

‘Big Data’라는 분야의 유행에 대한 현상 분석

빅 데이터(big data)에 대한 분야에 대해서 소개를 하고자 한다. 빅 데이터란 기존 데이터베이스 관리도구를 통해 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술을 의미한다. 이러한 데이터 분석 방식에는 텍스트 마이닝, 오피니언 마이닝, 소셜 네트워크 분석, 군집 분석 등이 있으며 이러한 분석을 통해 정치, 문화, 예술 등 다양한 분야에 걸쳐 활용하는 분야이다. 트위터 데이터 분석(소셜 네트워크 분석)을 통해 미국에서 어떤 후보자가 대선에서 승리할지를 정확히 예측한 것은 대표적인 빅 데이터 분석의 예로 들 수 있다. 하지만 앞서 이야기 했듯이 대량의 데이터를 분석하는 기법이나 알고리즘이 일반 사람들에게 이해하기 어려운 기술이고, 전문 분야임에도 불과하고 많은 사람들이 빅 데이터에 관심이 많다. 기초과학에 대한 지식의 경우, 그 단어만 아는 경우가 많은데 빅 데이터라는 단어를 알고 있는 사람들 중 다수가 어떠한 기술이고 그 기술이 어떻게 사용되고 있고 이렇게 사용하면 어떻게 되는 식은 약간의 통찰력을 가지고 있는 경우가 많다. 따라서 우리 조는 이러한 복잡한 기술에 대해서 사람들이 이렇게 큰 관심을 갖는 이유가 무엇인지, 그리고 그 관심을 갖게 만든 변인이 무엇인지 고민하고 분석하고자 한다.

우선, 혁신적인 이 기술이 어떻게 확산 되었는지를 에버렛 로저스(Everett M. Rogers)가 1926년에 제시한 혁신확산이론에 근거하여 설명하고자 한다. 혁신확산이론에 근거하여 혁신의 확산을 설명하면 새로운 아이디어, 관행, 서비스, 제품 등이 시간이 경과하면 다양한 구성원들의 커뮤니케이션 채널을 통해 확산된다. 이러한 확산에는 다양한 특성과 결정적인 요인이 있지만 로저스가 판단한 중요한 요인은 크게 상대적 이점과 적합성이 있다. 상대적 이점은 혁신이 이전의 것보다 더 낫다고 인식하는 정도를 이야기 하는데, 잠재적 채택자에게 상대적 이점이 있는 혁신의 경우 이익이나 명성을 주기도 한다. 적합성은 잠재적 채택자가 가지고 있는 기존의 가치, 과거 경험, 필요에 혁신이 부합한다고 생각하는 정도로서, 적합성이 높을 수록, 잠재적 채택자의 불확실성이 감소하게 된다.

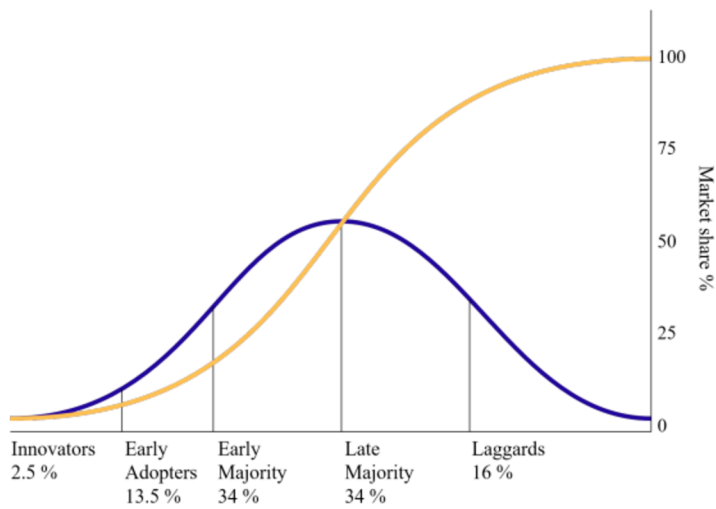
이러한 이론에 근거하여 빅 데이터의 확산에 대한 가설을 설정한다면,

1. 빅 데이터의 유행과 확산은 사람들에게 이전의 방식보다 더 큰 이점을 준다는 인식과 언론에서의 노출을 통해 자신에게 적합하다는 판단이 영향을 주었다고 가정한다. (Factorial ANOVA)

