

전략분야 현황분석

# 빅데이터





# 빅데이터

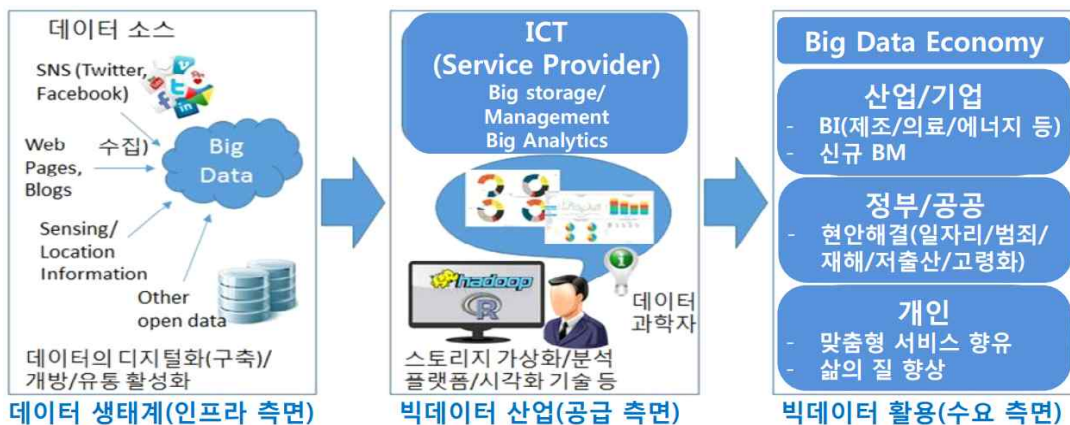
## 1. 개요

### 가. 일반적 정의

#### (1) 정의

- 빅데이터(Bigdata)란 형식과 규모가 다양하고 데이터 축적속도가 매우 빨라 기존의 데이터베이스로 처리할 수 있는 역량을 넘어서는 초대용량(테라바이트 또는 페타바이트급)의 정형, 비정형 데이터를 모두 포함하는 것을 말하며, 데이터의 생성, 수집, 저장, 관리 및 분석하여 가치를 추출하고 지능화 서비스의 기반을 지원하는 기술
- 빅데이터는 정형, 비정형을 포함한 다양한 종류의 대규모 데이터로부터 효율적으로 가치를 추출하고, 데이터의 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처로 이전에는 거대한 데이터 집합 자체만을 의미하였으나, 점차 그 범위가 확대되어, 도구, 플랫폼, 분석 기법 등을 포함한 포괄적 의미로 확대
  - 빅데이터에 대한 정의는 다양하게 설정되고 있으나, 일반적으로 양(Volume), 속도(Velocity), 다양성(Variety), 가치(Value)를 공통된 정의로 규정
- 빅데이터 분석 솔루션은 대용량의 정형 혹은 비정형 데이터로부터 숨겨진 패턴과 알려지지 않은 정보 간의 관계를 찾아내어, 비즈니스 의사결정을 지원할 수 있는 인사이트를 발굴하고 예측하는 소프트웨어 또는 하드웨어를 의미
  - 산업의 투명성 증대, 소비자 니즈 발견, 트렌드 예측, 성과향상, 고객 세분화, 의사결정 지원과 대행, 비즈니스 모델 및 상품/서비스의 혁신 등을 위하여 사용

#### [ 빅데이터 개념도 ]



\* 출처: “혁신성장동력 시행계획”, 관계부처 합동, 2018.05.28

## (2) 필요성

- 언택트 시대의 도래로 인한 빅데이터 수집 및 활용 급증 추세
  - 정부 및 다양한 기업들은 AI 등을 결합한 빅데이터 분석을 통해 코로나-19 방역에 기여
    - 코로나-19 확진자가 발생하는 경우 이동경로에 따른 모바일 이용 데이터, 신용카드 결제 데이터, CCTV 데이터 등을 분석하여 경로 및 접촉자 파악
  - 2020년 3월 기준으로 네트워크 트래픽은 전년도 대비 30%이상 증가하며 데이터가 빠르게 축적
    - 특히 전년도대비 SNS 버즈량은 15배 증가, 온라인 배송량은 약 3배 증가
  
- 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
  
- 전 세계적 디지털 트랜스포메이션 추세로 인해 데이터량이 급격히 증가하여 빅데이터 수집·분석 기술에 대한 니즈 대폭 증대 예상
  - 전 세계에서 연간 생성되는 데이터량은 2016년 96 EB<sup>1)</sup>에서 2021년 278 EB로 증가할 전망<sup>2)</sup>
  - 무선 네트워크의 고도화와 스마트폰의 보급 확대로 비정형 데이터 기반의 모바일 데이터는 2016년 7.2 EB에서 2021년 48.2 EB로 6.7배 폭증할 것으로 예상<sup>3)</sup>
  
- 민간·공공에서 축적되는 빅데이터의 활용으로 생산성을 향상시키고 비용을 대폭 절감
  - (민간) 정보서비스 부문의 빅데이터 DB 활용으로 1조 7,775억 원의 생산유발효과 및 1조 6,122억 원의 부가가치 창출 전망
  - (공공) 공공부문 빅데이터 활용으로 행정 효율성 제고, 세수증대 등을 통해 최대 4조 2천억 원(GDP의 0.4%)의 부가가치 창출 전망
  
- 빅데이터 산업은 오픈소스 중심의 소프트웨어 산업으로 중소기업에 적합한 산업인 동시에 비즈니스에 활용하려는 수요가 시장 성장을 견인하는 현 우리나라에 적합한 산업<sup>4)</sup>
  - 미국 기업은 78%의 회사가 오픈소스 기반으로 운영, SW가 필요할 때 66% 이상의 회사가 오픈소스가 있는지 먼저 확인하고, 39%의 회사가 오픈소스 프로젝트를 계획 중이며, 47%의 회사가 자사의 툴과 프로젝트를 오픈소스로 공개

1) Exabyte: 10<sup>18</sup> bytes

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

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

4) 오픈소스 중요성과 시사점(NIPA, 2019. 07)



## 나. 구축 범위

### (1) 가치사슬

- 빅데이터 산업은 데이터의 생산, 수집, 처리, 분석, 유통, 활용 등을 통해 가치를 창출하는 상품과 서비스를 생산·제공하는 산업으로, 데이터솔루션, 데이터구축, 데이터인프라, 분석서비스 등으로 구분 가능
- (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

### (2) 대표적 분류 방법

- 2018년 데이터 산업 현황 조사 보고서에 따르면, 데이터 산업은 데이터 솔루션, 데이터 구축 및 컨설팅, 데이터 서비스로 구분

#### [ 빅데이터 분류 ]

대분류	중분류	세부 기술
데이터 솔루션	데이터 수집	• 데이터 검색, 로그데이터 수집, 웹데이터 수집, 데이터통합/연계, 데이터 교환/개방
	DBMS	• RDBMS, NoSQL DBMS, 인메모리 DBMS, 기타 DBMS
	데이터 분석	• 정형/비정형 데이터 분석, 실시간 데이터 분석, 데이터 시각화 분석, 데이터 (전)처리
	데이터 관리	• 데이터 모델링, 마스터 데이터 관리, 데이터 품질 관리, DB 운영 및 성능 관리
	데이터 보안	• DB 보안, 개인 데이터 보안
	데이터 플랫폼	• 빅데이터 플랫폼 개발 및 공급
데이터 구축 및 컨설팅	데이터 구축	• DB 설계/구축, 데이터 이행, 데이터 구축/가공
	데이터 컨설팅	• 데이터 설계 컨설팅, 데이터 품질 컨설팅, DB 성능 개선 컨설팅, 데이터 거버넌스 컨설팅, 데이터 분석/활용 컨설팅
데이터 서비스	데이터 판매 및 중개	• 데이터 판매, 데이터 신디케이션
	정보 제공	• 포털/정보 매개 서비스, 정보 제공 서비스
	데이터 분석 제공	• 소셜 데이터 분석 정보 제공, 마케팅 데이터 분석 정보 제공, 리스크 데이터 분석 정보 제공, 기타 데이터 분석 정보 제공

\* 출처: 2018 데이터 산업 현황 조사, 2019, 한국데이터산업진흥원,

### (3) 기술로드맵 전략분야의 범위

- 빅데이터 산업 구분에 따라 본 전략분야에서는 데이터 구축 및 컨설팅은 기술의 영역에 보다 근접하므로 중소기업 전략기술 로드맵의 취지에 맞게 전략제품의 관점에서 데이터솔루션과 데이터 서비스 중심으로 분석을 진행할 예정

#### [ 빅데이터 전략분야의 범위 ]

\* : 본 전략분야의 범위

구분	상품 및 기술
데이터 솔루션	데이터 수집*
	DBMS
	데이터 분석*
	데이터 관리
	데이터 보안*
	데이터 플랫폼*
데이터 구축 및 컨설팅	데이터 구축
	데이터 컨설팅
	데이터 판매 및 중개
데이터 서비스	정보 제공
	데이터 분석 제공*

\* 출처: 2018 데이터 산업 현황 조사, 2019, 한국데이터산업진흥원, 네모아이씨지 재가공

## 2. 시장 분석

### 가. 세계 시장 분석

#### (1) 세계시장 동향 및 전망

##### ◎ 빅데이터 전체시장 전망

- Big Data Market이 2019년 발표한 세계 빅데이터 시장은 2018년 1,136억 달러 규모로 평가되었으며 연평균 성장률 10.6%로 2024년 2,075억 달러 규모로 성장이 전망됨
  - 기업들은 최소한의 인프라를 활용한 효율적인 빅데이터 시스템 구축을 위해 다양한 빅데이터 솔루션 및 기술을 도입하는 단계
  - IoT와 모바일 기기의 증가로 인한 데이터양 급증하여 이에 대응할 빅데이터 솔루션 수요 증가
  - 코로나-19의 여파로 인하여 비대면 산업 등 기존의 오프라인 활동을 온라인으로 대체하면서 데이터 생성량이 예상보다 빠르게 증가하는 추세

#### [ 빅데이터 세계 시장 규모 및 전망 ]

(단위: 십억 달러, %)

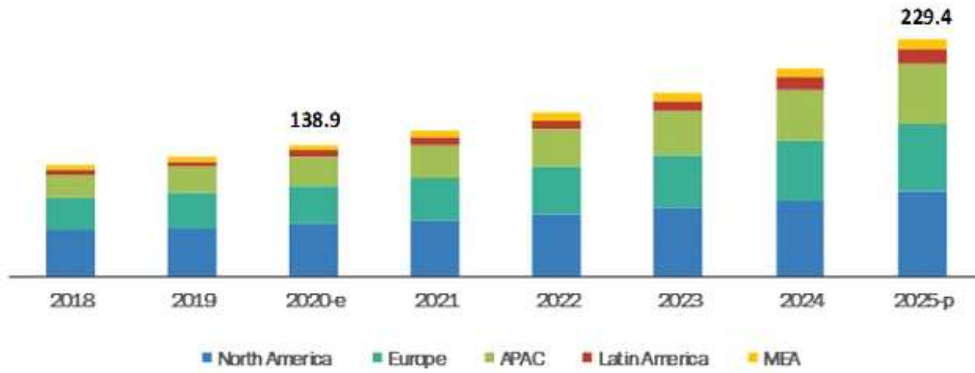
구 분	'18	'19	'20	'21	'22	'23	'24	CAGR
세계시장	113.6	125.6	138.9	153.6	169.8	187.7	207.5	10.6

\*출처: Big Data Market - Global Forecast to 2025, 2019, marketsandmarkets

- 미국, 중국, 일본 등의 선진국은 물론 아시아태평양 지역의 국가들까지 빅데이터와 인공지능 분야에 관심이 높으며, 특히 4차 산업혁명의 핵심으로 평가받고 있는 빅데이터 및 AI를 위해 빅데이터를 차세대 산업으로 선정하고 육성하기 위해 노력 중
  - 클라우드 기반 데이터 산업을 위주로 성장 중이며, 지역적으로는 아시아-태평양 시장의 가파른 성장세로 현재 북미 중심의 시장 규모를 추월할 가능성이 존재
  - 정부뿐만 아니라 민간기업 중에서도 Google, 애플, 텐센트, 알리바바 등의 세계적 기업의 경우 빅데이터, 인공지능 관련 핵심기술 확보를 위해 관련 투자 및 인수가 활발

[ 지역별 빅데이터 시장 성장 추이 및 예측 (2018-2025) ]

(단위: 십억 달러)



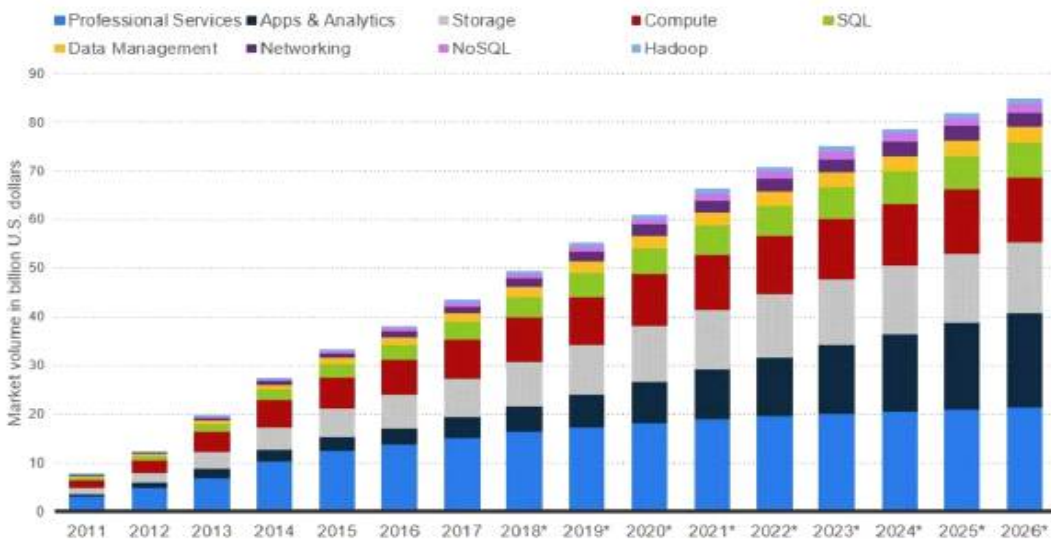
\* 출처: Big Data Market - Global Forecast to 2025, 2019, marketsandmarkets

◎ 분야별 시장 동향 및 전망

□ 빅데이터 시장은 크게 보안, 서비스, 데이터 엔지니어링으로 분류가 가능

[ 분야별 빅데이터 시장 성장 추이(2011-2026) ]

(단위: 십억 달러)



\* 출처: Wikibon and reported by Statista, 2018

- (빅데이터 보안 시장) 빅데이터 보안 시장은 2019년 149억 달러에서 2025년까지 382.3억 달러로 연평균 17% 성장 전망<sup>5)</sup>
  - 빅데이터 산업의 전체 발전 속도에 비해 보안 관련 산업은 더디게 개발되고 있었지만, 인텔 등 보안 관련 이슈가 발생하면서 기업들의 필요성 인지가 증가하는 추세
- (빅데이터 엔지니어링 시장) 데이터의 수집 및 가공을 포함하는 빅데이터 엔지니어링 시장은 2018년 344억 달러에서 연평균 17.6% 성장하며 2023년까지 773억 달러로 성장 전망
  - 데이터 시장 중 데이터 통합 시장은 '18년 130억 달러에서 '23년 약 275억 달러까지 성장 예상
  - 전 세계 빅데이터 및 데이터 엔지니어링 서비스(Big Data and Data Engineering Services) 시장은 2018년 약 344억 7,000만 달러에서 연평균 17.6% 성장하여 2023년 약 773억 7,000만 달러에 이를 것으로 전망

[ 전 세계 빅데이터 및 데이터 엔지니어링 서비스 시장 ]

(단위 : 백만 달러, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
데이터 통합	13,000	15,110	17,550	20,400	237.0	27,540	32,001	16.2
데이터 모델링	8,440	9,620	10,970	12,500	14,250	16,250	18,525	14.0
데이터 품질	7,620	9,050	10,750	12,780	15,180	18,030	21,420	18.8
분석	540	6,670	8,250	10,200	12,600	15,580	19,257	23.6
계	34,470	40,540	47,670	56,060	65,930	77,370	90,987	17.6

\* 출처: Big Data and Data Engineering Services Market - Global Forecast to 2023, 2018, MarketsandMarkets 연구개발특구진흥재단 제공

- (빅데이터 서비스 시장) 빅데이터 서비스(BDaaS, Big Data as a Service) 시장은 연평균성장률 30.5%로 '19년 113억불에서 '24년 427억 달러로 성장 예상
  - 데이터 누적의 가속화와 클라우드 기술의 발전으로 인한 빅데이터 관련 서비스의 수요 급증
  - 특히 아시아-태평양 지역의 성장률 급증하여 미래 빅데이터 서비스 시장 주도 예상
- 빅데이터 애플리케이션 및 분석 시장은 '18년 50억 달러에서 '26년 194억 달러로 크게 증가할 것으로 예측되며 연평균 15.49%의 성장률을 달성할 것으로 예상

5) Big Data Security Market - Growth, Trends, And Forecast(2020-2025),

## (2) 세계시장 핵심플레이어 동향

### ◎ 기업 인수합병 동향

- 글로벌 기업용 소프트웨어 업체들은 고객 경험에 관한 데이터 확보를 위해 대규모 인수합병(M&A)을 진행
  - (Google) 빅데이터 분석업체 '루커(Looker)'를 26억 달러에 인수하여 클라우드 컴퓨팅 사업을 강화
  - (Microsoft) 소셜 서비스 플랫폼인 틱톡을 인수를 추진하며 틱톡 가입자 8억 명에 대한 빅데이터를 확보를 노력
  - (SAP) '18년 11월, 미국 온라인 설문조사업체 퀵트릭스를 80억달러(약 9조원)에 인수
  - (Adobe) '18년 5월 기업용 전자상거래 솔루션업체인 마젠토를 16억 8000만 달러(약 1조 8000억원)에 인수
    - 마젠토는 코카콜라, 네슬레 등 세계 25만 개의 가맹점이 사용 중으로, 마젠토가 보유한 고객 경험 관련 데이터를 어도비의 클라우드 환경에 통합해 보다 체계적인 데이터 분석에 사용할 계획
  - (Salesforce) 고객관계관리(CRM) 업체로 '18년 3월 모바일 앱(응용프로그램) 데이터 통합 업체인 물소프트를 65억 달러(약 7조 원)에 인수하여 고객 경험을 보다 효과적으로 공유 및 연동한다는 계획 발표

### ◎ 분야별 주요기업 동향

- 클라우드 서비스 분야
  - (Amazon) AWS(Amazon Web Service)라는 클라우드 빅데이터 서비스를 제공하여 다양한 기업들의 광범위한 데이터베이스를 효율적으로 구축
  - (Oracle) '오라클 클라우드 데이터 사이언스 플랫폼'을 출시하여 기업 내 협력 기반의 머신러닝 모델 개발과 교육, 관리 및 구축을 통해 데이터 사이언스의 성공적 수행에 기여
    - '오라클 클라우드 데이터 사이언스 플랫폼'은 AutoML 자동화 알고리즘 선택 및 튜닝, 자동화된 예측형 특징(predictive feature) 선택, 모델 평가, 모델 설명 등의 기능 제공
- 유통·물류 분야
  - (Amazon) 거대 유통망과 데이터를 이용하여 제약·의료 시장까지 진입
    - 2018년 6월에 온라인 제약 스타트업 필팩(PillPack) 인수
    - 아마존은 빅데이터를 활용하여 개인 조건에 따른 선호 상품 추천 및 가격을 최적화
    - 아마존고(AmazonGo)는 계산대 없는 매장으로 모든 방문객들의 이동경로 및 실제 구매 제품 등 막대한 빅데이터 수집을 통해 유의미한 정보를 추출
  - (Netflix) 엔터테인먼트 기업 중에서 빅데이터를 가장 적극적으로 활용
    - 넷플릭스는 빅데이터 모니터링 및 분석을 활용하여 소비자들을 위한 콘텐츠 추천 알고리즘과 콘텐츠 제작 기획

- (Softbank) 비전펀드를 통해 미국 온라인 포워딩 업체인 Flexport에 10억 달러를 투자
  - Flexport의 사업 분야는 크게 3가지로, 출발지와 종착지까지의 end to end 서비스를 통해 실시간으로 물류를 확인할 수 있는 ‘해상’, ‘항공 서비스’와 수입업자, 수출업자, 해상 운송업자, 항공사, 창고, 관세 중개업자 등을 한 플랫폼에 모아 글로벌 무역을 수행하는 데 필요한 모든 데이터를 구조화함으로써 공급 체인 비용을 절감하는 ‘고객 중개’가 있음
  - Flexport는 자체 솔루션 시스템을 개발하여 고객들이 효율적이고 최적의 의사결정을 내릴 수 있도록 서비스를 제공

#### □ 의료 분야

- (Enlitic) 각종 의학 영상 자료를 토대로 환자의 암세포를 식별하고 5년 이내 완치율, 사망률 등을 예측
  - 암진단 실패율은 0%, 부정확한 진료 비율은 47%로 방사선전문의보다 정확한 결과값 도출
- (Microsoft) 미국의 대형 약국 체인인 ‘월그린 부츠 얼라이언스(Walgreens Boots Alliance)’와 헬스케어 관련 시스템 개발을 위한 전략적 제휴를 체결했다고 발표
  - 양사는 MS의 클라우드 서비스인 ‘애저(Azure)’와 MS의 AI(인공지능) 플랫폼을 활용하여 의약품 판매에 관련된 새로운 유통 모델을 개발한다는 계획
  - 2020 NRF에서 미국 최대 슈퍼마켓 체인인 월마트(Walmart), 2위인 크로거(Kroger) 등 마이크로소프트와 협업하고 있는 사례가 자세히 소개되었고, 크로거의 경우 마이크로소프트 클라우드 플랫폼인 애저(Azure) 기반의 RaaS(Retail as a Service)를 도입해 새로운 비즈니스 수익 모델을 창출<sup>6)</sup>

#### □ 금융 분야

- (JP Morgan Chase) 빅데이터를 활용하여 직원의 비리 혐의의 포착 및 은행 부동산 담보의 최적 가격 측정에 활용
- (Progressive) 자동차에 부착된 기기가 전송하는 데이터를 바탕으로 고객의 운전 패턴을 분석하고 미래 사고 가능성을 예측
- (Visa) 빅데이터를 활용하여 구입품목, 시점, 결제 위치 등을 실시간으로 파악하며 고객의 구매 성향을 토대로 할인쿠폰을 발송해주는 RTM(Real Time Messaging) 서비스를 제공
- (AMEX) 위치 기반 소셜네트워크 정보를 활용한 고객별 맞춤형 마케팅을 실시하며 AMEX Sync 프로그램을 활용
  - AMEX Sync 프로그램은 제휴를 맺은 소셜 플랫폼의 고객 계정을 AMEX카드와 연동시켜 고객에게 맞춤형 할인 혜택 제공
- (Lenddo) 신용평가 알고리즘을 개발할 때 온라인상 대출자의 평판에 대한 비정형 데이터를 추출 및 활용하여 신용도 평가
  - 개인이 아닌 소상공인의 경우 기업의 평판 및 영업 활성화 정도를 통해 대출 여부 판단

6) MS, 'NRF 2020' 참석...클라우드-AI 활용 디지털 트랜스포메이션 강조(뉴데일리경제, 2020.01.14)

## 나. 국내 시장 분석

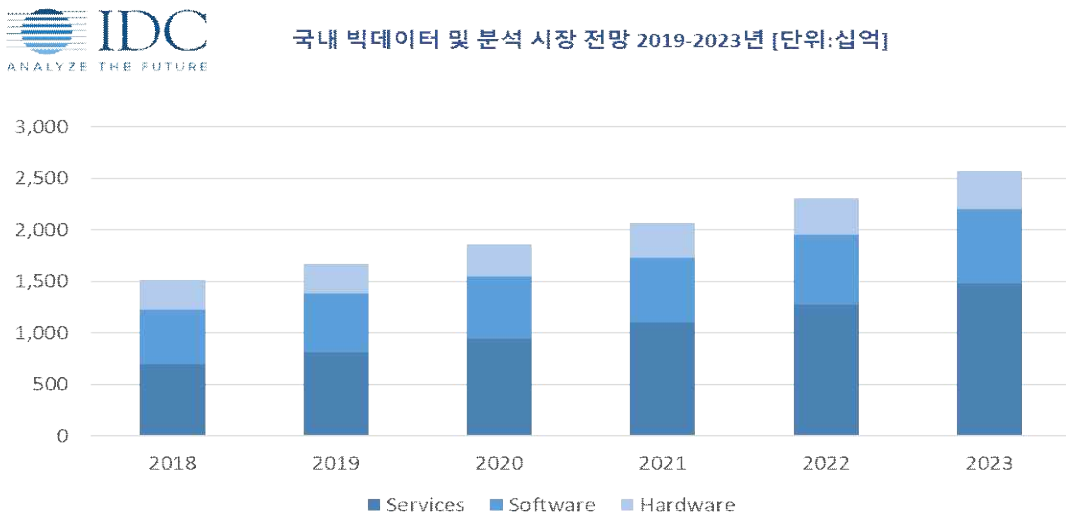
### (1) 국내 시장 동향 및 전망

#### ◎ 전체 국내 시장 전망

- 2019년 국내 빅데이터 및 분석 시장은 2018년 대비 10.9% 증가한 1조 6,744억 원으로 2024년까지 연평균 11.2% 성장하여 2조 8,569억 원 규모로 성장 전망
  - 기업 및 공공기관 등 모든 산업에서 비즈니스 혁신 및 인사이트 도출을 위한 데이터 활용의 중요성에 대한 관심 증대
  - 기업들은 디지털 트랜스포메이션을 위한 노력을 지속적으로 기술이고 다양한 형태의 클라우드 환경 전환을 활발히 고려
  - 빅데이터 기반의 고급 분석 및 인공지능 시스템 구축을 위한 데이터의 필요성 증가 요인에 의해 관련 시장은 향후 지속적 성장 예측

[ 국내 빅데이터 시장규모 추이(2018-2023) ]

(단위: 십억 원)



Source: IDC, 2019

\* 출처: 국내 빅데이터 및 분석 2019-2023 전망, 2019, 한국IDC

[ 빅데이터 국내 시장 규모 및 전망 ]

(단위: 억 원, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
국내시장	15,044	16,744	18,636	20,741	23,084	25,692	28,569	11.2

\* 출처: 국내 빅데이터 및 분석 2018-2024 전망, 2019, 한국IDC, 네모아이씨지 재가공

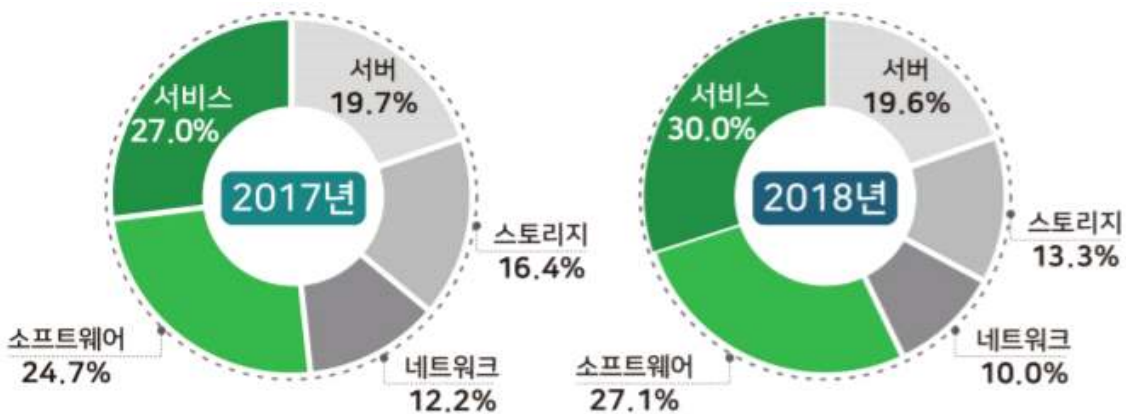


◎ 세부 시장 전망

- 부문별 빅데이터 투자 규모는 '17년과 '18년 모두 서비스 부문 1위, 소프트웨어 부문 2위를 차지
  - '17년과 '18년을 비교하면, 서비스와 소프트웨어의 비중은 증가했지만, 하드웨어 비중은 감소 (서비스 3.0%p, 소프트웨어 2.4%p 증가 / 하드웨어 5.4%p 감소)
  - 이는 빅데이터 관련 투자가 과거 하드웨어 중심으로 이루어졌던 것에서 빅데이터를 활용한 서비스, 이를 가동하기 위한 소프트웨어에 대한 투자로 변화

[ 국내 빅데이터 시장규모 추이(2017-2022) ]

(단위: %)



\* 출처: 2018년 데이터산업 현황조사, 2019, 과학기술정보통신부

- 국내 빅데이터 시장은 20% 이상의 고성장세를 보일 것으로 전망

[ 시장영역별 국내 빅데이터 시장 현황(2016-2018) ]

(단위: 억 원, %)

구분		2016년	2017년	2018년	CAGR (2016-2018)
공공시장	시장규모	999	1,338	2,014	42.0
	성장률	43.1	33.9	50.5	
민간시장	시장규모	2,441	3,209	3,829	25.2
	성장률	26.8	31.5	19.3	
합계	시장규모	3,440	4,547	5,843	30.3
	성장률	31.1	32.2	28.5	

\* 출처: 2018년 데이터산업 현황조사, 2019, 과학기술정보통신부

## (2) 국내 생태계 현황

### ◎ 유통·물류 분야

- 수송 규모의 확대와 소량·다품종 처리 시설 도입 등에 따라 물류 창고의 대형화 및 고도화
- 집하부터 배송까지 다양한 이해관계자들이 얽혀 있으며 톱니바퀴처럼 일사불란하게 움직이는 물류는 다른 산업에 비해 많은 트랜잭션이 발생
  - 인프라를 가지고 있는 사업자를 연결하는 포워딩(국제물류 주선사업)의 중요성 증대
- 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
  - 소프트뱅크는 디지털 포워딩 스타트업 Flexport에 투자
- 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 수립
  - 와이파이를 이용한 방문자 위치정보를 수집하여 유동인구를 상세히 파악

### ◎ 제조 분야

- 빅데이터를 활용한 스마트팩토리 설계·생산 과정에 자동화와 정보시스템 등을 도입하였으며, 제품의 품질과 공정의 효율성을 높이기 위하여 데이터를 분석하고 그 결과를 활용하는 스마트 팩토리 도입 사례 증가
  - 캡제미니의 보고서에 따르면 스마트 팩토리를 도입한 자동차 제조사는 향상된 물류, 재료비용, 장비 효율화, 생산 품질 개선 등을 통해 운영 마진을 평균적으로 36% 향상
- ERP SCM 등 어플리케이션 시장이 성숙기에 접어들었으나 빅데이터 플랫폼을 활용한 수평적, 수직적 통합이 이루어지면서 다시 성장의 계기가 마련
- 국내 주요 대기업들은 그룹 내 IT 계열사(LG CNS, 삼성 SDS, SK C&C, 포스코 ICT)를 통해서 어플리케이션 영역을 구축하고 클라우드나 빅데이터 플랫폼에 초점을 두고 진행 중
- 중소기업은 주로 SI 프로젝트의 형태로 중소기업 대상 비즈니스를 영위하거나 기업별 장점이 있는 분야의 니치 마켓에 주력하는 중

- 데이터를 수집, 저장과 분석 및 처리하고 그 결과를 시각화하여 눈으로 보는 관리(V/M)를 가능하게 함으로서 문제가 발생하기 전에 이를 인지하고 사전에 조치 가능
  - 불량 발생 환경 분석, 사전 조치하여 불량률 낮추고 결함 발생 최소화하며 사전 보수로 유지 및 수리 비용을 절감하며 적절한 피드백이 가능하고 자동으로 작업지시 및 최적의 투입량을 결정하여 에너지 절약이 가능
- 빅데이터를 분석 및 해석할 수 있는 컴퓨터 프로그래밍, 데이터베이스를 처리할 수 있는 자연계 역량과 인문학적 소양을 폭넓게 갖춘 융합형 인재의 육성 필요
  - 데이터 지식뿐만 아니라 산업 전반에 걸친 전문지식이 필요하므로 빅데이터 전문가와 산업 종사자 간 네트워크가 필요

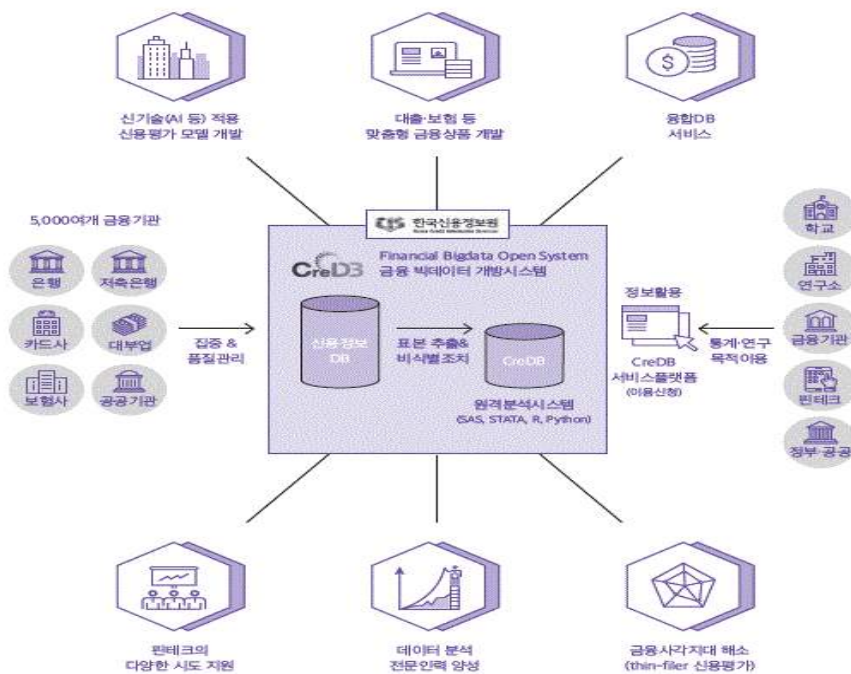
### ◎ 의료 분야

- 스마트폰, 스마트워치 등의 보급과 함께 센서기술이 고도화되어 의료산업의 디지털화의 필요성 증가
  - 건강정보, 의료 기초 정보를 보다 정교하게 일상 속에서 수집 가능
  - VR, AR, 5G, 로봇 기술의 진화로 의사의 의료서비스를 수행하는 관점에서 원격 의료 시술과 정밀 수술이 발전하였으며 신약 개발과 임상 실험을 하는 의학 기술의 진화에도 큰 기여
  - 단편적인 데이터의 수집이 아닌 데이터 시각화를 통해 데이터의 흐름을 보다 쉽게 파악하여 환자의 개인 이력 관리, 파악에 용이
- 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

◎ 금융 분야

- 금융 분야는 영업이 전산시스템에서 이뤄지기 때문에 데이터 집적량이 많고 정확도도 높은 편
  - 추진 중인 ‘정보 개방’ 흐름은 불합리한 데이터 격차를 해소에 금융산업 발전을 도모하는 방향성
- 마케팅, 상품개발, 리스크관리, 신용평가 등 금융 전반에 빅데이터를 활용
  - 오픈 API를 통한 혁신적인 대외 금융서비스의 구현
  - IT인프라 운영 전략, 사이버 보안 전략
- ‘통합 빅데이터 플랫폼’ 구축을 통한 마케팅 역량의 강화
  - 폭증하는 데이터 처리를 위한 다양한 형태의 클라우드 구현과 새로운 데이터 거버넌스의 정립
  - 인공지능(AI)기반의 레그테크를 통한 신속한 규제 대응
- 신용정보원은 ‘금융 빅데이터 개방시스템(CreDB)’을 구축해 핀테크 업계, 학계, 일반 기업 등에 순차적으로 개방
  - 기존에 대규모 고객을 가지고 있지 못하던 중소형 금융사, 핀테크사, 연구 기관에 새로운 상품 개발, 시장 분석, 연구 등에 나설 수 있는 데이터 기반을 마련

[ 금융권 빅데이터 활성화 및 생태계 조성 ]



\* 출처: 2019 데이터산업백서, 2019, 한국데이터산업진흥원

- 금융위원회는 2019년 6월부터 가동한 ‘금융 분야 빅데이터 인프라’ 구축 방안 발표
  - 금융 빅데이터 개방시스템의 구축, 금융데이터 거래소의 운영, 데이터 전문기관 출범

## ◎ 농축산 분야

- 반도체 산업과 같이 지속가능한 디지털 농업 시대의 발전을 위해서는 다양한 농업 데이터를 중심으로 하는 스마트팜 시장에 대한 투자 중요성 증대
- 스마트팜 데이터는 작물을 재배하는 시설(온실), 시설 내·외부의 환경 및 제어 정보, 작물정보, 농작업, 재배에 필요한 요구자원 등의 복합 생산 데이터, ‘생산-유통-가공-소비’에 이르는 전주기 데이터 등으로 구성
  - 노지작물의 농작물 상태 파악, 병충해 관리, 가뭄관리 등을 위해 드론이나 위성을 활용하고, 센서와 영상 카메라 등을 활용해 특정 유형 작물 생장의 모니터링 및 분석, 로봇 시스템 제어, 운영을 하기 위한 지능형 시스템 개발이 진행

## ◎ 자동차 분야

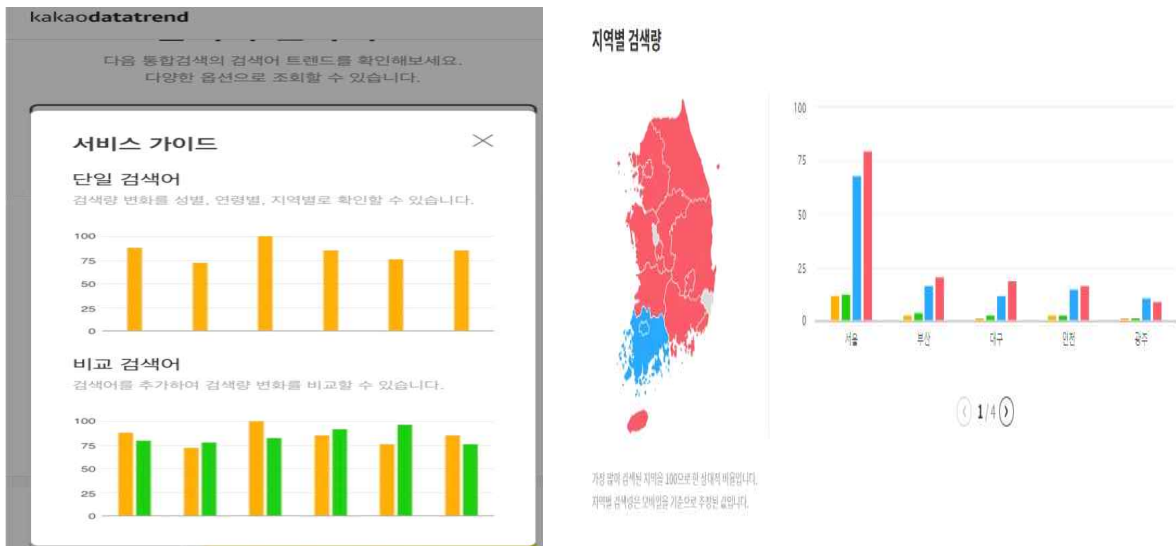
- 전 세계적으로 자율주행 자동차 시장이 지속해서 성장하는 가운데 관련 서비스 산업과 자율주행 자동차를 운행하며 사용되는 데이터의 양과 종류 역시 기하급수적으로 증가
  - 도로의 상태, 장애물, 날씨 등 주행환경 정보와 자율주행 자동차, 도로, 이용자의 정보를 수집하고, 수집된 정보를 융·복합하여 인공지능 기반으로 분석하고 기존 교통 체계 또는 자율주행서비스에 활용하는 기술 개발이 필요
- 자율주행 빅데이터 기술은 크게 대용량 데이터 수집, 인공지능 기반 분석,서비스 적용을 위한 기술 세 가지로 구분
  - (데이터 수집 및 저장 기술) 자율주행 자동차와 인프라의 대용량 빅데이터 수집·저장·관리 기술 및 교통 인프라 데이터의 클라우드 소싱 기술 등 개발 진행 중
  - (인공지능 기반 분석 기술) 자율주행 자동차 데이터와 인프라 데이터의 융합 모델링 기술 및 인공지능 기반 자율주행 빅데이터 분석기술 등 개발 진행 중
  - (빅데이터 기반 자율주행 서비스 응용 기술) 빅데이터와 클라우드 기반의 자율주행 자동차 이동성 최적화 기술 및 자율주행 자동차의 주행환경 인지 범위 확대를 위한 클라우드 기반의 실시간 정보공유 기술 등의 개발 진행 중

### (3) 생태계 핵심플레이어 동향

#### ◎ 빅데이터 분석·처리

- 빅데이터 분석·처리 분야에서 텍스트, 이미지, 음성 등 비정형 데이터 등 다양한 빅데이터의 융합분석과 예측을 통한 플랫폼 개발 중
  - (네이버) ‘네이버 데이터 랩’을 통해서 검색어 추이, 업종별 검색 사용자 수, 아파트 실거래 지표 등을 활용하여 사용자들에게 정보 제공
  - (카카오) 2020년 7월 ‘카카오 데이터 트렌드’를 선보여 다음(Daum)의 통합 검색어 정보를 기반으로 각 검색어의 데이터를 서로 비교해볼 수 있으며 조회한 내용을 파일로 내려 받아 직접 데이터를 가공해 연구나 리서치 등에 활용 가능

[ 카카오 ‘데이터 트렌드’ ]



\* 출처: 카카오데이터트렌드(kakaodatrend)

- (KT넥스알) 계열사인 KT넥스알을 통해 대용량 데이터를 수집 저장하고 컨테이너 클라우드 환경을 통해 인프라 관리, 데이터 처리, 고급분석, 시각화 및 데이터 마켓 플레이스 기능을 갖춘 빅데이터 분석 플랫폼을 개발 중
- (와이즈넷) 자연어처리, 머신러닝, 텍스트 마이닝을 요소 기술을 주력으로 대용량 빅데이터의 수집, 검색을 비롯해 분석을 위한 마이닝, 인공지능 기반 챗봇 등 인공지능 솔루션을 개발 중
- (솔트룩스) BigO는 빅데이터 분석플랫폼으로 빅데이터의 수집, 변환, 분석, 시각화를 통해 의사결정을 지원하는 플랫폼으로 자연어 처리기와 기계학습, 분산병렬처리 기술을 기반으로 동작

## ◎ 금융 및 보안

### □ 인터넷은행의 신용평가모델 기반으로 활용되는 빅데이터

- (카카오뱅크) 카카오뱅크의 신용평가모델은 주주사인 카카오의 정보를 활용하여 쇼핑, 택시 등의 온라인 활용 데이터, 고객 선호도, SNS 활동내역 등을 결합한 카카오 스코어를 산출하여 신용평가모델에 적용
- (케이뱅크) 주주사인 KT의 통신 데이터와 BC카드의 고객 데이터 및 가맹점 데이터를 활용하여 신용평가모델을 개발한 결과, 더 넓은 층의 고객을 확보했을 뿐만 아니라 시중 은행보다 낮은 연체율을 기록
- (핀크) 통신서비스 이용 정보를 활용해 ‘통신등급’이라는 새로운 개념을 만들고, 이를 대출 심사에 활용하고 대출 판매까지 중개하는 ‘통신료 납부정보 활용 용평가 서비스’를 제3차 혁신금융서비스로 승인받아 추진 중

### □ 기존 금융사의 빅데이터 활용

- (우리은행) 국내 최초로 빅데이터를 활용한 기업진단시스템 ‘빅아이’를 도입함으로써 비정형 데이터를 여신심사 시 활용할 수 있을 것으로 기대
- (삼성화재) 보험사기탐지시스템(IFDS, Insurance Fraud Detection system)을 개발하여 보험계약과 개인 보험정보 등 방대한 데이터를 활용하여 보험사기 고위험군을 추출 후, 조사 착수 시스템을 구축함. 이를 통해 허위 신고를 통해 보험금을 받으려는 사기 건을 사전 적발이 가능
- (현대카드) 현대카드의 데이터 동맹 ‘도메인 갤러리’를 활용하여 마케팅 효과가 높을것으로 예상되는 고객을 선별해 쿠폰 등 프로모션을 발송하는 데이터 플랫폼인 ‘트루 노스(true north)를 공개
- (신한카드) 빅데이터 분석 플랫폼 신한카드샵(#)을 통해 중소가맹점의 매장 운영과 홍보를 도와주는 서비스를 개발
- 이외에도 NH농협은행의 빅데이터 기반 위험 관리 시스템 구축, 신한은행의 ‘소매 고객 주거래지수 모형’ 개발, 국민은행의 데이터 분석부 신설 등 은행권에서 빅데이터 활용을 위한 사업이 전개되고 있음

### □ 금융·보안 분야에서는 이상금융거래탐지, 금융사기, 각종관제시스템 등 보안 문제를 고객정보 등 빅데이터와 머신러닝 기법 등을 활용한 솔루션 개발이 주도 중

- (삼성화재) 빅데이터로 보험사기를 적발하는 ‘보험사기 방지시스템(IFDS)’을 개발하여 축적된 보험계약과 사고, 보험금 지급정보 등의 정보를 바탕으로 보험사기 위험도 점수를 실시간으로 산출
- (지티원) 금융과 보안이 결합된 솔루션을 보유하고 있는 기업인 지티원은 통합 자금세탁방지 솔루션을 통해 주요 거래고객, 거래내역 등 금융거래 관련 데이터를 활용하여 부정 사용 및 혐의 거래 모니터링으로 자금세탁행위를 사전에 차단하기 위한 기술을 선도 중
- (노매드커넥션) 보안 분야를 선도하는 기업으로 블록체인 기술을 바탕으로 암호화폐지갑을 위한 통합경제솔루션, 상품의 생산, 가공, 유통에 이르는 공급망관리(SCM) 솔루션, 전자문서 인증 및 보안관리 솔루션, 신뢰성과 투명성을 강화한 디지털투표 솔루션을 선도 중

### ◎ 의료 분야

- 의료 분야에서는 빅데이터 기술을 도입하면서 비정형 데이터의 분석결과까지 포함하여 의료기술, 신약개발, 임상실험, 판매/마케팅 및 전 영역에 적용할 수 있게 됨
  - (삼성서울병원) 유전적, 의학적 현상, 평상시 생활 등 방대한 개인의 의료데이터를 기반으로 개인별 맞춤 치료서비스를 제공
    - 빅데이터 기술을 활용하여 의료데이터를 분석하고 개인에 맞는 건강관리 및 치료방법을 제안하고 더 나아가 질병 또한 예측하는 서비스로, 치료중심이 아닌 건강관리 및 예방에 집중
  - (동아ST) 아주대 병원과 함께 복합제 개발과정에서 빅데이터 기술을 활용
    - 처방전 및 약물의 부작용 사례를 통해 약 개발 시 비용과 리스크 감소 가능
    - 개발 시 무작위로 여러 약을 혼합해서 실험하는 것보다 빅데이터 기술을 활용한 정확한 수요 및 의료기술을 포함한다면 비용과 리스크 감소 가능
  - (휴먼스케이프) 개인 건강데이터를 신약개발과 임상시험 등에 활용하는 블록체인 커뮤니티 앱 ‘모아(MOAAH)’를 운영하여 의료 빅데이터 수집
  - (쓰리빌리언) 방대한 의료 빅데이터를 기반으로 AI 기술과 융합해 약 7800여 종의 희귀질환을 진단해주는 솔루션을 개발하여 기존에 5년 이상 걸리던 진단 기간을 20~40시간으로 단축 및 진단비용을 10분의 1 수준으로 절감

### ◎ 미디어 분야

- 미디어 분야에서는 다양한 엔터테인먼트 콘텐츠를 빅데이터 활용하여 소비자들에게 제공하는 플랫폼 위주
  - (왓차) 알고리즘으로 고객 데이터를 분석해 영화를 추천해주고 직접 볼 수도 있는 온라인 동영상 서비스(Over the Top,OTT) ‘왓차플레이’ 출시하여 약 570만 명의 가입자 확보
  - (데이블) 데독자, 성별, 연령, 행동패턴 등을 자동분석해 독자 개인별 맞춤콘텐츠를 추천하는 ‘개인화 뉴스 서비스(데이블 뉴스)’를 제공하며 아시아 주요 미디어 1800개사와 제휴

### ◎ 교육 분야

- 교육 분야에서는 빅데이터를 활용한 교육 서비스 플랫폼 다수
  - (뤼이드) ‘산타토익’이라는 AI 튜터 어플리케이션 출시하여 소비자들에게 빅데이터를 활용한 개인맞춤형 효율적 교육 로드맵 및 콘텐츠 제공
  - (매스프레소) ‘칸다’라는 딥러닝 교육 플랫폼을 개발하여 수학문제를 스마트폰으로 촬영하면 600만 건의 문제풀이 데이터를 통해 신속히 정답을 제공



### 3. 기술 분석

#### 가. 해외 기술 동향

##### (1) 주요 기술 동향

- (증강 데이터 관리) 머신러닝과 AI를 활용해 기업의 데이터 관리 카테고리를 생성하며 데이터 품질, 메타데이터 관리, 마스터데이터 관리, 데이터 통합 등이 포함
  - 2022년 말에 이르면 데이터 관리 수작업이 45% 가량 줄어들 것으로 전망
- (빅데이터 패브릭) 분산 데이터 환경에서 마찰없는 액세스와 데이터 공유를 가능하게 하며 사일로화 된 저장소를 탈피해 일관된 단일 데이터 관리 프레임워크 구축
  - 데이터 거버넌스, 데이터 웨어하우스/데이터 레이크, 데이터 가상화로 구성
  - 이기종 데이터를 하나의 물리적 저장소에 통합하지 않고 가상 통합
  - 빅데이터를 통해 비즈니스 가치창출을 위해 데이터 거버넌스가 반드시 필요
- (블록체인) 블록체인을 활용한 데이터 분석 기술 개발 추진
  - 블록체인은 원장 데이터베이스 관리 시스템(DBMS)을 통해 단일 기업의 데이터 소스에 대한 감사를 가능하게 하는 옵션을 제공하며 빅데이터 기술과 융합 기대
  - 가트너에 따르면 데이터 분석은 DBMS 인프라와 블록체인 기술 간의 기능 미스매치를 강조하여, 블록체인을 기존 DBMS 인프라의 부가 기술로 포지셔닝해야 한다는 주장
- (IoT 시스템) 모든 사물이 연결되는 초연결시대와 더불어 DT(Data Technology) 시대에서 IoT 기술이 데이터 수집에 핵심적인 역할을 할 것으로 전망
  - IBM, Google 등은 사물인터넷을 통한 솔루션을 제공하는 서비스 개발이 활발하며, 시스코, Sigfox Intel 등은 IoT의 인프라 측면에서 활발한 개발이 이루어지는 중
  - 국내에서도 IoT 플랫폼이 주요 서비스 사업자(SKT) 및 주요 연구기관(KETI, ETRI)에 의해서 개발되고 구축되고 있으며, 향후 IoT 센서들이 광범위하게 설치되어 데이터가 수집되면, 빅데이터 수요가 늘어날 전망
- (빅데이터 지식처리 플랫폼) 딥러닝 및 인공지능과 결합된 형태의 빅데이터 분석 플랫폼이 등장하면서 심층 질의응답이 가능하고 고속병렬처리 방식으로 연산소요시간을 대폭 단축시킨 차세대 지능형 지식처리 플랫폼으로의 진화가 진행 중
  - Google: 텐서플로(TensorFlow) 및 Knowledge Graph, IBM : Watson, Apple : Siri
  - 데이터의 급증에 따라 수집되는 데이터의 기계학습(Machine Learning)의 수요도 함께 증가하고 있으며, 이미지 분석을 필두로 딥러닝 기술이 발전하여 향후에는 IoT에 의해서 생성되는 시계열 예측 데이터에 대한 스트림 기계학습 등장이 예상

- (분석/시각화 상용 솔루션 개발) 하둡(Hadoop)에 의해 주도되었던 빅데이터 플랫폼에서, 기능과 성능을 개선하기 위한 클러스터 컴퓨팅 프레임워크방식의 스파크(Spark)로 대변되는 고속 메모리기반 분석 등 분석/시각화를 강조한 상용 솔루션들이 부각되는 중
  - 고속 분석을 보장하는 Apache Spark가 급격하게 자리 잡고 있으며, NFLab에서는 Spark용 분석 시각화 도구인 Zeppelin을 오픈소스로 공개하여 세계적으로 주목을 받고 있음
  - 하둡 기반 빅데이터 기업: 클라우데라(Cloudera), 호튼웍스(Hortonworks), 맵알(MapR)
  - 분산/시각화 솔루션: 스팟파이어(Spotfire), 타블로(Tableau)
  
- (클라우드형 빅데이터 플랫폼 서비스) IoT의 증가로 특히 시계열 빅데이터 처리 전용 솔루션들이 증가하고 있고, 특히 클라우드 형태로 빅데이터 플랫폼을 제공하는 서비스가 증가하는 추세
  - 시계열 빅데이터 처리: Par-Stream DB(상용솔루션), Influx-DB(오픈소스)
  - 클라우드형 빅데이터 플랫폼 서비스: Google, Amazon AWS, Microsoft Azure, Oracle 등 거의 모든 글로벌 회사가 클라우드형 빅데이터 서비스 제공
  
- (음성언어 처리기술) 스마트폰 중심으로 한 모바일 인터넷 환경에서의 많은 사람들이 음성인식을 사용함으로써 엄청난 규모의 사용자 로그 데이터를 확보하고 있는 상황
  - 음성언어 처리기술은 크게 불특정 다수화자의 다양한 발음 특성을 표현하는 것을 목적으로 통계적 방식으로 대표 패턴을 생성하는 빅데이터 기반의 음향모델과 n개 단어열에 대한 출현빈도를 확률값으로 나타내는 n-gram 기법을 활용한 음성데이터의 문법적 구조를 모델링 하는 것으로 구분
  - Google: Google은 현재의 음성인식 서비스를 확장하여 실제로 Google은 음성검색 서비스를 통해 하루 동안 한 사람이 2년 동안 쉬지 않고 얘기하는 양의 방대한 양의 음성데이터를 수집
    - 수집된 음성데이터는 텍스트 자료로 전사가 가능하도록 하여 100억 개 이상의 문법 구조를 학습하고 음성인식 성능을 개선하는데 활용
  
- 기술 인력 현황 : 글로벌 IT 기업은 빅데이터 기술 개발·확보 추진
  - (기술개발) 구글, 아마존 등 빅데이터 선도기업은 하둡(Hadoop) 등 빅데이터처리 기술을 개발하고 빅데이터 서비스 플랫폼 확보
    - 구글은 클라우드 상에서 빅데이터를 분석하는 ‘빅쿼리 서비스’를 제공하여 별도인프라 구축없이 빅데이터를 분석할 수 있도록 지원
  - (기술확보) IBM, HP 등 대형 IT 기업은 빅데이터 솔루션 확보를 위하여 빅데이터 전문업체를 인수하고 기존 제품과 결합
    - BM은 데이터 웨어하우징 및 분석기업인 네티자(Netezza) 인수(‘10.9), HP는 실시간 분석 플랫폼 기업 버티카(Vertica) 인수(‘11.3) 등
  - (인력) 미국은 ’18년에 44만명 이상 분석전문가가 필요하지만 관련 전공자 및 경력자는 30만명으로, 14만명 이상의 인력 부족 예상

- (영상 빅데이터 분석기술) 기존 영상 자체에 대한 인식의 범위를 뛰어넘는 의미 있는 정보추출과 내용분석을 통해 새로운 가치를 창출하며 미래 변화를 예측하는 좋은 기회로 작용할 것으로 예상
  - DARPA 주도의 VIRAT(Video and Image Retrieval and Analysis Tool)프로젝트가 대표적이며 labeled data를 가지고 특징을 추출하여 영향을 분류하는 학습으로 영상에 대한 인식/분류하는 감독학습기술과 Deep learning 기술로 객체를 인식/분류하는 비감독(unsupervised) 학습기술이 이슈
  - 미국: DARPA에서는 대규모 영상 DB에서 content-based searching 기능을 통해 빠른 속도로 대용량 영상콘텐츠를 검색하는 기능을 구현, 움직임의 형태와 속성에 대하여 검색 도구를 제공
  - 일본: 히타치 연구소에서는 영상 빅데이터 기술과 관련하여 이미지가 포함되어있는 유사한 장면을 자동으로 검출하고 해당 장면으로 이동, 특정객체 검출, 추출 영상에 대해 주석(annotation)을 자동으로 태깅하는 3가지 기술요소를 제시
    - 유사이미지검색 기술: 쿼리 이미지에 비슷한 이미지를 데이터베이스에서 찾아오는 기술로 대량의 영상 데이터 중에서 원하는 이미지를 추출할 수 있고, 특징량 벡터 클러스터링 기반의 고속유사 벡터 검색기법을 사용하여, 유사이미지 검색기능을 제공
    - 객체탐지 기술: 이미지 중에서 사람의 얼굴이나 자동차 등의 객체 영역을 식별하는 기술로 이미지의 부분 영역과 검출 대상의 사례 이미지를 유사이미지 검색의 특징량 기준으로 일치하는 객체 영역을 검출
    - 이미지 주석 기술: 이미지가 나타내는 내용에 해당하는 메타 데이터를 자동으로 부여하는 기술로 주어진 이미지 쿼리에 대하여 유사이미지 검색을 하고 검색결과에 이미지에 나오는 텍스트의 단어를 확률적 지표에 의해 평가하여 특별한 사전학습 없이 이미지에 의미를 부여하는 키워드를 추정

#### □ 주요 기술별 연구개발 현황

- 영상분석: 영상인식기술은 ObjectVideo, iOmni-Scient, AgentVi, Bosch 등 해외 업체들이 절대 강세를 보이고 있으며, 물체와 상황 인지 및 추적 기술 개발이 활발히 진행 중
- 공간분석: 미국을 중심으로 한 세계 각국의 공공 데이터 개방이 있고, Google, 야후, 트위터 등 기업들과 연구기관 및 정부기관에서 다양한 플랫폼과 시스템 구축 서비스 제공 중
- 스트림분석: 스트림 분석 플랫폼으로는 Oracle, Microsoft, IBM, Streambase 등의 외산 상용 솔루션이 주로 있으며, 데이터 스트림 마이닝을 위한 오픈소스 SW인 RapidMiner, MOA도 존재
- 고급분석: 해외업체들이 예측분석 솔루션 분야에서 강세를 보이고 있으며, 특히 IBM, SAS는 포레스터가 선정한 가장 경쟁력 있는 솔루션 업체
- 시각화기술: IBM, SAS, SAP, 오라클 등 대형 IT 기업들과 타블로나 스팟파이어 등 시각화 전문 업체들이 솔루션 출시
- 빅데이터 서비스 기술(BDaaS): Amazon, Google, Oracle, Microsoft는 클라우드형 빅데이터 분석 플랫폼 제공

## (2) 해외 기업 기술 동향

- 빅데이터 분석기술이 결합된 플랫폼의 주요 플레이어는 IBM, Google, Microsoft가 센싱디바이스, 웹, 음성 등의 데이터들을 결합하고 분석하여 플랫폼 서비스를 제공하고 있음
  - (IBM) 음성 및 이미지 인식, 자연어 처리, 번역, 데이터 분석 등에 인공지능 서비스를 제공하고 있음
  - (Google) 실시간의 빅데이터를 축적하고 전 세계에 설치한 IoT 기기를 Google 클라우드 플랫폼에 연결하여 실시간으로 빅데이터 분석하는 플랫폼을 개발
  - (Microsoft) 인프라 측면에서 클라우드 플랫폼 애저를 기반으로 SaaS솔루션을 제공하여 기업 및 고객의 IoT 시장 진입 장벽을 낮추고 신제품을 개발하는데 지원하는 등 고객측면에서 초점이 맞춰짐
- (Amazon) Amazon 웹 및 알렉사를 통해 클라우드 기반의 인공지능경망 기반분석 서비스와 음성인식 서비스를 융합적으로 제공 중
  - 기존의 사용자가 생성하는 문서, 웹, 음성 데이터들을 사물인터넷과 결합하고 인식기술을 고도화 서비스 제공
  - 기존의 WiFi, 3G/4G/LTE, Bluetooth, Zigbee, 시리얼 통신 등 사람·사물·서비스를 연결할 수 있는 유·무선 네트워크를 활용하며 IoT의 구성요소를 특정 기능을 수행하는 응용서비스와 연동하는 역할 수행
  - 정보의 센싱, 가공, 추출, 처리, 저장, 판단 기능을 의미하는 검출정보 기반기술과 위치정보 기반기술, 보안기술, 데이터마이닝 기술, 웹서비스 기술 등이 단위 기술로 구성
  - 온라인 구매추천 및 예측에 빅데이터를 활용하고, 클라우드 기반의 빅데이터 분석 서비스 제공
    - 누적된 구매 빅데이터를 이용, 구매 패턴을 분석해 관심 있을 만한 상품을 추천하고 있으며, Amazon 매출의 1/3 이상이 구매 추천 서비스에 의해 발생
    - 기존의 주문과 검색 내역, 위시리스트와 쇼핑카트에 담아놓는 상품, 반품 내역, 마우스 커서가 머무르는 시간 등 소비자들의 행동 및 구매 패턴 예측을 기반으로 예상 구매 물품을 미리 포장해서 고객과 가까운 물류창고에 옮겨 놓는 예측배송을 통해 배달 시간과 물류비용을 절감함
- (후지쯔) 농지작물 및 작업에 대한 이미지 등의 데이터를 분석하는 플랫폼서비스를 제공
  - 클라우드 기반의 센싱네트워크 장비를 통해 들어온 이미지 등 각종 빅데이터를 분석하여 농지작물의 실적과 수확량 품질 등을 확인할 수 있도록 함
  - 기후와 토양환경 등에 대한 센서로 수집되는 데이터와 과거 수확 실적 등 빅데이터를 융합·분석하여 최적의 파종, 농약 살포, 수확 시점을 제공
- (시스코) 인프라 측면에서는 IT 아키텍처, 네트워킹, 클라우드 인프라, 실시간 분석, 보안 플랫폼, 벤처 투자 등으로 신산업 생태계 창출
- (Intel) IoT 환경에 적합한 퀴크(Quark) 프로세서를 기반으로 윈드리버의 클라우드 서비스, 다양한 분석 기능 등을 포함한 새로운 제품군 공개

- (Walmart) 재고 분석 결과를 외부 협력업체 등에 공개하고 물류, 재고 관리 현황의 실시간 분석을 통해 경쟁력 강화
  - 월마트는 소셜데이터를 이용하여 지역주민의 관심을 분석하고, 캘리포니아 주 마운틴뷰 지역은 차전거, 아칸소 주 벤톤빌은 낚시 관련 상품을 보강
- (ConocoPhillips) 방대한 부품 데이터를 분석하여 장애발생 징후 패턴을 발견하고 필요 부품의 사전 교체로 가동중단 최소화
  - 가동중단 발생률을 80% 감소시켜, 생산성 20% 증가 및 운영비용 7억 달러 절감
- (Google) 검색 통계 분석을 통해 노후 차량 보상 프로그램의 호응도를 분석하여 정부 예산이 부족할 것을 예측
  - 미국 정부는 4개월 시행 목표로 10억 달러를 편성하였으나, 1주일만에 소진되고 20억 달러를 긴급 추가 편성
- (Aviva) 운전자의 운전 패턴에 기반을 둔 고객 맞춤형 보험 상품 ‘RateMyDrive’ 출시
  - 차량 내 운행기록 장치를 통해 실제 운전 행태를 수집 및 분석하여 주로 운전하는 시간과 지역 등을 감안해 보험료를 산정하는 ‘Pay-as-you-drive’ 상품인 ‘RateMyDrive’를 내놓아 고객들로부터 좋은 반응을 얻음
- (Southwest Airlines) 비행기 좌석 스크린에 승객별로 맞춤형 광고 제공
  - 전세계적으로 5억 명에 달하는 고객 정보를 갖고 있는 액시엄(Acxion)사의 DB에 저장되어 있는 항공기 탑승객의 쇼핑 습관과 구매 패턴 등을 분석한 후 승객별 최적화된 광고를 제공
- (Target) 여성 고객의 임신 여부를 선제적으로 파악하기 위해 빅데이터 분석을 활용하여 맞춤형 프로모션 제공
  - 고객의 임신을 높은 확률로 추정할 수 있게 해주는 주요 아이টে에 대한 검색이나 구매 등 온/오프라인에서 행해지는 고객의 모든 활동을 지속적으로 축적하여 고객의 임신 주기까지 알아맞힐 수 있는 수준의 정보를 확보
- (Zara) 빅데이터 분석을 활용해 전세계 매장의 판매 현황을 실시간으로 분석한 뒤 고객 수요가 높은 의류를 실시간으로 공급할 수 있는 물류망을 구축

## 나. 국내 기술 동향

- 빅데이터 분야의 기술경쟁력 평가 결과, 최고기술국은 미국으로 나타났으며 우리나라의 경우는 최고기술국 대비 71% 수준으로 나타났고 중소기업은 64% 수준으로 평가되었음
  - 최고기술국 대비 우리나라의 기술격차는 2.3년으로 평가되었으며 중소기업의 경우는 3.3년으로 평가되었음

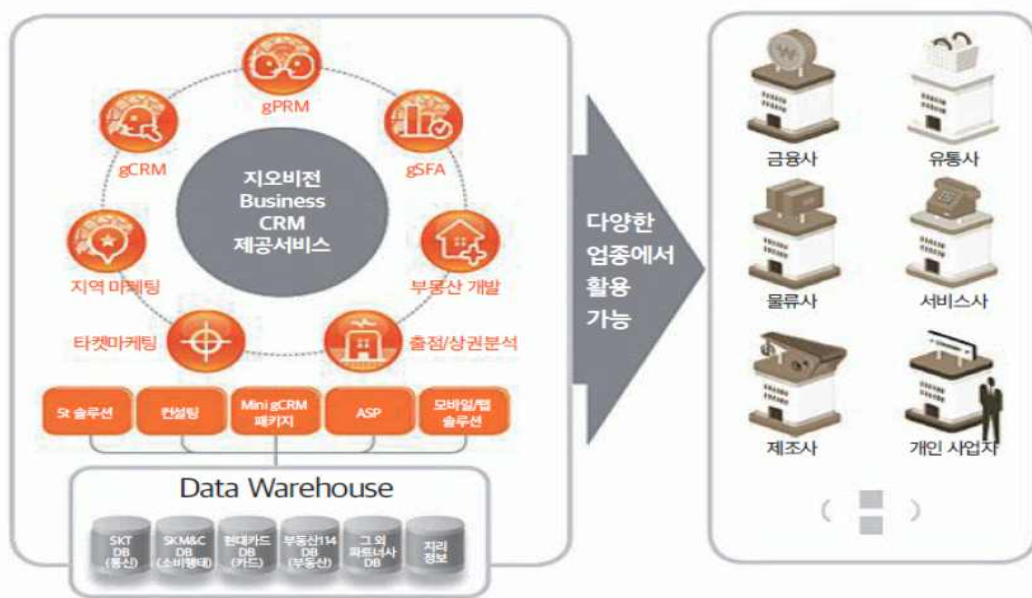
### (1) 기술 동향

- 국내 빅데이터 예측·분석, 이종소스 분석 등의 기술은 초기단계로 기술개발의 필요성 증대
  - 비정형적 데이터와 관련된 예측 연구는 시작단계이며, 시각화 기술은 데이터마이닝 작업에 기초한 정보 전달보다는 메시지 전달을 위한 시각 표현 위주의 작업이 이루어짐
- 국내 빅데이터 활용기술 수준은 낮은 편이며 주로 금융 및 통신 산업 관련 도메인 지식 기반 빅데이터 기술을 활용
- 텍스트 분석/음성 분석: ETRI에서는 웹데이터 및 대규모 코퍼스로부터 반자동으로 언어 분석에 필요한 지식 추출 방법을 개발하여 기술문서 자동번역 시스템 탑재
- 2019년 국내 중기신청과제 228개 중 28,832건 중 텍스트 분석 및 음성 분석기술을 포함하는 과제는 5개 과제 500건으로 전체 중 약 1.73%로 나타남
  - ETRI에서는 대규모 데이터 기반에 자동번역시스템의 성능을 개선하는 연구를 추진하고 있으며, 해외와 달리 규칙 기반의 자동번역방법(rule-based machine translation)을 채택한 것이 차별점이며 번역 지식의 규모와 품질이 성능에 영향을 미치는 핵심요소임
- 영상 빅데이터 분석: 국내 영상분석 솔루션 기업들은 자체 개발한 분석 알고리즘으로 기술을 고도화하면서 시장발전을 이끌고 있음
- 2019년 국내 중기신청과제 228개 중 28,832건 중 영상정보처리 기술을 포함하는 과제는 17개 과제 2,885건으로 전체 중 약 10%에 달함
  - 인텔리빅스: 자체 개발한 분석시스템을 탑재하여 오브젝트기반 사물인식 기술을 채택하고, 높은 인식률을 제공함 또한 딥러닝 기술을 적용한 차세대 기술 개발을 진행하고 있음
  - 비전인: 딥러닝을 이용한 객체인식 알고리즘을 자체 개발해 찾고자하는 객체를 정확하게 찾아내는 기술을 확보하고 있으며 빠르게 움직이는 객체뿐만 아니라 저조도 영상에서도 탐지가 가능하고 화재, 연기 등을 찾아낼 수 있음
- 공간분석: Daum, NHN, 솔트룩스, SK텔레콤 등의 업체들이 자체 관련 기술을 개발하여 공간 빅데이터 서비스를 제공하고 있으나, 공간 빅데이터 시장에서 전문인력 및 연구 역량을 체계적으로 확보하고 있지 않아 관련 연구는 아직 초기 단계라 평가됨
- 빅데이터 서비스 기술(BDaaS): 클라우드 환경에서 여러 사용자 간의 데이터 공유와 분석을 지원하기 위해 업체 중심으로 멀티테넌트 Hadoop 개발이 진행 중

(2) 국내 기업 기술 동향

- (솔트룩스) 지능형 빅데이터 분석 플랫폼 빅데이터 스위트(Bigdata Suite)을 개발하여 대규모 데이터를 분석·처리하는 기술을 갖추
  - 2018년까지 공간·행정·민간 정보 등 341종의 데이터를 수집하여, 놀이 시설 안전사고의 여러 융합 DB의 결과를 지도로 보여줘 현황을 쉽고 직관적으로 파악
  - 이 플랫폼은 시계열 정보가 담긴 융합 데이터셋과 기초 데이터셋의 시계열 정보에 따라 '화면 분할 방식' 및 '슬라이드 방식'으로 시각화 제공
  - 빅데이터의 분산처리를 돕는 공간하둡(Hadoop) 플랫폼으로 분석 결과를 지도 위에 시각화해 보여주고 미래 결과를 예측
  
- (SK텔레콤) 고객 데이터와 외부 데이터를 기반으로 상권 분석을 해주는 빅데이터 서비스 지오비전(Geovision) 개발 및 제공
  - 전국 유동인구를 5분 단위로 확인할 수 있는 국내 유일 서비스로 한 가지국의 통신 반경을 10m 단위로 쪼개 관리하는 '피셀(pCell)' 기술을 적용해 기존 '셀 ID 측위 방식' 대비 5배 이상 정밀한 위치 정보를 제공
  - 경북경찰청은 지오비전을 활용하여 지역 유동 인구 데이터를 확보하고 효율적인 코로나-19 확산 방지와 국민 체감안전 향상을 위한 핀포인트 순찰을 지원

[ SK 텔레콤 지오비전 개요도 ]



\* 출처: SK텔레콤 Biztworld 홈페이지

- (LG CNS) 클라우드 기반으로 인공지능을 활용해 데이터 수집에서 분석, 결과 시각화까지 일련의 과정을 빠르게 수행하여 기업들의 데이터 기반 의사결정을 돕는 '답(DAP)' 서비스를 출시



- (씨이랩) 빅데이터 튜닝 전문 플랫폼인 ‘아큐튜닝(Accu.Tuning) for 씨이랩(XIILAB)’을 출시해 AI 통합 기술 수준 향상에 기여
  - 아큐튜닝을 통해 별도 학습 없이 빅데이터 전처리 및 영상 속 자연어 자동 분류, 자동 데이터 탐색·분석, 최적의 분류 예측모델 자동 생성 등을 제공하며 영상인식 빅데이터 분석 효율성 향상
  - 데이터 전처리(Preprocessing)·데이터 증강·모델 생성·모델 분석 등 전문가용 고급 기능을 손쉽게 활용하게 해주는 기능 제공
- (신한카드) 빅데이터 분석을 바탕으로 고객 라이프 스타일 맞춤형 상품 추천 서비스 제공

[ 신한카드의 빅데이터 분석 기반 고객 라이프 스타일 분류 ]



\* 출처: 신한카드 홈페이지

- (KT 넥스알) 하둡 기반의 빅데이터 분석 플랫폼인 NDAP을 주력 솔루션으로 제공
  - 빅데이터 배치 처리 및 근-실시간 검색 플랫폼으로 빅데이터 분석을 위한 모든 작업 (수집/저장/분석/검색/관리 등) 및 실시간 데이터 질의 처리 가능
  - 실시간으로 발생하는 다양한 형태의 머신/휴먼데이터를 수집하여 데이터에 기반한 신속한 비즈니스 의사 결정을 지원하는 Lean Stream 솔루션 제공
- (데이터스트림즈) 분산 메모리 기반 실시간 데이터 통합 플랫폼 제공
  - 다양한 데이터를 쉽고 빠르게 수집/저장하는 분산 메모리 기술을 적용한 순수 국산 솔루션으로 초기 도입 비용을 최소화한 고성능/고효율의 빅데이터 통합 처리 운영 아키텍처 제공
  - TeraONE의 거버넌스 체계를 통하여 수집된 데이터(실시간/배치, 정형/비정형)의 품질을 개선한 양질의 데이터로 분석의 신뢰성 향상



□ (다음소프트) 소셜 네트워크 마이닝 분야의 전문 분석을 수행

- SNS 정보 기반 여론 진단 서비스, 소셜미디어, 트위터, 블로그 트렌드 분석 등을 서비스하며, 소셜미디어 상의 데이터에서 의미 있는 정보를 찾고, 조직화함으로써 정보간의 관계나 패턴, 트렌드 등을 분석하는 서비스 제공
- 버즈량 분석: 특정 이슈 및 아이템 관련 문서의 발생 건수를 시간별, 일별, 월별, 연도별로 분석하여 화제가 된 지점 포착
- 연관어 분석: 특정 이슈 및 아이템 관련 문서에서 사용된 단어를 텍스트 마이닝 기법으로 처리하여 통계적으로 분석
- 감성 분석: 특정 이슈 및 아이템 관련 문서에서 연관어 분석과 비슷한 방식으로 감정과 관련된 단어의 사용 비중 분석
- 빅마우스 분석: 특정 이슈 및 아이템 관련 문서 중에서 영향력 있는 사람의 작성 여부 파악

□ (비스텔) 반도체, 디스플레이, 자동차, 철강, 에너지 분야의 스마트 팩토리를 위한 빅데이터 솔루션 제공

- 산업 IoT, 클라우드, 빅데이터 기술을 기반으로 제조 환경의 실시간 데이터를 검출, 분석, 예측하여 고장 탐지, 데이터 분석, AI 기반의 운영 등을 실현하기 위한 솔루션 제공
- DFD(Dynamic Fault Detection): 실시간 센서 추적 분석, 거짓 경보 제거 등을 통한 실시간 모니터링 기능 제공
- WQP(Wafer Quality Predictor): 센서 추적 데이터를 평가해 각 웨이퍼에 대한 품질 점수를 부여하고 웨이퍼 재고 품질에 대한 요약 정보를 대화식 대시보드를 통해 제공
- CM(Chamber Matching): 최고의 챔버를 자동식별하고 모든 센서에 대해 완전 추적 분석을 수행하며 시간 기반으로 챔버 분석을 수행

□ (더아이엠씨) 비정형 데이터 일관처리솔루션(TEXTOM)과 텍스트 기반의 기업 및 공공기관 중심의 모니터링 및 마케팅 서비스(TEXTOMi)를 제공

- 비정형 데이터의 일관처리솔루션인 TEXTOM은 웹 환경에서 분야와 관계없이 누구나 비정형 데이터를 수집, 정제, 분석, 시각화에 이르는 분석과정을 쉽고 빠르게 처리할 수 있도록 빅데이터 분석에 사용자 편의성을 제공하고 있는 것이 강점임
- Web, Portal, SNS, 뉴스, 사용자요청사이트(커뮤니티 사이트 등)를 탑재된 크롤러에 의해 다양한 키워드의 수집이 가능
- 국문, 영문, 중문으로 이루어진 언어별 특성에 맞는 형태소분석기를 지원하고 있으며, 수집된 텍스트 데이터의 빠른 정제과정을 지원
- 정제된 데이터를 활용하여 시멘틱 네트워크(Semantic Network)와 기계학습(Machine Learning) 분석이 가능한 매트릭스, 감성 분석, 토픽모델링, N-gram, TF-IDF 등 분석의 확장성을 고려한 다양한 분석 결과물을 제공
- 빈도수, 공출현빈도, 토픽모델링 등 다양한 시각화를 제공함으로써 분석결과에 대한 직관적인 빅데이터 탐색이 가능함

## 4. 정책 분석

### 가. 해외 정책 동향

#### ◎ 미국

- 미국 연방 정부는 '19년 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
  - 10대 원칙은 윤리적 거버넌스, 의식적 설계 및 학습 문화라는 세 범주로 구성되고, 40대 실행 전략은 10대 원칙 실현을 위한 5~10년 간의 중장기 전략
  - 데이터의 가치를 육성하고 공공 부문에의 활용을 촉진하기 위한 실천 전략, 데이터의 관리와 보호를 위한 실천 전략 그 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 둔화 될 것으로 전망
- 대통령 직속기구인 국가과학기술위원회(National Science and Technology Council: NSTC) 에서 연방정부 부처 기관의 ICT 연구개발 프레임워크인 NITRD 프로그램을 운영
  - 2018년 빅데이터, 로봇, 인공지능, 고성능 컴퓨팅 시스템 인프라 운용으로 영역을 확대하였으며 2019년 사이버보안을 비롯한 개인정보 등으로 확대 변경하여 시행 중
  - 빅데이터와 인공지능 분야를 강화하기 위해 투자 확대 및 프로그램을 신설하였고, 실제로 예산은 인프라 및 소프트웨어에 비해 증가
- 2009년도부터 시작된 오픈데이터 정책의 대표적 서비스인 Data.gov는 트럼프 정부에 들어서도 계속 진행
  - 2019년 6월 말 기준 Data.gov에는 251,906개의 데이터 세트 등재되어 누구나 이 데이터를 활용하여 가치를 제고할 수 있도록 소프트웨어 어플리케이션 함께 제공
- 빅데이터 R&D전략 계획 [ BDIWG, '16.5 ]
  - 범부처 차원에서 빅데이터 7대 R&D 전략과 18개 세부과제를 제시하여 차세대 빅데이터 기술 역량 확보, 신뢰성 탐구 및 획기적인 발견을 위한 빅데이터 R&D 지원, 빅데이터 혁신이 가능한 사이버 인프라 연구 향상 등이 7대 전략으로 선정

## ◎ 유럽

### □ 개인정보보호 규정(GDPR, General Data Protection Regulation) [ '18 ]

- 2018년부터 시행되어 유럽 의회에서 유럽 시민들의 개인정보 보호를 강화하기 위해 만든 통합 규정

### □ 공통 유럽 데이터 공간 [ '18 ]

- 2017년 1월 채택한 유럽 데이터 경제 구축의 후속조치로 서로 다른 부문, 국가 및 분야의 데이터 통합을 제안하고 있으며, 과학적 정보의 접근 및 보존에 관한 2012년 권고안을 개정하고 있음
- B2B, B2G 환경에서의 민간 부문 데이터 공유에 대한 지침 등을 포함하고 있으며 2018년 5월 일반정보보호규정이 시행됨에 따라 데이터 관리와 개인정보보호를 위한 국제적 프로토콜이 될 가능성이 있어 국내에도 시사하는 바가 큼

