

'Big Data'라는 분야의 유행에 대한 현상 분석

12조 : 오준엽, 이종균, 허남균, 유대원

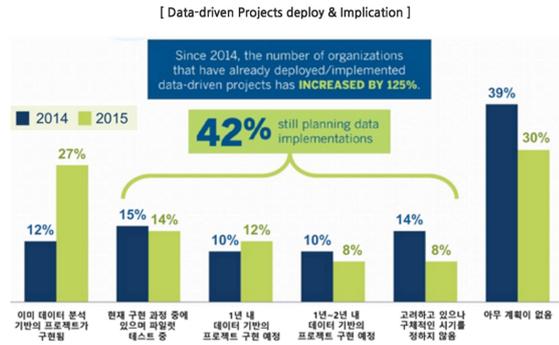
관심사에 대한 소개와 정리

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 예측(그림1)한 것은 대표적인 빅 데이터 분석의 예로 들 수 있다.

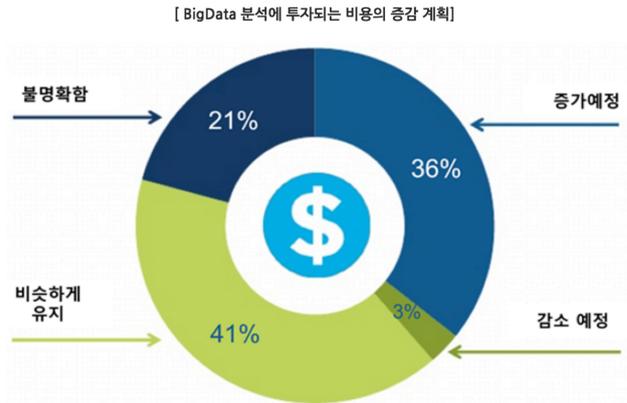


(그림1) 미국 대선을 예측한 트위터의 데이터

다음으로 최근 많은 기업들이 앞다투어 빅 데이터 기반으로 한 의사결정 시스템을 갖추기 시작했다.(그림2) 또한 기존에 빅 데이터에 투자하는 비용에 대한 비율 역시 늘리는 추세이다. 하지만 이러한 기업의 의사결정이 정말로 매출이나 다양한 이익을 근거하여 이루어진 판단인지 확실하지 않다. 대부분 앞으로의 전망을 통해 이러한 결정을 했다고 연구는 조사하고 있다. 더욱이 일반인들에게는 대량의 데이터를 분석하는 기법이나 알고리즘이 이해하기 어려운 기술이고, 전문 분야이다. 하지만 많은 사람들이 빅 데이터를 알고 그 관심이 높아지고 있다. 기초과학에 대한 지식의 경우, 예를 들어, 상대성 법칙의 경우 그 단어만 알고 의미는 모르는 경우가 많다. 반면에 빅 데이터에 관련해서는 많은 사람들이 어느 정도 개념을 알고 있고, 그 중요성에 대해서 많이 언급하고 있다. 따라서 저희 조는 이러한 대중과 기업의 관심이 정말로 어디에서 오고, 그러한 인식이 어디까지 사람들에게 영향력을 주는지 판단하고자 하였다.



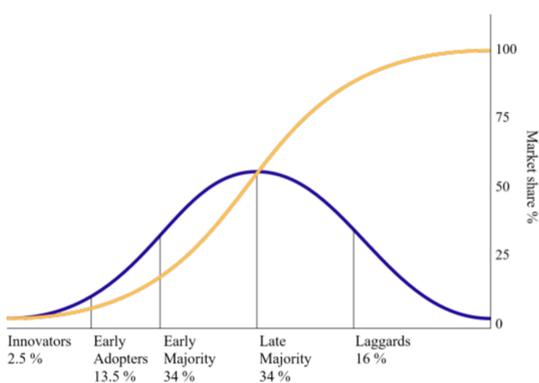
기업의 27%는 데이터 분석에 기반한 의사결정 시스템 구현이 완료



빅데이터 분석에 투자하는 비용을 증가시키겠다, 36%

(그림2)기업의 2015년 빅 데이터 투자 지표

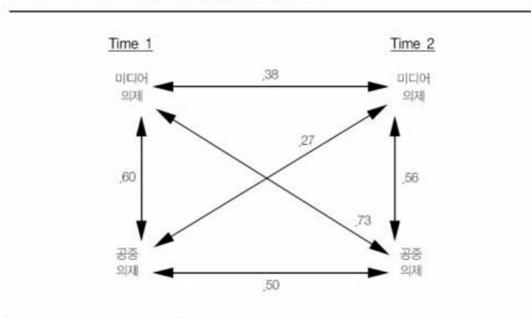
관심사에 대한 연구 문제와 가설



Diffusion of Innovations(Everett Rogers)

(그림3) 혁신확산이론

(그림 13-1) 시간 지체에 따른 의제의 상호지연 효과



Agenda Setting(Maxwell McCombs and Donald Shaw)

(그림4) 의제설정이론

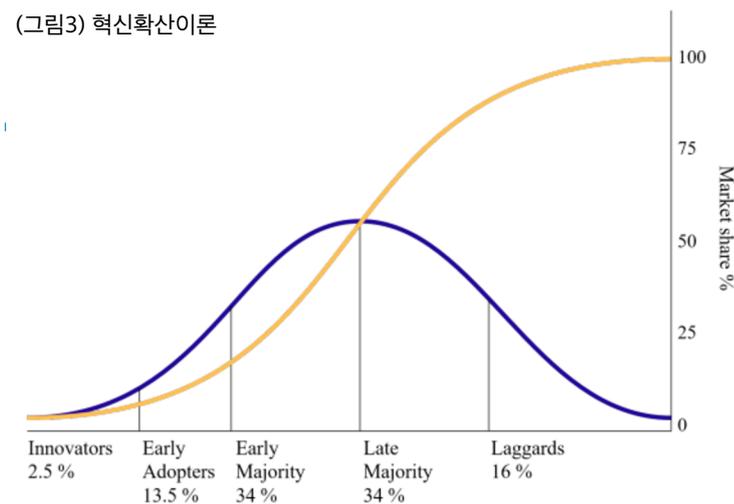
우선, 이러한 대중의 관심을 만든 것이 크게 오피니언 리더와 언론의 노출 때문이라고 생각했다. 우리의 주장을 혁신적인 기술이 어떻게 확산 되었는지를 설명한 **에버렛 로저스(Everett M. Rogers)**의 혁신확산이론에 근거 하여 설명하고자 한다. 혁신확산이론에 근거하여 혁신의 확산을 설명하면 새로운 아이디어, 관행, 서비스, 제품 등이 시간이 경과하면 다양한 구성원들의 커뮤니케이션 채널을 통해 확산된다. 이러한 확산에는 다양한 요인이 있지만 로저스가 판단한 중요한 요인은 크게 상대적 이점과 적합성이 있다. 상대적 이점은 혁신이 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. 또한 매스 미디어로 설득하는 것 보다 오피니언 리더 즉, 대인 채널이 설득에 더욱 용이하다고 주장하였다.

