**빅데이터**

**Big Data 개념 및 역할 그리기이(가) 표시된 사진

자동 생성된 설명**

**한국산업기술대학교**

**201715404 컴퓨터공학과 권한길**

|  |
| --- |
| **(제4차 산업혁명) 제4차 산업혁명과 연계하여 ‘Big Data’의 필요성과 비전에 관하여 교안의 내용들을 기반으로 정리해 보시오.** |
| 4차 산업혁명이란 학문 및 기술의 경계가 없어지고 여러 분야의 기술이 융합되어 새로운 기술 혁신을 뜻한다. 4차 산업혁명 요소 기술은 3가지가 있다. 사물인터넷, 빅데이터, 인공지능이다. 그 중 빅데이터는 단순히 대용량 데이터만을 의미하는 것이 아니라 대용량 데이터 활용 및 분석을 통해 가치 있는 정보를 추출하는 것이다.  먼저 빅데이터를 출현 배경을 이해하는 것이 중요하다. Ppt 18page를 참고하면 메인 프레임 컴퓨터 시대, PC시대, 인터넷 모바일 시대, 스마트 시대 로 변화하는 시대 속에서 현 시점인 스마트 시대에는 엄청난 데이터가 발생하게 된다. 인터넷의 환경의 변화, 사용자 규모의 변화, 사용자당 데이터 양의 변화에 따라 데이터의 전체 크기가 기하급수적으로 증가한 것이다. 예를 들어 트위터, 페이스북 등 SNS의 급격한 확산과 기업들의 데이터 수집 증가, 스마트 모바일 디바이스와 멀티미디어 콘텐츠의 증가를 예를 들을 수 있으며 이미 구글, 아마존, 페이스북, 애플 등 기업들은 핵심 서비스를 통해 방대한 데이터를 수집하고 활용하고 있다.  이 분석한 데이터만을 빅데이터라 하지 않는다. 앞서 언급했듯이 대용량 데이터 활용 및 분석을 통해 가치 있는 정보를 추출해야 한다. 가치 있는 정보란 과거에 축적된 데이터를 바탕으로 미래에 능동적으로 대응하거나 변화를 예측하기 위한 정보이다. 이 변화를 예측하는 것이 왜 필요한가는 빅데이터의 필요성을 의미한다. 자본주의 사회에서는 수요보다 공급이 많은 것이 당연하다. 그런데 이러한 수요보다 공급이 많은 특징은 기업의 손해를 가져올 수 있다. 이 때 고객의 수요를 파악하면 낭비를 줄일 수 있다. 수요량만큼만 생산하면 되기 때문이다. 데이터 속에서 주된 고객층의 흐름을 파악하여 거기에 맞는 상품을 제공하는 것이 고객과 기업 모두의 이익이 되기 때문에 빅데이터 활용은 이제 기업 입장에서 선택이 아닌 필수조건이 되는 것이다.  이는 이미 실제로 많은 기업들이 빅데이터를 적용하는 이유이다.  현재 2020년에 살고 있는 우리는 아주 쉽게 예시를 찾아볼 수 있을 정도로 많이 발전되었고 적용되었다. 구글이나 네이버 검색에 나이키 신발을 한번 검색했을 뿐인데 페이스북, 구글, 네이버 , 인스타그램 등 어떤 SNS를 접속해도 광고판은 나이키 신발을 홍보하고 있다.  빅데이터는 기업 매출 뿐만 아니라 불확실한 미래에 대해서 예측을 하는 모든 분야에서 쓰일 수 있다. 이는 굉장한 기술이라고 밖에 이해할 수 없다. 이는 인공지능 분야에서도 필수적이다. 직접 실험 및 수행을 해야 하는 것에 대한 실험을 가상으로 시뮬레이션 돌림으로써 수행비용을 절감하여 많은 위험요소를 없애 주기도 한다. 이처럼 빅데이터는 4차산업시대에 없어서는 안되는 가장 중요한 기술이다. 2015년부터 지금 2020까지도 엄청난 속도로 많은 데이터가 쌓이고 있으며 미래에는 더 많은 저장 공간과 데이터 처리 기술, 분석 기술 등 다양한 분야에서 널리 쓰일 것이라고 전문가들은 이야기한다. |