### □ 유럽 47개 주요 교통, 물류, IT 관련기업 및 기관으로 구성된 컨소시엄이 30개월 간 1,870억 유로의 예산으로 빅데이터 기반의 Transforming Transport 프로젝트 추진 [ '17.1 ]

### □ 유럽연합은 '15년부터 빅데이터에 대해 향후 5년간 산업 컨소시엄의 20억 유로와 공적 자금 5억을 합친 25억 유로(약 3조 3,870억 원) 투자

- 아토스(Atos), IBM, 노키아 솔루션 앤 네트워크스(Nokia Solution and Networks), Microsoft(N), 오렌지(Orange Telecom), SAP, Siemens 등의 기업들과 연구기관들 주도로 민관협력 추진 중

## ◎ 영국

### □ 데이터의 경제적 가치 보고서[ '18 ]

- 영국 재무부에서 발표한 보고서에는 데이터를 중심으로 한 영국 경제 발전의 5가지 과제\* 발표

\* 1. 데이터 소유권과 통제권에 대한 명확화 2. 개인데이터에 대한 보호 유지 3. 공공부문 데이터의 개방성 4. 데이터 상호운용 및 표준 강화 5. 안전하고 합법적인 데이터 공유 환경 구축

### □ 빅데이터 역량강화 종합전략 및 정부 변화전략[ '17 ]

- 공공서비스를 위한 정부기관간의 경계를 허물어 데이터를 효과적으로 사용할 수 있도록 하는 디지털 경제 법안이 제출되었으며 통과되고 나면, 데이터 공유 조항을 통해 전체 정부의 데이터를 효과적으로 사용할 수 있도록 함
- 정부차원에서 데이터 과학과 분석능력 구축, 데이터 공유 관련 윤리교육, 국가 데이터 인프라 구축 및 보안성 확보, 데이터 검색 도구 개선, 주요 리포지토리 저장 및 관리 방식 변화 등을 추진함

◎ 중국

□ '20년 9월 8일 데이터 안보에 관한 글로벌 이니셔티브를 발표

- 미국 정부가 중국산 소셜네트워크(SNS) 틱톡·위챗 사용을 금지한 조치에 대해서 대응하기 위해 발표한 글로벌 데이터 안보 표준안으로 데이터의 국적성을 강조
- 기업은 사업하는 국가의 법률을 존중하고, 기업이 해외에서 취득한 데이터를 그 국가(해외국가)에만 저장하도록 강요해서는 안 된다(4항)
- 다른 국가의 주권, 관할권 및 데이터 관리 권한을 존중하고 타국 데이터에 접근해선 안 된다(5항)

□ 교통관광서비스 빅데이터 활용 시범사업 추진 통지 [ '18 ]

- 교통, 관광서비스에 빅데이터 활용을 위한 플랫폼 구축 추진

□ 스마트시티 관련 빅데이터 및 클라우드 활용 요강 [ '17 ]

- 각 지역의 스마트시티 건설 시 빅데이터와 클라우드를 활용한 세부방안을 발표

□ 빅데이터 산업 발전 계획 ('16~'20) [ '17 ]

- 이미 일상생활에서 광범위하게 활용되는 빅데이터를 제조업 분야로 확대하며, 특히 생산모델 고도화 및 스마트 공장을 추진하는 과정에 빅데이터 응용 필요성 강조
- 빅데이터를 중국의 인구·경제 규모를 활용한 전 산업 발전 기회로 인식하고 10개 이상 글로벌 빅데이터 선도기업, 500개 응용서비스 기업 육성 목표 추진
- '20년까지 빅데이터 관련 제품 및 서비스업 매출을 1조 위안까지 확대하겠다는 목표 제시
- 데이터 개방 확대, 플랫폼·오픈소스 기술 지원, 빅데이터 전문 SW 수준 향상, 전문인재의 공급, 데이터거래소 등 생태계 조성
- 정확한 고객 수요 파악과 분석을 통해 R&D의 효율성을 높이는 한편 조립·생산 라인의 경쟁력을 높임으로써 생산예측과 관리를 강화하고 경영 효율화를 제고할 수 있을 것으로 기대

□ 빅데이터의 응용 기술 표준 및 통계 표준 제정

- 통계국은 국가통계국과 텐센트 등 6개 기업 공동으로 연구를 진행하여 정부 통계 데이터를 이용하여 빅데이터 기반 공동 수집·처리·분석·탐색·발표 기술 개발 표준과 통계 표준을 제정

## ◎ 일본

- 일본 정부는 2018년 6월 '데이터 활용의 기반 조성을 위한 국민 데이터 활용 추진 기본 계획' 수립
  - 일본은 5가지 사항\*을 담은 기본계획을 토대로 세계적인 데이터 활용 국가로 거듭나기 위한 의지 피력
    - \* 1. 국민 데이터 활용의 추진에 관한 기본적인 방침 2. 행정기관의 국민 데이터 활용에 관한 사항 3. 지방공공단체와 기업의 국민 데이터 활용 촉진에 관한 사항 4. 국민 데이터 활용과 관련해 정부가 중점적으로 강구해야 할 시책 5. 국민 데이터 활용의 추진에 관한 시책을 종합적이고 동시에 효과적으로 추진하기 위해 필요한 사항
- 정부 주도로 정부, 지자체, 민간 보유의 모든 데이터를 정리하여 개방할 계획 발표 [ '18 ]
  - '20년에 포털 사이트의 형태로 공개할 예정으로, 취합한 빅데이터 자료들을 모두 데이터베이스화 한 후 일반에게 공개하여 다방면에서 인공지능을 활용한 분석 자료로 사용할 예정
  - 국가가 보유한 각종 통계데이터, 지도데이터, 기상데이터와 민간에서 보유한 도로 상황 및 교통량과 관련된 데이터, 작물의 재배 상황 정보, 위치정보 등 광범위한 분야의 정보 취합이 목표
  - 인공지능 분석이 용이하도록 행정용어에서부터 산업용어까지 빅데이터 내에서의 용어를 통일·정리하여 취합할 예정
  - G20 정상회의에서 데이터베이스 구축에 미국과 유럽의 참가를 요청해 10억 명 이상의 정보가 집적된 초국가적인 데이터베이스 Pool 구축을 최종 목표로 함
  - '14년부터 추진한 공공기관의 데이터 공개를 통한 정부의 투명성, 경제 활성화, 공공서비스 향상 목적의 '빅데이터 활용을 통한 Open Data 정책'의 연장선으로 파악
- 미래투자전략 2017-Society 5.0의 실현을 위한 개혁 [ 총무성, '17.6 ]
  - 빅데이터를 ICT 융합 활성화 촉매로 활용하기 위해 이노베이션·벤처를 탄생시키는 선순환 시스템을 구축하여 빅데이터 산업 집중 투자를 추진
  - 건강수명연장, 이동혁명 실현, 공급망 첨단화, 쾌적한 도시 만들기, 핀테크의 5대 신성장 전략분야에 데이터를 활용한 기반구축
  - 새로운 사회의 인프라로서 '데이터 기반(현실데이터 플랫폼)'을 구축하고 데이터 활용을 향한 제도 정비, 교육·인재 역량 강화, 혁신벤처 선순환 시스템 구축
  - '20년까지 국민 데이터 활용추진 전략회의를 중심으로 공공데이터 개방 집중 실시, 데이터 활용 권한을 명확히 하는 기업 간 계약 가이드라인 전개, 데이터 유통을 위한 법제 정비 등 데이터 거래시장의 제도 정비 등을 실시

## 나. 국내 정책 동향

- 과학기술정보통신부는 데이터 활용을 통한 경제·사회 전반의 혁신 창출을 위해 3월 30일부터 데이터 바우처와 마이데이터 사업 등 `20년도 주요데이터 활용사업 공모를 추진

[ 2020년 주요 데이터 활용 지원 사업 내용 ]

사업명	예산규모	지원규모	지원내용
데이터 바우처	575억 원	1,420개 지원	중소기업, 소상공인, 1인창조기업 등에 데이터 구매 및 가공 지원
마이데이터 실증서비스	97억 원	8개 과제 지원	개인정보 보유 기업/기관과 컨소시움을 통해 다양한 실증서비스 개발 지원
데이터 플래그십	50억 원	10개 과제 지원	빅데이터·AI기반 혁신서비스 개발 지원
중소기업 빅데이터 분석 활용지원	7.5억 원	3개 지자체 지원 (지역별 25개 기업 지원)	중소기업의 빅데이터 활용 방법 및 데이터 수집, 가공, 분석 역량 등 컨설팅·교육

\* 출처: 대한민국 정책브리핑 브리핑룸, 과학기술정보통신부

- 정부는 2020년 7월 ‘한국판 뉴딜 종합계획’에서 10대 과제 중 하나로 ‘데이터 댐’ 발표
  - 분야별 빅데이터 플랫폼 확대, 공공데이터 14.2만개 신속 개방, AI학습용 데이터 1,300 구축 등 빅데이터 확충 추진
  - 분야별 데이터 확충·연계·활용, 5G·AI 융합서비스를 통한 글로벌 선도로 2025년까지 데이터시장 규모는 43조원으로 성장 예상

[ 데이터 댐 개념도 ]



\* 출처: 과학기술정보통신부

- 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 관련 법들이 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
  
- 정보화진흥원(NIA)의 데이터 거래를 위한 10개 빅데이터 플랫폼 사업 [ '19 ]
  - “데이터를 가장 잘 다루는 나라”라는 데이터 경제선언(18.8.31) 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, 신규사업창출, 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
  
- 데이터산업진흥원의 데이터 거래를 위한 바우처 사업 [ '19 ]
  - 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 200개 사를 선정하여 데이터 구매 부문과 일반 가공부문, Si가공 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



□ 데이터 산업 활성화 전략: I-Korea 4.0 데이터 분야 계획 I-Data [ 관계부처 합동, '18.6 ]

- 4차 산업혁명 시대 데이터가 모든 산업의 발전과 새로운 가치 창출의 촉매 역할을 하는 ‘데이터 경제(Data Economy)’로의 패러다임 전환에 입각하여 데이터 이용제도 패러다임 전환, 데이터 가치사슬 전방위 확산, 글로벌 데이터산업 육성기반 조성 등을 추진과제로 발표
- 공공데이터의 원칙적 개방(국가안보·개인정보 제외) 등을 통해 공공부문 핵심 데이터 (통계·연구·참조표준 등) 구축 개방 전면화
- 데이터의 효율적인 저장 관리를 위한 클라우드 본격 확산 및 데이터 유통 촉진을 위해 민간 공공을 연계한 개방형 데이터 기반을 구축하고 '22년까지 중소 벤처기업 1만개 이상에 클라우드 도입 지원
- 빅데이터 산업이 4차 산업혁명을 선도하는 핵심 성장주체로 거듭날 수 있도록 4차 산업혁명 요소기술 융합, 분석 전문인력 양성, 컴퓨팅파워 기반 기업성장 인프라 지원 등을 아우르는 역동적 산업 생태계 조성
  - 빅데이터 전문기업 성장 지원으로 데이터 강소기업 100개를 육성하고 빅데이터 개발에 필수적인 컴퓨팅 자원, 데이터셋 등이 집적된 판교 글로벌 ICT 혁신 클러스터 지속 고도화 및 활용 지원
  - K-ICT 빅데이터센터, K-ICT 클라우드 혁신센터, 글로벌 IoT 시험인증센터, 정보보호 클러스터, HPC 이노베이션 허브 등 세계 최고수준의 성능을 갖춘 초고성능 컴퓨터 5호기상도 연계

□ 대통령 직속 4차 산업혁명위원회에서는 데이터 산업 활성화 전략을 발표하여 개인정보 보호와 활용의 균형 및 데이터 경제 활성화 추진 [ '18 ]

- 정보주체인 개인이 데이터 통제·활용권을 가지게 하고 개인정보의 안전한 활용을 위한 법제화·안심존 구축·신기술 적용 등을 지원하며, 데이터 구축·개방, 저장·유통, 분석·활용의 빅데이터 활용 전 과정의 지원 체계 마련
  - 데이터 주권을 강화하여 개인정보가 해외에서도 안전하게 보호되도록 국외 이전 중단 명령권, 국외 재이전시 보호조치 의무 신설 등 제도 정비
  - 데이터 가치 사슬 전주기 혁신을 위하여 데이터 구축 개방(4차 산업혁명 핵심데이터 등, ~'22) → 저장 유통(데이터 거래의 비즈니스화, '18~) → 분석 활용(산업·사회 혁신 활용, '18~) 등 전 과정에 걸쳐 실제데이터 기반 영역별(의료·교통 등) 국가 빅데이터 지원체계 마련
- 빅데이터 핵심기술 확보 및 전문인력 확보를 통한 기반 조성 목적
- 국회 4차 산업혁명특별위원회는 이와 관련하여, 개인정보보호법과 정보통신망법, 신용정보법 등 중복 조항 정비 및 강력한 사후규제를 도입하여 가공에 대한 활용의 안전성을 보장하도록 입법 권고안 발표



## 5. 중소기업 전략제품

### 가. R&D 추진전략

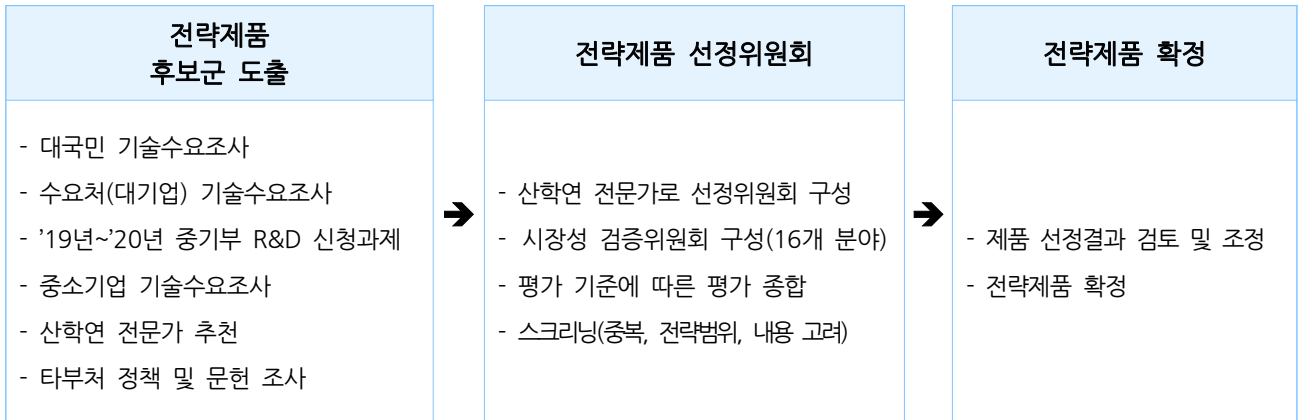
Factor	기회요인	위협요인
정책	<ul style="list-style-type: none"> <li>• 데이터 개방 여론 및 입법 추진</li> <li>• 공공부문에서의 제조업, 에너지, 농수산업, 스마트시티 등 분야 지능화 혁신을 추진</li> <li>• 정부주도의 중소기업 지원정책 및 핵심기술 지원과 전문인력 양성에 따라 기반조성을 전략적으로 추진</li> </ul>	<ul style="list-style-type: none"> <li>• 빅데이터에 내재되어 있는 개인정보 등 보안 문제에 따른 데이터 개방 정책의 이해관계자들 간의 갈등</li> <li>• 정부주도의 정책적 지원에도 불구하고 사회 전반으로 확산할 수 있는 정치적, 사회적 합의가 필요함</li> </ul>
산업	<ul style="list-style-type: none"> <li>• 우수한 ICT 인프라 확보로 다양한 융복합 빅데이터 플랫폼 서비스 개발 가능</li> <li>• 개별 산업분야 별로 내부 빅데이터 활용 특성이 잠재되어 있어 산업별, 산업간 활용 가능성이 무궁무진함</li> </ul>	<ul style="list-style-type: none"> <li>• 빅데이터의 원천이 대기업 및 일부 기업의 독점, 공유 및 거래 미흡</li> <li>• 빅데이터의 산업별 활용 가치에 비해 빅데이터 접근의 어려움, 개별 산업별로 또는 공공부문과 함께 협력적인 거버넌스가 필요</li> </ul>
시장	<ul style="list-style-type: none"> <li>• 전 산업 분야에 빅데이터 이슈가 자리잡고 있어 빅데이터를 축적 및 활용서비스 수요가 증가</li> <li>• 세계적인 빅데이터 분석 및 활용 시장의 확대</li> </ul>	<ul style="list-style-type: none"> <li>• 축적되는 데이터와 데이터 품질에 비해 시장의 빅데이터에 대한 지나친 기대감과 비현실적 요구</li> <li>• 빅데이터의 활용 가능성에 비해 높은 데이터 독과점 시장구조 및 거래제도</li> </ul>
기술	<ul style="list-style-type: none"> <li>• 빅데이터 활용 및 분석의 단위기술 개발이 활발하게 진행되고 있음</li> <li>• 빅데이터 단위기술은 대기업보다는 중소·벤처 기업이 진출하기 용이한 분야로 기술개발에 따른 기술선점 가능성이 높음</li> </ul>	<ul style="list-style-type: none"> <li>• 단위기술 연구개발 성과에 비해 융합적 기술개발의 한계</li> <li>• 데이터 단위의 표준화와 데이터 품질 미흡</li> <li>• 해외 대기업과의 R&amp;D 비용 및 기술력 격차</li> </ul>



### 중소기업의 시장대응전략

- 단순한 기능적 수행뿐만 아니라 안정성과 신뢰성을 동시에 확보해 높은 시장의 기대치에 대응
- 단위기술별 연구개발 확대를 통한 기술선점과 단위기술별 융합기술 개발을 통한 독창성 확보
- 이종간 산업과의 융복합 빅데이터 거버넌스를 통한 기술경쟁력 확보
- 빅데이터의 협력적 구축과 융합분석모형을 결합한 플랫폼 서비스로 사업화 모델 구상

## 나. 전략제품 선정 절차



### □ 전략제품 후보군 도출

- (최근 신청 과제) 중소벤처기업부 R&D 지원 사업 '19년~'20년 상반기 신청과제
  - (기술수요조사) 중소기업기술정보진흥원 주관 SMTECH(중소기업 기술개발사업 종합관리시스템) 성과분석 대상 중소기업으로부터 기술수요 수신
  - (대기업 의견) 전략분야 관련 대기업의 중소기업 유망 제품 관련 인터뷰
  - (산학연 전문가 추천) 분야별 전문가 대상 후보 추천 의뢰 의견수렴
  - (타부처 정책 및 문헌조사) 타 부처 정책사항 및 문헌조사를 통한 품목 발굴
- ※ (재밍, Jamming) 데이터 기반의 전략제품 발굴을 위하여 인공지능 전략분야에 시범적으로 도입

### □ 전략제품 선정위원회

- (선정방식) 중소기업 적합형 기술로드맵 수립 및 전략 강화를 위해 전략제품 선정위원회의 평가와 시장성 검증위원회의 평가를 종합하고, 전략분야에 따라 평가항목의 가중치를 조절하여 반영
- (전략제품 선정평가위원회) 분야별 산·학·연 전문가 위원회를 구성하여 전략제품에 대해서 각 5개 항목을 평가 및 검토 진행
- (시장성 검증위원회) 시장성 검증이 필요한 분야에 대해서 해당 전략분야에 관련성이 높은 전문가와 VC(투자심사역)으로 구성된 위원회가 전략제품 평가 진행
- (평가항목) 시장성, 기술난이도, 개발기간, 수입의존성 및 중소기업 적합성을 기준으로 평가
- (평가기준) 전략분야의 대구분(한국판 뉴딜 및 소부장·뿌리산업)에 따라 평가항목의 가중치를 조절

### □ 전략제품 확정

- (검토 및 조정) 선정된 전략제품들에 대해 최종적인 타당성 검증 및 분야 간 전략제품 검토 및 조정을 통해 전략분야별 전략제품 확정

## 다. 전략제품 선정결과

### ◎ SI기반 데이터 가치 고도화 플랫폼

- 수집된 다양한 형태와 유형의 데이터 생성규칙과 분포를 찾아 학습하고 학습 데이터를 근거로 오류 데이터를 판정하는 기술

- 다양한 데이터의 융복합을 통한 효과적인 데이터 활용을 위해 데이터 정제/전처리에 소요되는 많은 시간과 노력이 절감 필요
- 가비지 인 가비지 아웃의 원리가 적용되는 데이터 생태계에서 최적의 데이터 활용의 토대 마련이 필요

### ◎ 유통-물류 빅데이터 구축 및 분석 시스템

- 유통-물류 빅데이터 구축 및 분석 시스템은 유통, 물류 등의 산업 전반에서 빅데이터가 흐름을 파악하여 최적화하는 솔루션으로, POS 및 수-발주 데이터를 토대로 딥러닝을 이용해 과학적으로 물류를 유통, 상-하 역시 효과적인 적재 방법을 제시하는 시스템

- 새로운 데이터가 더욱 급증 할 것으로 예상되며, IoT를 생성하는 데이터의 순차 처리 및 실시간 처리가 가능하게 될 전망
- 빅데이터는 기존의 정보 시스템과 다른 데이터와 결합해 데이터 융합과 SI를 활용한 고급 분석으로 가속화될 전망

### ◎ 스마트 헬스케어를 위한 빅데이터 수집 시스템

- 의료와 기술이 융합된 형태로 환자 사용자의 개인별 건강상태를 시간과 장소의 제약 없이 실시간으로 모니터링 및 관리하고 건강정보 및 질병상태 등을 분석하여 최적화된 맞춤형 진료를 제공하는 서비스 또는 시스템

- 전 세계적으로 심각한 고령화 추세와 급증하는 의료비 부담에 대한 국가적 대응이 시급한 상황. 이로 인해 스마트 헬스케어가 새로운 부가가치로 부상 중
- ICT 기술과 융합하여 헬스케어 및 신약 개발에 소요되는 임상 기간 및 비용 절감

### ◎ 빅데이터 분석 및 시각화 플랫폼

• 빅데이터 분석 및 시각화 플랫폼은 통상적으로 사용하는 소프트웨어가 수용할 수 없는 크기의 데이터인 빅데이터로부터 가치를 추출하고 결과를 분석하는 기술과 데이터 분석 결과를 유의미한 정보로 표현하는 기술의 총체를 의미

- 빅데이터 분석 및 시각화 플랫폼은 기계학습 및 인공지능을 활용한 분석 영역을 확대하고 있으며 결과 정보 전달 측면에서 계속 연구 및 개발 중
- 인공지능 기술을 적용하기 위한 수요의 증가로 기반이 될 분석 기술이 지능정보사회에서 대표적인 기술로 부상

### ◎ 빅데이터 기반 마케팅 인텔리전스 플랫폼

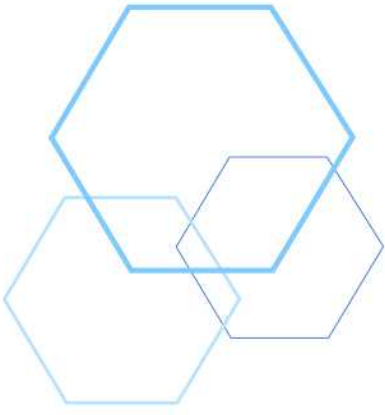
• 마케팅 인텔리전스 플랫폼이란 빅데이터와 정보분석 도구와 방법론을 활용한 마케팅 의사결정 지원 서비스로 데이터 기반의 시장 및 정책정보를 모니터링해 분석하여 다양한 통찰과 예측을 제공

- 제품의 모니터링, 신제품 기획 시기, 재고 조절 등 적절하고 신속한 대응이 생존할 수 있는 환경을 조성함에 따라 이를 모니터링, 예측하는 지능형 플랫폼의 요구가 증가
- 마케팅 인텔리전스 플랫폼의 사용을 통해 마케팅 과정 전반에 걸쳐 효율성 달성

### ◎ 공공 빅데이터 수집 및 분석 시스템

• 부처 및 기관별 유형과 형태가 상이한 공공 빅데이터를 수집하여 양질의 과학적 정책 서비스 설계와 활용이 가능하도록 가치 있는 공공 빅데이터의 정보를 추출하여 분석할 수 있는 시스템

- 정책 의제 설정, 정책설계, 집행의 과정에서 공공 데이터의 활용가치가 증가하고 있으며, 공공 빅데이터를 수집하고 가공하여 분석하는 기술에 대한 업무 연관성이 증가
- 협력-융합적 정책서비스 제공에 정책설계 다양화로 의사결정을 위한 분석지원 중요



전략제품 현황분석

# AI기반 데이터 가치 고도화 플랫폼





# 시 기반 데이터 가치 고도화 플랫폼

## 정의 및 범위

- 수집된 다양한 형태와 유형의 데이터 생성규칙과 분포를 찾아 학습하고 학습 데이터를 근거로 오류 데이터를 판정하는 기술
- 판정된 오류를 학습하여 오류데이터를 보정, 재생 추천을 통한 융합분석 및 데이터 활용 가치를 고도화할 수 있는 플랫폼

## 전략 제품 관련 동향

시장 현황 및 전망	제품 산업 특징
<ul style="list-style-type: none"> <li>• (세계) '18년 11억 9,500만 달러 규모에서 연평균 12.7%로 성장하여 '24년 24억 4,900만 달러 규모로 성장 전망</li> <li>• (국내) 2018년 15조 1,545억 원에서 연평균 12.4%로 성장하여 2024년 30조 5,406억 원 규모로 성장 전망</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터기반 경제가 활성화 되면서 데이터의 원천 가치에 대한 관심이 증가</li> <li>• 데이터품질관리에서 탈피해 데이터의 가치를 읽고 정비, 보정, 재생할 수 있는 기술이 필요</li> <li>• 데이터를 활용하는 다양한 분야에서 데이터 가치에 대한 필요성이 증가되고 있음</li> </ul>
정책 동향	기술 동향
<ul style="list-style-type: none"> <li>• 데이터품질관리 기술에 비해 데이터 가치 고도화 기술 개발이 미흡하여 원천기술 확보가 시급하며, 이에 대한 기초과학 연계정책이 필요함</li> <li>• 데이터산업의 직무별 부족률을 해결하기 위한 인재 양성정책과 함께 데이터과학의 융합적 지식을 공유할 수 있는 교육정책이 병행되어야 함</li> </ul>	<ul style="list-style-type: none"> <li>• 데이터의 유형과 특성에 따라 달리 생성되고 관리해석되는 데이터에 따른 이상치, 특이치, 결측치 보정 및 재생, 교정 등의 가치 고도화 연구가 이루어지고 있음</li> <li>• 시와 머신러닝을 활용한 이상치, 특이치, 결측치 처리 방법 및 모델을 탑재한 솔루션 개발이 이루어지고 있음</li> </ul>
핵심 플레이어	핵심기술
<ul style="list-style-type: none"> <li>• (해외) IBM, ORACLE, DATA LADDER</li> <li>• (대기업) 삼성에스디에스, 삼성전자, 케이티</li> <li>• (중소기업) 데이터스트림즈, 위세아이텍</li> </ul>	<ul style="list-style-type: none"> <li>• 시 및 머신러닝을 활용한 이상치 탐지 및 검출</li> <li>• 시계열 추세와 데이터 분포를 활용한 결측치 대체방법</li> <li>• 오류데이터 및 학습데이터 구축</li> <li>• 데이터 가치 모델과 결합한 데이터 가치 고도화 솔루션 개발</li> </ul>

## 중소기업 기술개발 전략

- 이상치, 결측치, 특이치 등에 대한 탐색적데이터분석을 기반으로 하는 데이터 가치 탐구에 대한 AI 및 머신러닝 등 방법론 개발
- 유형별, 특성별 생성형태별 데이터 수집과 오류데이터 등 교정, 재생 등 추천에 필요한 학습데이터 구축
- 산·학 연계 및 테스트베드 확보를 통한 개발기술 검증으로 국내시장 개척 및 선점 필요

# 1. 개요

## 가. 정의 및 필요성

### (1) 정의

- 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 SI기반의 데이터 오류 및 보정 자동화 기술
- 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

### (2) 필요성

- 다양한 데이터의 융복합을 통한 효과적인 데이터 활용을 위해 데이터 정제/전처리에 소요되는 많은 시간과 노력을 절감할 필요성
  - 정형/비정형/반정형 등의 데이터 유형과 특징은 데이터 통합의 어려움으로 작용되며, 융복합 분석시 데이터 결측과 편향의 문제를 해결하지 않으면 신뢰성 높은 분석결과를 도출하기 어려움
  - 이를 위해 기존의 다양한 분야의 데이터 유형에서 자주 발생하는 오류, 결측치, 이상치의 특징들을 근거로 정형/비정형/반정형 등의 데이터 전처리 및 정제 작업을 자동화
- 가비지 인 가비니아아웃의 원리가 적용되는 데이터 생태계에서 최적의 데이터를 활용할 수 있는 토대 마련이 필요
  - 표본의 수가 적어 버려지는 데이터와 결측치나 이상치 처리에 대한 객관적인 근거를 찾기 어려운 경우 원천데이터에서부터 정제, 가공, 개념데이터 셋에 이르기 까지 분석에 활용될 수 있는 수준의 품질을 보장하는 데이터가 필요함
- 데이터 편향을 제거하고 데이터 가치를 향상시키기 위해 최적의 데이터를 생산할 수 있는 SI기술 결합과 활용이 필요
  - 이를 위해 기존의 다양한 분야의 데이터 유형에서 자주 발생하는 오류, 결측치, 이상치의 특징들을 SI가 비교하고 학습함으로써 분석데이터로서의 품질수준을 보장하도록 분류하고 보정하고 재생값을 추천하여 사용자가 보정값 및 결측 및 이상치를 선택할 수 있는 지원도구가 필요



- 기존의 데이터품질관리 도구는 실제 데이터 값의 패턴과 유형의 연속성과 지속성, 유용성 등을 보장하는 반면, 결측치나 이상치 등 데이터 근본의 가치를 고려하지 않고 있음
  - 데이터를 수집하고 저장하여 통합하는데 필요한 표준화 기술의 활용은 활발하게 진행되어 왔음
  - 그럼에도 불구하고 데이터로서의 근본적인 가치를 관리하는 기술은 아직 낮은 수준에 머물러 있는 것이 현실
  - 빅데이터의 가치와 AI 기술의 발달로 데이터를 단순히 표준화하고 관리하는 직업에서 데이터를 학습하고 최적의 결측치와 이상치에 대한 보정값을 추천해줄 수 있는 기술적 기반이 다져짐
  - 이종간 데이터의 유형과 특징에 대한 상관성을 토대로 데이터를 추론 예측하는 방식은 고안되어 있음(예를들어 비정형지수와 정형지수의 상관성 패턴을 통해 상호간 결측 및 이상치를 탐지하는 방법)
  
- 통합데이터를 위한 수집, 저장된 데이터의 표준화 과정을 지원하는 데이터품질 솔루션을 보완하여, 데이터 원천의 근본적 가치의 문제를 해결하는데 활용가치가 매우 큼
  - 기술적 파급효과
    - 메타데이터 관리시스템으로 한정하던 기술적 영역에서 데이터 표준화와 데이터 통합이후 활용가치 극대화
    - 데이터 표준화 및 다양한 기관 및 부처에서 수집되는 융복합 데이터의 거버넌스 극대화
  - 경제적 파급효과
    - 버려지는 데이터의 보정과 재생을 통해 데이터의 가치를 지속적으로 유지하면서 빅데이터 분석의 효율성과 효과성을 달성할 수 있을 것으로 기대
    - 데이터품질관리에 있어서 사람이 확인하고 판단해야 하는 문제 및 재조사 비용을 절감하여 AI기반의 자동화를 통해 해결함으로써 데이터품질관리 담당자의 인력과 소요시간을 단축시킬 것으로 기대
    - 기관과 부처별로 산재되어 있는 다양한 형태와 유형의 빅데이터의 품질에 대한 가치를 제고하고 보정 및 재생하는 것으로 타 산업분야로의 활용 확산에 기여할 수 있을 것으로 기대
  - 사회적 파급효과
    - 버려지는 데이터의 가치를 재생산하여 시민사회와 국민에게 도움이 될 수 있는 새로운 가치를 창출하는데 기여하는 효과가 있음
    - 결측값이나 이상치가 있는 데이터에 대해서 가비지 인 가비지 아웃의 데이터의 인식에서 객관적이고 과학적인 보정과 재생산 될 수 있다는 인식 확산에 기여할 수 있으며 이를 통해 빅데이터의 인식과 가치가 극대화 할 수 있을 것으로 기대

## 나. 범위 및 분류

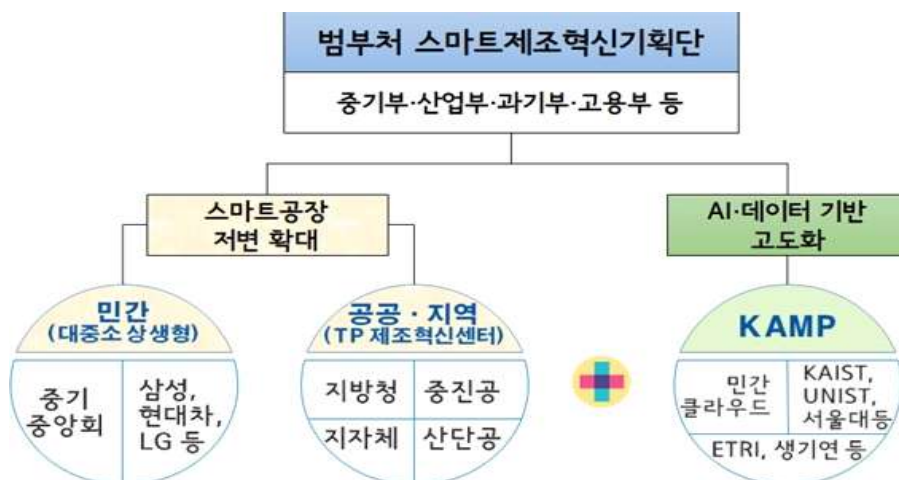
### (1) 가치사슬

- (가치사슬) 빅데이터 분석이 활발해지고 융복합 활용범위가 증가함에 따라 빅데이터를 활용하는 공공/민간의 전 분야에 걸쳐 데이터품질관리 효율 향상 기대
  - (후방산업) 데이터 가치 고도화는 데이터표준을 비롯한 데이터 품질 및 분석 등 데이터과학과 관련된, 모든 기술과 방법론 분야의 종합기술의 성격을 가지고 있으며 이를 어떻게 종합하고 구성하는지가 중요
    - 요소기술들 즉, 표준화관리기술, 데이터통합기술 AI 및 머신러닝, 통계, 분석방법 등에 있어서 존재하고 활용되고 있는 기술들이 존재
    - 기술들을 융합해서 데이터 과학적 특징들을 접목할 수 있는 기술결합 구성에 대해서는 미흡한 수준
  - (전방산업) 공공기관, 민간기업 등 의료, 국방, 행정, 농업 등 데이터를 활용하여 가치를 창출하는 전 산업 전 분야에 걸쳐 전방위적인 산업군을 형성할 수 있음
    - 데이터를 활용하는 곳에서는 항상 데이터 품질과 가치에 대한 문제를 가지고 있으므로 이에 대한 문제를 해결해야 하는 모든 산업군에서 활용될 수 있는 기술임

[ AI 기반 데이터품질 및 가치 고도화 산업구조 ]

후방산업	데이터품질 및 가치 고도화 분야	전방산업
AI 기반 머신러닝 등 분석기술, DBMS 등 데이터관리 기술, 데이터컨설팅, 통계자료 관리기술, 마이그레이션 기술, 데이터표준화기술	AI 기반의 정형/비정형 등 빅 데이터 형태와 유형 수집 및 학습, 결측치 및 이상치 등에 대한 보정 학습 등 정규화와 표준화, 데이터 다중대체 추천 및 재생	빅데이터 융합분석플랫폼, 데이터 축적 및 데이터 가치 표준화를 활용하는, 연구기관, 학교, 공공 및 민간 기업 등 전 분야

[ AI 제조 플랫폼 구조도 ]



\* 출처: 정부 인공지능(AI), 데이터 기반의 제조혁신 전략. 2020.07.23.

## (2) 용도별 분류

□ AI기반 데이터 가치 고도화 플랫폼은 데이터품질 및 가치 관리와 융복합 분석분야로 구분

- (데이터 가치 관리) 각 기관 및 분야별 데이터의 유형과 특성에 따른 가치 고도화
  - 다양한 유형의 데이터를 수집하고 통합하여 관리해야하는 데이터 가치의 표준화
  - 결측치, 이상치, 시계열상의 특이치 등 데이터 가치를 저해하는 모든 상황 파악
  - 데이터 가치를 회복하기 위한 조사 및 탐색적 데이터 분석(EDA)에 소모되는 시간을 절약
  - 데이터표준화와 함께 데이터가치에 대한 고도화를 통해 빅데이터 등 데이터 품질평가 지표로 활용
  - 즉각적인 데이터 분석 결과를 요구하는 플랫폼 및 시스템에서 데이터 가치를 자동으로 보정하고 재생해줌으로써 효율적인 분석 환경을 구성하는데 활용
- (데이터 융복합 분석) 정형/비정형 등 빅데이터의 즉각적인 분석이 필요하도록 데이터 가치 보정 및 재생으로 분석의 객관성 및 효율성 확보
  - 데이터 유형이 데이터 조사 및 수집의 상황별 학습을 통해 데이터 가치 품질을 고도화하는 방안을 추천
  - 데이터의 가치를 저해하는 상황과 유형을 사용자가 직접 파악하고 관리하여 결측치, 이상치 등을 보정하고 재생해야하는 문제를 자동화하여 데이터 가치관리 업무의 효율성을 확보
  - 이상치, 결측치, 특이치 등의 데이터 가치 보정과 재생을 추천하여 분야별 사용자의 특성에 따른 맞춤형 데이터 가치품질 확보
  - 공공 및 민간 등 빅데이터를 활용한 업무지원도구로 플랫폼화가 빠르게 진행됨에 따라 즉각적이고 객관적인 데이터 분석에 활용될 데이터 가치품질 확보에 활용
  - 연구자들의 연구결과에 대한 객관성 확보는 데이터품질 및 가치에 근거하므로 연구기관의 데이터 수집 및 분석 서버에서 자동으로 구현되도록 활용

### [ 데이터 융복합 분석의 예시 ]



\* 출처: 금융위원회, 2020.03.30

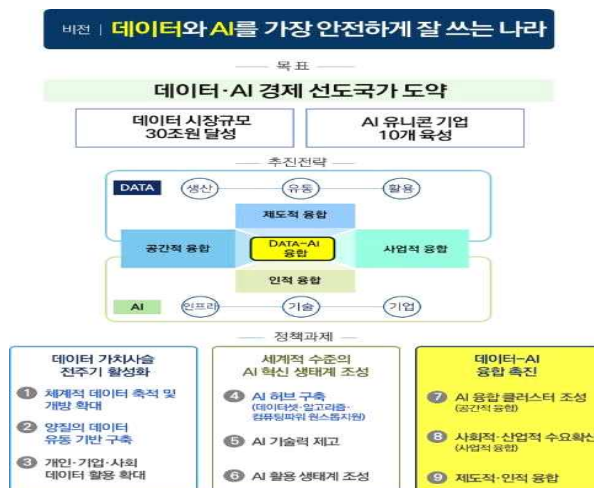
## 2. 산업 및 시장 분석

### 가. 산업 분석

◎ 국가 주도의 데이터 기반 산업정책으로 데이터 가치에 대한 인식과 정책지원 가속화

- 데이터경제를 달성하기 위한 데이터의 수집, 저장, 유통, 활용의 가치사슬을 기반으로 가치창출 체계를 마련하여 경제적 가치를 창출하는 제도적 지원
  - 정부는 2017년 관계부처 합동으로 혁신성장을 위한 사람중심의 4차 산업혁명 대응계획 (2017.11.30)을 발표
    - 데이터 기반의 산업생태계 조성을 목표로 금융, 교통, 의료, 제조, 농업 등 산업 전반에 걸친 데이터 기반 강화 정책 수립
    - 빅데이터 전문센터 육성을 통한 공공빅데이터 센터 설치로 민간 및 공공의 데이터를 구축하고 이를 AI 학습용 데이터 형태로 개방 정책 수립
  - 정부는 각 부처별로 데이터 관련 정책을 수립하고 체계적인 데이터 활용에 대한 정책지원을 가속화
    - (과기정통부) 빅데이터 플랫폼 및 센터 구축('19년 743억원)과 데이터 바우처 사업('19년 600억원), 그리고 국가연구데이터 플랫폼 구축을 통한 양질의 데이터 유통 등 데이터 거래기반 조성으로 전산업분야에서의 데이터 표준 및 가치 고도화에 대한 지원
    - (부처별 데이터 관련 정책) 산업통상자원부, 국방부, 행정안전부, 국토교통부, 통계청, 금융위원회 등에서는 빅데이터 환경구축을 통한 국민실생활과 서비스에 활용될 수 있는 데이터 생태계 구축 지원을 확대하고 있음
    - 혁신성장 전략투자 방향(기획재정부, 2018.08.13.), 데이터 경제활성화 계획(관계부처합동, 2018.08.31.), 데이터·인공지능 경제활성화의 이정표 제시(과학기술정보통신부, 2019.01.16.) 등 활발한 정책지원 계획을 수립
  - 데이터의 생산, 데이터 정제 가공을 통한 데이터 정보유통, 기업, 정부, 공공기관이 활용하여 소비자 및 시민들의 데이터 기반 혁신서비스를 달성하는 가치창출 지원

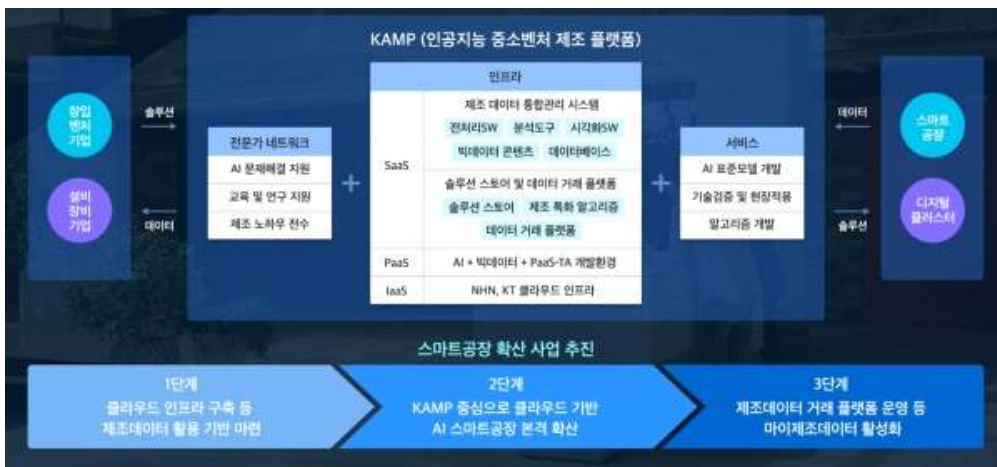
[ 대한민국 정책브리핑 ]



\*출처: 대한민국 정책브리핑. 2020.03.13.

- 데이터3법 개정 등 데이터 관련 제도 및 법령을 정비하여 데이터의 가치를 극대화 하기 위한 산업분야의 확산을 주도
  - 증가하는 빅데이터 활용 가치와 데이터 경제 기반의 정책지원 확대로 양질의 데이터 수집, 가공, 유통의 데이터생태계에 따른 데이터 품질 및 가치의 제도적 보완 확대
    - 혁신성장 전략투자 방향(기획재정부, 2018.08.13.), 데이터 경제활성화 계획(관계부처합동, 2018.08.31.), 데이터-인공지능 경제활성화의 이정표 제시(과학기술정보통신부, 2019.01.16.) 등 활발한 정책지원 계획을 수립
  - 정보통신망 이용촉진 및 정보보호 등에 관한 법률(법률 제17358호, 2020. 6. 9., 일부개정)을 통해 데이터 처리가 필수인 데이터경제기반의 데이터 산업생태계에서 개인정보보호 및 이전과 조치에 대한 데이터 표준처리에 대한 제도적 활용 가능성 확대
  
- 데이터의 생산, 수집, 처리, 분석, 유통, 활용 등 부가적인 가치를 창출하여 상품과 서비스를 생산 제공하는 데이터관련 산업과 업무에 활용
  - 데이터 가치 창출을 위해 데이터 솔루션, DBMS, 데이터 구축 및 컨설팅, 데이터 서비스 등 활용
  - 데이터경제 가치구조에서 양질의 데이터를 활용 상품과 서비스를 생산하는 전 산업분야에서의 데이터 수집, 데이터 분석, 데이터 관리 및 플랫폼에 활용
    - 데이터 아키텍트, 데이터 개발자 및 엔지니어, 데이터분석가, 데이터베이스관리자, 데이터컨설턴트, 데이터기획자들의 업무에 활용
  
- 민간분야에서는 비즈니스인텔리전스, 마케팅인텔리전스를 비롯한 제고 및 마케팅 관리 지원도구와 플랫폼의 활용이 증가하고 있으며, 공공분야에서는 정책 및 업무지원 플랫폼의 활용이 증가하고 있음
  - 민간과 공공분야 모두 다양한 데이터를 수집하고 가공 및 분석하여 업무 및 정책의 의사결정을 돕는 지원도구로서의 플랫폼화가 가속되고 있음
  - 데이터 기반의 플랫폼을 구성하기 위해서는 데이터 품질 및 가치의 중요성이 커짐에 따라 데이터의 수집, 가공, 분석과정에서의 데이터 가치 손실을 최소화하는 기술적 요구가 커짐

[ 민간분야 데이터 기반 플랫폼 구축 및 운영 계획 예시 ]



\* 출처: NHN - KAMP(Korea AI Manufacturing Platform), 2020

◎ 데이터품질 및 가치고도화 기술력 증대를 위한 연구 및 개발 추진

- 데이터통합 및 품질관리에 있어서 데이터품질관리 솔루션에 데이터 표준화에만 한정하고 있는 실정임
  - 데이터기반의 표준화 및 품질관리 도구는 상용화 되어 있는 솔루션이 시장에 많이 진출되어 있는 상황임
  - 그러나 데이터 품질관리 도구는 데이터의 이력추적, 연계, 비식별처리, 데이터모니터링, 데이터포맷, 컬럼명, 프로파일진단 만을 수행하여 가용데이터에 대해서만 표준화를 진행
  - 이상치, 특이치, 결측치에 대해서는 표준화 및 품질관리 대상에서 제외됨, 하지만 데이터로서의 가치는 이들 데이터를 보정하거나 재생하여 사용가능하도록 만들어야 함
  
- 데이터품질과 표준화 중심에서 데이터의 가치를 보정하고 재생하여 데이터로서의 활용을 가능하게 만들어주는 기술개발이 모듈별로 이루어지고 있음
  - 특히 데이터 가치를 보정하고 재생하는 핵심적인 EDA(Exploratory Data Analysis: 탐색적자료분석)방법론과 AI 및 머신러닝을 활용한 검증 방법을 중심으로 연구가 이루어져 오고 있음
  - 기존의 통계데이터 탐색적 분석인 왜/첨도 분석, 시계열분석, 분포확인 등의 기술적 부분에서, 특이치 분해, 이상치 탐지, 데이터 최적화, 결측치 대체 방법 등의 데이터마이닝 기술 부분으로 확대되어 가고 있음
  - 이상치, 특이치 등을 제외하는 방법으로는 데이터 본연의 의미와 가치를 찾아 분석의 품질을 높이는데 활용되기 어려우므로 값을 변경시키거나 가중치를 조정하는 등의 방법을 통한 영향력 감소와 추정하는 기술적 방법들이 개발되어 오고 있음
  
- 기초과학분야와 응용과학기술분야간의 융합적 모델이 만들어질 수 있는 정부의 정책적 지원이 필요함
  - 상대적으로 방법론개발 보다는 기술개발에 정부정책이 치우쳐 있으며, 방법론의 특성 때문에 기초과학분야와의 융합적 기술개발에 대한 정책적 지원이 필요함
  - 데이터 가치 고도화 플랫폼은 데이터 탐색과 보정 및 재생이라는 원천적 방법론에 기반한 기술이므로 이에 대한 각 분야별 융합환경이 필수임
  
- 다양한 형태와 특징을 갖는 민간 및 공공에서 생성되는 가공되지 않는 데이터의 수집과 오픈이 활성화 되어야 함
  - 데이터 본연의 의미와 가치를 고도화하기 위해서는 다양한 분야에서 활용되고 생성되는 데이터들의 수집과 특징을 찾는 것이 중요
  - 분야와 형태에 따라 달라지는 데이터 융합보정 및 재생에 활용될 데이터 셋(학습셋)을 구현하기 위한 민간 및 공공의 강력한 데이터 오픈 정책이 필요함



◎ AI기반의 데이터 활용에 대한 기대 및 관심 증대

- 데이터 가치 고도화는 데이터의 과학적 지식과 분석적 가치에 대한 인식이 중요하게 작용함
  - 전통적으로 데이터에 대한 기본적인 정보를 파악하는 것에서 데이터에 대한 특징을 있는 그대로 확인하려는 목적을 통해 분석적 가치를 찾는 것임
  - 따라서 통계뿐만 아니라 AI 및 머신러닝 기반의 데이터 분석기법의 활용성에 대한 관심이 증가
- 다양한 분야에서 데이터의 가치를 고도화시키기 위한 방법론적 시도들이 이루어지고 있음
  - 연구나 분석 초기 데이터의 현황과 특성을 파악하여 분석에 적합한 데이터로서의 가치를 가지는지를 확인하는 도구로 통계적 방법론에 기반한 탐색적 데이터 분석을 시도
  - 비정형 데이터에 대해서는 데이터의 질적 특성으로 통계적 방법론이 적합하지 않은 문제를 해결하기 위해 AI 및 머신러닝 기반의 데이터 탐지, 보정 등 최적화 방안이 고려되고 있음

[ 데이터 산업 SWOT 분석 ]

	긍정적	부정적
내부	<p style="text-align: center;">&lt; 강점(Strengths) &gt;</p> <ul style="list-style-type: none"> <li>• 정부주도의 데이터 관련 정책 지원</li> <li>• 글로벌 수준의 IT 경쟁력</li> <li>• 데이터기반 기술과 서비스 융합 기회 증대</li> <li>• 데이터품질 관련 솔루션 개발이 활발함</li> <li>• 데이터가치고도화 방법론에 대한 연구시도 확대</li> </ul>	<p style="text-align: center;">&lt; 약점(Weakness) &gt;</p> <ul style="list-style-type: none"> <li>• 방법론적 기술 지원의 한계</li> <li>• 데이터 품질관리도구와의 차별적 한계</li> <li>• 가설검증 등 데이터 분석에 매몰</li> <li>• 민간 데이터개방 미흡</li> <li>• 데이터 융합적 관점의 부재</li> </ul>
외부	<p style="text-align: center;">&lt; 기회(Opportunities) &gt;</p> <ul style="list-style-type: none"> <li>• 세계 데이터산업 시장의 성장</li> <li>• 지속적인 데이터 기반 서비스 산업수요</li> <li>• 데이터품질 및 가치에 대한 꾸준한 요구</li> <li>• AI 및 머신러닝 활용에 따른 경쟁력 증가</li> </ul>	<p style="text-align: center;">&lt; 위협(Threats) &gt;</p> <ul style="list-style-type: none"> <li>• 기술표준 및 원천방법론 등 선진국 주도</li> <li>• 데이터가치 표준에 대한 부재</li> <li>• 데이터 가치 검증에 대한 한계</li> <li>• 분야별 데이터 특성 및 특징에 대한 이해 부족</li> </ul>

## 나. 시장 분석

### (1) 세계시장

- industryARC가 2020년 발표한 세계 데이터 가치 고도화 플랫폼 시장규모는 '18년 11억 9,500만 달러 규모로 평가되었으며 연평균 12.7%로 성장하여 '24년 24억 4,900만 달러 규모로 성장이 전망됨

[ 세계 데이터 가치 고도화 플랫폼 시장규모 및 전망 ]

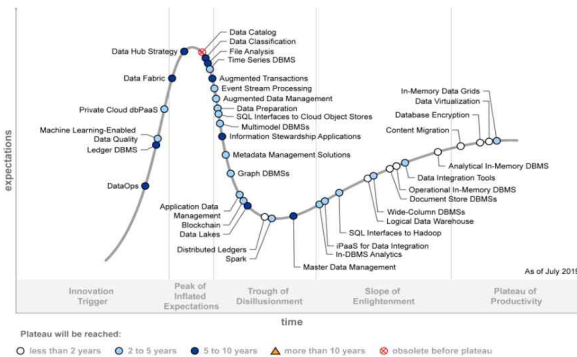
(단위 : 십억 달러, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
세계시장	1,195	1,347	1,518	1,711	1,928	2,173	2,449	12.7

\* 출처: industryARC, 2020을 바탕으로 네모아이씨지 재추경

[ 데이터관리 사이클 변화 및 데이터 품질도구 시장 ]

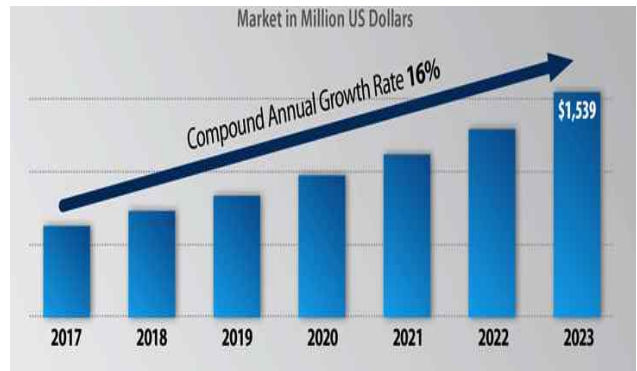
[ 데이터관리 사이클 변화 ]



\* 출처: Gartner, 2019. & Datamation, 2020.

[ 데이터 품질도구 시장 ]

(단위 : 백만 달러)



- 데이터기반의 정보서비스 시장은 분야별로 안정적인 성장세를 보이고 있으며 2021년까지 연평균 약 4.6%수준으로 전망하고 있어 향후에도 성장세는 지속할 것으로 보임

[ 글로벌 정보서비스 부문별 시장규모 ]

Science, Technical & Medical	\$37.7	4.6%
Health Information & Health IT	\$96.7	6.9%
Geophysical & Geomapping	\$14.1	4.5%
Education, Training & Human Capital Management	\$159.0	6.1%
Consumer Entertainment	\$711.7	4.7%
Content Technology & Distribution Services	\$88.1	7.5%
Total	\$1,752.8	4.6%

\* 출처: 2019데이터산업백서, 한국데이터산업진흥원, 2019. 재인용.



## (2) 국내시장

- 국내 데이터 가치 산업 시장규모는 2018년에 15조 1,545억 원 규모로 평가되었으며, 연평균 12.4%로 성장하여 2024년까지 30조 5,406억 원에 이를 것으로 전망됨
  - 국내 데이터산업은 데이터솔루션 시장으로 구분되며 DBMS, 데이터관리, 데이터플랫폼, 데이터분석, 데이터 수집, 데이터 보안으로 구분됨

### [ 국내 데이터 가치 산업 시장규모 및 전망 ]

(단위 : 억 원, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
국내시장	151,545	170,234	191,343	215,069	241,738	271,714	305,406	12.4

\* 출처: Marketsandmarkets, industryARC(2020)과 2019데이터산업백서(2019), 한국데이터산업진흥원.를 바탕으로 세계 빅데이터 시장과 국내 빅데이터 시장을 복합적으로 고려해 네모아이씨지 재추정

- 데이터기반의 거래, 정보제공, 분석제공에 대한 수요와 기술개발이 활발하게 진행됨에 따라 데이터 품질 및 고도화 방법론에 대한 요구가 지속적으로 증가될 것으로 보임
  - 데이터기반 서비스는 크게 데이터를 직접 판매하거나 중개하는 데이터 거래서비스, 데이터나 DB를 기반으로 정보를 제공하는 정보제공 서비스, 데이터를 분석해 제공하는 분석제공 서비스로 구분할 수 있음
  - 이 중 정보제공 서비스가 89.5%를 차지하는 가장 큰 시장을 구성하고 있음

### [ 국내 데이터산업 시장규모 현황 ]

(단위 : 억 원, %)

분류	시장규모						증감률		
	2017년		2018년		2019년		전체	직접매출	
	전체	직접매출	전체	직접매출	전체	직접매출			
데이터서비스	데이터거래	2,918	2,713	4,175	3,796	4,315	3,915	43.1%	39.9%
	정보제공	61,570	12,747	67,580	22,673	74,664	24,867	9.8%	77.9%
	데이터 분석제공	3,690	2,878	4,023	3,633	4,382	3,932	9.0%	26.3%

\* 출처: 2019데이터산업백서(2019), 한국데이터산업진흥원.

### 3. 기술 개발 동향

- 기술경쟁력
  - AI기반 데이터 가치 고도화 플랫폼은 미국이 최고기술국으로 평가되었으며, 우리나라는 최고기술국 대비 66.4%의 기술 수준을 보유하고 있으며, 최고기술국과의 기술격차는 2.7년으로 분석
  - 중소기업의 기술경쟁력은 최고기술국 대비 59.7%, 기술격차는 3.2년으로 평가
  - EU(78.5%)>중국(73.8%)>일본(73.7%)>한국의 순으로 평가
- 기술수명주기(TCT)<sup>8)</sup>
  - AI기반 데이터 가치 고도화 플랫폼은 6.35의 기술수명주기를 지닌 것으로 파악

#### 가. 기술개발 이슈

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

- 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

8) 기술수명주기(TCT, Technical Cycle Time): 특허 출원연도와 인용한 특허들의 출원연도 차이의 중앙값을 통해 기술 변화속도 및 기술의 경제적 수명을 예측

- 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 SI기반의 머신러닝 과정에 필요
- 데이터 유형과 형태 그리고 생성과정을 SI와 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

- 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: 경하영(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 ]

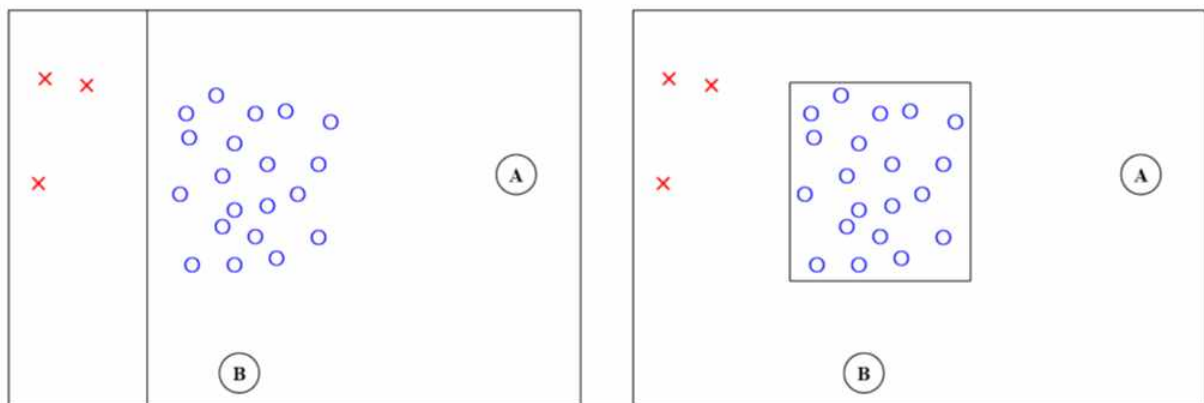
구분	부분	상세	
추정프로세스	시간데이터	시계열 분석	ADF 단위근 검정
		이상치 탐지	박스 플롯
	텍스트데이터	n-gram 벡터 생성	형태소 분석(MPH)
		유사성 분석	레벤슈타인 거리(LD)
임베딩	시각화	다차원척도법(MDS)	

\* 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 AI기반의 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

◎ 핵심 원천기술 및 국가경쟁력의 확보가 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

[ 2018~2022년 국내 SW분야 인력 부족 숫자 ]

(단위 : 명)

구분	AI	클라우드	빅데이터	AR-VR
인력 부족 수	9,986	335	2,785	18,727

\* 출처: 소프트웨어정책연구소, 유망 소프트웨어 분야의 미래 일자리 전망, 2019

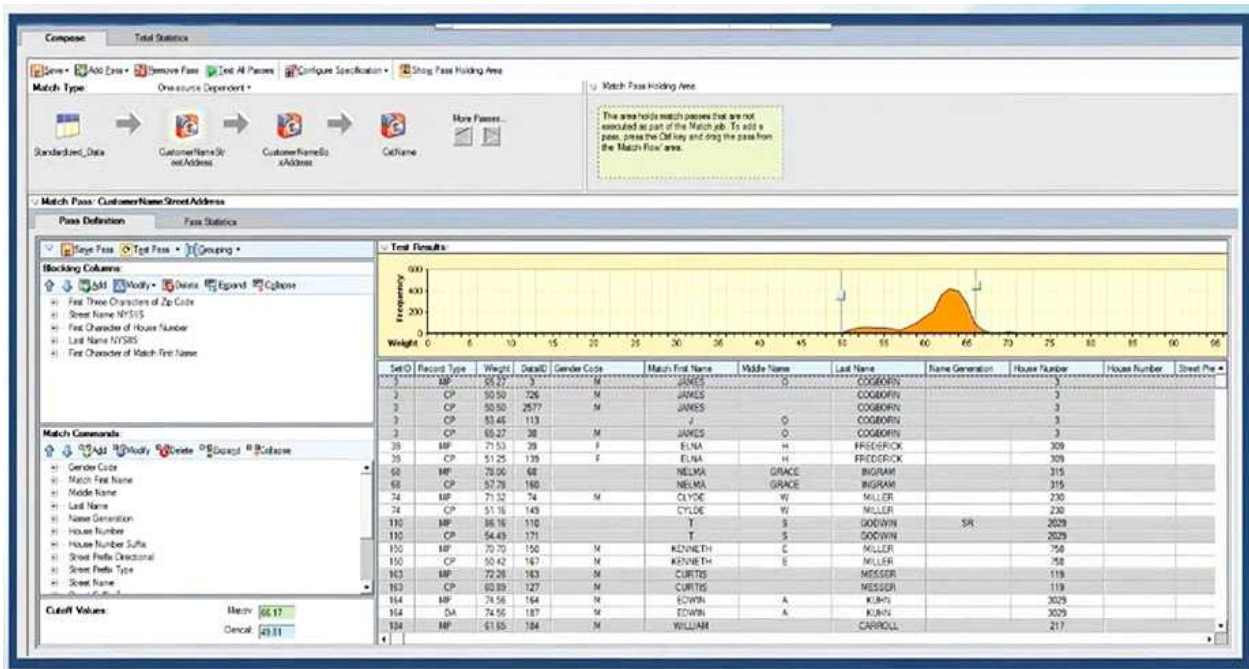
## 나. 생태계 기술 동향

### (1) 해외 플레이어 동향

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

- 데이터와 그 관계를 이해하는데 도움이되는 종단 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

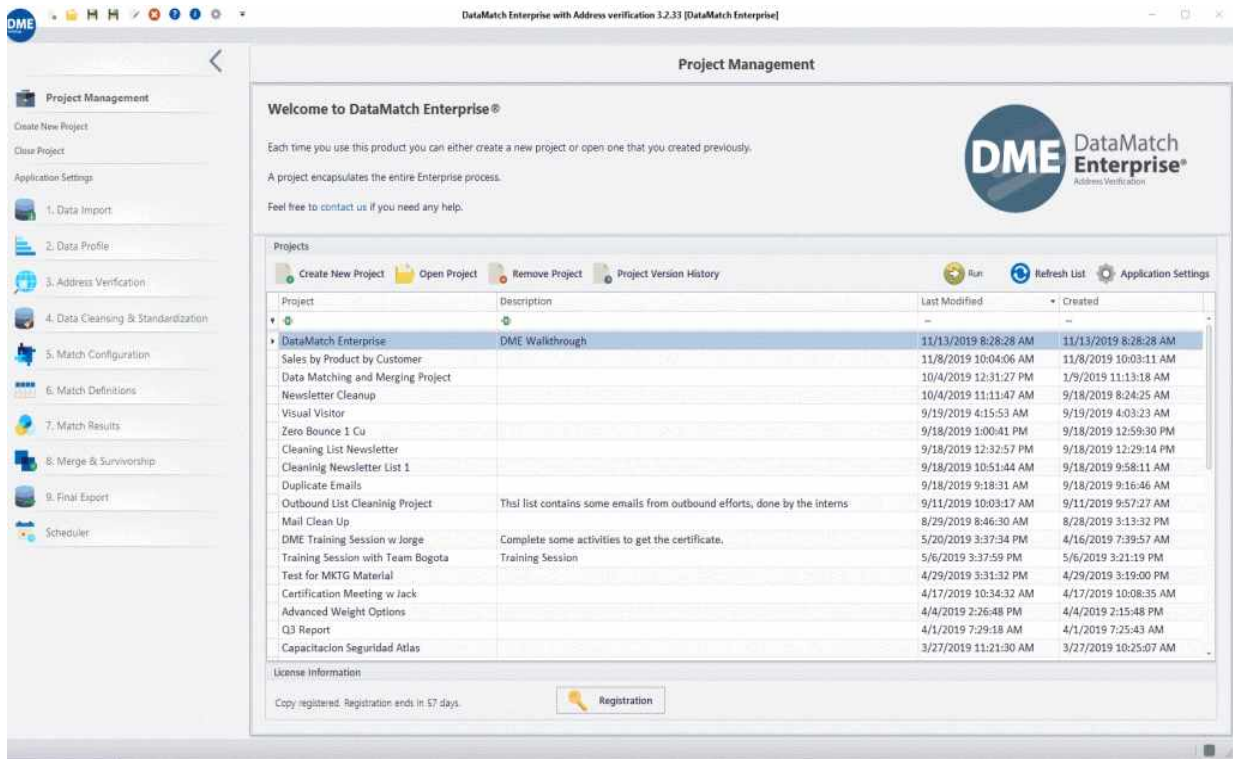
- 조직내 연구개발, 제조, 고객서비스, 규제, 공급업체에 대한 각기 다른 데이터를 취급하는 조직들의 맞춤형 품질관리 도구를 개발
- 조직내 연구개발, 제조, 고객서비스, 규제, 공급업체에 대한 각기 다른 데이터를 취급하는 조직내 생산된 데이터의 중복방지과 중복제거 및 표준화, 중복레코드 식별, 이상치 탐지 및 유사한 유형의 데이터로 여러 클라이언트를 처리할 수 있는 시스템 통합자 및 컨설턴트가 사용할 수 있는 데이터 인터페이스 사용



□ (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



## (2) 국내 플레이어 동향

- (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

## 다. 국내 연구개발 기관 및 동향

### (1) 연구개발 기관

[ 데이터 가치 고도화 분야 주요 연구조직 현황 ]

기관	소속	연구분야
한국데이터산업진흥원	유통기반실 유통지원팀 데이터자격검정센터	<ul style="list-style-type: none"> <li>• 데이터 유통 관련 표준화지원</li> <li>• 데이터 품질 시험체계구축 및 데이터 품질인증, 데이터 품질 컨설팅</li> </ul>
한국과학기술원	데이터베이스연구소 첨단 데이터베이스 엔진 및 응용연구실	<ul style="list-style-type: none"> <li>• 라이프로그 관리</li> <li>• 그래프데이터 관리 및 마이닝</li> <li>• 무선센서 네트워크에서 쿼리처리</li> <li>• 데이터 스트림 처리 및 시멘틱 웹데이터 처리</li> </ul>
충북대학교 산학협력단	-	<ul style="list-style-type: none"> <li>• 데이터 연관성 기반 이상치 평가 방법</li> <li>• 이력 데이터의 이상치 평가방법</li> </ul>
삼성에스디에스	-	<ul style="list-style-type: none"> <li>• 데이터의 결측치 보정방법</li> </ul>
한국전자통신연구원	인공지능연구소 KSB융합연구단	<ul style="list-style-type: none"> <li>• 시계열 데이터 처리</li> <li>• 결측치 대체 학습모델 생성</li> </ul>

### (2) 기관 기술개발 동향

- 한국데이터산업진흥원
  - 데이터 품질에 대한 전반적인 관리의 표준모델을 개발하고 규격화 하여 평가 및 검증 방법개발
- 한국전자통신연구원
  - 시계열 데이터의 특성에 따른 처리 및 방법 및 결측치를 포함하는 데이터에 학습모델을 적용하여 결측치를 대체하는 방법 개발
- 한국과학기술원
  - 이상속성을 포함하는 데이터를 연관패턴을 통해 이상속성을 검출하는 방법 개발
- 충북대학교 산학협력단
  - 테이블 형태로 저장되는 데이터 연관성이 있는 속성의 분포정보를 기반으로 이상치를 평가하는 방법 개발

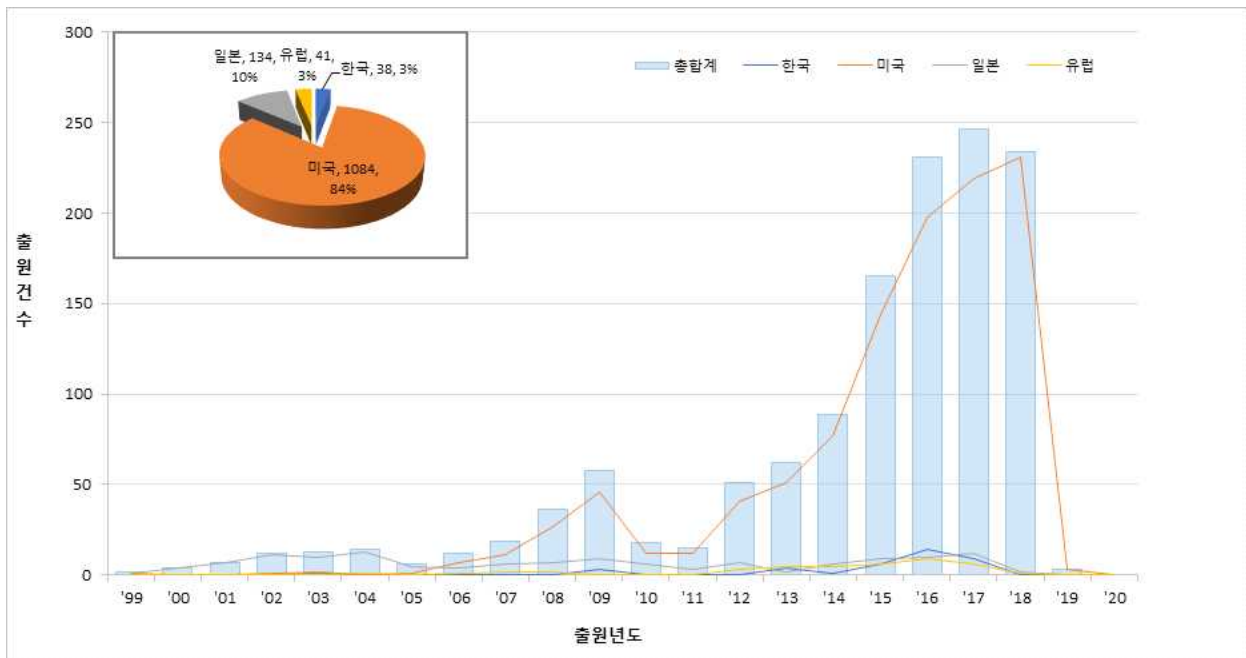
## 4. 특허 동향

### 가. 특허동향 분석

#### (1) 연도별 출원동향

- 시 기반 데이터 가치 고도화 플랫폼은 '12년부터 급격한 성장을 보임
  - 각 국가별로 살펴보면 미국이 가장 활발한 출원활동을 보이고 있음
- 국가별 출원비중을 살펴보면 미국이 전체의 84%의 출원 비중을 차지하고 있어, 최대 출원국으로 시 기반 데이터 가치 고도화 플랫폼 분야를 리드하고 있는 것으로 나타났으며, 일본은 10%, 유럽은 3%, 한국은 3% 순으로 나타남

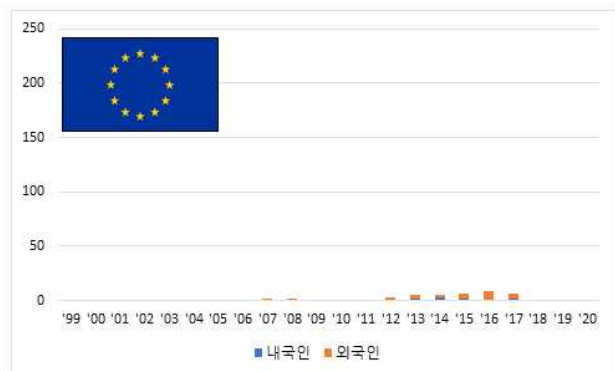
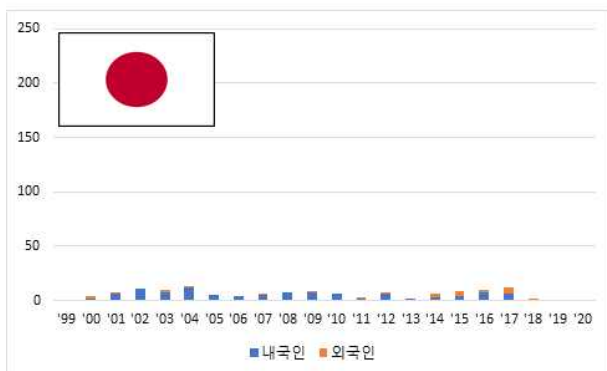
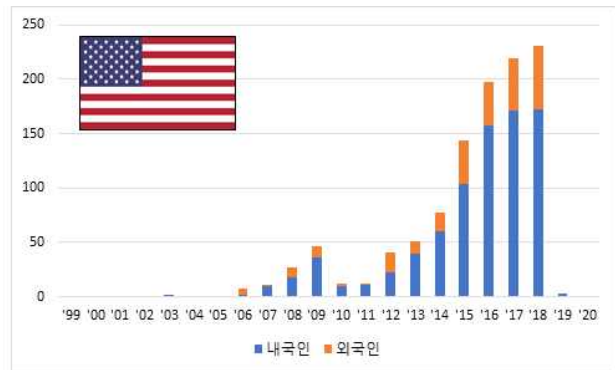
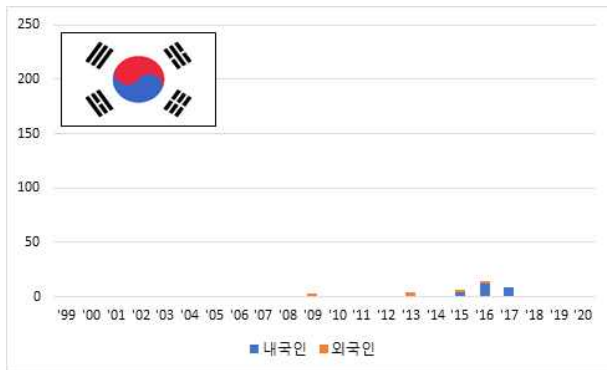
[ 시 기반 데이터 가치 고도화 플랫폼 연도별 출원동향 ]



## (2) 국가별 출원현황

- 한국의 출원현황을 살펴보면, '15년부터 해당 기술의 출원이 소폭 증가하는 추세
  - 내국인 위주의 출원이 진행되고 있음
  - 한국 기술의 양적 흐름은 미국에 비해 미미한 수준을 보임
  - 미국의 출원 수에 비해 4% 정도의 수준을 보임
- 미국의 출원현황을 살펴보면 분석구간 초기부터 전체 특허기술의 출원 증감 흐름에 영향을 주고 있는 것으로 나타남. 미국의 경우, 한국에 비해 외국인의 비중이 적은 것으로 나타남
- 유럽의 출원현황은 출원수가 매년 20건 이하로, 뚜렷한 증감 동향이 나타나지 않음
- 일본의 출원현황은 출원수가 매년 20건 이하로, 뚜렷한 증감 동향이 나타나지 않음

[ 국가별 출원현황 ]

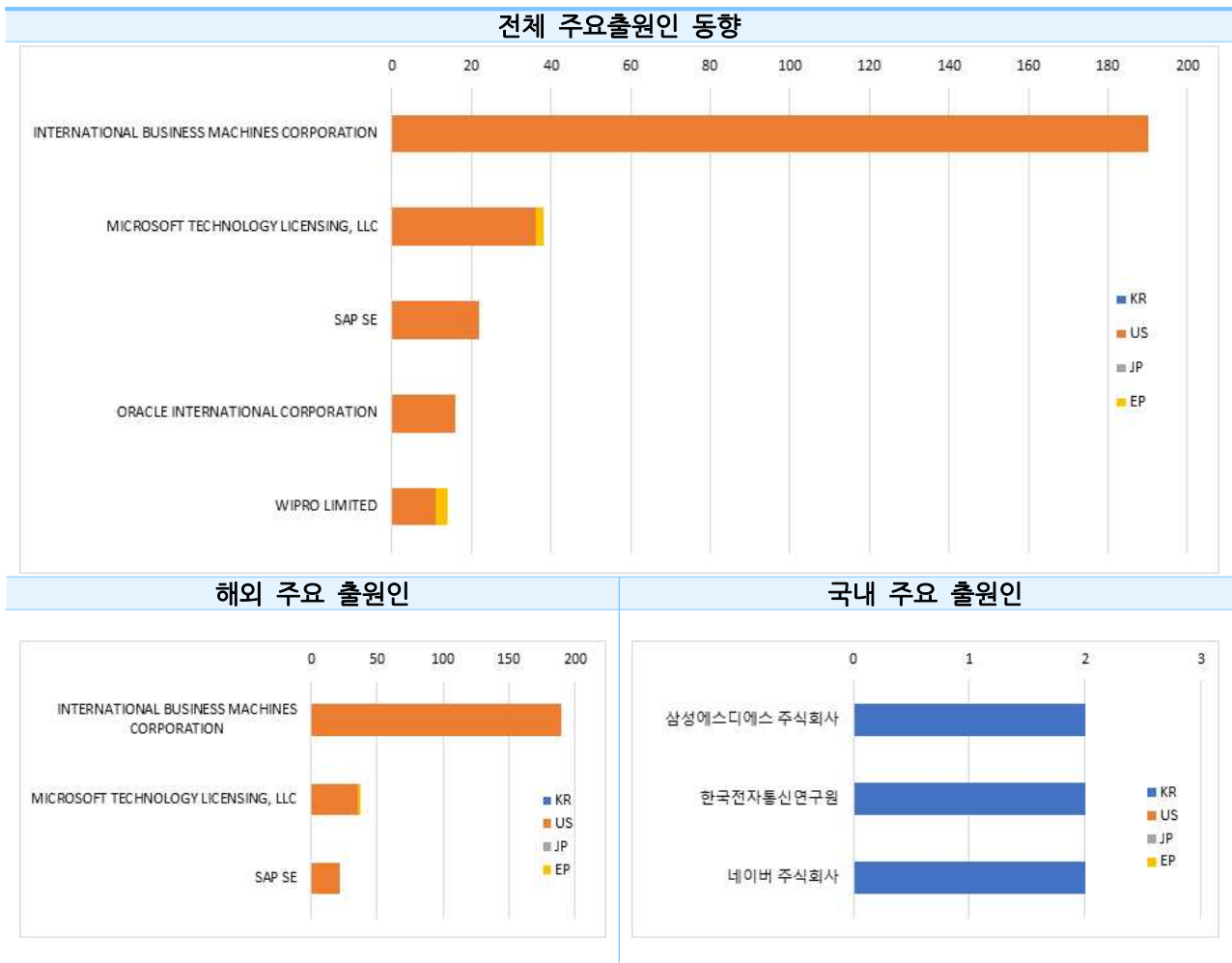




## 나. 주요 출원인 분석

- Si기반 데이터 가치 고도화 플랫폼의 전체 주요출원인을 살펴보면, 주로 미국 국적의 출원인이 다수 포함되어 있는 것으로 나타났으며, 제 1 출원인으로는 미국의 INTERNATIONAL BUSINESS MACHINES CORPORATION인 것으로 나타남
  - 제 1 출원인인 INTERNATIONAL BUSINESS MACHINES CORPORATION의 출원은 미국에 집중된 경향을 보임
- Si기반 데이터 가치 고도화 플랫폼 관련 기술로 빅데이터 분석을 다루는 대기업에 의한 출원이 대다수를 차지
  - 국내에서는 중소기업(개인), 연구기관/대학의 활발한 출원이 이루어짐