다음으로 의제 설정이론이 있다. 의제 설정이론의 경우, 공중의제가 중요한 것이 아니라 미디어에서 선정한 의제가 결국 공중의제가 된다는 것을 의미한다. 실제로 위의 그림(그림4)을 보면 시간이 지나고 미디어 의제가 공중 의제가 되는 비율이 73%로 실제 공중 의제가 미디어 의제가 되는 비율인 23% 보다 약 3배 가량 높은 것으로 판단되었다. 따라서 이러한 혁신확산이론과 의제설정이론이 빅 데이터에 대한 사람들의 인식 확산에 근거가 되었다고 판단하고 그에 근거한 변인을 설정하게 되었다. 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)

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

(그림3) 혁신확산이론



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.

<가설 종류 및 정리>

IV

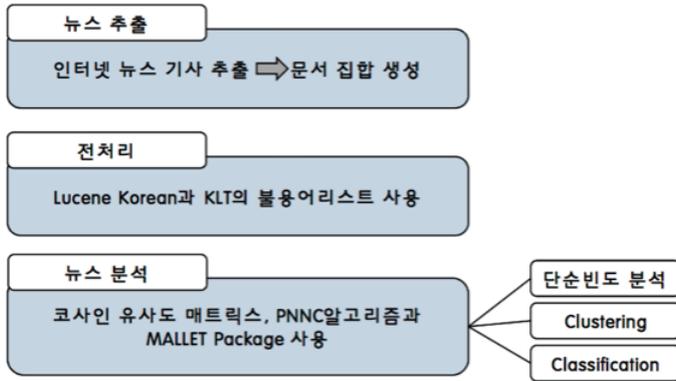
1. 언론의 노출횟수 (숫자)
2. 지식 창출, 공유, 습득 방식에 대한 변화 (사회 현상학적 연구자료 검색)
3. 트렌드 분석과 마케팅 활용 (실제 사용되는 기술 발전 추이와 사람들의 인식 설문조사)
4. 빅 데이터라는 키워드 노출 빈도 확인(실제 노출되는 데이터를 추출)
5. 기업의 빅 데이터를 이용 유무의 차이로 인한 매출 차이(데이터 분석을 하는 기업과 아닌 기업의 매출차이)
6. 오피니언 리더들의 언급과 주장 (사람들의 설문조사 & 실제 언급에 대한 횟수)

CV : 기업의 크기(중소기업) , 사람들 인식 (20-30대를 중심으로 실제 온라인 이용자)

DV : 빅 데이터의 유행과 확산

<데이터 추출 방법 및 문제시 차선택 방안>

1. 언론의 노출횟수 (숫자) : 언론의 노출횟수의 경우, 실제로 온라인 상에서 얼마만큼 빅 데이터가 언급 되었는지 언론사의 데이터를 가지고 와서 빅 데이터 관련 단어를 중심으로 얼마나 언급된 수가 늘어났는지 확인한다.



(그림5)

이와 같은 방식으로 뉴스를 추출하여 전처리 과정을 통해서 단어를 슈아낸 후에 뉴스를 분석한다. 이 그림의 과정에서는 다양한 알고리즘을 사용했지만 우리가 원하는 것은 이것이 긍정적인지, 부정적인지를 확인하는 것이 아니라 언급한 횟수에 대한 부분이기 때문에 그 횟수의 증가 추이를 확인하기 위해서 빈도수 검사를 실행한다.

2. 지식 창출, 공유, 습득 방식에 대한 변화 (사회 현상학적 연구자료 검색) : 사회 현상학적인 연구자료를 확인한다. 예를 들어, 빅 데이터를 실제로 이용하는 분야가 어떻게 증가 하였는지, 그 증가에 대한 부분이 어떻게 발전하게 되었는지 등 실제 사회에서 빅 데이터에 관심을 가질 수 밖에 없는 현상학적인 이유를 찾아본다. 예를 들어, '컴퓨터의 발전과 인터넷의 발전으로 데이터 베이스를 구축하는 서버가 발전하고, 그 서버를 많은 사람들이 이용하는 검색 엔진에서 이용하게 되고 그러한 매스 데이터가 쌓이게 되면서 자연스럽게 발전했을 것이다'라는 식의 컴퓨터 발전과 사회 발전의 관계를 분석한 논문 등을 참고하여 여러가지 사회 현상학적인 요인들을 결합하여 결과를 얻는다.

3. 트렌드 분석과 마케팅 활용 (실제 사용되는 기술 발전 추이와 사람들의 인식 설문조사) : 실제 기술이 발전된 추이를 보여주는 데이터를 보여주고, 사람들의 인식의 설문조사와 상관관계를 보여 준다. 이 경우, 년도 별 사람들의 설문 조사를 얻을 수 없을 것이다. 예를 들어, 2005년 사람들의 빅 데이터에 대한 인식을 확인하는 데이터는 얻을 수 없을 것이다.

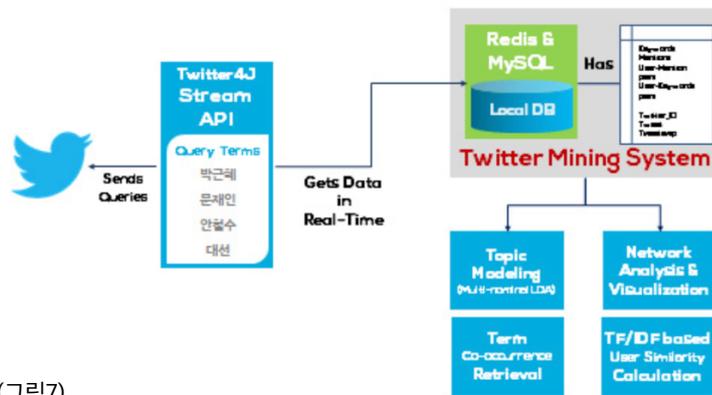
-> 따라서 이러한 경우는 이전에 빅 데이터라는 분야에 대해서 언급이 된 시점을 조사하여 그 당시 사람들이 빅 데이터에 대한 논문(2000년대)을 찾아보고, 어느정도의 예측으로 대신한다. 예를 들어, 아래 그림(2015년도 조사 자료)과 같이 업종별 미도입 사유에 대한 자료가 있다. 이 데이터가 현재(2016년)변화가 있다면 사람들이 혁신이라고 생각하는 빅 데이터의 상대적 이점, 적합성, 복잡성, 시험 가능성 등 다양한 요인의 변화가 있거나 혁신이 확산되는 단계가 변화 되었다고 볼 수 있다고 생각한다. 이처럼 기존의 다양한 데이터를 중심으로 기술 발전 추이와 함께 비교한다면 불가능한 설문조사나 부족한 질적조사를 대체할 수 있다고 생각한다.

