



한국보건복지인력개발원
KOREA HUMAN RESOURCE DEVELOPMENT INSTITUTE
FOR HEALTH & WELFARE



국가인적자원개발컨소시엄
CHAMP Consortium for HRD Ability Magnified Program

의약품 빅데이터 분석 과정

- 3 차시 -

구글 사례 등 소개

3차시. 구글 사례 등 소개

· 학습목표

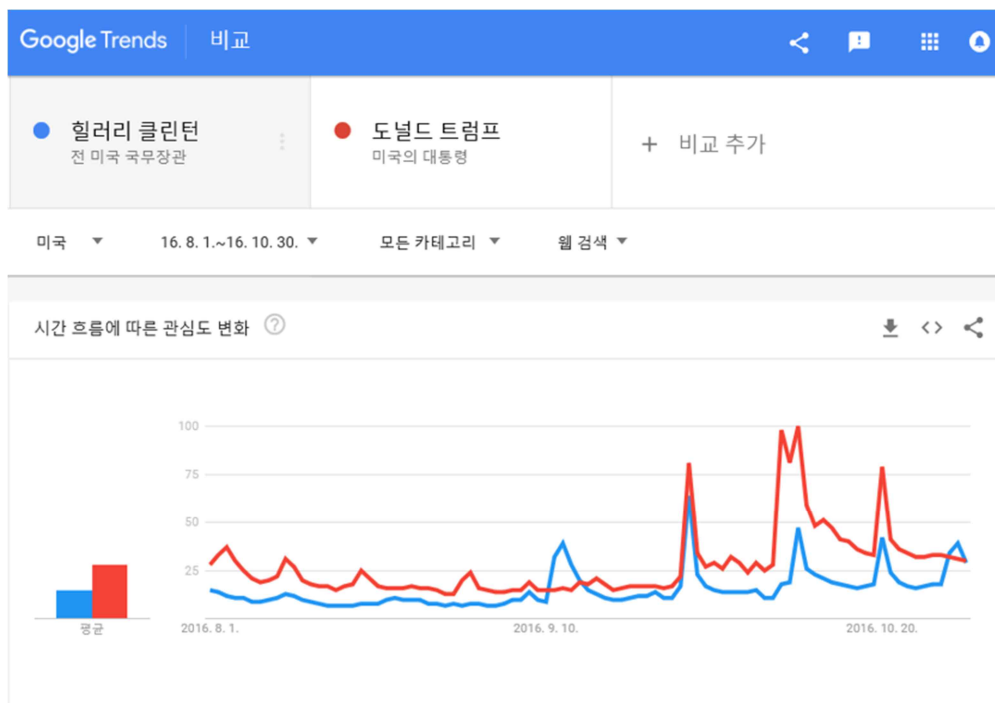
1. 구글 트렌드를 설명할 수 있다.
2. 빅데이터 활용 사례를 파악할 수 있다.

· 학습하기

1. 구글 트렌드

난 2016년 힐러리 클린턴과 도널드 트럼프의 미국 대선이 있었습니다.

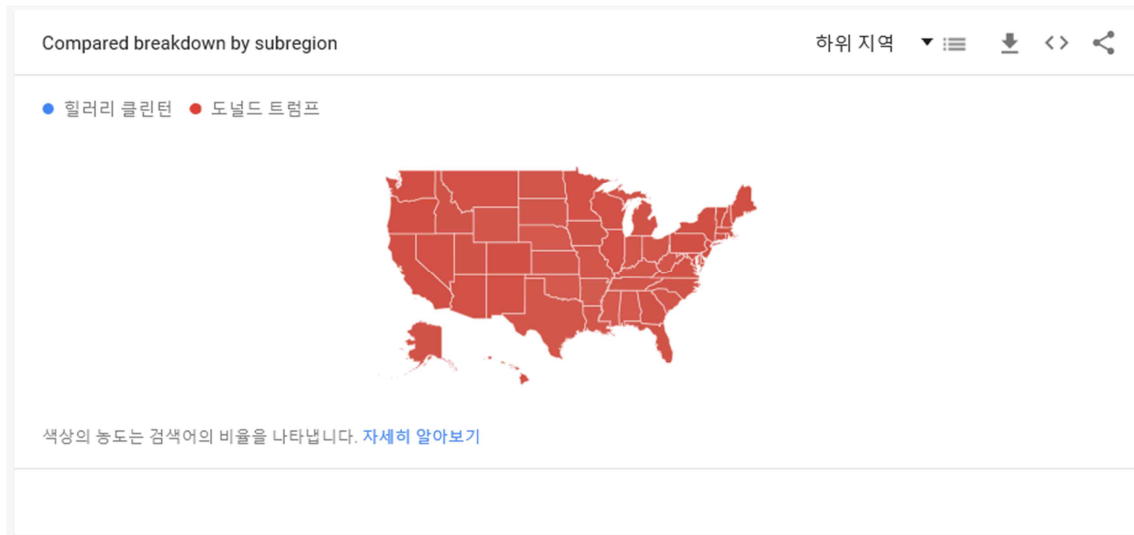
이때 미국의 대부분 여론조사 기관과 CNN 등의 뉴스에서는 힐러리의 낙승을 예측했습니다. 하지만 실제로는 우리가 다 알고 있는 것처럼 도널드 트럼프의 당선으로 결과가 달랐죠. 예상외의 결과에 여론 조사 기관들과 언론사들에서는 이러한 결과에 대한 논평 등을 쏟아냈습니다.