|  |
| --- |
| **(Big Data 적용) Big Data가 적용된 사례 및 서비스를 교안에 있는 것을 제 외하고 5가지만 들어서 간략히 소개해 보시오.** |
| **1. 로켓배송**  -한 소비자가 기저귀가 필요하여 주문을 하면, 기존의 소셜커머스 기업들 에서는 주문 정보가 들어가고 그때 포장을 하여 택배로 보내는 데까지 1~2일 정도 시간이 필요하다.  그러나 ‘로켓 배송’ 을 해 주는 소셜커머스 기업에 기저귀를 주문한다면, 주문 정보가 기업으로 가는 것이 아니라, 주문을 한 소비자의 지역에 있는 ‘캠프’ 라고 불리는 물류센터로 간다. 그렇게 되면 주문 정보를 받은 ‘캠프’ 에서는 미리 포장을 완료한 기저귀를 즉시 배송을 하여 불과 2~3시간 만에 소비자에게 기저귀가 배달된다.  ‘캠프’ 에서 미리 포장 완료된 기저귀를 준비할 수 있었던 이유가 바로 빅데이터 분석을 활용하였기 때문이다. ‘로켓 배송’ 에서는 각 지역별로 주문했던 지난 내역들을 빅데이터 분석을 통해 주문이 들어올 상품들을 미리 포장을 완료하여 ‘캠프’ 라는 물류창고에 보관하고 있었기 때문에 빠른 시간 안에 주문부터 배송까지 완료를 할 수 있었던 것이다.  빅데이터 분석을 통한 주문상품 예측으로 빠른 배송이 완료되어, 소비자들의 기업에 대한 만족도가 ‘로켓 배송’ 을 안 하는 다른 업체에 비해 월등하게 높게 나타났다.  **2. 알파고**  -알파고는 구글(Google)의 딥 마인드(DeepMind Technologies Limited)가 개발한 인공지능(AI, Artificial Intelligence) 바둑 프로그램이다. 1200여대의 중앙처리장치(CPU)가 연결된 슈퍼 컴퓨터로, 빅데이터 연산을 수행하는 소프트웨어다. 알파고는 초당 경우의 수 10만개를 검색할 수 있다.  알파고에는 CPU와 더불어 그래픽연산장치(GPU)도 500∼600장 투입됐다.  알파고는 사람의 두뇌처럼 작동한다. 알파고의 능력은 상상하기 힘든 속도의 연산 능력 뿐만 아니다. 알파고는 사람의 두뇌처럼 신경망 구조로 작동한다  신경망 구조의 핵심은, 중요한 것만 추려서 걸러내는 수법으로 효용성이 떨어지는 경우의 수를 빨리 내칠 수 있다는 점이다. 그러니 알파고는 모든 경우의 수를 연산하지 않고 가지치기를 통해 중요한 것만 걸러내 효율성을 극대화한다  **3. 서울시 심야버스 노선도** -주변을 둘러보면 빅데이터는 이미 우리의 일상 속 많은 분야에서 응용되고 있다. 대표적인 사례는 2013년 개통한 서울시 심야버스 노선이다. 심야버스의 노선도는 직관적으로 만들어진 것이 아니라 빅데이터를 통한 철저한 분석에 의해 도출된 것이다. 어떤 식으로 사용 되었는가를 간단히 살펴보면 다음과 같다.   우선 심야시간대의 통화량 등을 이용하여 서울시 각 지역의 유동인구 밀집도를 수집하고 이를 분석한다. 그리고 결과에 따른 최적의 노선을 도출한다. 이렇게 도출된 노선을 기준으로 다시 한 번 유동인구 빅데이터를 이용하여 배차 간격을 조절한다. 이러한 과정을 거쳐 탄생한 심야버스는 하루 평균 1000명 이상의 승객이 이용하고 있다. 즉, 빅데이터를 통한 분석과 의사결정이 사람들에게 좀 더 편안함을 준 사례이자 아주 모범적인 빅데이터의 활용 사례라고 할 수 있다.  **4. 신용평가모델**  -빅데이터는 일상 생활 뿐만 아니라 금융에서도 널리 이용되고 있다. 최근의 신용평가 모델은 기존의 복잡한 과정과 인터뷰 등을 거치지 않고 온라인상으로 제공하는 서비스가 등장하고 있다. 이는 빅데이터 기반의 핀테크 서비스 중 하나로 대두되고 있다. 기존의 신용평가는 여러 종류의 서류와 인터뷰 등을 통해 이루어졌다. 심사 대상자는 신용평가 등급을 받기 위해서는 복잡하고 귀찮은 준비 과정을 거쳐야만 했다.   그러나 빅데이터를 기반으로 한 새로운 신용평가 모델은 이러한 준비가 전혀 필요치 않다. 최소한의 신청자 정보만을 가지고 해당 신청자의 소셜 네트워크 활동 데이터를 수집하여 이를 기준으로 평가하는 방식이다. 예를 들면, 신청자가 소셜 네트워크상에서 “부정적인 단어의 사용이 많다” 라는 것이 빅데이터 수집, 분석을 통해 도출되었다면 신청자의 신용도에 대해서 다시 한번 고려해 볼 만한 사항이 될 수 있을 것이다.   또한 택배 이용률도 신용평가의 기준이 될 수 있다. 만약 A라는 사람이 꾸준히 택배를 이용하는 사람이라면 A는 일정한 소득이 있을 것이라고 유추할 수 있고 이것을 신용도를 평가하는 요소로 활용할 수 있을 것이다. 즉 빅데이터는 기존의 정형적인 데이터를 넘어서 비정형적인 데이터에서도 의미를 찾을 수 있도록 해 주었다고 할 수 있다.   **5. 플루 트렌드**  -구글의 ‘플루 트렌드’ 역시 빅데이터를 활용한 사례이다. 일반적으로 사람들은 감기에 걸리면 병원 약국에 가기 전 독감이나 같은 단어를 검색하는데 구글은 이러한 점에서 생각해서 2008년부터 검색 정보와 위치를 기반하여 미국의 감기바이러스 확산 상황을 알려주는 서비스를 제공 중에 있다. 즉, 이 시스템은 ‘감기’나 ‘독감’과 같은 검색빈도가 높은 지역을 지도에 표시 함으로써 독감의 확산을 예측할 수 있다. 대부분의 보건 기구는 일주일에 한번만 예상 수치를 업데이트하지만 Google 독감 트렌드는 18개 국가를 대상으로 매일 업데이트되므로 기존의 시스템을 보완한 것이다. |