[표 8] 업종별 빅데이터 미도입 사유 [중복 응답, 단위: 응답기업수]

구분	도입 관심 수준						
	빅데이터라고 할 만한 데이터가 없음	빅데이터 도입 효과에 대한 불신	관련 전문 인력 없음	CEO/CIO 무관심	빅데이터를 분석할 만큼 큰 기업이 아님	빅데이터 자체가 어떤 것인지 잘 모름	빅데이터 도입효과가 나타날 업무가 없음
공공	59	16	20	23	29	13	23
금융	6	4	2	6	3	2	3
유통/서비스	60	18	18	10	41	11	19
제조	119	26	26	60	74	36	59
의료	34	21	9	10	20	8	12
통신/미디어	11	7	3	3	9	2	6
Total	289	92	78	112	176	72	122

규모별로 살펴볼 경우 기업 규모가 작을수록 자사가 보유한 데이터양의 문제로 빅데이터 도입을 꺼리는 기업이 많았으며, 전문 인력의 문제 및 CEO/CIO의 무관심이 큰 것으로 나타났다. (그림6)

4. 빅 데이터라는 키워드 노출 빈도 확인(실제 노출되는 데이터를 추출) : 빅 데이터가 얼마나 자주 노출 되는지 확인하기 위해서 트위터 API를 이용해서 데이터를 추출한다. 트위터에서 데이터를 추출하는 이유는 트위터가 사람들이 가장 많이 실시간으로 자신의 생각을 쓰고, 텍스트로 되어 있는 자료이기 때문에 얼마나 많이 나오는지 확인하는 추이를 확인하기 용이하다고 생각했기 때문이다.



(그림7)

트위터API 사용 방식은 우선 트위터에서 데이터를 긁어서 가지고 온 후에 트위터 마이닝을 해주는 프로그램을 선정(자바, R 등)한 후에 그에 맞도록 데이터를 처리하여 빈도수 검사를 하여 사용한다. 실제로 트위터에서 데이터를 추출하여 검색하는 방식은 많이 사용되고 있기 때문에 추출하여 빈도를 찾아보는 빈도수 검사의 경우 어려움이 없을 것이라고 생각한다.

5. 기업의 빅 데이터를 이용 유무의 차이로 인한 매출 차이(데이터 분석을 하는 기업과 아닌 기업의 매출차이): 우선 데이터를 실제로 얼마만큼 보유하고 있는지 그 추이를 확인하고, 그것을 분석할 수 있는 능력의 차이를 조사한다. 그 조사를 통해서 기업의 매출의 차이가 그러한 데이터의 차이와 가공 기술 및 활용 능력의 차이로 인해 매출이 차이가 난다는 것을 보여 준다. 이러한 방법으로 우선 기존의 기업을 분석한 다양한 자료를 찾아본다. 이러한 자료는 우선 비교 할 기업을 선정해야 하는데, 실제로 빅 데이터를 이용하여 큰 매출을 이룬 회사는 IT 분야와 마케팅 분야가 많을 것이라고 생각하여 그러한 기업을 중심으로 선정한다. 앞서 이야기 한 것 처럼 대기업이 아닌 중소기업으로 통제하여 조사한다.

■ 미국 산업별 기업당 평균 데이터 보유량



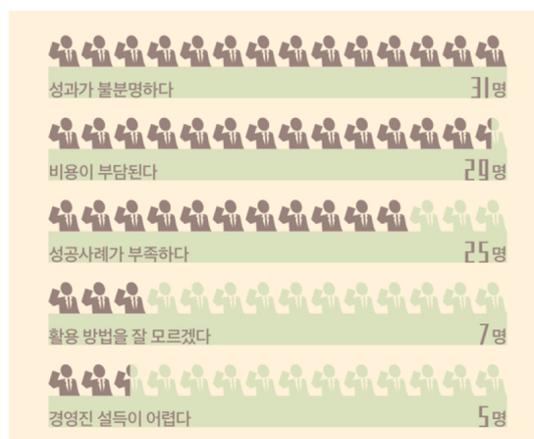
자료 : McKinsey&Company, 2011.06

(그림 8)

다음 그림은 2011년 미국 산업별 평균 데이터 보유량이다. 이와 같이 산업 별 데이터를 얼마만큼 보유 하고 있는지를 시대별로 정리하고 가장 많은 분야가 가장 데이터를 활용한다는 근거가 있다면 그 분야의 매출이 빠르게 증대한다는 것을 보여주는 방법도 있다고 생각한다. (그림 4).

3. 빅데이터 분석 솔루션을 마케팅에 활용한다고 가정했을 때, 가장 염려되는 점은 무엇인가?