이 당시 이슈로 등장한 것이 바로 구글 트렌드입니다.

구글 트렌드를 이용하여 2016년 8월 1일부터 10월까지 구글 트렌드를 이용하여 검색 결과를 비교해봤습니다.

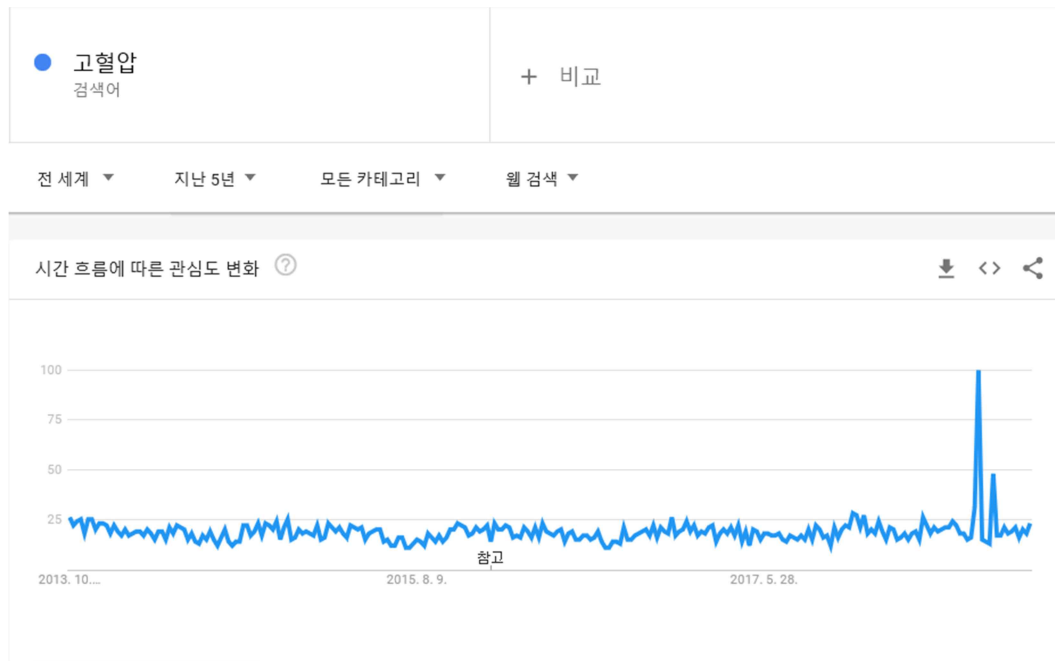
도널드 트럼프를 나타내는 빨간색 선은 힐러리 클린턴의 파란색보다 거의 대부분 위에 분포되어 있는 것으로 알 수 있습니다. 그리고 선거가 가까워질수록 검색 결과가 더 많아지는 상향 추이를 보입니다. 그래프에서 갑자기 솟구치는 구간은 TV 토론 등 정치적인 이슈가 있었던 때인 것으로 생각 되어집니다.



이번에는 미국의 각 주별로 검색어의 우선 순위를 확인해 봤습니다. 주별로 확인해 봐도 트럼프를 상징하는 빨간색이 모든 주에서 표시되어 있습니다. 즉, 미국의 모든 주에서 힐러리와 클린턴의 검색 순위에서도 트럼프가 우세하다는 것으로 알 수 있습니다.



