라는 플랫폼을 개발하고 패션 아이템, 색상, 패턴 등 관련 정보를 모니터링할 수 있으며, 패션 관련 키워드를 5주 단위로 출현 가능성을 예측 가능
  - 패션 텍스트ومي는 비정형 데이터 특히 텍스트 데이터를 활용하여 패션 관련 키워드로 이슈를 모니터링할 뿐만 아니라 패션유통기업 및 브랜드에 대한 평판을 감성 분석으로 모니터링하도록 하여 기업의 위험과 신뢰도를 모니터링 할 수 있도록 의사결정을 지원
  - 텍스트마이닝에 적합한 자연어처리기술과 키워드 정제를 통한 사용자 사전 및 감성 사전을 분야별로 구축하고 있는 것이 큰 장점
  - 시계열예측 방법인 ARIMA 모형을 통해 패션 아이템과 색상, 브랜드 등의 등장 빈도를 예측할 수 있음 인공지능 및 머신러닝 기법을 활용하여 예측결과를 정밀하게 할 수 있는 분석기법을 개발하고 있으며, 패션뿐만 아니라 지역경제, 교육 분야로도 확장

[ 패션 텍스트미 메인화면 ]



\* 출처: 패션 텍스트ومي(<http://fashion.textomi.co.kr>), 2019

- (데이터솔루션) 빅데이터 분석 플랫폼인 빅스테이션(BigStation)을 자체 개발하고 주요 공공 및 금융기관의 시스템통합 사업을 수행
  - 데이터솔루션은 데이터 컨설팅, 분석, 구축 및 운영에 이르는 분석과 인프라에 이르는 플랫폼 분야 과정을 종합적으로 수행
  - 빅데이터 플랫폼 인프라 부문에서는 빅데이터 수집 저장에 필요한 스토리지, 서버 및 네트워크, 가상화 플랫폼, 클라우드 포털 솔루션을 제공
- (SK텔레콤) 지오비전이라는 상권분석 및 타겟 마케팅 지원 서비스를 개발했으며, 통신 데이터와 유동인구, 지리정보, 소비업종과 상품판매 현황을 종합적으로 분석
  - 통신 데이터를 보유하고 있는 것이 강점이며 지리정보와 결합하여 신규 가치를 창출하고, 카드 및 금융, 유통정보 등 이종 간 데이터와 결합했을 때 시너지가 극대화되는 것이 장점
  - 한편, 소셜미디어 데이터, 검색키워드를 이용하여 고객, 사회이슈파악, 기업이미지 파악 등 모니터링에 활용하는 자체 빅데이터 솔루션을 개발

[ SK텔레콤의 상권분석 및 타겟 마케팅 지원서비스 ]



\* 출처: SK텔레콤 홈페이지, 2019

- (KB국민카드) 빅데이터 분석을 통해 카드이용 서비스 및 편의성을 제공하면서, 소비자의 행동 패턴을 분석하여 마케팅에 활용
  - 신규상품 개발 및 상품 추천에 활용, 또한 실시간 마케팅 시스템을 개발하여 고객이 자신의 니즈와 위치에 따라 최적화된 카드 혜택 및 맞춤형 정보를 실시간으로 검색
  - 카드 이용객의 행동 패턴을 분석해 혜택가맹점 앱을 개발하여 고객에서 맛집 추천 및 비슷한 고객의 상점 이용 패턴을 제공하고 있는 것이 특징
- (NC소프트) 사기탐지 알고리즘을 고도화하여 불법행위를 모니터링하는 시스템을 개발하여 불법적으로 다른 사람의 자산을 이용하거나 가져가는 행위에 대한 탐지
  - 수많은 이용자들이 게임을 이용하면서 남기는 방대한 규모의 비정형 로그데이터를 실시간으로 분석하는데 데이터마이닝과 기계학습기술을 활용

## IV. 결론 및 시사점

### ◎ 데이터 품질 및 가치향상과 안정성 확보를 통한 견고하고 유용한 데이터 댐 건설

#### ☐ 데이터의 유용성 확보와 다목적 활용가치 향상을 위한 품질가치 고도화

- 데이터 댐의 최종적인 목적은 유용한 데이터의 다목적 활용에 있으므로 수집되는 데이터의 형태와 특징, 유형 등에 구애받지 않는 데이터 수집과 탐색, 그리고 가공이 필요
- 데이터 댐의 특성상 다양한 횡/종단 정보와 정형/비정형 데이터를 융합하여 활용해야 하는 만큼 융복합 분석에 활용할 수 있는 데이터의 품질과 표준화가 선행되어야 함
- 또한, 목적에 따라 정량적 분석과 비정량적 분석 등 다양한 분석을 시도할 수 있도록 이상치, 결측값, 특이치 보정과 재생 등 분석 데이터로서의 가치를 향상시키는 고도화가 반드시 필요