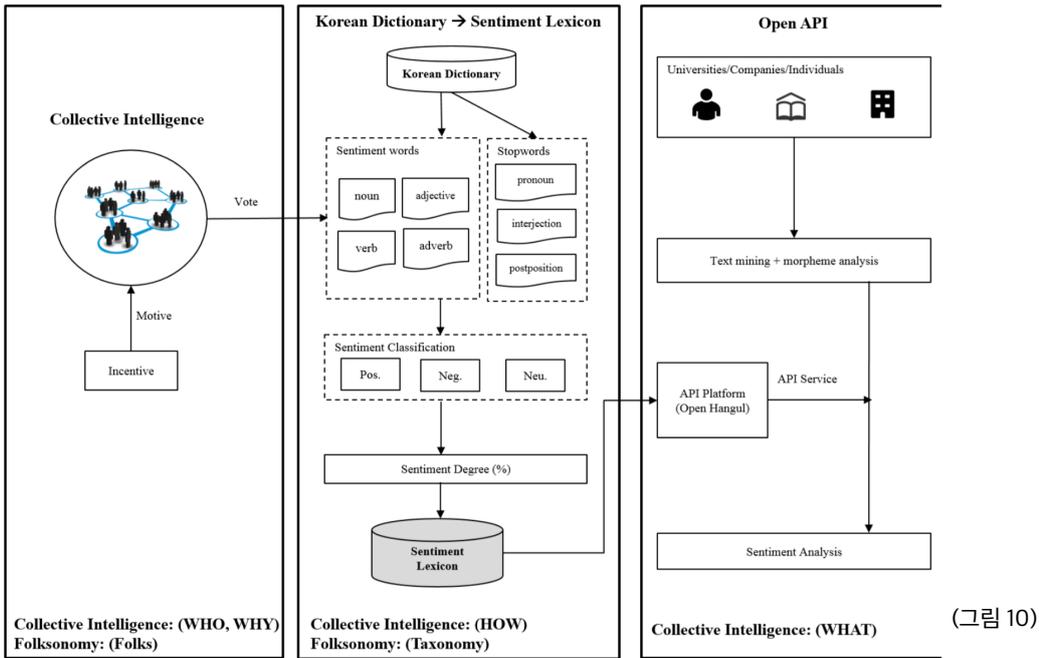
조사결과 마케팅들은 '성과가 불분명하다는 점', '비용 부담', '성공사례의 부족' 등을 염려하는 것으로 나타났다. 빅데이터 마케팅에 갈라잡이가 될 다양한 사례가 더욱 요구되는 시점이다.



(그림 9)

또한 실제로 적용되어야 하는 분야의 종사자의 설문조사 자료를 통해서 어떻게 변화하고 있는지를 조사한다면 양적인 데이터를 중심으로 하지만 질적인 부분 까지도 어느정도 충족할 수 있다고 생각한다. 하지만 이러한 데이터의 경우 없는 경우가 있을 수 있기 때문에 참고자료 정도로 이용하는 편이 좋을 것이라고 판단한다.

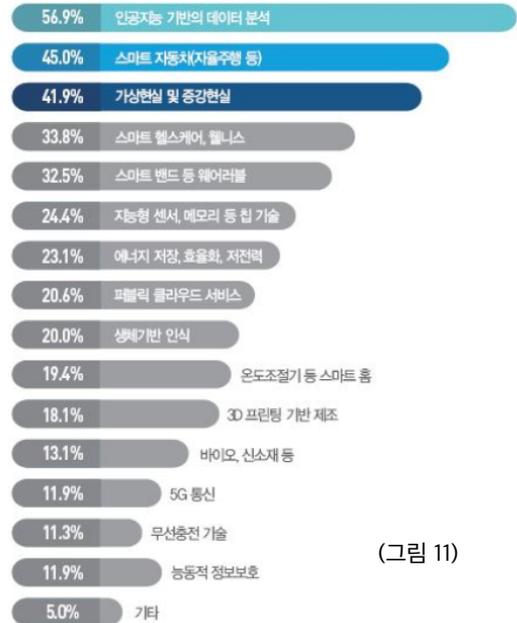
6. 오피니언 리더들의 언급과 주장 (사람들의 설문조사 & 실제 언급에 대한 횟수): 실제로 오피니언 리더들이 얼마나 언급 했는지 데이터를 얻기 위해서 가장 먼저 트위터의 오피니언 리더를 중심으로 확인한다. 그 방식으로 집단 지성과 '빅 데이터'라는 단어가 얼마나 연결되어 이루어지고 있는지를 확인한다. 또한 빅 데이터 관련 서적과 논문이 얼마나 많이 출판되고 있는지 조사한다. 또한 그 판매량이 얼마만큼인지를 확인한다. 이러한 자료를 바탕으로 전문가들이 얼마나 많이 빅 데이터를 언급하고 있는지 확인할 수 있고, 사람들이 얼마나 관심을 갖고 있는지를 서적의 판매량을 통해서 확인해 본다. 아래의 그림은 트위터에서 집단 지성과 단어를 어떻게 관계 하는지 보여주는 과정이다.(그림 6)



(그림 10)

이 외에도, 언론에서 빅 데이터를 정의하는 다양한 자료가 얼마나 많이 나오는지 확인해 본다. 의제설정이론(Agenda Setting)에 근거하여 이야기 하면 미디어의 의제설정이 실제로 공론화 되는 경우가 많은데, 빅 데이터의 경우에도 그런 의제 중 하나일 수도 있기 때문이다. 미디어가 새로운 기술 중 빅 데이터에 주목을 하고, 그것에 대해서 많이 언급함과 동시에 사람들에게 특정한 이미지를 부여했다면 그것 또한 하나의 가설 설정을 하여 확인해 볼수 있다고 생각한다. 하지만 이것 자체를 하나의 가설로 설정하여 주목하기 보다는 다중회귀분석의 하나의 요인으로 사용하면 좋을 것이라고 사료된다. (그림 7)

올해 부상할 핵심 기술 (복수응답)



(그림 11)

<데이터 수집 설계 및 측정>

앞서 많은 가설을 만들고 그에 따라서 변인을 설정했다. 하지만 우리 조가 변인을 측정하고자 했을 때 11조에서 얘기한 문제인 '변인에 대한 구체화' 즉, 조작적 정의가 정확하지 않다고 판단하여 그러한 판단을 조금 더 명확하게 하고자 하였다. 또한 다양한 API를 사용하는데에는 기술적인 한계가 되었다. 결과적으로 우리가 필요한 변인을 측정하기 위해 Mass Media의 DB와 기업인을 대상으로 하는 설문조사, 일반인을 대상으로 하는 심층 인터뷰를 통해서 미디어의 노출, 오피니언 리더의 대인채널, 일반인의 의식 및 기업인들의 의식 등을 조사해 보았다.

-Mass Media

Mass Media

각각의 DB를 이용하여 년도 별 미디어의 노출 건수의 추이를 확인한다.



