1. 빅데이터(Big Data)란 무엇인가?

몇 년 전부터 빅데이터와 관련해서 중요성을 강조하는 많은 주장이 많이 대두되고 있다. 이로 인해 사람들은 빅데이터에 관심을 가지게 되고 활용도에 대해 생각해보기 시작했다. 빅데이터란 다양한 형태의 데이터를 수집, 처리, 저장해서 목적에 맞게 분석하고 이를 전략적 의사결정에 활용하거나 지식을 시스템화하여 생산성 향상에 활용하고 새로운 비즈니스 모델의 창출에 활용하려는 패러다임이다. 예전에는 연역적 방법으로 지식을 추출했으나 최근에 컴퓨터의 발전으로 대용량의 데이터를 저장하고 처리할 수 있게 되는 등 지식을 추출하는 환경이 급격하게 변하게 되어 귀납적인 방법을 활용하고 있다. 빅데이터의 주요 핵심 요소는 3가지가 존재하는데 기술적 요소, 전략적 요소, 인적 요소가 이에 해당한다. 기술적 요소란 데이터, 데이터 수집 기술, 데이터 분석 기술로 구성되고, 전략적 요소는 빅데이터 분석 및 활용 기회를 탐색하는 것, 인적 요소는 데이터 분석에 대한 이해도가 높은 구성원이 얼마나 있는지 확인하는 것이다. 빅데이터의 중요성이 대두된 이유는 데이터 분석으로 중요한 정보나 지식이 추출될 수 있기 때문이다. 데이터 분석의 과정은 데이터 소스를 인식하고, 사람이 그 데이터를 수집하고 저장해서 처리, 분석, 활용하게 된다. 이렇게 데이터 분석에서 나오는 정보나 지식을 활용하자는 패러다임인 비즈니스 인텔리전스가 등장하였고, 그 후에 빅데이터가 등장하게 되었다. 이 비즈니스 인텔리전스와 빅데이터의 차이는 데이터양의 차이, 분석대상 원천 데이터의 다양화, 외부 데이터의 활용 여부 차이가 있다. 빅데이터는 최근에 새로 생겨난 개념이 아니라 데이터 분석 개념에서 발전했기에 데이터 분석에 대한 이해와 함께 빅데이터에 접근해야 할 것이다.

이렇게 빅데이터를 활용함에 따라 인공지능 또한 토픽으로 대두되고 있다. 인공지능 또한 최근에 등장한 개념이 아니라 1950년대 중반에 등장한 개념인데 존 매카시가 논문에서 인공지능을 거론한 이후로 인공지능을 주제로 하는 국제 컨퍼런스도 진행하게 되었다. 그러나 그 당시에는 인공지능을 구현하기에는 애로사항이 많았기 때문에 관심도가 점점 떨어졌었고, 1980년대에 지식을 if-then 룰의 형태로 추출해서 규칙 기반 시스템, 전문가 시스템을 개발하게 되면서 다시 인공지능의 관심도가 다시 올라가게 되었다. 그 후로 2010년대에 들어서 컴퓨터의 발전과 엄청난 데이터양, 데이터를 분석하는 방법론의 발전으로 인해 연역적 방법이 아닌 귀납적 방법으로 인공지능을 만드는 발상과 함께 개발을 시작하였고, 이것의 대표적인 예가 알파고였다. 이렇게 빅데이터를 기계로 학습해서 인공지능을 만들게 되면서 우리 삶이 점점 변화하기 시작했는데 첫째로 의사결정의 질을 높일 수 있고, 두 번째로는 운영효율을 굉장히 높일 수 있다. 마지막으로 새로운 가치의 창출도 가능하게 되었다. 이러한 특징 덕분에 인공지능을 이용해 새로운 비즈니스와 서비스를 제공할 수 있기에 이러한 혜택을 받을 수 있도록 노력을 해야 할 것이다.

이렇게 빅데이터와 인공지능이 무엇인지, 그리고 어떠한 관계인지 알아보았는데 확실히 예전보다 데이터를 수집하고 분석하고 활용하는 빈도가 많이 늘었다는 생각이 들었다. 옛날에는 다량의 데이터를 처리할 수 있는 기계나 방법이 존재하지 않았기에 데이터를 활용하는데 제약이 많았지만, 점점 기술의 발전으로 인해 데이터를 한 번에 많이, 그리고 빠르게 처리할 수 있게 됨에 따라 기업에서 양질의 컨텐츠를 제작할 수 있게 되고, 우리는 그것을 즐길 수 있게 되었다. 이렇게만 보면 빅데이터가 장점만 있는 것처럼 보이지만 만약 부정적인 내용의 데이터를 수집하여 그것을 이용해 컨텐츠를 만들거나 인공지능을 학습시킨다면 어떻게 될지 생각하면 끔찍하다. 결국 데이터를 무작정 수집하는 것이 아니라 우선 선별을 거쳐 활용해야 할 중요성이 더욱 커지지 않을까 생각한다. 미래에 데이터 선별 기술이 발전하여 데이터를 활용하는 데에 있어서 사람들이 불편해야 할 상황이 발생하지 않았으면 좋겠다.

