

감이 아닌 과학으로 성공하는 빅데이터 구축 및 활용 가이드 #청년인재 김영<mark>자</mark> #

빅데이터가 뭐하는 것이냐고 물어볼 때마다 대답하기 힘들다. 사실 나도 잘 모른다. 통계를 공부하고 빅데이터를 접한 지 3년이 다 되어가지만 빅데이터를 정의하기가 어렵다. 예전에 여자친구가 빅데이터가 대체 뭐냐고 물어본 적이 있다.

나는 '백대이터는 말그대로 큰 자료를 바탕으로 기업의 의사결정에 도움이 될 수

있도록 하는 거야 '라고 답을 하고싶었지만 그때 당시에 나의 대답은 정말 형편 없을 정도로 간단했다. '부년에이터는 부년에이터야~' 빅데이터를 공부하고 있지 만 빅데이터에 대해 정의 내리라고 하면 확실하게 답을 하지 못했다. 왜냐하면 통계를 배우면서 처음 들어본 단어이면서 프로그램으로 데이터를 돌려봐도 어느 정도를 빅(BIG)이라고 봐야하는지도 몰랐고, 또한 빅데이터라고 해도 이것이 정말 가치 있는 분야인지 피부에 와 닿지 않았기 때문이다. 생각해보면 정말로 빅데이터가 무엇인가? 라는 질문에 대한 고찰을 스스로 해보거나 정보를 찾아본 적이 없었다. 그저 프로그램으로 데이터를 다루는게 재미있었기에 나는 유망한 분야라는 빅데이터를 선택했을 뿐 좀 더 에에하게 역사나 탄생 배경 등

?

Н

그런 나에게 이번에 읽은 이 책은 빅데이터에 대한 정보가 담긴 첫 빅데이터 교과서인 셈이다.

빅데이터에 대한 책 하나라도 찾아서 본 적이 없었다.

책은 IT트렌드의 변화에 대해 대형 컴퓨터부터 시작해서 개인용 컴퓨터, 빅데이터 기술의 시초인 하둡(Hadoop) 의 탄생배경, 스마트폰의 모빌리티 시대와 센어의 시대까 지 발전 순어대로 서술하였다. 또한 디지타이징 비즈니스



를 강조했는데, 나는 디지타이징 비즈니스라는 말을 이 책을 통해 처음 알게 되었다.

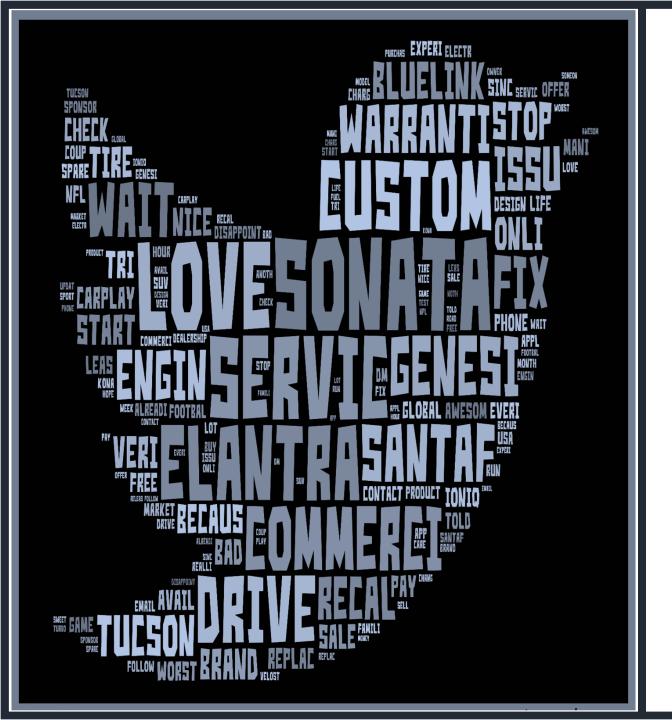
디지타이징이 무엇인가 하면, 현재의 5가지 기술(소열, 모바일, 빅데이터, 클라우드, 사물 인터넷)을 도입해서 비즈니스를 혁신하는 것이다. 여기서 말하는 디지타이징이 중요한 이유는 데이터양의 증가와 증가된 데이터를 활용하여 고객의 행동을 예측하고 그에 맞는 오퍼와 액션을 취할 수 있기 때문이다. 이로인해 자연스럽게 기업경쟁력이 강화되고, 상품의 가치 또한 증대될 것이기 때문이다. 영어로 표현 되어서 다소 어렵게 느껴질 수 있으나, 간단하게 말해서 데이터를 기반으로 한 의사결정 비즈니스라고 생각하면 될 것 같다. 그리고 2장에서는 빅데이터에 대해 나름 정의를 내려준다. '빅데이터란 기존 의 기술로 처리할 수 없는 데이터로, 이전에는 그 내용을 분석할 수 없었지만 기술이 발달함에 따라 내 용 분석이 가능…' 너무 길다. 한마디로 '분석을 통한 가치 있는 데이터들의 집합'이라고 생각한다. 또 한 흥미로우면서 중요한 글귀를 보았는데 '데이터란 과거의 기록이기 때문에 후행성을 지녔지만 동시 에 미래에 일어날 일을 내포하는 언행성을 가지고 있다'는 글귀였다. 미래를 예측한다는 행위는 영 화나 만화에서만 가능한 것이 아니라, 데이터를 바탕으로 충분히 가능한 만큼 그 가치 또한 굉장하다고 할 수 있다. 빅데이터를 배우길 잘 했다고 느낀 대목이었다. Git add . Git commit -m "First"