Book(교보문고 DB+ 국립중앙도서관 단행자료 DB)

(왼쪽 부터 그림 12,13,14,15) 각종 미디어 DB자료



News(Daum기사 DB)



Research(Naver 검색 DB)

우선, 우리가 필요하다고 생각하는 데이터를 모았다. 1차 자료를 얻는 것도 중요하지만 현재 학생이라는 신분으로 많은 데이터를 얻는다는 것이 시간적으로나 비용 적으로 불가능 했기 때문이다. 따라서 tv, 판매 도서 자료, 단행본, 온라인검색 등 실제로 사람들에게 얼마나 노출 되었는지를 판단하기 위해서는 사람들이 가장 많이 접하는 매체를 중심으로 그 양이 얼마나 늘었는지 보는 것이 중요하다고 판단하였다. 따라서 이와 같은 자료를 모으기 위해 교보문고db 와 국립중앙도서관 단행자료 db, 다음 뉴스 기사 db, Naver 검색 db 등의 년도 별 데이터 양을 판단하여 다각도로 매스 미디어 양의 변화를 판단하였다.

-Interview

Focus Interview

일반인을 대상으로 인터뷰를 통해 빈 데이터에 대한 노출과 인식정도를 확인한다

(왼쪽 부터 그림 16,17,18,19) Interviewee



20대 대학생 남자
(경영학과)



20대 대학생 여자
(일어일문학과)



50대 여자
(주부)

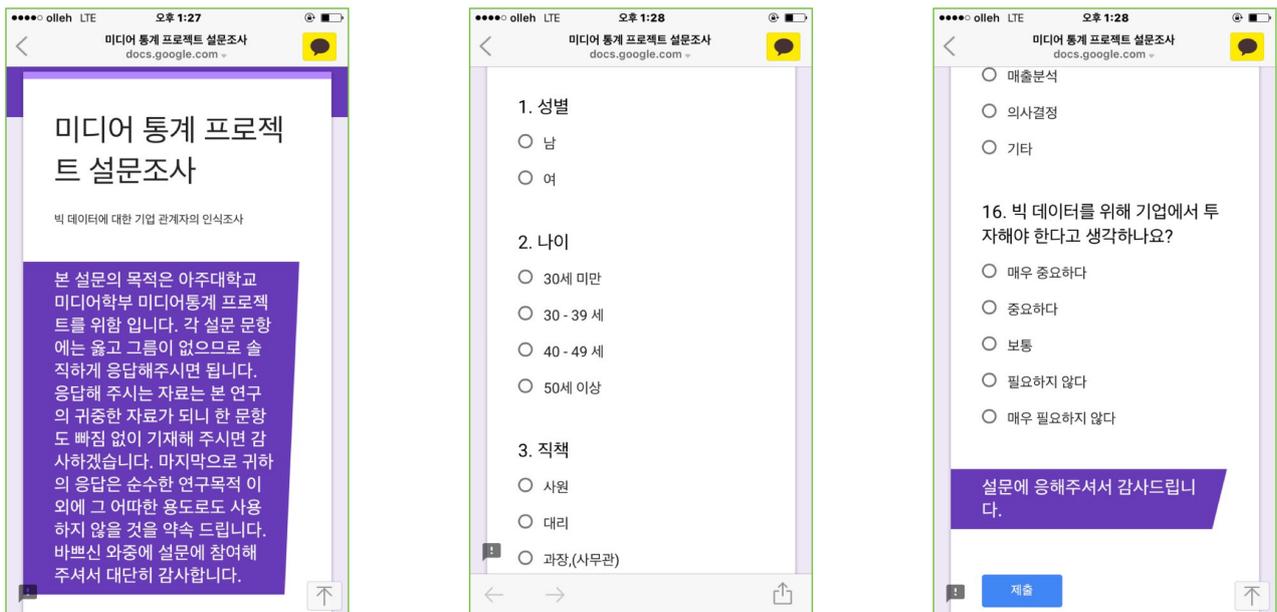


50대 남자
(의사)

다음으로, 사람들의 의견을 심층적으로 들어보아야 한다고 판단했다. 물론, 여기서 샘플은 미디어학과와 같이 관련 분야를 종사하는 사람이 아닌 나이별, 성별, 직업별 다양한 직종의 사람이 적합하다고 판단하였다. 그 이유는 미디어학과 학생의 경우, 기본적으로 공부를 하면서 많은 사람들에게 빅 데이터라는 분야를 접하고 중요성에 대해서 많이 듣기 때문에 그 인식이 충분히 각인 되었을 것이라고 판단했기 때문이다. 따라서 20대 대학생 중에서는 경영학과 3학년 학생과 일어일문학과 3학년 여학생을 선정 하였고, 50대 주부와 50대 직업이 의사인 남자를 선정하였다. 이외에도 단편적인 인터뷰(7명)를 했으나 그 사람들과의 인터뷰는 단편적인 내용과 짧은 시간 이었기 때문에 생략하였다. 총 4명의 사람과는 직접 문항을 만들어서 딱딱하게 이야기하기 보다는 사람들이 어떻게 생각하는지, 그 생각을 따라서 깊은 질문을 하는 방식으로 하였다. 이러한 방식으로 사람들이 자신이 가지고 있는 이야기를 편안하게 하도록 유도하였다. 인터뷰는 특정 장소나 특정 시간대에 만나서 실시하는 FGI로 진행된 것이 아니라 각각의 사람들과 약 30분에서 1시간 정도로 진행하였다. 인터뷰를 하는 동안 최대한 편안하게 하여 진솔한 이야기를 유도 하였고, 중립적인 자세로 인터뷰에 임하였다.

Questionnaire

실제 업무 종사자들이 생각하는 빅 데이터에 대한 노출과 중요도를 파악하기 위해 특정 집단을 위한 설문지를 작성한다.



-Questionnaire

(그림 20) App에서의 설문조사 프로세스

마지막으로 기업인을 중심으로 하는 설문지를 작성하였다. 실제로 다양한 업무에 종사하는 사람들이 어떻게 생각하는지를 알아야 실제 기업에서 전개하고 있는 변화가 얼마나 의미가 있는지 판단할 수 있기 때문이다. 이와 같이 판단한 이유는 현재 많은 기업에서 빅 데이터 중심의 의사 결정 시스템을 구축하고 있는데 그러한 판단이 실제 기업에서 종사하는 사람들의 업무에서 오는 불편함이나 필요성 때문인지, 아니면 단순히 언론이나 다양한 매체를 통해서 스스로 접하면서 중요하다는 인식이 강하게 생겨 이와같은 판단을 한 것인지 구별하기 위함이다. 따라서 설문지는 다양한 기업규모, 나이, 직책 등 다양한 업무와 다양한 포지션에 위치한 사람들의 다양한 의견을 얻기 위해 설문 문항으로 넣었고, 실제 업무에서 얼마나 중요하게 생각하는지, 그리고 그 생각만큼 빅 데이터를 알거나 쓰고 있는지 등을 설문지에 포함시켰다.