[ Si기반 데이터 가치 고도화 플랫폼 주요출원인 ]



(1) 해외 주요출원인 주요 특허 분석

◎ IBM(International Business Machines)

- IBM은 미국의 다국적 기업으로, 시 기반 데이터 가치 고도화 플랫폼과 관련하여, 빅데이터 중에서 에러 데이터를 보정하는 기술에 특화된 특허를 다수 출원. 그 중 등록된 특허는 100건
  - 주요 특허들은 시 기반의 기계 학습에서 에러 데이터를 보정하거나 유효성을 검사하는 기술 특허를 다수 출원하는 것으로 파악

[ IBM 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
US10095774 (2017.12.21)	Cluster evaluation in unsupervised learning of continuous data	데이터 관리자는 적어도 2개의 시간 간격들로부터 연속적인 데이터 사이의 시간적 안정성 척도에 기초하여 적절한 수의 클러스터들을 선택하는 기술	
US10031943 (2017.11.13)	Imputing data for temporal data store joins	두 개의 독립적인 데이터 세트들은 데이터 스토어(data store) 내에 포함되며, 데이터 구조에 대한 연산과 연관된 하나 이상의 널 값(null value)들이 존재하고 보정하는 기술	
US10078659 (2017.09.25)	Semantic database driven form validation	시맨틱 데이터베이스들을 사용하여 월드 와이드 웹으로부터 이용 가능한 데이터에 기초하여 양식 입력을 동적으로 검증하는 기술	
US10068019 (2017.02.15)	Efficient extraction of intelligence from web data	식별한 타겟팅된 인덱스에 관련된 웹 데이터를 포함하는 초기 데이터 유니버스(initial data universe)를 월드 와이드 웹으로부터 끌어당기는 표현 기반 소셜 미디어 모니터링(SMM) 툴 기술	
US10298683 (2017.01.04)	Consolidating data access in a dispersed storage network	에러 인코딩된 데이터를 분산시키기 위해 분산 저장 및 태스크 (DST) 유닛에 의한 실행하는 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출



◎ MICROSOFT TECHNOLOGY LICENSING, LLC

- MICROSOFT TECHNOLOGY LICENSING, LLC는 Microsoft의 자회사로, 회사의 특허 및 기술 이전 관련 기업으로, 다수의 AI기반 빅데이터 분석 관련 특허를 보유하고 있으며, AI기반 데이터 가치 고도화 플랫폼과 관련한 특허를 미국에 출원
  - 머신러닝을 활용한 빅데이터 분석에 관련하여 출원을 진행한 특허 모두 미국에 출원한 것으로, 자국 내 출원 성향이 높은 것으로 사료
  - AI 기반 데이터 가치 고도화 플랫폼 기술과 관련하여 38건의 미국 출원을 진행하였으며, 그 중 등록된 특허는 23건으로 파악

[ MICROSOFT TECHNOLOGY LICENSING, LLC 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
US10521482 (2017.04.24)	Finding members with similar data attributes of a user for recommending new social connections	프로파일 유사성 및 소셜 네트워크 내의 기존 상호접속들에 기초하여 새로운 접속들을 추천하기 위한 기술	
US10535018 (2016.10.31)	Machine learning technique for recommendation of skills in a social networking service based on confidential data	보상 벡터들의 가중 매트릭스는 기계 학습 스킬 통화 가치 예측 모델에 공급된 후보 벡터에 포함된 스킬에 대한 예측된 통화 가치를 출력하기 위해 기계 학습 스킬 통화 가치 예측 모델을 트레이닝 기술	
US10095753 (2016.09.28)	Aggregation and generation of confidential data insights with confidence values	하나 이상의 속성을 갖는 슬라이스에 대해 검색된 제1 기밀 데이터 유형의 복수의 이전에 제출된 기밀 데이터 값을 산출하는 기술	
US10129295 (2016.08.31)	Clustering approach for detecting ddos botnets on the cloud from ipfix data	머신 러닝을 이용하여 분류기를 트레이닝하여, 분산된 서비스 거부 공격의 일부인 엔티티에 대한 신뢰를 증가시키기 위해 엔티티들을 분류하는 기술	
US10157291 (2016.07.28)	Collection flow for confidential data	소셜 네트워킹 서비스의 멤버 프로파일로부터 누락된 속성 값들을 출력하기 위해 머신 러닝 알고리즘에 의해 트레이닝되는 속성 간섭 모델 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출



◎ SAP SE

- SAP SE는 독일의 데이터베이스를 이용하여 통합된 업무용 애플리케이션을 개발하는 기업으로 등록특허 7건 보유
  - 트레이닝된 머신 러닝 모델에서 트레이닝 데이터의 객체를 늘이기 위한 다양한 값을 출력하기 위한 응용 특허를 다수 보유

[ SAP SE 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
US10509782 (2017.12.11)	Machine learning based enrichment of database objects	트레이닝된 머신 러닝 모델에 의해, 데이터베이스 내의 제 1 객체가 데이터베이스 내의 제 2 객체와 동일 또는 유사하다고 결정하는 단계를 포함하는 데이터베이스 내의 객체를 풍부화하는 기술	
US10621496 (2016.12.21)	Management of context data	컨텍스트 프로파일은 애플리케이션 식별자에 대응하고, 컨텍스트 프로파일과 응용 프로그램 식별자의 매핑이 만들어지며, 매핑에 기초하여, 컨텍스트 프로파일은 대응하는 컨텍스트 제공자에게 전송하는 기술	
US10163038 (2015.12.24)	Error detection in recognition data	인식 데이터의 에러 검출을 용이하게 하기 위한 기술적 프레임워크를 사용하며, 검출된 오류 기록 및 결정된 정정을 제시하기 위해 리포트가 생성하는 기술	
US10235450 (2015.12.21)	Semantic layer for processing machine data	엔트리 타입 인식, 엔트리 타입들을 시맨틱 이벤트들에 매핑하는 것 및 시맨틱 이벤트-엔트리 타입 매핑에 기초하여 지식 베이스 내의 엔트리들을 구축하는 기계 데이터 처리 시스템 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

## (2) 국내 주요출원인 주요 특허 분석

### ◎ 삼성에스디에스 주식회사

- 삼성에스디에스 주식회사는 한국의 IT 솔루션 서비스 기업으로, AI 기반 데이터 가치 고도화 플랫폼 기술에 있어서, 데이터 분류기술과 관련된 특허를 2건 출원하였으며, 그 중 등록된 특허는 1건 보유

[ 삼성에스디에스 주식회사 주요특허 리스트 ]

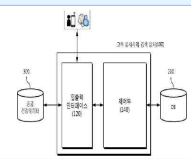
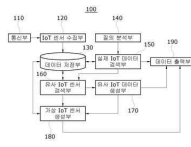
등록번호 (출원일)	명칭	기술적용분야	도면
KR1768438 (2013.10.30)	데이터 분류 장치 및 방법과 이를 이용한 데이터 수집 시스템	데이터 분류 성능을 향상시킬 수 있는 데이터 분류 장치 및 방법과 이를 이용한 데이터 수집 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ 한국전자통신연구원

- 한국전자통신연구원은 한국의 정보, 통신, 전자, 방송 및 성과 관련 융·복합기술 분야를 연구하는 연구 기관으로, '16년도 11월부터 출원을 시작하여 아직까지 등록된 특허는 없는 것으로 파악됨
  - 한국전자통신연구원의 공개특허를 대상으로 분석한 결과, AI기반 데이터 가치 고도화 플랫폼 기술과 관련된 특허를 2건 출원

[ 삼성에스디에스 주식회사 주요특허 리스트 ]

공개번호 (출원일)	명칭	기술적용분야	도면
KR2018-0062353 (2017.11.10)	대용량 다차원 시계열 건강 데이터의 복합 차원으로의 축소를 통한 고속 유사 사례 검색 방법 및 그 장치	하나 이상의 프로세서를 포함하는 장치에 의해 수행되는 다차원 건강 데이터에 대한 유사 사례 검색 방법	
KR2018-0049634 (2016.11.03)	IoT 데이터를 생성하는 장치 및 방법	사용자가 원하는 실제 IoT 데이터를 제공할 수 있는 실제 IoT 센서를 검색하고, 검색된 실제 IoT 센서로부터 사용자가 원하는 실제 IoT 데이터를 수집 및 사용자에게 제공하는 기술	

\* 공개특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ 네이버 주식회사

- 네이버 주식회사는 한국의 인터넷 서비스 기업으로, AI기반 데이터 가치 고도화 플랫폼 기술과 관련된 특허를 2건 출원하였으며, 출원을 진행한 2건의 특허 모두 등록된 것으로 파악

[ 네이버 주식회사 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
KR1768521 (2016.05.02)	이미지에 포함된 객체에 대한 정보 데이터를 제공하는 방법 및 시스템	특정 아이템에 대해 학습된 딥러닝 학습 모델을 이용하여 이미지에 포함된 객체에 대한 정보 데이터를 제공하는 기술	
KR1778679 (2015.10.02)	딥러닝을 이용하여 텍스트 단어 및 기호 시퀀스를 값으로 하는 복수 개의 인자들로 표현된 데이터를 자동으로 분류하는 방법 및 시스템	딥러닝을 이용하여 텍스트 단어 및 기호 시퀀스를 값으로 하는 복수 개의 인자들로 표현된 데이터를 자동으로 분류하는 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

## 다. 기술진입장벽 분석

### (1) 기술 집중력 분석

- AI기반 데이터 가치 고도화 플랫폼 관련 기술에 대한 시장관점의 기술독점 현황분석을 위해 집중률 지수(CRn: Concentration Ratio n, 상위 n개사 특허점유율의 합) 분석 진행
  - 상위 4개 기업의 시장점유율이 0.21로 AI기반 데이터 가치 고도화 플랫폼 분야에 있어서 독과점 정도는 낮은 수준으로 판단
  - 국내 시장에서 중소기업의 점유율 분석결과 0.75로 해당 기술에 대하여 중소기업이 진입하여 활동하고 있는 것으로 파악

[ 주요출원인의 집중력 및 국내시장 중소기업 집중력 분석 ]

주요 출원인 집중력	주요출원인	출원건수	특허점유율	CRn	n
	INTERNATIONAL BUSINESS MACHINES CORPORATION(미국)	190	14.6%	0.15	1
	MICROSOFT TECHNOLOGY LICENSING, LLC(미국)	38	2.9%	0.18	2
	SAP SE(독일)	22	1.7%	0.19	3
	ORACLE INTERNATIONAL CORPORATION(미국)	16	1.2%	<b>0.21</b>	<b>4</b>
	WIPRO LIMITED(인도)	14	1.1%	0.22	5
	SPLUNK INC.(미국)	13	1.0%	0.23	6
	SALESFORCE.COM, INC.(미국)	13	1.0%	0.24	7
	ACCENTURE GLOBAL SOLUTIONS LIMITED(아일랜드)	13	1.0%	0.25	8
	SAS INSTITUTE INC.(미국)	12	0.9%	0.26	9
	MICROSOFT CORPORATION(미국)	12	0.9%	0.26	10
	<b>전체</b>	<b>1297</b>	<b>100%</b>	<b>CR4=0.21</b>	
	국내시장 중소기업 집중력	출원인 구분	출원건수	특허점유율	CRn
중소기업(개인)		36	75.0%	0.75	
대기업		3	6.3%		
연구기관/대학		9	18.8%		
<b>전체</b>		<b>48</b>	<b>100%</b>	<b>CR중소기업=0.75</b>	

(2) 특허소송 현황 분석

- AI기반 데이터 가치 고도화 플랫폼 관련 기술 진입 장벽에 대한 분석을 위해 특허소송을 이력 검토
  - 2015년 2월 캘리포니아 북부지방법원에 원고 Sentius International, LLC와 Microsoft Corporation간의 콘텐츠에 특정 데이터 개체를 연결하는 Microsoft 제품에 대한 특허 침해소송이 종결

[ AI기반 데이터 가치 고도화 플랫폼 관련 특허소송 현황 ]

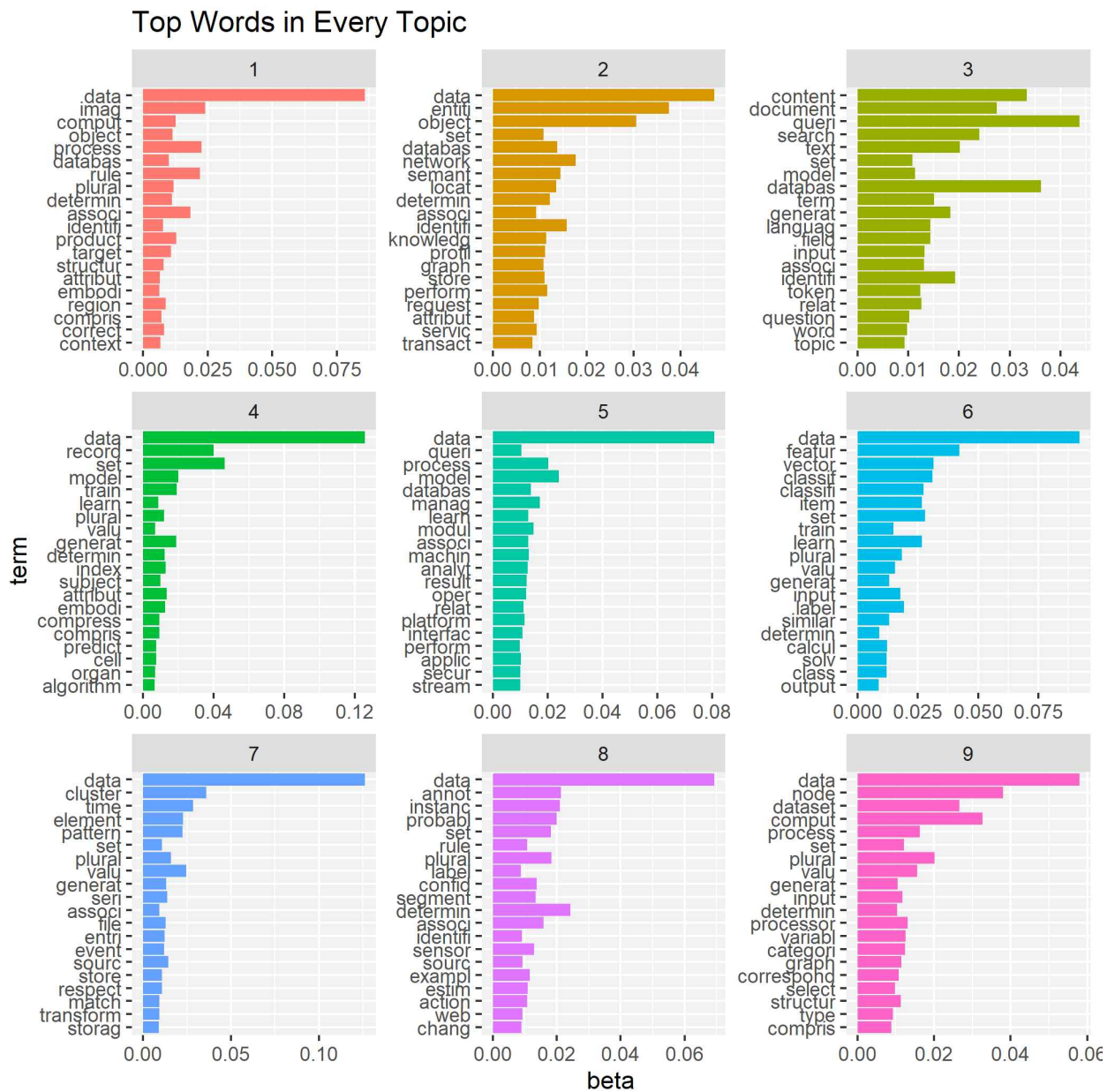
		명칭	출원인	원고 v. 피고
1	US7672985 (2010.03.02)	Automated creation and delivery of database content	Sentius International LLC	Sentius International, LLC v. Microsoft Corporation
		대상제품명	소제기일	소송종료일
		Microsoft Office suite and its component programs such as Microsoft Word that use a "contextual reference engine" to parse content and to associate specific data objects to content for display in pop-up screens	2013.02.22	2015.02.05

## 5. 요소기술 도출

### 가. 특허 기반 토픽 도출

- 1,226개의 특허의 내용을 분석하여 구성 성분이 유사한 것끼리 클러스터링을 시도하여 대표성이 있는 토픽을 도출

[ 시 기반 데이터 가치 고도화 플랫폼에 대한 토픽 클러스터링 결과 ]



## 나. LDA<sup>9)</sup> 클러스터링 기반 요소기술 도출

[ LDA 클러스터링 기반 요소기술 키워드 도출 ]

No.	상위 키워드	대표적 관련 특허	요소기술 후보
클러스터 01	data, image, process, rule, associate, product, computing, plural, object, determine	<ul style="list-style-type: none"> <li>• Method for computer-controlled processing of requests for the analysis of data in a storage medium</li> <li>• Patient administration system comprising intelligent interface device for the transmission of medical data</li> <li>• Method, system and computer program product for automating expertise management using social and enterprise data</li> </ul>	AI 및 머신러닝, 탐색적 데이터 분석 융합 기술
클러스터 02	data, entities, object, network, identification, semantic, database, location, determine, perform	<ul style="list-style-type: none"> <li>• Data unification device and method for unifying unstructured data objects and structured data objects into unified semantic objects</li> </ul>	오류 데이터 식별 및 탐색 기술
클러스터 03	queries, database, content, document, search, text, identification, generate, term, language	<ul style="list-style-type: none"> <li>• System and method for sequencing database-based content recommendation</li> <li>• System and method for automated hybrid sequencing database generation</li> </ul>	데이터 추출 및 변환 기술
클러스터 04	data, merchant, queries, set, pattern, communicate, variable, relate, morpheme	<ul style="list-style-type: none"> <li>• System and method for combining data sets</li> <li>• Operation and method for prediction and management of the validity of subject reported data</li> </ul>	데이터 추출 및 변환 기술
클러스터 05	data, model, process, manage, module, database, machine, learn, associate, analytic	<ul style="list-style-type: none"> <li>• Platform management of integrated access of public and privately-accessible datasets utilizing federated query generation and query schema rewriting optimization</li> <li>• Computer-implemented methods, computer readable medium and systems for a precision agriculture platform that integrates a satellite date model and an orchard data model</li> </ul>	AI 및 머신러닝, 탐색적 데이터 분석 융합 기술
클러스터 06	data, feature, vector, classify, set, classification, learn, item, label, plural	<ul style="list-style-type: none"> <li>• Data classification device and program</li> <li>• Data classification method, data classification program, and data classification device</li> </ul>	데이터 형태 분류기술
클러스터 07	data, cluster, time, value, element, pattern, plural, source, series, generate	<ul style="list-style-type: none"> <li>• Computerized data equidistant-mapping method and computer system using the same</li> <li>• Aggregation and generation of confidential data insights with confidence values</li> </ul>	AI 및 머신러닝, 탐색적 데이터 분석 융합 기술
클러스터 08	data, determine, annotation, instance, probable, plural, set, associate, confidential, segment	<ul style="list-style-type: none"> <li>• Method and system for notifying a user of a terminal of contextual data relating to elements identified in an address book application</li> <li>• Data synchronization architecture</li> </ul>	데이터 추출 및 변환 기술
클러스터 09	data, node, computing, dataset, plural, process, value, processor, variable, categories	<ul style="list-style-type: none"> <li>• Quantitative assessment of similarity of categorized data</li> <li>• Device and method for generating a crime type combination based on historical incident data</li> </ul>	데이터 품질관리기술

9) Latent Dirichlet Allocation



### 다. 특허 분류체계 기반 요소기술 도출

- AI기반 데이터 가치 고도화 플랫폼 관련 특허에서 총 10개의 주요 IPC코드(메인그룹)를 산출하였으며, 각 그룹의 정의를 기반으로 요소기술 키워드를 아래와 같이 도출

[ IPC 분류체계에 기반한 요소기술 도출 ]

IPC 기술트리		
(서브클래스) 내용	(메인그룹) 내용	요소기술 후보
(G06F) 전기에 의한 디지털 데이터처리	• (G06F-017) 디지털 컴퓨팅 또는 데이터 프로세싱 장비, 방법으로서 특정 기능을 위해 특히 적합한 형태의 것	-
	• (G06F-007) 취급하는 데이터의 순서 또는 내용을 조작하여 데이터를 처리하기 위한 방법 또는 장치	데이터 형태 분류기술
	• (G06F-015) 디지털 컴퓨터 일반	-
	• (G06F-019) 특수한 어플리케이션에 특히 적합한 디지털 컴퓨팅 또는 데이터 처리 장치 또는 방법	-
(G06K) 데이터의 인식; 데이터의 표시; 기록매체; 기록매체의 취급	• (G06K-009) 인쇄문자, 손으로 쓴 문자를 독취하거나 인식 또는 패턴을 인식하기 위한 방법 또는 장치, 예. 지문인식	AI 및 머신러닝, 탐색적 데이터 분석 융합 기술
(G06N) 특정 계산모델 방식의 컴퓨터시스템	• (G06N-005) 지식기반모델을 이용한 컴퓨터시스템	-
	• (G06N-099) 이 서브클래스의 다른 그룹으로 분류되지 않는 주제사항	-
	• (G06N-003) 생체모델기반 컴퓨터시스템	-
	• (G06N-007) 특정수리모델에 기반을 둔 컴퓨터시스템	-
(H04L) 디지털 정보의 전송, 예. 전신통신	• (H04L-029) 그룹 1/00에서 H04L 27/00의 하나에도 포함되지 않는 배치, 장치회로 또는 시스템	-

## 라. 최종 요소기술 도출

- 산업·시장 분석, 기술(특허)분석, 전문가 의견, 타부처 로드맵, 중소기업 기술수요를 바탕으로 로드맵 기획을 위하여 요소기술 도출
- 요소기술을 대상으로 전문가를 통해 기술의 범위, 요소기술 간 중복성 등을 조정·검토하여 최종 요소기술명 확정

### [ AI기반 데이터 가치 고도화 플랫폼 분야 요소기술 도출 ]

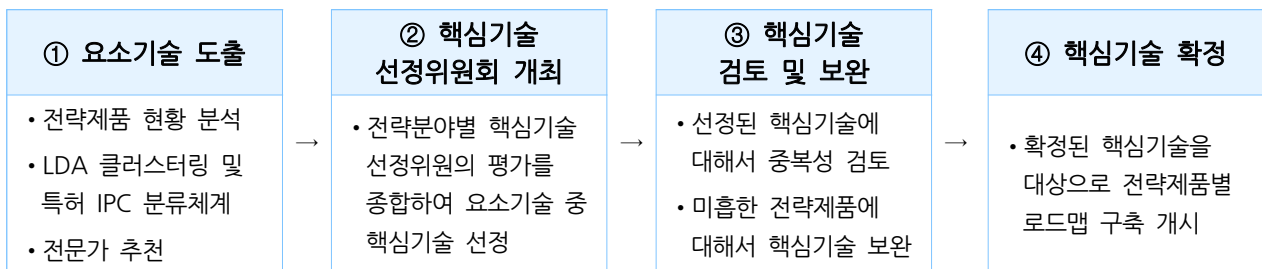
요소기술	출처
AI 기반 오류 데이터 보정 및 재생 기술	IPC 기술체계, 전문가추천
데이터 추출 및 변환 기술	특허 클러스터링, 전문가추천
데이터 형태 분류기술	특허 클러스터링, IPC 기술체계, 전문가추천
오류데이터 식별 및 탐색 기술	특허 클러스터링, 전문가추천
데이터 품질관리 기술	특허 클러스터링, 전문가추천
AI 및 머신러닝, 탐색적 데이터 분석 융합 기술	특허 클러스터링, 전문가추천

## 6. 전략제품 기술로드맵

### 가. 핵심기술 선정 절차

- 특허 분석을 통한 요소기술과 기술수요와 각종 문헌을 기반으로 한 요소기술, 전문가 추천 요소기술을 종합하여 요소기술을 도출한 후, 핵심기술 선정위원회의 평가과정 및 검토/보완을 거쳐 핵심기술 확정
- 핵심기술 선정 지표: 기술개발 시급성, 기술개발 파급성, 기술의 중요성 및 중소기업 적합성
  - 장기로드맵 전략제품의 경우, 기술개발 파급성 지표를 중장기 기술개발 파급성으로 대체
- AI기반 데이터 가치 고도화 플랫폼
  - 데이터 품질진단 및 관리 도구를 넘어 AI기반의 데이터 보정, 재생 등 데이터 자체로의 본질적 가치를 조정하고 관리하는 전반적인 데이터 품질 및 가치 고도화 기술로 발전
  - 데이터의 진단과 검토 AI기반의 오류 데이터 학습 등 융복합 등 다양한 기술적, 방법론적 결합

#### [ 핵심기술 선정 프로세스 ]



### 나. 핵심기술 리스트

#### [ AI기반 데이터 가치 고도화 플랫폼 분야 핵심기술 ]

핵심기술	개요
AI 및 머신러닝, 탐색적데이터분석 융합 기술	• AI에 기반한 다양한 형태의 데이터유형을 학습하여 상황과 형태, 그리고 종류에 따라 데이터 가치를 판단하고 오류를 식별하는 기술
AI 기반 오류 데이터 보정 및 재생 기술	• AI에 의해 학습된 다양한 데이터의 유형과 종류 중에서 특정 데이터의 결측치, 이상치, 특이치 등을 용도에 맞게 보정, 재생하는 기술
오류데이터 식별 및 탐색 기술	• 텍스트, 이미지, 영상, 음성 등 비정형 데이터에 대한 식별과 분석가능한 데이터로의 변환을 거쳐 데이터 가치 오류 여부를 탐색하고 모니터링 하는 기술
데이터 품질관리 기술	• 분석용 데이터로서의 적합성 여부 즉, 결측치, 신뢰성, 타당성, 유효성 등의 데이터 가치 품질에 대한 모니터링과 진단 기술

### 다. 중소기업 기술개발 전략

- 이상치, 결측치, 특이치 등에 대한 탐색적데이터분석을 기반으로 하는 데이터 가치 탐구에 대한 AI 및 머신러닝 등 방법론 개발
- 유형별, 특성별 생성형태별 데이터 수집과 오류데이터 등 교정, 재생 등 추천에 필요한 학습데이터 구축
- 산-학 연계 및 테스트베드 확보를 통한 개발기술 검증으로 국내시장 개척 및 선점 필요

### 라. 기술개발 로드맵

#### (1) 중기 기술개발 로드맵

[ SI기반 데이터 가치 고도화 플랫폼 분야 중기 기술개발 로드맵 ]

SI기반 데이터 가치 고도화 플랫폼	데이터의 진단과 검토 SI기반의 오류 데이터 학습 융복합 등 다양한 기술적, 방법론적 결합			
	2021년	2022년	2023년	최종 목표
AI 및 머신러닝, 탐색적데이터 분석 융합 기술				학습데이터와 신규데이터 내 오류데이터 탐색기술개발
AI 기반 오류 데이터 보정 및 재생 기술				오류데이터의 분류 및 보정, 재생 방법론 개발 및 코드개발
오류데이터 식별 및 탐색 기술				정형/비정형 데이터 통합 기술개발
데이터 품질관리 기술				데이터 가치와 품질 통합기술 개발

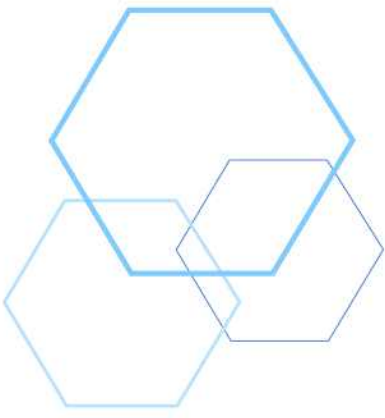
(2) 기술개발 목표

□ 최종 중소기업 기술로드맵은 기술/시장 니즈, 연차별 개발계획, 최종목표 등을 제시함으로써 중소기업의 기술개발 방향성을 제시

[ AI기반 데이터 가치 고도화 플랫폼 분야 핵심기술 연구목표 ]

핵심기술	기술요구사항	연차별 개발목표			최종목표	연계R&D 유형
		1차년도	2차년도	3차년도		
AI 및 머신러닝, 탐색적데이터분석 융합 기술	<ul style="list-style-type: none"> <li>AI기반 다양한 유형과 사례의 학습데이터구축</li> <li>EDA 방법론 융합</li> </ul>	다양한 유형과 사례의 학습데이터 구축	학습데이터와 신규데이터의 탐지 방법론개발	AI데이터 기반의 데이터 탐색기술 연계	학습데이터와 신규데이터 내 오류데이터 탐색기술개발	산학연
	<ul style="list-style-type: none"> <li>오류데이터 탐색 신뢰도(%)</li> </ul>	80	85	95	95	산학연
AI 기반 오류 데이터 보정 및 재생 기술	<ul style="list-style-type: none"> <li>AI기반 다양한 유형, 사례의 학습 데이터구축</li> <li>머신러닝을 활용한 데이터 추론방법론 결합</li> </ul>	데이터 형태와 특징에 따른 오류데이터 축적	특성별 상황별 오류데이터 보정 및 재생기술 개발	오류 탐지 및 보정, 재생기술 고도화	오류데이터의 분류 및 보정, 재생 방법론 개발 및 코드개발	산학연
	<ul style="list-style-type: none"> <li>보정 및 재생 유사도(%) 검증</li> </ul>	80	85	90	90	산학연
오류데이터 식별 및 탐색 기술	<ul style="list-style-type: none"> <li>다양한 형태의 오류(학습) 데이터 확보</li> <li>비정형 데이터의 오류탐지</li> </ul>	ETL등을 활용한 오류데이터 탐지기술 개발	데이터 변환기술 개발	이종간 데이터 통합 기술 고도화	정형/비정형 데이터 통합 기술개발	기술혁신
데이터 품질관리 기술	<ul style="list-style-type: none"> <li>데이터 가치를 고려한 데이터 품질평가</li> </ul>	데이터 가치 유형 분류	데이터 가치를 고려한 품질진단 평가 지표 개발	데이터 가치 유형 고도화	데이터 가치와 품질 통합기술 개발	기술혁신





전략제품 현황분석

# 유통/물류 빅데이터 구축 및 분석 시스템







# 유통·물류 빅데이터 구축 및 분석 시스템

## 정의 및 범위

- 유통·물류 빅데이터 구축 및 분석 시스템은 POS 및 수·발주 데이터를 토대로 딥러닝을 활용해 물류를 과학적, 합리적으로 분석하여 효율적 물류 구성을 통해 비용 절감과 효율증대를 확보할 수 있는 관리 시스템

## 전략 제품 관련 동향

시장 현황 및 전망	제품 산업 특징
<ul style="list-style-type: none"> <li>(세계) '18년 25억 8,000만 달러에서 연평균 17.3% 성장하여 '24년 67억 3,000만 달러 규모에 이를 것으로 전망</li> <li>(국내) '18년 380.4억 원 규모에서 연평균 성장률 16.8%로 '24년 965.8억 원 규모로 이를 것으로 전망</li> </ul>	<ul style="list-style-type: none"> <li>물류 트래킹 RFID 중심으로 성장 중</li> <li>공급망 전반에 걸쳐 의사결정자들은 빅데이터 소스를 효과적으로 관리하는 방법을 모색 중</li> <li>미래의 유통·물류 산업의 핵심기술은 빅데이터와 인공지능을 통한 데이터 분석</li> <li>국내 물류 빅데이터 활용 부문은 시작 단계</li> </ul>
정책 동향	기술 동향
<ul style="list-style-type: none"> <li>단순한 유통·물류 빅데이터 정책에 머무르지 않고 블록체인과 함께 4차 산업혁명의 기술을 융합한 정책을 추진</li> <li>유통상품의 상품정보 데이터베이스 구축 및 소비행태 분석시스템 실증을 진행하고 있으며, 온라인 유통사 등 업계와 지속 협의해 나갈 계획중</li> </ul>	<ul style="list-style-type: none"> <li>수송 규모의 확대와 소량·다품종 처리 시설 도입 등에 따라 물류 창고의 대형화 및 고도화</li> <li>물류공급망의 효율화를 위한 IT 시스템(LES) 접목</li> <li>개인 정보를 포함한 데이터 검색 기술 간 개인정보 보호에 대한 의무 사항 준수가 요구</li> </ul>
핵심 플레이어*	핵심기술
<ul style="list-style-type: none"> <li>(해외) Amazon, DHL, 트라센, UPS, 자라</li> <li>(대기업) 삼성SDS, CJ대한통운, 한국네트웍스</li> <li>(중소기업) 위비즈테크, 광원엘에스, 밸류체인씨엔티, 네오시스템즈, 택트레이서, 에이아이엠, 랩투마켓, 진코퍼레이션, 더블유앤이케이</li> </ul>	<ul style="list-style-type: none"> <li>개인 정보를 포함한 데이터 검색 기술</li> <li>데이터 분석 기술</li> <li>빅데이터 보안 기술</li> <li>데이터의 유지 보존 및 폐기 기술</li> <li>스마트 화물 추적 서비스의 동향 분석 기술</li> </ul>

\*생태계 취약 전략제품

## 중소기업 기술개발 전략

- 글로벌 플랫폼에 제공하기 어려운 유통·물류 실시간 데이터 수집 및 연계기술의 개발
- 유통·물류 산업 특화 빅데이터 기반 서비스 창출
- 플랫폼이 아닌 빅데이터를 활용한 서비스 개발과 유료 수익 모델 개발을 통한 수익 창출

## 생태계 강화방안

- 데이터 3법의 시행이 끝이 아니고 데이터 시장의 시작이며, 데이터 거래 활성화를 위한 다양한 방안과 데이터 축적 방안을 마련할 필요
- 빅데이터는 물류에서 새로운 경쟁력을 높일 수 있는 핵심적인 요소이지만 물류 산업에서 빅데이터를 어떻게 수집하고 활용해야 할지 국가적인 관점에서 데이터 개방과 활용을 위한 지원책 등이 논의되어야 함

# 1. 개요

## 가. 정의 및 필요성

### (1) 정의

- 유통·물류 빅데이터 구축 및 분석 시스템은 POS 및 수·발주 데이터를 토대로 딥러닝을 활용해 물류를 과학적, 합리적으로 분석하여 효율적 물류 구성을 통해 비용 절감과 효율증대를 확보할 수 있는 관리 시스템
- 인공지능(AI)과 정보통신(ICT) 등의 다양한 신기술을 활용해 물류 현장 자동화 및 무인화 설비를 구축해 효율적으로 운영하는 시스템

### (2) 필요성

- 빅데이터는 기술 발달로 유통·물류 분야를 포함한 다양한 산업에 영향력이 높아지는 추세
  - 대표적인 글로벌 물류 기업인 DHL에서 발간한 Logistics Trend Radar에서 기술적 트렌드(Technology Trend)와 사회경영 분야 트렌드(Social & Business Trend)를 제시하였는데 이 중 기술 트렌드 부분에서 가장 그 시급성과 기술파급력이 높은 것이 빅데이터로 나타남
  - 앞으로 모바일 디바이스의 보급과 IoT의 진전에 따라 GPS 데이터와 RFID 데이터, 센서 데이터 등의 새로운 비정형 데이터가 더욱 급증할 것으로 예상되며, IoT를 생성하는 다양한 비정형 데이터의 순차 처리 및 실시간 처리가 가능하게 될 것으로 전망
  - 빅데이터는 기존의 정보 시스템과 다른 데이터(모바일 및 IoT 등)와 결합하여 데이터 융합과 인공지능 기술(AI)을 활용한 고급 분석으로 가속화될 전망
  - 인공지능(Artificial Intelligence)은 인간의 지능을 인공적으로 실현하기 위한 연구 분야로 2000년대부터 빅데이터를 이용하여 인공지능 스스로 지식을 습득하는 기계 학습(Machine Learning)이 실용화되었고, 인공지능이 스스로 습득하는 심층 학습(Deep Learning)이 등장하고 응용 연구와 개발이 활발
- 첨단 정보통신기술 발전이 촉발하는 4차 산업혁명의 흐름이 국내외 유통에서 발생하고 있으며 초지능, 초실감, 초연결화를 구현하는 ‘리테일 4.0’ 개념이 등장
  - 개인화된 소비 확대, 라이프스타일의 다양화 등으로 나타나는 소비자의 변화
    - 글로벌 유통기업들은 인공지능 도입을 통해 보다 개인화된 고객 경험 제공(54%), 고객 이해도 제고(54%), 고객과의 커뮤니케이션 실시간 관리(52%) 등을 실현할 수 있을 것으로 기대
  - 오프라인/온라인 유통 채널과 점포가 4차 산업혁명 신기술과 만나 진화함에 따라 유통(Retail)과 기술(Technology)이 융합된 ‘리테일 테크’ 개념이 등장

- 물류업계는 e-Commerce의 폭발적인 성장과 고객 서비스의 요구 증가로 운영의 복잡성과 난이도가 높아지고 있으며, 고령화, 임금인상의 사회적 환경요인으로 많은 어려움 직면
  - 이에 업무환경의 자동화와 운영의 최적화에서 해법을 찾고자 많은 노력을 기울이고 있으며, 이는 인력 의존도를 최소화하고 공정별 정확도 제고 및 Human Error 감축, Time Loss의 제거를 통한 생산성 향상에 직접적인 효과가 있다고 판단
  - 물류 시스템 개선으로 인한 생산성 개선과 IoT 데이터 분석 시장에 대한 기대수요가 높으며, 고객 관리 및 마케팅 기법에 대한 수요도 존재하며, 다음의 세 가지 측면에서 활용될 전망
    - 기업 간의 물류의 제 기능을 고도화하는 것과 동시에 조달, 생산, 판매, 회수 등 분야 간 정보의 상호 활용을 촉진
    - 기업 간의 수요와 공급의 동기화로 관련된 기업끼리 서플라이 체인(supply-chain)을 형성하여 생산성 향상과 수익 확대에 이용
    - 기업이 사회적 책임을 완수하는데 있어 공공적 문제를 해결하기 위한 효율적인 경영 전략의 도입을 목표로 설정
  - 다만 물류와 유통은 전통적인 분야인 만큼 데이터의 공유와 거래에도 보수적인 입장을 취하는 민간 기업이 많아 빅데이터 적용 성공사례가 선행되어야 민간 기업을 빅데이터 시장으로 끌어들이는 유인책이 가능

[ 물류 자동화 개념도 ]



\* 출처: LG유플러스

## 나. 범위 및 분류

### (1) 가치사슬

- (후방산업) 물류 빅데이터가 생성되는 ICT 인프라, IoT, 클라우드 산업, 활용 가능한 지식처리 및 시각화 플랫폼, 인공지능, 로봇틱스, AR/VR, 3D BIM 맵핑 분야 등
  - ERP 시스템, 바코드 및 RFID 스캔 데이터, 센서 카메라, 아카이브, 인터넷, 소셜 네트워킹 망 등에서 발생
  - 데이터 유형, 크기, 성장률, 데이터 구조, 데이터 품질, 그리고 소유권 등의 다양한 형식
  - 스마트 물류 관련 키워드를 분석한 연구결과, DHL은 제 4차 산업혁명 시대에서 물류 기술 트렌드와 사회, 비즈니스 측면의 트렌드를 제시하여 확장을 전망
  - 5년 내 적용 가능한 물류 기술로 빅데이터, 클라우드, 자율물류, 3D 프린팅, 로봇틱스 및 자동화, 사물인터넷, 위치탐지, 웨어러블 기술, 증강현실, 저비용 센서 기술, 암호결제를 제시
- (전방산업) 상품을 고객에게 직접 전달하는 운송업과 물건의 유통을 담당하는 유통업, 자재 및 완제품을 활용하는 제조업 등 전 산업

#### [ 유통·물류 빅데이터 분야 산업구조 ]

후방산업	유통·물류 빅데이터 분야	전방산업
ICT 인프라, IoT, 클라우드 산업, 지식처리 및 시각화 플랫폼, 인공지능, 로봇틱스, AR/VR, 3D BIM	유통·물류 스마트 물류, 암호결제	유통, 제조 등 전산업

## (2) 용도별 분류

### [ 용도별 분류 ]

구분	적용 사례
소매·유통(Retail)	• AI의 초보 단계인 선형신경망에 대한 활용도가 가장 높고, 다음으로 순환신경망 기술을 적용
수송·물류(Transport and Logistics)	• 선형신경망과 합성곱신경망을 이용하여 산업의 효율성을 개선
SCM·제조 (Supply-Chain Management and Manufacturing)	• 선형신경망과 함께 강화학습 및 합성곱신경망을 이용한 공급망관리가 유효
마케팅·판매(Marketing and Sales)	• 선형신경망을 통한 소비자의 구매패턴 분석
서비스 운영(Service Operations)	• 선형신경망과 순환신경망 기술을 활용하여 소비자들의 피드백 적용 및 효율성 증대

\* 출처 : 스마트 유통·물류 산업에서의 인공지능 서비스(정보통신기획평가원), 2019

## (3) 유통·물류 세부 단계별 분류

### [ 유통·물류 단계별 빅데이터 적용 ]

적용 방식	역할
Scheduling	• 재고 수준, 수요 및 생산능력의 가시성 제고, 따라서 생산 및 유통 일정의 정확성 향상
Warehousing	• 창고 ERP 시스템 내의 빅데이터 실시간 분석 및 재고 수준, 납품 불일치 및 입고 납품 확인
Demand forecasting	• 매출, 시장 동향, 경쟁사 데이터, 관련 지역 및 세계 경제요인 등에 대한 접근으로 더욱 정확한 수요 추정
Delivery	• 배달 경로, 교통 데이터, 실시간 날씨 모니터링 및 용량 및 자산 공유가 필요한 경우 재라우팅
Inventory Planning	• SKU(Stock Keeping Unit) 레벨의 완벽한 투명성, 완전히 자동화된 보충 시스템과 재고 부족 범위를 제거하고 재고 범위를 최적화하는 수요 예측 데이터
Distributing	• 자재 흐름 데이터를 기반으로 유통 허브, 플랜트 및 창고의 복잡한 웹을 실시간으로 최적화

## 2. 산업 및 시장 분석

### 가. 산업 분석

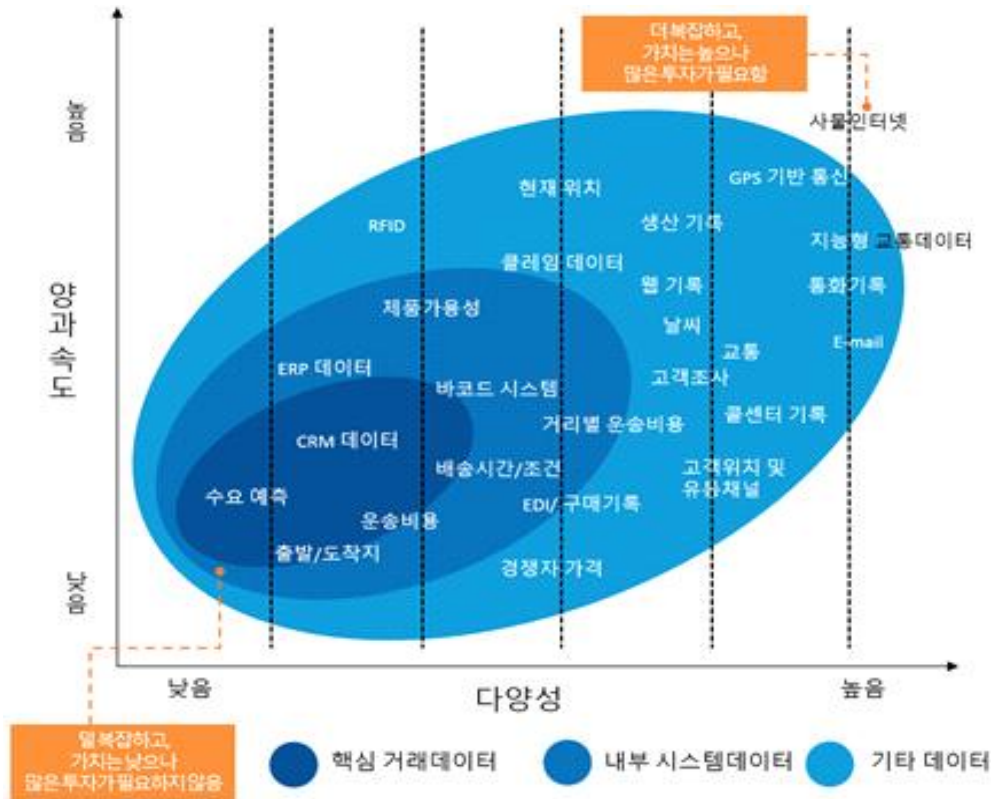
#### ◎ 물류 트래킹 RFID 중심으로 성장 중

- 물류 산업에서의 RFID의 도입은 2000년대 중반 이후부터 활발히 진행되고 있는 분야라 최신 기술이라고 하기는 어렵지만, 그 응용과 적용 범위가 현재도 증가하는 추세로, 눈여겨보아야 하는 기술
  - 소매유통, 물류, 우편 등의 분야에서 화물 모니터링과 추적 및 효율적인 재고관리 등 물류 가시성 확보를 위한 시스템의 기반기술로서 RFID 기술이 많이 활용
  - RFID 적용 분야가 전반적으로 물류 및 유통산업에서 다소 감소하는 추세로 예상되지만, 전체 RFID 시장은 꾸준히 증가할 것으로 예상
- 국내 시장의 동향은 대기업을 중심으로 자체적 RFID 성공적 사례를 기반으로 RFID에 대한 수요가 꾸준히 증가
  - 특히 국내의 RFID 응용서비스 분야 중 물류 유통 분야에서의 적용이 대부분을 차지할 정도로 물류 유통 분야에서 널리 쓰이는 기술
  - 물류 처리량 방대 기업들을 중심으로 해당 기술에 기반, 물류 최적화를 위한 물류 가시성 확보중
- 국내 유통, 물류 시장은 그 규모가 꾸준히 증가하고 있으며, 이에 따라 물류 시장에 관여하는 여러 기업이 물류 시스템의 효율화를 위한 투자가 지속될 것으로 예상
  - 꾸준히 증가하는 물류 시장과 다양화되는 고객들의 수요, 그리고 기업의 안정적인 물류 시스템 운영을 위해 물류 가시화에 대한 투자가 증대될 것으로 예상

#### ◎ 공급망 전반에 걸친 빅데이터의 침투

- 공급망의 의사결정자들은 빅데이터 소스를 효과적으로 관리하는 방법을 모색하고 있으며, 빅데이터 솔루션을 적용하는 공급망 운영 프로세스의 개선을 도모
- 빅데이터는 물류, 재고관리, 상품 구성, 마케팅 등 다방면으로 활용 중
  - 모바일 결제, 오프라인 재화의 디지털화, 앱을 활용한 회원제 등 구매 고객의 온·오프라인 데이터 수집 및 고객의 소비 패턴 추적, 분석
  - 모바일 결제 시스템은 오프라인 비즈니스 내에서 판매, 운영, 회원 관리의 디지털 전환을 가속화했고, 모바일 결제 생태계는 판매자와 구매자 사이에 데이터가 오가는 기반으로서 기능
  - 이를 통해 기업은 모바일 결제를 통해 소비자들의 디테일한 구매 데이터를 더욱 빠르게 확보하고 분석 가능

[ 공급사슬 상의 빅데이터 활용 ]



\* 출처: KPMG 공급사슬 빅데이터 분석보고서(2017)

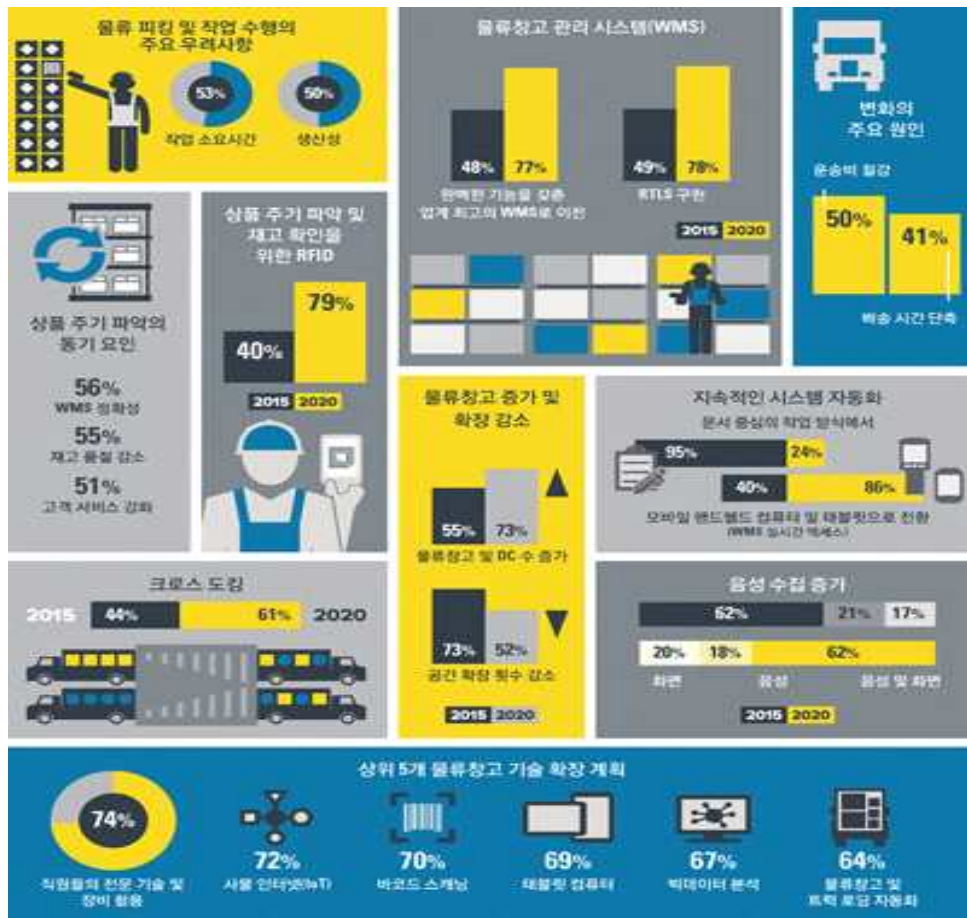
◎ 미래의 유통·물류 산업의 핵심기술은 빅데이터와 인공지능을 통한 데이터 분석

- 빅데이터 활용과 오프라인 매장 시스템 디지털화, 인공지능 시스템 도입 등 다양한 시도를 통해 해외 선진 기업들의 물류 시스템은 최적화되는 추세
  - 소비자가 직접 제품을 고르고 운송하는 개념이 아닌 소비자는 제품을 구매하고 유통기업이 필요한 시기에 필요한 만큼을 원하는 곳으로 배송하는 개념으로 변화
  - 매장은 계절, 시간, 지역 등을 기반으로 제품이 판매되는 상황을 파악하여 필요한 물량만큼만 발주할 수 있음. 매장은 당일 매출을 예상하고 재고 및 상품정보를 통합적으로 제어 가능
  - 공급업체는 제공된 정보를 통해 생산관리를 최적화하여 운영 효율을 향상 가능
  - 통합 비즈니스 플래닝을 지원하는 빅데이터 솔루션은 현재 시장 동향과 고객 선호를 분석하여 조직이 보다 공급망에 적극적 대응하여 운영
  - 다수의 정적 및 동적 데이터의 특정 포인트에서 다양한 시장 판매, 소셜 미디어, 그리고 인구통계 데이터의 삼각관계 수치를 통해 공급망 활동을 예측하고 능동적으로 계획할 수 있는 인사이트 확보 가능



- 유통·물류 산업의 경쟁력 향상을 위해 빅데이터와 인공지능을 통해 고객 가치를 향상하기 위한 스피드, 유연성과 비용 절감을 추구하는 경향을 보임
  - 전통적으로 유통·물류 산업의 경쟁력은 운송비와 보관비 등 직접적으로 투자한 비용의 최소화에서 있어 저비용 전략을 중심으로 규모의 경제, 대량 운송 수단 등이라고 여겨짐
  - 비용보다는 유연성과 스피드에 초점을 맞춘 ZARA와 같은 Fast-fashion의 등장, 제품 라인업을 단순화하여 비용을 절감하기보다 다양한 화면크기를 통해 고객에게 더 많은 선택권을 제공하는 삼성의 스마트폰 성공전략 등 고객 중심의 SCM 개념이 등장하며 전통적인 시각에서 벗어나는 추세
  
- 물류 창고의 대형화 및 고도화는 물류 산업의 노동 주체를 인력에서 로봇·자동화 기술로 이행시키는 현상을 주도
  - 선진국 및 선도 물류 기업들은 물류 로봇 등 첨단 ICT를 활용해 물류 자동화·효율화 및 비용 절감을 추진
  - Amazon은 물류 창고 로봇 '키바'를 통해 물류센터 운영비용 20% 절감, 재고 보관 공간 50% 증대 등의 성과를 발표
  - MOL, NYK 등 일본선사는 2025년까지 원격조정(무인)선박 도입을 계획하고 있으며 일본 정부는 2030년까지 정부 주도로 인공지능(AI)을 투입해 물류 산업을 '완전 무인화'한다는 로드맵을 발표

[ 물류산업에서의 빅데이터 활용 가능성 ]



\* 출처: LM물류매거진, 2016.08



## ◎ 우리나라는 물류 빅데이터 활용에 있어 시작 단계

- 사업적 가치가 있는 데이터가 부족하며 별다른 데이터 거래시장이 전무
  - 해외의 경우 AggData, Content for Sale, Data Market 등 사업적 가치를 지닌 데이터를 판매하는 기업이 다양하게 존재하나, 국내는 개인정보 보호법으로 인해 공공데이터 활용이 낮은 수준에 머물러 있으며 데이터를 생산하는 통신, 금융 등 핵심기업 또한 데이터의 유출에 아주 보수적인 태도를 보이고 있어 데이터 거래시장이 제대로 형성되어 있지 않음
- 전사적 자원관리 시스템을 시작으로 물류에서도 쓴, WMS, OMS 등 관련 시스템에 대한 투자가 되어 데이터는 축적되나 활용도는 미흡
  - 물류산업에서 운송장 정보는 매일 축적되고 있으나 활용에 진전이 없고, 운송장 정보의 생성에서 화물의 이동 등에 대한 실시간 정보가 누적되고 있으나 이것들 속에서 큰 의미를 찾기 위한 노력이 부족했거나 의미를 찾지 못함
  - 물류창고 WMS에서도 제품 입출하 정보에서 시작하여 보관되는 제품의 사이즈, 종류, 물류창고의 온도, 습도 정보 등 다양한 정보가 실시간으로 측정 가능한 상황이지만 동적인 데이터의 흐름은 활용되지 못하고 버려지고 있으며, 요약정리된 자료만 정적인 데이터로 활용되고 있는 실정
- 활발한 기술공유를 통해 급성장하는 해외시장에 비해 국내시장의 보수성은 기술발전을 저해하는 가장 큰 요인
  - 유통·물류시스템 개선으로 인한 생산성 개선과 IoT 데이터 분석 시장에 대한 기대가 높으며 고객 관리 및 마케팅 기법에 대한 수요도 높음
  - 다만 물류와 유통은 전통적인 분야인 만큼 데이터의 공유와 거래에도 보수적인 입장을 취하는 민간기업이 많아 성공사례가 선행되어야만 민간기업을 빅데이터 시장으로 끌어들이는 유인이 될 것
- 다국적 기업의 공급사슬관리(SCM) 및 e-Commerce의 발달이 지속적으로 진행되어, 부가가치를 창출할 수 있는 유통·물류 관련 서비스의 수요가 증대되는 양상
  - 지식기반(Knowledge-based) 사업과 전자상거래의 발달로 세계화 기업의 공급망의 효율성을 제고하기 위해 다양한 물류시스템의 개선이 핵심 요소. 특히 다국적 기업이 물류 산업의 중심이 된 현재, 해외 배송 등 장거리 배송이 폭발적으로 증가
  - 국제 물류가 활발해짐에 따라 빅데이터 기반의 유통 및 물류 배송 추적의 필요성이 증대되고 있으며, 다양한 업체에서 물류망의 가시성 확보를 위한 서비스를 제공

## 나. 시장 분석

### (1) 세계시장

- 물류 관련 빅데이터 분석 세계 시장은 2018년에 25.8억 달러로 평가되었으며, 연평균 17.3% 성장하여 2024년까지 67.3억 달러에 이를 것으로 전망됨
  - 정보 기술의 발전으로 기업은 이제 막대한 양의 데이터에 액세스, 저장 및 처리가 가능하게 됨
  - 조직에서는 데이터 세트를 분석하고 운영에 적용할 주요 통찰력을 파악해 빅데이터 분석 기술을 이용하여 운영

[ 빅데이터 기반 물류 트래킹 시스템 세계 시장규모 및 전망 ]

(단위 : 백만 달러, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
세계시장	2,580	3,030	3,550	4,170	4,890	5,740	6,730	17.3

\* 출처: Supply Chain Big Data Analytics Market - Mordor Intelligence(2019)

- 유통산업의 다변화 추진 및 온라인, 오프라인, 모바일을 융합한 옴니채널(Omni-Channel)의 등장으로 소비채널이 급속히 확장하였으며 기능 또한 다양해져서 빅데이터를 통한 유연한 대응이 필요
  - 사업 포트폴리오 확대 전략을 수립하고 변화된 소비자의 편의 증대와 소비자 경험에 대해 데이터를 통한 차별화 시도 필요
  - 멀티 채널 유통이 확산되면서 채널 별로 분리 운영되는 재고들을 통합 관리하여 고객이 원하는 제품을 가장 효과적인 곳에서 제공할 수 있어야 하므로 고객에 대한 통합된 정보를 제공하는 빅데이터 플랫폼을 구축 필요
- GPS 및 위치정보산업의 시장 환경을 분석한 결과, 스마트 화물 추적 서비스의 동향 분석이 매우 중요한데, 이는 유통에 활용되는 스마트 화물을 추적함으로써 물류 루트의 가시성을 확보할 수 있는 핵심기술이기 때문
  - 한국전자통신연구원의 위치정보서비스(LBS) 기술 및 시장 동향 분석 연구에 따르면, 2017년 세계 LBS 시장은 215.1억 달러로 평가되며, 연평균 성장률 38.4%로 성장하여 2022년까지 LBS 시장은 1091.96억 달러에 이를 것으로 예상
  - 특히, 북미와 유럽이 세계 LBS 시장의 50% 이상을 점유
  - 차량운행관리는 GPS 및 추적기술을 통해 차량의 현재 위치와 운행 정보 등에 관한 실시간 데이터를 제공하는 솔루션임. 특히 Bi Intelligence에 따르면, 2020년에는 북미에서 상업용 차량 1억 8,000만대에 적용 할 것으로 예견, 이는 전체 시장의 90%가량을 차지하는 매우 큰 수치
  - 관련 기업들은 트럭을 통한 물적자산의 운송 모니터링에 GPS 기술을 이용한 솔루션을 매우 효과적으로 사용 중

## (2) 국내시장

- 빅데이터 기반 트래킹 시스템의 국내 시장 규모는 2018년 380.4억 원 규모로 평가되었으며, 연평균 16.8% 성장하여 2024년까지 965.8억 원에 이를 것으로 전망됨

[ 빅데이터 기반 트래킹 시스템 국내 시장규모 및 전망 ]

(단위 : 억 원, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
국내시장	380.4	444.3	518.9	606.1	708.0	826.9	965.8	16.8

\* 출처 : Big Data Market Size - MarketsandmarketsSupply, Chain Big Data Analytics Market Mordor Intelligence (2019)을 바탕으로 전 세계 빅데이터 시장과 국내 빅데이터 시장의 관계를 복합적으로 고려해 네모아이씨지 재추정

- 한국인터넷진흥원 '2019년 국내 위치정보 산업 동향조사'에 따르면 국내 위치정보 산업 LBS(Location-Based Services) 매출액 규모가 2018년대비 2,500억원 증가한 1조 8,418억 원이 될 것으로 전망
  - IoT 업계에서는 스마트 더스트 관련 연구개발(R&D)이 지속적으로 이뤄지고 있으며 향후 활용 사례가 확산될 것으로 예측 중
  - 물류 분야에서는 화물 팔레트나 택배상자 등에 장치를 부착해 배송 물품의 위치를 파악하는 IoT 활용 방안이 연구 중

### 3. 기술 개발 동향

- 기술경쟁력
  - 유통/물류 빅데이터 구축 및 분석 시스템은 미국이 최고기술국으로 평가되었으며, 우리나라는 최고기술국 대비 69.8%의 기술 수준을 보유하고 있으며, 최고기술국과의 기술격차는 2.4년으로 분석
  - 중소기업의 기술경쟁력은 최고기술국 대비 60.4%, 기술격차는 3.2년으로 평가
  - EU(78.9%)>중국(72.4%)>한국>일본(68.7%)의 순으로 평가
- 기술수명주기(TCT)<sup>10</sup>
  - 유통/물류 빅데이터 구축 및 분석 시스템은 7.04의 기술수명주기를 지닌 것으로 파악

#### 가. 기술개발 이슈

##### ◎ 유통·물류 산업 기술 트렌드

- 수송 규모의 확대와 소량·다품종 처리 시설 도입 등에 따라 물류 창고의 대형화 및 고도화
  - 수요 측면에서 온라인 구매의 증가, 산업 측면에서는 M&A를 통한 화주 기업의 대형화, 인프라 측면에서 교통 시설 발전에 따른 거점 집중화가 진행되면서 물류 창고의 대형화가 진행
  - 온라인 거래 확대와 1인 가구의 증가, 신선 물류에 대한 수요 증대 등으로 물류 창고에서 처리해야 하는 물품의 유형, 크기가 다양화되고 이를 처리하기 위한 물류 창고가 고도화
- 집하부터 배송까지 다양한 이해관계자들이 얽혀 있으며 톱니바퀴처럼 일사불란하게 움직이는 물류는 다른 산업에 비해 많은 트랜잭션이 발생
  - 기업들은 경쟁에 앞서기 위해 해외 시장을 개척하고 저렴한 생산지를 찾아 공장을 이전하면서 더욱 복잡해진 공급망의 효율적 관리가 핵심 성공 요소로 부각
  - 물류 산업은 역사는 오래되었으나 산업특성 상 오프라인 위주의 산업 특성으로 빅데이터와 같은 ICT 융합에 대한 저항 발생
    - 물류에서 빅데이터의 활용도를 향상시키기 위해서, 우선 물류에서 활용 가능한 다양한 센서들을 개발하고 보급하여 물류 현장에서 빅데이터를 구축할 만큼의 데이터 생성이 요구
    - 운송수단의 발달과 컴퓨터와 인터넷이라는 혁명적 기술이 도입되고, 복잡한 공급망과 물류의 관리 방식이 한층 발전됨으로써 이러한 변화에 대응
    - 고속통신망과 스마트폰의 보급 확산으로 점화된 e-Commerce는 쇼핑의 무대를 세계로 확대하고, 직접 소비자를 대상으로 하는 물류 시장이 재편성

10) 기술수명주기(TCT, Technical Cycle Time): 특허 출원연도와 인용한 특허들의 출원연도 차이의 중앙값을 통해 기술 변화속도 및 기술의 경제적 수명을 예측

- 물류산업이 사상 유례없는, 산업 경계가 허물어지고 과거에 없었던 새로운 유형의 경쟁이 나타나는 격변의 시대에 진입
  - 신기술 출현, 신규 시장진입 등 변화 추세 뚜렷한 산업 경계가 허물어지면서 산업 생태계가 대변혁을 겪고 있는 가운데 물류산업 역시 과거에 없었던 새로운 유형의 경쟁이 나타나는 환경 변화에 직면
  
- 물류공급망의 효율화를 위한 IT 시스템 접목
  - 많은 기업들이 LES(Logistics Execution System)<sup>11)</sup>를 도입하여 일정 부분 성과를 확인
  - 정보통합은 로봇, 자율주행, 가상현실, 블록체인 등 4차 산업혁명 기술과 융합 필요
  - 물류 정보의 통합, 공유와 관계된 매우 중요한 기술로, 물류공급망 내의 공공데이터뿐만 아니라 이해관계자들의 물류 정보까지 통합하기 위해서는 정책적 해결방안이 함께 제시 필요
  
- 유통·물류 기술 트렌드는 물류공급망, 물류공급체계, 물류시설에 해당되는 기술개발로 주종을 이뤄 좁게는 물류 운영, 물류 작업공간, 물류 작업자에 해당되는 기술개발로 이어질 전망
  - 물류시설의 자동화, 물류 운영의 고도화, 물류공급망의 효율화, 물류시설의 친환경화, 작업공간의 안전화, 물류공급체계의 표준화, 물류 작업자의 노동 편의화, 화물의 보호/보안화 등으로 기술 트렌드가 존재
  
- 운송 부문에서는 대량 수송에서 맞춤형 운송 서비스 중심으로 수요가 이동중
  - 소규모의 개인화된 물류서비스의 증가 및 운송 조건의 복잡화, 하역 보관부문에서는 수송 규모의 확대와 소량 다품종 처리 시설 도입 등에 따른 물류 창고의 대형화와 고도화가 지속중.
  - 정보처리 부문에서는 거래 규모의 증가와 수요자의 물류관리 니즈 증대로 인한 정보관리 제공 필요성이 높아지는 중

[ 맞춤형 운송 서비스의 예시 ]



\* 출처: 한두한, 2019

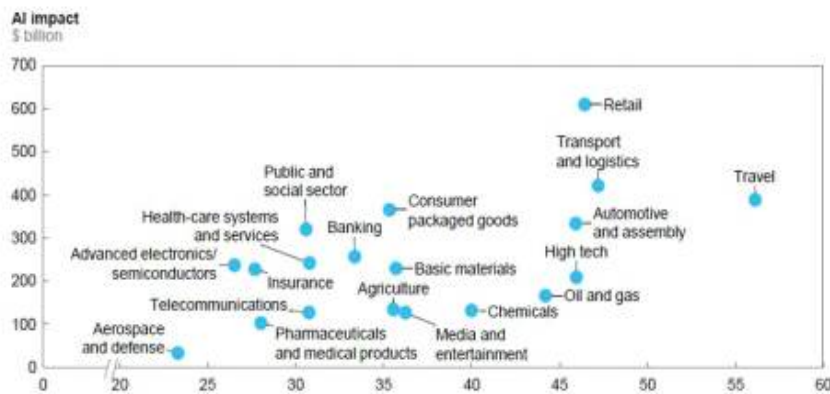
11) ERP의 하위시스템으로 실제 물류 업무의 정보처리를 담당하는 시스템으로 OMS(Order Management System), TMS(Transportation Management System), WMS(Warehouse Management System)으로 구성

### ◎ 유통·물류 산업의 빅데이터 기술

#### □ 개인 정보를 포함한 데이터 검색 기술

- 모바일 장치 및 IoT 디바이스의 보급에 의해 개인정보가 포함된 데이터도 쉽게 수집 가능하여 개개인에 특화된 빅데이터 응용서비스가 예상되지만, 개인정보를 포함한 데이터 취득에는 미리 이용 목적을 확인하고 본인의 동의 또는 이용 목적을 미리 공지하는 등의 개인정보 보호에 대한 의무 사항 준수가 요구

[ 물류산업에서의 빅데이터/인공지능 영향력 ]



\* 출처: McKinsey&Company, 2019

#### □ 데이터 분석 기술

- 분석 결과를 이용자에게 제공하는 데 있어 분석가는 결과를 이해할 수 있는 이유와 근거를 확인하고 사용자에게 제시할 수 있어야 하지만, 인공지능·기계 학습·심층 학습에 의한 분석의 경우, 도출된 결과의 근거 확인이 어렵고, 분석가가 전문가가 아닌 상황에서 책임을 가지는 점이 실용화를 방해
- 또한, 유통 물류 산업의 경우, IT분야 전문가가 아닌 부분을 감안하여 인공지능·기계 학습·심층 학습을 이끌어내는 추론 결과의 이유와 근거를 나타내는 기술의 개발이 필요

#### □ 빅데이터 보안 기술

- 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
- 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
- 또한, 은닉 검색은 공개키 암호화 또는 대칭 키 알고리즘을 이용하여 AND 검색, OR 검색, 범위 검색, 유사 검색 등을 구현할 수 있는 기술 적용
- 은닉 암호화는 암호화 상태 처리에서 얻은 결과를 이용자가 소유하는 개인 키로만 해독한 후 암호화된 결과로 변환하여 이용자에 제공함으로써 빅데이터 시스템 운영자가 데이터 유출 방지

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

재고관리와 창고보관은 특히 보관된 부품과 제품의 움직임을 감시하고 지속적으로 추적하여 올바른 재고가 올바른 장소로 출하되었는지 확인하기 위해 IoT를 이용하여 대폭 개선 가능

- RFID 태그와 센서를 사용하여 품목의 정확한 위치, 창고 및 위치 이동, 그리고 기타 중요한 정보를 입수하여 기록하고 이를 추적하여 빅데이터 분석으로 비용 절감의 기회를 찾고 납기 지연 등을 미연에 방지하는 등의 성과를 기대
- 창고에 있는 컨베이어 로봇의 RFID 태그와 센서는 전 창고 내에서 로봇의 움직임을 동기화하고, 다운 타임과 작업 시간 및 운반하는 재고량을 모니터링하여 유틸리티와 일정을 최적화

◎ 유통·물류 분야 빅데이터 도입을 위한 준비

유통 데이터는 기본적으로 2차 데이터가 필요

- 다른 목적을 위해서 구축된 시스템으로부터 자동적으로 수집되어 보존되고 있는 로그 데이터, 혹은 트랜잭션 데이터로 데이터의 양에 따라 데이터 가치가 정비례로 증가
- 빅데이터 활용의 문맥에서 수집, 분석, 이용을 검토하는 데이터의 경우, 다른 목적으로 수집된 2차 데이터를 활용하여야 함. 예를 들면, POS 데이터는 매장 관리나 상품발주를 위한 판매 시점 관리가 본래의 목적이며, 그 때문에 수집되고 있는 데이터임. 다른 예로 신상품 개발에 POS 데이터를 이용하는 경우, POS 데이터는 다른 목적으로 수집된 데이터(=2차 데이터)

유통 빅데이터 응용 시스템에서 저장 · 관리하는 데이터 구조에 따라 적절한 데이터베이스 관리 시스템의 선정이 요구

- 수집하는 데이터 이용 목적을 명확히 하고, 레거시 시스템을 통해 즉시 사용할 수 있는 데이터 (딥 데이터)로 변환하고 효율적인 추적·관리 및 분석, 이용 효율화로 연결 필요
- 수집된 데이터의 선별 및 가공의 경우, 그 타당성에 대한 근거 필요 여부를 검토
- 수집되는 데이터의 종류와 수집 방법 등에 있어, 중복 제거 및 표현의 통일 등의 표준화를 통해 데이터 품질 향상
- 데이터 송신자와 전송 장비를 속인 가짜 데이터나 데이터 송신자와 전송 장비가 보낸 데이터가 변조되는 경우는 빅데이터 분석 활용에 치명적인 영향을 미칠 수 있음. 사이버 공격에 대한 대응은 물론 데이터 송신자, 장비 정보 등 데이터가 도중에 변경되지 않았는지 확인 등이 요구



□ 유통 빅데이터 분산 처리 시스템 요구 사항

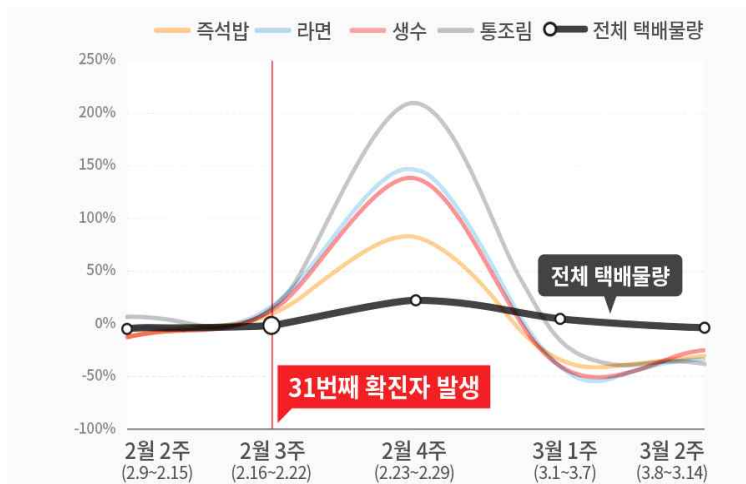
- 첫째, IoT 센서를 포함한 빅데이터 분산 처리 시스템이 처리하는 데이터의 양과 서버의 규모는 기존과 비교가 되지 않을 정도로 규모가 큰 점을 고려하여, 비용적인 측면에서의 효율성, 지속적으로 증가하는 데이터들의 수용, 빈번하게 발생하는 장애에 대한 대처와 인프라 관리의 편리성이 요구
- 둘째, 다양한 사용자들의 요구를 충족시키기 위해 데이터의 입·출력과 높은 처리 성능이 요구됨. 구체적으로 대용량 파일에 대한 신속한 입출력 성능, 네트워크 구조 인식을 통한 저장되는 데이터들의 최적 배치, 보다 빠른 처리를 위한 효과적인 캐시 사용, 순간적으로 집중되는 부하의 유연한 대처 등이 존재
- 셋째, 데이터에 대한 암호화, 데이터 영역에 대한 사용자 간 엄격한 접근 제어, 사용자 데이터에 대한 관리자의 접근 제어와 같은 사항들이 요구

□ 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

□ CJ대한통운의 유통 물류 빅데이터 분석환경은 데이터 관점에서의 전사 데이터 허브 및 데이터 거버넌스 영역, 분석관점에서의 분석 서비스 모델 개발 영역, 인프라 관점에서의 데이터 수집, 처리, 저장, 분석 영역으로 구성

[ CJ대한통운의 코로나19 31번 확진자 발표 전후 택배 데이터 분석 ]



\*출처 : CJ대한통운, 2020



□ KPMG의 리테일 테크 4.0에서는 각 분야별로 사용될 수 있는 5가지 분류를 제시

- 전략 수립 단계: AI는 인간을 대신해 데이터를 분석하고 전략을 수립해 경영자의 의사 결정에 도움을 줄 수 있는데, 수요 예측부터 상권 분석, 소비자 분석에 이르기까지 AI가 경영 전략 수립에 기여하는 영역을 다각화
- 물류관리 단계: 재고 및 창고 관리, 재고 실사, 매장 관리 등의 영역에서 자동화가 이루어지고 있으며, 특히, IoT 센서의 발달은 재고관리의 정확성을 높이고, 모든 상품의 이동과 상태 정보를 실시간으로 공유하고 관리하는 ‘Real-Time SCM’화
  - 로봇이 사람을 대신하게 해 관리의 효율을 극대화하려는 목적
- 마케팅 단계: AR과 VR 등 각종 신기술이 고객 경험을 강화하고, 오프라인 매장이 디지털화되면서 ‘피지털(phygital)’이라는 패러다임이 나타나고 있으며, AI와 IoT는 빅데이터를 통해 특정 소비자를 주목하는 타겟 마케팅과 실시간(real-time) 마케팅을 구현 중
- 결제 및 배송 단계: 결제의 간편화, 배송의 효율화를 목적으로, 유통 기업은 VR 페이, 무인 결제 시스템 등으로 소비자가 구매 결정 후 결제하기까지 걸리는 시간을 최소화하는 데 중점을 두고 있음. 드론을 비롯한 각종 로봇 기술을 상품 배송에 활용하면서 기업의 인건비 절감 효과와 더불어 배송 속도를 향상시키고 배송의 정확도를 개선
- 고객 관리 단계: 반복 구매 및 재구매를 간편하게 함으로써 자사 플랫폼에서 고객의 지속적인 구매를 유도하고 있으며, IoT 및 블록체인 기반의 제품 이력 추적 시스템을 배송 및 반품 과정에 접목해 소비자의 신뢰도를 높이는 동시에 브랜드 이미지 제고 효과를 통해 고객 로열티를 고취

[ KPMG의 유통단계 레벨 성장 개요도 ]



\* 출처: 삼성KPMG, 2017

## 나. 생태계 기술 동향

### (1) 해외 플레이어 동향

- (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 기사에 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 남는 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

## (2) 국내 플레이어 동향

### ◎ 스마트 물류 대기업

- (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

◎ 스마트 물류 중소기업

- (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, 검사일 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 창고운동을 용이하게 하며, 이노비즈 기업 인증을 보유

## 다. 국내 연구개발 기관 및 동향

### (1) 연구개발 기관

[ 유통·물류 빅데이터 구축 및 분석 분야 주요 연구조직 현황 ]

기관	소속	연구분야
한국교통연구원	물류기술연구센터 화물운송시장 정보센터 우수물류기업 인증센터 동북아 북한교통 연구센터	<ul style="list-style-type: none"> <li>• 국내외 택배업계 빅데이터 등 첨단기술 활용 기술</li> <li>• 차세대형 물류시설 동향 전망</li> <li>• 물류 가치사슬 변화 연구</li> </ul>
국토교통부, 해양수산부	국토도시실 주택토지실 건설정책국 교통물류실 항공정책실 도로국 철도국	<ul style="list-style-type: none"> <li>• 국가물류시행 계획</li> <li>• 고부가가치 물류산업, 물류 신산업 육성</li> <li>• 스마트 물류기술 개발</li> <li>• 해운산업 장기발전 계획</li> </ul>
농림축산식품부	농림부 농수산부 농림수산부 해양수산부 농림수산식품부	<ul style="list-style-type: none"> <li>• 농산물 시장 물동량/가격 예측 및 물동량 최적분산을 위한 지능형 알고리즘 개발 및 검증</li> <li>• 물량분산 지능형 알고리즘이 탑재된 유통업무 모니터링 및 관리 시스템 개발</li> <li>• 개발된 지능형 유통업무 관리 시스템의 현장 적용 및 검증</li> </ul>

### (2) 기관 기술개발 동향

#### 한국교통연구원

- 제조·유통·물류 산업은 4차 산업혁명의 영향으로 산업간 영역 파괴 현상이 심화되면서 단순한 융합이 아닌 경계가 무너진 신산업이 다수 등장
- 사람 중심의 4차 산업혁명 구현을 목표로 기술뿐만 아니라 우리나라 산업여건에 적합한 국가적 대응전략을 수립
- 관련 R&D 과제, 인공지능 기반의 미래교통운영 기반기술 개발 및 활용(총 연구기간: '18.7.~'22.12. (54개월))
- 관련 R&D 과제, 사업용 차량을 이용한 도로교통 정보 수집 및 활용기술 개발(총 연구기간: '18.7.~'20.12. (30개월))
- 관련 R&D 과제, 국가교통체계 혁신기술개발사업 기획(총 연구기간 : '19.10. ~ '20.10. (12개월))

□ 국토교통부/해양수산부

- ‘국토교통혁신성장’의 핵심 정책과제 발표 후 2019년 6월 국토교통부는 산업 지원체계·성장기반·시장질서 혁신을 위한 물류 산업 방안 제시 및 과제 공고
- 물류 4.0 시대 도래에 따른 물류 산업 구조변화와 미래 전망을 통하여 IT·유통업·제조업 등과의 융복합 촉진과 선도적 물류 산업 발전 중장기 전략 및 지원방안을 마련

□ 한국전자통신연구원

- 비대면 접수/배달 서비스 및 본인인증 방안 연구
- 분류작업 노동부하 감소를 위한 차세대 자동구분 시스템 개념 설계 및 개발
- IoT활용 기반 스마트 배송시스템 플랫폼 기술 개발
- 우편물배송 사륜 전기차 구조변경 등 기술 규칙 도출 및 빅데이터 기반 배송시스템과 IoT 기반 스마트 배송시스템을 통합한 수요자 맞춤형 통합배송시스템 플랫폼 기술 개발
- 국내 우정IT 기업의 해외 수출 지원을 위한 협력 강화

□ 산업통상자원부

- 빅데이터 서비스 플랫폼 프로토타입 구현
- 생산 및 물류 운영 최적화를 위한 스케줄링 알고리즘 개발
- 제품출하 납기 지연 최소화를 위한 자동창고 운영(할당/적재)정책 수립
- 자동창고 운영최적화 평가를 위한 시뮬레이션 기반의 자동창고 모델 설계