대도시별로 세분화한 결과에서도 트럼프의 검색이 힐러리보다 더 많다는 것을 알 수 있습니다. 이상의 구글 트렌드로 일부 학자들 사이에서는 트럼프가 당선될 것이다. 라고 예측한 학자들도 소수가 있었습니다.



이번에는 실제 사례로 고혈압을 주제로 지난 5년간의 검색 관심도 추이를 살펴보겠습니다.

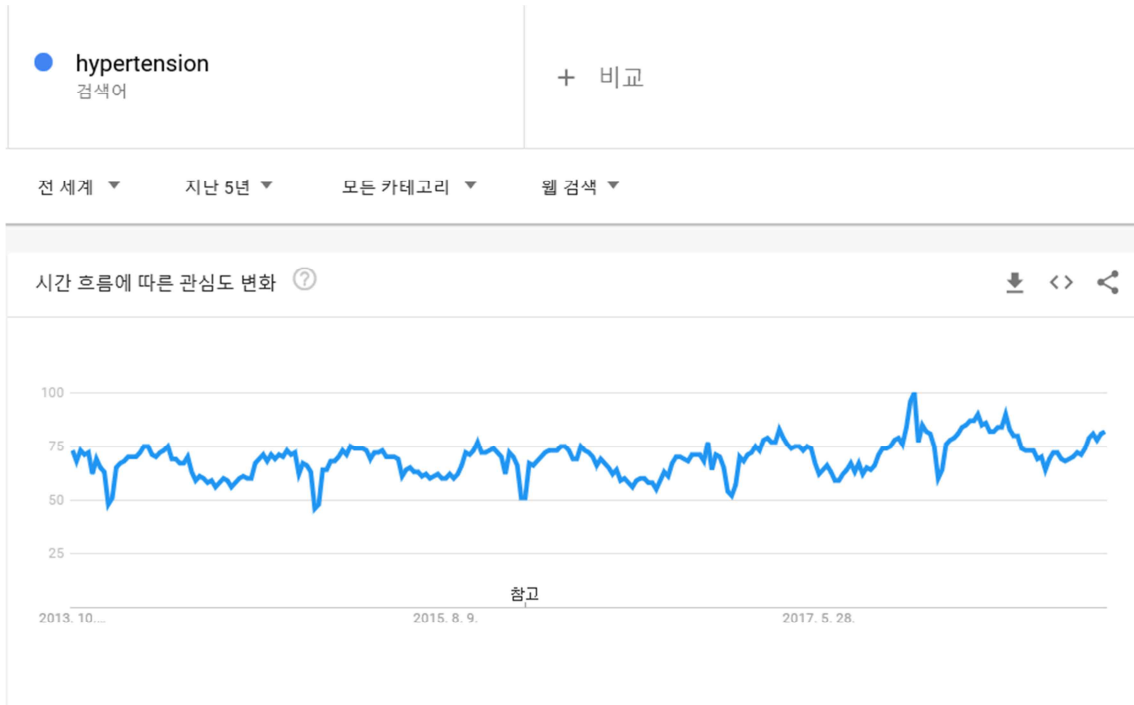
지난 5년간 고혈압에 대한 관심도는 일정하게 꾸준히 검색되는 것을 알 수 있습니다. 그런데 최근에 급격한 관심도가 높은 것으로 확인할 수 있습니다.

관련 주제 ?	급상승 ▼	↓	<>	↔	관련 검색어 ?	급상승 ▼	↓	<>	↔
1 발암 물질 - 주제	급등				1 고혈압 발암 물질	급등			
2 대한민국 식품의약품안전처 - 주제	급등				2 발암 고혈압 약	급등			
3 식품의약품안전처 - 대한민국 청원...	급등				3 고혈압 약 발암 물질	급등			
4 치매 - 주제	급등				4 식약 처 고혈압	급등			
5 건강보험 - 주제	급등				5 식약 처	급등			
< 25개 주제 중 1-5번 표시 중 >					< 25개 검색어 중 1-5번 표시 중 >				

관련 주제와 검색어를 확인해 본 결과 발암물질이 나오고 있습니다.

식품의약품안전처에서 지난 8월 5일에 고혈압약 원료에서 발암물질이 검출된 사건 때문에 이와 같은 관심도와 관련 검색어가 나온 것이죠.