#### ☐ 데이터의 기본적인 리스크인 데이터 무결성 검증이나 데이터 보완에 대한 리스크에 대한 방안이 선제적으로 필요

- 데이터 수집부터 표준화, 가공, 거래, 결합 등 데이터 생태계 전 분야에 걸쳐 이루어지며, 데이터 수집부터 가공, 그리고 데이터의 삭제 및 폐기까지의 데이터 생애 전부가 데이터 댐에서 이루어지기 때문에 데이터 댐 역시 데이터의 기본적인 속성 문제에서 자유로울 수 없음
- 수집된 데이터의 진위 검증 가능 여부 및 품질 확보 이슈, 확보된 데이터가 수집 과정에서 악의적 목적으로 제3자에 의하여 위, 변조 되지 않았는지, 또는 혹은 시스템 상의 오류로 데이터의 훼손이 발생하지 않았는지 등에 대한 위협이 있는지를 수집된 데이터의 관리 및 감독할 필요가 존재

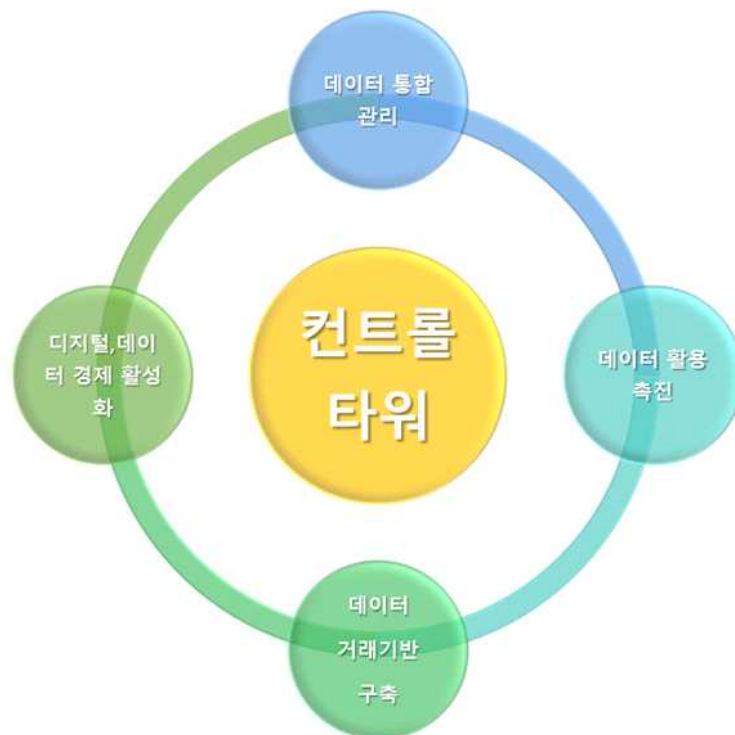
#### ☐ 따라서, 데이터 댐은 다양한 형태와 유형의 데이터를 탐색하고 통합하여 수집할 수 있어야 하며 수집된 데이터의 품질과 분석적 가치향상을 고도화 할 수 있어야 함은 물론, 수집된 데이터가 외부에 노출되어 발생하는 피해 및 데이터 속에 저장된 개인정보를 포함한 기밀정보의 유출을 예방할 수 있는 선제적 방안 마련이 필수적



◎ 공공 빅데이터 수집 및 분석시스템 분야의 컨트롤타워 구축 및 제도적 한계 극복 등 국가경쟁력의 확보가 시급

- 공공기관 또는 민간에서 공공의 목적으로 생성하고 관리하고 있는 공공 빅데이터는 태생적 한계로 기술적인 수집과 연계가 어려워 이를 가능하게 할 제도적, 조직적체계 기반이 선행되어야 함
  - 부처간 데이터 거버넌스의 부재로 데이터 칸막이를 해소하기 위한 제도적인 규칙마련과 정책적 지원이 선행되어야 함
  - 각기 다른 공공데이터의 생성규칙을 표준화하고 중복생성의 문제 등 데이터 관리 업무의 조직 체계 개선이 필요
  - 공공 빅데이터의 수집을 용이하게 하기 위한 전산파일 생성의 표준화와 규격화 문제도 해결해야 함
- 공공과 민간의 데이터를 수집탐색하고, 표준화 근거와 규칙, 데이터 가치에 대한 정책적 지원을 담당할 국가 데이터 컨트롤타워 및 조직신설을 통한 데이터 댐 관리
  - 공공데이터(행안부)와 민간데이터(과기부)로 이원화 되어 있는 데이터 관리 관점과 정책을 통합관리할 수 있는 조직이 필요
  - 데이터 생산기관(지자체, 정부부처, 공공기관 등)의 데이터를 표준화하고 가치판단 규칙을 정책적으로 판단할 수평적인 정책조정기능을 수행할 기관과 부서에서 전담할 필요가 있음

[ 공공 빅데이터 수집 및 분석 시스템에서 컨트롤 타워의 역할 ]