□ 미래창조과학부

- 물류 및 교통분야에서 발생하는 빅데이터의 처리/분석 및 플랫폼에 집중하여 국내 빅데이터 기술의 국제적 경쟁력 상승
- 물류 및 교통 빅데이터 분석 및 플랫폼 기술 개발을 통해 차세대 성장 시장인 빅데이터 시장 주도권 확보
- 물류 교통 빅데이터 분석을 통한 다각적 분석 도출
- 오픈소스 기반 기술을 우선적으로 연구하여 물류교통 플랫폼 기술성과의 수준을 향상

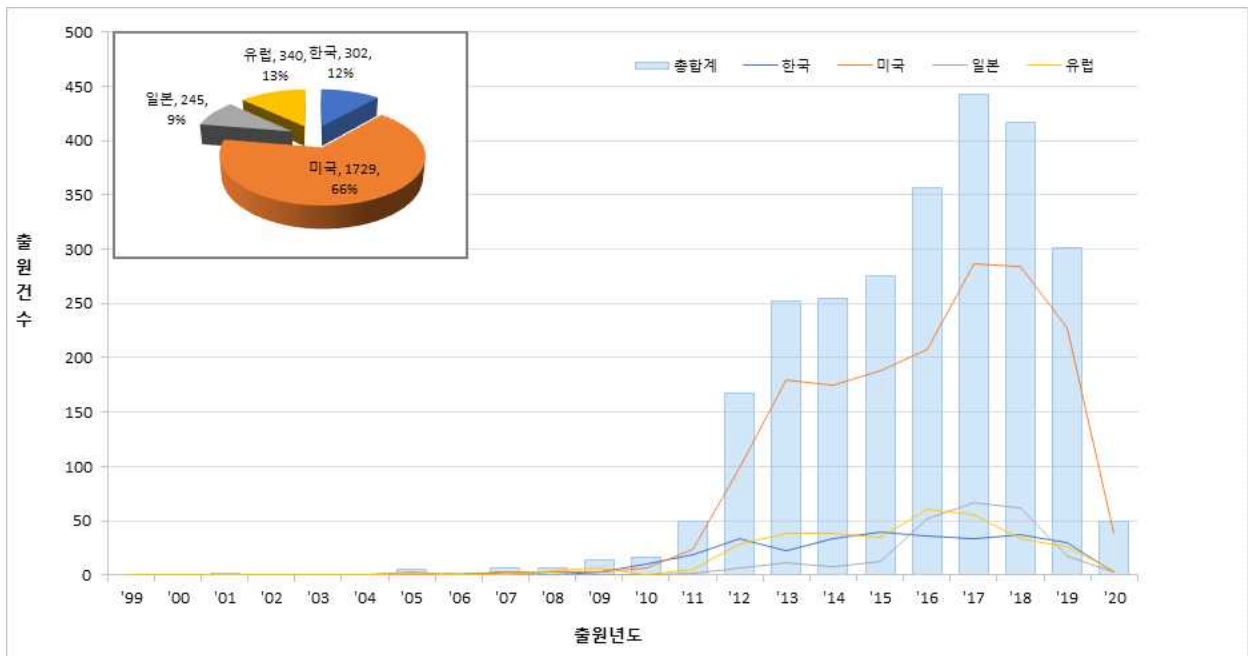
## 4. 특허 동향

### 가. 특허동향 분석

#### (1) 연도별 출원동향

- 유통·물류 빅데이터 구축 및 분석 시스템은 '12년부터 급격한 성장을 보임
  - 각 국가별로 살펴보면 미국이 가장 활발한 출원활동을 보이고 있음
- 국가별 출원비중을 살펴보면 미국이 전체의 66%의 출원 비중을 차지하고 있어, 최대 출원국으로 유통·물류 빅데이터 구축 및 분석 시스템 분야를 리드하고 있는 것으로 나타났으며, 유럽은 13%, 한국은 12%, 일본은 9% 순으로 나타남

[ 유통·물류 빅데이터 구축 및 분석 시스템 연도별 출원동향 ]

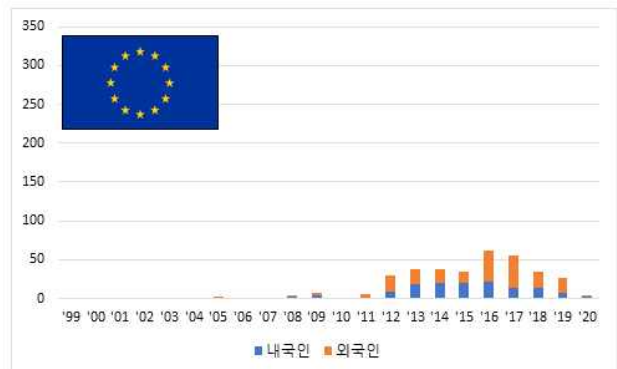
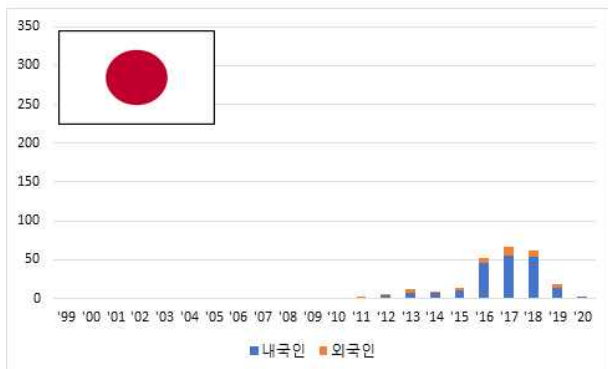
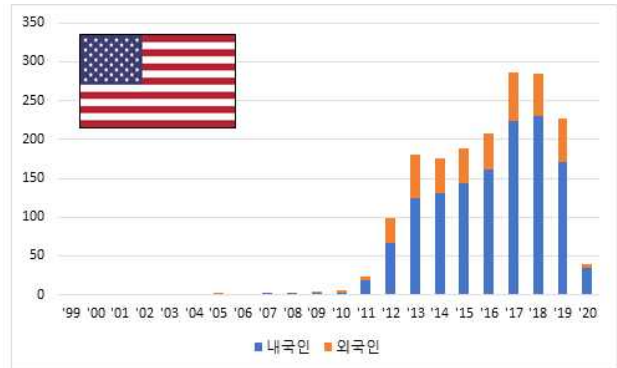
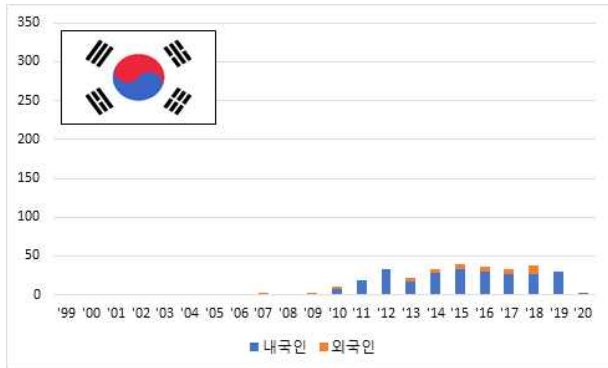




## (2) 국가별 출원현황

- 한국의 출원현황을 살펴보면, '11년부터 해당 기술의 출원이 완만히 증가하는 추세
  - 내국인 위주의 출원이 진행되고 있음
  - 한국 기술의 양적 흐름은 유럽과 상당히 유사
  - 미국의 출원 수에 비해 17% 정도의 수준을 보임
- 미국의 출원현황을 살펴보면 분석구간 초기부터 전체 특허기술의 출원 증감 흐름에 영향을 주고 있는 것으로 나타남. 미국의 경우, 한국에 비해 외국인의 비중이 큰 것으로 나타남
- 유럽의 출원현황을 살펴보면 한국과 유사한 동향을 보임
- 일본의 출원현황은 살펴보면, '15년 이후 급격히 증가하며, 미국과 유사한 동향을 보임

[ 국가별 출원현황 ]



### (3) 기술 집중도 분석

□ 전략제품에 대한 최근 기술 집중도 분석을 위한 구간별 기술 키워드 분석 진행

- 전체 구간(1999년~2020년)에서 Computer Program Product, Mobile Device, Computer Implemented 등 키워드가 다수 도출
- 최근 구간 분석 결과, 최근 1구간(2012년~2015년)과 비교할 때, 2구간(2016년~2020년)에서 대부분의 키워드가 비슷하여 큰 변화가 없음을 알 수 있었으나 Sensor Data 키워드가 많이 등장하는 것으로 보아 유통물류 빅데이터 구축 및 분석 시스템 분야에도 위치나 물류에 대한 센싱 관련 연구개발이 활발한 것으로 추정

#### [ 특히 키워드 변화로 본 기술개발 동향 변화 ]

#### 전체구간(1999년~2020년)



- Computer Program Product, Mobile Device, Computer Implemented, Rfid Tag, Computing Device, Inventory Management, Real Time, Sensor Data, Supply Chain, User Interface

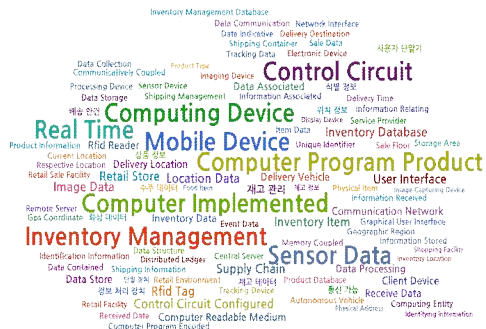
#### 최근구간(2012년~2020년)

##### 1구간(2012년~2015년)



- Computer Program Product, Mobile Device, Rfid Tag, Computer Implemented, Computing Device, Real Time, Inventory Data, Inventory Management, Supply Chain, Communication Network

##### 2구간(2016년~2020년)

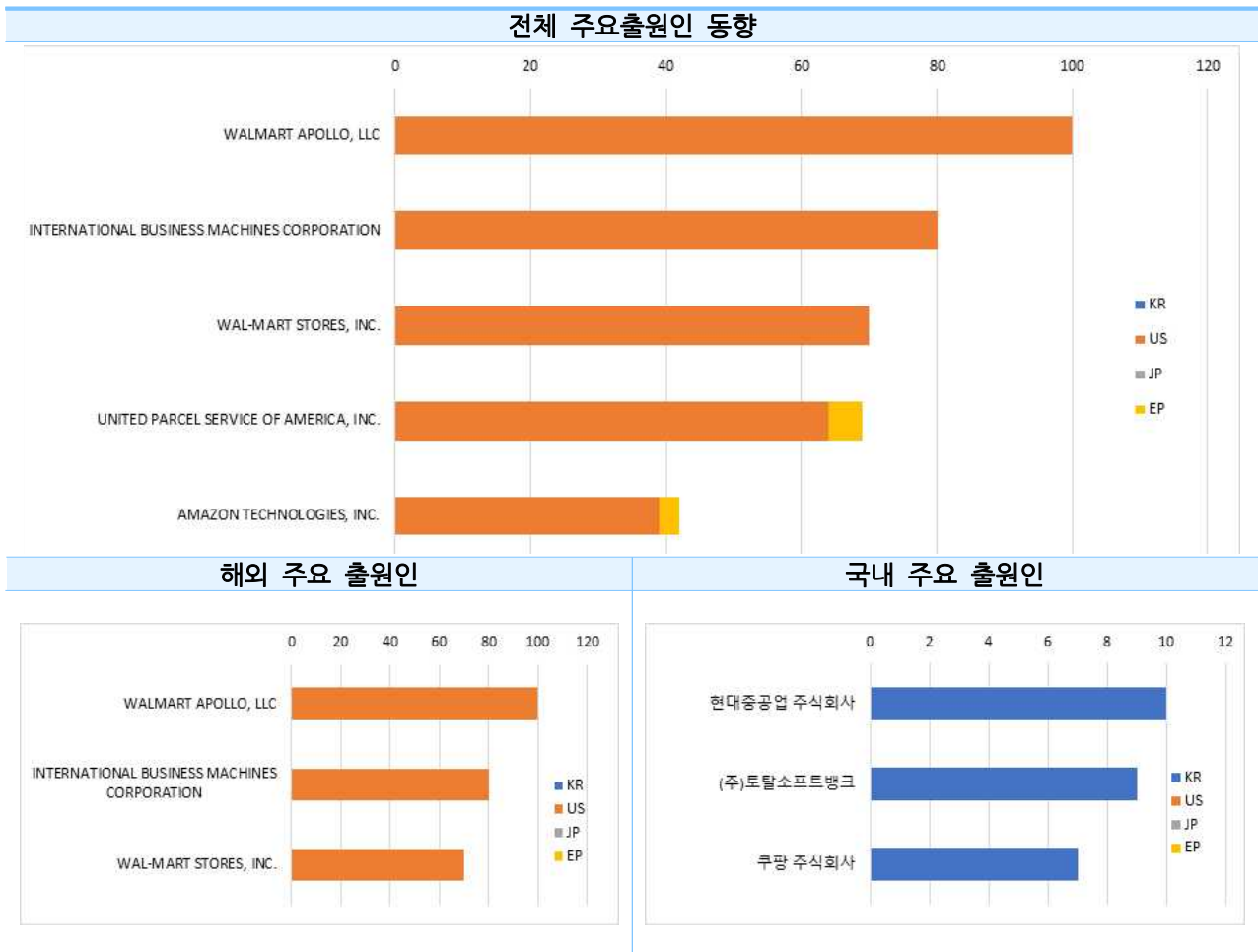


- Computer Program Product, Computer Implemented, Mobile Device, Inventory Management, Computing Device, Sensor Data, Real Time, Control Circuit, Rfid Tag, Supply Chain

## 나. 주요 출원인 분석

- 유통·물류 빅데이터 구축 및 분석 시스템의 전체 주요출원인을 살펴보면, 주로 미국 국적의 출원인이 다수 포함되어 있는 것으로 나타났으며, 제 1 출원인으로는 미국의 WALMART APOLLO, LLC인 것으로 나타남
  - 제 1 출원인인 WALMART APOLLO, LLC의 출원은 미국에 집중된 경향을 보임
- 유통·물류 빅데이터 구축 및 분석 시스템 관련 기술로 유통망을 보유한 대기업에 의한 출원이 대다수를 차지
  - 국내에서는 대기업, 중소기업(개인)의 활발한 출원이 이루어짐

[ 유통·물류 빅데이터 구축 및 분석 시스템 주요출원인 ]

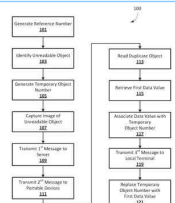
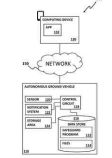
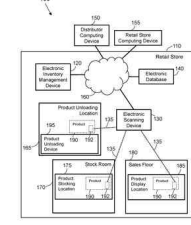
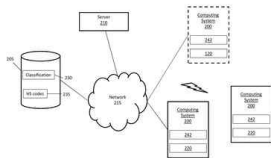
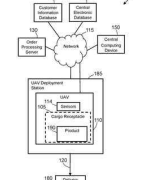


(1) 해외 주요출원인 주요 특허 분석

◎ WALMART APOLLO, LLC

- WALMART APOLLO, LLC은(는) 미국에 본사는 둔 유통 전문 기업으로 유통·물류 빅데이터 구축 및 분석 시스템과 관련하여, 전자 스캐닝 장치 및 데이터 베이스를 통한 물류 식별 및 관리 기술에 특화된 특허를 다수 출원. 그 중 등록된 특허는 33건 보유
  - 주요 특허들은 스캐닝 장치를 통한 물류에 대한 식별 및 분류하고, 무인 장치를 통한 물품을 고객에게 전달 하는 기술 관련 특허를 다수 출원하는 것으로 파악

[ WALMART APOLLO, LLC 주요특허 리스트 ]

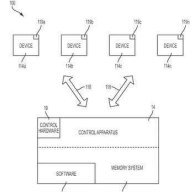
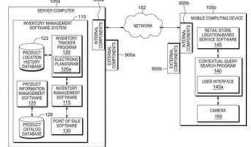
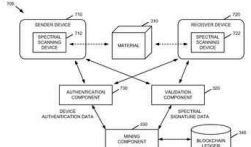
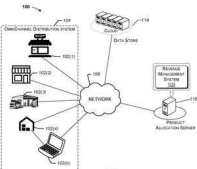
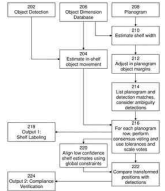
등록번호 (출원일)	명칭	기술적용분야	도면
US10515338 (2019.07.12)	Systems, devices, and methods for machine reading	컴퓨터-판독가능 매체들이 물건들의 머신 판독가능 특징들을 판독하고, 로컬 단말은 판독 불가능한 객체에 대응하는 임시 객체 번호를 생성하고, 판독 불가능한 객체를 촬영하는 기술	
US10445686 (2019.01.02)	Systems and methods for enabling delivery of commercial products to customers	AGV(자율적인 지상 차량)를 통해 경로 상에 고객에게 상업적 상품의 전달하기 위한 것으로, AGV는 센서들에 통신가능하게 결합된 제어 회로 포함하는 기술	
US10496955 (2018.09.14)	Systems and methods for identifying and remedying product mis-shipments to retail stores	전자 스캐닝 장치, 전자 데이터베이스 및 전자 재고 관리 장치에 대한 기술	
US10769585 (2018.09.11)	Systems and methods for automated harmonized system (hs) code assignment	공급자의 품목 견적을 분류 데이터베이스와 결합하는 데이터 저장 시스템 포함하는 HS 코드 할당을 위한 기술	
US10586202 (2018.07.30)	Systems and methods for validating products to be delivered by unmanned aerial vehicles	무인 항공기(aerial vehicle)들을 통해 고객들에게 전달될 제품들을 검증하기 위한 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ IBM(International Business Machines)

- IBM은 미국의 다국적 기업으로, 블록체인을 통한 제품의 공급망을 관리하기 위한 특허를 보유하고 있으며, 유통·물류 빅데이터 구축 및 분석 시스템과 관련한 특허를 미국에 출원
  - 유통·물류 빅데이터 구축 및 분석 시스템 관련하여 출원을 진행한 특허 모두 미국에 출원한 것으로, 자국 내 출원 성향이 높은 것으로 사료
  - 모바일 장치를 활용한 물품 관리 기술과 관련하여 80건의 미국 출원을 진행하였으며, 그 중 등록된 특허는 31건으로 파악

[ IBM 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
US10671972 (2019.02.26)	Automated zone location characterization	사용자의 모바일 장치를 통해 특정 구역 내 제품의 위치 및 정보를 데이터를 분석하여 정보를 제공	
US10223668 (2018.03.23)	Contextual searching via a mobile computing device	쿼리 개체 및 전자 플랜노그램은 잘못 배치된 제품을 식별하고, 고객 요구에 맞게 적절한 재고 수준을 유지하며, 소매점 내의 제품 위치를 추적하는 기술	
US10685323 (2017.12.11)	Blockchain ledgers of material spectral signatures for supply chain integrity management	블록체인을 통한 제품과 연관된 스펙트럼 데이터 통해 공급망 관리 기술	
US10423923 (2016.09.13)	Allocating a product inventory to an omnichannel distribution supply chain	옴니채널 공칭 수요 예측 모델에 기초하여 타겟 상품 제품에 대한 할당이 생성되고, 옴니채널 분배 시스템을 위한 목표 상품 할당의 최악의 시나리오가 식별하고, 이를 방지하는 기술	
US10552787 (2016.09.04)	Store planogram to image registrations	이미지 등록에 자동으로 플랜노그램들을 준수하기 위해 스토어 선반들을 체크하기 위한 자동화된 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ WAL-MART STORES, INC.

□ WAL-MART STORES, INC.는 미국에 본사를 둔 유통 전문 기업으로 등록특허 1건 보유

[ WAL-MART STORES, INC. 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
US9830573 (2013.05.16)	Retail inventory management services through personal mobile devices	소매점 작업자의 개인 스마트폰을 통해 통합 재고 관리 서비스에 소매점 제어 가능한 액세스를 제공하는 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

(2) 국내 주요출원인 주요 특허 분석

◎ 현대중공업 주식회사

- 현대중공업 주식회사는 '11년도 6월부터 출원을 시작하여 아직까지 등록된 특허는 없는 것으로 파악됨
  - 현대중공업 주식회사의 공개특허를 대상으로 분석한 결과, 유통·물류 빅데이터 구축 및 분석 시스템 기술에 있어서, 강재 운송 시스템 및 강재의 재고를 예측하는 기술과 관련된 특허를 10건 출원

[ 현대중공업 주식회사 주요특허 리스트 ]

공개번호 (출원일)	명칭	기술적용분야	도면
KR2017-0073785 (2015.12.18)	가상 컨테이너 적재 검사 방법 및 가상 컨테이너 적재 검사 시스템	크레인을 이용하지 않고 컨테이너 적재 검사를 수행할 수 있는 가상 컨테이너 적재 검사 방법 및 가상 컨테이너 적재 검사 시스템	
KR2014-0062934 (2012.11.15)	모바일을 이용한 강재 운송시스템	강재의 차량 적재 작업이 진행되는 동안 모바일 장치에 입력되는 강재의 자재정보와, 운송정보 등이 자재 관리시스템으로 자동으로 전송	
KR2014-0054897 (2012.10.30)	강재 적정 재고 예측 시스템 및 방법	선박을 제조하는 조선소에서 선박 생산에 소요되는 강재의 적정 재고를 신속하고도 경제성 있게 예측하도록 하는 강재 적정 재고 예측 시스템	

\* 공개특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출



◎ (주)토탈소프트뱅크

□ (주)토탈소프트뱅크는 한국의 물류 패키지 소프트웨어 전문 기업으로, 유통·물류 빅데이터 구축 및 분석 시스템과 관련된 특허를 9건 출원하였으며, 9건의 특허 중 등록된 특허는 6건을 보유

[ (주)토탈소프트뱅크 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
KR2061788 (2015.12.17)	클라우드 서비스를 이용한 적하계획 통합 제공 방법	선박의 컨테이너 하역을 위해 필요한 적하계획 목록을 클라우드 서비스를 통해 상호 공유하여, 컨테이너의 적하와 관련된 사용자측 단말기들이 소정의 적하계획을 통합적으로 제공받는 기술	
KR1783840 (2015.09.24)	부분적 멀티 쓰레드 방식을 이용한 경합관리자 시스템 및 경합관리 방법	경합관리자 개념을 도입한 부분적 멀티 쓰레드 방식을 이용한 경합관리자 시스템 및 경합관리 방법	
KR1644104 (2014.12.19)	컨테이너 터미널의 시뮬레이션 평가 시스템 및 방법	컨테이너 터미널 운영에 필요한 작업 계획, 실행 및 관리를 수행하는 터미널 운영서버에 구축된 기능 및 프로그램들을 가상의 시뮬레이션 서버와 함께 공유함으로써, 컨테이너 터미널의 시뮬레이션 평가 기술	
KR1690276 (2014.12.11)	컨테이너 터미널 내 장비들의 스케줄 조정 방법	컨테이너 터미널 내 각종 물리적 객체정보, 각 장비들의 작업 통계정보 및 각 장비들의 동작정보 등을 바탕으로 하여 각 장비들이 스케줄을 예측하여, 조정하는 기술	
KR1409907 (2013.08.23)	컨테이너 터미널에서의 작업 통제용 메시지 관리방법	하나의 작업 메시지를 통해서 복수의 클라이언트와 관리자가 동일한 정보를 제공받을 수 있도록 지원하여 컨테이너 터미널에서의 작업 통제용 메시지를 관리하는 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출



◎ 쿠팡 주식회사

□ 쿠팡 주식회사는 한국의 온라인 쇼핑몰 기업으로, 유통·물류 빅데이터 구축 및 분석 시스템 기술과 관련된 특허를 7건 출원하였으며, 등록 특허 1건 보유

[ 쿠팡 주식회사 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
KR1735018 (2015.11.09)	배송 서비스 제공 장치 및 방법	구매자의 주문 정보를 기반으로 프로세서는 상기 최우선순위의 물류센터로부터 상기 주문 정보에 따른 상품을 제공하는 기술	<pre> graph TD     S101[주문 정보 수신 S101] --&gt; S103[주문 정보에 따른 상품을 제공할 수 있는 물류센터 추출 S103]     S103 --&gt; S105{복수의 물류센터들 추출? S105}     S105 -- NO --&gt; S109[최우선순위 물류센터 선택 S109]     S105 -- YES --&gt; S107[소정 기준에 따라 복수의 물류센터들간의 순위 결정 S107]     S107 --&gt; S109             </pre>

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

## 다. 기술진입장벽 분석

### (1) 기술 집중력 분석

- 유통·물류 빅데이터 구축 및 분석 시스템 관련 기술에 대한 시장관점의 기술독점 현황분석을 위해 집중률 지수(CRn: Concentration Ratio n, 상위 n개사 특허점유율의 합) 분석 진행
  - 상위 4개 기업의 시장점유율이 0.12로 유통·물류 빅데이터 구축 및 분석 시스템 분야에 있어서 독과점 정도는 낮은 수준으로 판단
  - 국내 시장에서 중소기업의 점유율 분석결과 0.75로 해당 기술에 대하여 중소기업이 진입하여 활동하고 있는 것으로 파악

[ 주요출원인의 집중력 및 국내시장 중소기업 집중력 분석 ]

주요 출원인 집중력	주요출원인	출원건수	특허점유율	CRn	n
	WALMART APOLLO, LLC(미국)	100	3.8%	0.04	1
	INTERNATIONAL BUSINESS MACHINES CORPORATION(미국)	80	3.1%	0.07	2
	WAL-MART STORES, INC.(미국)	70	2.7%	0.10	3
	UNITED PARCEL SERVICE OF AMERICA, INC.(미국)	69	2.6%	<b>0.12</b>	<b>4</b>
	AMAZON TECHNOLOGIES, INC.(미국)	42	1.6%	0.14	5
	OBIC CO LTD(일본)	35	1.3%	0.15	6
	UNITED STATES POSTAL SERVICE(미국)	29	1.1%	0.16	7
	SAP SE(독일)	18	0.7%	0.17	8
	FEDEX CORPORATE SERVICES, INC.(미국)	17	0.6%	0.18	9
	ACCENTURE GLOBAL SOLUTIONS LIMITED(아일랜드)	15	0.6%	0.18	10
	<b>전체</b>	<b>2616</b>	<b>100%</b>	<b>CR4=0.12</b>	
국내시장 중소기업 집중력	출원인 구분	출원건수	특허점유율	CRn	n
	중소기업(개인)	206	74.9%	0.75	
	대기업	40	14.5%		
	연구기관/대학	29	10.5%		
	<b>전체</b>	<b>275</b>	<b>100%</b>	<b>CR중소기업=0.75</b>	

**(2) 특허소송 현황 분석**

- 유통·물류 빅데이터 구축 및 분석 시스템 관련 기술 진입 장벽에 대한 분석을 위해 특허소송을 이력 검토
  - 2016년 2월 캘리포니아 중앙지방법원에 원고 Secured Mail Solutions LLC와 Universal Wilde, Inc.간의 광역네트워크를 통해 전자메일을 전송하는 시스템에 대한 특허 침해소송이 종결
  - 2015년 8월 캘리포니아 중앙지방법원에 원고 FOURSEVENS, LLC 와 Eclipse IP, LLC간의 온라인 주문 시스템의 전자 메시지 시스템에 대한 특허 침해소송이 종결
  - 2015년 7월 캘리포니아 중앙지방법원에 원고 ELEVENGEAR LLC 와 Eclipse IP, LLC.간의 온라인 주문 시스템의 전자 메시지 시스템에 대한 특허 침해소송이 종결
  - 온라인을 통해 전자 메시지(문자) 전송하는 것과 관련하여 기업 간에 특허소송을 진행하는 등 국내기업이 미국시장에 진입하는 경우, 진입장벽으로 작용할 수 있음

**[ 유통·물류 빅데이터 구축 및 분석 시스템 관련 특허소송 현황 ]**

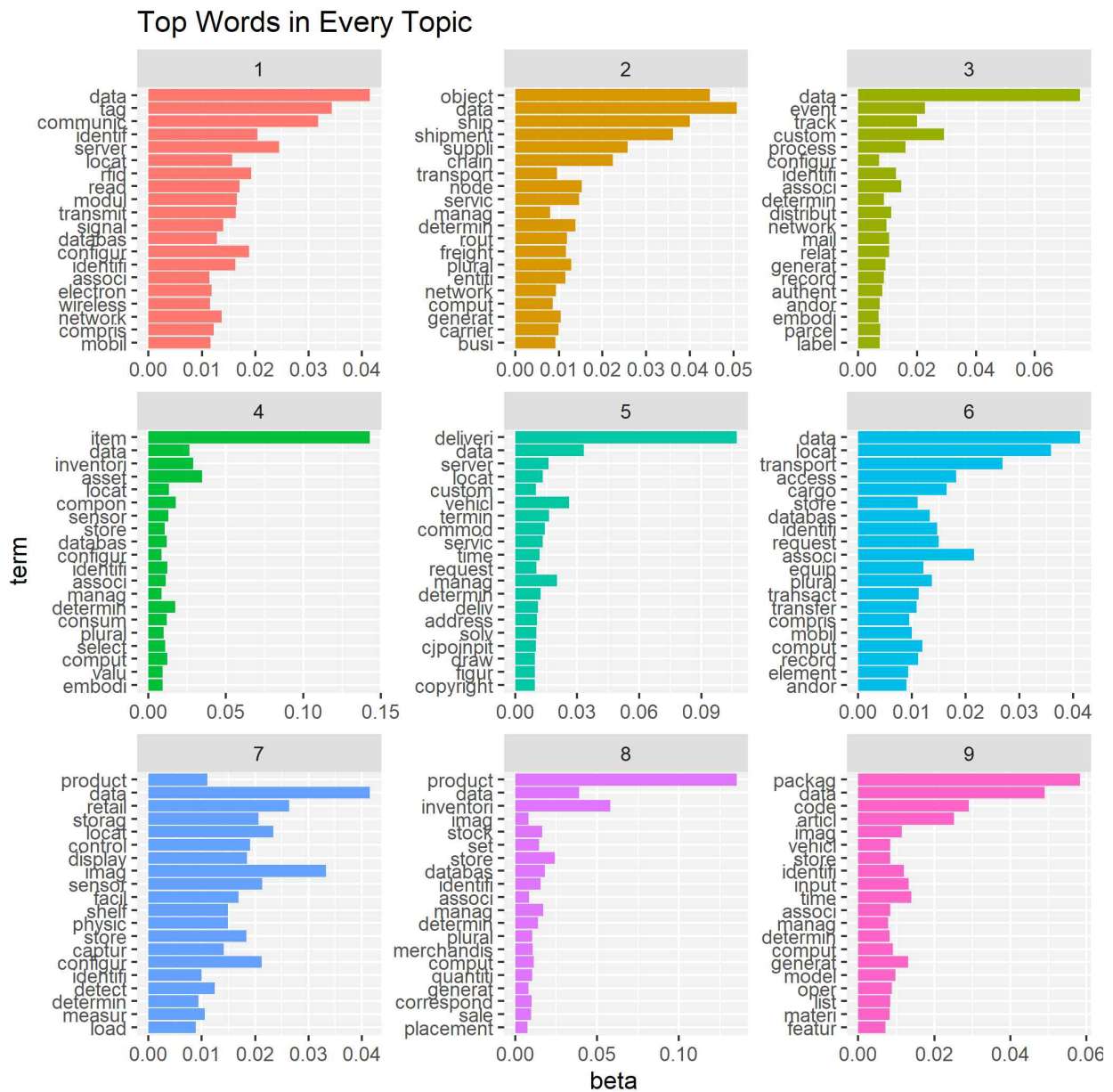
1	US9105002 (2015.08.11)	<b>명칭</b>	<b>출원인</b>	<b>원고 v. 피고</b>
		System and method for providing information to a recipient of a physical mail object	Secured Mail Solutions, LLC	Secured Mail Solutions LLC v. Universal Wilde, Inc.
		<b>대상제품명</b>	<b>소제기일</b>	<b>소송종료일</b>
		System and method for generating, affixing, storing, and processing mail identification data - UW uses software and hardware	2015.09.25	2016.02.16
2	US9019130 (2015.04.28)	<b>명칭</b>	<b>출원인</b>	<b>원고 v. 피고</b>
		Notification systems and methods that permit change of time information for delivery and/or pickup of goods and/or services	Eclipse IP, LLC	FOURSEVENS, LLC v. Eclipse IP, LLC
		<b>대상제품명</b>	<b>소제기일</b>	<b>소송종료일</b>
		Electronic messaging features of online ordering system	2015.05.22	2015.08.12
3	US9019130 (2015.04.28)	<b>명칭</b>	<b>출원인</b>	<b>원고 v. 피고</b>
		Notification systems and methods that permit change of time information for delivery and/or pickup of goods and/or services	Eclipse IP, LLC	ELEVENGEAR LLC v. Eclipse IP, LLC
		<b>대상제품명</b>	<b>소제기일</b>	<b>소송종료일</b>
		www.elevengear.com	2015.05.20	2015.07.31

## 5. 요소기술 도출

### 가. 특허 기반 토픽 도출

- 2,223개의 특허의 내용을 분석하여 구성 성분이 유사한 것끼리 클러스터링을 시도하여 대표성이 있는 토픽을 도출

[유통 물류 빅데이터 구축 및 분석 시스템에 대한 토픽 클러스터링 결과]



## 나. LDA<sup>12)</sup> 클러스터링 기반 요소기술 도출

[ LDA 클러스터링 기반 요소기술 키워드 도출 ]

No.	상위 키워드	대표적 관련 특허	요소기술 후보
클러스터 01	data, tag, communicate, server, identification, rfid, configure, read, module, transmit	<ul style="list-style-type: none"> <li>• System and method for piece picking or put-away with a mobile manipulation robot</li> <li>• Autonomous mobile bin storage and retrieval system</li> </ul>	배송경로 최적화 분석 기술
클러스터 02	data, object, ship, shipment, supplies, chain, node, service, determine, plural	<ul style="list-style-type: none"> <li>• A method, computer program and system for detecting a discrepancy arising from a mismatch between incoming and outgoing flows of resource</li> <li>• Device for allocating ships, method for allocating ships, program and recording media thereof</li> </ul>	배송경로 최적화 분석 기술
클러스터 03	data, custom, event, track, process, associate, identification, distribute, mail, relate	<ul style="list-style-type: none"> <li>• Systems and method for tracking enterprise events using hybrid public-private blockchain ledgers</li> <li>• System for capturing verified consumption profiles for food items</li> </ul>	상품구매 및 고객 데이터 연관성 분석 기술
클러스터 04	product, stock, store, report, bom, package, supplier, node, message, transport	<ul style="list-style-type: none"> <li>• System and method for inventory sharing in a laboratory management system</li> <li>• Inventory management system</li> </ul>	상품 정보 빅데이터 구축 기술
클러스터 05	delivery, data, vehicle, manage, terminal, server, commodify, location, service, determine	<ul style="list-style-type: none"> <li>• Method for processing confidential document</li> <li>• Delivery management system</li> </ul>	배송경로 최적화 분석 기술
클러스터 06	data, location, transport, associate, access, cargo, request, identification, plural, database	<ul style="list-style-type: none"> <li>• System and method for an internet-enabled marketplace for commercial warehouse storage and services</li> <li>• Systems and methods for distributed ledger-based floorplanning</li> </ul>	스마트 물류 관리 시스템 기술
클러스터 07	data, image, retail, location, sensor, configure, storage, control, display, store	<ul style="list-style-type: none"> <li>• Assigning boxes to storage bays based on available spaces</li> <li>• Out of stock sensor</li> </ul>	실시간 데이터 수집 기술
클러스터 08	product, inventory, data, store, database, manage, stock, identification, set, determine	<ul style="list-style-type: none"> <li>• Intelligent warehousing management method, apparatus, system and unmanned intelligent warehousing device</li> <li>• Method and system for validating a device that uses a dynamic identifier</li> </ul>	스마트 물류 관리 시스템 기술
클러스터 09	package, data, code, article, time, input, generate, identification, image, model	<ul style="list-style-type: none"> <li>• Method and system for progressive picking</li> <li>• Fast branch-based hybrid bom system and method</li> </ul>	실시간 데이터 수집 기술

12) Latent Dirichlet Allocation

### 다. 특허 분류체계 기반 요소기술 도출

- 유통·물류 빅데이터 구축 및 분석 시스템 관련 특허에서 총 1개의 주요 IPC코드(메인그룹)를 산출하였으며, 각 그룹의 정의를 기반으로 요소기술 키워드를 아래와 같이 도출

[ IPC 분류체계에 기반한 요소기술 도출 ]

IPC 기술트리		요소기술 후보
(서브클래스) 내용	(메인그룹) 내용	
(G06Q) 관리용, 상업용, 금융용, 경영용, 감독용 또는 예측용으로 특히 적합한 데이터 처리 시스템 또는 방법; 그 밖에 분류되지 않는 관리용, 상업용, 금융용, 경영용, 감독용 또는 예측용으로 특히 적합한 시스템 또는 방법	• (G06Q-010) 경영; 관리	상품구매 및 고객 데이터 연관성 분석 기술

## 라. 최종 요소기술 도출

- 산업·시장 분석, 기술(특허)분석, 전문가 의견, 타부처 로드맵, 중소기업 기술수요를 바탕으로 로드맵 기획을 위하여 요소기술 도출
- 요소기술을 대상으로 전문가를 통해 기술의 범위, 요소기술 간 중복성 등을 조정·검토하여 최종 요소기술명 확정

### [ 유통 물류 빅데이터 구축 및 분석 시스템 분야 요소기술 도출 ]

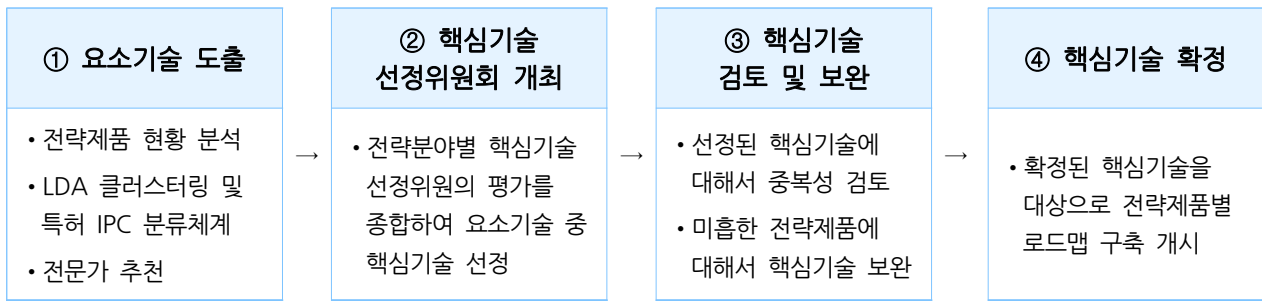
요소기술	출처
배송경로 최적화 분석 기술	특허 클러스터링, 전문가추천
상품구매 및 고객 데이터 연관성 분석 기술	IPC 기술체계, 전문가추천
지능형 상품 관리 기술	전문가추천
스마트 물류 관리 시스템 기술	전문가추천
상품 정보 빅데이터 구축 기술	특허 클러스터링, 전문가추천
서비스 시각화 기술	전문가추천
실시간 데이터 수집 기술	특허 클러스터링, 전문가추천
대용량 트랜잭션 데이터 저장 기술	특허 클러스터링, 전문가추천

## 6. 전략제품 기술로드맵

### 가. 핵심기술 선정 절차

- 특허 분석을 통한 요소기술과 기술수요와 각종 문헌을 기반으로 한 요소기술, 전문가 추천 요소기술을 종합하여 요소기술을 도출한 후, 핵심기술 선정위원회의 평가과정 및 검토/보완을 거쳐 핵심기술 확정
- 핵심기술 선정 지표: 기술개발 시급성, 기술개발 파급성, 기술의 중요성 및 중소기업 적합성
  - 장기로드맵 전략제품의 경우, 기술개발 파급성 지표를 중장기 기술개발 파급성으로 대체

#### [ 핵심기술 선정 프로세스 ]



### 나. 핵심기술 리스트

#### [ 유통·물류 빅데이터 구축 및 분석 시스템 분야 핵심기술 ]

핵심기술	개요
배송경로 최적화 분석 기술	• 분석 시스템 최적화 기술을 적용 전과 후를 비교하여 배송경로(거리, 시간, 경비 등등)가 얼마나 향상되었는지에 대한 최적화 성능 측정
상품구매 및 고객 데이터 연관성 분석 기술	• 상품 Item의 특징과 고객의 특징과 상호 비교하여 고객이 선호하는 또는 과거 고객이 주로 구매했던 성향과 얼마나 잘 적합한지를 분석하고 이에 대한 성능을 측정
지능형 상품 관리 기술	• 유통/구매/아이템 분석 등을 통한 상품 큐레이션 성능 • 지능형 관리를 통한 신규 상품 추천에 대한 고객 히팅 비율, 추천 정확도, 그 결과 실제 구매로 이어진 비율 등 • 상기 개념에 대해 기존 적용 전후를 비교
스마트 물류 관리 시스템 기술*	• 스마트 기술 적용 전후의 물류처리 비율을 측정 비교
상품 정보 빅데이터 구축 기술	• 유통·물류 분석에 필요한 유효한 빅데이터 규모/종류 및 빅데이터 처리 시스템 성능

\* 표시는 생태계 취약 기술을 의미



### 다. 중소기업 기술개발 전략

- 글로벌 플랫폼에 제공하기 어려운 유통·물류 실시간 데이터 수집 및 연계기술의 개발
- 유통·물류 산업 특화 빅데이터 기반 서비스 창출
- 플랫폼이 아닌 빅데이터를 활용한 서비스 개발과 유료 수익 모델 개발을 통한 수익 창출

### 라. 기술개발 로드맵

#### (1) 중기 기술개발 로드맵

[ 유통·물류 빅데이터 구축 및 분석 시스템 분야 중기 기술개발 로드맵 ]

유통·물류 빅데이터 구축 및 분석 시스템	물류의 과학적이고 합리적인 관리와 분산			
	2021년	2022년	2023년	최종 목표
배송경로 최적화 분석 기술				배송효율 최적화
상품구매 및 고객 데이터 연관성 분석 기술				고객-상품 연관성 분석 정확도 90%
지능형 상품 관리 기술				지능형 관리 대상 상품 관리 최적화
스마트 물류 관리 시스템 기술				물류처리 최적화
상품정보 빅데이터 구축 기술				상품 정보 빅데이터 구축 기술 개발

(2) 기술개발 목표

□ 최종 중소기업 기술로드맵은 기술/시장 니즈, 연차별 개발계획, 최종목표 등을 제시함으로써 중소기업의 기술개발 방향성을 제시

[ 유통·물류 빅데이터 구축 및 분석 시스템 분야 핵심요소기술 연구목표 ]

핵심기술	기술요구사항	연차별 개발목표			최종목표	연계R&D 유형
		1차년도	2차년도	3차년도		
배송경로 최적화 분석 기술	배송효율 향상율	-	10%	20%	배송효율 최적화	상용화
상품구매 및 고객 데이터 연관성 분석 기술	고객-상품 연관성 분석 정확도	-	70%	80%	고객-상품 연관성 분석 정확도 90%	산학연
지능형 상품 관리 기술	지능형 관리 대상 상품에 대한 처리 향상율	-	20%	30%	지능형 관리 대상 상품 관리 최적화	상용화
스마트 물류 관리 시스템 기술	물류처리효율 향상율	-	10%	20%	물류처리 최적화	기술혁신
상품 정보 빅데이터 구축 기술	상품 정보 빅데이터 구축 기술	40%	70%	100%	상품 정보 빅데이터 구축 기술 개발	산학연



전략제품 현황분석

# 스마트 헬스케어 위한 빅데이터 수집 시스템





# 스마트헬스케어를 위한 빅데이터 수집 시스템

## 정의 및 범위

- 고령화 사회로 진입한 시대에 사용자 헬스케어를 위한 데이터 수집과 메디데이터의 임상 분석을 통해 신약 개발에 소요되는 임상 기간을 단축할 수 있고, 임상3상 때 1인당 들어가는 비용도 줄일 수 있는 시스템
- 4차 산업혁명의 핵심 ICT 기술인 빅데이터, 인공지능(AI) 사물인터넷(Internet of Things, IoT), 클라우드 컴퓨팅 등을 헬스케어와 접목한 빅데이터 분야

## 전략 제품 관련 동향

시장 현황 및 전망	제품 산업 특징
<ul style="list-style-type: none"> <li>(세계) '18년 1,236억 달러 규모에서 연평균 16.2%로 성장하여 '24년 3,042억 달러 규모로 성장 전망</li> <li>(국내) '18년 약 185억 6,000만 원 규모에서 연평균 15.7%로 성장하여 '24년 445억 2,000만 원 규모로 성장 전망</li> </ul>	<ul style="list-style-type: none"> <li>ICT 핵심기술[빅데이터 및 인공지능 등]의 기반 기술로 메디데이터 수집, 정제, 분석 및 기술 융합 활용에 제한적</li> <li>스마트 헬스케어가 부상함에 따라 인공지능, 사물인터넷, 웨어러블 디바이스, 스마트폰, 클라우드 컴퓨팅 등 기존 의료시스템 밖에 있었던 디지털 기술이 의료 분야에 빠르고 광범위하게 접목 중</li> </ul>
정책 동향	기술 동향
<ul style="list-style-type: none"> <li>복지부, '첨단의료기술 신속진입 평가트랙'을 포함하는 '첨단의료기술 별도평가 실행방안'에 대한 연구를 진행 중이며, 2018년 하반기 시범사업 실시 예정 [2018.4]</li> <li>과기정통부, 의료기기 R&amp;D 범부처(과기정통부, 산업부, 복지부) 전주기 통합 사업으로 R&amp;D 사업의 효율화 추진 [2018.4]</li> <li>과기정통부, 제3차 생명공학육성기본계획(2017~2026)을 수립하여 헬스케어 산업과 바이오 연구 산업을 융합형 신산업으로 육성 및 의료 빅데이터 규제를 개선</li> </ul>	<ul style="list-style-type: none"> <li>국내 의료기관 및 기업체들이 정밀의료 서비스를 위해 기술개발을 추진 중</li> <li>EHR, PHR, 빅데이터, 인공지능 기반 의료 서비스 등 다양한 응용 분야에서 활용될 수 있는 플랫폼 기술에 대한 개발이 활발</li> <li>글로벌 ICT 기업의 웨어러블 디바이스 시장경쟁의 본격화로 헬스케어 분야에 우선적 도입과 적용이 진행중</li> </ul>
핵심 플레이어	핵심기술
<ul style="list-style-type: none"> <li>(해외) Google, IBM, Apple, Epic Systems, Philips, Baidu, Quell, Zephyr</li> <li>(국내) 삼성전자, 씨젠, 뷰노, 루닛, 아산병원, 웰트</li> </ul>	<ul style="list-style-type: none"> <li>인공지능/빅데이터 기반 스마트 헬스 응용 기술</li> <li>바이오 데이터 수집용 IoT 기술</li> <li>실시간 데이터 분석을 통한 위험상황 인지 기술</li> <li>개인용 헬스 디바이스/바이오 데이터 보호 기술</li> </ul>

## 중소기업 시장대응전략

- 글로벌 업체들의 비즈니스 전략을 면밀히 분석하고, 초기 시장을 선점할 수 있는 비즈니스 전략
- 산학연 및 대-중견-중소기업 협력에 기반해 자체적으로 진행하기 어려웠던 R&D 진행
- 정부 R&D 사업과 현장실습 프로그램 연계로 부족한 인력 확충하여 경쟁력 제고

# 1. 개요

## 가. 정의 및 필요성

### (1) 정의

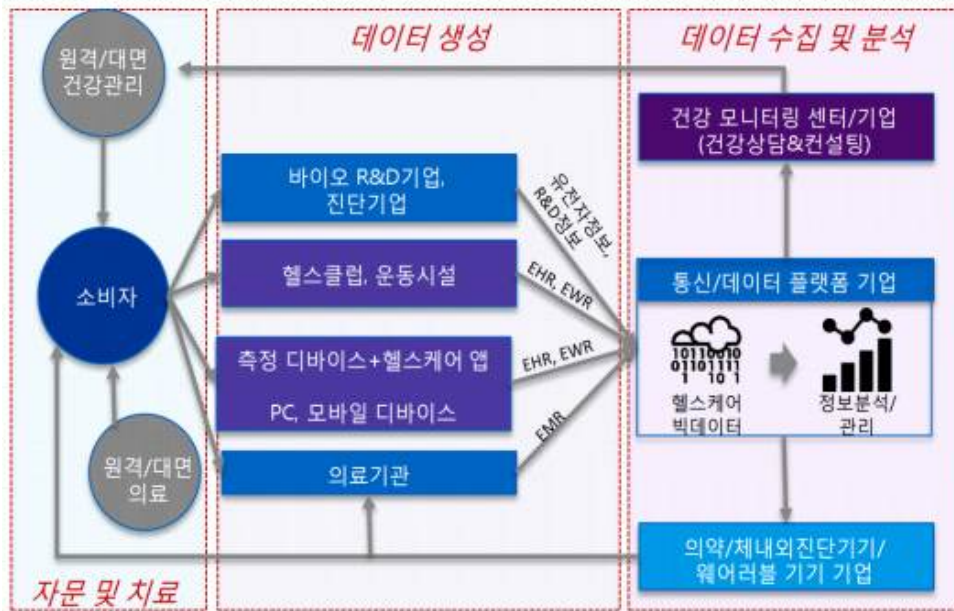
- 의료와 기술이 융합된 형태로 환자 사용자의 개인별 건강 상태를 시간과 장소의 제약 없이 실시간으로 모니터링 및 관리하고 건강정보 및 질병 상태 등을 분석하여 최적화된 맞춤형 진료를 제공하는 서비스 또는 시스템
- 기존 u-헬스의 개념이 포괄하고 있던 u-메디컬, u-실버, u-웰니스는 물론 건강관리, 영양, 운동 처방, 환자 교육 등을 포함한 용어로 최신기술을 활용하여 병원과 가정 등 언제 어디서나 환자의 상태를 지능적으로 모니터링하면서 관리하고 환자 정보와 질병 정보 등을 분석하여 실시간으로 맞춤형 서비스가 제공되는 것<sup>13)</sup>
- 4차 산업혁명의 핵심 ICT 기술인 빅데이터, 인공지능(AI) 사물인터넷(Internet of Things, IoT), 클라우드 컴퓨팅 등을 헬스케어와 접목한 분야
- 개인의 건강과 의료에 관한 정보, 기기, 시스템, 플랫폼을 다루는 산업 분야로서 건강 관련 서비스와 의료IT가 융합된 분야
  - 개인 맞춤형 건강관리서비스를 제공, 개인이 소유한 휴대형, 착용형 기기나 클라우드 병원 정보 시스템 등에서 확보된 생활습관, 신체 검진, 의료이용정보, 인공지능, 가상현실, 유전체 정보 등의 분석을 바탕으로 제공되는 개인 중심의 건강관리생태계
  - 의료 서비스와 건강관리서비스가 모두 제공되어 의료 서비스를 요구하는 환자는 물론 건강에 관심이 있는 일반인 대상의 상시적인 케어서비스와 필요에 따라 제공되는 의료 서비스를 포함

13) 헬스케어 新시장 창출을 위한 정책연구, 산업자원통상부

(2) 필요성

- 전 세계적으로 심각한 고령화 추세와 급증하는 의료비 부담에 대한 국가적 대응이 시급한 상황. 이로 인해 스마트 헬스케어가 새로운 부가가치로 부상 중
  - 의료비의 경우 한국은 지난 10년간 의료비 비중이 연평균 3.71%로 OECD 주요국 중 의료비용 부담이 가장 빠르게 증가하여 2020년 GDP 대비 12% 수준까지 확대 전망
- 4차 산업혁명(Industry 4.0)은 기술과의 융합을 기반으로 산업 시스템을 변화시켜 삶의 정의와 접근 방식을 새롭게 제시함에 따라 건강에 대한 접근 방식이 변화되었으며 스마트 헬스케어를 기반으로 하는 산업 또한 변화하는 중
  - 의료와 ICT 융합으로 의료데이터의 접근 및 확보 방식이 변화되고 있는 스마트 헬스케어는 융합된 기술의 의료기기, 서비스, App을 사용하는 환자(일반인) 개개인의 질병 관리, 건강관리, 식습관 관리 등의 다양한 형태의 서비스로 발전 중
  - 이를 기반으로 정밀의료, 원격의료 모니터링 서비스 등 개인별 맞춤형 치료서비스와 스마트 헬스케어 서비스가 종합적으로 생애주기 전반을 관리해주는 범위로 확대될 것으로 예상

[ 스마트 헬스케어 산업 구조 ]



\* 출처: ETRI 미래전략 연구소

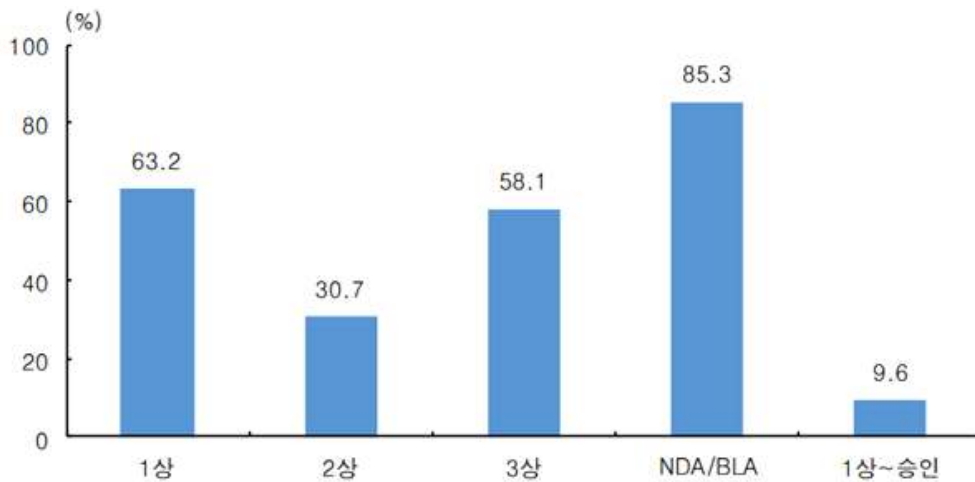
□ 스마트 헬스케어에 대한 관심이 확대되고 있는 배경은 크게 세 가지로 구분

- 의료 서비스의 패러다임이 질병이 발생한 후에 치료를 받는 치료·병원 중심에서 스스로 건강을 관리하는 예방·소비자 중심으로 변화
- 스마트기기와 센서 기술을 통해 일상에서 손쉽게 자신의 식사량이나 혈압, 운동량 등 건강 상태를 기록하고 관리하는 ‘자가 건강 측정(Quantified Self)’ 트렌드의 확산
- 기술발전으로 인하여 웨어러블 디바이스는 우리 몸에 밀착되어 지속해서 생체정보를 파악할 수 있게 만들어주고 있으며, 이는 ‘자가 건강 측정’ 트렌드를 확산시키는 요인
- 이뿐 아니라 다양한 ICT 기술, 의료기술, 빅데이터는 인공지능과 결합하여 헬스케어 산업에서의 혁신서비스를 창출

□ 또한, ICT 기술을 융합하여 사용자 헬스케어를 위한 데이터 수집과 메디데이터의 임상 데이터 수집 및 분석을 통해 신약 개발에 소요되는 임상 기간 및 비용 절감 가능

- 인체를 대상으로 후보 의약품 또는 백신에 대한 임상시험을 처음으로 수행하는 경우 일반적으로 소수의 건강한 지원자들에게 약물이 제공. 그러나 암 등의 말기 질환을 위한 치료제의 경우 해당 질환을 앓고 있는 지원자를 대상으로 임상시험이 수행되는 경우도 존재
- 신약을 판매하기 위해서는 반드시 임상 절차를 거쳐야 하는데 이게 총 4상(Phase)으로 이루어짐
- 임상 1상에서 신약 판매 허가까지의 성공 확률은 약 10% 정도로 하이 리스크-하이 리턴 구조

[ FDA 임상 시 단계별 성공 가능성 ]



\* 출처: 제약바이오 2020년 투자전략, 한국투자증권, 2019.11



□ 임상시험 실시 시기에 따른 분류 체계

- 임상 1상 : 소수의 건강한 일반인 대상으로 약물 독성 테스트를 진행하여 약물의 체내 흡수, 분포, 대사, 배설 등에 대한 자료를 수집하면서 안전성을 평가(소요기간 1~3년/ 대상 인원 20~100명)
- 임상 2상 : 환자를 대상으로 적정용량의 범위(최적의 투여량 등)와 용법을 평가하면서 약물의 유효성 여부 확인 (소요기간 2~4년 / 대상 인원 100~500명)
- 임상 3상 : 수백 명 이상의 환자를 대상으로 약물의 유효성과 안전성을 최종적으로 검증(소요기간 3~5년 / 대상 인원 1,000~5,000명)
- 임상 4상 :약물 시판 후 부작용을 추적하여 안전성을 재고하고, 추가적 연구를 시행

[ 신약 개발 과정 중 임상 연구의 위치 ]



\* 출처: 한국임상시험포럼 K-CLIC

□ 실제 임상시험을 진행하다 보면 시간과 비용면에서 여러 가지 어려운 점이 발생하게 되는데, 그 중 중요한 사항들을 빅데이터 기반의 테스트 데이터를 이용하여 그 데이터에 머신러닝 알고리즘을 적용하면 시간과 비용을 절감 가능

- 생체반응이 다양하여 같은 약에 대한 반응이 환자 간에는 물론이고 동일 환자에게서도 시간, 환경, 본인의 기분 등에 따라 달리 작용하는 변수가 존재
- 연구팀 구성이 어려움, 즉, 구성 요원마다 관심, 협조, 성의 등의 정도가 다르며, 특히 병원에서의 근무시간이 제한되고 이동이 빈번한 전공의들이 참여하는 경우 상호 협조에 문제 많음
- 시험의 계획과 시행, 그리고 자료 분석에 이르기까지 종합적인 감독과 지휘를 할 수 있는 전문지식과 리더십을 갖춘 사람을 구하기가 어려움
- 신약을 투약할 때 그 대상 선정에 있어 비록 랜덤하게 한다고 하더라도 도의적인 문제가 소지가 존재
- 임상시험은 장기간에 걸친 관찰이 필요, 특히 짜임새 있는 3상 임상시험에 해당하면 기획에만 3~6개월 소요

## 나. 범위 및 분류

### (1) 가치사슬

- 의료산업 분야에 있어서 의료보험, 발병 예측서비스, 의료기기 산업, 생활개선 서비스 등의 인간 건강에 영향을 미칠 수 있는 부분에서 활용가치가 증대될 것으로 예상
- 수집된 데이터에 기반하여 제약 분야, 이상 시험, 원격 의료 모니터링, 개인별 맞춤형 치료와 같이, 생애주기 전반을 관리해주는 서비스가 전방산업에 포함
  - 전방산업의 기술과 하드웨어의 결합이 또 다른 후방산업들의 발전을 야기. 제스처, 신체 자세, 목소리 톤, 표정과 같은 인간의 감정을 포착하기 위한 마이크 및 카메라와 같은 기타 구성요소와 함께 심장박동 및 온도를 모니터링하는 생물학적 센서가 장착된 하드웨어가 늘어나면서 향후 감성 분석 시장의 성장을 이끌 것으로 예상
  - 후방산업은 사용자로부터 빅데이터를 수집할 수 있는 헬스케어 디바이스 분야, 각종 센서를 기반으로 사용자 데이터를 수집할 수 있는 센서 분야 및 네트워크 구성을 위한 통신 인프라 등의 산업을 포함

#### [ 스마트 헬스케어의 가치사슬 ]

전방산업	스마트 헬스케어를 위한 빅데이터 수집 시스템	후방산업
의료, 헬스케어, 의료보험, 생활개선 서비스, 제약 개발, 이상 시험, 진단/예측/치료 모니터링 등의 데이터를 활용하는 전 분야	빅데이터, 인공지능 및 클라우드 관련 기술을 활용을 위한 데이터 수집 체계 시스템 영상(의료) 문서, 영상(의료) 데이터, 전자 처방전, FHIR 등의 정보교환기술 유전자 데이터의 전송, 공유 기술 등 헬스케어 IoT 지원을 위한 통신 기술 중 보안, 개인정보관리 관련 기술	의/약학, 생명공학, 의용공학 등 소재, 기기 및 생물정보 관련 연구개발 및 산업, 스마트헬스케어 디바이스 분야

- 전방산업과 후방산업의 상호관계가 기술개발의 매우 직접적인 영향을 미치기 때문에 후방산업의 기술발전이 곧 전방산업의 발전속도를 좌우
  - 각각 후방산업의 개별적인 기술 즉, 텍스트, 음성, 이미지, 생체 신호 분석기술이 더 발전하고 정교한 분석도 가능한 상태에서 상호 간의 보완이 잘 이뤄진다면 전방산업의 발전도 비약적으로 될 것이라 예상

(2) 용도별 분류

□ 스마트 헬스케어 산업은 일반적으로 하드웨어, 소프트웨어 및 서비스 분야로 분류

[ 용도별 분류 ]

분류	상세분류	설명	관련 제품 및 용도
하드웨어	제품/서비스 일체형, 단품 형태 제품	<ul style="list-style-type: none"> <li>• (개인건강관리기기) 건강관리를 위해 건강 생체신호를 측정하는 (의료)기기, 식약처 승인이 필요한 기기</li> <li>• (웨어러블 기기) 건강증진/개선을 위해 신체에 착용되어 생체신호 측정과 모니터링을 하는 기기</li> </ul>	게이트웨이
			혈당, 혈압, 심전도, 활동량 측정
			헤모글로빈 측정
			체성분/체지방 측정
			의료용 센서 삽입 스마트기기
			현장검사 기기(POCT)
부품	부품, 장치, 시약		밴드/목걸이형, 부착(패치)형, 인체 삽입형
			시약, 바이오센서
			저장 및 디스플레이 장치
소프트웨어	의료/건강관리 콘텐츠	<ul style="list-style-type: none"> <li>• (건강정보 제공 App) 일반적 의학정보, 운동정보, 영양정보 등 건강정보 제공</li> <li>• (맞춤형 건강관리 App) 개인 건강정보를 수집하여 맞춤형 건강관리 제공</li> </ul>	통신 장치
			웰니스(휴식방법, 요가, 뷰티팁 등) App
			영양관리 및 정보 제공 App
			의학적 정보(약품, 질병, 복약등) 제공 App
			개인 건강기록(PHR) App
			병원기록 관리 App
	미들웨어, 플	<ul style="list-style-type: none"> <li>• (의료정보관리 플랫폼/DB) 의료기관 의료정보 통합 저장/관리 시스템</li> <li>• (개인 건강정보관리 플랫폼/DB) 건강/의료정보 통합 저장, 관리</li> </ul>	피트니스 또는 운동 관리 App
			의료정보관리 플랫폼(EMR, EHR)
			개인 건강정보관리 플랫폼
			기타
서비스	진단서비스	<ul style="list-style-type: none"> <li>• (진단서비스) 유전자, 의료진단 서비스</li> </ul>	체외진단 서비스
			유전자/유전체 분석 서비스
	건강관리	<ul style="list-style-type: none"> <li>• (건강관리서비스) 하드웨어 기기의 건강정보 및 의료정보 분석, 건강관리서비스</li> <li>• (원격 의료 서비스) 원격으로 행해지는 의료 서비스 및 진단</li> </ul>	개인 건강검진 관리 서비스
			개인 건강기록(PHR) 관리 및 맞춤형 서비스
			노인건강관리 서비스
		건강관리 포털 서비스	
		원격상담, 원격모니터링 서비스	

## 2. 산업 및 시장 분석

### 가. 산업 분석

#### ◎ 새로운 스마트 헬스케어 시스템의 관점

- 스마트 헬스케어 산업은 의료 서비스 패러다임이 발병 후 치료가 아닌, P4(Preventive: 예방, Predictive: 예측, Personalized: 개인 맞춤, Participatory: 참여) 중심의 새로운 의료패러다임으로 전환하는 환경에서 발전되었으며 이러한 전환은 개별 환자 질병에 대한 통합적 진단 치료 및 예방에 초점을 맞춘 시스템 관점의 총체적 접근을 의미
- 기술과 의료 서비스가 융합된 산업으로 실시간으로 수집되는 개인 건강정보, 유전체 정보 전자 의무기록 등의 데이터들이 방대해짐에 따라 헬스케어 서비스가 빅데이터 기술, 인공지능 기술 등과 결합하여 정밀의료 개인 맞춤형 서비스와 같이 개별적으로 활용이 가능한 새로운 부가가치를 창출하는 산업으로 재탄생
- 스마트 헬스케어 산업은 정보통신 및 기타 최첨단 과학기술을 빠르게 융합시키는 특징을 갖고 있으며, 기존 의료기기와 달리 네트워크를 이용하여 다른 기기 또는 앱 등과 밀접하게 연결
- 스마트 헬스케어에서 디바이스가 네트워크에 연결될 수 있다는 것은 역으로 다양한 위협요소들을 내포하고 있다는 뜻
- 스마트 헬스케어 산업의 후방산업은 의/약학, 생명공학, 의용공학 등 소재, 기기 및 생물정보 관련 연구개발 및 산업이며, 전방산업은 의료 서비스 등 보건 및 의료산업, 질병 진단/예측 치료 등이 있음

[ 신약 개발 단계별 활용 방안 ]



\* 출처 : 과학기술정보통신부 제공

## ◎ 스마트헬스케어와 스마트 · ICT 기술의 접목

- 스마트기기와 센서 기술을 통해 일상에서 손쉽게 자신의 식사량이나 혈압, 운동량 등 건강 상태를 기록하고 관리하는 ‘자가 건강 측정(Quantified Self)’ 트렌드의 확산
- 스마트기기 기술과 ICT 기술이 건강관리 및 의료 서비스와 융합된 산업으로서 소프트웨어 기업, 하드웨어 제조업 서비스, 기업 정부 부처 등으로 구성
  - 소프트웨어 기업은 건강관리 관련 애플리케이션, 서비스 등에서 수집된 데이터들을 분석하며, 의료·건강정보 솔루션, 개인 건강기록 솔루션, AI 기반 분석 툴, 플랫폼 기술 등이 있음
  - 하드웨어 제조업은 건강관리 관련 데이터들을 수집하기 위해 하드웨어, 웨어러블 스마트기기 등을 제조하고 데이터 수집을 하는 개인건강기기, 웰니스기기, 통신기기, 센서 등이 있음
- 디지털 헬스케어의 발전은 웨어러블 기기, 유전자 분석, 인공지능의 적용 등으로 기존의 제약 및 의료기기 산업에 더해 보완적 가치사슬을 확대할 전망
- 그러나 헬스케어 생태계의 지배자는 여전히 기존 의료기관과 의사일 가능성이 크므로, 병원-의료보험-환자로 이어지는 중심적 가치사슬은 공고히 유지될 것
- 병원 기능의 재편 및 기술혁신으로 인한 전문병원이 출현할 가능성이 있으나, 건강보험 지불체계와 같은 제도적 요소가 헬스케어 생태계 변화에 변수로 작용할 전망
- 제약회사 및 의료기관과 같은 전통 사업자뿐만 아니라, 웨어러블 디바이스, 모바일 소프트웨어 및 통신사들과 같은 신규 사업자와의 협업이 요구될 것

## ◎ 스마트헬스케어의 부상 배경

- 의료 서비스의 전체적 패러다임 변화
  - 예방중심 의료 서비스로의 전환: 질병이 발생하면 치료를 받는 치료·병원 중심에서 스스로 건강을 관리하는 예방·소비자 중심으로 변화
- 의료데이터들의 빠른 증가
  - IDC에 의하면 의료데이터의 양이 2012년 500PB에서 2020년 25,000PB로 8년간 약 50배 증가
- 사회적 니즈의 증가
  - 고령화 및 만성질환자 증가로 의료비 부담이 가중되고 있으며, 이에 대응하여 스마트헬스케어를 통한 의료비 경감 수요 증가

## 나. 시장 분석

### (1) 세계시장

- 세계 스마트 헬스케어 시장 규모는 '18년 1,236억 달러 규모에서 연평균 성장률 16.2%로 '24년 3,042억 달러 규모로 성장 전망

[ 세계 디지털 헬스케어 시장규모 ]

(단위 : 십억 달러, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
세계시장	123.6	143.6	166.9	193.9	225.3	261.8	304.2	16.2

\* 출처: GRAND VIEW RESEARCH, 2020.06

- 스마트 헬스케어 산업은 글로벌 제약사 또는 글로벌 IT 기업 등의 다양한 업체들이 시장 성장 잠재력에 대한 큰 기대를 가지고 참여하고 있음
  - 글로벌 헬스케어 시장규모는 제약·바이오, 의료기기, 의료영상장비, 체외진단, 환자 모니터링, 디지털 헬스 분야로 나누는 것이 가능

[ 글로벌 헬스케어 분야별 시장규모 ]

(단위 : 십억 달러)

	제약·바이오	의료기기	의료영상장비	체외진단	헬스케어 IT
2020	1,355.50	448.90	33.80	87.10	130.20
2021	1,413.80	472.20	35.30	95.80	138.50
전년대비 성장률 (%)	4.3	5.2	4.4	10.0	6.4

\* 출처: 생명공학정책연구센터(2018)을 바탕으로 2020년 이후 네모아이씨지 추정

- IT, IoT 등의 기술이 발전하고 헬스케어와 접목되면서, 고령화, 1인당 의료비 지출 증가 등 현시대의 문제점에 대한 해결책으로 의료와 ICT 기술의 융합이 주목받으면서 새로운 서비스 영역 창출 기반 산업으로서 가치가 높아지고 있음
- McKinsey 조사에 의하면 2050년 선진 및 신흥시장의 65세 이상 인구 비중이 각각 26%, 14%까지 높아질 것으로 예상하며 주요 OECD 국가의 평균 인당 의료비 지출 증가율은 3.7%지만 한국의 경우는 7.7%로 예상
- IDC에 의하면 의료 데이터양이 2012년 500PB에서 2020년에는 25,000PB로 약 50배가 증가할 전망이며 폭발적으로 증가하는 의료데이터를 분석하고 활용하는 방안이 중요한 이슈로 주목

## (2) 국내시장

- 스마트 헬스케어 빅데이터 국내 시장규모는 2018년에 185.6억 원 규모로 평가되었으며, 연평균 15.7%로 성장하여 2024년까지 445.2억 원에 이를 것으로 전망됨

[ 국내 디지털 헬스케어 시장규모 ]

(단위 : 억 원, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
국내시장	185.6	214.7	248.5	287.5	332.6	384.8	445.2	15.7

\* 출처: Big Data Market Size - MarketsandmarketsSupply, GRAND VIEW RESEARCH(2020.06)를 바탕으로 세계 빅데이터 시장과 국내 빅데이터 시장을 복합적으로 고려하여 네모아이씨지 재추정

- 의료진들이 기업들과 협업해 1,500만 쪽 이상의 전문자료, 200종 이상의 의학서를 습득하여 이를 바탕으로 의료진에게 가장 적합한 치료법을 추천하는 임상 결정 지원 시스템 구축에 활용되어, 국내에서도 부산대 병원, 조선대 병원, 가천 길병원 등에서 사용하는 등 국내시장에서의 확산이 빠름
- 시장을 선도하고 있는 글로벌 기업들이 대다수 위치한 북미 지역이 가장 큰 시장점유율을 보이나, 향후 노인 인구 증가 및 가처분소득 증가, 참여 업체 수 증가 등이 기대되는 아시아 지역이 성장률은 가장 높을 것으로 전망

### 3. 기술 개발 동향

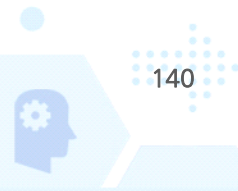
- 기술경쟁력
  - 스마트 헬스케어 를 위한 빅데이터 수집 시스템은 미국이 최고기술국으로 평가되었으며, 우리나라는 최고기술국 대비 79.5%의 기술 수준을 보유하고 있으며, 최고기술국과의 기술격차는 1.8년으로 분석
  - 중소기업의 기술경쟁력은 최고기술국 대비 69.1%, 기술격차는 2.4년으로 평가
  - 한국, 중국(79.5%)>EU(79%)>일본(70.9%)의 순으로 평가
- 기술수명주기(TCT)<sup>14)</sup>
  - 스마트 헬스케어 를 위한 빅데이터 수집 시스템은 6.57의 기술수명주기를 지닌 것으로 파악

#### 가. 기술개발 이슈

##### ◎ 기술개발 트렌드 분석

- 클라우드 컴퓨팅
  - 세계적으로 클라우드 컴퓨팅을 활용하여 원격 의료를 하고 환자의 데이터를 공유하는 등 클라우드를 헬스케어 산업에 도입하고자 노력
  - 클라우드 컴퓨팅은 데이터 분석과 저장 능력의 향상에 있어 중요한 역할을 함
  - 본 서비스는 기능에 따라, 인프라(IaaS, Infrastructure as a Service), 플랫폼(Platform as a Service), 소프트웨어(Software as a Service) 세 가지로 분류
  - 보건의료 서비스의 질과 운영 효율성의 향상, 다른 지역과의 정보 공유촉진, 관리 비용 절감을 통해 데이터 기반 헬스케어의 적용뿐만 아니라 데이터의 저장, 정보 공유, 손실 예방, 환자 정보 기록 등 다방면에 적용될 수 있음
  - 글로벌 컨설팅 업체 Forst&Sullivan에 따르면 글로벌 헬스케어 클라우드 컴퓨팅 시장은 2018년 기준 51억 3백만 달러(6조 1,899억 원)에서 연평균 24.2%씩 성장해 2023년에는 151억 8백만 달러(18조 3,260억 원)에 이를 것으로 전망
  - 병원에서만 가능했던 헬스케어가 일상생활에서도 가능해짐에 따라 폭넓은 개인 헬스케어 정보를 저장하고 처리하기 위해 클라우드 인프라의 중요성이 대두중

14) 기술수명주기(TCT, Technical Cycle Time): 특허 출원연도와 인용한 특허들의 출원연도 차이의 중앙값을 통해 기술 변화속도 및 기술의 경제적 수명을 예측





□ 디지털 센서와 웨어러블 디바이스

- 디지털 센서는 목적과 제작자에 따라 여러 가지 형태로 제작 가능한 무선 소형 감지 및 데이터 수집 장치
- 이 디지털 센서와 웨어러블 기기는 유헬스케어(Ubiquitous Healthcare)와 wBAN(wireless Body Area Network) 기술을 융합하여 신체 정보를 환자나 의사에게 전달하는 역할을 함
- 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

□ 인공지능(AI: Artificial Intelligence) 기술

- 인공지능 기술은 의료 현장에서 첨단 데이터 분석을 통해 최적의 치료법을 선택하고 환자별 치료결과 예측에도 활용될 예정
- 시스템 근간이 되는 데이터는 환자와 사용자의 개인 의료 기록과 유전정보는 물론 건강관리 기기와 각종 센서에서 얻은 실시간 데이터를 활용해 진단의 정확성을 제고 함
- 국내에서는 인공지능 기술을 사용하여 기존 의료기관에 축적된 의료영상 데이터를 분석하는 기술을 개발하고 있음

□ 유전분석 기술

- 유전체학에서 가장 실용적인 활용 분야로 대두되는 것은 약물 유전체학(pharmacogenomics)으로 약을 복용하는 사람에게 가장 잘 맞는 효과적인 표적 약물을 만드는 것이 목적

[ 스마트헬스케어의 진화 방향성 ]

구분	Tele 헬스	E 헬스	U 헬스	Smart 헬스
시기	1990년 중반	2000년	2006년	2012년 이후~
서비스 내용	원내 치료	치료 및 정보제공	치료/예방 관리	자가측정/치료/예방/복지/안전
주 Player	병원	병원	병원, ICT기업	병원, 보험사, ICT기업, 서비스기업 등
주 이용자	의료인	의료인, 환자	의료인, 환자, 일반인	의료인, 환자, 일반인
주요 시스템	병원운영 (HIS, PACS)	의무기록(EMR) 웹사이트	건강기록(EHR) 모니터링	개인건강기록(PHR) 맞춤형 서비스 인공지능 연계 서비스

\* 출처: '스마트 헬스케어산업 활성화 방안', 2015년 1월 30일, 산업통상자원부

◎ 인공지능 스마트헬스케어 산업의 동향

- 패러다임의 변화가 가능하게 된 동인은 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
- 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 방대한 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
- 기하급수적으로 늘어나는 의료데이터 및 Insight 도출을 통한 임상실험 기초 데이터 분석과 활용으로, 스마트 의료케어의 질적 수준 향상이 직면한 문제점을 다소 해결

[ 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

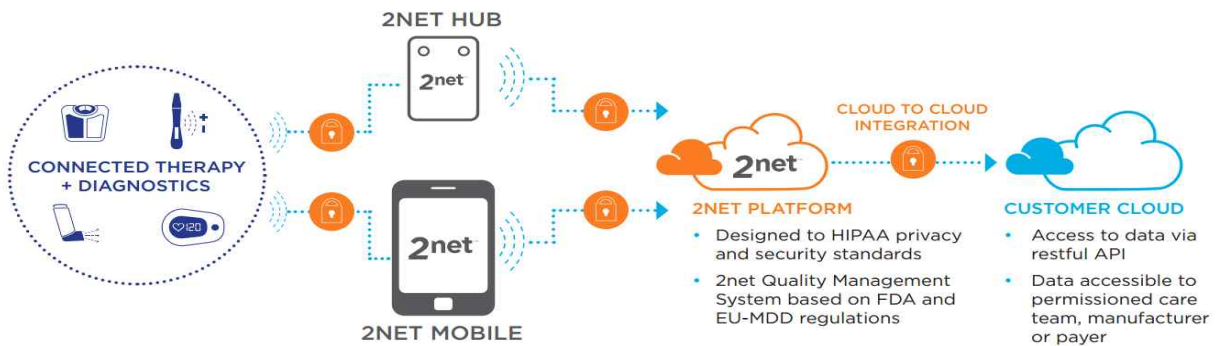
- 헬스케어 산업의 패러다임이 질병이 발생한 후에 치료를 받는 치료·병원 중심에서 스스로 건강을 관리하는 예방·소비자 중심으로 변화하면서 헬스케어 산업 내 빅데이터 분석이 더 중요해지고 있음
- 개개인이 생성해내는 방대한 양의 비구조화 데이터를 분석하게 되면 기존의 정형화된 치료방식이 아닌 개인에게 맞춰진 정밀 치료를 할 수 있게 되고 효과는 극대화될 수 있음
- 헬스케어 웨어러블 디바이스는 사용자가 자신의 건강 상태를 모니터링하기 위한 자가관리 웨어러블 디바이스와 만성질환자나 심장질환자들과 같은 특정 목적을 달성하기 위해 개발된 메디컬 웨어러블 디바이스, 소형 센서가 내장되어 생체 신호를 모니터링하는 피트니스 및 웰니스 웨어러블 디바이스로 구분함. 특히, 피트니스 웨어러블 디바이스는 인포테인먼트와 결합하여 실시간 건강정보를 활용하는 형태로 발전 중

◎ 스마트헬스케어 개발 이슈

- 스마트 헬스케어는 임상 실험 기초 데이터 및 정밀의료의 필수 요소에 있어서, 건강정보 기록, 스마트 헬스케어 기술, 빅데이터 분석기술 등 핵심 분야에 관련됨
  - 스마트 헬스케어는 정밀의료에 있어서 유전체 분야와 함께 가장 중요한 요소로 데이터의 추출(웨어러블, IoT 등)에서 관리, 분석까지의 전반적인 과정에서의 핵심 역할을 담당
  - 스마트 헬스케어는 빅데이터 분석을 통해 빠르고 정확한 진단과정을 가능하게 할 뿐만 아니라, 개인의 생활습관과 행동 분석을 통해 개개인에게 맞는 효과적인 의료 서비스의 적용에도 기여할 수 있음
  