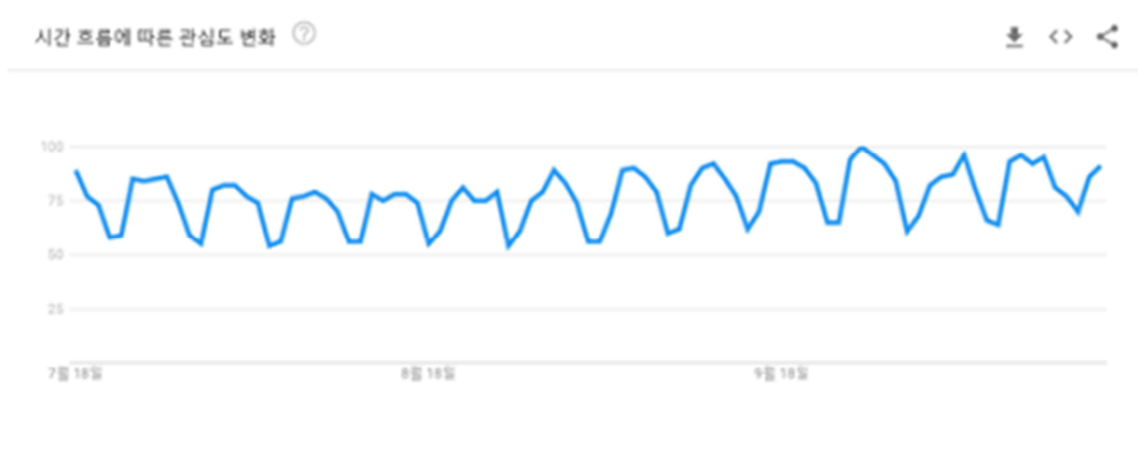
또한 전 세계의 지역별 관심도에서는 북미 지역과 한국, 일본, 호주 등에서 관심도가 높았다는 것으로 알 수 있습니다.



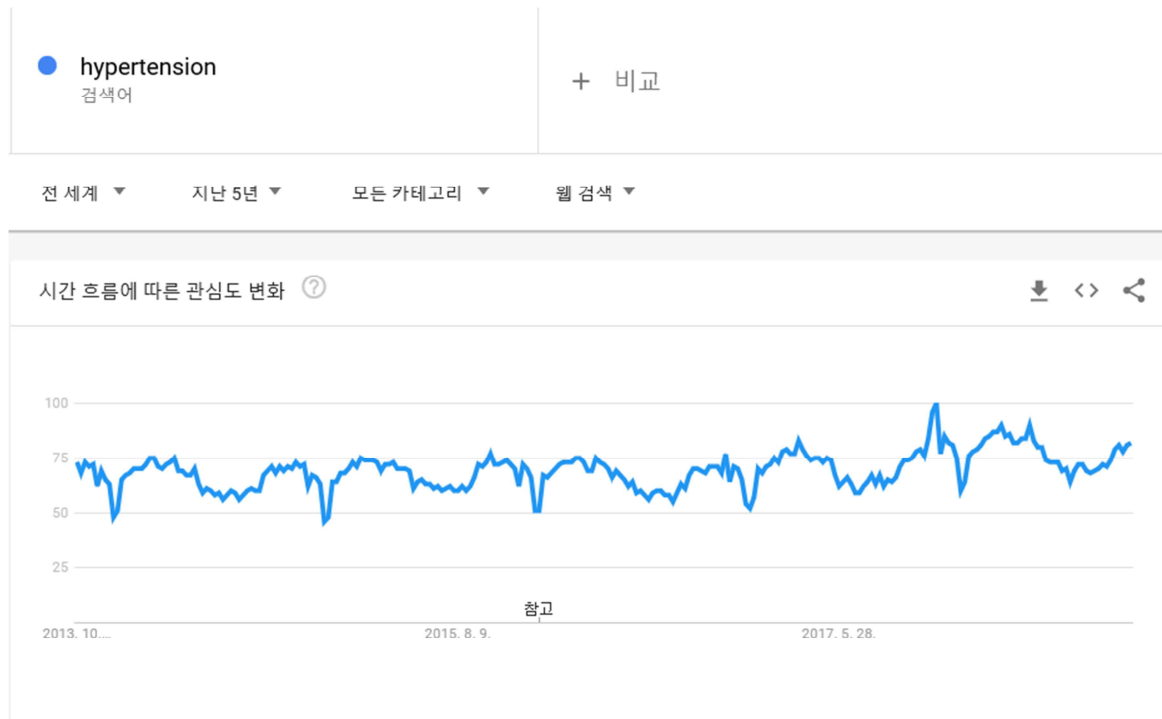
한글로 검색한 결과이기 때문에 지역적인 관심도가 부정확할 수 있으므로 hypertension으로 지난 5년간 검색을 해봤습니다.