책은 IT트렌드의 변화에 대해 대형 컴퓨터부터 시작해서 개인용 컴퓨터, 빅데이터 기술의 시호인 하둡(Hadoop) 의 탄생배경, 스마트폰의 모빌리티 시대와 센서의 시대까 지 발전 순서대로 서술하였다. 또한 디지 타이징 비즈니스



를 강조했는데, 나는 디지타이징 비즈니스라는 말을 이 책을 통해 처음 알게 되었다.

디지타이징이 무엇인가 하면, 현재의 5가지 기술(소열, 모바일, 빅데이터, 클라우드, 사물 인터넷)을 도입해서 비즈니스를 혁신하는 것이다. 여기서 말하는 디지타이징이 중요한 이유는 데이터 양의 증가 와 증가된 데이터를 활용하여 고객의 행동을 예측하고 그에 맞는 오퍼와 액션을 취할 수 있기 때문이 다. 이로인해 자연스럽게 기업경쟁력이 강화되고, 상품의 가치 또한 증대될 것이기 때문이다. 영어로 표현되어서 다소 어렵게 느껴질 수 있으나, 간단하게 말해서 데이터를 기반으로 한 의사결정 비즈니스라 고 생각하면 될 것 같다. 그리고 2장에서는 빅데이터에 대해 나름 정의를 내려준다. '빅데이터란 기 존의 기술로 처리할 수 없는 데이터로, 이전에는 그 내용을 분석할 수 없었지만 기술이 발달함에 따라 내용 분석이 가능…' 너무 길다. 한마디로 '분석 가능하며 가치 있는 제상 모든 정도들의 집합'이다. 또 한 흥미로우면서 중요한 글귀를 보았는데 '데이터란 과거의 기록이기 때문에 후행성을 지녔지만 동시 에 미래에 일어날 일을 내포하는 언행성을 가지고 있다'는 글귀였다. 미래를 예측한다는 행위는 영 화나 만화에서만 가능한 것이 아니라, 데이터를 바탕으로 충분히 가능한 만큼 그 가치 또한 굉장하다고 할 수 있다. 빅데이터를 배우길 잘 했다고 느낀 대목이었다.



이왕이면 랜덤 추출인 로또 번호까지 예상 가능하다면 정말 좋겠다는 생각이 든다.

3장에서는 소셜미디어의 중요성을 다룬다. 특히 소셜미디어가 스마트폰에 의해 확산되었으며, 텍스 트 분석을 통해 많은 가능성을 제시하고 있다. 경험상 트위터와 같은 텍스트 데이터를 크롤링하여 분석한 후 소비자들의 감정 분석을 통해 기업의 이미지와 개선점을 파악하였을 때 문제점과 해결 해야할 방향성에 대해 어느정도 알 수 있었으며, 그만큼 텍스트 데이터의 활용가능성은 무궁무진하 다는 것을 느꼈었다. 4장에서는 빅데이터를 도입했을 때의 이슈에 대해 짚었는데, 경영자, 현업 부서, IT 부서 등의 빅데이터에 대한 이애와 기대치가 동일하지 않으므로, 빅데이터 프로젝트 중에 많이 부딪치게 된다고 하였다. 아직 기업에서 어떻게 빅데이터 프로젝트가 진행되는지 경험해 보지 못하였기에 4장을 읽을 때좀더 관심있게 읽었다. 그리고 현 기업들의 빅데이터에 대한 인식을 알 수 있었으며, 아직까지 빅데이터가 확실히 시스템적으로 자리잡기에는 시간이 걸리겠다는 생각을 하였다. 5장에서는 빅데이터 테마 언정을 위한 도출방법인데, 이 부분은 생략하였다. 다음으로 6 장에서는 구글과 넷플릭스, 아마존 등의 기업들이 빅데이터를 어떻게 활용 했는지에 대해 쓰여져 있었다. 뒷장으로 갈수록 더 이해하기 어렵고, 용어적으로도 어려워서 읽기 힘들었다. 다만 전반적으 로 빅데이터의 배경과 기업이나 사회에 어떤 의미를 갖는지, 빅데이터가 무엇인지 어느정도 알 수 있는 계기가 되었다.