- 단일 산업으로는 가장 큰 규모를 차지하고 있을 뿐만 아니라, 시장 변동성의 위험이 낮고, 고령화 등 선진국, 후진국을 가리지 않고 지속해서 성장하고 있는 헬스케어 산업은 글로벌 ICT 기업들에는 피하기 어려운 유혹
  - Verizon은 2012년 12월, ‘Health Platform & Partnership’을 발표하며 본격적으로 스마트 헬스케어 시장에 진출하였고, Qualcomm은 ‘2net Health’를 기반으로 스마트 헬스케어 플랫폼 사업을 본격화

[ Qualcomm의 2net 플랫폼 ]

MEDICAL-GRADE PLATFORM FOR CONNECTED THERAPY MANAGEMENT



\* 출처: Capsule Technologies, Inc (2019)

- 스마트 헬스케어 분야는 다른 산업 분야와는 다르게 미국과 기타국가의 기술 및 사업적 격차가 점차 확대되고 있다는 점은 우려할 만한 사안
  - 미국은 스마트 헬스케어를 위한 적합한 시장 여건을 갖추고 있는데, 세계에서 가장 큰 4,000조 원에 달하는 헬스케어 산업 규모, 의료보험 자간의 경쟁이 이루어지는 민간의료보험 체제로 의료보험사가 신기술 도입에 적극적이고, 지속해서 대규모 투자가 가능한 자금시장 등 헬스케어 산업에 유리한 시장환경을 가지고 있음. 심지어 고비용 의료시스템으로 인해 스마트 헬스케어의 잠재적 먹거리가 풍부하다는 점에서 모험적인 투자와 진출이 앞다투어 이루어지고 있음
  - 유럽이나 일본 등 대부분의 다른 선진국에서도 스마트 헬스케어는 아직 연구나 시범단계에 있거나, 일부 실제 적용 사례가 제한적 이거나 시험적으로 지급되는 보험급여 등 일반적인 헬스케어 서비스 과정에 반영되고 있지 않으나, 미국은 시범적이나 제한적으로 이루어지고는 있지만, 그 중 Intel은 독자적으로 출시한 홈헬스 게이트웨이에서 큰 실패를 경험한 후, GE Healthcare와 공동으로 조인트벤처를 설립하여 새롭게 사업을 추진하고 있음

## 나. 생태계 기술 동향

### (1) 해외 플레이어 동향

- Allied Market Research에서 발표한 통계에 따르면, 세계 디지털 헬스케어 산업의 시장규모는 2016년 960억 달러 규모로 추정되며, 이후 연평균 21.1% 성장률을 보이고 2020년에는 2,060억 달러 규모까지 확대될 것으로 전망
  - Transparency Market Research에서는 기기 등을 대상으로 하는 글로벌 스마트 헬스케어 제품 시장이 2016년 317억 달러 규모에서 2023년에 578억 달러 규모로까지 성장하며, 2015년부터 2023년까지 매년 성장률은 8.84% CAGR(Compound Annual Growth Rate)일 것으로 전망
- (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 명의 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

## (2) 국내 플레이어 동향

- 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, 닥터다이러리 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

## 다. 국내 연구개발 기관 및 동향

### (1) 연구개발 기관

[ 스마트헬스케어 분야 주요 연구조직 현황 ]

기관	소속	연구분야
한림대학교	소프트웨어융합대학	• 무선통신·네트워크, 사물인터넷(IoT)
가톨릭대학교	가톨릭대학교 서울성모병원	• 의료 인공지능 (의료 AI), 빅데이터 분석 기술
가천대학교	가천대학교 인공지능 헬스케어 연구센터	• 인공지능 원천기술 연구, 바이오 및 의료분야 응용기술개발

### (2) 기관 기술개발 동향

- 한림대학교, 실시간 상황정보 빅데이터 기반의 개인 맞춤형 헬스케어 시스템 개발 (2016/06/01 ~ 2019/05/31)
  - 헬스케어 센서 및 통신 인터페이스를 포함하는 스마트 헬스케어 디바이스
  - 실시간 헬스케어 빅데이터 수집·분석을 위한 사용자영역 로컬 클라우드·빅데이터 시스템
  - 3rd Party 개방형 데이터 연동 및 헬스케어 서비스 프로비저닝, 관리를 위한 서비스 및 관리 플랫폼
  - 헬스케어 정보 연동을 위한 스마트 사용자 단말 (User Mobile)
  
- 가톨릭대학교, 모바일 기반 통합 개인 건강 맞춤형 만성질환 관리서비스 개발 및 헬스케어 빅데이터 분석 기술개발 사전기획연구 (2019/07/16 ~ 2019/08/31)
  - 모바일 기반 헬스케어 디바이스 연동 데이터 누적 관리 플랫폼 기술
  - 헬스케어 데이터 연동 기술
  - PHR 데이터 누적 관리 기술
  - 개인별 최적 치료 서비스 제공을 위한 임상 및 헬스케어 데이터 통합 표준화 전략 및 기술 개발
  - 모바일 기반 개인별 맞춤 만성질환 관리 프로그램 개발
  - 지역별 1차 의료기관 중심 실증 시스템 구축
  - 1차 의료기관 중심 맞춤 관리 서비스 임상 효과 실증
  
- 가천대학교, 우울증 예방 로봇·챗봇 개발 추진 (2019/12)
  - ‘미스터마인드’와 제휴하여 초고령화 시대를 대비한 인공지능 시니어 헬스케어 모델을 제시할 계획
  - 가천대가 개발한 시니어 우울증 예방 알고리즘을 미스터마인드가 개발한 로봇에 적용해 활용할 예정



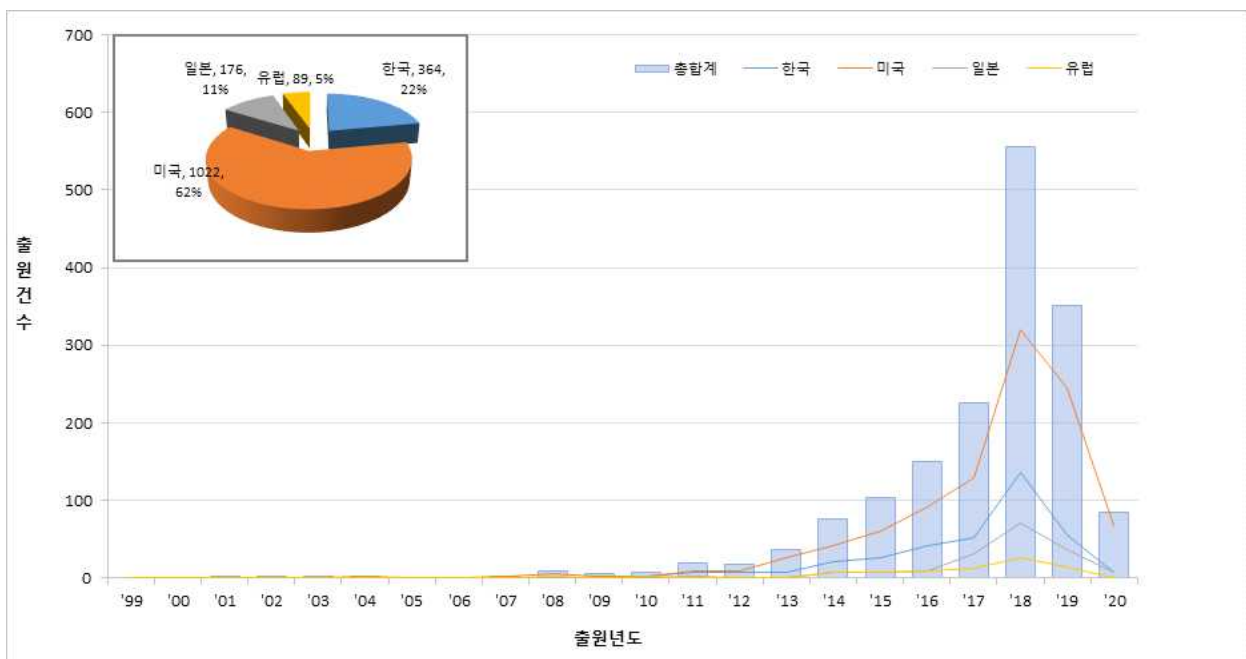
## 4. 특허 동향

### 가. 특허동향 분석

#### (1) 연도별 출원동향

- 스마트헬스케어를 위한 빅데이터 수집 시스템은 '17년부터 급격한 성장을 보임
  - 각 국가별로 살펴보면 미국이 가장 활발한 출원활동을 보이고 있음
- 국가별 출원비중을 살펴보면 미국이 전체의 62%의 출원 비중을 차지하고 있어, 최대 출원국으로 스마트헬스케어를 위한 빅데이터 수집 시스템 분야를 리드하고 있는 것으로 나타났으며, 한국은 22%, 일본은 11%, 유럽은 5% 순으로 나타남

[ 스마트헬스케어를 위한 빅데이터 수집 시스템 연도별 출원동향 ]

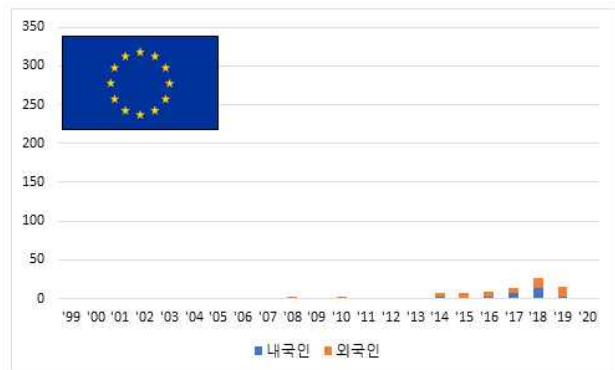
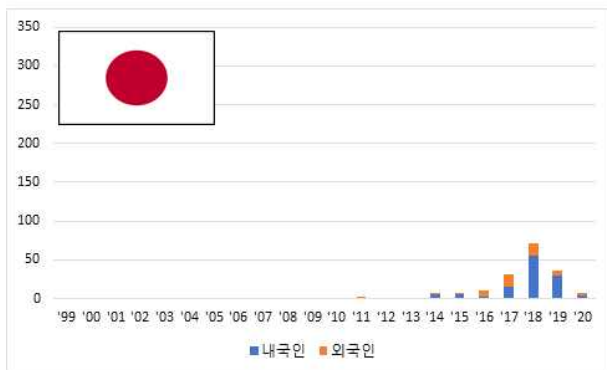
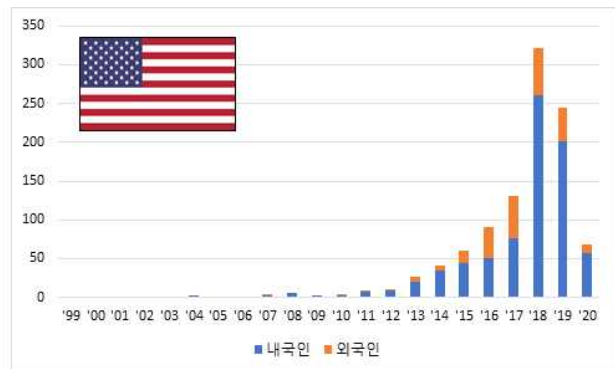
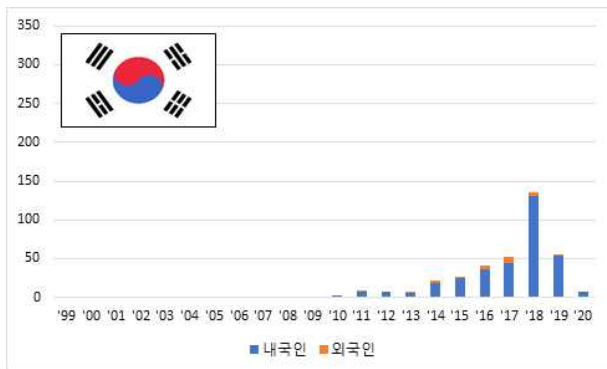




## (2) 국가별 출원현황

- 한국의 출원현황을 살펴보면, '18년부터 해당 기술의 출원이 급격히 증가하는 추세
  - 내국인 위주의 출원이 진행되고 있음
  - 한국 기술의 양적 흐름은 미국과 상당히 유사
  - 미국의 출원 수에 비해 36% 정도의 수준을 보임
- 미국의 출원현황을 살펴보면 분석구간 초기부터 전체 특허기술의 출원 증감 흐름에 영향을 주고 있는 것으로 나타남. 미국의 경우, 한국에 비해 외국인의 비중이 큰 것으로 나타남
- 유럽의 출원현황을 살펴보면 출원수가 매년 30건 이하로, 뚜렷한 증감 동향이 나타나지 않음. 해당 기술 분야에서 유럽 시장에 대한 관심도가 높지 않은 것으로 보임
- 일본의 출원현황은 살펴보면 한국과 유사한 동향을 보임

[ 국가별 출원현황 ]





(3) 기술 집중도 분석

□ 전략제품에 대한 최근 기술 집중도 분석을 위한 구간별 기술 키워드 분석 진행

- 전체 구간(1999년~2020년)에서 Healthcare Provider, Electronic Health Record, Health Information 등 키워드가 다수 도출
- 최근 구간 분석 결과, 최근 1구간(2012년~2015년)과 비교할 때 2구간(2016년~2020년)에서 Healthcare Provider, Electronic Health Record 키워드가 많이 등장하는 것으로 보아 스마트 헬스케어에 위한 빅데이터 분야에도 인공지능/빅데이터 기반 스마트 헬스 응용 기술, 바이오 데이터 수집 기술 관련 연구개발이 활발한 것으로 추정

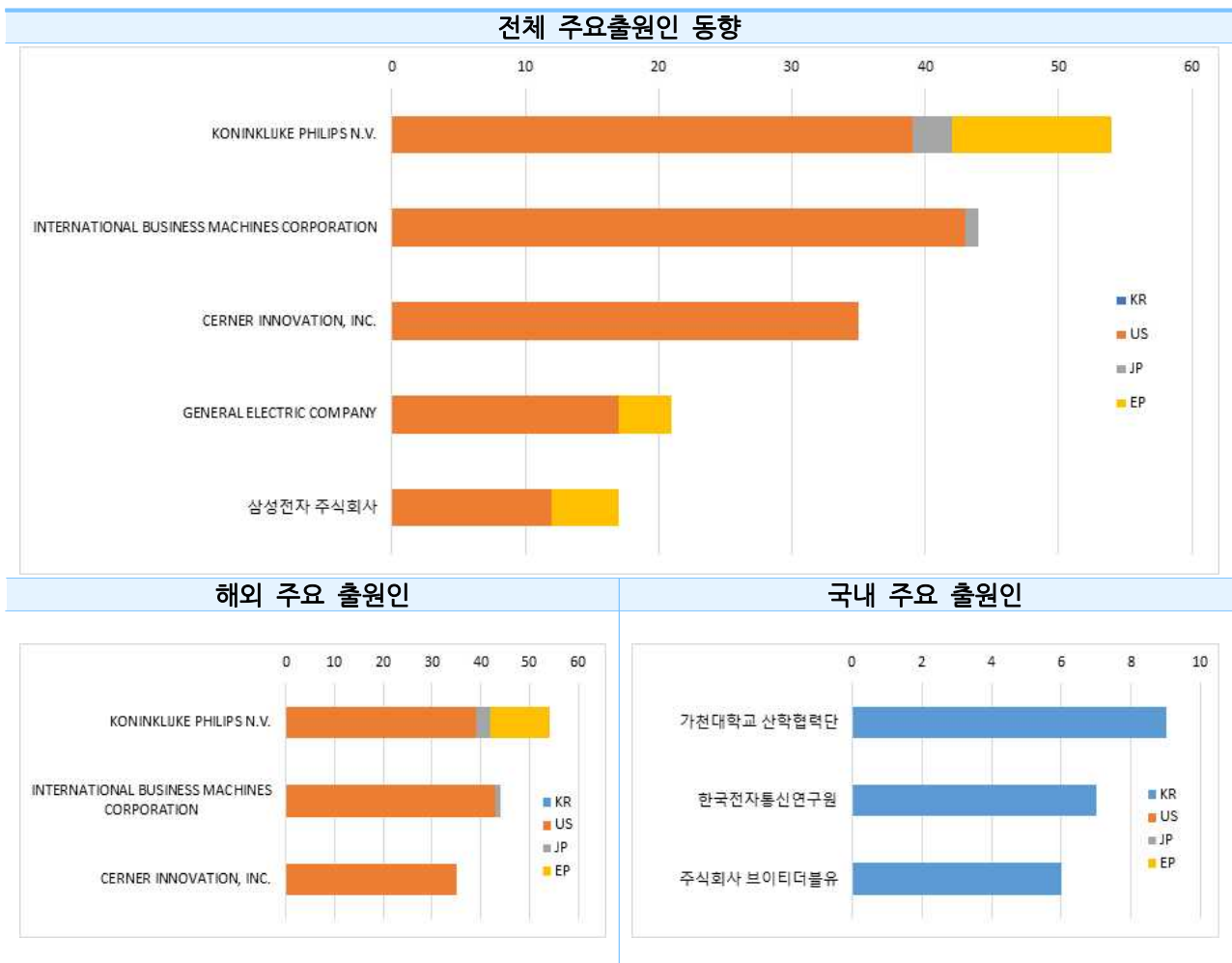
[ 특허 키워드 변화로 본 기술개발 동향 변화 ]

전체구간(1999년~2020년)	
<ul style="list-style-type: none"> <li>• Healthcare Provider, Electronic Health Record, Health Information, Computer Implemented, 건강 상태, User Interface, Patient Data, Real Time, Healthcare Facility, Health Data</li> </ul>	
최근구간(2012년~2020년)	
1구간(2012년~2015년)	2구간(2016년~2020년)
<ul style="list-style-type: none"> <li>• 사물 인터넷 센서, Computer Readable Medium, 신체의 상태, Health Condition, 신체의해부 캐릭터, 센싱 단계, 스마트 모바일 기기, 데이터 전송 단계, 클라우드 서버, Mobile Device</li> </ul>	<ul style="list-style-type: none"> <li>• Healthcare Provider, Electronic Health Record, Computer Implemented, 건강 상태, Health Information, Healthcare Facility, Computing Device, User Interface, Health Data, Patient Data</li> </ul>

## 나. 주요 출원인 분석

- 스마트헬스케어를 위한 빅데이터 수집 시스템의 전체 주요출원인을 살펴보면, 주로 미국 국적의 출원인이 다수 포함되어 있는 것으로 나타났으며, 제 1 출원인으로는 네덜란드의 KONINKLIJKE PHILIPS N.V.인 것으로 나타남
  - 제 1 출원인인 KONINKLIJKE PHILIPS N.V.의 출원은 미국에 집중된 경향을 보임
- 스마트헬스케어를 위한 빅데이터 수집 시스템 관련 기술로 의료 장비를 다루는 대기업에 의한 출원이 대다수를 차지
  - 국내에서는 연구기관/대학, 중소기업(개인)의 활발한 출원이 이루어짐

[ 스마트헬스케어를 위한 빅데이터 수집 시스템 주요출원인 ]

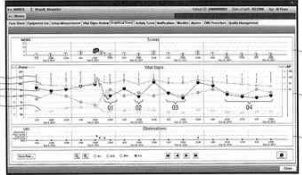
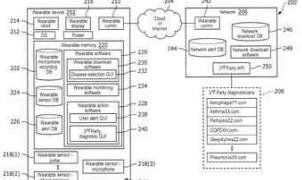
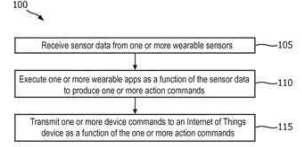
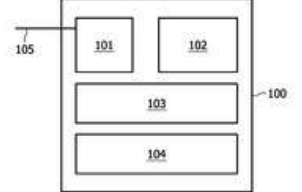


(1) 해외 주요출원인 주요 특허 분석

◎ KONINKLIJKE PHILIPS N.V.

- KONINKLIJKE PHILIPS N.V.은 네덜란드 기업으로, 스마트헬스케어에 위한 빅데이터 수집 시스템과 관련하여, 환자 관련 데이터를 모니터링하고 이를 가시화하거나 지원하는 것에 특허된 특허를 다수 출원. 그 중 등록된 특허는 4건 보유
  - 주요 특허들은 보다 사용자(의사, 간호사, 간병인 등)에게 환자의 건강상태를 쉽게 인식하기 위해서 환자 관련 데이터를 모니터링하고 이를 가시화하거나 지원하는 것에 활용하는 기술 특허를 다수 출원하는 것으로 파악

[ KONINKLIJKE PHILIPS N.V. 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
US10636523 (2016.05.20)	Device, system and method for visualization of patient-related data	환자의 건강 상태를 사용자가 빠르고 쉽게 인식할 수 있도록, 가시화를 위해 획득된 환자-관련 데이터를 처리하기 위한 프로세서를 포함한 기술	
US10229754 (2016.03.08)	Wearable device obtaining audio data for diagnosis	질병과 관련된 증상들에 대한 웨어러블 센서 판독들을 모니터링하고, 질병의 진단을 위해 건강 전문가에게 전송하는 기술	
US10359807 (2016.03.07)	Wearable health interface for controlling internet of things devices	웨어러블 디바이스가 사물 인터넷(IoT) 디바이스를 제어하는 기술	
EP2183694 (2008.07.11)	Decision support system for acute dynamic diseases	패혈증과 같은 급성인 질환을 고치는 환자의 치료를 위해 치료 개입의 선택에 있어서 임상 의사, 간호부 또는 다른 사용자를 지원하는 의료 장치	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ IBM(International Business Machines)

- IBM은 미국의 다국적 기업으로, 다수의 빅데이터 분석 관련 특허를 보유하고 있으며, 스마트헬스케어에 위한 빅데이터 수집 시스템을 활용한 헬스 케어 기술과 관련한 특허를 미국에 출원
  - 컴퓨팅 장치를 활용한 개인 헬스 케어 관련하여 출원을 진행한 특허 대부분 미국에 출원한 것으로, 자국 내 출원 성향이 높은 것으로 사료
  - 클라우드 시스템 기술을 활용한 헬스 케어 기술과 관련하여 44건의 미국 출원을 진행하였으며, 그 중 등록된 특허는 4건으로 파악

[ IBM 주요특허 리스트 ]

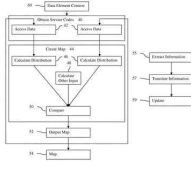
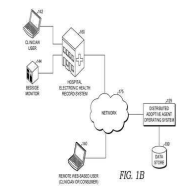
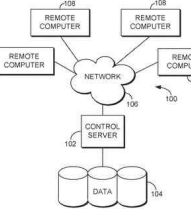
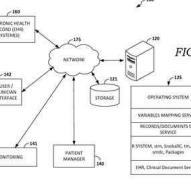
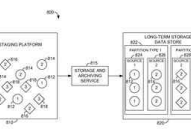
등록번호 (출원일)	명칭	기술적용분야	도면
US10529446 (2016.12.22)	Continuous health care plan coordination between patient and patient care team	개인화된 헬스 케어 관리 시스템을 구현하기 위한 메커니즘 기술	
US10586615 (2016.11.01)	Electronic health record quality enhancement	건강 관리 제공자에 의해 환자의 전자 건강 기록에 대한 업데이트를 식별하고 식별된 업데이트의 품질을 평가하는 전자 건강 기록 관리 기술	
US9274850 (2012.03.09)	Predictive and dynamic resource provisioning with tenancy matching of health metrics in cloud systems	컴퓨팅 자원들은 클라우드 컴퓨팅 환경에서 서비스 레벨 목표들에 부합하도록 동적으로 제공하고, 클라우드 컴퓨팅 환경에 프로비저닝할 수 있는 자원을 결정하여 품질을 모니터링하는 기술	
US10679749 (2008.08.22)	System and method for virtual world biometric analytics through the use of a multimodal biometric analytic wallet	멀티모달 바이오메트릭 분석 wallet의 사용을 통해 가상 세계 바이오메트릭 분석을 위한 시스템 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ CERNER INNOVATION, INC.

- CERNER INNOVATION, INC.은 미국 기업으로 헬스케어 사업의 지능형 솔루션을 제공하는 기업으로 등록특허 17건 보유
  - 스마트헬스케어를 위한 빅데이터 수집 시스템에서 환자 정보 데이터를 활용한 헬스케어 시스템 관련 응용 특허를 다수 보유

[ CERNER INNOVATION, INC. 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
US10565315 (2019.05.10)	Automated mapping of service codes in healthcare systems	헬스케어 내의 시맨틱의 맵핑기술로, 시맨틱 시스템들에 관한 맵(map)이 생성되어 제1 및 제2 시맨틱 시스템들 사이에서 통신하는데 사용	
US10672516 (2016.12.29)	Decision-support recommendation optimization	컨텍스트-특정 시스템-생성 추천 환자의 사용자 흡수의 예측된 확률을 제공하고, 관련 추천을 방출할지 여부를 결정하기 위해 상기 정보를 적용하기 위한 시스템 기술	
US10740439 (2016.04.14)	Mobile application for medication reminders	컴퓨터 기반 헬스케어 정보를 위해 제공되어, 환자의 모바일 장치 상의 적절한 시간에 처방 약물을 복용하도록 환자에게 미리 알림을 자동으로 제공하는 기술	
US10770184 (2015.12.04)	Determining patient condition from unstructured text data	전자 건강 기록으로부터의 비구조화 텍스트 데이터에 기초하여 하나 이상의 환자 상태의 존재 또는 부재의 가능성을 결정하기 위해 제공되며, 이는 케어 서비스 제공	
US10373712 (2015.05.08)	Aggregation, partitioning, and management of healthcare data for efficient storage and processing	헬스케어 데이터의 배치(batch)에 대한 액세스가 필요한 솔루션을 컴퓨팅함으로써 효율적인 검색 및 처리를 허용하는 헬스케어 데이터 저장 시스템 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

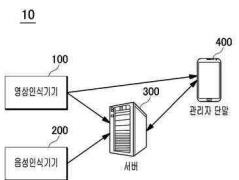

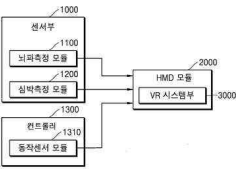
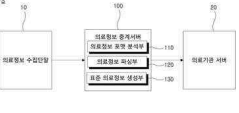
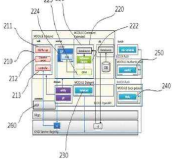
## (2) 국내 주요출원인 주요 특허 분석

### ◎ 가천대학교

□ 가천대학교는 스마트헬스케어 위한 빅데이터 수집 시스템에 있어서, 사용자의 생체 정보를 활용한 가상 현실 시스템과 관련된 특허를 9건 출원

- 가천대학교는 스마트헬스케어 위한 빅데이터 수집 시스템에 있어서, 사용자의 생체 정보를 활용한 가상 현실 시스템과 관련된 특허를 국내에만 특허를 출원. 등록건수는 7건으로 파악

[ 가천대학교 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
KR2133943 (2020.01.16)	시니어의 건강을 위한 홈 헬스케어 제공방법 및 제공장치	영상인식기와 음성인식기를 통하여 현재 건강이상 여부를 확인하고, 관련 정보를 전송 받아 시니어의 신체 건강에 이상 여부를 모니터링 하는 기술	
KR2151492 (2019.11.12)	사용자의 집중을 통한 뇌 건강을 위한 몰입형 피드백 가상현실 시스템 및 구현방법	가상현실 콘텐츠를 활용하여, 사용자의 집중을 통한 뇌 건강을 위한 몰입형 피드백 가상현실 시스템 및 구현 기술	
KR2151494 (2019.11.12)	사용자의 활동을 통한 뇌 건강 및 신체건강을 위한 활동형 피드백 가상현실 시스템 및 구현방법	사용자로부터 측정되는 생체정보를 실시간으로 적용하여 가상현실 콘텐츠를 제공하는 사용자의 활동을 통한 뇌 건강 및 신체 건강을 위한 활동형 피드백 가상현실 시스템	
KR1966752 (2017.12.21)	모바일 헬스 데이터의 의료기관 중계 시스템	모바일 단말에서 수집된 의료정보를 의료기관 서버에서 이용하는 표준 의료정보 데이터로 변환하여 의료기관 서버로 제공하는 기술	
KR1865073 (2016.04.19)	개인화된 건강 데이터 중계 시스템 및 방법	건강 데이터를 개방함에 따라 서비스 제공자와 데이터 제공자 간의 클라우드 스토리지에 저장된 건강 데이터를 중계하는 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ 한국전자통신연구원

□ 한국전자통신연구원은 한국의 연구기관으로, 스마트헬스케어를 위한 빅데이터 수집 시스템과 관련된 특허를 7건 출원하였으며, 7건의 특허 중 등록된 특허는 1건을 보유

[ 한국전자통신연구원 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
KR1961165 (2015.07.21)	개방형 건강 관리 장치 및 방법	건강사례 빅데이터를 구축하며, 건강사례 빅데이터를 기반으로 개인 맞춤형 건강 서비스를 제공하는 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출



◎ 주식회사 브이티더블유

- 주식회사 브이티더블유는 한국의 빅데이터 전문 분석 및 컨설팅 기업으로, 스마트헬스케어를 위한 빅데이터 수집 시스템과 관련된 특허를 6건 출원하였으며, 6건의 특허 중 등록된 특허는 3건을 보유

[ 주식회사 브이티더블유 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
KR2110388 (2020.01.13)	지역적 블록체인 기반의 cphr 서비스 운용 방법	복수의 진료 기록을 저장하고, 복수의 진료 기록은 블록체인으로 연결되어 지역적 블록체인 기반의 cPHR 서비스 운용 방법	
KR2144540 (2020.01.13)	응급 상황을 위한 블록체인 기반의 cphr 서비스 운용 방법	응급 상황에 적용할 수 있는 블록체인 기반으로 관리 서버가 응급 상황을 응급 상황을 나타내는 정보를 획득하고, 획득된 응급 상황에 대응하는 응답 신호를 생성 및 전송하는 기술	
KR2144532 (2020.01.13)	블록체인 기반의 cphr 서비스 운용 방법	블록체인 기반의 cPHR 서비스 운용 방법에 의하면 개인 건강 기록에 대해 환자가 주권을 가지는 정보 활용 패러다임을 제공	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출



## 다. 기술진입장벽 분석

### (1) 기술 집중력 분석

- 스마트헬스케어를 위한 빅데이터 수집 시스템관련 기술에 대한 시장관점의 기술독점 현황분석을 위해 집중률 지수(CRn: Concentration Ratio n, 상위 n개사 특허점유율의 합) 분석 진행
  - 상위 4개 기업의 시장점유율이 0.09로 스마트헬스케어를 위한 빅데이터 수집 시스템 분야에 있어서 독과점 정도는 매우 낮은 수준으로 판단
  - 국내 시장에서 중소기업의 점유율 분석결과 0.73으로 해당 기술에 대하여 중소기업이 진입하여 활동하고 있는 것으로 파악

[ 주요출원인의 집중력 및 국내시장 중소기업 집중력 분석 ]

주요 출원인 집중력	주요출원인	출원건수	특허점유율	CRn	n
	KONINKLIJKE PHILIPS N.V.(네덜란드)	54	3.3%	0.03	1
	INTERNATIONAL BUSINESS MACHINES CORPORATION(미국)	44	2.7%	0.06	2
	CERNER INNOVATION, INC.(미국)	35	2.1%	0.08	3
	GENERAL ELECTRIC COMPANY(미국)	21	1.3%	<b>0.09</b>	<b>4</b>
	삼성전자 주식회사 (한국)	17	1.0%	0.10	5
	HILL-ROM SERVICES, INC.(미국)	10	0.6%	0.11	6
	FINC TECHNOLOGIES INC(일본)	10	0.6%	0.12	7
	가천대학교 산학협력단(한국)	9	0.5%	0.12	8
	APPLE INC.(미국)	8	0.5%	0.13	9
	SIEMENS HEALTHCARE GMBH(독일)	8	0.5%	0.13	10
	<b>전체</b>	<b>1651</b>	<b>100%</b>	<b>CR4=0.09</b>	
	국내시장 중소기업 집중력	출원인 구분	출원건수	특허점유율	CRn
중소기업(개인)		274	72.9%	0.73	
대기업		19	5.1%		
연구기관/대학		83	22.1%		
<b>전체</b>		<b>376</b>	<b>100%</b>	<b>CR중소기업=0.73</b>	

## (2) 특허소송 현황 분석

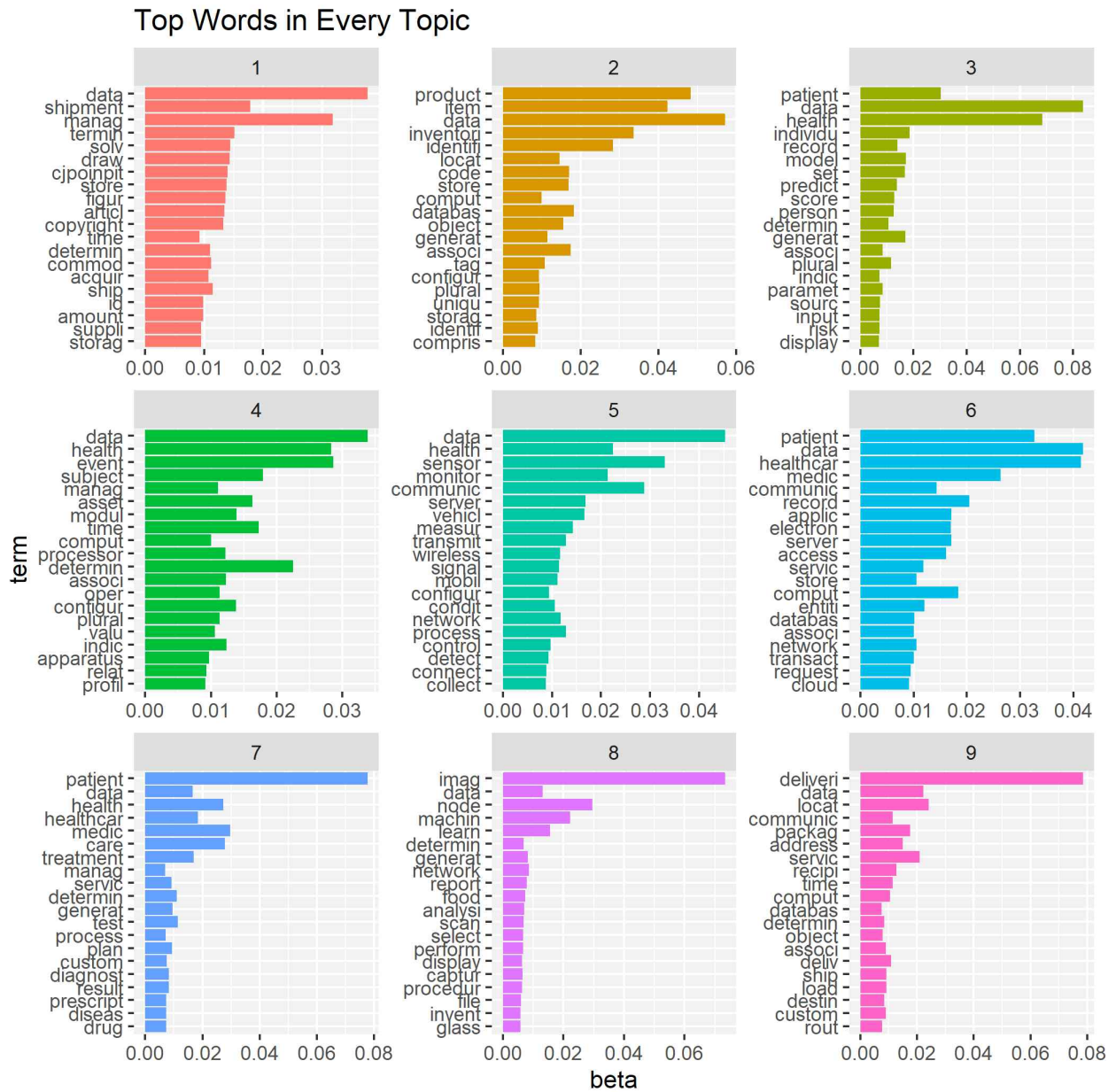
- 스마트헬스케어를 위한 빅데이터 수집 시스템 분야 관련 특허소송 이력은 검색되지 않음
  - 따라서 국내기업이 미국시장에 진입하는 경우, 해당 분야를 선점할 수 있을 것으로 판단
  - 다만, 빅데이터 수집과 관련해서는 4개의 소송이 검색

## 5. 요소기술 도출

### 가. 특허 기반 토픽 도출

- 2,223개의 특허의 내용을 분석하여 구성 성분이 유사한 것끼리 클러스터링을 시도하여 대표성이 있는 토픽을 도출

[ 스마트헬스케어를 위한 빅데이터 수집 시스템 도구에 대한 토픽 클러스터링 결과 ]



## 나. LDA<sup>15)</sup> 클러스터링 기반 요소기술 도출

[ LDA 클러스터링 기반 요소기술 키워드 도출 ]

No.	상위 키워드	대표적 관련 특허	요소기술 후보
클러스터 01	data, manage, shipment, terminal, solve, draw, store, figure, article	<ul style="list-style-type: none"> <li>Product placement planning system and product placement planning method</li> <li>Article management data creation apparatus, article management data creation method, and article management data creation program</li> </ul>	-
클러스터 02	data, product, item, inventories, identification, database, associate, code, store, object	<ul style="list-style-type: none"> <li>Method and apparatus for storage of data for tracking manufactured items</li> <li>Method and system for validating a device that uses a dynamic identifier</li> </ul>	-
클러스터 03	data, health, patient, individual, model, generate, set, record, predict, score	<ul style="list-style-type: none"> <li>Method and apparatus for prediction of complications after surgery</li> <li>Generating data in standardized formats and providing recommendations</li> </ul>	인공지능/빅데이터 기반 스마트 헬스 응용 기술
클러스터 04	product, patient, stock, store, bomb, package, healthcare, event, account, node	<ul style="list-style-type: none"> <li>Systems and methods for response calibration</li> <li>Intelligent grouping of peers in motivational programs</li> </ul>	스마트 헬스 정보 교환 표준화 기술
클러스터 05	data, sensor, communicate, health, monitor, server, vehicle, measure, process, transmit	<ul style="list-style-type: none"> <li>Portable intelligent driver's health monitoring system for safety on the road</li> <li>Systems and methods for monitoring transported cargo</li> </ul>	개인용 헬스 디바이스
클러스터 06	data, healthcare, patient, medicine, record, computing, server, application, electron, access patient, medicine, care, health, healthcare, treatment, data, test, determine, generate	<ul style="list-style-type: none"> <li>Blockchain framework for enforcing regulatory compliance in healthcare cloud solutions</li> <li>Payer governed american medical information network (amin): patient, provider and payer federated healthcare information hub.</li> </ul>	클라우드 기반 스마트 헬스 응용 기술
클러스터 07	healthcare, treatment, data, test, determine, generate	<ul style="list-style-type: none"> <li>Dynamically determining risk of clinical condition</li> <li>Decision support with clinical nomenclatures</li> </ul>	실시간 데이터 분석을 통한 위험상환 인지 기술
클러스터 08	image, node, machine, learn, data, network, generate, report, food, analysis	<ul style="list-style-type: none"> <li>First trainable robot doctor named aipd using new methods and systems for its artificial brain</li> <li>Enforcement of rules for the shipping of items</li> </ul>	인공지능/빅데이터 기반 스마트 헬스 응용 기술
클러스터 09	delivery, location, data, service, package, address, recipient, communicate, time, delivery	<ul style="list-style-type: none"> <li>Methods for coordinating the delivery of parcels by travelers</li> <li>Administration system and method for product delivery</li> </ul>	-

15) Latent Dirichlet Allocation

### 다. 특허 분류체계 기반 요소기술 도출

- 스마트헬스케어를 위한 빅데이터 수집 시스템 관련 특허에서 총 10개의 주요 IPC코드(메인그룹)를 산출하였으며, 각 그룹의 정의를 기반으로 요소기술 키워드를 아래와 같이 도출

[ IPC 분류체계에 기반한 요소기술 도출 ]

IPC 기술트리		
(서브클래스) 내용	(메인그룹) 내용	요소기술 후보
(A61B) 진단; 수술; 개인 식별	• (A61B-005) 진단을 위한 측정	-
(G06F) 전기에 의한 디지털 데이터처리	• (G06F-019) 특수한 어플리케이션에 특히 적합한 디지털 컴퓨팅 또는 데이터 처리 장치 또는 방법	-
(G06Q) 관리용, 상업용, 금융용, 경영용, 감독용 또는 예측용으로 특히 적합한 데이터 처리 시스템 또는 방법; 그 밖에 분류되지 않는 관리용, 상업용, 금융용, 경영용, 감독용 또는 예측용으로 특히 적합한 시스템 또는 방법	• (G06Q-050) 특정 사업 부문에 특히 적합한 시스템 또는 방법, 예. 공익사업 또는 관광	개인용 헬스 디바이스
(G16H) 헬스케어 인포매틱스, 즉. 의료 또는 건강 관리 데이터의 취급 또는 처리에 특히 적합한 정보통신 기술	• (G16H-050) 의료 진단, 의료 시뮬레이션 또는 의료 데이터마이닝에 특히 적합한 ICT; 전염병이나 유행성병 검지, 감시 또는 모델화를 행하기 위해 특히 적합한 ICT	-
	• (G16H-010) 환자 관련의 의료 또는 건강관리 데이터의 취급 또는 처리에 특히 적합한 ICT	스마트 헬스 정보 교환 표준화 기술
	• (G16H-080) 의사, 환자 사이의 커뮤니케이션을 용이하게 하기 위해 특히 적합한 ICT, 예. 협조적 진단, 치료 또는 건강 감시	-
	• (G16H-020) 치료 또는 건강개선계획에 특히 적합한 ICT, 예. 처방전의 취급, 치료를 진행하는 것 또는 환자 컴플라이언스(compliance)를 감시하는 것	생활 습관 및 의료 데이터 수집 및 정제 기술
	• (G16H-040) 헬스케어 자원 또는 설비의 관리 또는 운영에 특히 적합한 ICT; 의료기기 또는 장치의 관리 또는 조작에 특히 적합한 ICT	-
	• (G16H-030) 의료화상의 취급 또는 처리에 특히 적합한 ICT	-
	• (G16H-015) 의학적 레포트에 특히 적합한 ICT, 예. 그 작성 또는 전송	-

## 라. 최종 요소기술 도출

- 산업·시장 분석, 기술(특허)분석, 전문가 의견, 타부처 로드맵, 중소기업 기술수요를 바탕으로 로드맵 기획을 위하여 요소기술 도출
- 요소기술을 대상으로 전문가를 통해 기술의 범위, 요소기술 간 중복성 등을 조정·검토하여 최종 요소기술명 확정

### [ 스마트헬스케어를 위한 빅데이터 수집 시스템 분야 요소기술 도출 ]

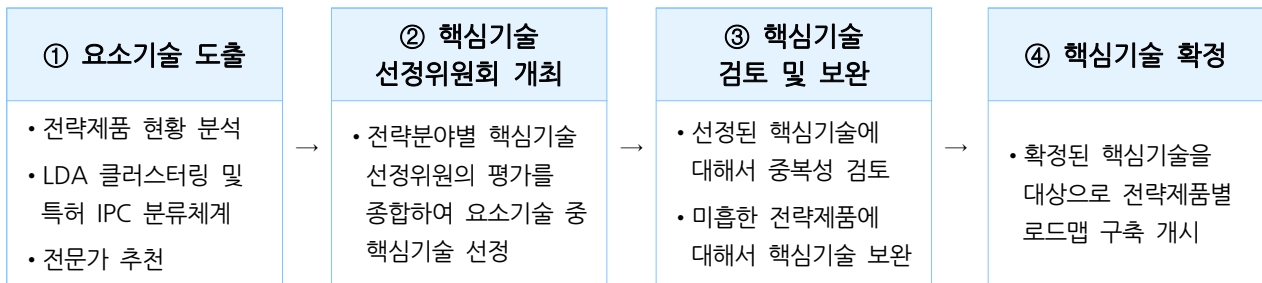
요소기술	출처
인공지능/빅데이터 기반 스마트 헬스 응용 기술	전문가추천
바이오 데이터 수집용 IoT 기술	특허 클러스터링, 전문가추천
실시간 데이터 분석을 통한 위험상황 인지 기술	특허 클러스터링, 전문가추천
개인용 헬스 디바이스	IPC 기술체계, 전문가추천
바이오 데이터 보호 기술	전문가추천
분석 처리속도 향상을 위한 엣지컴퓨팅 기술	전문가추천
스마트 헬스 정보 교환 표준화 기술	특허 클러스터링, IPC 기술체계, 전문가추천
생활 습관 및 의료 데이터 수집 및 정제 기술	IPC 기술체계, 전문가추천
클라우드 기반 스마트 헬스 응용 기술	전문가추천
진료 데이터 수집 및 저장 기술	특허 클러스터링, 전문가추천

## 6. 전략제품 기술로드맵

### 가. 핵심기술 선정 절차

- 특허 분석을 통한 요소기술과 기술수요와 각종 문헌을 기반으로 한 요소기술, 전문가 추천 요소기술을 종합하여 요소기술을 도출한 후, 핵심기술 선정위원회의 평가과정 및 검토/보완을 거쳐 핵심기술 확정
- 핵심기술 선정 지표: 기술개발 시급성, 기술개발 파급성, 기술의 중요성 및 중소기업 적합성
  - 장기로드맵 전략제품의 경우, 기술개발 파급성 지표를 중장기 기술개발 파급성으로 대체

#### [ 핵심기술 선정 프로세스 ]



### 나. 핵심기술 리스트

#### [ 스마트헬스케어에 위한 빅데이터 수집 시스템 분야 핵심기술 ]

핵심기술	개요
인공지능/빅데이터 기반 스마트 헬스 응용 기술	• 정보제공 허용가능한 개인의 건강정보(혈압, 키, 몸무게, 맥박, 건강검진정보 등)를 바탕으로 헬스 이상 징후 등 지속 트래킹하는 시스템
바이오 데이터 수집용 IoT 기술	• IOT 기반 실시간 바이오정보 수집
실시간 데이터 분석을 통한 위험상황 인지 기술	• 헬스정보 활용 및 주요헬스 지수개발을 통한 Warningsign 기술적용
개인용 헬스 디바이스	• 헬스 디바이스 개발 및 탑재를 통한 실시간 헬스 정보수집
바이오 데이터 보호 기술	• 민감 건강정보 등의 보안 기술 개발

### 다. 중소기업 기술개발 전략

- 글로벌 업체들의 비즈니스 전략을 면밀히 분석하고, 초기 시장을 선점할 수 있는 비즈니스 전략
- 산학연 및 대-중견-중소기업 협력에 기반해 자체적으로 진행하기 어려웠던 R&D 진행
- 정부 R&D 사업과 현장실습 프로그램 연계로 부족한 인력 확충하여 경쟁력 제고

### 라. 기술개발 로드맵

#### (1) 중기 기술개발 로드맵

[ 스마트헬스케어를 위한 빅데이터 수집 시스템 분야 중기 기술개발 로드맵 ]

스마트헬스케어 위한 빅데이터 수집 시스템	사용자 헬스케어를 위한 데이터 수집과 메디데이터의 임상 분석을 통해 신약 개발에 소요되는 임상 기간을 단축 및 삶의 질 향상			
	2021년	2022년	2023년	최종 목표
인공지능/빅데이터 기반 스마트 헬스 응용 기술				의료계, 보험 산업계 등에서 활용 가능한 지표 개발 개인화 서비스 수준
바이오 데이터 수집용 IoT 기술				바이오 수집 정보의 실시간, 유연한 수집 수준 헬스 관련 지표 및 모델링  데이터 loss 방지
실시간 데이터 분석을 통한 위험상황 인지 기술				상황별/ 고객 건강 수준별 위험지수 개발 및 알람
개인용 헬스 디바이스				개인용 헬스 정보의 의미 지표화 및 트래킹  바이오 데이터 전송 가능 디바이스 15종 이상
바이오 데이터 보호 기술				민감 정보 안전장치를 위한 철저한 보안 기술개발



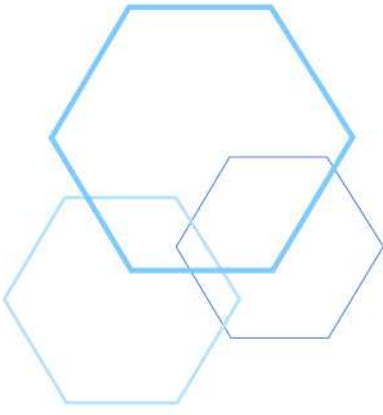
(2) 기술개발 목표

- 최종 중소기업 기술로드맵은 기술/시장 니즈, 연차별 개발계획, 최종목표 등을 제시함으로써 중소기업의 기술개발 방향성을 제시

[ 스마트헬스케어를 위한 빅데이터 수집 시스템 분야 핵심요소기술 연구목표 ]

핵심기술	기술요구사항	연차별 개발목표			최종목표	연계R&D 유형
		1차년도	2차년도	3차년도		
인공지능/빅데이터 기반 스마트 헬스 응용 기술	정보 수집 범위 및 수집 방법 수집 항목	헬스정보 define, 사업목표 수립, 정보 수집 기관별 정보 수집	지속 정보 수집 및 정보 간 연계 정보 업데이트 수준 지표개발	지표 개발 수준 점검 사용자 분석 Test, 활용수준 Test	의료계, 보험 산업계 등에서 활용 가능한 지표 개발 개인화 서비스 수준	상용화
	실시간 정보 변화 트래킹 및 적재					
	헬스 관련 주요 지표 개발					
	개인화 서비스 가능성					
바이오 데이터 수집용 IoT 기술	수집대상 (개인,의료기관)에 따른 IOT 장착 기술	바이오 정보 정의, 수집 기관, 수집 방법 정의, 활용 목적 기술 등	바이오 정보에 따른 IOT 개발 또는 적용 및 수집 처리	데이터 활용 기술, 분석 방법, 모델링 헬스 관련 이상징후 indexing 등	바이오 수집 정보의 실시간, 유연한 수집 수준, 헬스 관련 지표 및 모델링	상용화
	바이오 데이터 수집 데이터 loss율	10%	5%	1%	데이터 loss 방지	산학연
실시간 데이터 분석을 통한 위험상황 인지 기술	다양한 형태별 warning index 개발	위험 상황 정의에 대한 의학 전문가 관여 필요. 위험 상황 지표 개발	연령별/보유건강상태별 위험 지표 차별화  위험 상황 지표의 실시간 트래킹 및 네트워크를 통한 자동알람	결과 검증 및 추가 지표 보완 Pilottest	상황별/고객건강수준별 위험지수 개발 및 알람	산학연
	다양한 형태별 warning index 알람을 통한 사전 위험 예방					
	다양한 형태별 warning 처치 프로세스 연계					
개인용 헬스 디바이스	개인 헬스 정보 정의 (인바디 측정정보 등 포함) 디바이스 형태, 정보의 정확도	개인용 헬스 정보 형태 분석	개인 헬스 정보 측정을 위한 디바이스 유형, 정확도 분석 등	정보측정	개인용 헬스 정보의 의미 지표화 및 트래킹	상용화
	바이오 데이터 전송가능 디바이스 종류	5 종	8 종	15 종	바이오 데이터 전송 가능 디바이스 15종 이상	기술혁신
바이오 데이터 보호 기술	해킹 위험 대비 장치 기술	국내외 데이터 보안 기술 분석	방어벽, 보안기술개발	해킹위험 Test	민감 정보 안전장치를 위한 철저한 보안 기술 개발	창업성장
	데이터 접속자 개별관리 기술					





전략제품 현황분석

# 빅데이터 분석 및 시각화 플랫폼





# 빅데이터 분석 및 시각화 플랫폼

## 정의 및 범위

- 빅데이터 분석 및 시각화 플랫폼은 통상적으로 사용하는 소프트웨어가 수용할 수 없는 크기의 데이터인 빅데이터로부터 가치를 추출하고 결과를 분석하는 기술과 데이터 분석 결과를 유의미한 정보로 표현하는 기술의 총체를 의미
- 데이터 시각화는 빅데이터를 분석하여 이용자가 필요로 하는 정보를 도출, 정보를 쉽게 이해할 수 있도록 텍스트 혹은 이미지 등, 내용 전달에 효과적이고 시각적 형태를 활용

## 전략 제품 관련 동향

시장 현황 및 전망	제품 산업 특징
<ul style="list-style-type: none"> <li>(세계) 세계 데이터 통합시장은 2018년 130억 달러에서 연평균 16.2% 성장하여 2024년 320억 달러까지 성장 전망</li> <li>(국내) 2018년 1,584억 원 규모에서 연평균 16.3%로 성장하여 2024년 3,920억 원 규모로 성장 전망</li> </ul>	<ul style="list-style-type: none"> <li>사물인터넷, 웨어러블 디바이스의 이용 증가와 미디어 매체의 증가에 따라 비정형 데이터가 빅데이터 시장의 90% 이상을 차지하기에 시각화 플랫폼의 역할이 증대</li> <li>미적 형태와 기능성 두 가지를 모두 포괄하는 데이터 시각화로 데이터들의 연결과 그룹핑을 표현하는데 초점</li> </ul>
정책 동향	기술 동향
<ul style="list-style-type: none"> <li>규제혁신, 규제 샌드박스 활성화로 ICT/SW 제도 개선 및 수요자 중심의 ICT 중소·벤처 고성장화 지원</li> <li>정부는 4차 산업 혁명 정책의 중심인 빅데이터 산업 활성화를 체계적 지원을 위해 '17년 제3차 민간 합동 빅데이터 TF회의를 개최하여 '21년까지 112개 분야 지능, 융합형 공공데이터 구축 및 개방을 확대하고, 창업 콜라보 프로젝트, 창업 경진대회 등을 통해 공공데이터 활용 기업을 육성하고 창업을 지원</li> </ul>	<ul style="list-style-type: none"> <li>플랫폼 기술 관련 산업의 주요 이슈는 SQL의 재등장, 데이터 가상화</li> <li>데이터 분석 기술 관련 산업 내 자동화된 분석 도구의 이슈화</li> <li>국립과학기술센터, 한국의 과학공학 분야 연구개발 논문 성과는 6만 4천 편으로 전 세계 평균 성장률보다 0.4% 높은 4.2%를 나타냄('19.12.)</li> </ul>
핵심 플레이어	핵심기술
<ul style="list-style-type: none"> <li>(해외) Google, Oracle, Spotfire, DataPlanet</li> <li>(대기업) KT, 엑셈, 솔트룩스, 효성인포메이션시스템</li> <li>(중소기업) 에스씨플랫폼, 투그램시스템즈, 그루터, 데이터스퀘어</li> </ul>	<ul style="list-style-type: none"> <li>대용량 및 실시간 데이터 분석 플랫폼 기술</li> <li>데이터 시각화 기술</li> <li>머신러닝 기반 데이터 분석 기술</li> <li>실시간 데이터 처리 기술</li> <li>내외부 데이터 연동 기술</li> </ul>

## 중소기업 기술개발 전략

- IoT의 발전이 보편화됨에 따라 수집되는 비정형 데이터의 양은 폭발적으로 증가하여 비정형 데이터 분석의 중요도는 앞으로도 높아질 전망, 따라서 산업적으로 활용할 수 있도록 비정형 빅데이터를 분석하고 가공하는 기술개발을 통해 시장 선점
- 글로벌 선두 그룹들의 실시간 데이터 처리, 배치 데이터 처리 등의 기술을 도입하여 세계시장에 뒤처지지 않도록 함과 동시에 각 산업 분야에 적합한 다양한 데이터 분석과 시각화 기술개발로 중소기업의 경쟁력 강화
- 플랫폼이 아닌 빅데이터를 활용한 서비스 개발과 유료 수익 모델 개발을 통한 수익 창출

# 1. 개요

## 가. 정의 및 필요성

### (1) 정의

- 빅데이터 분석 및 시각화 플랫폼은 통상적으로 사용하는 소프트웨어가 수용할 수 없는 크기의 데이터인 빅데이터로부터 가치를 추출하고 결과를 분석하는 기술과 데이터 분석 결과를 유의미한 정보로 표현하는 기술의 총체를 의미
- 데이터 시각화는 빅데이터를 분석하여 이용자가 필요로 하는 정보를 도출, 정보를 쉽게 이해할 수 있도록 텍스트 혹은 이미지 등, 내용 전달에 효과적인 시각적 형태를 활용

### (2) 필요성

- 빅데이터 분석 및 시각화 플랫폼은 기계학습과 인공지능을 활용한 분석 영역을 확대하고 있으며 분석 결과 정보 전달 측면에서 계속 연구 및 개발 중
  - 빅데이터 분석 및 시각화는 데이터 수집, 저장, 처리 단계와 직접적인 관련이 있으므로 상호 연관하여 연구 및 개발 필요성 증대
- 지능화 관점에서 인공지능 기술을 적용하기 위한 수요는 증가하고 있으며 이의 기반이 되는 빅데이터 분석 기술은 지능정보사회에서 대표적인 기술로 `부상
  - 미국의 정보기술 연구 및 자문회사인 가트너는 지능(Intelligence) 분야, 디지털(Digital) 분야, 메시(Mesh) 분야를 2019년 전략 기술의 3대 핵심 주제로 선정
  - 지능 분야의 개별적인 기술로 자율 사물(Autonomous Things), 증강 분석(Augmented Analytics), 인공지능 주도 개발(AI-Driven Development)을 선정
    - 자율 사물은 로봇, 드론, 자율주행차 등과 같은 사물이며 인간이 수행하던 일을 인공지능을 이용하여 자동화하는 것
    - 증강 분석은 특정 영역의 증강 지능에 초점을 맞추어 기계학습을 활용하여 분석 콘텐츠 개발, 소비, 공유의 혁신을 의미
    - 인공지능 주도 개발은 전문 데이터 과학자와 앱 개발자가 협력하는 기존 방식이 아닌 전문 개발자들이 단독으로 개발 및 운영할 수 있도록 인공지능을 앱에 내장하기 위한 도구, 기술, 최적화된 프로세스에 관한 연구와 개발과정 자체에서 사용되는 인공지능 기반의 도구 개발 연구를 의미
  - 빅데이터 분석은 기계학습과 인공지능 기술을 이용한 예측 분석 영역으로 확장

## 나. 범위 및 분류

### (1) 가치사슬

데이터 분석 및 시각화 플랫폼의 산업구조

- 후방산업으로 ICT 인프라, 클라우드, IoT를 꼽을 수 있으며 전방산업으로는 3D 변환 시각화 도구, 인공지능, AR/VR 기술로 볼 수 있음

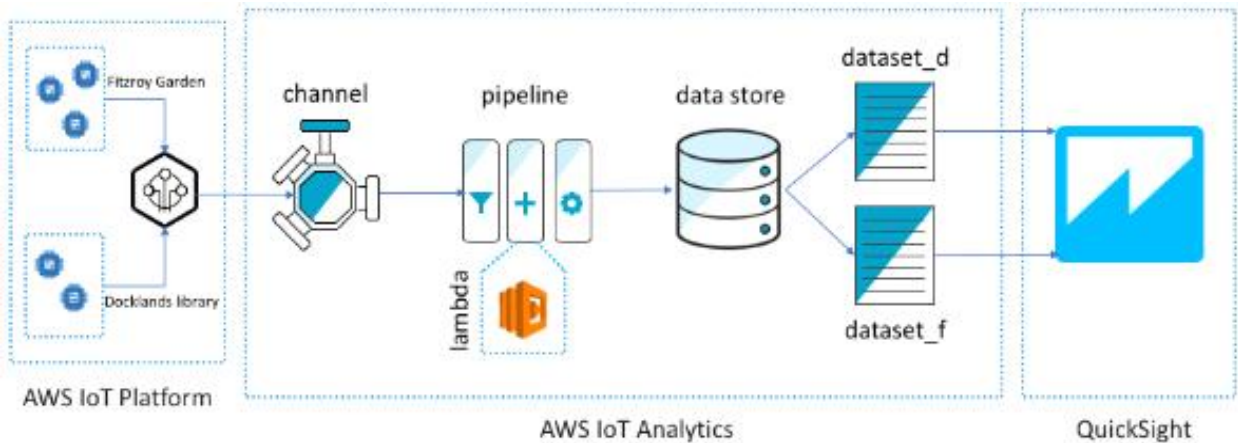
[ 데이터 분석 및 시각화 플랫폼 분야 산업구조 ]

후방산업	데이터 분석 및 시각화 분야	전방산업
ICT 인프라, 클라우드, IoT 초연결 시대 등	빅데이터 수집 및 시스템 구축, 시계열 빅데이터 처리, 분석 시각화 플랫폼	3D 변환 시각화 도구, 인공지능, AR/VR, 데이터 판매 및 증개

데이터 발생에서부터 시각화 플랫폼까지의 플로우

- 정형 정보 데이터 발생, 이를 수집하고 저장. 여기서 메타 데이터는 정형, 반정형, 비정형 데이터로 구분
- SQL, NoSQL, Hadoop을 통해 데이터를 분석하고 분류한 뒤 딥러닝 알고리즘을 접목해서 의미 있는 분석 결과를 도출, 그리고 그 결과를 시각화, 도식화

[ 데이터 발생에서부터 시각화 플랫폼까지의 플로우 ]



\* 출처: AWS(<http://aws.amazon.com>), 2019.

### (2) 용도별 분류

- 대민 서비스 고도화를 위한 공공부문 분석, 대량의 데이터 분석을 통한 금융 특화 솔루션, 자연어처리, 머신러닝, 언어학적 규칙을 활용한 비정형 데이터 분석 및 시각화 서비스로 나뉘볼 수 있음

### (3) 세부 용도별 기술 분류

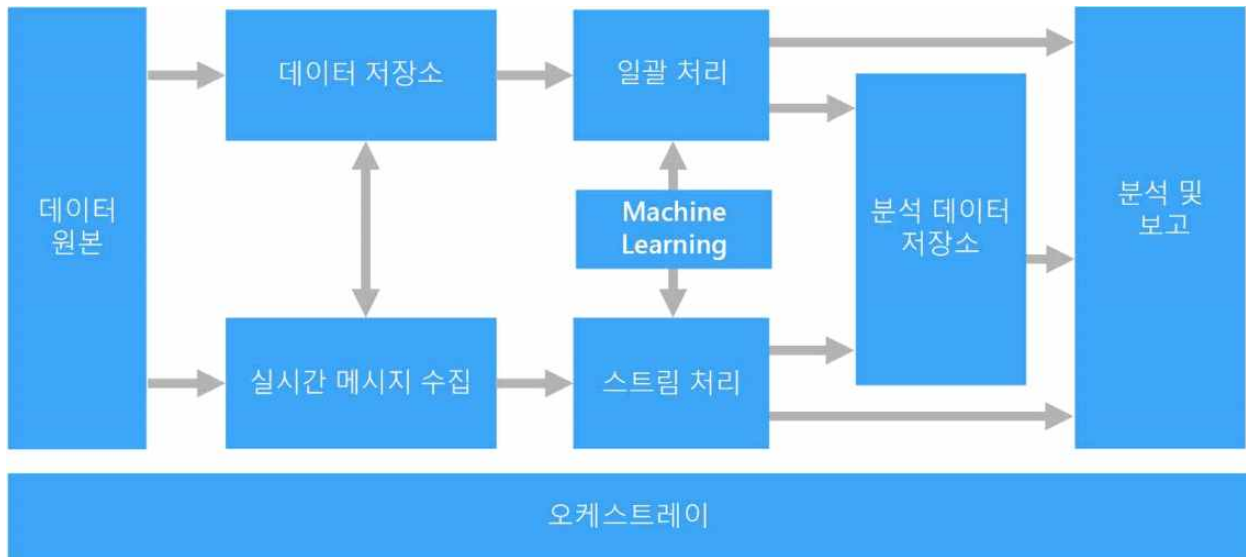
#### [ 기술별 분류 ]

분류	상세 내용
분산 스토리지 기술	<ul style="list-style-type: none"> <li>• 대용량 파일을 다양한 형식으로 저장하는 기술로 분산 파일 시스템과 오브젝트 스토리지가 대표적인 솔루션</li> <li>• 분산 파일 시스템은 여러 호스트가 참여하여 마치 단일 파일 시스템인 것처럼 파일 서비스를 제공하는 것으로 HDFS(Hadoop File System), GlusterFS 등의 솔루션이 포함</li> <li>• 오브젝트 스토리지는 블록 스토리지인 파일 시스템과 달리 데이터를 오브젝트로 관리하여 엑사바이트 범위도 손쉽게 확장할 수 있도록 Ceph, Lustre, AWS S3 등의 솔루션이 포함</li> </ul>
NoSQL 데이터베이스 기술	<ul style="list-style-type: none"> <li>• 관계형 데이터베이스와 달리 비정형, 반정형 데이터를 빠르게 분석하도록 데이터 형태에 따라 Columnar DB, Document DB, Key-Value DB, Graph DB 등으로 나뉘는 저장 기술</li> <li>• 클라우드 환경에서 오픈소스 기반의 NoSQL 데이터베이스인 HBase, MongoDB 등을 대부분 사용했지만 클라우드 솔루션의 확장으로 AWS Dynamo와 같이 클라우드 기업에서 직접 제공하는 솔루션의 사용이 증가</li> </ul>
배치 데이터 처리 기술	<ul style="list-style-type: none"> <li>• 대용량 데이터를 MapReduce와 같이 분산병렬 처리하여 상대적으로 장시간에 걸쳐 원하는 데이터를 추출하는 기술</li> <li>• MapReduce는 대용량의 배열 및 행렬로 표현된 데이터를 통계 분석하거나 정형 데이터와 비정형 데이터 간 연계 분석, 기계학습에 활용</li> </ul>
실시간 데이터 처리 기술	<ul style="list-style-type: none"> <li>• 주어진 짧은 시간(보통 0.1초~1분) 내에 데이터 처리를 보장하는 기술</li> <li>• 배치 데이터 처리와 실시간 데이터 처리를 분리하는 람다 아키텍처가 소개된 이후 계층별 적합한 솔루션이 제공</li> <li>• Apache Spark, Apache Storm 등의 솔루션이 활용</li> </ul>
기계학습 기반 데이터 분석 기술	<ul style="list-style-type: none"> <li>• 기존 통계 기반 데이터 분석 기술과 달리 기계학습과 인공지능 기술을 이용해 예측 분석 등을 하는 기술</li> <li>• 기계학습 기반 분석 알고리즘이 효과적으로 동작하기 위해서는 충분한 학습이 필요한데 학습할 데이터가 많아질 경우 학습 시간이 매우 길어지므로 병렬처리 기법을 이용해 처리 속도를 향상하는 RapidMiner 등과 같은 빅데이터 분석 플랫폼에 대한 개발이 활발</li> </ul>
데이터 시각화 기술	<ul style="list-style-type: none"> <li>• 유용한 데이터 분석 결과를 시각적으로 표현하는 기술로 기업용 보고 도구부터 시작해 오픈 소스 이용 확대로 R, Python을 이용한 다양한 시각화 도구로 확장</li> </ul>



- 빅데이터 솔루션 아키텍처는 데이터를 수집, 저장, 처리, 분석, 시각화와 같은 일련의 단계를 위한 각종 솔루션의 논리적 구성
  - 빅데이터 분석은 저장 및 처리된 데이터를 이용해 모델링 및 학습하는 단계이므로 이전 단계 기술인 분산 스토리지 기술, NoSQL 데이터베이스 기술, 배치 데이터 처리 기술, 실시간 데이터 처리 기술에 영향을 받음

[ 빅데이터 아키텍처 ]



\* 출처: Big Data Architecture Style - Microsoft(2019)

- 빅데이터의 중요도가 높아지면서 주목받고 있는 데이터 활용 역량은 크게 ‘데이터 분석’과 ‘시각적 스토리텔링’으로 이루어짐
  - 데이터 분석은 기술적으로 데이터를 수집하고 정제하는 데이터 가공 단계부터 분석기법을 활용한 데이터 분석에 이르는 전 과정에 필요한 능력
  - 시각적 스토리텔링은 데이터 분석 결과를 시각적으로 표현해 스토리텔링을 하는 능력
- 위 두 역량을 아우르는 ‘데이터 시각화’에 대한 중요성이 강조
  - 기술적으로 데이터를 다루는 전문 역량이 없어도, 시각적 스토리텔링을 위한 디자인 역량이 없더라도 데이터를 활용한 인사이트 발굴, 스토리텔링을 할 수 있도록 도움

## 2. 산업 및 시장 분석

### 가. 산업 분석

#### ◎ 빅데이터 산업의 특징과 전망

- 빅데이터 분석 솔루션을 포함하는 빅데이터 산업의 특징은 규모의 경제, 중소기업 적합 산업, 기술·노동집약적 산업, 경기 동행산업, 라이프 사이클이 짧은 산업, 진입 장벽이 낮은 산업 등으로 요약
  - 중소기업의 설문 조사결과 63.7%는 4차 산업혁명에 대해 잘 알지 못하는 단계라고 응답<sup>16)</sup>
- 텍스트, 동영상, 음성, 센서, GPS, SNS 등의 활용으로 기하급수적인 비정형 데이터가 생성
  - 주요 통신사는 CDR 분석, 네트워크 및 시스템관리 등에 빅데이터 기술을 적용해 서비스 품질 강화에 활용하고 있으며 내부 빅데이터 적용 성과를 사업 모델화해 클라우드 기반 빅데이터 분석 플랫폼을 제공하고 있으며, 공공 분야에 적용해 공익사업에 참여하는 추세
  - 3세대 플랫폼에 기반한 혁신과 가치 창조가 크게 증가할 것으로 전망되며 새로운 핵심 기술이 혁신 가속기로 작용하면서 이 단계를 견인할 것이고 그러면 전 산업에 걸쳐 3세대 플랫폼의 힘과 애플리케이션이 확대될 전망
- 빅데이터가 머신 생성 데이터에 기반을 두고 있는 모든 경제 부문의 새로운 자동화와 파괴적인 혁신의 형태로 가시화된 혜택을 가져오기 위한 빅데이터의 잠재력 실현 조건
  - 빅데이터 기반의 경제성장을 위한 인프라 구축
  - 많은 머신 데이터를 입수
  - 빅데이터 활용 관련 신뢰 및 보안
  - 교통망, 분산형 발전 및 공급, 의료, 비재생 자원의 재사용
  - 산업자동화, 농업 등에 있어 복잡한 최적화 문제를 해결할 때

#### ◎ 데이터 시각화의 중요성과 필요성

- 빅데이터 시장 내에서 비정형 데이터가 전체의 90% 이상을 차지, 사물인터넷, 웨어러블 디바이스의 이용 증가와 미디어 매체의 증가에 따라 하루 수집되는 빅데이터의 양은 약 1,000만 건을 상회할 것으로 예측 (1년이면 약 40억 건 이상의 데이터 수집)
- 데이터 시각화의 장점
  - 많은 양의 데이터를 시각적 요소를 활용해 요약하여 표현하므로 한눈에 볼 수 있으며 흐름을 파악하기 쉬움

16) 중소기업에게 4차 산업혁명이란?(뉴스웨이, 2019.11)

[ 같은 양의 데이터를 엑셀에 입력했을 때(좌)와 데이터 시각화를 했을 때(우) ]



\* 출처: 서울 열린데이터 광장(2019)

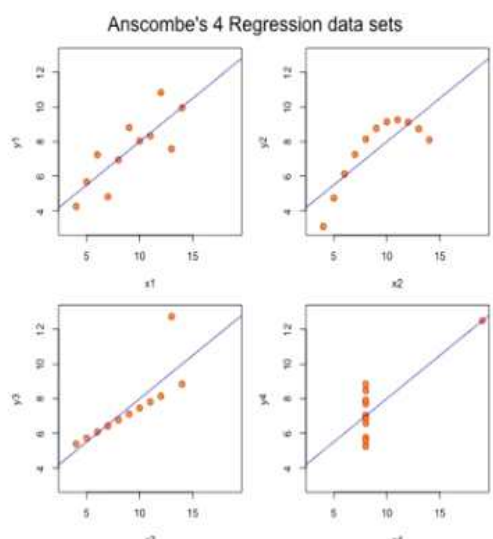
- 데이터 분석에 대한 전문 지식이 없어도 누구나 쉽게 데이터 인사이트를 찾을 수 있음
  - 시각화 요소의 형태, 크기, 위치나 색을 근거로 시각적 패턴을 찾고자 하므로 차트의 시각적 패턴을 근거로 데이터의 추세는 어떤지, 다른 값들과 구별되는 값을 빠르게 찾을 수 있음
  - 시각화를 통해 데이터 활용 범위가 넓어지고, 나아가 새로운 가치 창출의 범위가 확대
- 요약 통계보다 정확한 데이터 분석 결과를 도출 가능
  - 1973년 F.J.Anscombe가 개발한 Anscombes' Quarter는 동일한 요약 통계(평균, 표준편차, 상관관계)를 가진 4개의 데이터셋을 산점도로 시각화하였을 때, 시각적 패턴이 명확히 다르다는 것을 입증하는데 이는 요약 통계정보만으로는 데이터를 정확하게 볼 수 없다는 것을 의미

[ 동일한 요약통계를 가진 4개의 데이터셋(좌)과 이를 시각화를 했을 때(우) ]

	I	II	III	IV
10	8.04	9.14	7.46	6.58
8	6.95	8.14	6.77	5.76
13	7.58	13.74	12.74	7.71
9	6.81	8.77	7.11	8.84
11	8.33	9.26	7.81	8.47
14	9.96	8.1	8.84	7.04
6	7.24	6.13	6.08	5.25
4	4.26	3.1	5.39	12.5
12	10.84	9.13	8.15	5.56
7	4.82	7.26	6.42	7.91
5	5.68	4.74	5.73	6.89

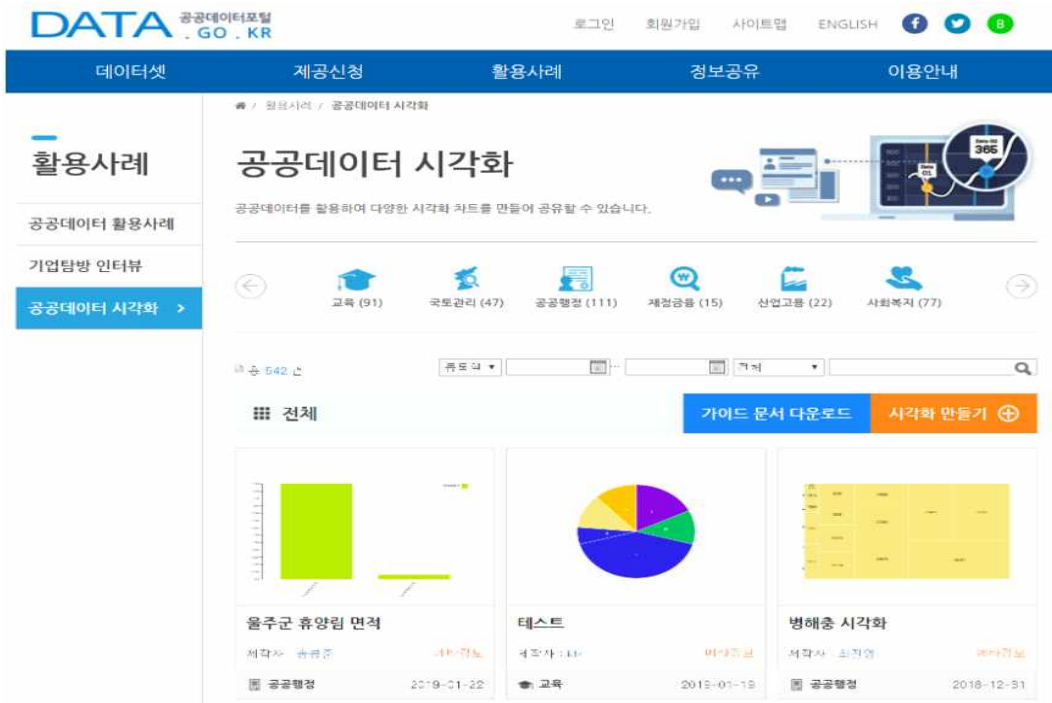
Mean of X	11.0	Correlation between X and Y	0.875
Variance of X	10.0	Linear regression	$y=3.0+0.5x$
Mean of Y	7.5		
Variance of Y	3.75		



\* 출처: Anscombe's Quartet

- 데이터 시각화를 활용할 수 있는 분야와 방법이 무궁무진

[ 공공데이터 포털의 시각화 서비스 ]



\* 출처: 공공 데이터 포털

- 산업 및 기업 차원에서는 보유 데이터의 활용도를 높이기 위해 데이터 시각화를 도입하여 사내 KPI 지표 등 성과 지표 추적, 데이터 기반의 마케팅 활성화 등을 위한 목적으로 데이터 대시보드를 구축하고 활용

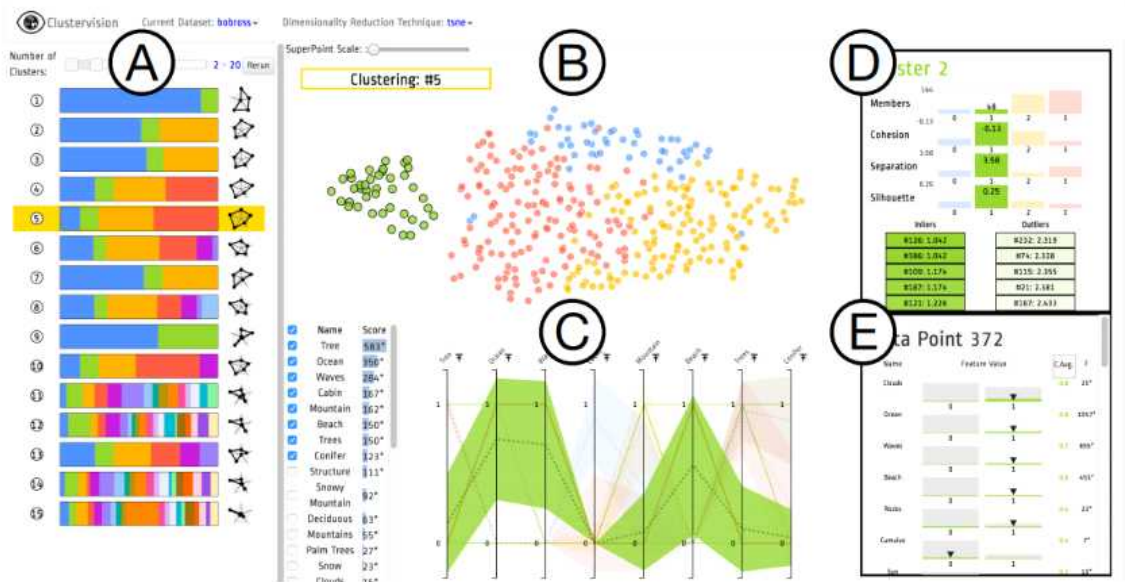
[ 하나은행의 데이터 시각화 기술 기반의 분석 플랫폼 '하나빅 인사이트' ]



\* 출처: 하나 금융 그룹(2018)

- 인공지능 분야에서도 데이터 시각화를 활용한 시각적 분석을 시도하고, 인공지능 학습 결과를 시각적으로 보여주는 방법 등으로 활용

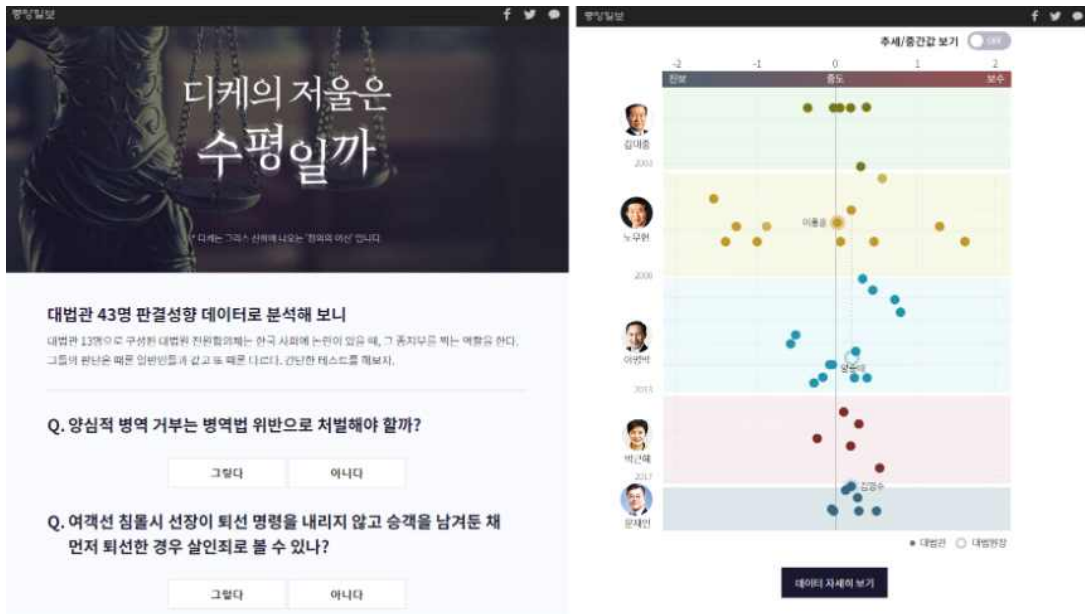
[ 클러스터링을 위한 시각적 분석도구 '클러스터비전' ]



\* 출처: 클러스터비전(2018)

- 언론은 데이터를 기반으로 한 스토리텔링에 방점을 두고, 이름하여 ‘데이터저널리즘’을 실현하는 데 데이터 시각화를 적극적으로 활용

[ 중앙일보의 데이터 시각화 활용 기사 ]



\* 출처: 중앙일보

- 예술의 영역에서도 데이터의 시각적 패턴 그 자체를 예술로 보는 새로운 예술의 형태가 출현하여 이를 두고 데이터아트, 이 분야의 예술가를 데이터 아티스트라고 일컬음



[ 시각화 패턴이 그려진 도자기(좌), 기상 데이터를 조형물로 만든 'Sculptural Musical Score'(우) ]



\* 출처: 중앙일보

- 선진국은 국가 경제의 지속성장 및 일자리 창출을 위해 빅데이터 접목을 통한 주력산업의 재도약과 혁신성장을 도모
  - (독일) 기계산업 '인더스트리 4.0', (일본) 로봇제조산업, (미국) ICT산업, (이스라엘) 항공·방위 산업
  
- 주요국은 미래 경쟁력을 좌우하는 데이터의 중요성을 인식, 데이터산업 활성화를 위해 전략 수립과 투자 확대 등 데이터 패권 경쟁 본격 돌입
  - (미국) 빅데이터 R&D 전략('16) / (EU) 데이터 경제 육성 전략('17) / (일본) Society 5.0 실현 데이터 활용 전개('17) / (중국) 빅데이터 산업 발전계획('17)

## (2) 산업의 구조

- 빅데이터 관련 국내 산업의 구조는 데이터 솔루션, 데이터 구축 및 컨설팅, 데이터 서비스로 분류

### [ 데이터 산업 범위 ]

구분	범위
데이터 솔루션	<ul style="list-style-type: none"> <li>DBMS, 데이터 수집, 데이터모델링, 데이터분석, 검색엔진, 데이터품질, 데이터통합, 데이터보안, 빅데이터플랫폼 등 데이터 처리 및 관리 솔루션 제품을 판매하는 비즈니스</li> <li>라이선스, 개발/커스터마이징, 유지보수에서 매출 발생</li> <li>데이터산업분류 '1.데이터 처리 및 관리 솔루션 개발/공급업'에 해당하는 사업체</li> </ul>
데이터구축/ 컨설팅	<ul style="list-style-type: none"> <li>DB시스템 구축, 기존 DB 정제 및 재구축을 위한 데이터이행, 데이터가공/구축 등 데이터/DB 관련 SI 용역 및 컨설팅 비즈니스를 의미하며, 구축/개발, 유지보수/운영관리, 컨설팅, 제품판매 등을 통해 매출 발생</li> <li>데이터산업분류 '2.데이터구축 및 컨설팅업'에 해당하는 사업체</li> </ul>
데이터 서비스	<ul style="list-style-type: none"> <li>데이터/DB를 원천데이터나 분석 및 활용이 가능한 상태, 또는 수요 맞춤형 데이터/DB를 판매·중개하거나, 데이터를 가공/활용/분석해 온오프라인(모바일, 앱 등 포함)으로 주제·분야별 정보제공서비스나 분석정보제공서비스 등을 제공하는 비즈니스를 의미함</li> <li>데이터 이용료/수수료 등의 직접매출과 광고료 등의 간접매출로 수익이 발생</li> <li>데이터산업분류 '3.데이터 판매 및 제공 서비스업'에 해당하는 사업체</li> </ul>
데이터 인프라	<ul style="list-style-type: none"> <li>데이터 기반 솔루션 구축과 서비스를 위해 반드시 필요한 하드웨어(서버, 스토리지, 네트워크 장비 등)를 직접 수요자에게 공급하거나 클라우드(IaaS, PaaS, DaaS) 비즈니스를 통해 인프라 서비스를 제공하는 비즈니스를 의미</li> <li>데이터인프라 부문에서는 HW판매, 클라우드 구축비/이용료 등을 통한 매출 발생</li> </ul>

\* 출처: 데이터 산업 현황 조사 보고서, 한국데이터진흥원, 2018.