재미있는 사실을 몇 가지 발견할 수 있습니다.

첫째, 지난 8월 5일의 고혈압약 원료 발암물질 사태가 있을 당시 전 세계적인 관심도에서는 특별히 관심도가 튀지 않는다는 사실입니다.



최근 90일간의 관심도로 확인해 보아도 뚜렷한 현상은 보이지 않습니다. 여러 가지 시사점이 있겠지만 가장 먼저 생각할 수 있는 문제는 “복제약” 으로 생각됩니다. 발암물질이 검출된 고혈압약 원료들은 상당수가 오리지널 약이 아닌 복제약에서 주로 사용된 원료이었다는 것으로 생각할 수 있습니다.



두 번째 흥미로운 것은 특정 주기마다 관심도가 낮다는 것으로 알 수 있습니다. 좀 더 자세히 확인하면 주로 12월 중순과 말입니다. 이는 계절적인 문제로 생각됩니다. 마지막 세 번째는 최근 2년을 보면 관심도가 조금씩 상향되고 있다는 것입니다.

하위 지역별 관심도 ②
하위 지역

관련 주제 ②

급상승 ▼ ▴ <> ↻

관련 검색어 ②

급상승 ▼ ▴ <> ↻

관련 주제 ②	급상승 ▼ ▴ <> ↻	관련 검색어 ②	급상승 ▼ ▴ <> ↻
1 발암 물질 - 주제	급등	1 고혈압 발암 물질	급등
2 대한민국 식품의약품안전처 - 주제	급등	2 고혈압 약 발암	급등
3 식품의약품안전처 - 대한민국 청원...	급등	3 발암 물질 고혈압 약	급등
4 뇌출혈 - 주제	급등	4 식약 처 고혈압 약	급등
5 부정맥 - 주제	급등	5 판매 중지 고혈압 약	급등

< 24개 주제 중 1-5번 표시 중 >
< 13개 검색어 중 1-5번 표시 중 >

최근 1년간 지역별 관심도와 관련 주제와 검색어입니다.

지역적인 관심도에서 보면 특히 강원도 지역에서 관심도가 높은 것으로 확인할 수 있습니다.

하위 지역별 관심도 ②
하위 지역

관련 검색어인 고혈압 발암 물질로 검색을 해보겠습니다.

고혈압 발암 물질로 검색한 결과에서 의미 있는 현상을 확인할 수 있습니다.

서울, 인천, 경기와 부산 지역에서는 고혈압 발암 물질에 대한 관심도가 높은데 비하여, 기타 지역에서는 관심도가 매우 낮다는 사실입니다. 이는 여러 가지 시사점이 있을 수

있습니다. 지역별 관심도의 차이인지? 지역별 정보의 차이인지? 등 여러 가지를 생각해
게 합니다.

이렇듯 구글 트렌드를 이용하면 최근 관심도가 높은 검색어에 대한 추이와 연관성, 지역
적 특성 등을 생각할 수 있습니다.

2. 사례

■ 올빼미 버스



Why? 빅데이터, 예원출, 2017

우리 나라에서 빅데이터를 활용한 공공분야 성공 사례 중에는 서울시 심야 버스, 일명
올빼미 버스가 있습니다. 올빼미 버스의 노선을 위해서 자정부터 새벽 5시까지의 휴대폰
통화량 빅데이터와 밤시간대 유동인구 데이터를 이용하여 분석한 결과입니다.

새벽 시간대 유동인구가 많으면서 휴대폰 통화량, 신용카드 결제 내역 그리고 택시 교통
카드 결제 내역의 빅데이터 자료를 이용하여 최적의 노선과 배차 간격 등을 조율한 올빼
미 버스가 탄생했습니다.

빅데이터는 기존의 정형 데이터가 아닌 비정형 데이터까지 이용하여 MB, GB 단위보다
훨씬 더 큰 TB 이상의 데이터를 이용하여 분석한 것입니다.

■ 구글의 인공지능

구글의 인공지능은 초기 개와 고양이의 사진을 구분하지 못하는 시절이 있었습니다. 하
지만 사진 이미지가 계속 축적되고 딥러닝을 이용한 빅데이터 분석이 사용되면서 이제는
개, 고양이, 사람을 서로 구분할 수 있는 시대가 되었으며, IBM 왓슨은 체스를 구글의 알

파고는 바둑으로 사람을 이기는 시대가 되었습니다.



