빅데이터 독후감

202121043 오희진

이 책의 제목 ‘슈퍼크런처’란 대규모 데이터 세트를 분석해서 외관상 무관한 일들 사이의 경험적 상관관계를 발견하는 넘버 크런처, 즉 대규모의 데이터를 수집하고 분석하여 통계를 통하여 다음을 예측하는 사람들을 슈퍼크런처라고 말한다.

슈퍼크런처라는 제목 답게 이 책에는 여러 가지 데이터를 분석한 통계학을 통하여 활용하는 사례들을 많이 소개한다. 통계는 우리 생활에 밀접하게 관련되어 있다. 대표적으로 우리에게 친숙한 슈퍼크런칭 사례는 넷플릭스이다.

넷플릭스를 실행시키고 무엇을 볼지 고민하다 맞춤 추천 서비스를 이용하여 영상을 시청한 경험이 있지 않은가? 이는 슈퍼크런칭을 이용한 협업 필터링인데 이 기능을 통하여 총 5만개의 영화 가운데 90% 이상이 최소한 한 달에 한 번씩 대여되도록 만들어 데이터를 통하여 새로운 수요를 창출하고 있다. 이러한 데이터 기법을 통하여 고객들이 평소에 잊고 있거나 유행에 휩쓸려 지나간 영상들을 다시 재발굴하고, 틈새시장까지 공략할 수 있도록 하였다.

하지만 이렇게 유용해 보이는 협업 필터링에도 이면이 존재한다. 개인 맞춤 서비스를 통하여 자신의 취향에 맞는 영상과 새로 찾아볼 필요가 없는 편리함도 생기지만, 우리가 원하는 부분만의 정보를 취득하는 것은 일종의 편협한 시각과 의사결정을 지배하게 될 수 있다.

데이터는 계속 쌓인다. 우리가 아침에 브라우저를 열기도 전에 구글은 이미 데이터 마이닝을 통해 내가 컴퓨터를 열자마자 들어갈 사이트를 이미 알고 있다. 데이터를 통하여 범죄를 예방하고 잡는 적절한 방향이 있을 수 있지만 프라이버시 침해. 데이터 남용 문제 등이 있을 수 있다는 점을 알고 있어야 한다.

슈퍼 크런칭에는 협업 필터링 말고도 여러 가지 활용 사례가 많다. 통계와 관련이 없어 보이는 의학계에서도 슈퍼크런칭의 사례를 찾아 볼 수 있다.

탁월한 데이터 기반 의사결정인 근거중심의학은 체계적인 통계연구를 토대로 하였다.

베타 차단제가 실제로 심부전 환자들에게 도움이 될 수 있음을 밝혀낸 것, 에스트로젠 요법이 폐경기 여성들에게 도움이 되지 않는다는 것을 입증한 것 모두 슈퍼 크런칭에 의해서였다. 또한 도널드 베릭의 ‘10만 생명 구하기 캠페인’은 환자들이 어떻게 사망하는지 반복하여 지켜본 다음, 광범위한 통계적 근거들을 통해 특정 위험을 감소시킬 만한 조처를 했는지 파악하려 했다. 이를 통해 베릭은 중환자실에 있는 환자들에게 중심정맥관을 삽입하여 감염으로 사망한다는 사실이 많다는 것에 주목하였다. 그 이후 체계적인 손 세척이 감염의 위험을 90% 이상 줄일 수 있다는 것을 알아내었다. 이 캠페인을 통해 많은 환자를 구출해 낼 수 있었다. 물론 캠페인이 시행되지 않았어도 업무 수행방식을 개선하여 환자들의 생명을 살릴 수 있었겠지만 이는 통계적으로 사망 위험을 감소시키는 것으로 보이는 조치를 모색하여 근거중심의학의 결과를 의료서비스 제공자들의 대규모 활동으로 전환할 수 있었다.

이러한 사례들처럼 데이터를 이용한 통계는 양날의 검, 즉 어떻게 쓰이냐에 따라 달라질 수 있다. 통화기록을 수집하여 범죄자를 잡거나, 이것이 프라이버시 침해가 될지는 활용하는 방법에 따라 달라질 수 있다.

기술이 발전 함에 따라 슈퍼크런칭의 활용은 점점 발전할 것이다. 우리의 직관과 경험, 그리고 거기에 통계를 더하여 더 좋은 선택을 내리는 것에 공조해야 할 것이다. 데이터의 결과만 믿을 것뿐만 아니라 우리의 생각에 통계를 합쳐 보완을 해야 한다고 생각한다.

빅데이터 개론 과목을 수강하고 교수님의 강의를 들으면서 빅데이터를 이용한 사이트, 기술 등의 예시를 많이 보았는데 책을 통하여 더 자세한 사례, 기술의 발전이 이끌 수 있는 문제점, 어떻게 활용해야 할지 더욱 배울 수 있었다.

앞으로 있을 빅데이터개론 팀 프로젝트, 캡스톤 등 후에 내가 해야 할 활동들에 대한 밑거름이 된 것 같았다.

이 책뿐만 아니라 다른 사례들을 많이 찾아보고 어떤 데이터를 이용하여 어떻게 사용할 수 있을지 더 생각해 보아야겠다.