- 데이터 분석에는 정형 데이터 분석, 비정형 데이터 분석, 실시간 데이터 분석, 데이터 시각화 분석이 포함

[ 데이터 분석의 종류 ]

분류	정의	관련 예시
정형/비정형 데이터 분석	<ul style="list-style-type: none"> <li>• (데이터리포팅) DBMS 데이터를 사용자의 이용목적에 맞도록 간편하게 보고서를 생성 및 활용할 수 있도록 지원하는 솔루션</li> <li>• (OLAP) 미리 정의된 관점에 따라 사용자가 기준을 유연하게 변경하면서 데이터를 분석할 수 있도록 지원하는 솔루션</li> <li>• (데이터마이닝) 데이터 집합을 통계적으로 분석하여 의미있는 인사이트를 찾아내는 과정을 지원하는 솔루션 등</li> <li>• (웹데이터분석) 웹사이트 데이터를 추출 및 분석하여 기술 트렌드 및 기업 평판 등을 분석하기 위해 적용하는 솔루션</li> <li>• (텍스트마이닝) 주로 문서 및 웹사이트의 데이터에서 키워드를 찾아내고 빈도 및 연관 관계 등을 분석하는 솔루션 등</li> <li>• (정보추출) 신문기사, 논문, 특허 등의 동향정보로부터 주요 개체 및 개체와 관련된 사실 정보를 정형화된 형태로 추출하여 분석에 활용하는 솔루션</li> <li>• 음성 및 영상 등의 데이터로부터 의미 있는 데이터를 추출하여 분석하는 솔루션</li> <li>• 소셜미디어 내의 인맥 정보를 이용하여 사회관계망을 분석하고, 포스팅되는 내용을 통해 최신이슈와 트렌드를 분석하는 솔루션</li> </ul>	Reporting솔루션, OLAP(OnLine AnalyticalProcessing),통계분석 패키지, 웹데이터분석, 텍스트마이닝,비디오데이터 가공솔루션, 음성데이터가공 솔루션, 트위터,페이스북 등의 API연계 솔루션(Social Analytics 솔루션)
실시간 데이터 분석	<ul style="list-style-type: none"> <li>• 데이터 분석을 통해 위험 신호 또는 징후를 감지하여 위기나 재난에 대처할 수 있도록 예측하는 솔루션</li> </ul>	CEP(Complex EventProcessing), BAM(Business Activity Monitoring)
데이터 시각화 분석	<ul style="list-style-type: none"> <li>• 데이터 세트의 의미를 시각적으로 표현하거나 탐색하는 과정을 지원하는 솔루션</li> </ul>	시각화 및 비주얼 분석 솔루션(Visual Analytics 솔루션)

\* 출처: 데이터 산업 현황 조사 보고서, 한국데이터진흥원, 2018.

- 세계 빅데이터 분석 솔루션 산업은 2009년 세계 금융위기 이후 모바일 통신 환경이 보편화되고 효율적인 기업 경영 활동의 중요성이 높아짐에 따라 글로벌 IT 솔루션 업체를 중심으로 성장
- 기존 데이터베이스 산업을 선도하고 있는 글로벌 IT 솔루션 업체들은 빅데이터 분석 솔루션 전문 업체들을 인수·합병하여 경쟁력 확보



## 나. 시장 분석

### (1) 세계시장

- 데이터 시장 중 시각화 플랫폼과 가장 가깝다고 볼 수 있는 데이터 통합시장으로 '18년 130억 달러에서 연평균 16.2%로 성장하여 '24년 320억 달러까지 성장이 전망됨
  - 전 세계 빅데이터 및 데이터 엔지니어링 서비스(Big Data and Data Engineering Services) 시장은 2018년 약 344억 7,000만 달러에서 연평균 17.6% 성장하여 2024년 약 909억 9,000만 달러에 이를 것으로 전망

#### [ 빅데이터 및 데이터 엔지니어링 서비스의 세계 시장규모 및 전망 ]

(단위 : 백만 달러, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
데이터 통합	13,000	15,110	17,550	20,400	23,700	27,540	32,000	16.2
데이터 모델링	8,440	9,620	10,970	12,500	14,250	16,250	18,530	14.0
데이터 품질	7,620	9,050	10,750	12,780	15,180	18,030	21,420	18.8
분석	540	6,670	8,250	10,200	12,600	15,580	19,260	23.6
계	34,470	40,540	47,670	56,060	65,930	77,370	90,990	17.6

\* 출처: Big Data and Data Engineering Services Market - Global Forecast to 2023, Markets and Markets(연구개발특구진흥재단, 2018. 09)를 기반으로 '24년 네모아이씨지 추정

### (2) 국내시장

- 과학기술정보통신부의 2018 데이터산업 현황 조사(2019.03)에 따르면, 빅데이터 분석 시장은 데이터산업 세부 시장 중 데이터솔루션 중분류별 시장에 속하며 2018년 1,584억 원 규모에서 전체 빅데이터 세계시장 대비 국내시장의 비율로 추산한 연평균 16.3%로 성장하여 2024년에는 3,920억 원 규모로 성장 전망

#### [ 데이터 분석 분야의 국내 시장규모 및 전망 ]

(단위 : 억 원, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
국내시장	1,584	1,842	2,142	2,492	2,898	3,370	3,920	16.3

\* 출처: 2018 데이터 산업 현황 조사, 과학기술정보통신부(2019.03). 인용. '19년~'24년은 전체 빅데이터 세계시장 대비 국내시장의 비율로 추산한 CAGR로 네모아이씨지 재추정

### 3. 기술 개발 동향

- 기술경쟁력
  - 빅데이터 분석 및 시각화 플랫폼은 미국이 최고기술국으로 평가되었으며, 우리나라는 최고기술국 대비 70.6%의 기술 수준을 보유하고 있으며, 최고기술국과의 기술격차는 2.2년으로 분석
  - 중소기업의 기술경쟁력은 최고기술국 대비 65.2%, 기술격차는 2.8년으로 평가
  - EU(81.3%)>중국(73.8%)>한국>일본(70.5%)의 순으로 평가
- 기술수명주기(TCT)<sup>17)</sup>
  - 빅데이터 분석 및 시각화 플랫폼은 5.88의 기술수명주기를 지닌 것으로 파악

#### 가. 기술개발 이슈

- 기반 기술
  - 보편적으로 통용되고 있는 하둡, 아파치 등 기존의 오픈소스 플랫폼을 기반으로 비정형 데이터 수집, 분석, 시각화 및 거버넌스 기능을 제공하는 기술에 대한 개발에 집중
  - ETRI에서는 클라우드컴퓨팅 환경에서 데이터 접근 제어 및 효율적 시스템관리, 자원 할당이 가능하도록 멀티테넌시 기능을 확장한 하둡 플랫폼 기술을 개발
  - 데이터스트림즈에서는 빅데이터와 RDB Microsoft를 동시에 처리할 수 있는 아키텍처 플랫폼 제공
  - IMC에서는 하둡 기반의 데이터 수집 및 전처리, 텍스트 마이닝 및 시각화에 특화된 솔루션 텍스트를 출시
- 데이터 관리 기술
  - 빅데이터 환경에서 데이터의 수집, 등록, 유통, 거래, 관리를 위한 프레임워크 및 개념 모델에 대한 연구가 ETRI를 중심으로 시작
  - 한국데이터진흥원에서는 데이터 스토어를 개방형 플랫폼으로 전환하여 공식 거래소를 개장하였으며 데이터 거래를 위한 기술 개발을 추진 중

17) 기술수명주기(TCT, Technical Cycle Time): 특허 출원연도와 인용한 특허들의 출원연도 차이의 중앙값을 통해 기술 변화속도 및 기술의 경제적 수명을 예측

□ 데이터 분석 기술

- 기본적으로 하둡, NoSQL 등의 빅데이터 분석 인프라 기술이 필요하고, 그 위에 다양한 통계처리, 데이터 마이닝, 텍스트 마이닝, 오피니언 마이닝, 그래프 마이닝 등 다양한 분석 방법 및 기계학습, 인공지능 기법을 적용, 이 중에서 비정형 텍스트 빅데이터에 내재한 가치를 효과적으로 알아내는 텍스트 마이닝, 오피니언 마이닝, 소셜네트워크 분석 기술들이 큰 주목을 받고 있음

[ 빅데이터 분석 기술 ]

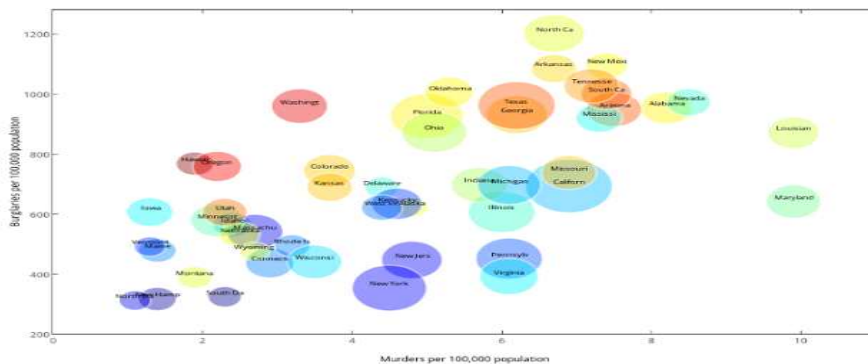


\* 출처: 빅데이터 분석 플랫폼 Truestory

□ 데이터 시각화 기술

- 전통적인 시각화 기술은 주로 시스템 로그나 실험 분석 결과 등에 대한 통계정보를 그래프로 보여주는 방식인데, 빅데이터의 시각화 측면에서는 모든 데이터를 살펴보는 것에 제약이 따르기 때문에 시각화의 기술적인 요소와 더불어 데이터를 요약하고, 한눈에 살펴볼 수 있도록 돕는 시각화 방법론적 요소의 중요성이 커지고 있음
- 데이터 시각화에서는 미적 형태와 기능성 두 가지를 모두 가져가는 것으로 대개 데이터들의 연결과 그룹핑을 표현하는데 초점

[ 데이터 관계 시각화 버블 차트의 예시 ]

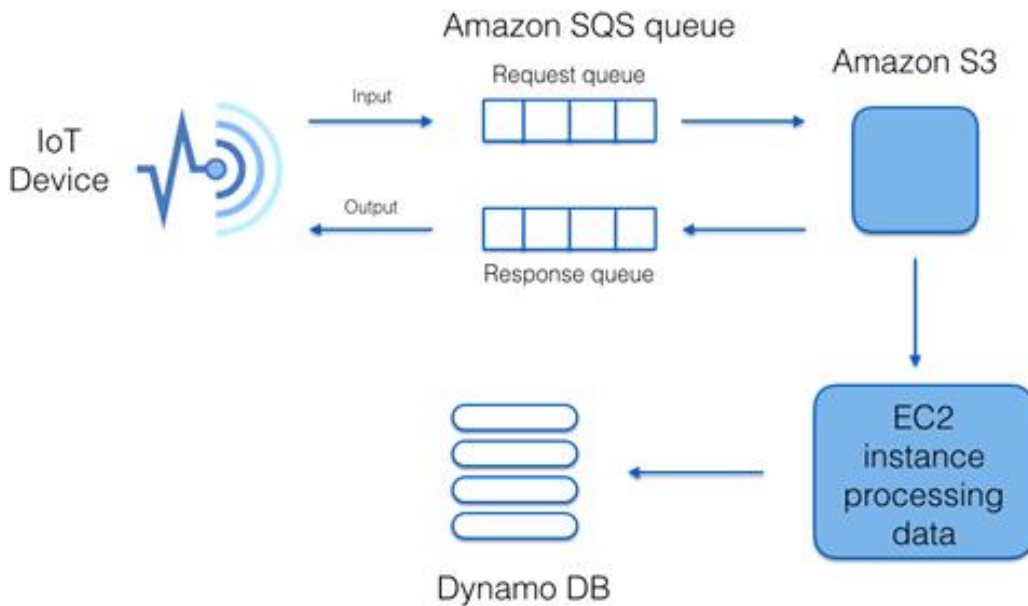


\* 출처: LG CNS - IT Solutions/Big Data, 2016.09

## 나. 기술개발 트렌드

- 기계학습은 일반적으로 통제된 학습 환경에서 저장된 데이터를 사용하였으나 덜 통제된 학습 환경인 IoT 스트리밍 데이터를 실시간으로 학습하는 새로운 모델의 요구
  - 아키텍처와 플랫폼, 구현 요건을 선택하기 이해서는 타깃화된 분석에 대해 이해하고, 데이터 소스, 데이터 처리 요건을 확인
  - 사물은 인간과의 의사소통에 중점을 두면서 다양한 상황에서 유연하고 적절하게 대응함에 따라 스트리밍 데이터를 적용한 복잡하고 정교한 기계학습 알고리즘이 필요
  - 스트리밍 분석은 데이터와 가까운 엣지에서 1차적으로 분석되고 2차적으로 클라우드 서버에 보내어 분석되는 엣지 컴퓨팅 환경에서 이루어질 것으로 예상

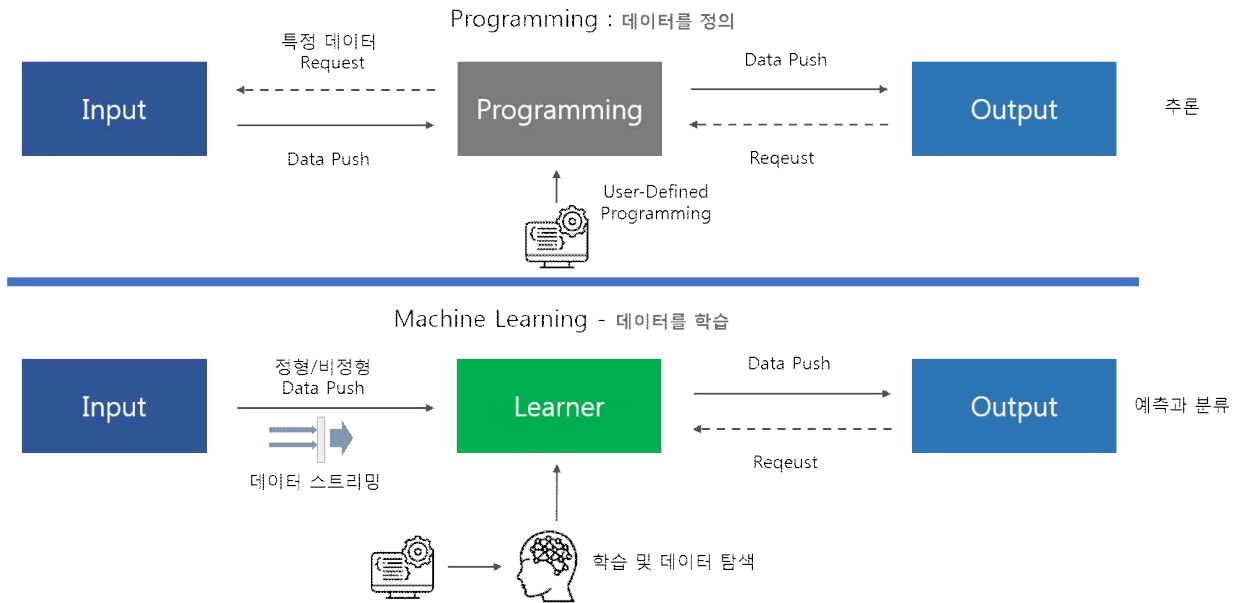
[ 예측 분석을 활용한 IoT 아키텍처 예제 ]



\* 출처: <https://becominghuman.ai>, 2019.

- 빅데이터에 AI와 머신러닝을 융합하여 비즈니스 가치를 대규모로 제공하는 기술
  - AI는 기업들이 데이터를 가장 소중한 기업 자산으로 전환할 수 있도록 해주는 것
  - 인공지능 플랫폼은 기본 작업을 자동화하여 단순하지만, 시간 소모적인 활동을 줄이는 등 여러 가지 방법으로 비용을 절감
  - 인공지능은 신뢰할 수 있는 조연자가 되어 작업이 더욱더 빨리 완료될 수 있도록 지원

[ 프로그램 작성을 위한 기계학습 ]



\* 출처: <http://www.lewware.co.kr>, 2019.

- 데이터 큐레이션이란 데이터 수집과 정제에서 어노테이션과 분류, 학습용 데이터 생성 등 데이터의 활용 가치를 높이기 위한 모든 활동을 의미
  - 인터넷에 널린 정보들을 주제별로, 혹은 관련된 연계성, 연관성을 지닌 무엇인가를 모아서 정리해서 알기 쉽고 접근하기 쉽게 만드는 작업

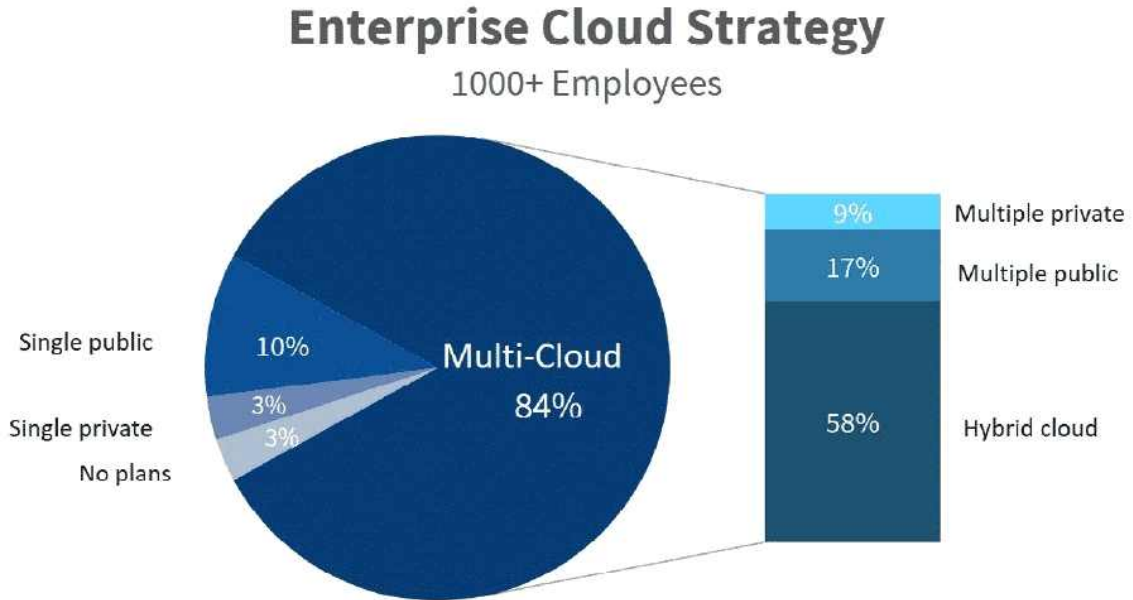
[ 데이터 큐레이션의 과정 ]



\* 출처: [www.prowebscraper.com](http://www.prowebscraper.com), 2019.

- 하이브리드 클라우드는 둘 이상의 퍼블릭 클라우드 또는 프라이빗 클라우드 환경을 결합한 것으로 민감 데이터를 내부에 안전하게 유지하고 다른 데이터를 외부에 유지하는 방법을 제공하는 것
  - 대부분의 퍼블릭 클라우드 서비스는 데이터센터에 설치된 다수의 서버와 스토리지를 가상화 기술을 통해 사용자가 필요한 만큼 할당하는 방식으로 제공되기에 서버당 사용자가 몰릴수록 성능이 저하

[ 하이브리드 클라우드 이용률 ]



\* 출처: RightScale 2019 State of the Cloud Report, Flexera

- 초연결 지능화, 분석 신뢰성 향상, 활용도 제고 등 빅데이터 기술발전 전망을 고려하여 핵심 원천기술 개발
  - 빅데이터 관련 분석(오픈 데이터 플랫폼, 엣지 분석 기술, 공간 빅데이터 분석 기술 등)
  - 인공지능(상황인지형 데이터 관리, 초대규모 모사현실 기술 등)
  - 초연결 지능화(초연결 데이터 관리, 현실-가상공간 실시간 연계 기술, 지능형 집단협업 플랫폼 기술 등) 분야 R&D 지원 강화
  
- 초대용량 데이터의 고속 수집·처리·분석·예측기술 및 빅데이터 기술과 4차 산업혁명 요소기술(AI·IoT·클라우드·정보보호)을 연계한 융합 선도기술 개발 요구
  
- 책임 있는 AI (Responsible AI)
  - 2024년 말에 75%의 기업이 AI 운영으로 전환해 스트리밍 데이터 및 분석 인프라가 5배 증가할 전망
  - 현재 전염병이 발생한 시점에서 머신러닝 및 자연어처리와 같은 일반 AI 기술은 바이러스 확산과 대책의 효과 및 영향에 대해 뛰어난 통찰력과 예측력을 제공한다. 강화 학습 및 분산 학습과 같은 더 똑똑한 AI 기술은 적응력이 뛰어난 시스템을 구축해 복잡한 비즈니스 상황을 처리하는 데 적합
  
- 의사결정 인텔리전스 (Decision intelligence)
  - 2023년에 33%의 대기업이 의사결정 인텔리전스를 사용하는 분석가를 보유할 전망
  - 의사결정을 짓기 위해 여러 논리적 및 수학적 기술이 필요하거나 자동화가 되어야 하거나 문서화가 되어야 하는 경우 의사결정 관리 및 모델링 기술을 통해 탐색 가능

## 다. 생태계 기술 동향

### (1) 해외 플레이어 동향

- (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 ]



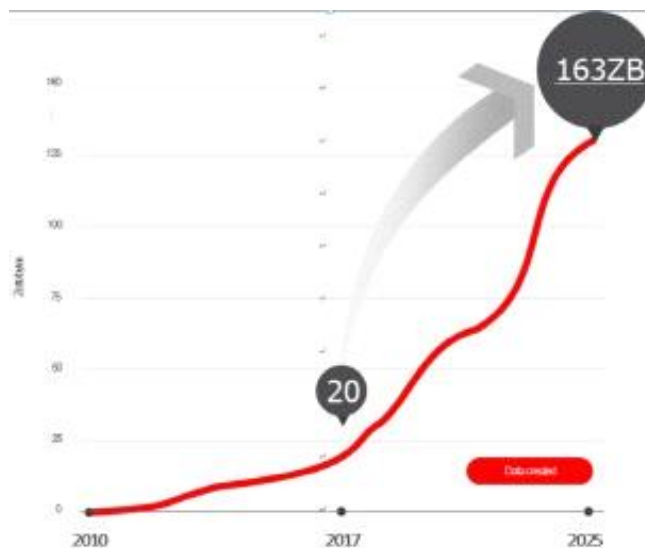
\* 출처: Learnbigdatatools



## (2) 국내 플레이어 동향

- (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

### (3) 국내 중소기업 동향

- (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

## 다. 국내 연구개발 기관 및 동향

### (1) 연구개발 기관

[ 기관별 주요 추진내용 ]

기관	소속	연구 분야
한양대학교	한양대학교 산학협력단 컴퓨터소프트웨어학부	NVRAM aware Database Techniques, Object Systems, Storage Systems, FTL for NAND-Flash based SSD, E-Learning
과학기술정보통신부	울산과학기술원	딥러닝, 단일세포, 세포분류, 유전자 네트워크, 랜덤포레스트
한국과학기술원	식품의약품안전처	빅데이터, 위해정보, 인공지능, 알고리즘
아주대학교	아주대학교 건축학과	데이터시각화, 알고리즘, 설계자동화, 머신러닝, 지능형 시스템

### (2) 기관 기술개발 동향

#### 빅데이터 주요 분야별 구축 방향

- 의료: 의료 인공지능 개발을 위한 학습용 의료데이터 셋 및 의료정보 시스템 구축('18~)
- 제조: 스마트공장 내 스마트센서 기반 제조 공정 빅데이터 구축('18~)
- 농수산업: 농식품 통합 데이터 지도 구축('18~), 어선 정보 등 수산정보 제공 ('18~)
- 도시: 방법, 방재, 3차원 공간정보 등 도시운영 통합 플랫폼 기반 도시 데이터 수집 마련('18~)
- 교통: 자율차, 스마트 도로 관련 차량, 신호등 등 실시간 교통 DB 구축('18~)
- 환경: 대기오염, 수질오염 등 환경 빅데이터 통합·개방 플랫폼 구축('18~)

- 모든 공공데이터는 원시 데이터(raw data) 형태로 최대한 모으고, 이 중 민간 수요가 높은 데이터를 국가중점 데이터로 선정('17. 48개 분야 → '22. 128개 분야), 조기 개방 확대

- (한양대학교) 초연결사회 위험 관리를 위한 빅데이터 기반 사회 환경 실시간 모니터링/사회 시뮬레이션 시스템 개발 - 수집되는 빅데이터와 분석 및 예측 결과, 개발된 사회 모델, 다양한 분석 및 시뮬레이션 기법들을 사회현상 모니터링 시스템과 사회 시뮬레이션 시스템에 지속적으로 통합하고, 이를 시각화함으로써 그룹내 초학제적 융복합연구를 활성화 (2018.08.01. ~ 2025.02.08.)

- (울산과학기술원) 단일세포 빅데이터 분석을 위한 AI 알고리즘 개발 (2020.09.01. ~ 2021.08.31.)
  - 자동 세포분류를 위한 딥러닝 알고리즘 개발
  - 배치효과 보정을 위한 딥러닝 알고리즘 개발
  - 클러스터 시각화
  - 데이터의 아유형별로 유전자 네트워크를 분석하여 질병 상태 또는 아유형별로 유전자 네트워크가 어떻게 변화를 보이는지를 분석
  
- (식품의약품안전처) 빅데이터 기반 위해정보의 인공지능 분석·평가 솔루션 기술 개발 (2020.11.27. ~ 2021.05.26.)
  - 빅데이터 자료원 확보 및 데이터 댐 구축
  - 위해 예측모델 개발을 위한 인공지능 학습용 데이터 구축
  - 위해 이슈 도출·평가를 위한 위해 지표 및 지수 개발
  - 위해지수 산출을 위한 인공지능 알고리즘 개발
  - 위해 감시·관리를 위한 대시보드 개발
  - 인공지능 통합위해정보 감시관리체계 시스템 구축을 위한 중장기로드맵 마련
  
- (아주대학교) 데이터 시각화 기법을 활용한 지능형 외피설계 시스템 구축 연구(2020.06.01. ~ 2023.02.28.)
  - 데이터 시각화를 활용한 지능형 외피 설계 시스템 구축에 관한 연구이며, 연구는 총 3개년에 걸쳐 진행을 계획하고 있으며, 총 6단계의 과정으로 이루어 진행
  - 데이터시각화 알고리즘 구성
  - 알고리즘 분석
  - 외피설계 설계 알고리즘 / 정보연동
  - 성능평가
  - 설계자동화시스템 - 스크립팅 구성 및 활용
  - Program Prototype과 테스트 모형 제작

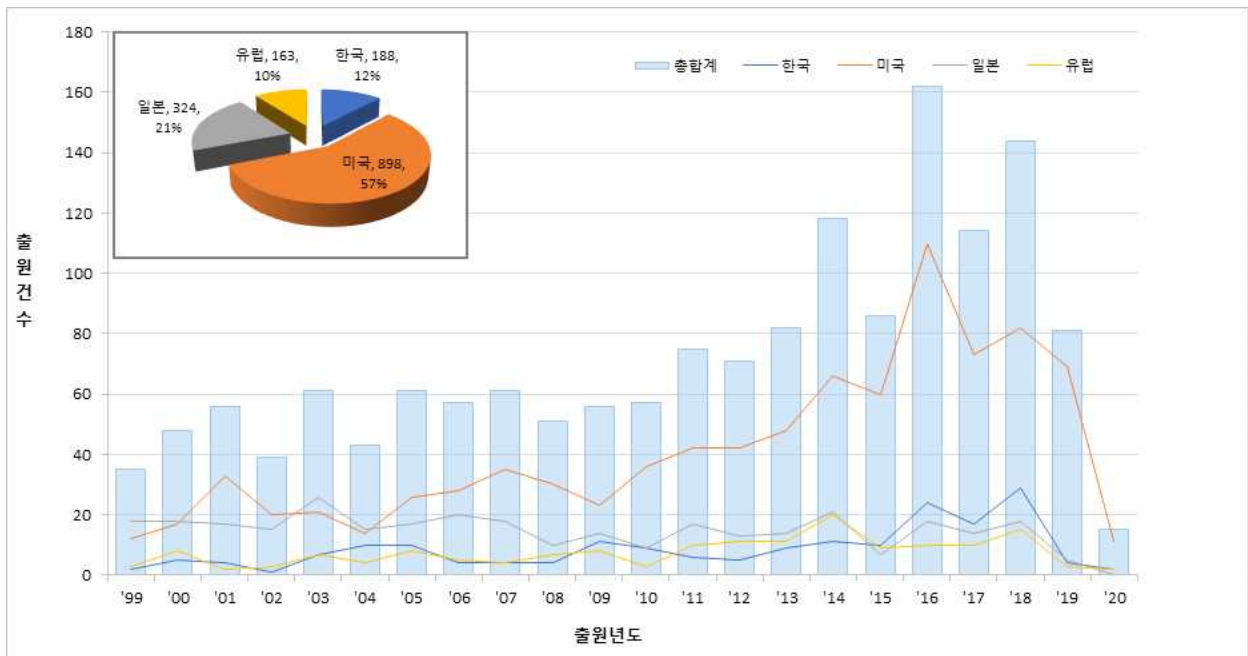
## 4. 특허 동향

### 가. 특허동향 분석

#### (1) 연도별 출원동향

- 빅데이터 분석 및 시각화 플랫폼은 '14년부터 높은 성장을 보임
  - 각 국가별로 살펴보면 미국이 가장 활발한 출원활동을 보이고 있음
- 국가별 출원비중을 살펴보면 미국이 전체의 57%의 출원 비중을 차지하고 있어, 최대 출원국으로 빅데이터 분석 및 시각화 플랫폼 분야를 리드하고 있는 것으로 나타났으며, 일본은 21%, 한국은 12%, 유럽은 10% 순으로 나타남

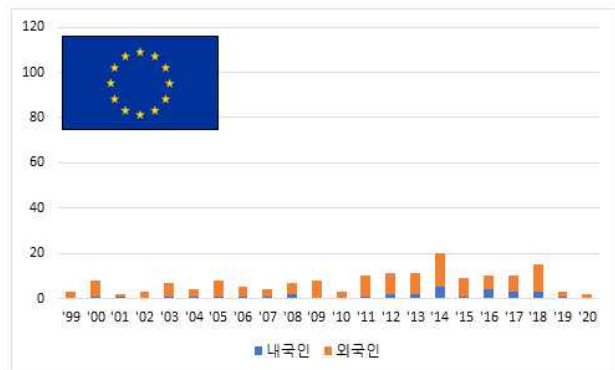
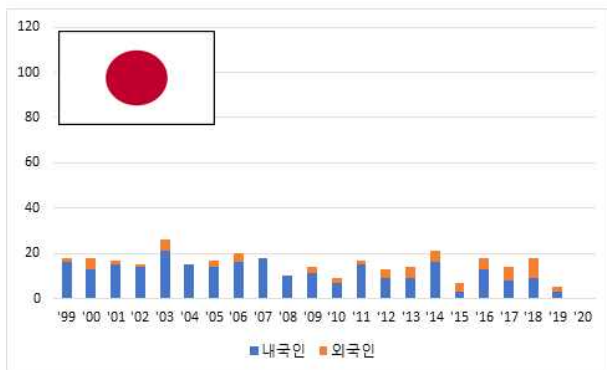
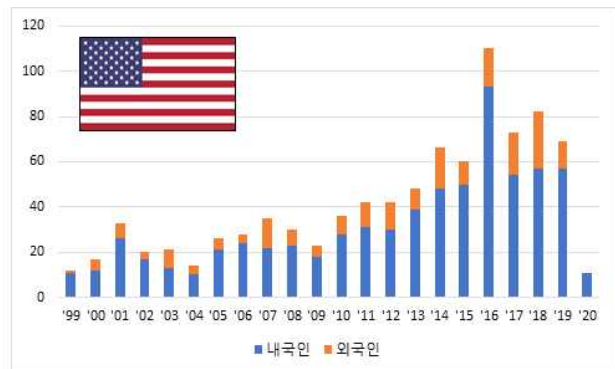
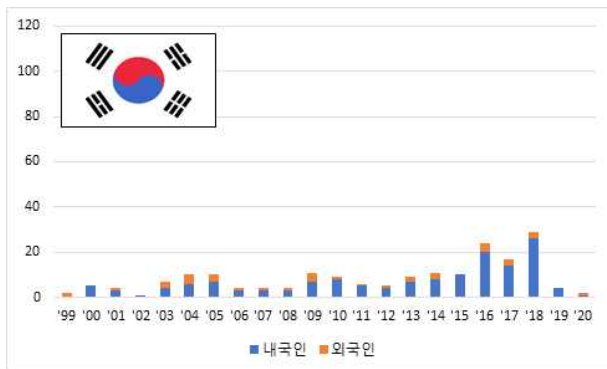
[ 빅데이터 분석 및 시각화 플랫폼 연도별 출원동향 ]



## (2) 국가별 출원현황

- 한국의 출원현황을 살펴보면, '16년부터 해당 기술의 출원이 급격히 증가하는 추세
  - 내국인 위주의 출원이 진행되고 있음
  - 한국 기술의 양적 흐름은 미국과 상당히 유사
  - 미국의 출원 수에 비해 21% 정도의 수준을 보임
- 미국의 출원현황을 살펴보면 분석구간 초기부터 전체 특허기술의 출원 증감 흐름에 영향을 주고 있는 것으로 나타남. 미국의 경우, 한국에 비해 외국인의 비중이 큰 것으로 나타남
- 유럽의 출원현황을 살펴보면 매년 20건 이하로, 뚜렷한 증감 동향이 나타나지 않음
- 일본의 출원현황은 출원수가 매년 30건 이하로, 뚜렷한 증감 동향이 나타나지 않음. 해당 기술 분야에서 일본 시장에 대한 관심도가 높지 않은 것으로 보임

[ 국가별 출원현황 ]



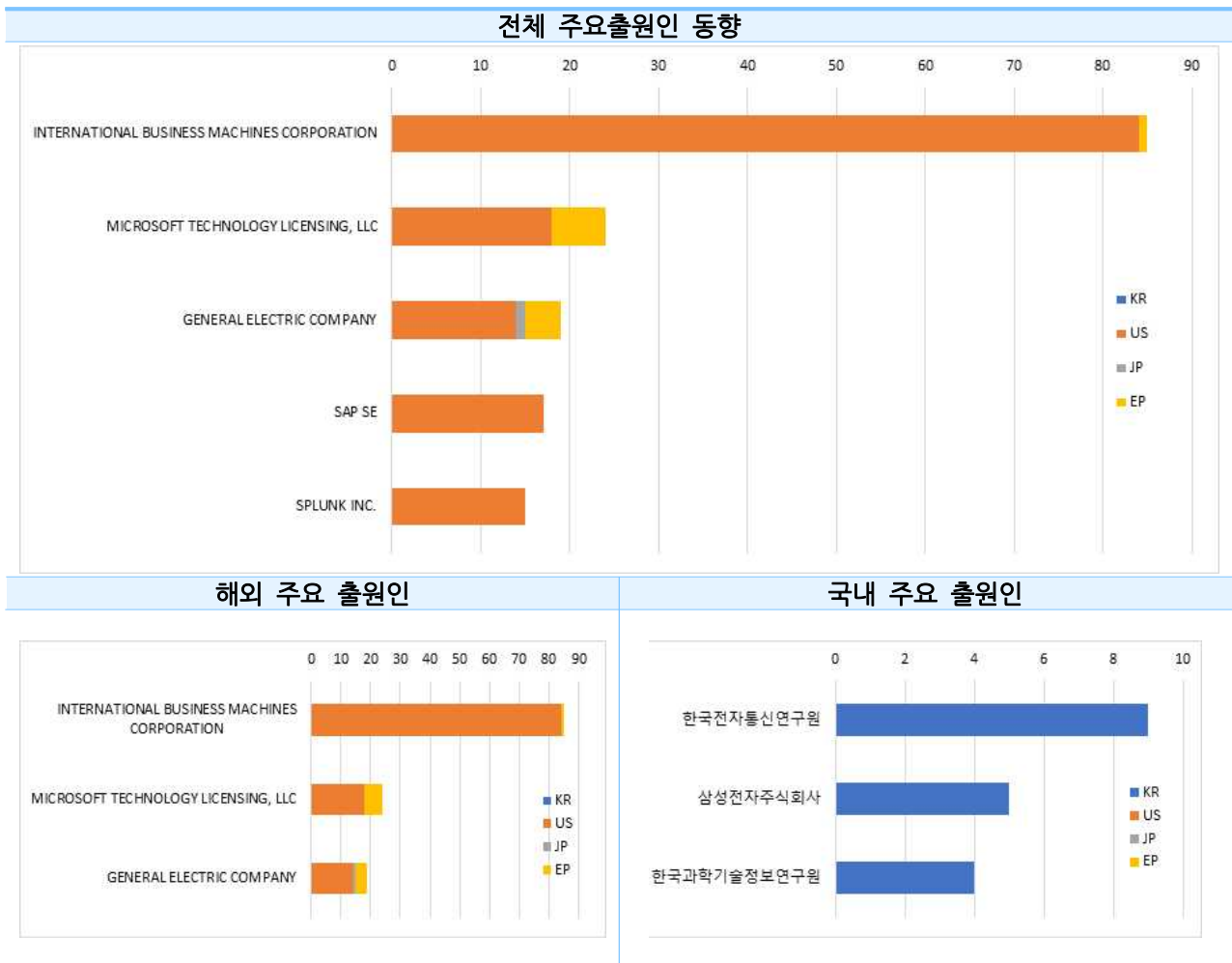




## 나. 주요 출원인 분석

- 빅데이터 분석 및 시각화 플랫폼의 전체 주요출원인을 살펴보면, 주로 미국 국적의 출원인이 다수 포함되어 있는 것으로 나타났으며, 제 1 출원인으로는 미국의 INTERNATIONAL BUSINESS MACHINES CORPORATION인 것으로 나타남
  - 제 1 출원인인 INTERNATIONAL BUSINESS MACHINES CORPORATION의 출원은 미국에 집중된 경향을 보임
- 빅데이터 분석 및 시각화 플랫폼 관련 기술로 빅데이터 솔루션을 다루는 대기업에 의한 출원이 대다수를 차지
  - 국내에서는 연구기관/대학, 대기업의 활발한 출원이 이루어짐

[ 빅데이터 분석 및 시각화 플랫폼 주요출원인 ]



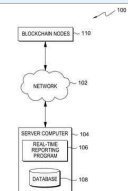
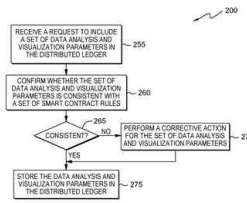
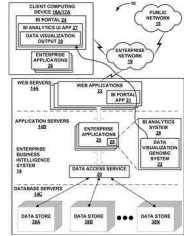
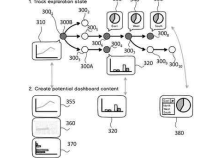
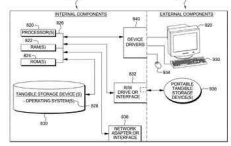


(1) 해외 주요출원인 주요 특허 분석

◎ IBM(International Business Machines)

- IBM은 미국의 다국적 기업으로, 빅데이터 분석 및 시각화 플랫폼과 관련하여, 블록체인을 통한 데이터 분석 및 시각화 기술에 특화된 특허를 다수 출원. 그 중 등록된 특허는 57건
  - 주요 특허들은 블록 체인 패브릭 분산 원장(distributed ledger)과 같은 분산 원장에 데이터 분석 및 시각화 파라미터들의 세트에 기초하여 리포트를 생성하는 기술 특허를 다수 출원하는 것으로 파악

[ IBM 주요특허 리스트 ]

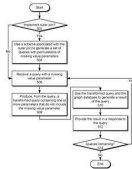
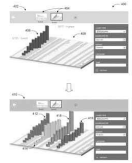
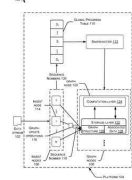
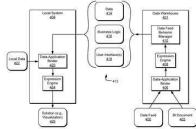
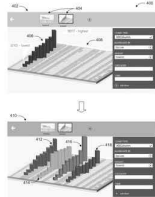
등록번호 (출원일)	명칭	기술적용분야	도면
US10691678 (2019.02.13)	Consistent reporting using blockchain	블록체인 패브릭 분산 원장(distributed ledger)과 같은 분산원장에 데이터 분석 및 시각화 파라미터들의 세트를 포함하는 기술	
US10289716 (2017.10.31)	Consistent reporting using blockchain	하나 이상의 프로세서들은 데이터 분석 및 시각화 파라미터들의 세트를 분산 원장(distributed ledger)에 저장하고, 데이터 분석 및 시각화 파라미터들의 세트에 기초하여 데이터세트의 리포트를 생성하는 기술	
US10769162 (2017.03.16)	Generating business intelligence analytics data visualizations with genomically defined genetic selection	데이터 시각화 요소를 암호화하는 디지털 유전자 및 데이터 시각화의 자동화된 생성을 위한 유전자 선택 프로세스의 기초로서 요소에 대한 잠재적인 증분 변화를 유전학적으로 정의하는 기술	
US10706598 (2016.12.14)	Interface for data analysis	데이터 수집과 관련된 적어도 하나의 대시보드 인스턴스는 사용자 상호 작용 데이터 및 잠재적 시각화 세트에 기초하여 정의하는 기술	
US9760913 (2016.10.24)	Real time usability feedback with sentiment analysis	키오스크 장치로부터 사용자 소프트웨어 피드백 및 충돌 해결하고, 키오스크 장치의 사용자와 소프트웨어 사이의 상호작용 동안 사용자로부터 오디오 및 시각화 데이터를 캡처하는 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ MICROSOFT TECHNOLOGY LICENSING, LLC

- MICROSOFT TECHNOLOGY LICENSING, LLC는 Microsoft의 자회사로, 회사의 특허 및 기술 이전 관련 기업으로, 3차원 데이터 시각화 관련 특허를 보유하고 있으며, 빅데이터 분석 및 시각화 플랫폼 기술과 관련한 특허를 미국 및 유럽에 출원
  - 데이터를 변환하여 시각화 하는 기술과 관련하여 24건의 미국 및 유럽 출원을 진행하였으며, 그 중 등록된 특허는 11건으로 파악

[ MICROSOFT TECHNOLOGY LICENSING, LLC 주요특허 리스트 ]

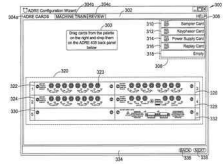
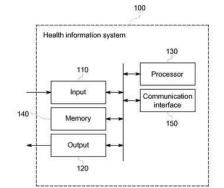
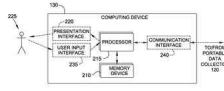
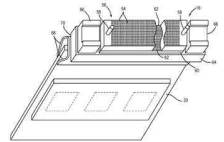
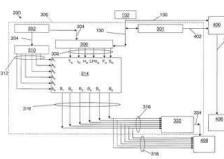
등록번호 (출원일)	명칭	기술적용분야	도면
US10417230 (2016.10.26)	Transforming and evaluating missing values in graph databases	그래프 데이터베이스의 질의를 처리하기 위한 시스템	
US10067635 (2016.05.19)	Three dimensional conditional formatting	평판 형태의 데이터를 3차원 영상으로 변환하거나, 기초 데이터와 관련된 영상 또는 그래픽을 선택하여 배경 영상을 생성하는 기술	
US9589069 (2016.01.21)	Platform for continuous graph update and computation	스트리밍 데이터 처리하기 위한 그래프 마이닝 연산을 수행하는 기술	
US10379711 (2015.07.08)	Data feed having customizable analytic and visual behavior	데이터 송신 장치 회사가 분석적 및 시각적인 특징을 가지는 사용자 정의 프리뷰 적용에 대한 데이터 피드를 연합시키도록 허용하는 데이터 시장 인프라 기술	
US9360992 (2013.07.29)	Three dimensional conditional formatting	3차원의 조건적 포매팅 기반의 데이터 시각화 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ GENERAL ELECTRIC COMPANY

- GENERAL ELECTRIC COMPANY는 미국의 다국적 기업으로 빅데이터 분석 관련 특허를 19건 출원하였으며, 그 중 등록 특허 6건 보유
  - 빅데이터 분석 및 시각화 플랫폼에서 빅데이터를 획득하고, 이를 분석하여 가시화 하는 응용 특허를 다수 보유

[ GENERAL ELECTRIC COMPANY 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
US10359912 (2016.11.11)	Methods and apparatus for configuring a data analyzer	데이터 분석기의 그래픽 표현을 프로세서에 연결된 프리젠테이션 인터페이스 상에 사용자에게 디스플레이하는 기술	
US10216902 (2015.08.31)	Methods and systems for improving connections within a healthcare ecosystem	헬스케어 에코시스템에서 소스 및 타겟 시스템들 사이에서 데이터를 분석하는 기술	
US9507484 (2012.12.27)	Methods and apparatus for configuring a data analyzer	프로세서를 사용하여, 구성 파일은 데이터 분석기 구성 및 기계 트레인의 그래픽 표현에 기초하여 생성하는 기술	
US8761479 (2010.11.08)	System and method for analyzing and visualizing spectral ct data	스펙트럼 CT 데이터를 포함하는 환자로부터 획득된 이미지 데이터를 활용하여 스펙트럼 CT 데이터를 분석하고 가시화하기 위한 기술	
US8095588 (2007.12.31)	Method and system for data decomposition via graphical multivariate analysis	하나 이상의 변수에 대한 모델 데이터 및 실제 데이터를 생성하기 위한 수학적 모델을 이용한 다변량 분석 기술	

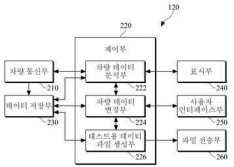
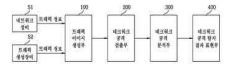
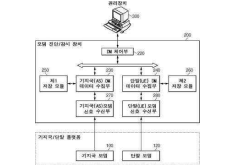
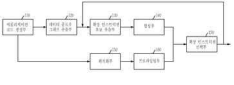
\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

(2) 국내 주요출원인 주요 특허 분석

◎ 한국전자통신연구원

- 한국전자통신연구원은 한국의 정보, 통신, 전자, 방송 및 성과 관련 융·복합기술 분야를 연구하는 연구 기관으로, 빅데이터 분석 및 시각화 플랫폼에 있어서, 빅데이터를 기반으로 분석하고, 시각화 하는 기술과 관련된 특허를 다수 출원
  - 한국전자통신연구원은 빅데이터 분석 및 시각화 플랫폼에 있어서, 데이터 분석 기술과 관련된 특허를 국내를 중심으로 특허를 출원. 등록 건수는 4건으로 파악

[ 한국전자통신연구원 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
KR1586051 (2011.05.31)	제품 테스트용 차량 데이터 제공 장치 및 방법	차량 데이터 파일을 분석하여, 차량 데이터 파일에 포함된 기준 데이터에 대한 기준 데이터 그래프 생성 기술	
KR1219538 (2009.07.29)	비주얼 데이터 분석 기반의 네트워크 공격 탐지 장치 및 그 방법	대량의 트래픽 데이터를 실시간으로 처리 할 수 있고, 트래픽 이미지를 생성할 때 다양한 정보들을 삽입하여 기존의 트래픽 볼륨 기반의 네트워크 공격 탐지 기법들이 탐지하지 못했던 공격들을 탐지하는 기술	
KR0658565 (2005.12.10)	기지국/단말 모뎀 진단/감시 시스템 및 그 방법	기지국 모뎀과 단말 모뎀의 연동 시험 시에 발생하는 오류 원인을 효율적으로 분석하여 사용자 또는 관리자에게 제공하는 기술	
KR0552675 (2003.12.26)	확장 인스트럭션 선택 장치 및 그 방법	확장 인스트럭션 선택 장치에 있어서, 어플리케이션 코드의 데이터 플로우를 분석하여 데이터 플로우 그래프를 추출하는 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ 삼성전자 주식회사

□ 삼성전자는 빅데이터 분석 및 시각화 플랫폼 기술에 있어서, 데이터 처리 및 시각화하는 기술과 관련된 특허 출원, 등록 건수는 2건으로 파악

[ 삼성전자 주식회사 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
KR1615707 (2009.09.17)	데이터 처리 장치 및 방법	데이터 처리 장치 및 당된 인터페이스를 이용하여 상기 입력 데이터를 시각화하는 기술	
KR0571960 (2004.01.09)	화상형성장치의 제어장치 및 방법	인쇄 대상 페이지 데이터를 병렬로 처리함으로써 인쇄 대상 데이터를 고속으로 처리하는 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ 한국과학기술정보연구원

- 한국과학기술정보연구원은 한국의 국가 연구기관으로, 빅데이터 분석 및 시각화 플랫폼 기술과 관련된 특허를 4건 출원하였으며, 등록 건수는 2건으로 파악

[ 한국과학기술정보연구원 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
KR1629394 (2014.12.09)	대용량 데이터 분석을 위한 r의 연계방법, 시각화 방법 및 프로그램을 저장한 저장매체	R의 연계를 통해 대용량 데이터 분석하고, R로부터 수신한 분석 결과를 그래프 형태로 디스플레이하는 기술	
KR1551320 (2013.07.10)	다중 리소스 분석을 기반으로 연구자 내면 역량을 분석하는 데이터 분석 장치, 그 방법 및 그 프로그램을 저장하는 저장매체	연구자와 관련된 개체정보 데이터를 분석하고, 분석하여 내면역량지표들을 시각화하여 표출하는 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

## 다. 기술진입장벽 분석

### (1) 기술 집중력 분석

- 빅데이터 분석 및 시각화 플랫폼관련 기술에 대한 시장관점의 기술독점 현황분석을 위해 집중률 지수(CRn: Concentration Ratio n, 상위 n개사 특허점유율의 합) 분석 진행
  - 상위 4개 기업의 시장점유율이 0.09로 빅데이터 분석 및 시각화 플랫폼 분야에 있어서 독과점 정도는 매우 낮은 수준으로 판단
  - 국내 시장에서 중소기업의 점유율 분석결과 0.58로 해당 기술에 대하여 중소기업의 진입이 용이하다고 판단됨

[ 주요출원인의 집중력 및 국내시장 중소기업 집중력 분석 ]

	주요출원인	출원건수	특허점유율	CRn	n
주요 출원인 집중력	INTERNATIONAL BUSINESS MACHINES CORPORATION(미국)	85	5.4%	0.05	1
	MICROSOFT TECHNOLOGY LICENSING, LLC(미국)	24	1.5%	0.07	2
	GENERAL ELECTRIC COMPANY(미국)	19	1.2%	0.08	3
	SAP SE(독일)	17	1.1%	<b>0.09</b>	<b>4</b>
	SPLUNK INC.(미국)	15	1.0%	0.10	5
	HITACHI LTD(일본)	15	1.0%	0.11	6
	PALANTIR TECHNOLOGIES INC.(미국)	14	0.9%	0.12	7
	MICROSOFT CORPORATION(미국)	14	0.9%	0.13	8
	TABLEAU SOFTWARE, INC.(미국)	13	0.8%	0.14	9
	ACCENTURE GLOBAL SOLUTIONS LIMITED(아일랜드)	12	0.8%	0.14	10
	<b>전체</b>	<b>1573</b>	<b>100%</b>	<b>CR4=0.09</b>	
국내시장 중소기업 집중력	출원인 구분	출원건수	특허점유율	CRn	n
	중소기업(개인)	99	57.9%	0.58	
	대기업	26	15.2%		
	연구기관/대학	46	26.9%		
	<b>전체</b>	<b>171</b>	<b>100%</b>	<b>CR중소기업=0.58</b>	

## (2) 특허소송 현황 분석

- 빅데이터 분석 및 시각화 플랫폼 관련 기술 진입 장벽에 대한 분석을 위해 특허소송을 이력 검토
  - 2019년 8월 대법원에 원고 InvestPic, LLC와 SAP America, Inc.간의 통계학적 분석을 디스플레이 하는 장치에 대한 특허 침해소송이 종결
  - 2020년 10월 텍사스 서부지방법원에 원고 Slyce Acquisition Inc. 와 Syte - Visual Conception Ltd.간의 데이터 기반 이미지 분석을 통한 시각화 시스템에 대한 특허 침해소송이 종결
  - 2018년 7월 뉴저지 지방방법원에 원고 DIOPSY, INC. 와 KONAN MEDICAL USA, INC..간의 전자망막검사 장치에 대한 특허 침해소송이 종결
  - 데이터를 분석하고, 이를 시각화 하거나 디스플레이 하는 것과 관련하여 기업 간에 특허소송을 진행하는 등 국내기업이 미국시장에 진입하는 경우, 진입장벽으로 작용할 수 있음

[ 빅데이터 분석 및 시각화 플랫폼 관련 특허소송 현황 ]

번호	사건번호 (연도.월.일)	명칭	출원인	원고 v. 피고
		Method and system for analysis, display and dissemination of financial information using resampled statistical methods	Attractor Holdings LLC	InvestPic, LLC v. SAP America, Inc.
1	US96349291 (2002.02.19)	대상제품명	소제기일	소송종료일
		Enterprise GRC, and Access Risk Management and other similar products (Declaratory Judgment)	2019.03.14	2019.08.23
2	US9152624 (2015.10.06)	명칭	출원인	원고 v. 피고
		Systems and methods for visual presentation and navigation of content using data-based image analysis	Retail Optimization International, Inc	Slyce Acquisition Inc. v. Syte - Visual Conception Ltd.
3	US9026189 (2015.05.05)	대상제품명	소제기일	소송종료일
		Scan & Shop tool of Kohl's mobile app for iOS devices through Apple App Store and also for Android devices through Google Play Store	2019.04.11	2020.10.02
3	US9026189 (2015.05.05)	명칭	출원인	원고 v. 피고
		Electrode sensor assembly for electroretinography and pattern electroretinography	Diopsys, Inc.	DIOPSY, INC. v. KONAN MEDICAL USA, INC.
3	US9026189 (2015.05.05)	대상제품명	소제기일	소송종료일
		Instruments for vision testing (Counterclaim Filed)	2015.07.30	2018.07.23

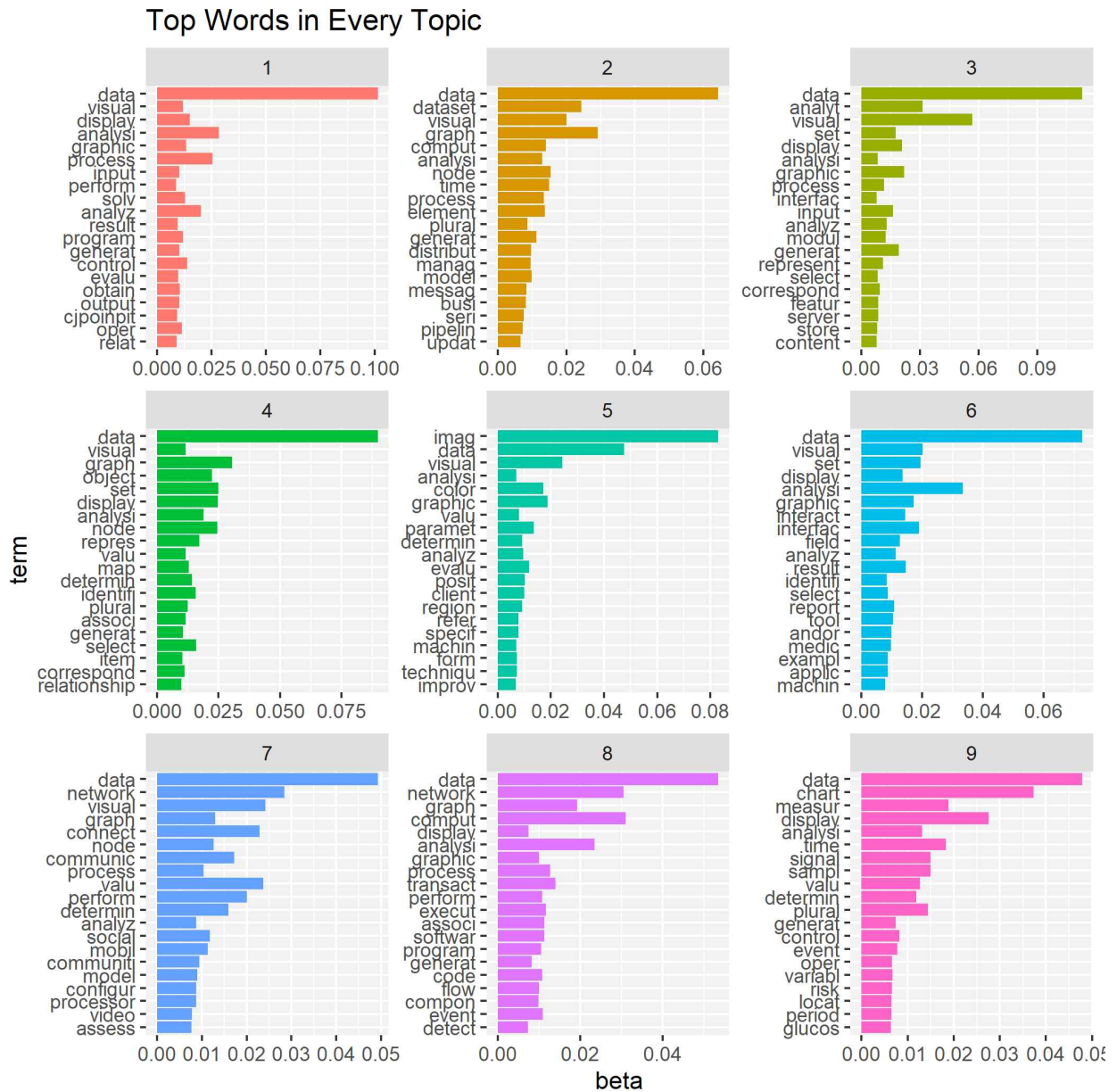


## 5. 요소기술 도출

### 가. 특허 기반 토픽 도출

- 1,230개의 특허의 내용을 분석하여 구성 성분이 유사한 것끼리 클러스터링을 시도하여 대표성이 있는 토픽을 도출

[빅데이터 분석 및 시각화 플랫폼에 대한 토픽 클러스터링 결과]



## 나. LDA<sup>18)</sup> 클러스터링 기반 요소기술 도출

[ LDA 클러스터링 기반 요소기술 키워드 도출 ]

No.	상위 키워드	대표적 관련 특허	요소기술 후보
클러스터 01	data, analysis, process, analyze, display, control, graphic, solve, program, visual	<ul style="list-style-type: none"> <li>Apparatus for measuring superhigh frequency partial discharge and partial discharge position for high-voltage power apparatus</li> <li>Train monitor data analyzer</li> </ul>	통계 기반 데이터 처리 기술
클러스터 02	data, graph, dataset, visual, node, time, computing, element, process, analysi	<ul style="list-style-type: none"> <li>Pipeline management supporting server and pipeline management supporting system</li> <li>Visualization tool for parallel dependency graph evaluation</li> </ul>	데이터 시각화 기술
클러스터 03	data, visual, analytic, graphic, display, generate, set, input, analyze, module	<ul style="list-style-type: none"> <li>Intellectual property analysis and report generating system and method</li> <li>Vao productivity suite</li> </ul>	통계 기반 데이터 분석 기술
클러스터 04	data, visual, element, graphic, pipeline, object, secret, trade, axis, node	<ul style="list-style-type: none"> <li>Algorithm complexity identification</li> <li>Data processing, analysis, and visualization system for use with disparate data types</li> </ul>	배치 데이터 처리 기술 데이터 시각화 기술
클러스터 05	image, data, visual, graphic, color, parameter, evaluate, posit, client, analyze	<ul style="list-style-type: none"> <li>Image-processing apparatus, image-capturing apparatus, image-processing method and image-processing program</li> <li>Image processor, printer, image processing method and printing method</li> </ul>	대용량 및 실시간 데이터 분석 플랫폼 기술
클러스터 06	data, analysis, visual, set, interface, graphic, result, interact, display, field	<ul style="list-style-type: none"> <li>Method and system for joint evaluating a medicinal image database</li> <li>Apparatus, system and methods for assessing drug efficacy using holistic analysis and visualization of pharmacological data</li> </ul>	데이터 시각화 기술
클러스터 07	data, network, visual, value, connect, perform, communicate, determine, graph, node	<ul style="list-style-type: none"> <li>Method for determining a dioptric parameter of an ophthalmic lens to be provided to a person</li> <li>Systems and methods for conducting more reliable assessments with connectivity statistics</li> </ul>	대용량 및 실시간 데이터 분석 플랫폼 기술
클러스터 08	data, computing, network, analysis, graph, transact, process, execute, software, associate	<ul style="list-style-type: none"> <li>Method and apparatus for documentation, analysis, auditing, accounting, protection, registration, and verification of trade secrets</li> <li>Information security analysis system</li> </ul>	내외부 데이터 연동 기술
클러스터 09	data, chart, display, measure, time, signal, sample, plural, analysi, value	<ul style="list-style-type: none"> <li>Data signal quality evaluation method and apparatus using high speed sampling</li> <li>Operation arrangement proposal preparation information and operation arrangement proposal preparation device</li> </ul>	머신러닝 기반 데이터 분석 기술

18) Latent Dirichlet Allocation

## 다. 특허 분류체계 기반 요소기술 도출

- 빅데이터 분석 및 시각화 플랫폼 관련 특허에서 총 10개의 주요 IPC코드(메인그룹)를 산출하였으며, 각 그룹의 정의를 기반으로 요소기술 키워드를 아래와 같이 도출

[ IPC 분류체계에 기반한 요소기술 도출 ]

IPC 기술트리		
(서브클래스) 내용	(메인그룹) 내용	요소기술 후보
(G06F) 전기에 의한 디지털 데이터처리	• (G06F-015) 디지털 컴퓨터 일반	-
	• (G06F-016) 정보 검색; 그 데이터베이스 구조; 그 파일 시스템 구조	-
	• (G06F-017) 디지털 컴퓨팅 또는 데이터 프로세싱 장비, 방법으로서 특정 기능을 위해 특히 적합한 형태의 것	-
	• (G06F-009) 프로그래머를 위한 장치, 예. 제어장치	-
	• (G06F-003) 컴퓨터로 처리할 수 있는 형식으로 전송된 데이터를 변환하는 입력기구; 처리장치로부터 출력장치로 데이터를 전송하기 위한 출력기구, 예. 인터페이스 기구	-
	• (G06F-019) 특수한 어플리케이션에 특히 적합한 디지털 컴퓨팅 또는 데이터 처리 장치 또는 방법	-
(G06K) 데이터의 인식; 데이터의 표시; 기록매체; 기록매체의 취급	• (G06K-009) 인쇄문자, 손으로 쓴 문자를 독취하거나 인식 또는 패턴을 인식하기 위한 방법 또는 장치, 예. 지문인식	-
(G06Q) 관리용, 상업용, 금융용, 경영용, 감독용 또는 예측용으로 특히 적합한 데이터 처리 시스템 또는 방법; 그 밖에 분류되지 않는 관리용, 상업용, 금융용, 경영용, 감독용 또는 예측용으로 특히 적합한 시스템 또는 방법	• (G06Q-010) 경영; 관리	통계 기반 데이터 분석 기술
	• (G06Q-050) 특정 사업 부문에 특히 적합한 시스템 또는 방법, 예. 공익사업 또는 관광	배치 데이터 처리 기술
(G06T) 이미지 데이터 처리 또는 발생 일반	• (G06T-011) 이차원 (2D) 화상 생성	-

## 라. 최종 요소기술 도출

- 산업·시장 분석, 기술(특허)분석, 전문가 의견, 타부처 로드맵, 중소기업 기술수요를 바탕으로 로드맵 기획을 위하여 요소기술 도출
- 요소기술을 대상으로 전문가를 통해 기술의 범위, 요소기술 간 중복성 등을 조정·검토하여 최종 요소기술명 확정

### [ 빅데이터 분석 및 시각화 플랫폼 분야 요소기술 도출 ]

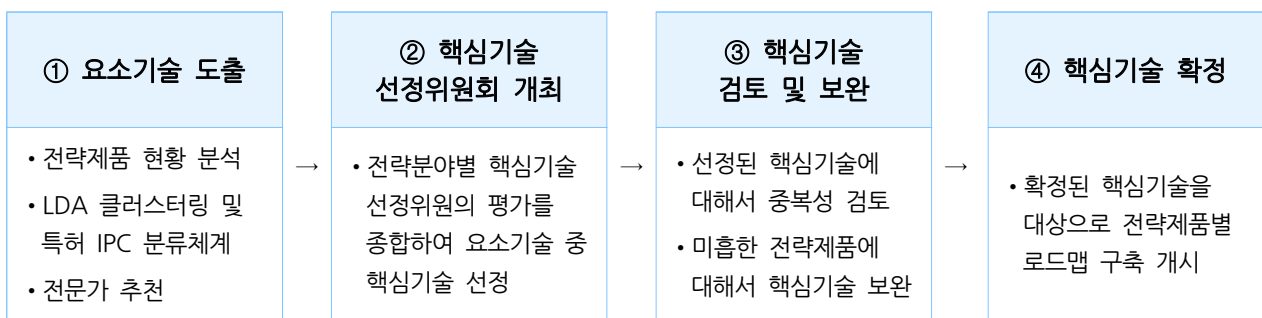
요소기술	출처
대용량 및 실시간 데이터 분석 플랫폼 기술	특허 클러스터링, 전문가추천
데이터 시각화 기술	특허 클러스터링, 전문가추천
머신 러닝 기반 데이터 분석 기술	전문가추천
실시간 데이터 처리 기술	특허 클러스터링, 전문가추천
내외부 데이터 연동 기술	전문가추천
통계 기반 데이터 분석 기술	기술분류체계, 전문가추천
NoSQL 데이터베이스 기술	전문가추천
배치 데이터 처리 기술	기술분류체계, 전문가추천
분산 스토리지 기술	전문가추천
실시간 데이터 저장 기술	특허 클러스터링, 전문가추천

## 6. 전략제품 기술로드맵

### 가. 핵심기술 선정 절차

- 특허 분석을 통한 요소기술과 기술수요와 각종 문헌을 기반으로 한 요소기술, 전문가 추천 요소기술을 종합하여 요소기술을 도출한 후, 핵심기술 선정위원회의 평가과정 및 검토/보완을 거쳐 핵심기술 확정
- 핵심기술 선정 지표: 기술개발 시급성, 기술개발 파급성, 기술의 중요성 및 중소기업 적합성
  - 장기로드맵 전략제품의 경우, 기술개발 파급성 지표를 중장기 기술개발 파급성으로 대체

#### [ 핵심기술 선정 프로세스 ]



### 나. 핵심기술 리스트

#### [ 빅데이터 분석 및 시각화 플랫폼 분야 핵심기술 ]

핵심기술	개요
대용량 및 실시간 데이터 분석 플랫폼 기술	• 대용량의 데이터를 사용자가 분석하고, 결과에 대한 정보를 표현하는 기술
데이터 시각화 기술	• 데이터 분석 결과를 시각화하여 인간이 이해할 수 있도록 유의미한 정보를 표현하는 기술
머신러닝 기반 데이터 분석 기술	• 추천, 분류, 군집 등과 같은 머신러닝 기반 빅데이터 분석 기술
실시간 데이터 처리 기술	• 고성능 인메모리를 이용한 실시간 데이터 처리 기술
내외부 데이터 연동 기술	• 문서, 그래프, 키 값, 인 메모리, 검색 등과 같은 데이터 모델을 지원하는 빅데이터 처리를 위한 비관계형 데이터베이스 관리 시스템(NoSQL) 기술

### 다. 중소기업 기술개발 전략

- IoT의 발전이 보편화됨에 따라 수집되는 비정형 데이터의 양은 폭발적으로 증가하여 비정형 데이터 분석의 중요도는 앞으로도 높아질 전망, 따라서 산업적으로 활용할 수 있도록 비정형 빅데이터를 분석하고 가공하는 기술개발을 통해 시장 선점
- 글로벌 선두 그룹들의 실시간 데이터 처리, 배치 데이터 처리 등의 기술을 도입하여 세계 시장에 뒤처지지 않도록 함과 동시에 각 산업 분야에 적합한 다양한 데이터 분석과 시각화 기술 개발로 중소기업의 경쟁력 강화
- 플랫폼이 아닌 빅데이터를 활용한 서비스 개발과 유료 수익 모델 개발을 통한 수익 창출

### 라. 기술개발 로드맵

#### (1) 중기 기술개발 로드맵

[ 빅데이터 분석 및 시각화 플랫폼 분야 중기 기술개발 로드맵 ]

빅데이터 분석 및 시각화 플랫폼	정형/비정형 등 다양한 빅데이터를 실시간 시각적으로 분석하여 사용자가 내용을 이해할 수 있도록 시각적 형태를 극대화			
	2021년	2022년	2023년	최종 목표
대용량 및 실시간 데이터 분석 플랫폼 기술				빅데이터를 실시간으로 다양한 분석 가능
데이터 시각화 기술				정형/비정형 등 다양한 데이터 분석결과 시각화 주요 정보별 최적의 시각화
머신러닝 기반 데이터 분석기술				다양한 데이터에 대한 분석 알고리즘 개발 데이터 분석 정확도 향상
실시간 데이터 처리 기술				내/외부 빅데이터에 대한 실시간 통합 및 경제처리 자동화
내외부 데이터 연동 기술				정형/비정형, 내/외부 데이터 연동 및 통합 데이터 분석 정확도 향상

(2) 기술개발 목표

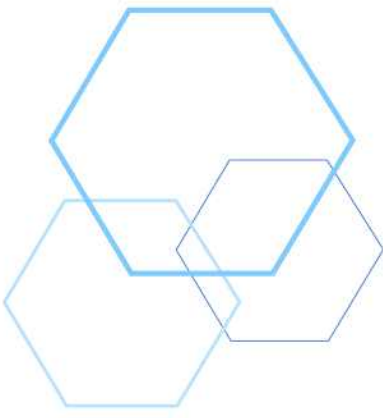
□ 최종 중소기업 기술로드맵은 기술/시장 니즈, 연차별 개발계획, 최종목표 등을 제시함으로써 중소기업의 기술개발 방향성을 제시

[ 빅데이터 분석 및 시각화 플랫폼 분야 핵심요소기술 연구목표 ]

핵심기술	기술요구사항	연차별 개발목표			최종목표	연계R&D 유형
		1차년도	2차년도	3차년도		
대용량 및 실시간 데이터 분석 플랫폼 기술	분석플랫폼 구축 기술	대용량 데이터 구축 (정형/비정형)	분석모듈 및 시각화 모듈 개발	분석플랫폼 고도화	빅데이터를 실시간으로 다양한 분석 가능	창업성장 상용화
데이터 시각화 기술	시각화 기술	데이터 시각화 제공방법 개발	시각화 방법 다양화(질적/양적 데이터 표현)	시각화 고도화	정형/비정형 등 다양한 데이터 분석결과 시각화	창업성장 상용화 산학연
	데이터정확도 기반의 시각화 기술 표준화 등	분류 데이터 유형별 최적의 시각화 기획 및 설계	시각화 개발 (단일시점 기준 외에 기간별 트렌드 파악 가능)	시각화 결과 개별 정보의 의미를 스크립트화 사용자 관점의 test 및 개선점 보완	주요 정보별 최적의 시각화	산학연
머신러닝 기반 데이터 분석기술	머신러닝 알고리즘	비정형/정형 데이터 분류 알고리즘 개발	데이터 특성에 따른 추가 분류모델 및 군집 모델 개발	머신러닝 모델 성능 향상 고도화	다양한 데이터에 대한 분석 알고리즘 개발	산학연
	모델링 기술					
	ML기반 데이터 분석 정확도 (F1 Score)	60	70	80	데이터 분석 정확도 향상	
실시간 데이터 처리 기술	실시간 데이터 처리 성능향상	내/외부 데이터의 실시간 연동 및 통합 처리 기술개발	데이터의 정제, 가공, 분석에 대한 실시간 처리 알고리즘 개발	실시간 데이터 처리 성능(속도) 고도화	내/외부 빅데이터에 대한 실시간 통합 및 정제처리 자동화	상용화
내외부 데이터 연동 기술	빅데이터 내외부 연동 기술	내/외부 및 이종간 데이터 연동을 위한 데이터 표준화	내/외부 및 이종간 데이터 연동 자동화	내/외부 및 이종간 데이터 연동 성능 고도화	정형/비정형, 내/외부 데이터 연동 및 통합	기술혁신 상용화
	데이터 연관분석 정확도 (F1 Score)	60	70	80	데이터 분석 정확도 향상	산학연







전략제품 현황분석

# 빅데이터 기반 마케팅 인텔리전스 플랫폼





# 빅데이터 기반 마케팅 인텔리전스 플랫폼

## 정의 및 범위

- 빅데이터 기반 마케팅 인텔리전스 플랫폼이란 온라인과 오프라인의 텍스트 빅데이터를 활용하여 마케팅 이슈를 실시간으로 모니터링하고 추이를 파악하고 나아가 시계열 예측모형과 머신러닝을 통해 예측 정보를 제공하는 빅데이터 분석기술이 종합적으로 구현된 플랫폼

## 전략 제품 관련 동향

시장 현황 및 전망	제품 산업 특징
<ul style="list-style-type: none"> <li>(세계) 2018년 1,852억 달러에서 연평균 8.9%로 성장하여 2024년 3,096억 달러 규모로 성장 전망</li> <li>(국내) 2018년 2,725억 원 규모에서 연평균 8.7%로 성장하여 2024년 4,495억 원 규모로 성장 전망</li> </ul>	<ul style="list-style-type: none"> <li>빅데이터 단위요소기술 개발의 발전과 플랫폼 수요증가에 따라 단위기술과 융복합한 플랫폼 요구 증가</li> <li>빅데이터의 확보가 요소기술만큼 중요한데도 데이터제공 규제에 따른 거버넌스와 독과점 데이터의 거래 미흡</li> </ul>
정책 동향	기술 동향
<ul style="list-style-type: none"> <li>플랫폼 사업을 좌우할 데이터 개방정책 및 법안 개정 활동이 활발히 진행 중</li> <li>빅데이터 융복합 전략 기술로 인공지능, 데이터 전문가 등 기업과 연계한 실무 전문인력 양성 정책</li> </ul>	<ul style="list-style-type: none"> <li>빅데이터 요소단위 기술들의 융복합 분석을 통한 직관적이고 통찰력 있는 정보제공을 위한 플랫폼 역할의 증대</li> <li>빅데이터 분석을 위한 통계, 시계열 예측모형 및 인공지능 등 머신러닝 기법의 활용 요구</li> </ul>
핵심 플레이어*	핵심기술
<ul style="list-style-type: none"> <li>(해외) Amazon, IBM, Anexinet, REI</li> <li>(국내) SK텔레콤, 다음소프트, 데이터솔루션</li> </ul>	<ul style="list-style-type: none"> <li>마케팅 인텔리전스</li> <li>감성 및 사용자 사전 분류기술</li> <li>텍스트 데이터 분석기술</li> <li>자연어 처리 기술</li> <li>개인화 마케팅</li> </ul>

\*생태계 취약 전략제품

## 중소기업 기술개발 전략

- 빅데이터 단위기술에 대한 융복합 기술이 공백 상태인 만큼 중소벤처기업의 기술 선점 가능
- 데이터 개방정책에 기반한 빅데이터 확보 가능성이 커짐에 따라 대기업 독점의 데이터 시장에서 플랫폼 경쟁력 제고
- 빅데이터 확보의 보편성 추세는 대기업 위주에서 중소벤처 기업으로 빠르게 시장 구조가 이동 중이며, 이에 따른 플랫폼 사용의 경로 의존성 확보에 따른 시장 선점 효과 기대