■ IBM의 인공지능 시스템, 왓슨

구글은 번역, 딥러닝, 자율주행 분야에서 선두하고 있으며 상업적인 부분에서는 IBM의 왓슨이 앞선다고 볼 수 있습니다.

왓슨은 학교, 제약회사, 병원 등에 시스템을 구축하고 있으며 우리 나라에서는 가천대 길병원 등 8개 병원, 중국은 1000여개의 의료기관에 도입되어 암진단 등에 사용되고 있습니다. 하지만 아직 갈 길은 험난하고, 2018년의 경우 국내 병원 도입 0건으로 의약 분야에서는 암 진단의 정확도가 기대에 못미친다는 평가가 나오고 있는 것도 현실입니다.

왓슨이 주로 미국, 유럽 지역의 환자와 논문을 학습했기 때문에 한국과 같은 동양인에 대한 정보가 부족하고 자주 발생하는 암의 종류나 형태 등이 다르기 때문에 아직은 암 진단의 정확도 등이 떨어지는 것입니다.

하지만 구글의 알파고가 초기에 유럽의 프로 바둑 기사와 대국을 하면서 정보를 축적하고 학습을 통하면서 발전하여 이세돌 9단에게 이긴 것과 같이 동양인 환자와 논문 등을 꾸준히 학습하게 된다면 그 정확도는 상당히 올라갈 것으로 기대하고 있습니다.

■ 파리바게뜨 A점의 매출 변화

기업에게 있어서 빅데이터는 이익을 극대화 시켜주는 존재이기도 합니다.

“날씨에 따라 고객의 입맛이 변화한다”는 사실을 이용하여 국내 대표 베이커리 전문매장인 파리바게뜨는 5년간 169개 지역의 일별 매출과 기상자료를 통계기법으로

지수화하여 “날씨 판매지수”를 활용하였고, 2012년부터 날씨 마케팅을 시작하였습니다. 이를 통해 파리바게뜨는 판매량을 예측하고 주문량을 조절할 수 있었는데요. 그래서 찬스로스, 즉 판매할 제품이 없어 발생하는 손실을 방지하고 재고 부담을 줄일 수 있었다고 합니다. #4 전체적인 영업이익이 증가한 것은 말할 필요도 없습니다. 빅데이터를 잘 활용하면 기업의 매출 향상에도 크게 도움이 됩니다.

■ 2016년 올림픽의 성공적 개최

빅데이터의 순기능은 비단 기업에서만 일어나는 것은 아닙니다. 빅데이터가 주는 다른 이점에는 어떤 것이 있는지 좀 더 자세히 살펴보겠습니다. 2016 올림픽 개최지인 브라질의 리우데자네이루는 올림픽을 잘 치를 수 있도록 빅데이터를 활용했습니다. 바로 지능형 운영센터를 만든 것인데요. 이 운영센터는 도시 관리 기능과 긴급 대응 시스템을 구축하여 운영되었습니다. 시스템에서는 홍수와 산사태, 빅데이터를 활용한 지능형 운영 센터를 통해 도시 관리와 긴급 대응 시스템을 구축해 두었습니다. 본 시스템은 홍수, 산사태 등의 자연재해에 인적, 물적 피해를 최소화할 수 있도록 하였고 교통이나 통신, 전력 등 도시 기능을 유지하는 인프라가 문제 없이 돌아갈 수 있도록 관리합니다. 빅데이터를 이용해 세계적인 행사를 무사히 치뤘던 좋은 사례라고 할 수 있습니다.

■ 독일 축구대표팀을 승리로 이끈 빅데이터

또 다른 사례는 빅데이터를 활용하여 우승을 거머쥔 축구 국가 대표팀의 경우입니다. 2014년 월드컵은 강력한 우승 후보국이었던 브라질에서 열렸습니다. 브라질은 축구 강국으로 세계적으로 이름 높은데요. 브라질을 이기기 위해 독일은 다름아닌 빅데이터를 활용했습니다. 우승을 위해 독일은 비즈니스 솔루션을 제공하는 세계적인 IT기업인 SAP와 협력하였습니다. 이 기업이 개발한 플랫폼은 매 분마다 1만 2천여 개의 데이터가 생성하며 센서 데이터, 대용량 데이터를 분석하는 능력을 갖추고 있는데요. 플랫폼을 이용하여 독일 축구대표팀 선수들의 능력치인 운동량, 심박수, 방향 감각, 슈팅 동작, 순간속도와 같은 구체적인 데이터를 분석하였습니다. 데이터를 통해 개개인의 역량을 파악한 다음 각각에 맞는 훈련 프로그램과 전략을 짰는데요. 빅데이터를 활용한 보다 과학적이고 능률적인 전략을 통해 결국 독일팀은 우승 후보국인 브라질을 제치고 2014년 월드컵에서 당당히 우승할 수 있었습니다. 빅데이터를 활용하여 팀의 승리까지 거머쥐게 된 것입니다.

