독후감 - <대량살상 수학무기>

데이터 애널리틱스 – R기반 통계

201521466 김남현

책 제목인 대량살상 수학무기(Weapons of Math Destruction)는 대량살상 무기(Weapons of Mass Destruction)을 재미있게 패러디한 말이다. 컴퓨터를 비롯한 과학기술의 발달과 더불어서 빅데이터와 머신러닝 같은 단어가 익숙해지는 요즘시대에 데이터가 가지는 잠재적인 부작용을 경고한다. 어린 시절부터 수학에 재능이 있었던 책의 저자는 관련 분야에서 대학교수였고 이후에 금융시장에서 퀀트로서 활약을 했다. 분명 그녀는 숫자에 관한 전문가이기 때문에 세상에서 벌어지는 일들을 계산하고 예측하는 수학세계에 자신 있는 사람이다. 하지만 그녀는 지금 데이터과학자로서 빅데이터 기반의 알고리즘을 통한 머신러닝 세계에 회의감을 가지는 집단의 대표자이다. ‘워싱턴 당국의 나쁜 교사 색출하기’, ‘금융위기 예측실패’와 같은 사례들은 저자가 빅데이터의 허점을 지적할 때 사용하는 대표적인 예시이다. 이에 따르면 우리 사회가 만드는 알고리즘과 예측모델들이 처음에 의도한 목표를 완벽하게 성취하지 못할 뿐더러 최악의 경우 전혀 반대의 부정적인 결과를 낳을 수 있다. 잘못 설계된 모형으로 인해서 예상치 못한 사각지대에서 피해자가 발생하고 사회의 효율성이 감소한다. 사실 알고리즘을 바탕으로 한 예측모델들을 만들기 전에 데이터를 분석하고 현상의 원인을 파악하는 게 먼저인 이유가 이 때문인 것 같다. 우리가 수업시간에 배울 때 모형을 만드는 것보다 자료분석 하는 연습이 선행되는 것은 어떠한 요인들이 지금의 현상을 만들어 내는지 알 필요가 있기 때문이다. 수많은 변수들이 시시각각 발생하는 현실에서 특정 결과를 발생시키기까지 무엇이 개입되었는지 아는 것은 분명 어렵지만 그런 것들을 알지 못한다면 현실에 적용가능한 알고리즘을 개발할 수 없다. 문제는 언뜻 보면 그리 복잡해 보이지 않는 현상조차 우리가 간과하고 있는 요인에 의해 결정된다는 것이다. 예측하고 싶은 목표현상이 잠깐 생각해도 떠올릴 수 있는 원인들에 의해서만 결정되는 것이 아니며 심지어 우리가 원인이라고 생각했던 부분이 사실 현상에 전혀 영향력이 없는 존재일 수도 있다. 따라서 저자가 말하는 빅데이터 사회의 허점은 우리가 가진 기술적인 한계에 의해 발생하는 것이 아니라 잘못된 판단에 의해 발생할 확률이 더 높다고 볼 수 있다. 책 내용에 이런 내용이 있다. 일반적으로 사법체계에는 혹시나 있을 무고한 피해자가 발생하는 것을 막고 공정함을 유지하기 위하여 느리지만 꼼꼼한 일처리 과정에서 많은 시간과 돈을 들이기 때문에 의도적으로 효율성을 감소시키는 반면에, 지금 빅데이터 세상에서 설계하고자 하는 모형들은 효율성을 우선목표로 설정한 존재들이라는 것이다. 간단히 말하면 현실이랑 우리가 만들고자 하는 모델은 서로 상반되는 입장에 있다는 것이다. 사람들은 비록 감수해야할 비용이 늘어나더라도 일을 처리할 때 자신이 각자 할 수 있는 범위 안에서 그 순간에 발생할 수 있는 최대한 많은 변수들을 통제하려 한다. 그렇지만 설계된 예측모델은 어쩌면 결정적인 영향을 끼칠 수도 있는 예외적인 변수를 미처 고려하지 못하거나 한 번 더 상황을 살펴보는 인간적인 센스를 보여주지 못하는 것이다. 머신러닝에 있어서 완벽하게 인간과 동일하게 사고할 수 있는 능력을 구현하는 것은 궁극적 목표일 것이다. 결국 어찌 되었든 현재로서는 빅데이터 활용 능력에 허점이 많다는 것은 지극히 당연한 현실임에도 불구하고 우리는 그러한 사실을 쉽게 잊어버리고 빅데이터가 만들어내는 허상을 너무 맹신하는 경향이 있는 것이다. 게다가 선한 목적을 가지고 설계한 모델만 있는 것이 아니라 악의적으로 만들어진 모델을 이용해서 결과를 조작하는 행위도 발생한다. 저자는 전문가로서 빅데이터 세계의 위험을 양심적으로 경고하고 있다.