## 생태계 강화방안

- 마켓 인텔리전스가 성과로 연결되기 위한 전제조건은 올바른 데이터의 확보이나 빅데이터 플랫폼 중소기업에서는 좋은 데이터 확보가 어려움이 있어 이 부분에 대한 지원 필요
- 다양한 사회 현안 이슈 해결의 핵심 기술로 빅데이터 플랫폼 기술을 활용하여 정부 주도의 빅데이터 플래그십 투자 확대 추진 필요

# 1. 개요

## 가. 정의 및 필요성

### (1) 정의

- 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
- 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
  - 이 세 가지 기술이 유기적으로 결합 되었을 때 플랫폼의 사용자 성능이 필요에 의해 시너지 효과를 달성 가능

## (2) 필요성

- 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

## 나. 범위 및 분류

### (1) 가치사슬

- 전 산업 분야에 걸쳐서 영업 및 마케팅, 정책 PR 등에서 활용가치가 증가할 것으로 예상되며, 기존의 마케팅 과정 전반에 걸친 업무의 시간과 비용을 획기적으로 줄일 것으로 보여 부가가치 창출에도 기여할 것으로 전망
  - 기존의 마케터가 수행하여온 소모적인 모니터링과 분석업무에서 자유로워지고 이에 따라 분석결과와 통찰을 통해 기업의 매출 상승에 기여할 수 있는 아이디어 도출업무에 집중 가능
  - 제품 및 서비스의 위기관리 대응이 빨라지고 이에 따른 고객들의 서비스 만족도 상승은 기업의 매출 증가로 이어질 것으로 기대

#### [ 마케팅 인텔리전스 플랫폼 분야 ]

후방산업	마케팅 인텔리전스 플랫폼 분야	전방산업
텍스트 분석, 융합분석, 머신러닝, 통계 분석, 지수화 알고리즘	온라인포털, 뉴스, 소셜미디어 등 텍스트 데이터 기반의 모니터링 및 감성/평판 분석, 잠재고객 발굴, 통계자료 등 2차 자료와 비교, 비정형 지수화를 통한 정형자료와의 교차분석 및 상관성/연관성 분석과 예측서비스 이슈 분석, 제품 및 기업 평판, 수요 및 고객 요구 예측 등에 대한 모니터링 서비스	패션, 의료, 자동차, 금융, 교육, 농식품, 공공정책 등 활용 분야가 다양함

## (2) 용도별 분류

- 마케팅 인텔리전스 플랫폼은 제품 및 서비스의 모니터링과 기업위험(평판) 등 민간부문을 비롯해 정책 순응 및 불응, 정책개발 등 공공부문까지 확장할 수 있으며, 전 산업 분야와 정책과정 전반에 활용 가능하여 사실상 분야별 부문별 제한이 없는 것으로 분석
- 빅데이터의 성장, 인프라기술의 발전, 분석방법론의 다양화로 민간과 공공부문 모두 활용가치가 증가하고 있으며 이에 대한 요구도 빠르게 대응

### [ 용도별 분류 ]

공급망 관점	세부기술
패션	<ul style="list-style-type: none"> <li>• 온라인(블로그, 포털, SNS 등)데이터를 통해 패션 제품에 대한 패턴, 색상, 디자인, 기장, 아이템 등 모니터링을 통해 현재 유행 이슈를 파악하고, 경쟁제품의 패션 특성과 특징을 조사 없이 확인</li> <li>• 판매 형태, 재고 수준, 고객 성향 등을 파악하고 예측을 통해 신제품 및 계절 아이템의 선제 대응과 시장선점 등에 도움 제공</li> </ul>
농업	<ul style="list-style-type: none"> <li>• 온라인에서의 농업 관련 키워드를 모니터링 하여 농식품 수급 등 유통정책에 대응할 수 있으며, 질병 키워드를 파악하고 예측하여 위험요인에 대응 가능</li> <li>• 센싱디바이스 등 내부데이터와 기상데이터의 결합분석을 통해 농작물 피해를 최소화하는 대응방안을 제시하고, 적절한 작물관리 및 재배 시기를 조절 가능</li> </ul>
방위산업	<ul style="list-style-type: none"> <li>• 내부데이터를 사용하여 방산 물자의 품질진단과 위험판단, 재고량 조절에 대한 의사결정을 지원함으로써 방산 물자의 효율적인 관리가 가능</li> </ul>
지역경제	<ul style="list-style-type: none"> <li>• 지역 기반의 온라인데이터, 2차 자료와의 결합적 분석결과를 제공하여, 특정 지역의 경제 상황, 관광, 유통, 지역 화폐의 사용처를 한눈에 직관적으로 추적할 수 있도록 도움 제공</li> <li>• 지역경제 지표와 연관성이 높은 텍스트 데이터를 분석하여 지역경제 상황을 모니터링, 예측하고 대응할 수 있는 아이디어를 제공</li> </ul>
교육	<ul style="list-style-type: none"> <li>• 교육에 대한 정책적 이슈와 사회적 요구를 파악하고, 교육정책에 대한 공공성과 시민들의 평가를 모니터링, 시기별, 교육과정별 패턴 등을 파악하여 교육정책 수립 및 개선에 도움</li> </ul>
환경	<ul style="list-style-type: none"> <li>• 미세먼지 등 대기 환경에 대해 시민들이 생성한 온라인데이터를 기반으로 정책에 필요한 요구사항을 파악하여 정책과정에 반영할 수 있으며, 이를 통해 대기 환경정책에 대한 시민들의 순응을 유도</li> <li>• 다양한 내부환경 정보를 플랫폼화하여 에너지 사용량과 오염물질 배출 상황을 한눈에 모니터링하여 기업별 업종별 규제정책에 활용 가능</li> </ul>

## 2. 산업 및 시장 분석

### 가. 산업 분석

#### ◎ 산업의 특징

- 인텔리전스 플랫폼의 사용처는 영업활동, 마케팅, 고객센터, 금융, 제품개발 등 경영활동 전반에 걸쳐 활용가치가 증가
  - 특히, 시장 상황을 모니터링하고 이슈에 민감한 영업활동 분야가 가장 크게 활용할 것으로 보이며, 다음으로 판매 및 마케팅의 활용도가 증대
  - 이들 분야는 기업 활동에 있어서 전반에 위치한 분야로 모니터링 및 시장 상황 예측 등에 민감하여 마케팅 인텔리전스 플랫폼에 적합
- 마케팅 인텔리전스 플랫폼은 전 세계적으로 성장을 요구하도록 시장의 압력을 받고 있는 상황이며, 의사결정 지원에 대한 기업들의 투자 요구사항도 증가
  - 마케팅 인텔리전스 플랫폼은 기업운영의 통찰력과 효율성 향상이라는 측면에서 크게 인정받고 있으며 모바일기술 및 소셜컴퓨팅과의 연계도 채택
  - 특히, 직관적인 통찰을 제공해주기 때문에 실시간성과 시각화에 관한 관심이 크게 부각
- 인텔리전스 플랫폼은 비즈니스 인텔리전스를 중심으로 글로벌 기업들이 소프트웨어 개발 유도
  - 해외에서는 Good Data, IBM, Information Builders, Microsoft, Oracle, SAS, TIBCO Software 등이 주도해왔으나 국내에서는 중소기업에서도 인텔리전스 플랫폼 개발을 시도
- 협소한 의미의 마케팅에서부터 공공분야의 정책PR에 이르기까지 광범위한 마케팅 인텔리전스 플랫폼이 적용될 것으로 전망
  - 온라인포털이나 소셜미디어 등 쉽게 획득할 수 있는 빅데이터뿐만 아니라 웨어러블, 센싱디바이스, 제3자 데이터(통신내역, 신용카드 사용정보) 등 비식별 개방 데이터의 수요와 요구가 증가하고 있고, 정책적으로 규제 완화의 움직임이 있어 향후 데이터 개방이 마케팅 인텔리전스 플랫폼 시장의 급격한 성장을 이끌 것으로 기대
- 전 분야에 걸친 마케팅 인텔리전스 플랫폼의 효용성은 공공데이터 및 내부데이터의 빅데이터 개방 요구를 촉진해 빅데이터 분석에서 중요한 양질의 정형/비정형 데이터 확보를 통해 분석력과 예측력을 향상할 수 있을 것으로 예상
  - 마케팅 인텔리전스 플랫폼에서는 정량적 지표와 자료 비정형 자료를 융합 분석할 수 있도록 비정형 지수를 산출하는 것이 중요한 기술 중 하나로 이에 대한 지수산출방법이 중요해질 것으로 예상



## ◎ 산업의 구조

- 전방산업은 패션, 의료, 자동차, 농식품, 방위산업, 교육, 공공정책 등 독립적인 산업 분야에서 성장하고 있으나 향후 이종 간 산업 결합 등 활용 분야가 다양하게 적용될 것으로 기대
  - 패션 부문은 온라인포털과 소셜미디어의 텍스트 데이터를 구매량과 함께 분석하여 특정 패션의 아이템, 색상과 패턴 등 패션 특징에 대한 이슈를 파악하고 향후 패션 트렌드에 대하여 예측
  - 의료 부문에서는 계절 및 환경변화 미용의 트렌드 추적 등을 온라인 텍스트 데이터와 건강보험 등 의료 데이터를 융합적으로 활용하여 치료와 미용 목적의 수요와 요구 등을 파악할 뿐만 아니라 잠재고객을 발굴할 수 있을 것으로 기대
  - 자동차 부문에서는 고객의 불만, 디자인과 실용성 요구사항 등을 텍스트 데이터로부터 파악하여 자동차의 부분별 성능개선이나 보완, 신차개발 등 의사결정에 반영할 정보를 제공
  - 농식품 부문에서는 기상 및 환경데이터와 온라인 및 소셜미디어의 텍스트 데이터를 기반으로 농식품 판매량을 예측함과 동시에 농식품 재료의 트렌드를 추적하여 신제품 출시 시점과 개발기획에 아이디어를 제공
  - 공공정책 부문에서는 집행 중인 정책에 대한 홍보와 피드백을 온라인과 소셜미디어의 텍스트 데이터를 활용해 정책과정 전반에 이르는 정책 순응과 불응을 모니터링할 수 있도록 지원하며, 향후 정책개발과 보완 제도 정비 등에 활용
  - 마케팅 부문에서는 필요로 하는 빅데이터 분석 활용을 위해 개발되는 본 시스템은 온/오프라인에서 생산되는 막대한 정보 중 시장 및 산업 동향에 맞는 경쟁전략을 다양한 지표를 통해 분석 가능
  - 기존의 마케팅 분석은 텍스트 기반의 키워드 분석에 집중되어 있지만, IoT 응용 제품·서비스의 급속한 확산으로 발생하는 빅데이터에 대한 과학적/효과적 분석이 어려운 점에 대해 해결방안이 될 수 있음
- 기존의 빅데이터 분석기술들은 영상, 이미지, 텍스트, 음성, 정량자료 등 단편적인 분석만을 제공해 왔다면 마케팅 인텔리전스 플랫폼에서는 데이터의 다양성을 결합하여 직관적인 시각화를 보여주는 것이 특징
  - 또한, 부가적으로 기업의 필요사항에 맞게 선택하여 활용할 수 있는 다양한 분석 도구를 API 형태로 제공하여 손쉬운 이해와 폭넓은 활용을 지원하는 마케팅 분석 지원이 가능해질 것으로 예상함
  - 패션을 예측하는 경우 패션 아이템 중 원피스의 온라인 검색량과 언급량이 이슈가 되고 있다는 분석에서 더 나아가 기상데이터에 따른 계절요인과 여행통계, 구매량을 비교분석해 원피스 종류에 따른 유행 시점, 동시구매 패턴에 따른 추가 아이템 선별 등 예측력과 활용도를 증대
- 마케팅 인텔리전스 플랫폼은 기존에 정착된 후방산업과 전방산업의 상호관계가 뚜렷할 뿐만 아니라 상호의존성도 높아 데이터와 분석의 융합, 적절한 시각화와 결과에 대한 의사결정 해석의 신뢰성과 타당성이 전후방 산업을 동시에 발전
  - 후방산업의 개별적 방법론과 분석기술들은 전방산업이 가지고 있는 빅데이터의 개방과 관련이 있으며, 개방된 데이터를 기반으로 분석결과의 검증이 이루어지기 때문에 공진화(coevolution)할 수밖에 없어 기술개발 못지않게 생태계를 조성하는 것도 중요

## 나. 시장 분석

### (1) 세계시장

- Globe Newswire은 2020년 발표한 비즈니스 인텔리전스 시장규모에서 2018년 비즈니스 인텔리전스 세계 시장규모를 1,852억 달러로 평가하였고 연평균 8.9%로 성장하여 2024년 3,096억 달러 규모로 성장할 것으로 전망

[ 비즈니스 인텔리전스 세계 시장규모 및 전망 ]

(단위 : 십억 달러, %)

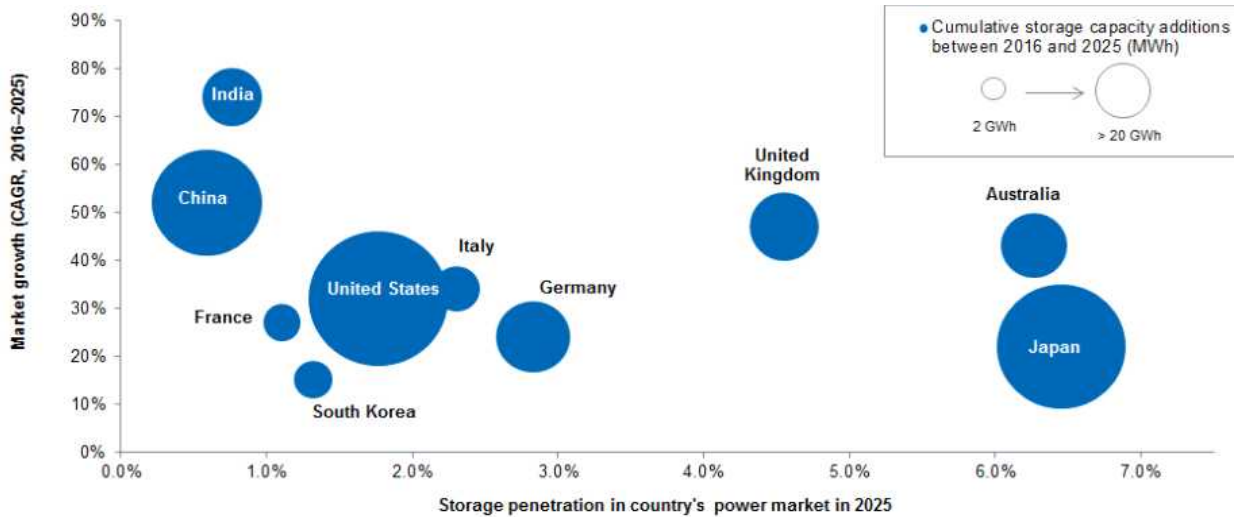
구분	'18	'19	'20	'21	'22	'23	'24	CAGR
세계시장	185.2	201.7	231.0	248.6	267.5	287.8	309.6	8.9%

\* 출처: Globe Newswire (2020) 기반으로 네모아이씨지 재추정

- 비즈니스 등 마케팅 인텔리전스 플랫폼은 사용자에게 쉬운 데이터 탐색과 통찰력을 얻기 위한 분석 도구들이 활용되는 만큼 개별 요소기술의 성장도 중요

[ 비즈니스 인텔리전스 지역별 시장 전망 ]

(단위: %)



\* 출처: Markets and Markets, 2018

- 마케팅 인텔리전스 플랫폼은 기존에 정착된 빅데이터 분석기술 등의 요소기술을 활용함에 따라 자연어 처리, 머신러닝 기술 등의 시장규모는 매년 증가 추세
  - 머신러닝 관련 시장은 2021년 576억 달러로 지난 2017년 120억 달러에서 증가할 것으로 전망하고 있으며 벤처 캐피탈 투자, 사모펀드(PE) 자금조달, 합병·인수의 지적재산권(IP) 및 특허 분야에서도 우위를 차지
  - 자연어 생성 시장규모는 2023년까지 8억 2530만 달러 규모로 성장할 것으로 예상
  - 분석소프트웨어 시장에서도 2019년 21.54억 달러에서 2024년 32.39억 달러로 성장이 예상

[ Total BI & Analytics SW Market ]

(단위: 백만 달러, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
Data Science Platforms	339	383	433	489	553	624	706	13.0
BI & Analytics	1,986	2,154	2,337	2,536	2,751	2,985	3,239	8.5

\* 출처: 코트라, 해외시장뉴스, 2018 기반으로 네모아이씨지 재추정/재가공

(2) 국내시장

- 빅데이터 기반 마케팅 인텔리전스 플랫폼 국내 시장규모는 2018년에 2,725억 원 규모로 평가되었으며, 연평균 8.7%로 성장하여 2024년까지 4,495억 원에 이를 것으로 전망됨

[ 국내 빅데이터 기반 마케팅 인텔리전스 플랫폼 시장규모 전망 ]

(단위: 억 원, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
국내시장	2,725	2,962	3,220	3,500	3,804	4,135	4,495	8.7

\* 출처: Big Data Market Size - MarketsandmarketsSupply, Globe Newswire (2020)을 바탕으로 세계 빅데이터 시장과 국내 빅데이터 시장을 복합적으로 고려하여 네모아이씨지 재추정

- 마케팅 인텔리전스 플랫폼의 핵심 분석기술 중 하나인 머신러닝의 경우 국내시장 규모는 2018년 670억 원에서 연평균 44.1%로 성장해 2024년 6,027억 원으로 증가할 것으로 예상
  - 머신러닝의 성장률이 급격한 상승을 보이는 이유는 머신러닝을 통해 통찰력을 쉽게 얻을 수 있는 기술개발의 응용력이 높기 때문

[ 국내 머신러닝 시장 ]

(단위: 억 원, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
국내시장	670	970	1,398	2,014	2,902	4,182	6,027	44.1

\* 출처: KISTI 마켓리포트, 2017 기반으로 네모아이씨지 재추정/재가공

### 3. 기술 개발 동향

- 기술경쟁력
  - 공공 빅데이터 수집 및 분석 시스템은 미국이 최고기술국으로 평가되었으며, 우리나라는 최고기술국 대비 84%의 기술 수준을 보유하고 있으며, 최고기술국과의 기술격차는 1.4년으로 분석
  - 중소기업의 기술경쟁력은 최고기술국 대비 74.1%, 기술격차는 2.3년으로 평가
  - EU(92.5%)>한국>일본(79.5%)>중국(67.3%)의 순으로 평가
- 기술수명주기(TCT)<sup>19)</sup>
  - 공공 빅데이터 수집 및 분석 시스템은 6.13의 기술수명주기를 지닌 것으로 파악

#### 가. 기술개발 이슈

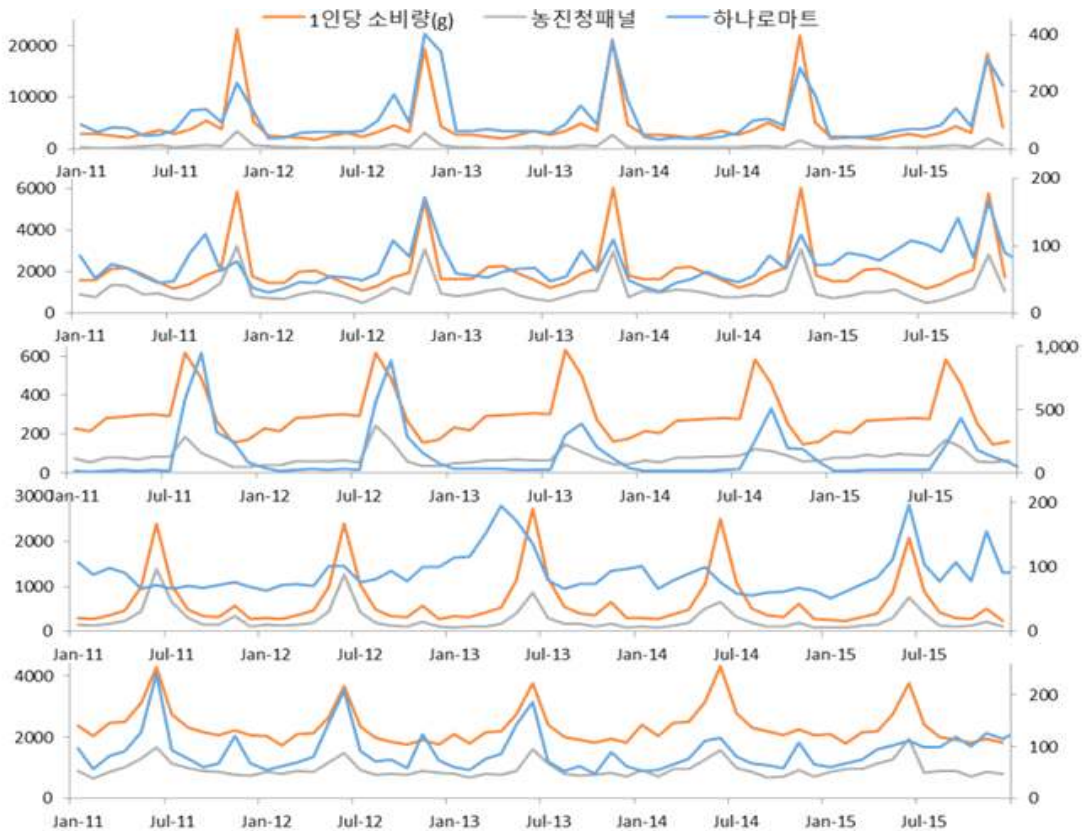
- 인공지능, 머신러닝, 텍스트 마이닝, 통계 분석 등 정착된 분석 기법과 공공부문과 민간부문에 존재하는 내/외부 빅데이터를 융합해 분석하는 것이 주요 역할인 마케팅 인텔리전스 플랫폼은 비정형 데이터의 사용자 사전 및 감성사전 자동화와 비정형 지수 개발이 이슈
  - 텍스트 기반의 내/외부 빅데이터를 수집하여 분석 목적과 특성에 맞게 정제 가공처리 한 후 사용자사전과 감성사전을 도출하고, 이를 통계 및 제3자 데이터와 통합하기 위한 지수화 과정을 거쳐 최종적인 분석결과를 도출하기 때문에 자연어 처리와 사전분류모델 기술, 사용자 데이터 특성에 맞는 크롤링 개발 등의 이슈가 존재
- 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
- 전통적인 제조업이나 서비스업 등 업종별, 산업별 데이터 통합을 촉진해 이종 간 데이터 융복합 분석 시도가 증가할 것이며 이를 통해 기존에는 없었던 새로운 제품 개발이나 서비스를 창출할 것으로 기대
- 빅데이터 분석에 필요한 단위요소기술 개발과 함께 데이터 확보의 요구와 수요가 증가할 것으로 예상됨에 따라 양질의 공공데이터 개방을 촉진할 것으로 기대
- 기술적으로는 마케팅 기획, 제작, 실행, 피드백에 이르는 전 과정에 걸쳐 인텔리전스 플랫폼의 역할은 광범위할 것이며, 기업 내부의 데이터와 융합될 경우 그 시너지 효과와 효용은 극대화될 것으로 예상

19) 기술수명주기(TCT, Technical Cycle Time): 특허 출원연도와 인용한 특허들의 출원연도 차이의 중앙값을 통해 기술 변화속도 및 기술의 경제적 수명을 예측

□ 지금까지 빅데이터 분석은 단위요소별로 독립적으로 수행되어 왔으나 데이터의 형태가 다양해지고 분석기술 및 방법론의 발달로 단위요소별로 이루어지던 빅데이터 분석이 융합되어 활용되기 시작

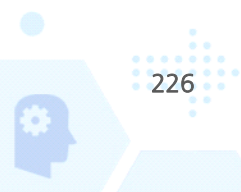
- 복수의 데이터를 통해 통찰력 있는 분석결과를 제공
- 플랫폼 서비스의 발달과 시장의 요구에 따라 다각적인 분석을 시각화하여 직관적으로 화면에 노출시킴으로써 사용자의 데이터 활용과 분석을 용이하게 할 수 있도록 텍스트 데이터를 정형화 시키려는 노력 지속
- 통계자료나 조사자료 등 정형데이터와 텍스트, 영상, 음성, 이미지 등 비정형 데이터의 복수 사용과 융합, 응용분석기술들을 플랫폼에서 구현할 수 있도록 발전
- 단위요소 기술 간, 이종 간 데이터의 결합을 가능하게 하기 위해서 정형 데이터와 비정형 데이터의 결합이 중요해졌고, 이에 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

- 고객 만족, 제품 및 서비스의 수요와 요구사항, 구매 및 사용자 선호패턴의 예측 등을 분석하여 고객의 주요 요구사항을 기획에 반영하거나, 기존 제품의 보완과 개선, 재고량 관리(선입선출 등), 홍보 마케팅 시점 등 분석과 대응의 영역으로 구분되며, 이에 대한 통찰을 제공하는 추세
- 온라인 포털과 소셜미디어의 비정형 텍스트 데이터 및 내부 데이터 등 개방형 데이터와 폐쇄형 데이터의 자동화 크롤러 등 수집기술의 요구가 높아지고 있음
  - 쉽게 획득할 수 있는 온라인 데이터에 비해 마케팅 데이터는 영업비밀 등 보안이나 민감한 측면 때문에 기업 내부에 보관되어 있으며 이를 자동화하여 수집하고 전처리에 대한 기술적 요구가 증가
- 개인정보보호 등 비식별 처리를 활용한 빅데이터 활용이 정책적으로 가능성을 인정받고 있으며 이에 대한 비식별 처리 기술 개발이 필요
- 비정형 텍스트 데이터 기반의 비정형 지수산출 기술 등 지수화의 반자동화 과정과 빅데이터 분석기술을 통합하고 예측 가능성을 높이는 인공지능 기술이 발전
- 기존의 비즈니스 인텔리전스 플랫폼에서는 정형 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





## 나. 생태계 기술 동향

### (1) 해외 플레이어 동향

- (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

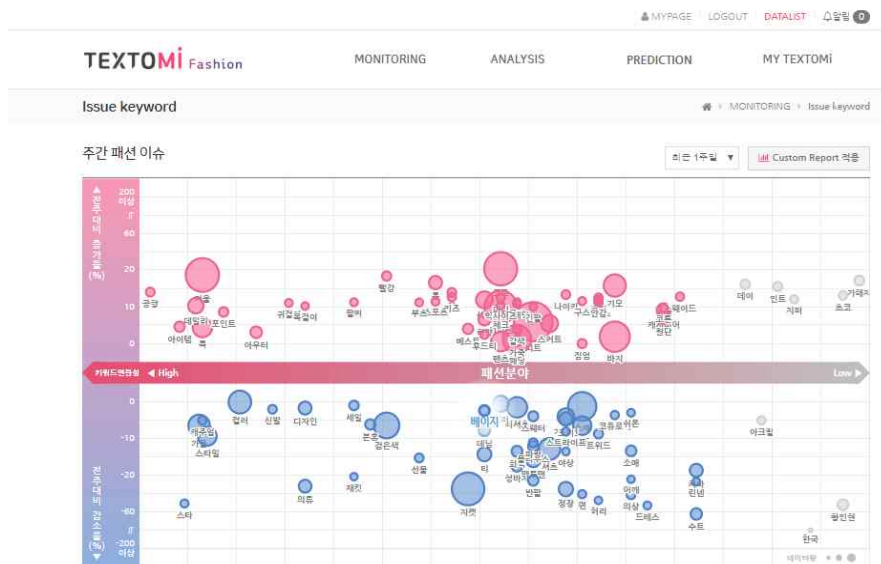
- (Anexinet) ListenLogic Omni-Channel Analytics Platform을 개발하여 고객경험에 대한 심층적인 이해와 행동 패턴을 분석하여 마케팅에 활용할 수 있도록 함
  - 기업이 운영하는 콜센터기록, 이메일, 채팅, 텍스트 및 CRM정보와 같은 다양한 데이터소스를 통해 고객의 패턴을 분석할 수 있도록 함
  - 비정형 데이터로부터 통찰력을 얻을 수 있도록 직관적인 시각화 정보를 세부적으로 제공하며, 기계학습을 사용하여 고객과의 관계를 이해하고 예측할 수 있도록 함으로써 향후 구매 가능성을 예측하는데 기여
  - 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 시를 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 동향

- (더아이엠씨) 온라인 포털 및 뉴스, 소셜미디어의 텍스트 데이터 기반의 패션 인텔리전스 플랫폼인 패션 텍스트미 라는 플랫폼을 개발하고 패션 아이템, 색상, 패턴 등 관련 정보를 모니터링할 수 있으며, 패션 관련 키워드를 5주 단위로 출현 가능성을 예측 가능
  - 패션 텍스트미는 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 ]

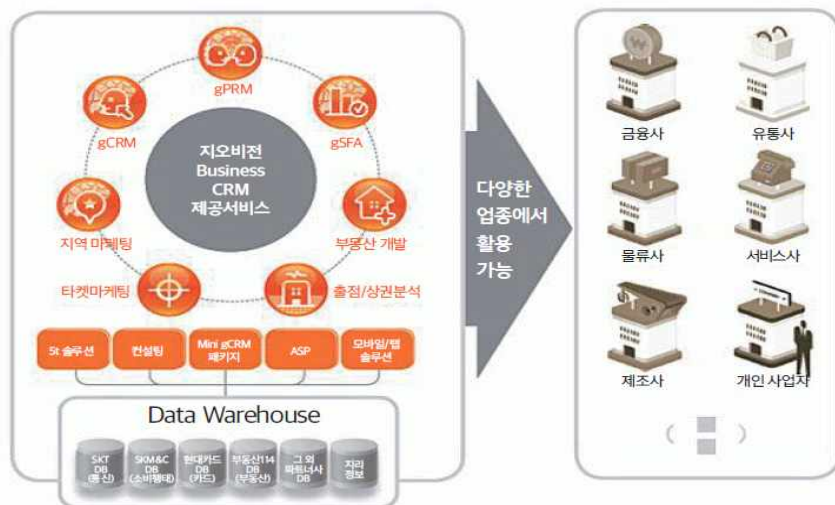


\* 출처: 패션 텍스트미(<http://fashion.textomi.co.kr>), 2019

- (다음소프트) 자체 보유하고 있는 막대한 빅데이터를 활용하여 통신 및 카드데이터와 융복합 교차분석을 시도하여 사회이슈를 분석
  - 비정형 데이터뿐만 아니라 정형 데이터와 결합하여 SI를 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 SI와 결합시켜 내재된 패턴을 찾는데 주력하고 있음. 이를 적용하여 관광동향, 여론, 주식시장 분석 등에 활용

- (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

## 다. 국내 연구개발 기관 및 동향

### (1) 연구개발 기관

[ 마케팅 인텔리전스 분야 주요 연구조직 현황 ]

기관	소속	연구분야
한국전자통신연구원	인공지능연구소 통신미디어연구소 지능화융합연구소 KSB융합연구단	<ul style="list-style-type: none"> <li>• 복합인공지능, AI슈퍼컴퓨터</li> <li>• 5G, Post5G이동통신, 디지털 콘텐츠</li> <li>• 교통/물류/의료/도시 등 지능화 솔루션, 국방ICT융합</li> <li>• 예측/예방/최적화 지능서비스, 지식융합 인공지능 플랫폼</li> </ul>
한국과학기술정보연구원	데이터분석플랫폼센터 계산과학플랫폼센터	<ul style="list-style-type: none"> <li>• 대용량 데이터 센터 구축, 플랫폼 연구개발</li> <li>• 초고성능컴퓨터, 계산과학공학, 과학기술 정보융합</li> <li>• 미래기술 탐색 인텔리전스</li> <li>• 빅데이터 공유 융합, 빅데이터 기술개발</li> </ul>
정보통신산업진흥원	AI융합산업팀 IoT산업팀	<ul style="list-style-type: none"> <li>• 지능정보서비스, 지능정보산업 인프라 조성</li> <li>• 클라우드컴퓨팅산업 육성</li> </ul>

\* 출처: 한국전자통신연구원, 한국과학기술정보연구원, 정보통신산업진흥원 각 홈페이지

### (2) 기관 기술개발 동향

- 한국전자통신연구원(ETRI) 지능화 솔루션 및 지식융합 인공지능 플랫폼 부문 등
  - 시간의 흐름에 따라 변화하는 다양한 정보를 수집, 기록 분석하여 위험 상황에 대응할 수 있는 지식 정보를 제공하는 ‘다중 로그 기반 멀티모달 데이터 융합 분석 및 상황대응 플랫폼 기술개발’(2017~2020, 진행 중)
  - ICT장비에 대한 응용서비스개발, 시험환경홍보 및 서비스개발인력 양성 정보 등 중소기업의 차세대 기술에 대한 개발-테스트-홍보의 원스톱 지원을 목적으로 하는 ‘차세대 네트워크 및 컴퓨팅 플랫폼연구 기반 구축’(2014~2019,종료)
  - 에너지 및 환경 데이터를 수집하여 융합분석을 통해 모니터링과 위험감지 및 탐색 정보를 제공하는 ‘에너지 사용 밀집지역용 융합형 안전관리 플랫폼 개발’(2015~2018, 종료)
- 한국과학기술정보연구원(KISTI) 대용량 데이터 플랫폼 개발, 수요탐색 인텔리전스, 빅데이터 융합 기술개발 등
  - 국가연구데이터를 수집하고 분류하고 가공하여 연구활용지원 정보를 제공하는 ‘국가연구데이터플랫폼 구축 및 운영’(2019~2020, 진행 중)
  - 국가차원에서의 연구데이터 공유 확산 체제를 구축하여 연구데이터 플랫폼 활성화를 위한 융합분석, 활용주제 발굴의 연구데이터 공유 확산체제 구축(2018~2020, 진행 중)
- 정보통신산업진흥원(NIPA) 지능정보서비스, 클라우드컴퓨팅산업육성 등
  - 문제발굴, 정의, 사회적 이슈탐색, 기술개발적용, 문제해결을 위한 융합적 지식정보제공을 위한 ‘지역균형발전 SW ICT융합기술개발’(2019~2023, 진행 중)

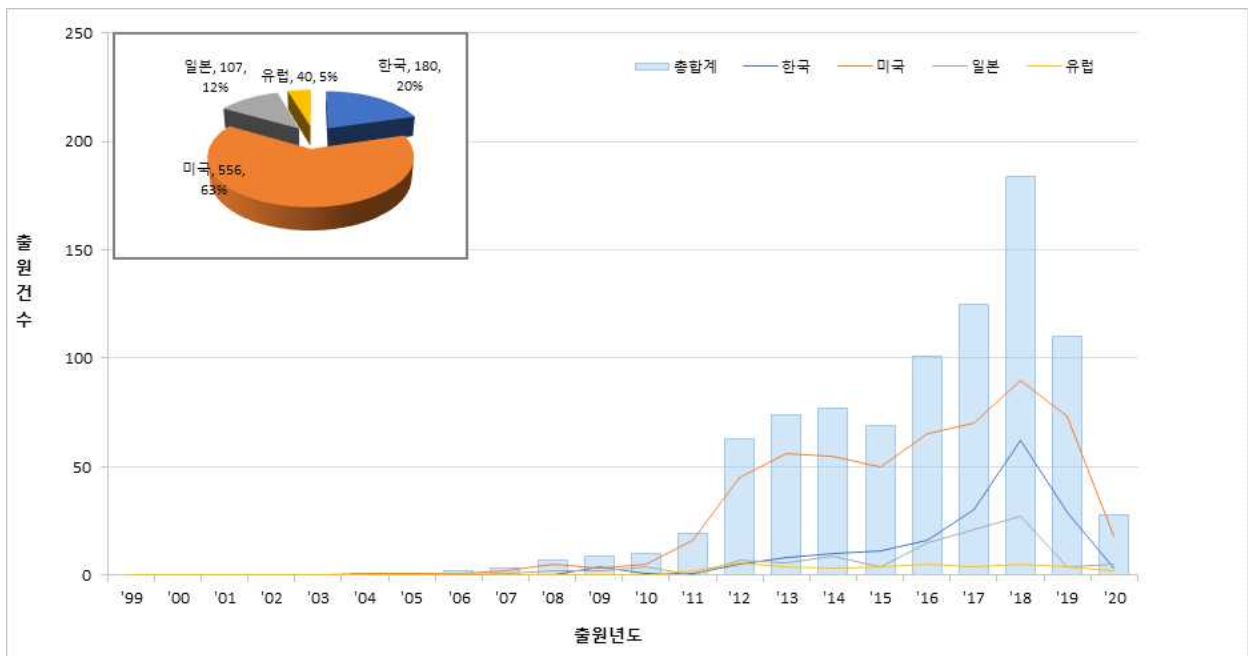
## 4. 특허 동향

### 가. 특허동향 분석

#### (1) 연도별 출원동향

- 빅데이터 기반 마케팅인텔리전스 플랫폼은 '12년부터 급격한 성장을 보임
  - 각 국가별로 살펴보면 미국이 가장 활발한 출원활동을 보이고 있음
- 국가별 출원비중을 살펴보면 미국이 전체의 63%의 출원 비중을 차지하고 있어, 최대 출원국으로 빅데이터 기반 마케팅인텔리전스 플랫폼 분야를 리드하고 있는 것으로 나타났으며, 한국은 20%, 일본은 12%, 유럽은 5% 순으로 나타남

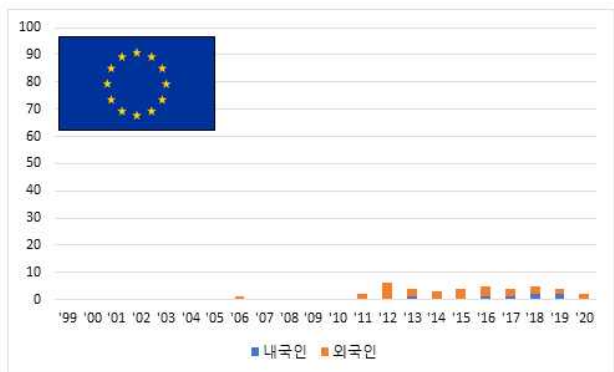
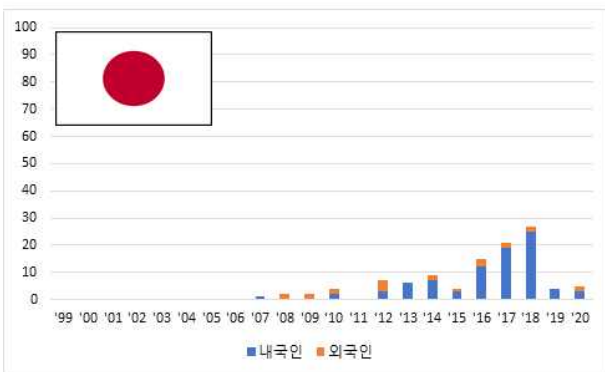
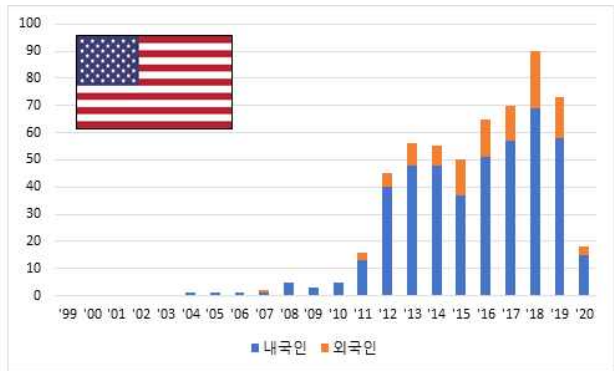
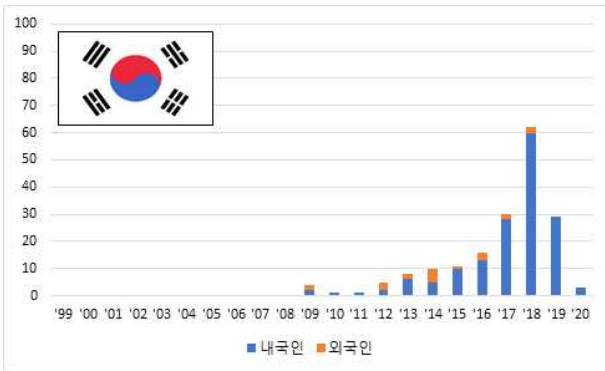
[ 빅데이터 기반 마케팅인텔리전스 플랫폼 연도별 출원동향 ]



## (2) 국가별 출원현황

- 한국의 출원현황을 살펴보면, '16년부터 해당 기술의 출원이 급격히 증가하는 추세
  - 내국인 위주의 출원이 진행되고 있음
  - 한국 기술의 양적 흐름은 미국과 상당히 유사
  - 미국의 출원 수에 비해 32% 정도의 수준을 보임
- 미국의 출원현황을 살펴보면 분석구간 초기부터 전체 특허기술의 출원 증감 흐름에 영향을 주고 있는 것으로 나타남. 미국의 경우, 한국에 비해 외국인의 비중이 큰 것으로 나타남
- 유럽의 출원현황을 살펴보면 출원수가 매년 10건 이하로, 뚜렷한 증감 동향이 나타나지 않음
- 일본의 출원현황을 살펴보면, '18년까지 증가세를 보이며, 한국과 유사한 동향을 보임

[ 국가별 출원현황 ]



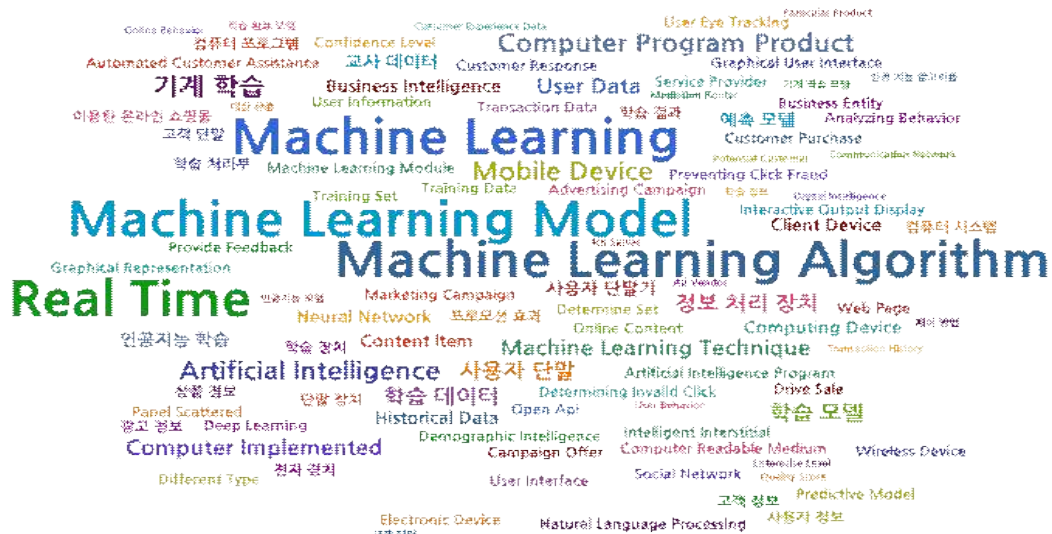
### (3) 기술 집중도 분석

□ 전략제품에 대한 최근 기술 집중도 분석을 위한 구간별 기술 키워드 분석 진행

- 전체 구간(1999년~2020년)에서 Machine Learning, Machine Learning Model, Real Time 등 키워드가 다수 도출
- 최근 구간 분석 결과, 최근 1구간(2012년~2015년)과 비교할 때 2구간(2016년~2020년)에서 나온 키워드가 대부분 유사하며 사용자 단말, Artificial Intelligence, 학습 모델 키워드가 많이 등장하는 것으로 보아 빅데이터 기반 마켓 인텔리전스 플랫폼 분야에도 AI를 이용한 자연어처리나 텍스트 데이터 분석 관련 연구개발이 활발한 것으로 추정

[ 특히 키워드 변화로 본 기술개발 동향 변화 ]

전체구간(1999년~2020년)



- Machine Learning, Machine Learning Model, Real Time, Machine Learning Algorithm, 기계 학습, Computer Program Product, Artificial Intelligence, Mobile Device, 사용자 단말, 학습 모델

최근구간(2012년~2020년)

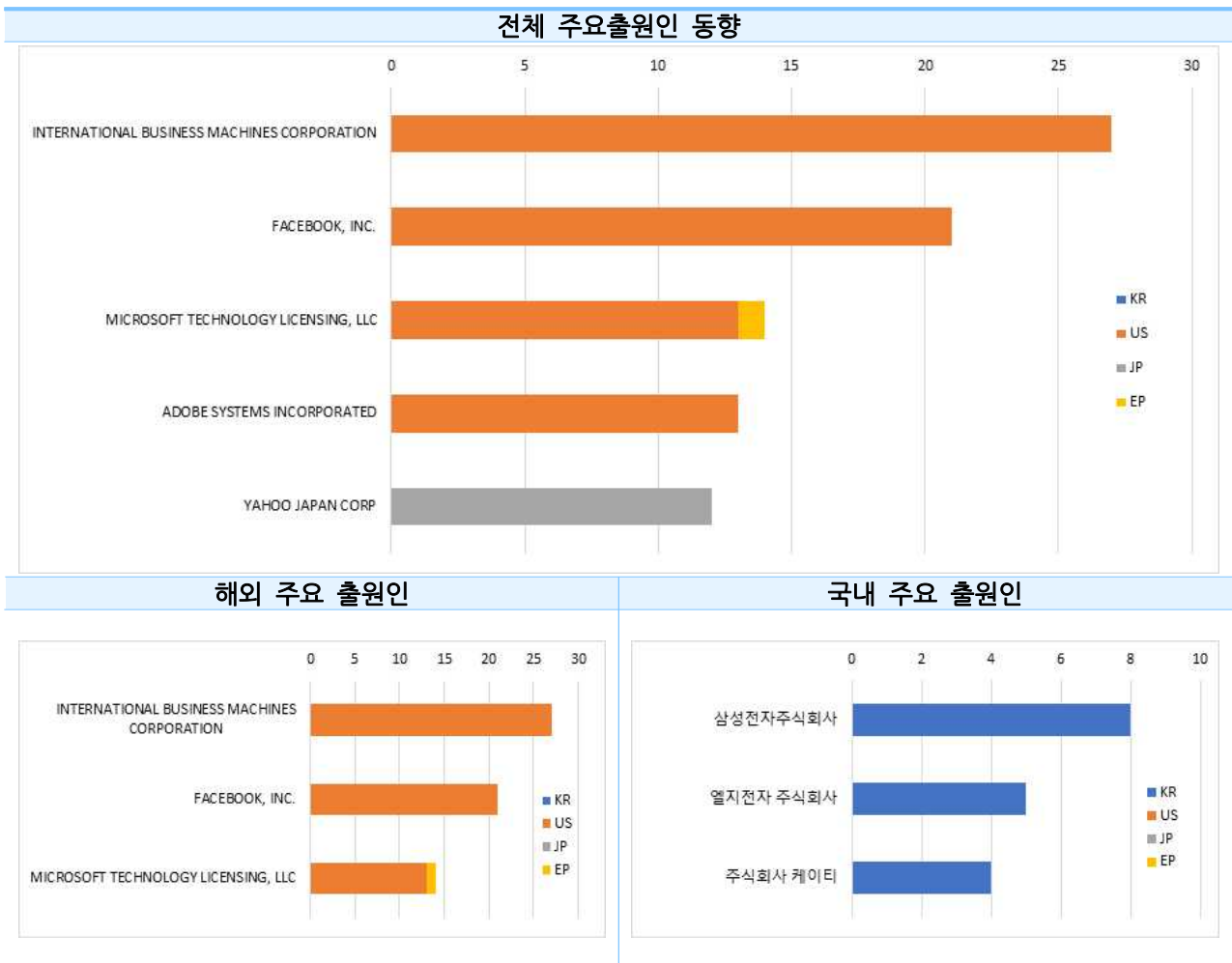
1구간(2012년~2015년)	2구간(2016년~2020년)
<ul style="list-style-type: none"> <li>• Machine Learning, Real Time, User Data, Mobile Device, 기계 학습, Computer Program Product, Computer Implemented, Machine Learning Technique, Preventing Click Fraud, Determining Invalid Click</li> </ul>	<ul style="list-style-type: none"> <li>• Machine Learning, Machine Learning Model, Real Time, Machine Learning Algorithm, 사용자 단말, Artificial Intelligence, 기계 학습, 학습 모델, Computer Program Product, Computing Device</li> </ul>



## 나. 주요 출원인 분석

- 빅데이터 기반 마케팅인텔리전스 플랫폼의 전체 주요출원인을 살펴보면, 주로 미국 국적의 출원인이 다수 포함되어 있는 것으로 나타났으며, 제 1 출원인으로는 미국의 INTERNATIONAL BUSINESS MACHINES CORPORATION인 것으로 나타남
  - 제 1 출원인인 INTERNATIONAL BUSINESS MACHINES CORPORATION의 출원은 미국에 집중된 경향을 보임
  
- 빅데이터 기반 마케팅인텔리전스 플랫폼 관련 기술로 빅데이터 분석을 다루는 대기업에 의한 출원이 대다수를 차지
  - 국내에서도 대기업의 활발한 출원이 이루어짐

[ 빅데이터 기반 마케팅인텔리전스 플랫폼 주요출원인 ]

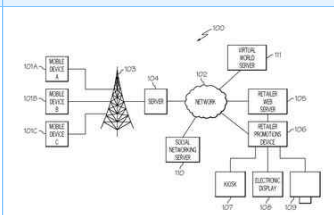


(1) 해외 주요출원인 주요 특허 분석

◎ IBM(International Business Machines)

- IBM은 미국의 다국적 기업으로, 빅데이터 기반 마케팅인텔리전스 플랫폼과 관련하여, 소매점의 매출을 증가시키기 위한 기술에 특화된 특허 27건 출원

[ IBM 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
US9208518 (2014.01.07)	Generating targeted group based offers to increase sales	빅데이터를 지능형 분석에 기초하여, 제품을 구매하기 위한 사용자/고객 세트에 대한 그룹 오퍼(group offer)를 생성하는 기술	


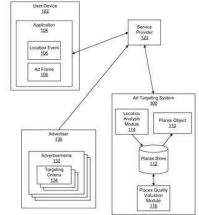
\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출



◎ FACEBOOK, INC.

- FACEBOOK, INC.는 미국 소셜네트워크 기업으로, 사용자의 모바일을 통한 결제 및 광고를 용이하게 하기 위한 특허를 보유하고 있으며, 빅데이터 기반 마케팅인텔리전스 플랫폼 기술과 관련한 특허를 미국에 출원
  - 모바일을 통한 결제 시스템 및 사용자를 타겟팅한 광고를 모바일로 제공하는 것과 관련하여 21건의 미국 출원을 진행하였으며, 그 중 등록된 특허는 2건으로 파악

[ FACEBOOK, INC. 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
US10417702 (2013.09.25)	Dynamically providing a third-party checkout option	전자상거래 애플리케이션과 결제 중계 사이의 중개자의 역할을 하는 전자상거래 결제 퍼실리테이터를 통해 체크아웃 프로세스의 용이성 및 속도를 증가시키는 기술	
US10332150 (2013.03.18)	Location event advertising	사용자의 지리적 위치 등의 정보를 수집하고, 모바일 애플리케이션을 실행하는 사용자를 타겟팅하여 광고를 제공하는 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ MICROSOFT TECHNOLOGY LICENSING, LLC

- MICROSOFT TECHNOLOGY LICENSING, LLC는 Microsoft의 자회사로, 회사의 특허 및 기술 이전 관련 기업으로 머신러닝 알고리즘을 활용한 광고 선택 기술 관련 특허 14건을 출원하였으며, 등록 특허는 4건 보유
  - 빅데이터 기반 마케팅인텔리전스 플랫폼에서 머신러닝을 사용한 광고의 품질을 평가하고 서비스하는 기술의 응용 특허를 다수 보유

[ MICROSOFT TECHNOLOGY LICENSING, LLC 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
US10546232 (2017.10.02)	Image recognition with promotion of underrepresented classes	사람의 얼굴을 인식하기 위한 기술로, 몇 개의 트레이닝 샘플로 이미지(얼굴)를 인식하는 기술	
US9589277 (2013.12.31)	Search service advertisement selection	머신 러닝 알고리즘의 관련성 메트릭을 최대화함으로써 스코어링을 위한 선형 모델을 구성하고, 기계 학습 스코어러는 WAND 알고리즘 기반 광고 선택하는 기술	
US9177333 (2010.06.17)	Ad copy quality detection and scoring	머신 러닝 알고리즘을 사용하여, 개별 광고에 대한 품질 점수(quality score)를 생성하는 기술	
US10235679 (2010.04.22)	Learning a ranker to rank entities with automatically derived domain-specific preferences	검색엔진의 도메인에 속하는 개체를 등급을 매기기 위해 구성될 수 있는 영역 특정의 랭커를 배우는 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

(2) 국내 주요출원인 주요 특허 분석

◎ 삼성전자 주식회사

- 삼성전자 주식회사는 빅데이터 기반 마케팅인텔리전스 플랫폼 기술에 있어서, 콘텐츠에 대한 사용자의 선호도를 예측하기 위한 기술과 관련된 특허를 다수 출원

[ 삼성전자 주식회사 주요특허 리스트 ]

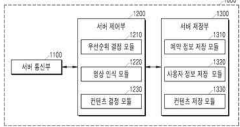
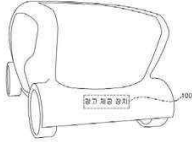
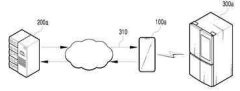
등록번호 (출원일)	명칭	기술적용분야	도면
KR1792587 (2009.01.23)	콘텐츠에 대한 선호도 예측 방법 및 장치와, 샘플 콘텐츠 선정 방법 및 장치	기계 학습 알고리즘의 입력으로 하여 사용자의 콘텐츠에 대한 선호도를 예측하는 콘텐츠 선호도 예측하는 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ 엘지전자 주식회사

□ 엘지전자 주식회사는 한국의 전자제품 전문 기업으로, 빅데이터 기반 마케팅인텔리전스 플랫폼 기술과 관련된 특허를 5건 출원

[ 엘지전자 주식회사 주요특허 리스트 ]

공개번호 (출원일)	명칭	기술적용분야	도면
KR2019-0121273 (2019.10.07)	차량 정보 연계 콘텐츠 제공 장치 및 방법	차량으로부터 획득한 사용자 소지 상품 정보, 사용자 착용 의상 정보, 차량 위치 정보 등 차량에서 제공하는 정보에 기초하여 관련 캐릭터를 판별하고, 판별된 캐릭터와 연계된 콘텐츠를 제공하는 기술	
KR2019-0107286 (2019.08.30)	차량용 광고 제공 장치 및 차량용 광고 제공 방법	인공 지능 알고리즘을 이용하여, 서버에서 생성된 데이터를 수신하고, 상기 데이터에 기초하여, 차량이 주행 중인 도로의 외곽에 매칭되는 증강 현실 광고 데이터를 생성하는 기술	
KR2019-0101327 (2019.08.12)	구독 제품 가격 산정 방법 및 가격 산정 장치	사물 인터넷을 위해 연결된 5G 환경에서 인공지능 알고리즘 또는 기계학습알고리즘을 실행하여 구독 중인 제품의 인수 가격을 산정하여 사용자에게 인수를 제시한 결과를 보상으로서 강화 학습하는 학습 모델 생성하는 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ **주식회사 케이티**

- 주식회사 케이티는 '15년도 1월부터 출원을 시작하여 아직까지 등록된 특허는 없는 것으로 파악됨
  - 주식회사 케이티의 공개특허를 대상으로 분석한 결과, 한국의 유무선 통신 서비스 기업으로, 빅데이터 기반 마케팅인텔리전스 플랫폼 기술과 관련된 특허를 4건 출원

[ **주식회사 케이티 주요특허 리스트** ]

공개번호 (출원일)	명칭	기술적용분야	도면
KR2020-0094850 (2019.01.24)	서비스 상품 해지 원인 분석 시스템 및 방법	서비스 상품 해지 원인 분석 시스템은, 유·무선 통신 서비스 상품에 대한 고객별 잠재적 해지 원인을 분석하는 기술	
KR2020-0063646 (2018.11.28)	음성 광고 제공 방법 및 시스템	인공 지능 스피커 등 음성 대화 서비스를 제공하는 단말에서 사용자의 특성이나 사용 환경 등을 고려하여 타겟팅된 음성 광고를 제공하는 기술	
KR2020-0025054 (2018.08.29)	광고를 제공하는 서버, 사용자 단말 및 방법	사용자 단말로 제공된 제 1 차 관심 광고에 대한 피드백 정보를 수집하고, 수집된 피드백 정보를 분석하여 유의미한 광고를 학습한 후, 학습된 유의미한 광고에 대한 정보에 기초하여 제 2 차 관심 광고를 검색하는 기술	

\* 공개특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

## 다. 기술진입장벽 분석

### (1) 기술 집중력 분석

- 빅데이터 기반 마케팅인텔리전스 플랫폼 관련 기술에 대한 시장관점의 기술독점 현황분석을 위해 집중률 지수(CRn: Concentration Ratio n, 상위 n개사 특허점유율의 합) 분석 진행
  - 상위 4개 기업의 시장점유율이 0.08로 빅데이터 기반 마케팅인텔리전스 플랫폼 분야에 있어서 독과점 정도는 매우 낮은 수준으로 판단
  - 국내 시장에서 중소기업의 점유율 분석결과 0.75로 해당 기술에 대하여 중소기업이 진입하여 활동하고 있는 것으로 파악

[ 주요출원인의 집중력 및 국내시장 중소기업 집중력 분석 ]

	주요출원인	출원건수	특허점유율	CRn	n
주요 출원인 집중력	INTERNATIONAL BUSINESS MACHINES CORPORATION(미국)	27	3.1%	0.03	1
	FACEBOOK, INC.(미국)	21	2.4%	0.05	2
	MICROSOFT TECHNOLOGY LICENSING, LLC(미국)	14	1.6%	0.07	3
	ADOBE SYSTEMS INCORPORATED(미국)	13	1.5%	<b>0.08</b>	<b>4</b>
	YAHOO JAPAN CORP(일본)	12	1.4%	0.10	5
	BANK OF AMERICA CORPORATION(미국)	8	0.9%	0.11	6
	삼성전자 주식회사(한국)	8	0.9%	0.12	7
	YAHOO! INC.(미국)	8	0.9%	0.13	8
	MICROSOFT CORPORATION(미국)	8	0.9%	0.13	9
	WALMART APOLLO, LLC(미국)	7	0.8%	0.14	10
	전체	<b>883</b>	<b>100%</b>	<b>CR4=0.08</b>	
	국내시장 중소기업 집중력	출원인 구분	출원건수	특허점유율	CRn
중소기업(개인)		129	75.4%	0.75	
대기업		26	15.2%		
연구기관/대학		16	9.4%		
전체		<b>171</b>	<b>100%</b>	<b>CR중소기업=0.75</b>	

## (2) 특허소송 현황 분석

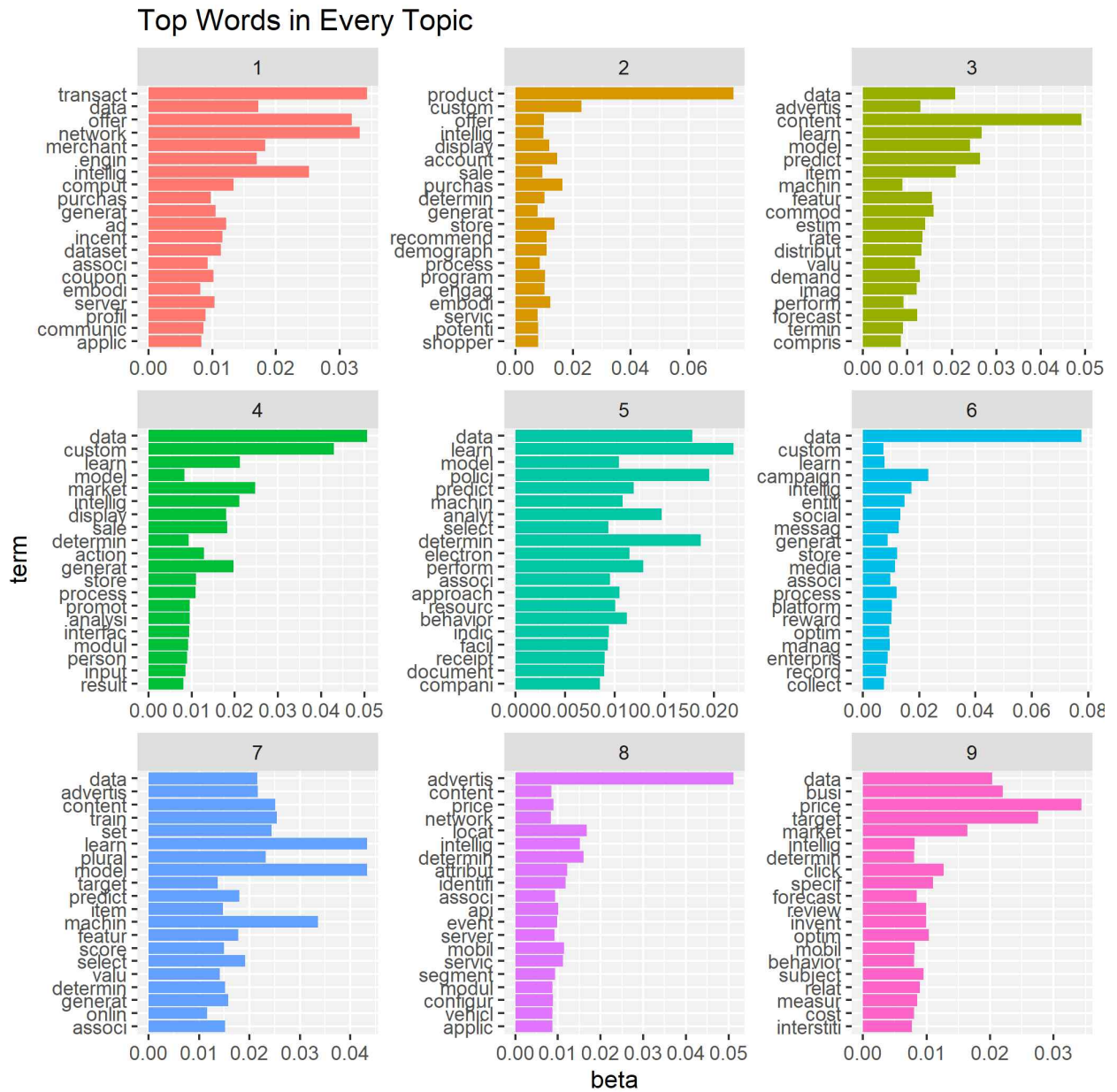
- 빅데이터 기반 마케팅인텔리전스 플랫폼 분야 관련 특허소송 이력은 검색되지 않음
  - 따라서 국내기업이 미국시장에 진입하는 경우, 해당 분야를 선점할 수 있을 것으로 판단
  - 다만, 빅데이터 수집과 관련해서는 4개의 소송이 검색

## 5. 요소기술 도출

### 가. 특허 기반 토픽 도출

- 650개의 특허의 내용을 분석하여 구성 성분이 유사한 것끼리 클러스터링을 시도하여 대표성이 있는 토픽을 도출

[ 빅데이터 기반 마케팅 인텔리전스 플랫폼 도구에 대한 토픽 클러스터링 결과 ]





## 나. LDA<sup>20)</sup> 클러스터링 기반 요소기술 도출

### [ LDA · 클러스터링 기반 요소기술 키워드 도출 ]

No.	상위 키워드	대표적 관련 특허	요소기술 후보
클러스터 01	transact, network, offer, intelligence, merchant, data, engin, computing, ad, incent	<ul style="list-style-type: none"> <li>• App social network via linket and ads for mobile deep links</li> <li>• Intelligent clearing network</li> </ul>	마케팅 인텔리전스
클러스터 02	product, custom, purchase, account, store, embodied, display, recommend, demography, program	<ul style="list-style-type: none"> <li>• Immersive crowdfunding and crowdsourcing system</li> <li>• Process and automated system for discovering and optimizing leadership and constituency correlations to increase participation and sustainability for election outcomes of governments and other organizations</li> </ul>	-
클러스터 03	content, learn, predict, model, item, data, commodify, feature, estimate, rate	<ul style="list-style-type: none"> <li>• Information processing apparatus, information processing method, and program</li> <li>• Demand forecasting system with improved process for adapting forecasts to varying time slots</li> </ul>	감성 및 사용자 사전 분류기술
클러스터 04	forecast, select, api, deal, asset, pip, custom, advertise, network, analytic	<ul style="list-style-type: none"> <li>• Sales promotion information providing service and sales promotion information providing program</li> <li>• Action data identification system</li> </ul>	마케팅 인텔리전스
클러스터 05	learn, policies, determine, data, analytic, perform, predict, electron, behavior, machine	<ul style="list-style-type: none"> <li>• Flow line prediction device, flow line prediction method and program</li> <li>• Statistically and ontologically correlated analytics for business intelligence</li> </ul>	-
클러스터 06	data, campaign, intelligence, entities, social, message, store, process, media, platform	<ul style="list-style-type: none"> <li>• Methods and apparatus for campaign mapping for total audience measurement</li> <li>• Multi-channel campaign planning</li> </ul>	-
클러스터 07	learn, model, machine, train, content, set, plural, advertise, data, select	<ul style="list-style-type: none"> <li>• Recommender systems and methods using cascaded machine learning models</li> <li>• Training and utilizing multi-phase learning models to provide digital content to client devices in a real-time digital bidding environment</li> </ul>	개인화 마케팅
클러스터 08	advertise, location, determine, intelligence, attribute, identification, mobile, service, api, event	<ul style="list-style-type: none"> <li>• Apparatus for setting advertisement time slot and method thereof</li> <li>• Api pricing based on relative value of api for its consumers</li> </ul>	마케팅 인텔리전스
클러스터 09	price, target, business, data, market, click, specific, optimize, review, invent	<ul style="list-style-type: none"> <li>• Allaroundearth</li> <li>• System and method for click fraud protection</li> </ul>	텍스트 데이터 분석 기술

20) Latent Dirichlet Allocation

### 다. 특허 분류체계 기반 요소기술 도출

□ 빅데이터 기반 마케팅인텔리전스 플랫폼 관련 특허에서 총 5개의 주요 IPC코드(메인그룹)를 산출하였으며, 각 그룹의 정의를 기반으로 요소기술 키워드를 아래와 같이 도출

[ IPC 분류체계에 기반한 요소기술 도출 ]

IPC 기술트리		
(서브클래스) 내용	(메인그룹) 내용	요소기술 후보
(G06N) 특정 계산모델 방식의 컴퓨터시스템	• (G06N-005) 지식기반모델을 이용한 컴퓨터시스템	마케팅 인텔리전스, 텍스트 데이터 분석 기술, 자연어 처리 기술, 비정형 텍스트 데이터 처리기술
	• (G06N-003) 생체모델기반 컴퓨터시스템	
	• (G06N-007) 특정수리모델에 기반을 둔 컴퓨터시스템	개인화 마케팅
	• (G06N-099) 이 서브클래스의 다른 그룹으로 분류되지 않는 주제사항	개인화 마케팅
(G06Q) 관리용, 상업용, 금융용, 경영용, 감독용 또는 예측용으로 특히 적합한 데이터 처리 시스템 또는 방법; 그 밖에 분류되지 않는 관리용, 상업용, 금융용, 경영용, 감독용 또는 예측용으로 특히 적합한 시스템 또는 방법	• (G06Q-030) 거래, 예. 쇼핑 또는 전자상거래	마케팅 인텔리전스, 개인화 마케팅

## 다. 최종 요소기술 도출

- 산업·시장 분석, 기술(특허)분석, 전문가 의견, 타부처로드맵, 중소기업 기술수요를 바탕으로 로드맵 기획을 위하여 요소기술 도출
- 요소기술을 대상으로 전문가를 통해 기술의 범위, 요소기술 간 중복성 등을 조정·검토하여 최종 요소기술명 확정

### [ 빅데이터 기반 마케팅 인텔리전스 플랫폼 분야 요소기술 도출 ]

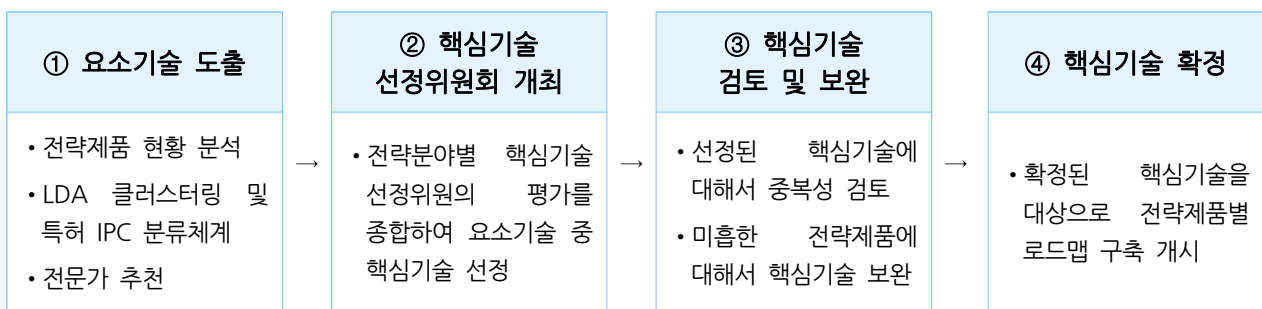
요소기술	출처
마케팅 인텔리전스	특허 클러스터링, IPC 기술체계, 전문가추천
감성 및 사용자 사전 분류기술	특허 클러스터링, 전문가추천
텍스트 데이터 분석 기술	특허 클러스터링, IPC 기술체계, 전문가추천
자연어 처리 기술	IPC 기술체계, 전문가추천
개인화 마케팅	특허 클러스터링, IPC 기술체계, 전문가추천
비정형지수산출 기술	전문가추천
비정형 텍스트 데이터 처리기술	IPC 기술체계, 전문가추천
비정형 텍스트 추출 기술	전문가추천

## 6. 전략제품 기술로드맵

### 가. 핵심기술 선정 절차

- 특허 분석을 통한 요소기술과 기술수요와 각종 문헌을 기반으로 한 요소기술, 전문가 추천 요소기술을 종합하여 요소기술을 도출한 후, 핵심기술 선정위원회의 평가과정 및 검토/보완을 거쳐 핵심기술 확정
- 핵심기술 선정 지표: 기술개발 시급성, 기술개발 파급성, 기술의 중요성 및 중소기업 적합성
  - 장기로드맵 전략제품의 경우, 기술개발 파급성 지표를 중장기 기술개발 파급성으로 대체

#### [ 핵심기술 선정 프로세스 ]



### 나. 핵심기술 리스트

#### [ 빅데이터 기반 마케팅 인텔리전스 플랫폼 분야 핵심기술 ]

핵심기술	개요
마케팅 인텔리전스	• 고객이 생성한 텍스트 데이터를 기반으로 성향과 감정, 요구사항을 파악하고 지수화된 정보와 융합하여 패턴과 미래를 예측해 마케팅 전략에 활용
감성 및 사용자 사전 분류기술	• 텍스트 데이터에서 분석 분야와 목적에 맞는 주요 단어들을 자동 또는 반자동화 하여 분류하고 강도를 구분하여 중요도에 따라 단어를 분류하는 기술
텍스트 데이터 분석 기술	• 비정형 텍스트 기반으로 추출한 단어들에 고객의 성향, 요구, 감정 등의 상태를 그룹화하여 다양한 유형의 텍스트 데이터 분석이 가능하도록 돕는 기술
자연어 처리 기술	• 고객의 성향과 감정, 니즈 등의 특징을 타당성 높은 단어들을 추출하는 방식으로 기존의 자연어 처리 방법을 고도화하는데 집중되고 있는 기술
개인화 마케팅	• 영상, 음성, 문서 등 다양한 형태의 비정형데이터에 대한 개인정보 탐지 및 비식별처리를 통해 마케팅 분석에 활용하는 기술

### 다. 중소기업 기술개발 전략

- 빅데이터 단위기술에 대한 융복합 기술이 공백 상태인 만큼 중소벤처기업의 기술 선점 가능
- 데이터 개방정책에 기반한 빅데이터 확보 가능성이 높아짐에 따라 대기업 독점의 데이터 시장에서 플랫폼 경쟁력 제고
- 빅데이터 확보의 보편성 추세는 대기업 위주에서 중소벤처 기업으로 빠르게 시장 구조가 이동 중이며, 이에 따른 플랫폼 사용의 경로 의존성 확보에 따른 시장선점효과 기대
- 정부 R&D사업과 현장실습 프로그램 연계로 부족한 인력 확충하여 경쟁력 제고

### 라. 기술개발 로드맵

#### (1) 중기 기술개발 로드맵

[ 빅데이터 기반 마케팅 인텔리전스 플랫폼 분야 중기 기술개발 로드맵 ]

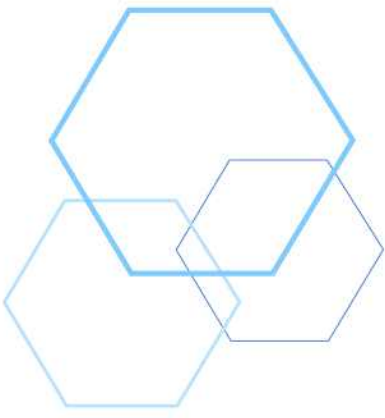
빅데이터 기반 마케팅 인텔리전스 플랫폼	비정형/정형 데이터 비식별 고도화 및 데이터 표준화를 통한 다양한 통찰 제공			
	2021년	2022년	2023년	최종 목표
마케팅 인텔리전스				마케팅에 적합한 5개 융복합 모델 개발
감성 및 사용자 사전 분류기술				융복합 분석이 가능한 5개 감성 및 사용자 사전 구축
텍스트 데이터 분석 기술				AI 및 머신러닝을 활용한 비정형지수 산출과 분류 기술 개발, 이종 간 데이터 결합
자연어 처리 기술				업종별 자연어 처리 성능 향상 및 광고 필터 등 타당성 개선
개인화 마케팅				비정형/정형 데이터 비식별 고도화 및 데이터 표준화

(2) 기술개발 목표

□ 최종 중소기업 기술로드맵은 기술/시장 니즈, 연차별 개발계획, 최종목표 등을 제시함으로써 중소기업의 기술개발 방향성을 제시

[ 빅데이터 기반 마케팅 인텔리전스 플랫폼 분야 핵심요소기술 연구목표 ]

핵심기술	기술요구사항	연차별 개발목표			최종목표	연계R&D 유형
		1차년도	2차년도	3차년도		
마케팅 인텔리전스	업종별 융복합 모델링	1개 업종에 대한 융복합 모델 개발	2~3개 업종에 대한 융복합 모델 개발	4~5개 업종에 대한 융복합 모델 개발	마케팅에 적합한 5개 융복합 모델 개발	기술혁신
	마케팅 추천 지능 엔진 정확도	0.7	0.8	0.9	0.9	기술혁신
감성 및 사용자 사전 분류기술	업종별 사용자사전 구축	1개 업종에 대한 감성 및 사용자 사전 구축	2~3개 업종에 대한 감성 및 사용자 사전 구축	4~5개 업종에 대한 감성 및 사용자 사전 구축	융복합 분석이 가능한 5개 감성 및 사용자 사전 구축	산학연
	업종별 사용자사전 분류					
	감성 및 사용자 사전 분류 지식 규모(Triples)	1억	10억	50억		
텍스트 데이터 분석 기술	텍스트 데이터 정량화	항목별 비정형지수 산출 방법론 개발	고객성향, 요구, 감정 등에 대한 머신러닝 분류 기법 개발	AI 및 머신러닝을 활용한 비정형지수와 정형데이터 융합 방법 개발	AI 및 머신러닝을 활용한 비정형지수 산출과 분류 기술 개발과 이종간 데이터 결합	산학연
	의미자질 추출 정확도	0.8	0.85	0.9	0.9	산학연
자연어 처리 기술	업종별 자연어 처리 결과에 대한 성능개선(%)	0.9	0.95	0.98	업종별 자연어 처리 성능 향상 및 광고 필터 등 타당성 개선	산학연
개인화 마케팅	비식별 처리	비정형 데이터 내 개인정보 비식별 처리 기술 추가	비정형/정형의 결합데이터의 표준화 기술 개발	영상, 음성, 이미지 등 다양한 비정형 데이터 비식별 고도화	비정형/정형 데이터 비식별 고도화 및 데이터 표준화	상용화
	데이터 품질관리					
	맞춤형 마케팅 추천 정확도	0.7	0.8	0.9		



전략제품 현황분석

# 공공 빅데이터 수집 및 분석 시스템







# 공공 빅데이터 수집 및 분석 시스템

## 정의 및 범위

- 다양한 공공 데이터 수집을 통해 대량의 정형 또는 비정형 빅데이터를 분석함으로써 국민건강 예방 및 정책서비스 설계가 가능하도록 가치 있는 공공기관의 정보를 추출할 수 있는 시스템

## 전략 제품 관련 동향

시장 현황 및 전망	제품 산업 특징
<ul style="list-style-type: none"> <li>(세계) '18년 1,688억 달러 규모에서 연평균 12.9%로 성장하여 '24년 3,500억 달러 규모로 성장 전망</li> <li>(국내) '18년 1,584억 원 규모에서 연평균 12.5%로 성장하여 '24년 3,211억 원 규모로 성장 전망</li> </ul>	<ul style="list-style-type: none"> <li>공공 빅데이터의 활용가치가 증가함에 따라 독점적 데이터의 수집과 생성에 관한 관심 증가</li> <li>데이터 수집, 관리, 품질표준, 분석 및 시각화에 대한 전반적인 융합기술 기대감 증가</li> </ul>
정책 동향	기술 동향
<ul style="list-style-type: none"> <li>공공 빅데이터에 관한 관심이 증가하면서 데이터 품질에 대한 제도적 개선의 움직임이 활발함</li> <li>부처와 기관별 데이터 칸막이에 따른 데이터 통합 문제를 해결하기 위한 데이터 거버넌스 측면에서의 제도개선의 요구 증가</li> </ul>	<ul style="list-style-type: none"> <li>다양한 형태와 유형의 공공 빅데이터 수집을 위한 특화된 수집기 기술개발</li> <li>대용량의 공공 빅데이터의 통합 수집, 저장관리를 위한 인프라 기술개발 요구</li> </ul>
핵심 플레이어*	핵심기술
<ul style="list-style-type: none"> <li>(해외) Clickworker, Qburst, ScienceSoft 등</li> <li>(대기업) 삼성에스디에스, 삼성전자, 케이티</li> <li>(중소기업) 위엠비, 빅인사이트, 와이즈넷</li> </ul>	<ul style="list-style-type: none"> <li>데이터 유형과 형태에 따른 검색기 및 크롤러</li> <li>빅데이터 연계 통합 기술</li> <li>빅데이터 연계 통합 기술</li> <li>정형/비정형 융합분석 기술</li> <li>빅데이터 수집 및 분석 플랫폼</li> <li>통합데이터 품질표준 기술</li> </ul>

\*생태계 취약 전략제품

## 중소기업 기술개발 전략

- 공공 빅데이터에 대한 부처와 기관별로 상이한 데이터 형태의 연계수집 및 통합관리 해결
- 기존에 구축된 인프라와 개발된 기술을 활용하여 진입장벽을 해소하고 융합분석에 특화
- 산·학 연계 및 테스트베드 확보를 통한 개발기술 검증으로 국내시장 개척 및 선점 필요

## 생태계 강화방안

- 시장 자체가 아직 초기 단계로 데이터 3법의 시행과 더불어 발전할 것이며, 이를 위해 스타트업 육성 등 정책적 지원이 필요
- 데이터 수집 및 통합이 특정 기관이나 부처에 한정되어 있으며, 더욱 많은 다양한 기관과 부처의 공공 빅데이터를 연계할 수 있는 데이터 거버넌스 정책 필요