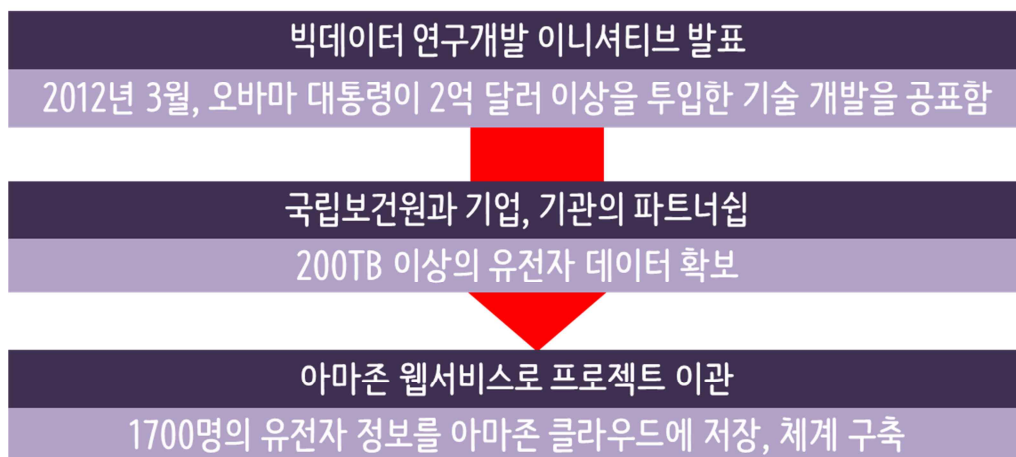
■ 미국 국립보건원의 질병치료체계

이번에는 미국의 사례를 살펴보겠습니다. 미국의 국립보건원은 다양한 질병을 연구하기 유전자 데이터를 공유하고 분석할 수 있는 시스템을 마련했는데요. 1700명의 유전자 정보를 아마존 클라우드에 저장하여 누구나 데이터 이용이 가능한 본 시스템을 이용하면 주요관리대상에 해당하는 질병을 관리하거나 예측할 수 있습니다.

인간은 유전자 수가 35000개로 그 수가 너무 많아 같은 질환이라도 증상이 다르게 나타나며 약물 반응도 다르게 발생한다고 합니다. 그래서 미국에서는 한 해 220만 명이상이 약물 부작용 때문에 입원하고 있는 형편입니다. 그래서 미국 외 여러 국가들은 약물 부작용으로 인한 인명 피해를 막기 위해 인간 유전체의 다양성을 파악하기 위한 국제 협력 프로젝트를 진행하게 되었습니다. 본 프로젝트를 통해 세계 여러 인종들의 유전자를 조사하고 이를 빅데이터화 하여 주요 질병에 대한 관리와 예측을 진행할 예정입니다.

■ 질병치료체계 추진과정

질병치료체계 추진과정



유전자 공유를 위한 질병치료체계의 추진과정을 살펴보겠습니다. 2012년 3월 오바마 대통령은 빅데이터 연구개발 이니셔티브를 발표하며 2억 달러 이상을 투입한 기술 개발을 공표했습니다. 그런 다음 미국 국립보건원이 기업, 기관과의 파트너십을 통해 200TB의 유전자 데이터를 확보했습니다. 국립보건원은 빅데이터 연구개발 이니셔티브 발표에 따라 본 프로젝트를 아마존 웹서비스로 이전하여 저장했습니다. 아마존

웹서비스는 1700명의 유전자 정보를 아마존 클라우드에 저장하게 됩니다. 그래서 현재와 같은 체계를 구축할 수 있었는데요.

■ 질병치료체계 기대효과

본 서비스를 통해 기대할 수 있는 효과로는 유전자 정보를 공유하여 새로운 질병에 대한 빠른 진단 서비스를 제공할 수 있습니다. 난치병과 불치병에 관련된 유전자 정보를 공유하고 분석함으로써 새로운 치료제를 개발할 수 있습니다. 최신 IT 기술을 결합하여 기존에 치료하기 힘들거나 치료에 오랜 시간이 걸리는 질병에 대한 치료 확률을 높일 수 있는 기회를 마련할 수 있습니다.