## ◎ 핵심 원천기술 및 국가경쟁력의 확보가 시급

- ☐ 데이터 수집, 저장관리, 표준화, 분석 및 시각화에 이르는 공공 빅데이터 수집 및 분석 전과정에 필요한 특화 된 요소기술의 세계적 수준으로의 개발이 요구됨
  - 공공 빅데이터에 적합한 수집 및 처리속도 개선 및 데이터품질 관리 변경 등 기술개발이 필요함
  - 수집된 통합연계데이터에 대한 통계분석과 AI, 머신러닝을 활용한 최신의 분석/분류 기법을 융합적용할 수 있는 방법론 연계기술 개발도 시급
- ☐ 데이터산업의 IT 투자규모는 증가하고 있지만 투자항목에서 데이터 가치 고도화에 대한 투자는 이루어지고 있지 않아 투자활성화가 필요함
  - 데이터산업 IT투자규모는 데이터 솔루션이 3.7%증가 하였고, 데이터구축/컨설팅은 2.2% 증가하였으나 데이터서비스에 대한 투자는 -0.1%로 감소하였음
  - 데이터솔루션, 구축/컨설팅, 서비스 산업의 근간이 되는 데이터 가치 고도화에 대한 민간 및 정부의 활발한 투자가 시급함
- ☐ 데이터 시장의 전반적인 성장세에 따라 데이터 품질과 가치에 대한 수요도 증가될 것으로 보여 데이터 가치를 발굴하고 교정하는 기술개발이 시급함
  - 일반적으로 데이터 품질 진단과 수정에 대한 기술은 데이터 가치 고도화 기술에 비해 상대적으로 많은 발전을 보여왔음
  - 데이터품질과 달리 데이터 가치 고도화 기술은 방법론적 접근이 필요하므로 데이터 과학과 기초과학 분야에서 논의되는 데이터 보정 및 교정, 재생에 대한 개발 방법론이 필요함
- ☐ 데이터산업의 성장함에 따라 기존의 인력부족률을 해소할 수 있는 인력양성 정책과 계획이 필요함
  - 데이터 산업의 직무별 평균부족률은 8.7%이며, 가장 부족한 직무는 데이터 과학자(34.2%), 데이터 분석가(14.8%), 데이터 개발자(13.5%)의 순으로 나타남
  - 데이터 가치 고도화는 각 직무별 협업을 통해서 해결 가능하기 때문에 각 분야별 인력양성 과정에서 데이터 가치 고도화에 대한 이해와 교육이 병행되어야 할 것임
- ☐ 민간의 데이터 관련 기술 및 분석 인력과 함께 정책적인 데이터 거버넌스와 파트너십을 촉진할 수 있는 공공부문에서의 데이터 전문인력 육성 필요
  - 데이터 기반의 행정적 목적과 정책적 가치를 알고 빅데이터의 특성과 기술적 이해가 높은 공공부문에서의 전문가 필요
  - 개편된 공무원 직렬, 직류체계에서 데이터 직류 시험과목에 데이터베이스, 알고리즘, 인공지능을 포함한 데이터 거버넌스, 데이터표준과 분석 등 다양한 빅데이터 관점의 이론과 실무를 강화할 필요가 있음

## 참고문헌

1. [한국판 뉴딜 종합계획], 관계부처 합동, 2020.07.14.
2. [보도자료] 통합 데이터지도 확대개편, 과학기술정보통신부, 2021.02.03.
3. 디지털뉴딜의 핵심축인 데이터 댐 사업 성과보고회, 과학기술정보통신부, 2020.12.16.
4. 디지털 뉴딜의 핵심, [데이터 댐] 프로젝트 본격 착수. 과학기술정보통신부, 2020.09.02.
5. [보도자료] AI 학습용 데이터 구축 사업 과제 확정, 과학기술정보통신부, 2020.06.19.
6. [디지털 트랜스포메이션 트렌드] 코로나19와의 전쟁, 일등공신은 AI와 빅데이터, 모비인사이드, 2021.02.05.
7. 중소기업 전략기술로드맵(빅데이터), 중소벤처기업부, 2021
8. 포스트 코로나19 시대의 데이터 주권과 데이터 거버넌스, 정보통신정책연구원, 2020
9. 디지털 뉴딜, '데이터 댐'과 블록체인, 블록체인경영협회 9월 이슈리포트, 2020.09
10. [이슈리포트] 데이터 댐의 주요사업과 향후 추진방향, 정보통신정책연구원, 2020
11. 글로벌 데이터 안보 이니셔티브(全球数据安全倡议) 디지털거버넌스 심포지엄 발표, 2020
12. [입법·정책보고서] 데이터 경제 활성화를 위한 입법정책 방안, 국회입법조사처, 2020.12
13. [이슈분석] 데이터 정책, 다른 나라는 어떻게 펼쳤나, 전자신문, 2020.06.11.
14. 유럽연합 디지털 정책의 동향과 전망: “유럽의 디지털 미래” · “유럽 데이터 전략” · “인공지능 백서”의 주요 내용과 의의, 이상윤 고려대학교, 2020.06.11.
15. EU일반개인정보보호법 가이드북, 방송통신위원회, 2020
16. 유럽연합 「데이터 거버넌스 법안」의 주요 내용과 시사점, HIIC 이창범 동국대학교 교수, 2021.02.06.
17. 중국, 데이터안전법 개정, 세계법제정보센터, 2021.06.24.