궁금한 점이 있다면 현재 귀납적 방법을 활용해서 빅데이터를 다루는데 귀납적 방법으로 데이터를 처리할 시에 단점이 존재하는지, 그리고 존재할 때 어떤 단점인지 알고 싶습니다.

2. 빅데이터 활용전략

최근에 빅데이터가 많은 기업에서 활용됨에 따라 풍부한 컨텐츠가 제공되고 있는데 빅데이터 활용의 성공 요인이 무엇인지 살펴보면 빅데이터 분석을 위한 기획이었다. 해당 조직이 처한 문제를 정확하게 정의하고 문제해결을 위해서 어떻게 분석하고 활용할 것인지 미리 정해야지 데이터를 효율적으로 활용할 수 있는 것이었다. 특히 분석 시나리오를 작성하는 것이 중요한데 어떤 데이터를 어떻게 모아서 어떤 방법으로 분석해서 어떤 가치를 제공할지 정의해야 하는 것이었다. 이러한 분석 시나리오를 활용하여 빅데이터를 활용한 예시를 살펴보면 다음과 같다.

첫 번째로 온라인 검색어 분석을 통한 독감 예보 서비스가 있다. 이는 검색엔진에 독감을 검색하게 되면 독감 관련 검색어의 조회 수의 급증 여부를 보고 독감이 유행하는지 판별하는 것이었다. 두 번째로는 SNS를 통한 영화 흥행 예측이다. 영화가 개봉되거나 영화에 대한 정보가 나오게 되면 사용자들이 영화에 대한 의견을 올리게 되는데 텍스트 마이닝의 오피니언 분석을 사용하여 흥행 수익을 예측했다. 세 번째로 빅데이터 분석을 통한 심야버스 노선 정책 지원인데 콜이 많이 나오는 지역을 분석하여 그 지역을 분석으로 노선을 그리는 것이었다. 네 번째로 빅데이터 기반 고객의 소리 분석 시스템을 통한 서비스 혁신이다. 콜센터를 통한 질의, 민원 등을 텍스트로 변환하여 텍스트 마이닝을 통해 분석하는 것이었다. 다섯 번째로 금융기관에서 만든 신용평가 시스템인데 기업의 재무 데이터와 거기에 수반되는 부도, 비부도 데이터를 활용하여 돈을 빌려주고자 하는 기업이 부도가 날 것인지 예측하는 시스템이다. 마지막으로 감사 정보 시스템을 구축하는 것인데 여러 형태로 이루어진 기업의 운영 데이터를 가져와 모델링하여 어떤 조건에서 문제가 많이 생기는지 분석하는 것이다.

이렇게 분석을 기획하고 빅데이터를 활용한 예시를 봤는데 분석 기획을 위한 주요 요소를 살펴보면 첫 번째로 목적을 수립하는 것이 있다. 데이터 분석이 실질적인 성과로 연결되기 위해서는 제약조건이 존재하기에 목적을 명확하게 하고 분석을 진행해야 할 것이다. 두 번째로 데이터를 어떻게 확보할 것인지 정하는 것이다. 개인정보 보호법에 따라 데이터를 수집하는데 제약이 있어 데이터 수집 방법을 생각해봐야 한다. 그리고 데이터 분석을 제대로 진행할 수 있는 역량을 갖추고 있는지 봐야 한다. 마지막으로 데이터로부터 도출된 모형을 가지고 의사결정 하는 것 또한 주요 요소이다.

분석 기획을 위한 주요 요소가 충족되었을 때 분석 기획을 진행하게 되면 세 단계로 진행하는 데 첫 번째로 시나리오를 한번 생각해보고 조직에서 활용할 수 있는 분석 기회를 찾아보는 것이다. 두 번째로 분석 기회를 어떻게 구체화해서 실현할 것인지 상세 설계를 하는 것이고, 그 후로 로드맵이나 일정 작성 등 마스터플랜을 수립하는 것이다.

이렇게 빅데이터를 활용하기 위한 분석 기획의 중요성과 이를 활용한 예시를 살펴봤는데 빅데이터를 활용하는 데 기획을 위한 시나리오를 작성하고 기획대로 진행해야 하는 필요성을 느끼게 되었다. 아무래도 다량의 데이터를 활용하는 것이라 중간에 하나라도 잘못된 데이터를 수집하면 큰 참사가 일어날 수 있어 신중하게 활용해야 하기 때문이라고 생각한다. 또한 빅데이터를 활용한 예시가 우리 주변에서 많이 볼 수 있는 것이라 신기했다. 빅데이터를 많이 활용한다는 것을 알고 있었지만 이렇게 활용되는지는 전혀 몰랐기 때문에 신선한 충격을 받았다. 위의 사례 외에도 우리 생활에서 빅데이터가 활용되는 사례가 많이 존재한다고 생각하는 데 무엇이 있을지 궁금하고, 과연 미래에는 어떤 곳에 빅데이터가 활용되어 우리 생활을 윤택하게 만들 것인지 기대된다. 여러모로 이번 빅데이터 강의를 통해서 빅데이터에 대해 많은 생각을 해볼 수 있어서 좋았다.

궁금한 점으로는 위의 사례는 사람들에게 도움이 되는 경우가 많았는데 반대로 분석 계획을 수립하지 않고 빅데이터를 활용했을 때 사람들에게 피해를 끼친 사례가 있는지 궁금합니다.