■ 보건복지부의 행복e음 서비스

우리나라 보건복지부의 행복이음 서비스입니다. 이 서비스는 지자체 공무원의 복지 행정 처리를 지원하는 정보 시스템인데요. 지방자치단체에서 집행하는 120여 가지 복지급여와 서비스 이력 데이터를 이용하고 있습니다. 복지 대상자의 선정을 정확하고 효율적으로 할 수 있어 복지행정의 획기적인 변화를 기대할 수 있습니다.

우리나라는 점점 복지국가로서의 모양을 갖추가면서 복지에산이 증가하고 복지 서비스를 다양화하게 되었습니다. 이에 따라 복지가 꼭 필요한 사람에게 중복이나 누락 없이 필요한 도움을 줄 수 있는 통합 관리망이 필요하게 되었습니다. 그래서 행복이음 서비스가 구축되게 되었는데요.

행복이음에서는 각종 사회복지 급여와 서비스 지원 대상자의 자격, 이력에 관한 정보를 통합 관리하고 있습니다. 또, 지방 자치단체에서 집행하는 120여 가지의 복지급여와 서비스 이력을 개인별, 가구별로 통합관리합니다. 민원인의 욕구를 정확히 파악하여 필요한 서비스를 꼭 필요한 서비스를 누락 없이 맞춤형으로 설계하여 신청하도록 안내하고 있습니다.

■ 행복e음 서비스의 기대효과

- ✓ 복지급여지급의 부정 소지를 차단하고 재정의 투명성을 제고함
- ✓ 전국 어디서나 관리망에 접속하여 연계 서비스를 제공함
- ✓ 복지수급자를 통합 관리하여 우선 대상자를 정확하게 선정함
- ✓ 사회복지통합관리망 구축을 통한 업무효율성을 극대화함

■ 빅데이터의 가치 창출

빅데이터가 언제나 좋은 결과를 초래하는 것은 아닙니다. 일반인에게는 개인 맞춤 서비스를 제공하고 기업에는 수익을 극대화 시켜주며 국가에는 국가 경쟁력을 향상시켜 주는 빅데이터, 다방면에서 가치창출을 하는 빅데이터의 이면에 있는 문제점에 대해서도 살펴보겠습니다.

■ 빅데이터로 인한 피해 사례

우선 가장 문제시되는 것이 바로 사생활 침해입니다. 개인이 인터넷에 검색했던 내역, SNS에 올린 글과 사진, 핸드폰으로 주고 받았던 대화들이 모두 데이터 베이스에 남아 이 데이터들을 조합하면 새로운 정보로 재생성 할 수 있는데요. 비식별화된 데이터를 수집한 후 여러 번의 가공 과정을 거치고 이것을 분석, 처리하면 실제 개인을 식별할 수 있는 정보로 재탄생 하는 셈입니다. 그러니까 무방비 상태로 웹을 통해 공개된 나의 단편적 개인 정보가 합쳐지면 내가 누구인지 알 수 있는 정보로 바뀐다는 것이죠.

예를 들어보겠습니다. 미국의 한 대형 소매 유통업체는 고객 맞춤형 상품 추천 서비스를 제공했는데요. 고객 중 한 명이 임신부용 속옷과 튜닝 방지 크림을 구매했던 내역이 있어 이를 토대로 고객의 임신 사실을 알아내고 임신부용 쿠폰을 우편으로 발송하였습니다. 그런데 이 고객은 알고 보니 미성년자였고 자신의 자녀가 임신을 했다는 사실을 알게 된 부모는 경악할 수 밖에 없었습니다. 부모보다 먼저 임신 사실을 알게 된 유통업체는 본의 아니게 고객의 프라이버시를 침해하게 된 것입니다. 매우 똑똑하고 정보 처리를 빠르게 한다는 것이 무조건 좋은 결과를 초래하는 것은 아니라는 것을 알 수 있는 대표적인 사례라고 할 수 있습니다.

이와 같이 비식별화된 데이터를 여러 번의 수집과 분석을 통해 개인을 식별할 수 있는 자료로 만들 수 있다는 점, 편리하고 혁신적인 신기술인 빅데이터 분석 기술의 이면에 숨겨진 어두운 부분이라고 할 수 있겠습니다. 하지만 지능정보사회와(빅데이터, 인공지능 시대) 관련한 새로운 보안 및 윤리 이슈에 대하여 국가가 별도로 법을 제정비하는 등 대안을 마련하고 있습니다.