이러한 혁신의 모델은 또한 앞서 주장한 것처럼 **커뮤니케이션 채널**이 중요한 요인이다. 당시 처음 연구는 매스 미디어의 영향력에 주목 했지만, 실제로 연구를 통해 사람들에게 확산되는 실제 요인은 매스미디어 채널보다 대인 채널이 더 큰 영향력을 가진다는 사실을 알게 되었다. 매스미디어 채널은 잠재적 채택자에게 혁신의 인지도를 높이고 지식을 전달하는 가장 신속, 효율적인 수단이지만 그 이상 설득까지는 이어지지 않는다는 사실을 알게 되었다. 반면 대인채널이 혁신을 받아들일도록 타인을 설득하는 더 효과적인 채널이라는 것을 알게 되었다. 따라서 오피니언 리더가 중요하다.

이러한 연구 결과를 통해 빅 데이터의 확산에 대한 가설을 설정한다면,

2. 빅 데이터의 유행과 확산은 오피니언 리더들의 빅 데이터에 대한 언급과 주장의 증가로 인해 사람들이 빠르게 이 기술에 대해서 받아들이고 있다.(F-test)



이와 같은 가설의 연구 방법은 트위터와 서적을 중심으로 오피니언 리더들의 빅 데이터 분야에 대한 언급 노출의 증가로 인해 사람들이 많이 검색하는 네이버나 서적의 판매 추이를 통해 확인할 수 있다. 사실, 이 이론의 시작은 대인 과 대인의 관계를 통해 이루어진 채널을 이야기 하고 있지만, 현재의 사회는 SNS라는 새로운 방식의 대인 채널이 도입 되었고, 그 안에서 오피니언 리더가 형성되고 그들과 인터렉션이 활발하기 때문에 이와 같은 연구 방법이 적합하다고 생각하였다.

3. 과거와는 다른 지식 창출, 공유, 습득의 기술이 빅 데이터에 관한 관심에 영향을 줬을 것이다.(F-test)

지식사회는 지식이 경쟁력이다. 지식의 지리적 경계는 멀어져서 글로벌한 세계 속에서 지식 창출, 공유, 습득의 기술이 과거와는 확연하게 달라졌다. 세미나, 컨퍼런스와 같은 오프라인에서 이메일, 포탈, 소셜미디어와 같은 온라인에 이르기까지 지식의 네트워크가 절대적으로 필요한 상황인 것이다. 이러한 시대 변화에 대응하여 빅 데이터에 대한 전반적인 이해와 더불어 빅 데이터를 활용하여 어떠한 지식을 창출하여 활용할 것 인가에 대한 문제의식을 갖는 것은 매우 중요하다고 볼 수 있다.

한건희 진서훈(2014),빅데이터에 대한 소개와 활용사례에 관한 연구. Journer of the Korean Data Analysis Society. Vol.16 No.3.

4. 빅 데이터를 활용한 트렌드 분석과 마케팅 활용은 빅 데이터에 관한 관심에 영향을 주었을 것이다.(F-test)

문화산업분야에서 빅 데이터는 주로 소셜데이터를 이용하여 이미지분석(브랜드분석), 트렌드 분석, 위기관리, 스토리텔링의 도구, 마케팅활용 등을 위한 목적으로 활용된다. 데이터의 분석방법은 데이터마이닝, 텍스트마이닝, 평판분석, 소셜분석 등이 사용된다. 빅 데이터 관련 회사들은 소셜미디어에 대한 모니터링을 통해 위험요소를 발견하고 대응하기도 하며, 고객의 취향이나 특정인의 이미지 또는 브랜드를 분석하여 마케팅 요소로 활용하기도 한다. 이를 통해 수익을 창출 하고 새로운 트렌드를 이끌어 나가는데 활용한다.