# 1, 개요

## 가. 정의 및 필요성

### (1) 정의

- 다양한 공공 데이터 수집을 통해 대량의 정형 또는 비정형 빅데이터를 분석함으로써 국민건강 예방 및 정책서비스 설계가 가능하도록 가치있는 공공기관의 정보를 추출할 수 있는 시스템
  - 공공분야의 데이터를 활용하여 국민의 건강, 안전, 복지, 생활 등 다양한 분야에서의 정책설계와 집행에 활용
  - 부처별, 공공기관별 산재되어 있는 공공데이터를 식별하고, 공공분야에 특화되어 있는 데이터를 통합 수집 할 수 있는 기술
- 부처별로 산재되어 있거나 중복되어 분야별 데이터의 교차 및 융합분석이 가능하도록 정책 의사결정을 지원하는 시스템
  - 다양한 분야의 공공데이터를 통합 수집할 수 있어 다양한 정책문제를 확인하고 분석하는 가공 및 분석하는 기술
  - 보유한 공공데이터의 정형 및 비정형 등 데이터의 유형과 특징에 따라 다양한 빅데이터를 수집하고 분석결과에 대한 시각화가 가능한 기술

### (2) 필요성

- 정책의제 설정, 정책설계, 집행의 과정에서 공공데이터의 활용가치가 증가하고 있으며, 공공 빅데이터를 수집하고 가공하여 분석하는 기술에 대한 업무연관성이 증가
  - 공공데이터의 활용가치가 증가하고 있음에도 불구하고 부처별로 산재되어 있고 중복되어있는 데이터를 일일이 수집하고 저장해서 분석해야하는 번거로움 존재
  - 산재되어 있고, 중복되어있는 공공데이터를 일괄적으로 수집하여 가공해야 하는 필요성이 제기됨
- 다양한 이해관계자와 정책의 영역이 분야별로 모호해지고 있으며, 협력적이고 융합적인 정책서비스 제공에 정책설계 다양화로 정책 의사결정을 위해서 분야별 공공데이터의 융복합 분석지원이 중요해짐
  - 이해관계자들이 다양해 졌고 정책문제도 분야별로 얽혀 복잡해짐에 따라 다양한 분야의 공공데이터의 상관관계를 살펴 정책에 활용해야 함
  - 이종 간 분야의 데이터를 교차 및 상관성을 분석하기 위해서는 이종 간 분야의 공공데이터를 통합 및 일괄 수집해야 하고 분석을 시도할 수 있는 시스템이 필요함

- 공공 빅데이터를 활용하여 새로운 서비스 및 공공정책에 활용할 기관과 민간 등 연구기관에서 활용 가능
  - 데이터를 활용하는 이용자가 개별적으로 공공데이터를 체계적으로 분류하는 과정을 거쳐야 하므로 시간과 비용이 들고 신속한 서비스 개발이 어려움
  - 데이터가 전산화되어 있지 않고 수기로 작성되어 있거나, 정형화되어있지 않고, 비정형으로 데이터를 생산하고 있는 사례가 많아 이를 정리하고 관리할 수 있어야 함
  - 공공데이터 포털은 일부 제한적인 데이터만을 적재하고 있는 반면에 중복되어있거나 자료의 생성이 유사한 정도, 융합활용범위 및 추천 등을 통해 공공데이터 활용을 극대화할 수 있는 기능이 요구됨
  - 공공정책 카테고리에 따른 부서별 데이터 활용 정도와 생성현황 식별을 통해 공개할 수 있거나 검색 및 수집을 할 수 있는 데이터의 구별 등 공공데이터 전반에 대한 품질평가와 데이터 보정 등에 대한 정보제공이 절실함
  
- 다양한 정책 수요를 가지고 있는 국민의 만족도 높은 증거기반의 정책 및 행정 서비스제공을 위한 기술, 사회, 경제 전반의 파급 및 연쇄효과가 기대됨
  - 기술적 파급효과
    - 산재하여 있으나 공유되고 있지 않은 부처와 기관별 공공데이터에 대한 검색과 수집을 통한 데이터 관리 기능
    - 공공데이터가 필요한 민간, 공공, 연구기관 등 다양한 분야에 활용 가능한 데이터 보정, 중복 및 유사처리 기능
    - 검색을 통한 부처와 기관별 데이터의 융복합 분석이 가능한 활용범위 및 데이터 형태 추천기능
  - 경제적 파급효과
    - 부처별 기관별로 데이터를 수집하기 위해 투입되는 인력 및 시간을 공공데이터 검색과 수집기능을 통해 효율성을 달성할 수 있음
    - 데이터의 유사성 및 중복성 비교가 가능해짐에 따라 데이터 생성과 관리 측면에서의 조직 및 업무 재설계로 효율적인 조직운영이 가능해짐
    - 공공 빅데이터의 성격이 강하면서 쉽게 공개되기 어려웠던 정보들을 제공함에 따라 일부 독점적(카드데이터, 통신데이터 등)인 공공데이터를 대체하거나 보완할 수 있는 데이터를 제공하여 타 산업 및 데이터 경계를 활성화할 수 있음
  - 사회적 파급효과
    - 다양한 사회문제, 공공의 행위 때문에 생성된(공공기관의 행정문서 및 IoT 데이터 등) 데이터를 활용할 수 있어 양질의 정책 및 공공서비스를 창출할 수 있을 것으로 기대
    - 다양한 공공데이터의 품질과 가치를 높일 수 있어 근거기반의 좋은 정책을 설계하는 데 활용하여, 사회문제를 해결하는 데 많은 도움을 줄 수 있을 것으로 기대

## 나. 범위 및 분류

### (1) 가치사슬

- (가치사슬) 공공데이터 활용을 필요로 하는 정부부처 및 기관과 민간기업 등에서 공공데이터 검색, 수집, 탐색, 분석, 시각화 결과를 제공하여 공공데이터의 활용과 가치를 극대화하는 시스템
  - (후방산업) 부처와 기관별로 산재하여 있는 공공데이터를 탐색하고 수집할 수 있는 기술과 수집된 공공데이터에 대한 분석을 통해 활용가치를 극대화할 수 있도록 인프라 및 소프트웨어 구축
    - 부처와 기관별로 관리되고 있는 데이터의 유사성과 중복성 그리고 품질 문제를 해결하기 위해 데이터 품질 기술이 연계되어야 함
    - 수집된 공공 빅데이터 활용을 위한 데이터 간 융복합 분석을 추천하거나 분석을 하여 결과를 시각화 해주는 기술 및 방법론의 개발이 필요
  - (전방산업) 공공데이터를 활용하는 의료, 복지, 고용, 경제 정책과 민간에서의 시장조사와 마케팅 분석에 활용하는 시스템 및 플랫폼
    - 공공데이터를 통한 정책기획 및 설계, 시장조사와 상품기획, 마케팅 등에 활용될 수 있는 시스템 및 플랫폼 기술

#### [ 공공 빅데이터 수집 및 분석 시스템 분야 산업구조 ]

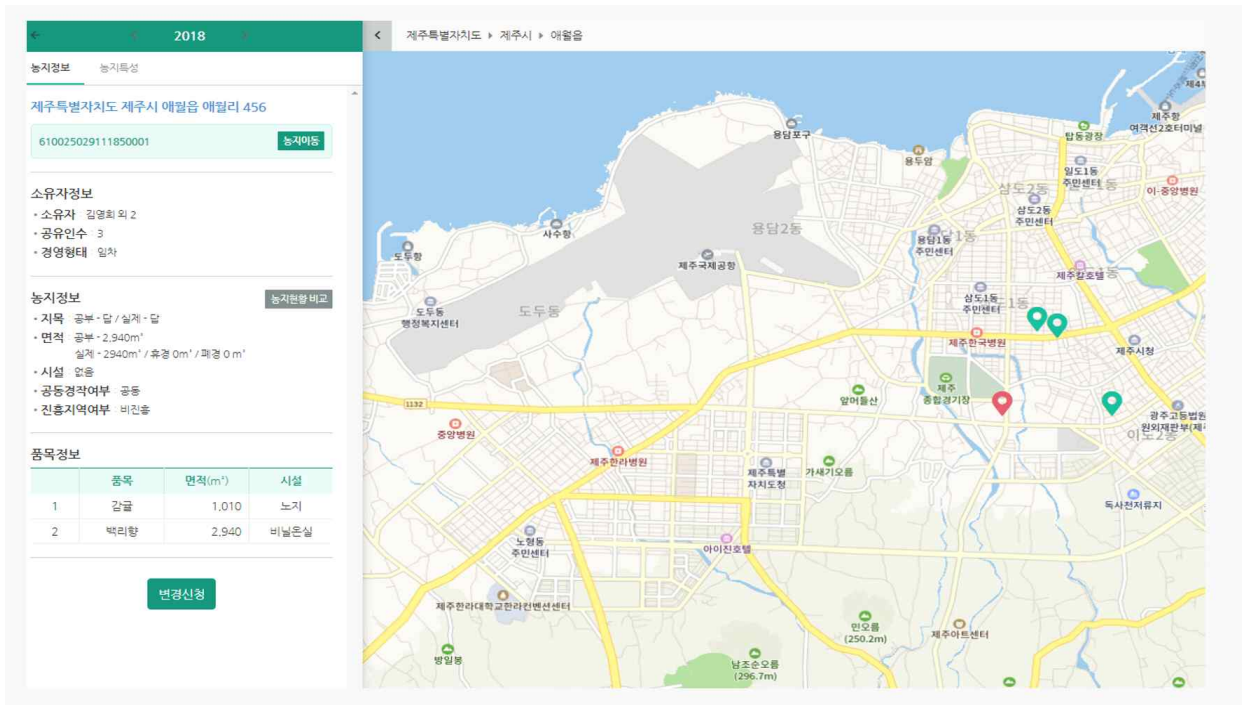
후방산업	공공 빅데이터 수집 및 분석 시스템	전방산업
데이터 수집기술, 데이터 품질 및 고도화 기술, 분석 및 시각화 기술, 데이터 관리, 데이터 컨설팅 등	부처와 기관별 공공데이터를 검색, 수집하여 분야별로 데이터 특성에 따른 분석방법론을 추천하고, 분석결과에 대한 시각화를 통해 정책 및 마케팅에 의사 결정해줄 수 있는 시스템	재난, 의료, 교통, 복지, 고용, 경제, 행정 등 근거기반 정책설계 및 분석, 민간분야 시장조사 및 기획과 마케팅 분석 시스템(플랫폼)

## (2) 용도별 분류

### ◎ 공공 빅데이터 수집 및 분석 시스템은 활용 측면에서 공공활용과 민간활용 분야로 분류

- (공공분야) 국민 생활과 안전 및 경제 등 행정과 정책서비스를 극대화하기 위해 필요한 공공데이터의 활용
  - 부처별로 산재하여 있는 중복 및 유사 데이터를 탐색하여 필요에 따라 활용 가능
  - 분야별 공공데이터를 탐색할 수 있어 시간 및 비용을 절감
  - 분야별 공공데이터를 탐색하고 수집 정리할 수 있어 관련 정책에 활용도가 높은 공공데이터를 분류하고 관리할 수 있음
  - 데이터의 유형과 특징에 따라 연계 또는 융합 분석할 수 있는 데이터인지 판단할 수 있도록 지원함
  - 분석방법 등을 지원하여 전통적인 분석방법으로는 알 수 없었던 통찰력과 상관성 높은 정책분석방법을 추천하는 기능
  - 공공데이터뿐만 아니라 다양한 외부 데이터들과도 연계 가능 여부를 체크할 수 있도록 지원
  - 공공데이터의 활용가치가 증가함에 따라 공공데이터를 활용한 행정업무지원서비스 및 플랫폼 개발 활성화

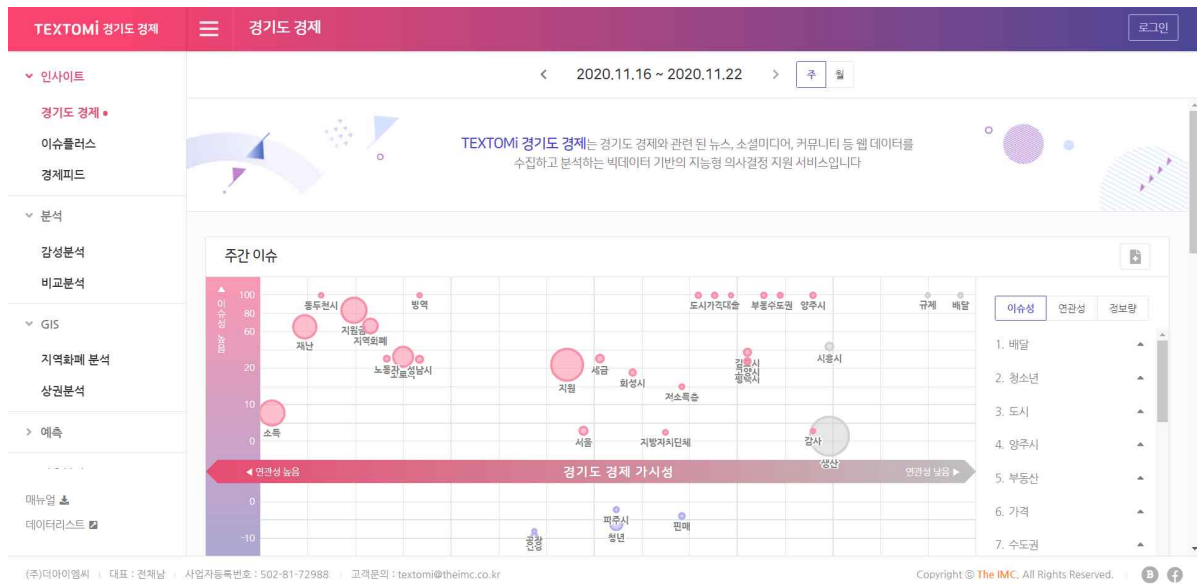
### [ GIS 기반 농경정보 서비스 ]



\* 출처: GIS 기반 농경정보 서비스(2018)

- (민간분야) 소비자들의 니즈와 유형, 생활양식의 변화와 패턴을 확인하여 상품기획이나 마케팅을 극대화하는 데 필요한 공공데이터 활용
  - 기업 입장에서 고객 및 시장 창출을 위해 공공데이터를 활용하여 대중의 생활양식 및 패턴 등을 확인할 수 있고 신시장 창출과 발굴에 활용
  - 민간 등 기업이 가지고 있는 데이터와 공공데이터와 결합하거나 융합하여 새로운 통찰을 제공할 분석의 기회를 제공
  - 공공의 가치와 연계된 서비스 창출과 분석에 활용

[ 대한민국 정책브리핑 ]



\* 출처: 경기도 경제 textomi 플랫폼(2020)

## 2. 산업 및 시장 분석

### 가. 산업 분석

◎ 공공데이터 가치와 활용에 관한 관심이 증가하고 있으며 이에 대한 정책적 지원이 가속화

- 정보는 데이터 경제 달성을 목표로 공공데이터에 대한 수집과 저장 그리고 유통, 활용 전반에 걸친 가치 창출 체계를 마련
  - 정부는 한국판 뉴딜 정책의 일환으로 분야별 데이터의 확충과 연계, 활용, 5G, AI 융합 서비스를 실현하기 위한 데이터 댐 정책을 시행
  - 빅데이터 플랫폼 및 센터구축사업을 통해 공공데이터뿐만 아니라 민간데이터를 통합 구축하는 정책을 시행
    - 공개하기 어렵고, 잠자고 있던 민간의 데이터를 수집 통합 구축하여 보다 활발한 데이터 활용을 이끌어내는 정책지원
  - 다양한 빅데이터 수집 통합 구축 정책을 통해 공공의 사회문제 해결과 공공데이터 품질은 물론 공공 빅데이터 활용이 확대되고 있음

#### [ 대한민국 정책브리핑, 데이터 댐 ]



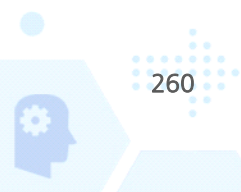
\* 출처: 대한민국 정책브리핑(과학기술정보통신부 블로그), 2020.03.13.



- 공공데이터 활성화를 위한 제도 및 규제 완화 등 정책적 활용을 위한 데이터 거버넌스 중심의 법령 정비가 활발
  - 부처 간 데이터 칸막이를 없애고 공공데이터의 자유로운 유통을 가능케 하려는 데이터 전담조직에 대한 논의가 활발함
  - 데이터의 통합관리, 데이터의 활용촉진, 데이터 거래기반 구축, 디지털, 데이터 경제 활성화 등에 대한 다양한 논의가 진행 중
  - 데이터 관련 부처 간의 기능 중복 및 협업 문제 등 조정에 대한 논의가 진행 중
- 공공 빅데이터의 활용을 극대화하기 위한 개인정보보호 등의 문제 논의
  - 개인정보 보호법, 정보통신망법, 신용정보법과 함께 공공 빅데이터의 활용가치를 증대시킬 공공데이터법과 데이터 기반 행정법 등이 새로 개정됨에 따른 공공데이터 비식별 처리 및 품질관리에 대한 기술적 요구 증대
  - 특히 행정안전부의 데이터 기반 행정법은 공공데이터에 대한 관리체계구축, 표준화, 데이터 통합관리 플랫폼 운영 등 데이터 기반 행정의 기반 구축에 따라 공공데이터 수집 및 분석에 대한 기술적 요구가 커짐

◎ 공공 빅데이터 수집 및 분석 시스템 기술력 증대를 위한 연구 및 개발 추진

- 공공데이터 포털을 통해 공공데이터를 개방하고 활용을 독려하고 있지만, 데이터 품질의 수준이 낮아 민간데이터 등 타 데이터와 연계하는데 어려운 실정
- 공공 빅데이터를 수집하는 데 필요한 부처별 기관별 데이터 통합을 위한 수집연계기술 개발이 이루어지고 있음
  - 정부 기관과 부처별로 정보화 또는 빅데이터 기반 구축사업을 통해 내부에 보유하고 있는 공공 빅데이터를 활용하기 위한 플랫폼 전략을 시도하고 있음
  - 정보화 기반을 통해 데이터를 식별하고 수집할 수 있는 인프라를 조성하고 있으며 산재하여 있는 데이터를 통합 저장하고 구축하는 데 필요한 기술을 도입하고 있음
  - 데이터 수집 및 통합이 특정 기관이나 부처에 한정되어 있으며, 더욱 많은 다양한 기관과 부처의 공공 빅데이터를 연계할 수 있는 데이터 거버넌스 기술이 선행되어야 함
- 공공 빅데이터를 통합 수집하기 위해 이해관계기관별 데이터 거버넌스 정책이 필요
  - 공공 빅데이터 전담조직 또는 위원회를 설치하여 공공데이터 중복문제와 협업 문제를 해결하는 정부의 제도적 지원이 필요
  - 데이터 관련 조직뿐만 아니라 공공 빅데이터가 외부에서 수집될 수 있도록 데이터 형태와 유형을 표준화하고 수집 가능하도록 제도를 정비하는 등 기술적 정책지원도 필요

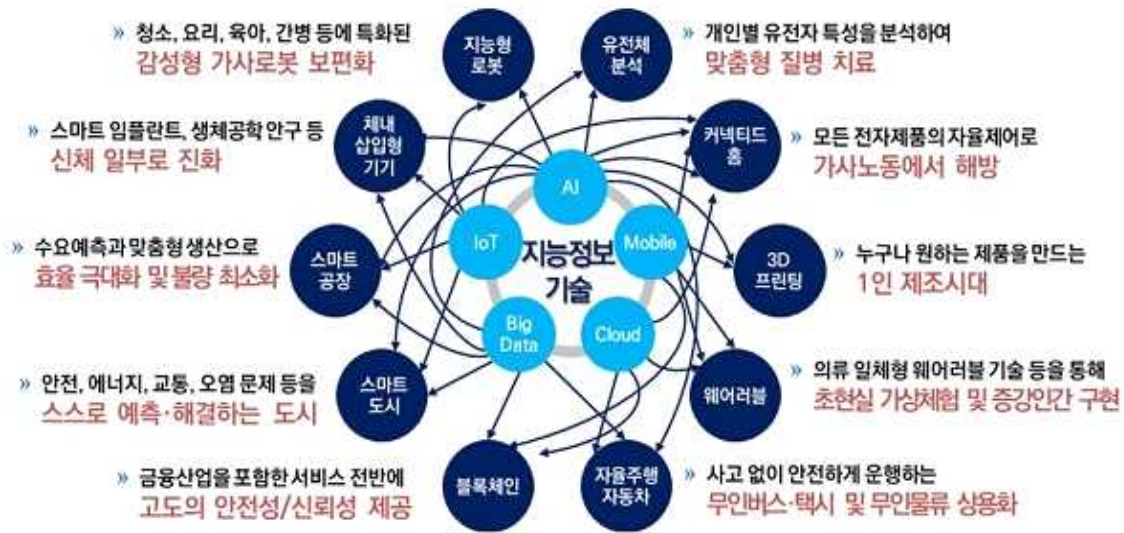




◎ 공공 빅데이터 활용에 대한 기대 및 관심 증대

- 공공데이터 개방과 데이터 관련 제도가 정비됨에 따라 공공데이터를 활용한 새로운 정책서비스 및 민간활용에 대한 기대가 증가 중
  - 공공 빅데이터 활용촉진을 위한 데이터 품질에 대한 수요가 꾸준히 증가하고 있음
  - 공공기관을 중심으로 빅데이터 기반구축사업 등 공공 빅데이터를 수집하고 통합 관리하여 새로운 분석 모델을 창출하는 기술적 요소의 관심이 증대
- 공공 빅데이터를 수집하고 연계하여 저장하는데 필요한 인프라 측면의 기술과 빅데이터 분석과 품질관리를 할 수 있는 AI 및 머신러닝 등 융합분석 기술에 대한 활용도 증가
  - IoT기반, 문서기반, 이미지 기반, 영상기반, 통계기반 등 다양한 공공 빅데이터를 수집하기 위한 대용량 수집기와 저장장치 등 인프라 구축 기술고도화 등의 기술개발이 기대
  - 정형/비정형 등 다양한 형태의 데이터를 통합 관리하기 위한 데이터 품질과 분석 기술에 대한 고도화 개발도 기대

[ 빅데이터, AI, 머신러닝 등의 융합분석 기술 활용 사례 ]



\* 출처: 카카오 정책산업 연구, 2017.04.14

## 나. 시장 분석

### (1) 세계시장

- Statista가 2020년에 발표한 빅데이터 분석 분야의 세계시장 규모는 '18년 1,688억 달러 규모로 평가되었으며 연평균 12.9%로 성장하여 '24년 3,500억 달러 규모로 성장 전망

[ 데이터 분석 분야의 세계 시장규모 및 전망 ]

(단위 : 십억 달러, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
세계시장	168.8	190.6	215.2	243.0	274.3	310.0	350.0	12.9

\* 출처: Statista, Revenue from big data and business analytics worldwide(2020)을 바탕으로 '23년 이후 '18~'22의 CAGR로 적용하여 네모아이씨지 재추정

### (2) 국내시장

- 과학기술정보통신부의 2018 데이터산업 현황 조사(2019.03)에 따르면, 빅데이터 분석 시장은 데이터산업 세부 시장 중 데이터솔루션 중분류별 시장에 속함
- 빅데이터 분석 분야의 국내 시장규모는 2018년 1,584억 원 규모에서 연평균 12.5%로 성장하여 2024년 3,211억 원 규모로 성장이 전망됨

[ 데이터 분석 분야의 국내 시장규모 및 전망 ]

(단위 : 억 원, %)

구분	'18	'19	'20	'21	'22	'23	'24	CAGR
국내시장	1,584	1,782	2,004	2,255	2,537	2,854	3,211	12.5

\* 출처: 2018 데이터 산업 현황 조사, 과학기술정보통신부(2019.03). 인용. '19년~'24년은 전체 빅데이터 세계시장 대비 국내시장의 비율로 추산한 CAGR로 네모아이씨지 재추정

- 국내 데이터 시장의 세부 시장은 연평균 20.7%의 성장추세를 보이고 있음
  - 데이터 구축 및 컨설팅 시장은 2019년까지 12.4%의 성장률을 기록하였으며 데이터 서비스는 2019년까지 33.6%의 연평균 성장률을 보이고 있어 성장세는 지속될 것으로 전망

[ 국내 데이터산업 세부 시장 현황 ]

(단위 : 억 원, %)

분류	중분류	2017년		2018년		2019년(E)		증감률 '17-'18		CAGR '17-'19(E)	
		전체	직접 매출	전체	직접 매출	전체	직접 매출	전체	직접 매출	전체	직접 매출
데이터 구축 컨설팅	데이터 구축	57,207	29,106	58,993	34,713	62,576	36,625	3.1%	19.0%	4.6%	12.1%
	데이터 컨설팅	1,687	1,687	2,297	2,297	2,346	2,346	36.1%	36.1%	17.9%	17.9%
	소계1	58,894	30,847	61,290	37,009	64,922	38,971	4.1%	20.0%	5.0%	12.4%
데이터 서비스	데이터 거래	2,918	2,713	4,175	3,796	4,315	3,915	43.1%	39.9%	21.6%	20.1%
	정보제공	61,570	12,747	67,580	22,673	74,664	24,867	9.8%	77.9%	10.1%	39.7%
	데이터 분석제공	3,690	2,878	4,023	3,633	4,382	3,932	9.0%	26.3%	9.0%	16.9%
	소계2	68,179	18,339	75,778	30,102	83,361	32,714	11.1%	64.1%	10.6%	33.6%
합계		127,073	49,186	137,068	67,111	148,283	71,685	7.9%	36.4%	8.0%	20.7%

\* 출처: 2019년 데이터산업 현황 조사(2019), 과학기술정보통신부

[ 공공 빅데이터 수집 및 분석시스템 SWOT 분석 ]

	긍정적	부정적
내부	<p>&lt; 강점(Strengths) &gt;</p> <ul style="list-style-type: none"> <li>• 정부주도의 공공데이터 개방 및 제도정비</li> <li>• 데이터 경제 등 정부의 정책적 공공데이터 지원</li> <li>• 빅데이터 기반 구축 등 공공부문의 사업 활성화</li> <li>• 공공 빅데이터의 생성과 수요 증가</li> <li>• 글로벌 수준의 IT 경쟁력</li> </ul>	<p>&lt; 약점(Weakness) &gt;</p> <ul style="list-style-type: none"> <li>• 부처 및 기관의 데이터 중복 문제</li> <li>• 활용가치에 비해 공공데이터의 품질 미흡</li> <li>• 부처 및 기관의 데이터 칸막이</li> <li>• 공공 빅데이터를 연계수집할 수 있는 인프라 부족</li> </ul>
외부	<p>&lt; 기회(Opportunities) &gt;</p> <ul style="list-style-type: none"> <li>• 세계 빅데이터 서비스 시장의 성장</li> <li>• 민간영역에서의 공공데이터 수요확장</li> <li>• 공공 빅데이터 품질 및 서비스의 꾸준한 요구</li> <li>• 빅데이터 인프라 및 분석 시장의 경쟁력 증가</li> </ul>	<p>&lt; 위협(Threats) &gt;</p> <ul style="list-style-type: none"> <li>• 기술표준 및 원천분석개발이 선진국 주도</li> <li>• 개인정보 처리 데이터에 대한 수집권한 등 미흡한 제도개선 필요</li> </ul>

### 3. 기술 개발 동향

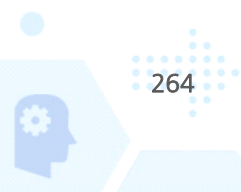
- 기술경쟁력
  - 공공 빅데이터 수집 및 분석 시스템은 미국이 최고기술국으로 평가되었으며, 우리나라는 최고기술국 대비 84%의 기술 수준을 보유하고 있으며, 최고기술국과의 기술격차는 1.4년으로 분석
  - 중소기업의 기술경쟁력은 최고기술국 대비 74.1%, 기술격차는 2.3년으로 평가
  - EU(92.5%)>한국>일본(79.5%)>중국(67.3%)의 순으로 평가
- 기술수명주기(TCT)<sup>21)</sup>
  - 공공 빅데이터 수집 및 분석 시스템은 6.13의 기술수명주기를 지닌 것으로 파악

#### 가. 기술개발 이슈

##### ◎ 다양한 데이터 형태의 공공 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 필요로함
  - 공공 빅데이터가 존재하는 위치에 따라 개방적일 수도 있고 폐쇄적일 수도 있으므로 수집 방법을 달리하는 수집기 기능(사용자가 원하는 데이터를 탐색할 수 있도록 다양한 검색조건을 지정할 수 있고 데이터 세트를 만들고 활용할 수 있는 기능)
- 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 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

21) 기술수명주기(TCT, Technical Cycle Time): 특허 출원연도와 인용한 특허들의 출원연도 차이의 중앙값을 통해 기술 변화속도 및 기술의 경제적 수명을 예측



□ 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 등 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
- 수집 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

□ 정형/비정형의 공공 빅데이터를 수집하여 정제, 분석 시각화 등의 모든 기능을 패키지형태로 제공해주는 시스템 또는 플랫폼 기술 개발

◎ 다양한 형태와 유형의 대용량 공공 빅데이터의 통합 수집 저장관리를 위한 인프라 기술 개발

- 공공 빅데이터의 특성상 데이터 형태의 다양성 문제를 통합하고 표준화해야 하는 문제를 해결하기 위해 메타데이터를 관리하고 데이터의 변경이 일관된 규칙을 따르는 기능 요구
- 공공 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
- 공공 빅데이터의 통합수집과 연계시 필요한 데이터베이스 관리 기술
  - RDBMS, NoSQL DBMS 등 관계형 모델과 비 관계형 구조를 갖는 데이터 관리 기술
  - 공공 빅데이터의 대용량 처리를 위한 데이터 메모리와 하드웨어 저장장치의 성능을 향상시킬 수 있는 인메모리 DBMS와 어플라이언스, 클라우드, 하이브리드DBMS 등 인프라 기술

◎ 공공 빅데이터 수집 및 분석시스템 분야의 핵심 원천기술 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
- 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

## 나. 생태계 기술 동향

### (1) 해외 플레이어 동향

- (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 ]



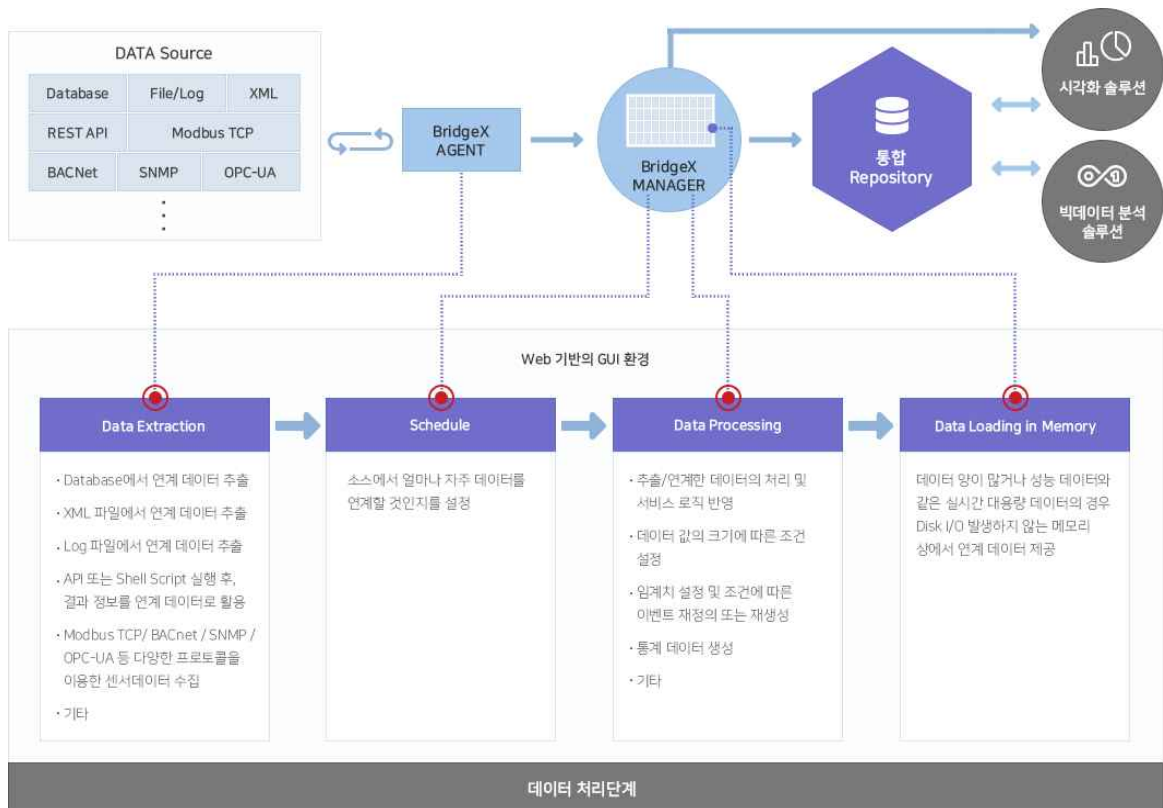
\* 출처: ScienceSoft

## (2) 국내 플레이어 동향

### □ (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) AI기반 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 ]



\* 출처: 와이즈넷

## 다. 국내 연구개발 기관 및 동향

### (1) 연구개발 기관

[ 공공 빅데이터 수집 및 분석시스템 분야 주요 연구조직 현황 ]

기관	소속	연구분야
한림대학교	한림대학교 산학협력단	<ul style="list-style-type: none"> <li>• 텍스트 데이터 통합수집</li> <li>• 분석에 사용될 데이터 셋 형성 및 데이터베이스 저장</li> </ul>
한국전자기술연구원	정보통신미디어연구본부 융합시스템연구본부 지능정보연구본부	<ul style="list-style-type: none"> <li>• 비정형 빅데이터를 기준에 따라 정형화 및 융합가공</li> </ul>
한국과학기술정보연구원	데이터분석플랫폼센터 계산과학플랫폼센터	<ul style="list-style-type: none"> <li>• 대용량 데이터 센터 구축, 플랫폼 연구개발</li> <li>• 빅데이터 공유 융합, 빅데이터 기술개발</li> </ul>
한국과학기술원	데이터베이스연구소 첨단 데이터베이스 엔진 및 응용연구실	<ul style="list-style-type: none"> <li>• 라이프로그 관리</li> <li>• 그래프데이터 관리 및 마이닝</li> <li>• 무선센서 네트워크에서 쿼리처리</li> <li>• 데이터 스트림 처리 및 시멘틱 웹데이터 처리</li> </ul>

### (2) 기관 기술개발 동향

한림대학교

- 웹(Web)상의 데이터를 크롤링하여 수집하고 데이터 분석에 사용될 기반 데이터 셋을 형성하는 장치 등 비정형 데이터 수집과 데이터 통합에 대한 연구수행

한국전자기술연구원

- 비정형 빅데이터를 기준처리하여 정형화하고 융합가공할 수 있는 데이터셋을 구축하는 연구수행

한국과학기술정보연구원

- 대용량 데이터 플랫폼 개발, 수요탐색 인텔리전스 연구를 수행하며, 데이터 수집과 분류, 가공을 통한 플랫폼 구축과 융합분석 활용주제 공유 확산체제 구축연구 진행

한국과학기술원

- 시멘틱 웹 데이터, 무선센서등 IoT데이터, 라이프 로그, 그래프데이터 등 다양한 데이터 처리와 마이닝에 대한 기술 연구

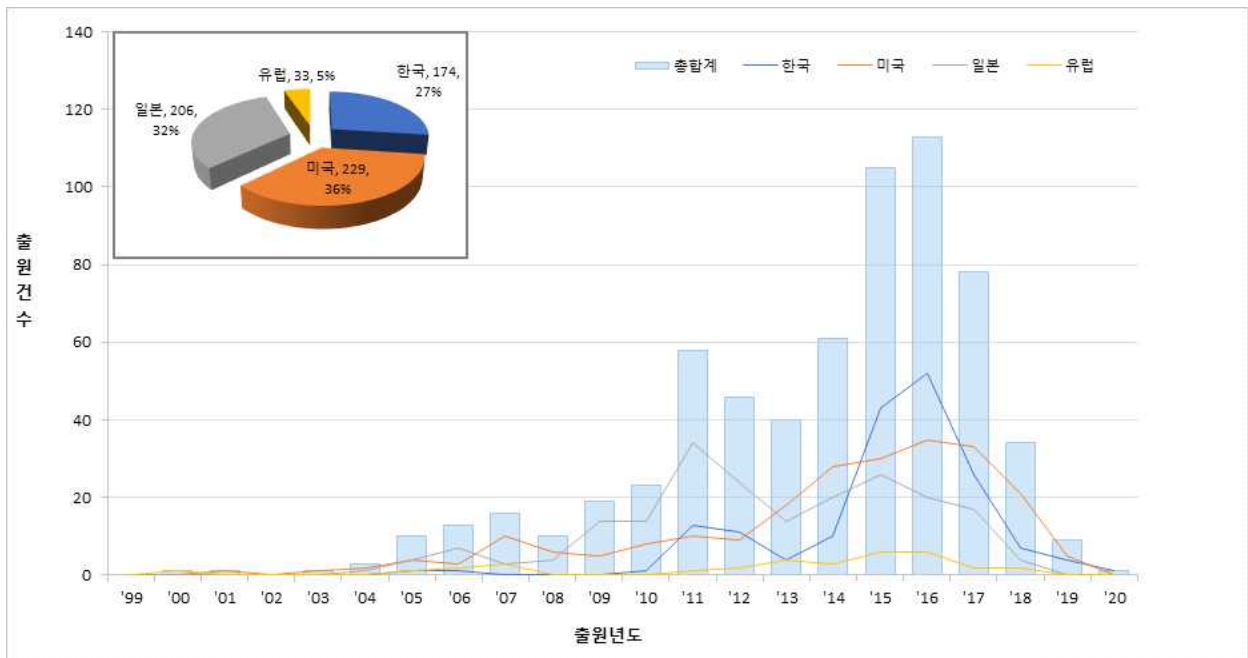
## 4. 특허 동향

### 가. 특허동향 분석

#### (1) 연도별 출원동향

- 공공 빅데이터 수집 및 분석 시스템은 '11년부터 급격한 성장을 보임
  - 각 국가별로 살펴보면 미국이 가장 활발한 출원활동을 보이고 있음
- 국가별 출원비중을 살펴보면 미국이 전체의 36%의 출원 비중을 차지하고 있어, 최대 출원국으로 공공 빅데이터 수집 및 분석 시스템 분야를 리드하고 있는 것으로 나타났으며, 일본은 32%, 한국은 27%, 유럽은 5% 순으로 나타남

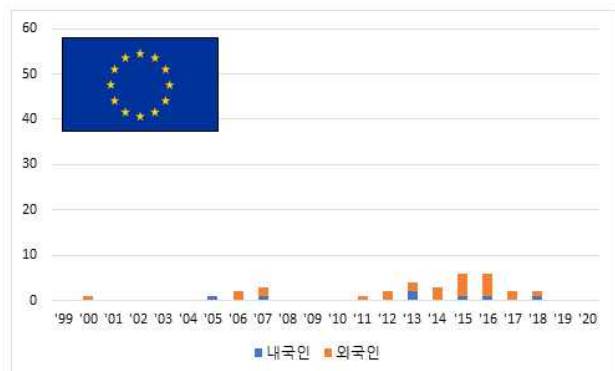
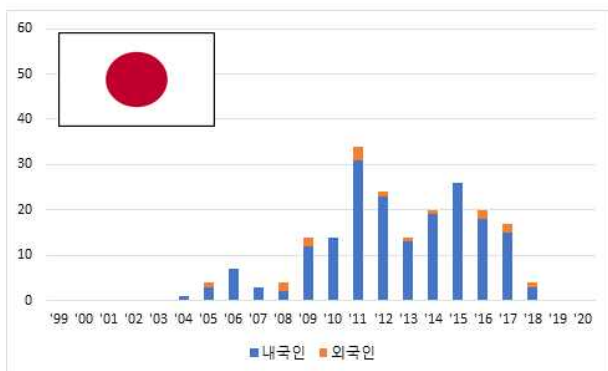
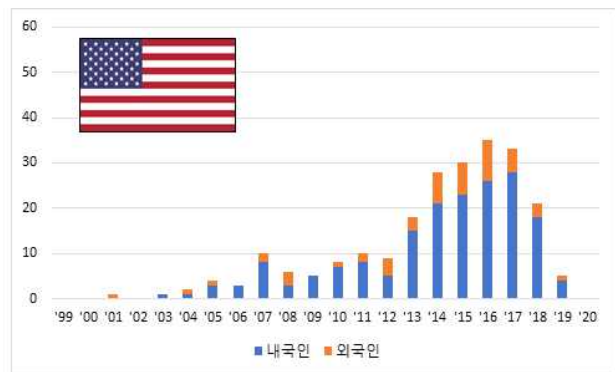
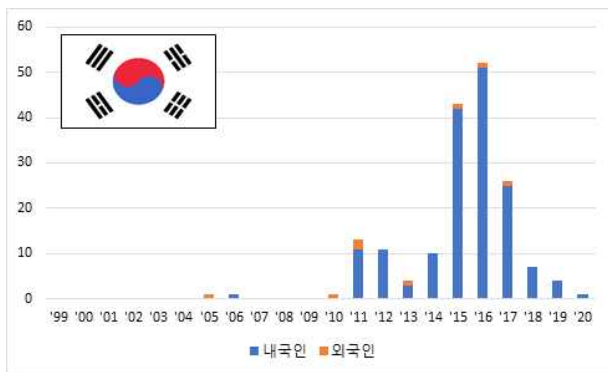
[ 공공 빅데이터 수집 및 분석 시스템 연도별 출원동향 ]



## (2) 국가별 출원현황

- 한국의 출원현황을 살펴보면, '15년부터 해당 기술의 출원이 급격히 증가하는 추세
  - 내국인 위주의 출원이 진행되고 있음
  - 한국 기술의 양적 흐름은 미국과 상당히 유사
  - 미국의 출원 수에 비해 76% 정도의 수준을 보임
- 미국의 출원현황을 살펴보면 분석구간 초기부터 전체 특허기술의 출원 증감 흐름에 영향을 주고 있는 것으로 나타남. 미국의 경우, 한국에 비해 외국인의 비중이 큰 것으로 나타남
- 유럽의 출원현황은 출원수가 매년 10건 이하로, 뚜렷한 증감 동향이 나타나지 않음. 해당 기술 분야에서 일본 시장에 대한 관심도가 높지 않은 것으로 보임
- 일본의 출원현황을 살펴보면, '11년에 일시적으로 급격히 증가하다가 이후 완만한 감소하는 추세

[ 국가별 출원현황 ]



### (3) 기술 집중도 분석

□ 전략제품에 대한 최근 기술 집중도 분석을 위한 구간별 기술 키워드 분석 진행

- 전체 구간(1999년~2020년)에서 정보 처리 장치, 사용자 단말, User Interface 등 키워드가 다수 도출
- 최근 구간 분석 결과, 최근 1구간(2012년~2015년)과 비교할 때 2구간(2016년~2020년)에서 사용자 단말, Data Source, 정보 관리 서버 키워드가 많이 등장하는 것으로 보아 공공 빅데이터 수집 및 분석 시스템 분야에도 데이터 유형과 형태에 따른 수집과 분석 방법 관련 연구개발이 활발한 것으로 추정

#### [ 특히 키워드 변화로 본 기술개발 동향 변화 ]

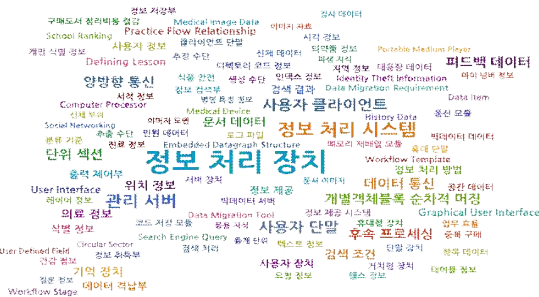
##### 전체구간(1999년~2020년)



- 정보 처리 장치, 사용자 단말, User Interface, 의료 정보, 정보 처리 시스템, 검색 결과, 검색 조건, 관리 서버, 식별 정보, 사용자 단말기

##### 최근구간(2012년~2020년)

###### 1구간(2012년~2015년)



- 정보 처리 장치, 정보 처리 시스템, 관리 서버, 사용자 클라이언트, 사용자 단말, 단위 섹션, 데이터 통신, 후속 프로세싱, 양방향 통신, 피드백 데이터

###### 2구간(2016년~2020년)

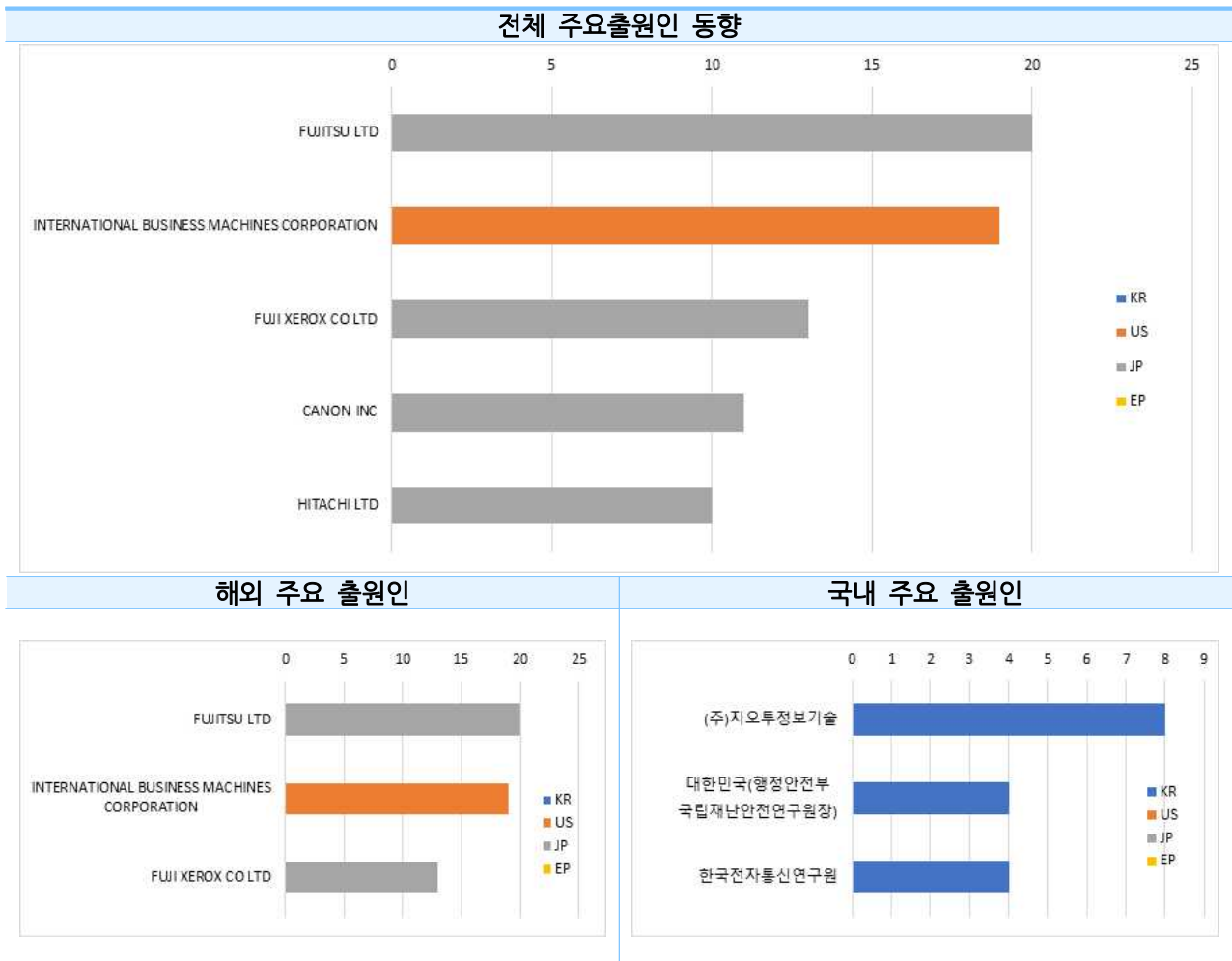


- 정보 처리 장치, 사용자 단말기, User Interface, Data Source, 사용자 단말, 컴퓨터 프로그램, 정보 관리 서버, 건강 데이터, Computing Device, Data Processing

## 나. 주요 출원인 분석

- 공공 빅데이터 수집 및 분석 시스템의 전체 주요출원인을 살펴보면, 주로 일본 국적의 출원인이 다수 포함되어 있는 것으로 나타났으며, 제 1 출원인으로는 일본의 FUJITSU LTD인 것으로 나타남
  - 제 1 출원인인 FUJITSU LTD의 출원은 일본에 집중된 경향을 보임
- 공공 빅데이터 수집 및 분석 시스템 관련 기술로 정보통신기술을 다루는 대기업에 의한 출원이 대다수를 차지
  - 국내에서는 중소기업(개인), 연구기관/대학의 활발한 출원이 이루어짐

[ 공공 빅데이터 수집 및 분석 시스템 주요출원인 ]



(1) 해외 주요출원인 주요 특허 분석

◎ FUJITSU LTD

- FUJITSU LTD는 일본의 정보통신기술 기업으로, 공공 빅데이터 수집 및 분석 시스템과 관련하여, 공공의 DB를 활용하여 데이터를 수집 및 분석하는 기술에 특화된 특허를 다수 출원. 그 중 등록된 특허는 4건 보유
  - 주요 특허들은 공공의 DB에서 공공의 데이터를 보다 효과적이고 신속하게 정보를 제공하기 위한 기술 특허를 다수 출원하는 것으로 파악

[ FUJITSU LTD 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
JP6102434 (2013.03.29)	서적정보 관리장치, 서적 정보 관리 프로그램 및 서적 정보관리 방법	기간마다 키워드를 기억한 계절 DB와 이용자가 과거에 등록한 전자 서적에 관한 서적 정보를 기억한 개인 책장 DB를 활용하여 시기에 따른 서적을 정리하여 이용자에게 제공하는 기술	
JP5958198 (2012.08.31)	프로그램, 정보처리 장치, 및 처리 방법	주민기록 시스템으로부터 취득한 어떤 세대의 구성의 변경을 나타내는 구성 변경 정보가, 국민건강보험 시스템에 등록 가능한 조건을 만족하는 구성 변경 패턴에 대응하는지 여부를 판정하는 기술	
JP5668619 (2011.06.29)	정보 제공 장치, 정보 제공 프로그램 및 정보 제공 방법	검색부는 이용자가 조작하는 단말로부터 이용한 요구를 받아들이고, 출력부는 통지 정보가 존재하는 취지를 나타내는 정보를 이용자가 조작하는 단말로 출력하는 기술	
JP5278177 (2009.06.08)	특허 정보 구축 장치, 특허 정보 구축 프로그램, 및 특허 정보 구축 방법	특허문헌의 명세서 문서 데이터로부터, 특허문헌의 명칭 및 인용 관계 및 인용처 문헌의 내용 설명 및 미해결 과제를 추출하고, 특허문헌으로 대응되어 과제 관계성 데이터베이스에 축적하는 기술	

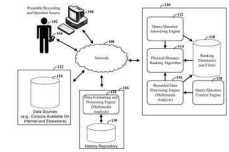
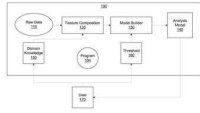
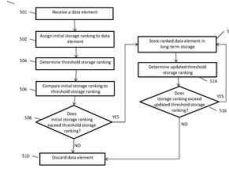
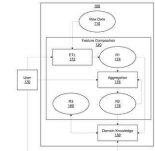
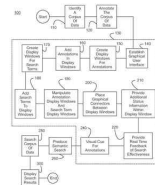
\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출



◎ IBM(International Business Machines)

- IBM은 미국의 다국적 기업으로, 다수의 데이터 모델 생성 관련 특허를 보유하고 있으며, 공공 빅데이터 수집 및 분석 시스템 기술과 관련한 특허를 미국에 출원
  - 공공 빅데이터 수집 및 분석 시스템에 관련하여 출원을 진행한 특허 모두 미국에 출원한 것으로, 자국 내 출원 성향이 높은 것으로 사료
  - 데이터의 코퍼스(corpus)를 활용하여 데이터를 분석하는 기술과 관련하여 19건의 미국 출원을 진행하였으며, 그 중 등록된 특허는 12건으로 파악

[ IBM 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
US10042892 (2017.08.16)	Question answer system using physical distance data	일련의 답들은 데이터의 코퍼스(corpus)로부터 질문에 대해 생성될 수 있고, 식별된 물리적 객체들 및 개체와 관련 물리적 객체들 사이의 물리적 거리에 기초하여 등급(rank)을 매기는 기술	
US10394931 (2016.01.13)	Balancing provenance and accuracy tradeoffs in data modeling	모델의 정확성과 확률의 트레이드오프(tradeoff)에 기초하여 데이터 모델 생성하는 기술	
US10395331 (2015.12.04)	Selective retention of forensic information	랭크된 데이터 엘리먼트를 생성하기 위해 모니터링된 데이터 엘리먼트에 초기 스토리지 랭킹을 할당하는 기술	
US9275425 (2013.12.19)	Balancing provenance and accuracy tradeoffs in data modeling	원시 데이터 세트를 수신하고, 3개의 데이터 세트를 기초하여 3개의 리포지토리를 생성하는 데이터 모델을 생성하는 기술	
US8972440 (2013.06.25)	Method and process for semantic or faceted search over unstructured and annotated data	데이터의 코퍼스 내에 포함된 복수의 데이터 요소와 연관된 주석의 그래픽 표현 및 복수의 데이터 요소 내에 포함된 검색어의 그래픽 표현의 집합을 생성하기 위해 그래픽 사용자 인터페이스를 사용한 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

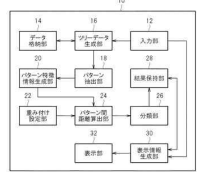
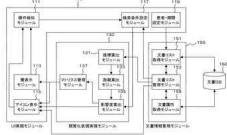
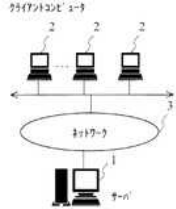


◎ FUJI XEROX CO LTD

□ FUJI XEROX CO LTD는 일본의 다국적 기업으로, 정보처리 장치 및 정보처리 프로그램 관련하여 등록특허 3건 보유

- 공공 빅데이터 수집 및 분석 시스템에서 데이터베이스 내 정보처리 관련 응용 특허를 다수 보유

[ FUJI XEROX CO LTD 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
JP5471372 (2009.12.01)	프로그램 및 정보처리 시스템	트리 구조의 정보군으로부터 부분 트리의 복수의 패턴을 추출하고, 추출된 각 패턴의 특징 정보를 생성하고, 생성된 각 패턴의 특징 정보에 기초하여, 각 패턴을 분류하고, 그 분류 결과에 따라 각 패턴을 표시하는 표시 정보를 생성하는 기술	
JP5347788 (2009.07.14)	정보처리 장치 및 정보처리 프로그램	표 내의 셀로 문서를 나타내는 정보를 표시하는 경우에, 그 문서를 배치하는 위치에 따른 셀의 크기로 할 수 있는 정보처리 장치의 문서 정보 추출 수단 기술	
JP4428110 (2004.03.26)	경험 지식정보처리 장치	경험 지식 데이터베이스에 축적된 정보에 기초한 경험 지식 처리부의 처리에 의해, 활동 및 지식 창조 프로세스를 2축으로 한 지식 총수의 이원표나, 지식의 전파 상황 그래프 등이 표시 출력 기술	


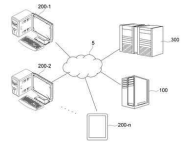
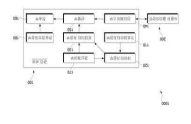

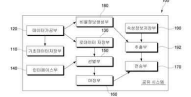
\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

## (2) 국내 주요출원인 주요 특허 분석

### ◎ (주)지오투정보기술

- (주)지오투정보기술은 한국의 지리정보 빅데이터 분석 기업으로, 공공 빅데이터 수집 및 분석 시스템에 있어서, 도시 계획 정보 분석 기술과 관련된 특허를 8건 출원하였으며, 출원을 진행한 8건의 특허 모두 등록된 것으로 파악

[ (주)지오투정보기술 주요특허 리스트 ]

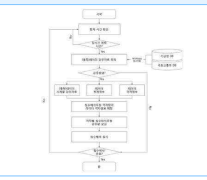

등록번호 (출원일)	명칭	기술적용분야	도면
KR1621428 (2015.11.05)	순차적 머징을 수행하는 토지적성평가용 도시계획정보 표출시스템	관리 서버 및 사용자 클라이언트로 구성된 순차적 머징을 수행하는 토지적성평가용 도시계획정보 표출시스템 기술	
KR1632066 (2015.11.05)	동일 주제도 데이터의 공통 아이디 정보를 이용한 토지적성평가용 도시계획정보 표출시스템	방대한 용량으로 이루어지는 속성정보, 위치정보 및 주제도 데이터 등의 도시계획정보를 단위 섹션별로 개별화시키고 이를 기준으로 데이터 전송, 이미지 머징 등의 후속 프로세싱을 처리하는 기술	
KR1621430 (2015.11.05)	공통 아이디 정보의 공유 적용을 통한 토지적성평가용 도시계획정보의 시계열 표출시스템	데이터 통신의 효율성을 극대화시킬 수 있고, 양방향 피드백을 통하여 더욱 사용자 지향적 환경으로 도시계획정보를 표출할 수 있는 시스템 기술	
KR1621427 (2015.11.05)	순차적 전송을 이용한 토지적성평가를 위한 도시계획정보 표출시스템	순차적 전송을 이용한 토지적성평가를 위한 도시계획정보 표출시스템 기술	
KR1632065 (2015.08.19)	데이터 분산처리 기반에 의한 도시계획정보의 주제도 공유 시스템	도시계획정보 내지 도시계획정보체계의 주제도 등을 사용자 클라이언트와 효과적으로 공유하는 시스템 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ 대한민국(행정안전부 국립재난안전연구원장)

□ 대한민국(행정안전부 국립재난안전연구원장)은 공공 빅데이터 수집 및 분석 시스템 기술과 관련된 특허를 4건 출원하였으며, 4건의 특허 중 등록된 특허는 2건을 보유

[ 대한민국(행정안전부 국립재난안전연구원장) 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
KR1670903 (2015.10.29)	레이더 강우자료를 이용하여 지표유출을 해석하는 방법	이동성 평면 레이더 강우자료의 좌표와 지표유출 해석자료의 좌표를 고려하여 시간별 레이더 강우자료를 지표유출 해석시에 입력 강우자료로 변환 활용하는 기술	
KR1685335 (2015.05.12)	빅데이터 의미 기반 재난 감성분류방법	재난 유형별 발생 과정 또는 대응 방안에 따라 사람들이 느끼는 감정이 어떻게 변화하는지 모니터링할 수 있는 재난 감성분류 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

◎ 한국전자통신연구원

- 한국전자통신연구원은 한국의 정보, 통신, 전자, 방송 및 성과 관련 융·복합기술 분야를 연구하는 연구 기관으로, 공공 빅데이터 수집 및 분석 시스템과 관련된 특허를 4건 출원하였으며, 등록특허 1건 보유

[ 한국전자통신연구원 주요특허 리스트 ]

등록번호 (출원일)	명칭	기술적용분야	도면
KR1903522 (2015.11.25)	다차원 건강데이터에 대한 유사사례 검색 방법 및 그 장치	다차원 건강데이터에 대한 검색을 위한 계산 복잡도를 줄이고, 효율적인 유사사례 검색이 가능한 기술	

\* 등록특허 기준, 피인용문헌수 및 패밀리 국가수가 큰 특허를 주요특허로 도출

## 다. 기술진입장벽 분석

### (1) 기술 집중력 분석

- 공공 빅데이터 수집 및 분석 시스템 관련 기술에 대한 시장관점의 기술독점 현황분석을 위해 집중률 지수(CRn: Concentration Ratio n, 상위 n개사 특허점유율의 합) 분석 진행
  - 상위 4개 기업의 시장점유율이 0.10로 공공 빅데이터 수집 및 분석 시스템 분야에 있어서 독과점 정도는 매우 낮은 수준으로 판단
  - 국내 시장에서 중소기업의 점유율 분석결과 0.7로 해당 기술에 대하여 중소기업의 진입이 용이하다고 판단됨

[ 주요출원인의 집중력 및 국내시장 중소기업 집중력 분석 ]

주요 출원인 집중력	주요출원인	출원건수	특허점유율	CRn	n
	FUJITSU LTD(일본)	20	3.1%	0.03	1
	INTERNATIONAL BUSINESS MACHINES CORPORATION(미국)	19	3.0%	0.06	2
	FUJI XEROX CO LTD(일본)	13	2.0%	0.08	3
	CANON INC(일본)	11	1.7%	<b>0.10</b>	<b>4</b>
	HITACHI LTD(일본)	10	1.6%	0.11	5
	RICOH CO LTD(일본)	9	1.4%	0.13	6
	(주)지오투정보기술(한국)	8	1.2%	0.14	7
	TOSHIBA CORP(일본)	7	1.1%	0.15	8
	NEC CORPORATION(일본)	6	0.9%	0.16	9
	KONICA MINOLTA MEDICAL & GRAPHIC INC(일본)	6	0.9%	0.17	10
	<b>전체</b>	<b>642</b>	<b>100%</b>	<b>CR4=0.10</b>	
국내시장 중소기업 집중력	출원인 구분	출원건수	특허점유율	CRn	n
	중소기업(개인)	117	69.6%	0.70	
	대기업	11	6.5%		
	연구기관/대학	40	23.8%		
	<b>전체</b>	<b>168</b>	<b>100%</b>	<b>CR중소기업=0.70</b>	

(2) 특허소송 현황 분석

- 공공 빅데이터 수집 및 분석 시스템 관련 기술 진입 장벽에 대한 분석을 위해 특허소송을 이력 검토
  - 2020년 1월 뉴욕 남부지방법원에 원고 BCM International Regulatory Analytics LLC와 Arker Inc.간의 데이터베이스를 위한 컴퓨터 프로그램 제품에 대한 특허 침해소송이 종결
  - 2019년 7월 델라웨어 지방방법원에 원고 West View Research, LLC와 BMW Manufacturing Co. LLC간의 컴퓨터화 정보 및 표시 장치에 대한 특허 침해소송이 종결

[ 공공 빅데이터 수집 및 분석 시스템 관련 특허소송 현황 ]

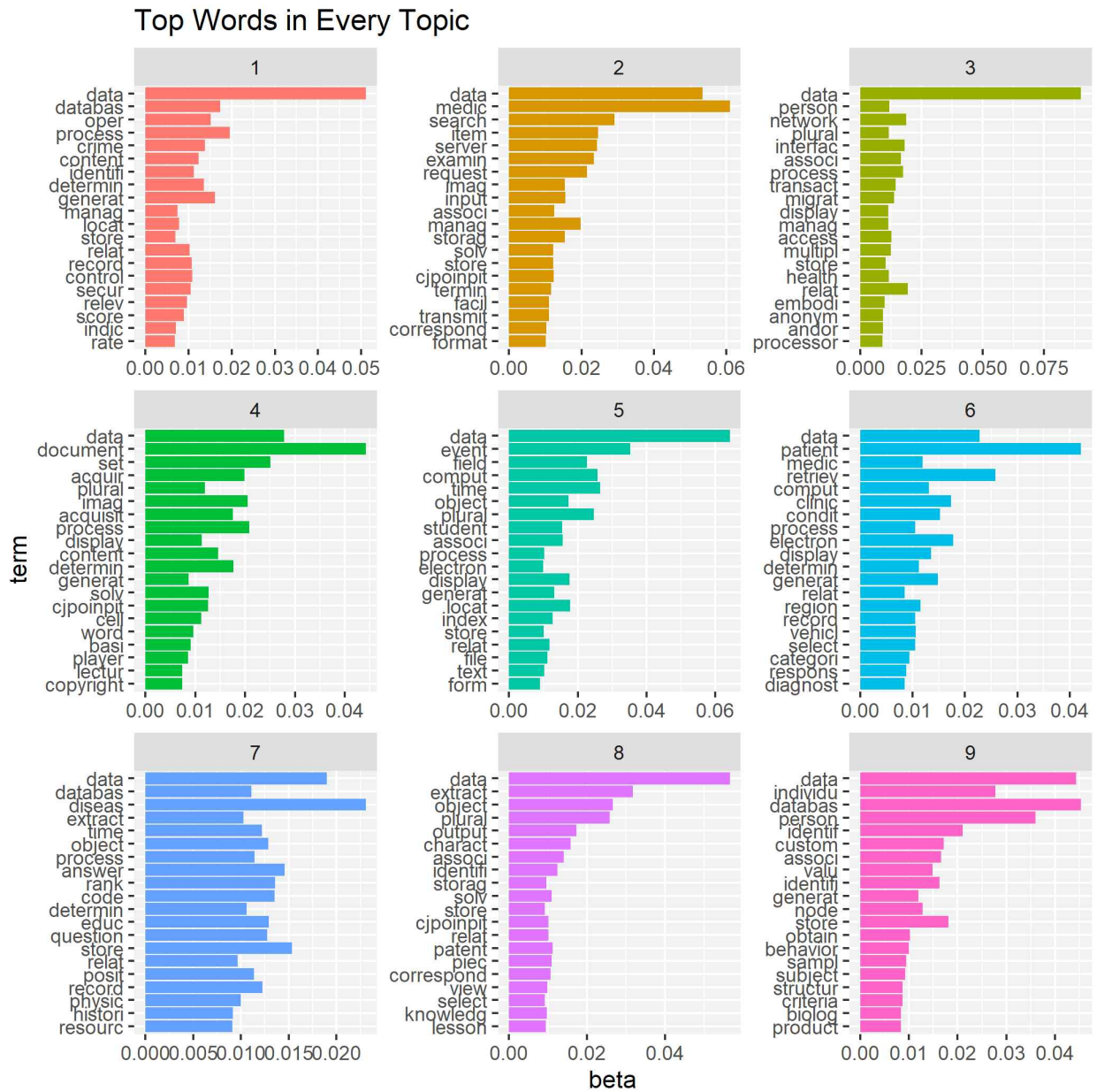
1	US9436726 (2016.09.06)	<b>명칭</b>	<b>출원인</b>	<b>원고 v. 피고</b>
		System, method and computer program product for a behavioral database providing quantitative analysis of cross border policy process and related search capabilities	BCM International Regulatory Analytics LLC	BCM International Regulatory Analytics LLC v. Arker Inc.
		<b>대상제품명</b>	<b>소제기일</b>	<b>소송종료일</b>
		Arker Product	2019.11.16	2020.01.04
2	US9412367 (2016.08.09.)	<b>명칭</b>	<b>출원인</b>	<b>원고 v. 피고</b>
		Computerized information and display apparatus	West View Research, LLC	West View Research, LLC v. BMW of North America LLC
		<b>대상제품명</b>	<b>소제기일</b>	<b>소송종료일</b>
		BMW ConnectedDrive, BMW Assist and BMW Online - BMW iDrive Products	2018.02.06	2019.07.18

## 5. 요소기술 도출

### 가. 특허 기반 토픽 도출

- 404개의 특허의 내용을 분석하여 구성 성분이 유사한 것끼리 클러스터링을 시도하여 대표성이 있는 토픽을 도출

[ 공공 빅데이터 수집 및 분석 시스템에 대한 토픽 클러스터링 결과 ]



## 나. LDA<sup>22)</sup> 클러스터링 기반 요소기술 도출

[ LDA 클러스터링 기반 요소기술 키워드 도출 ]

No.	상위 키워드	대표적 관련 특허	요소기술 후보
클러스터 01	data, process, database, generate, oper, crime, determine, content, identification, control	<ul style="list-style-type: none"> <li>Information management and assessment system for political and other types of candidates</li> <li>Providing controlled access to proprietary fisheries data</li> </ul>	빅데이터 수집 및 분석 플랫폼
클러스터 02	medicine, data, search, item, server, examine, request, manage, input, storage	<ul style="list-style-type: none"> <li>Medical information search system, medical information search gateway device, medical information search method, and program</li> <li>Entrance examination information retrieval device, system, and program</li> </ul>	데이터 유형과 형태에 따른 검색기 및 크롤러
클러스터 03	data, relate, network, interface, process, associate, transact, migrat, access, multipl	<ul style="list-style-type: none"> <li>Enterprise migration planning information repository</li> <li>System and method for handling data captured by a body worn camera</li> </ul>	-
클러스터 04	data, node, patent, extract, education, examine, medicine, field, migrate, disease	<ul style="list-style-type: none"> <li>Information processing apparatus, information processing method, and program</li> <li>Expert evaluation information management system</li> </ul>	정형/비정형 융합 정제 및 가공 기술
클러스터 05	data, event, time, computing, plural, field, location, display, object, associate	<ul style="list-style-type: none"> <li>Systems and methodologies providing controlled voice and data communication among a plurality of computing appliances associated as team members of at least one respective team or of a plurality of teams and sub-teams within the teams</li> <li>Systems and methods for creating a form for receiving data relating to a health care incident</li> </ul>	통합데이터 품질표준 기술
클러스터 06	patient, retrieve, data, electron, clinic, condition, generate, display, computing, medicine	<ul style="list-style-type: none"> <li>Apparatus, method, and program for clinical information processing</li> <li>Information processing apparatus, information processing method, and program</li> </ul>	데이터 유형과 형태에 따른 검색기 및 크롤러
클러스터 07	disease, data, store, answer, rank, code, education, object	<ul style="list-style-type: none"> <li>System and method for producing evaluation data</li> <li>Information processing device and program</li> </ul>	빅데이터 수집 및 분석 플랫폼
클러스터 08	data, extract, object, plural, output, character, associate, identification, patent, piece	<ul style="list-style-type: none"> <li>Self-construction of content in adaptive e-learning datagraph structures</li> <li>Dynamic contribution accounting in adaptive e-learning datagraph structures</li> </ul>	빅데이터 연계 통합 기술
클러스터 09	database, data, person, individual, identification, store, custom, associate, identification, value	<ul style="list-style-type: none"> <li>Method of generating, storing and mining data related to key opinion leaders in scientific fields and computer system configured for presenting an explorable graphical user interface</li> <li>Phenotypic integrated social search database and method</li> </ul>	데이터 유형과 형태에 따른 검색기 및 크롤러

22) Latent Dirichlet Allocation



### 다. 특허 분류체계 기반 요소기술 도출

- 공공 빅데이터 수집 및 분석 시스템 관련 특허에서 총 10개의 주요 IPC코드(메인그룹)를 산출하였으며, 각 그룹의 정의를 기반으로 요소기술 키워드를 아래와 같이 도출

[ IPC 분류체계에 기반한 요소기술 도출 ]

IPC 기술트리		
(서브클래스) 내용	(메인그룹) 내용	요소기술 후보
(G06F) 전기에 의한 디지털 데이터처리	• (G06F-021) 부정행위로부터 프로그램 또는 데이터, 그 컴퓨터 부품을 보호하기 위한 보안 장치	-
	• (G06F-017) 디지털 컴퓨팅 또는 데이터 프로세싱 장비, 방법으로서 특정 기능을 위해 특히 적합한 형태의 것	-
	• (G06F-019) 특수한 어플리케이션에 특히 적합한 디지털 컴퓨팅 또는 데이터 처리 장치 또는 방법	-
	• (G06F-007) 취급하는 데이터의 순서 또는 내용을 조작하여 데이터를 처리하기 위한 방법 또는 장치	데이터 유형과 형태에 따른 검색기 및 크롤러
(G06Q) 관리용, 상업용, 금융용, 경영용, 감독용 또는 예측용으로 특히 적합한 데이터 처리 시스템 또는 방법; 그 밖에 분류되지 않는 관리용, 상업용, 금융용, 경영용, 감독용 또는 예측용으로 특히 적합한 시스템 또는 방법	• (G06Q-050) 특정 사업 부문에 특히 적합한 시스템 또는 방법, 예. 공익사업 또는 관광	-
	• (G06Q-040) 금융; 보험; 세무 전략; 법인세 또는 소득세의 처리	정형/비정형 융합 정제 및 가공 기술
	• (G06Q-030) 거래, 예. 쇼핑 또는 전자상거래	빅데이터 수집 및 분석 플랫폼
	• (G06Q-090) 특별한 데이터 처리를 수반하지 않는 관리용, 상업용, 금융용, 경영용, 감독용 또는 예측용으로 특히 적합한 데이터 처리 시스템 또는 방법	-
	• (G06Q-010) 경영; 관리	-
(H04L) 디지털 정보의 전송, 예. 전신통신	• (H04L-029) 그룹 1/00에서 H04L 27/00의 하나에도 포함되지 않는 배치, 장치회로 또는 시스템	-

## 라. 최종 요소기술 도출

- 산업·시장 분석, 기술(특허)분석, 전문가 의견, 타부처 로드맵, 중소기업 기술수요를 바탕으로 로드맵 기획을 위하여 요소기술 도출
- 요소기술을 대상으로 전문가를 통해 기술의 범위, 요소기술 간 중복성 등을 조정·검토하여 최종 요소기술명 확정

### [ 공공 빅데이터 수집 및 분석 시스템 분야 요소기술 도출 ]

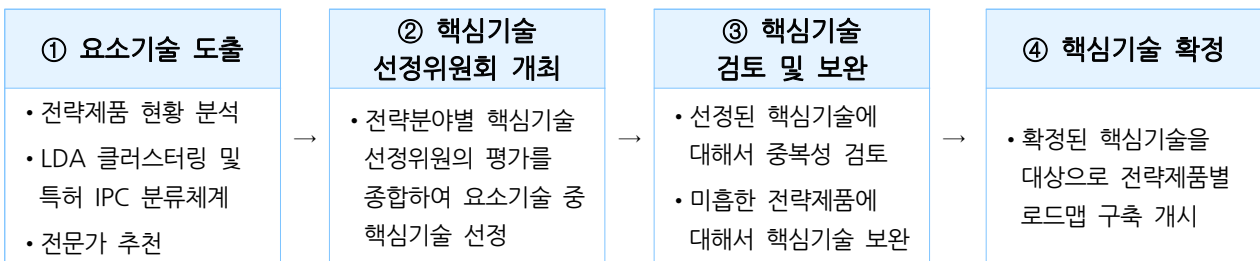
요소기술	출처
데이터 유형과 형태에 따른 검색기 및 크롤러	특허 클러스터링, IPC 기술체계, 전문가추천
빅데이터 연계 통합 기술	특허 클러스터링, 전문가추천
정형/비정형 융합 정제 및 가공 기술	특허 클러스터링, IPC 기술체계, 전문가추천
정형/비정형 융합분석 기술	특허 클러스터링, 전문가추천
빅데이터 수집 및 분석 플랫폼	특허 클러스터링, IPC 기술체계, 전문가추천
통합데이터 품질표준 기술	전문가추천

## 6. 전략제품 기술로드맵

### 가. 핵심기술 선정 절차

- 특허 분석을 통한 요소기술과 기술수요와 각종 문헌을 기반으로 한 요소기술, 전문가 추천 요소기술을 종합하여 요소기술을 도출한 후, 핵심기술 선정위원회의 평가과정 및 검토/보완을 거쳐 핵심기술 확정
- 핵심기술 선정 지표: 기술개발 시급성, 기술개발 파급성, 기술의 중요성 및 중소기업 적합성
  - 장기로드맵 전략제품의 경우, 기술개발 파급성 지표를 중장기 기술개발 파급성으로 대체
- 공공 빅데이터 수집 및 분석 시스템
  - 기관별 부처별 또는 민간에 산재해 있는 다양한 종류와 유형의 공공 빅데이터를 검색하고 추출하여 융합분석이 가능하도록 시각화 및 데이터 가공할 수 있는 시스템 개발
  - 공공 빅데이터의 데이터품질을 극대화 하여 분석에 활용할 수 있도록 공공데이터 가치를 재조정할 수 있는 기술개발

#### [ 핵심기술 선정 프로세스 ]



### 나. 핵심기술 리스트

#### [ 공공 빅데이터 수집 및 분석 시스템 분야 핵심기술 ]

핵심기술	개요
정형/비정형 융합분석기술*	• 다양한 유형의 공공 빅데이터를 분석 가능한 형태의 정량화 및 표준화하고 융합 및 연계 가능하도록 하는 기술
데이터 유형과 형태에 따른 검색기 및 크롤러	• 숫자, 문서, 영상, 음성 등 다양한 유형의 공공데이터를 탐색하고 수집할 수 있는 기술
통합데이터 품질 표준 기술*	• 다양한 유형의 데이터를 통합하는 과정에서 발생하는 데이터 품질 문제를 해결하고 보정하는 표준기술
빅데이터 연계 통합 기술	• 분석이 가능하도록 수집된 공공데이터의 연계 방법 추천 등 연계통합 자동화 기술
정형/비정형 융합 정제 및 가공 기술	• 통합데이터의 탐색적 분석과 관리 및 분석데이터로서 가치를 확보할 수 있는 정제와 가공 기술

\* 표시는 생태계 취약 기술을 의미

### 다. 중소기업 기술개발 전략

- 공공 빅데이터에 대한 부처 및 기관별로 상이한 데이터형태의 연계수집 및 통합관리 해결
- 기존에 구축된 인프라와 개발된 기술을 활용하여 진입장벽을 해소하고 융합분석에 특화
- 산-학 연계 및 테스트베드 확보를 통한 개발기술 검증으로 국내시장 개척 및 선점 필요

### 라. 기술개발 로드맵

#### (1) 중기 기술개발 로드맵

[ 공공 빅데이터 수집 및 분석 시스템 분야 중기 기술개발 로드맵 ]

공공 빅데이터 수집 및 분석 시스템	공공 빅데이터의 데이터품질을 극대화 하여 분석에 활용할 수 있도록 공공데이터 가치를 재조정할 수 있는 기술개발			
	2021년	2022년	2023년	최종 목표
정형/비정형 융합분석기술				비정형지수를 결합한 통합데이터 기술개발
데이터 유형과 형태에 따른 검색기 및 크롤러				다양한 유형의 공공데이터 수집 및 연계 기술개발
통합데이터 품질표준 기술				분석데이터 셋 구축을 위한 데이터 품질관리기술
빅데이터 연계 통합 기술				척도, 비율 등 분석유형에 따른 연계 추천기술 개발
정형/비정형 융합 정제 및 가공 기술				수집-정제-가공-시각화 자동화 기술 개발

(2) 기술개발 목표

□ 최종 중소기업 기술로드맵은 기술/시장 니즈, 연차별 개발계획, 최종목표 등을 제시함으로써 중소기업의 기술개발 방향성을 제시

[ 공공 빅데이터 수집 및 분석 시스템 분야 핵심기술 연구목표 ]

핵심기술	기술요구사항	연차별 개발목표			최종목표	연계R&D 유형
		1차년도	2차년도	3차년도		
정형/비정형 융합분석기술	정형/비정형 데이터 표준화기술	비정형 데이터의 지수 개발	정형과 비정형 데이터 매칭 기술 개발	데이터간 매칭 및 통합 기술 고도화	비정형지수를 결합한 통합데이터 기술개발	산학연
데이터 유형과 형태에 따른 검색기 및 크롤러	다양한 형태와 유형의 데이터 검색과 사이트 수집 기술	수집 사이트 탐색 및 수집정책 해결	수집 및 연계DB 구축	사이트별, 데이터별 상이한 크롤러 개발	다양한 유형의 공공데이터 수집 및 연계 기술개발	창업형
	수집속도(kb/sec)	6이상	7.5이상	8이상	8이상	창업형
통합데이터 품질표준 기술	데이터 품질관리 기술	정형/비정형 데이터 통합품질	통합데이터 품질개선 기술개발	분석데이터 가치 고도화	분석데이터 셋 구축을 위한 데이터 품질관리기술	산학연
빅데이터 연계 통합 기술	데이터 연계/통합추천 기술	수집데이터의 척도, 비율, 명목 등 변수화 구분 방법 개발	SI를 활용한 분석데이터 연계 통합 매칭기술 개발	연계통합 매칭 고도화	척도, 비율 등 분석유형에 따른 연계 추천기술 개발	산학연
정형/비정형 융합 정제 및 가공 기술	데이터 정제 및 가공, 시각화 기술	비정형데이터의 정제 자동화 개발	정제성능 고도화	정형/비정형 결합데이터의 시각화 기술 개발	수집-정제-가공-시각화 자동화 기술 개발	기술혁신