|  |
| --- |
| **□ (Big Data 연계기술) 효율적인 데이터 분석을 위해서 필요한 기술력, 특히 본 강의에서 다루길 희망하는 기술이나 내용은 무엇인지 A4 1/2 분량으로 정리해 보시오. (자신의 주관적인 생각 기재)** |
| 1. 텍스트 마이닝 기술  -현재 빅데이터 기술이 많이 발전하고 있고 실제로 기업에서도 많은 채용을 진행하는 것으로 알고 있습니다. 이에 발 맞춰 비정형 텍스트 데이터에서 가치와 의미가 있는 정보를 찾아내는(Mining)기술인 텍스트 마이닝 기술을 배우고 싶습니다.  2. 파이썬 프로그래밍  -교수님께서 말씀하셨듯이 현재 파이썬 언어가 떠오르고 있는데 기본적인 문법이나 간단한 프로그래밍하는 법을 배우고 싶습니다. 추가로 텍스트 마이닝 기술을 파이썬을 통해서 할 수 있다고 알고 있는데 이 분야도 배우고 싶습니다.  3. 빅데이터 분석 기술(SNS 분석 기술)  - 빅데이터 분석 기술 중 실제 일상생활에서 가장 가깝게 접하는 SNS 분석 기술을 좀 더 자세히 배우고 싶습니다. 동작하는 원리가 궁금합니다.  4. 빅데이터 시각화 기술  - 분석된 자료를 시각적으로 묘사하는 것이 중요한데 정보 시각화 기술에 대해서도 얕게 나마 배우고 싶습니다.  5. 데이터베이스 기술  -방대한 양의 데이터를 처리하려면 저장 또한 중요한 기술인데 어떻게 그렇게 많은 데이터를 저장하고 사용하는지 배우고 싶습니다. (클라우드)  6. 데이터 분석가  -데이터 분석가가 구체적으로 어떤 일을 하는지, 분석가가 되려면 어떤 스펙과 능력이 필요로 한지 궁금합니다. 컴퓨터공학을 전공하여서 어떤 점이 유리한지도 궁금합니다. 혹은 석사 이상을 주로 뽑는다고 알려져 있는데 사실 여부와 이유가 궁금합니다. |