윤홍근 (2013). 문화산업에서 빅데이터의 활용방안에 관한 연구. 글로벌문화콘텐츠, 157-179.

5. 비즈니스, 행정, 정치, 외교, 안보 등 여러분야에 걸친 빅데이터 분석의 중요도가 빅데이터 관심도 증가에 영향을 미칠 것이다.(Regression)

인터넷이 상용화 된 지 어언 20년을 넘어서는 지금, 사이버 공간에서 벌어지는 현상 중에서 가장 중요한 것을 하나 들라면, 아마도 빅 데이터 환경의 출현을 빼놓을 수 없을 것이다. 빅 데이터 환경의 출현은 비즈니스, 행정, 정치, 그리고 외교, 안보 등의 분야에 영향을 미치고 있다. 빅 데이터 환경의 출현은 21세기 신흥권력을 놓고 벌이는 국가 간 경쟁의 양식 뿐만 아니라 더 나아가 세계질서의 변화에도 영향을 미칠 가능성이 있다.

김상배(2015). 빅데이터의 국가전략:21세기 신흥권력 경쟁의 개념적 성찰, 국가전략. Vol.21 No.3.

6. 언론에 빅데이터 라는 키워드가 노출되는 빈도수가 높을수록 더많은 관심을 가질 것이다. (Regression)

회사나 정부가 아닌 개인이 관심을 가지려면 미디어가 중요한 역할을 한다고 생각하여 미디어에서 빅 데이터에 관해 자주 언급되면 개인의 관심도가 높아질 것이라고 생각해 위의 가설을 세웠다.

7. 빅 데이터를 이용한 기업의 매출과 그렇지 않은 기업의 매출의 차가 빅 데이터에 관한 관심에 영향을 주었을 것이다.(t-test)

기업의 가치 창출에 활용하기 위해 빅 데이터 기술을 적용 하였을 때 긍정적 효과가 나타나는 것을 검증할 수 있었다. 뿐만 아니라 기업이 나아갈 수 있는 지향점으로 내·외부에서 발생하는 정형데이터와 비정형데이터를 적절히 결합 함으로 분석 결과를 극대화 해 사용할 수 있다. 대규모의 정보가 현 시점 뿐만 아니라 과거의 데이터로 부터 시계열적인 연속성을 갖는데 이를 통해 시간 흐름 상의 추세분석을 가능하게 해 미래의 새로운 사업 모델일 제시할 수 있었다.

이재성, 홍성찬 (2014). 기업의 빅데이터 적용방안 연구. 인터넷정보학회논문지, 15(1), 103-112.

8. 빅데이터 솔루션을 적용한 회사의 매출과 저장된 데이터의 양, 그 회사가 사람들에게 노출된 정도 등이 빅 데이터에 대한 관심에 영향을 끼칠 것이다.(multiple regression)

빅 데이터 솔루션을 적용한 회사의 매출이 늘어난다면 사람들에게 투자에 대한 이점 때문에 관심을 갖게 된다. 또한 이 회사가 저장된 데이터의 양이 많을 수록 데이터의 이점이 있다고 판단 하였는데 이 결과가 유의미하게 되면 데이터 양이 많을 수록 그 분석이 정확도를 높여주고, 이것이 관심으로 이어질 수 있는 다른 변인에 큰 영향을 준다고 판단 하였다. 또한 이와 같은 회사들이 다양한 매체에 방법론적인 부분과 매출 등 다양한 정보가 빅 데이터라는 큰 틀 안에서 노출된 정도에 따라서 사람들이 더욱 관심을 갖는다고 판단 하였다. 이러한 판단은 앞서 이야기한 오피니언 리더들의 이야기와 연관될 수 있어서 그러한 오피니언 리더의 언급 횟수 등도 또 하나의 요인으로서 작용할 수 있다고 생각한다. 결과적으로 continuous + categorical variables가 동시에 변인으로 작용하는 다중회귀분석이 필요하다고 판단한다.