설문 문항

미디어 통계 프로젝트 설문조사

빅 데이터에 대한 기업 관계자의 인식조사

본 설문지의 목적은 아주대학교 미디어학부 미디어통계 프로젝트를 위함입니다. 각 설문 문항에는 옳고 그름이 없으므로 솔직하게 응답해주시면 됩니다. 응답해 주시는 자료는 본 연구의 귀중한 자료가 되니 한 문항도 빠짐 없이 기재해 주시면 감사하겠습니다. 마지막으로 귀하의 응답은 순수한 연구목적 이외에 그 어떠한 용도로도 사용하지 않을 것을 약속 드립니다. 바쁘신 와중에 설문에 참여해 주셔서 대단히 감사합니다.

설명(선택사항)

1. 성별

- 남
- 여

2. 나이

- 30세 미만
- 30 - 39 세
- 40 - 49 세
- 50세 이상

3. 직책

- 사원
- 대리
- 과장(사무관)
- 자장, 부장
- 팀장(국장)
- 임원

4. 기업규모

- 대기업
- 중소기업
- 정부기업, 협회
- 기타

5. 기업 업종

- 제조
- 금융,보험
- 유통,서비스
- 통신,IT
- 그룹(종합)
- 기타

6. 기업 형태

- B2C
- B2B
- 기타

7. 당신은 빅 데이터의 의미를 알고 있나요?

- 정확한 프로세스를 알고 있다
- 어떤 의미인지 정확히 알고 있다
- 단어가 의미하는 것이 무엇인지 정도 알고 있다
- 어디서 들어본 기억은 있지만 실제로 어떤 의미인지는 모르겠다
- 처음 들어본다

8. 당신은 어떠한 경로로 빅 데이터를 알게 되었나요?

- TV
- 신문, 라디오, 서적
- 인터넷
- 주변 사람
- 기타
- 모름

9. 당신은 빅 데이터가 중요하다고 생각하시나요?

- 매우 중요하다
- 중요하다
- 보통이다
- 중요하지 않다
- 매우 중요하지 않다

10. 당신은 실제 업무에서 빅 데이터를 사용하시나요?

- 사용한다(12번 문항 생략)
- 사용하지 않는다(11번 문항 생략)

11. 사용하신다면, 어느정도 활용하시나요?

- 가공된 자료를 실제 업무의 예로 활용
- 간단한 API를 이용하여 데이터를 분석 및 활용
- 실제로 필요한 데이터를 모아 분석 및 활용
- 다양한 데이터와 2차 자료를 통합적으로 분석 및 활용
- 사용하지 않는다

12. 사용하지 않는다면, 왜 사용하지 않으시나요?

- 빅 데이터라는 개념 자체를 정확하게 알지 못한다
- 빅 데이터를 이용하는 것은 내가 할 수 있는 분야가 아니다
- 빅 데이터를 이용하는 방법(프로그램 사용)이 어렵다
- 데이터를 모으는 것이 현실적으로 힘들다고 생각한다
- 실제 업무와 크게 관련이 없다고 생각한다.
- 기타

13. 자신의 업무에서 빅 데이터가 필요하다고 생각하나요?

- 매우 필요하다
- 필요하다
- 보통이다
- 필요하지 않다
- 매우 필요하지 않다(14번 항목 생각하세요)

...

14. 자신에 업무에 필요하다면 그 이유가 무엇인가요?

- 실제 업무와 직결되기 때문에(기초 자료)
- 매출과 연관되기 때문에
- 사람들이 필요하다는 말을 많이 듣기 때문에
- 언론에서 중요하다는 말을 많이 듣기 때문에
- 기타

15. 빅 데이터가 가장 필요한 업무가 무엇이라고 생각하나요?

- 자료 조사 및 분석
- 마케팅
- 매출분석
- 의사결정
- 기타

16. 빅 데이터를 위해 기업에서 투자해야 한다고 생각하나요?

- 매우 중요하다
- 중요하다
- 보통
- 필요하지 않다
- 매우 필요하지 않다

설문에 응해주셔서 감사드립니다.

설문(선택사항)

실제 데이터를 얻은 결과

제목 없는 설문지(응답) ☆ ■

파일 수정 보기 삽입 서식 데이터 도구 설문지 부가기능 도움말

타입스탬프

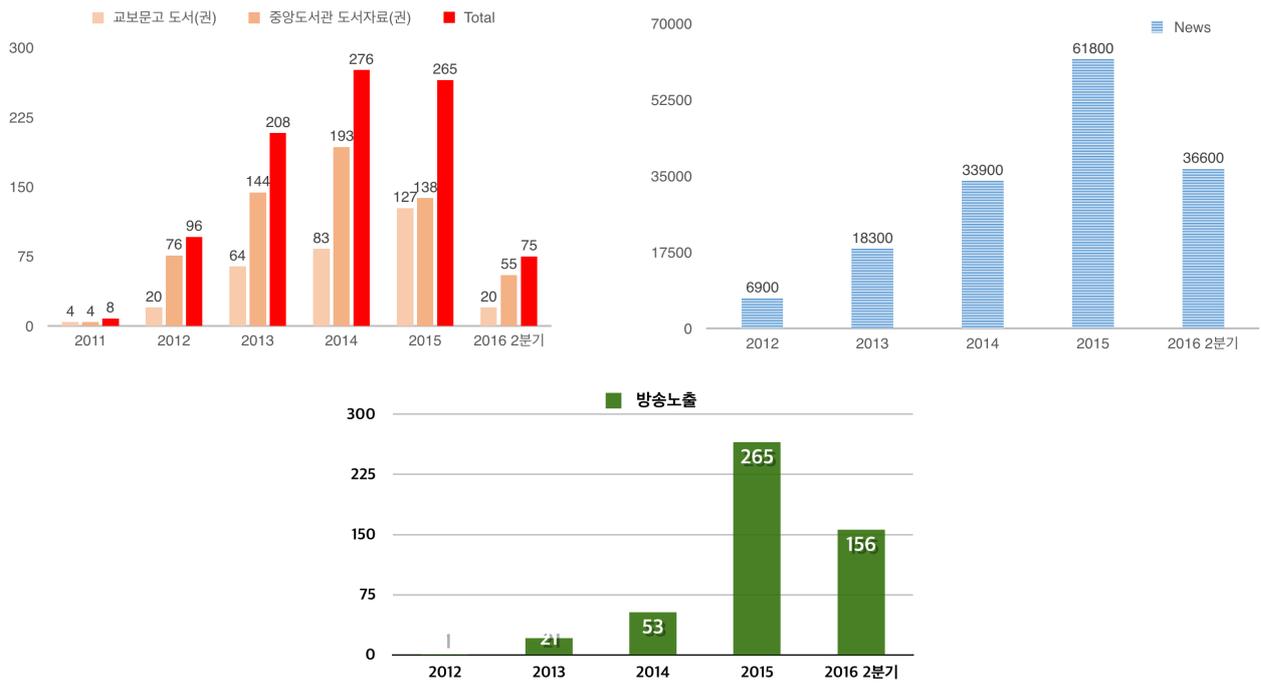
	A	B	C	D	E	F	G	H	I
1	타입스탬프	1. 성별	2. 나이	3. 직책	4. 기업규모	5. 기업 업종	6. 기업 형태	7. 당신은 빅 데이터의 의미를	8. 당신은 어떠한 정도로
2	2016. 6. 4 오후 12:43:33	남	30세 미만	과장,(사무관)	대기업	통신, IT	B2B	단어가 의미하는 것이 무엇인 TV	
3	2016. 6. 4 오후 12:44:20	남	30 - 39 세	팀장,(국장)	중소기업	통신, IT	B2C	단어가 의미하는 것이 무엇인 인터넷	
4	2016. 6. 4 오후 12:45:26	여	40 - 49 세	대리	정부기업, 협회	기타	기타	어디서 들어본 기억은 있지만 TV	
5	2016. 6. 4 오후 12:46:18	남	50세 이상	임원	대기업	유통, 서비스	B2B	어디서 들어본 기억은 있지만 신문, 라디오, 서적	
6	2016. 6. 4 오후 12:47:11	남	40 - 49 세	차장, 부장	중소기업	금융,보험	B2C	어떤 의미인지 정확히 알고 잇 인터넷	
7	2016. 6. 4 오후 12:48:16	여	30세 미만	사원	중소기업	제조	B2C	정확한 프로세스를 알고 있다 인터넷	
8	2016. 6. 4 오후 12:49:04	남	40 - 49 세	과장,(사무관)	대기업	유통, 서비스	B2B	처음 들어본다	모름
9	2016. 6. 4 오후 12:51:56	여	50세 이상	차장, 부장	정부기업, 협회	통신, IT	B2B	어디서 들어본 기억은 있지만 TV	
10	2016. 6. 4 오후 12:53:31	남	30 - 39 세	대리	대기업	그룹(종합)	B2B	정확한 프로세스를 알고 있다 인터넷	
11	2016. 6. 4 오후 12:54:07	남	30세 미만	사원	중소기업	통신, IT	B2C	정확한 프로세스를 알고 있다 TV	
12	2016. 6. 4 오후 12:55:19	남	30 - 39 세	사원	정부기업, 협회	기타	기타	어떤 의미인지 정확히 알고 잇 신문, 라디오, 서적	
13	2016. 6. 4 오후 12:55:54	여	30 - 39 세	대리	중소기업	통신, IT	B2C	정확한 프로세스를 알고 있다 신문, 라디오, 서적	
14	2016. 6. 4 오후 12:56:31	여	40 - 49 세	과장,(사무관)	대기업	그룹(종합)	B2B	어떤 의미인지 정확히 알고 잇 TV	
15	2016. 6. 4 오후 12:57:52	남	50세 이상	팀장,(국장)	대기업	통신, IT	B2B	어디서 들어본 기억은 있지만 신문, 라디오, 서적	
16	2016. 6. 4 오후 12:58:45	여	30세 미만	대리	중소기업	그룹(종합)	B2C	정확한 프로세스를 알고 있다 TV	
17	2016. 6. 4 오후 1:01:37	여	40 - 49 세	차장, 부장	정부기업, 협회	유통, 서비스	B2B	어떤 의미인지 정확히 알고 잇 TV	
18	2016. 6. 4 오후 1:05:33	남	50세 이상	팀장,(국장)	정부기업, 협회	그룹(종합)	B2B	단어가 의미하는 것이 무엇인 주변 사람	
19	2016. 6. 4 오후 1:05:42	여	30세 미만	사원	중소기업	통신, IT	B2C	어떤 의미인지 정확히 알고 잇 인터넷	
20	2016. 6. 4 오후 1:06:05	여	40 - 49 세	차장, 부장	중소기업	금융,보험	B2B	단어가 의미하는 것이 무엇인 인터넷	
21	2016. 6. 4 오후 1:06:42	남	30 - 39 세	대리	중소기업	유통, 서비스	B2C	어떤 의미인지 정확히 알고 잇 신문, 라디오, 서적	
22	2016. 6. 4 오후 1:06:47	남	30세 미만	대리	대기업	그룹(종합)	B2B	정확한 프로세스를 알고 있다 TV	
23	2016. 6. 4 오후 1:08:11	남	30세 미만	사원	정부기업, 협회	기타	기타	정확한 프로세스를 알고 있다 인터넷	
24	2016. 6. 4 오후 1:09:01	남	30 - 39 세	과장,(사무관)	중소기업	유통, 서비스	기타	어디서 들어본 기억은 있지만 인터넷	
25	2016. 6. 4 오후 1:09:09	여	30 - 39 세	대리	정부기업, 협회	통신, IT	B2B	정확한 프로세스를 알고 있다 TV	
26	2016. 6. 4 오후 1:10:13	여	30세 미만	대리	대기업	금융,보험	기타	단어가 의미하는 것이 무엇인 TV	
27	2016. 6. 4 오후 1:11:07	남	40 - 49 세	대리	정부기업, 협회	통신, IT	B2B	어떤 의미인지 정확히 알고 잇 신문, 라디오, 서적	
28	2016. 6. 4 오후 1:11:43	여	30 - 39 세	팀장,(국장)	중소기업	유통, 서비스	B2B	단어가 의미하는 것이 무엇인 인터넷	
29	2016. 6. 4 오후 1:12:42	여	30 - 39 세	임원	정부기업, 협회	통신, IT	B2C	정확한 프로세스를 알고 있다 인터넷	
30	2016. 6. 4 오후 1:14:06	남	30세 미만	사원	대기업	제조	기타	단어가 의미하는 것이 무엇인 신문, 라디오, 서적	
31	2016. 6. 4 오후 1:14:11	남	40 - 49 세	대리	정부기업, 협회	통신, IT	B2B	어떤 의미인지 정확히 알고 잇 신문, 라디오, 서적	
32	2016. 6. 4 오후 1:15:11	여	30세 미만	팀장,(국장)	중소기업	유통, 서비스	B2C	어떤 의미인지 정확히 알고 잇 인터넷	

<데이터 분석 및 결과>

-Mass Media

Mass Media

데이터 분석을 통해 2012년부터 미디어의 노출이 빠르게 증가하고 있음을 알 수 있었다.



(왼쪽 위부터 그림 21,22,23,24) 각종 DB 분석 결과

인터넷 뉴스, 방송, 도서 등 매스 미디어에서 빅 데이터를 키워드로 하는 자료의 양을 분석하였다. 분석 결과, 2012년 이후로 빠르게 증가하는 것을 볼 수 있다. 현재(2016년 2분기)까지 자료를 기준으로 보았을 때, 계속해서 미디어의 노출 양이 증가한다는 사실을 알 수 있었다. 이러한 상황으로 분석해 볼 때 초기의 경우, 빅 데이터를 미디어에서 인위적으로 노출 하였지만, 시간이 지나면서 그러한 의제가 공중 의제가 되었고 결과적으로 많은 사람들이 그 개념에 대해서 알고자 하는 욕구가 생겨나면서 자연스럽게 언론에서 더 많은 양의 정보를 다루게 되었다고 추정할 수 있었다. 특히, 인터넷 뉴스의 경우에는 2012년에 6900건 정도의 기사에서 2015년에 61800건의 기사로 약 9배 이상의 양이 늘어난 것을 볼 수 있다. 이와 같이 미디어 의제가 더욱 늘어남에 따라 당분간 빅 데이터에 대한 사람들의 공중 의제 또한 유지 될 것으로 예측한다.

-Focus Interview

심층 인터뷰 결과, 미디어의 노출과 오피니언 리더가 주는 영향력이 크다는 사실을 알 수 있었다. 특히, 새로운 사실은 오피니언 리더 혹은 대인 채널에 따라서 그 개념에 대한 평가가 다르게 이루어 진다는 점이다. 예를 들어, 20대 일어를 전공 하는 여학생의 경우, 공대를 다니는 남자친구(대인 채널)가 앞으로 인문대 생이 하는 다양한 직군이 빅 데이터로 인해 사라진다는 이야기를 듣고 현재 취업으로 인해 걱정이 많은 시기적인 특성과 맞물려 빅 데이터에 대한 인식을 부정적으로 하게 되었다. 이처럼 단순히 오피니언 리더가 혁신을 확산시키는 것을 넘어서 그 개념에 대한 긍정적, 부정적인 정도까지도 영향을 주는 것을 알 수 있었다.

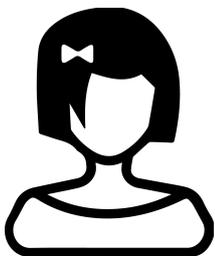


(그림 25) 오피니언 리더에 따른 빅 데이터의 인식 차이



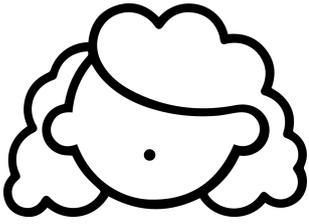
20대 대학생 남자
(경영학과)

빅 데이터에 대한 정보는 대학교 1학년(2012)년 수업시간에 교수님(Opinion Leader)께서 이야기하셔서 알게 되었다. 교수님께서 말씀하신 것이 너무 인상적이어서 그 이후로 군대에서 빅 데이터 관련 서적을 읽고, 제대하면 공부하고 싶다는 생각을 했다. 지금은 4학년인데 간단한 분석정도를 할 수 있다. 앞으로 경영학을 하면서 데이터 분석이 반드시 필요하다고 생각한다. 트위터와 같은 텍스트를 분석하는 능력을 기르고 싶다. 솔직히 말하면 빅 데이터라는 것이 말이 어렵지 분석을 하는데 더 많은 데이터를 쓰는 정도라고 생각한다. 충분히 문과생들도 공부를 통해 할 수 있는 분야라고 생각한다.



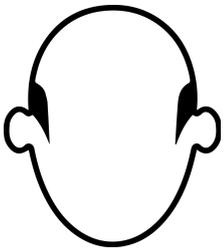
20대 대학생 여자
(일어일문학과)

빅 데이터라는 단어를 접한 것은 남자친구이다. 남자친구의 전공이 컴퓨터 공학인데 과제를 하는 모습을 지켜 보면서 빅 데이터에 대해서 알게 되었다. 솔직히 빅 데이터가 무엇인지도 모르겠고, 왜 쓰는지 잘 모르겠다. 트위터 같은 것을 분석한다고 하는데 트위터를 어떻게 분석하는지도 모르겠고, 아무나 막 올리는 글이 무슨 의미가 있는지도 모르겠다. 이번에 알파고(정확한 이름은 몰랐음)가 이기는 것을 보고 정말 필요하다는 생각보다는 무섭고, 내가 앞으로 할 일이 없어지는 것은 아닌가 하는 마음에 취업에 대해서 더 걱정하게 되었다. 남자친구가 농담처럼 이제 내가 할 일이 없겠다는 말에 많이 심란했다.



50대 여자
(주부)

빅 데이터라는 단어를 ‘아침마당’에서 본 것 같다. 강연자가 재미있게 이야기를 해줘서 어떤 내용인지 대충은 안다. 그리고 실제로 사용된 예시가 재미있었던 기억이 난다. 그것을 보고 아들에게 빅 데이터에 대해서 공부하라고 얘기했던 기억이 난다. 요즘 tv에서 빅 데이터라는 단어를 많이 들었던 것 같다. 나는 앞으로 배워서 써먹을 수 없지만 젊은이들은 꼭 알아야 하는 분야가 아닌가 하는 생각도 든다. 그때 강연자가 주부들의 고충을 모아서 그것을 통해서 새로운 기계들을 만들어서 주부들을 편하게 할 수 있다는데 그런게 있나 궁금하다. 실제로 아직까지 삶에 큰 영향을 주는 것 같지는 않다.



50대 남자
(의사)

의사: 약 3~4년 전에 빅 데이터에 대한 책을 구매해서 접했다. 그 전에 접했던 것 같긴 한데 정확히 기억이 안난다. 최근에 알파고의 승리를 뉴스로 접하면서 검색을 했던 기억이 난다. 어떤 식으로 돌아가고 있는지 자세히는 모르지만 중요한 것이고 미래를 바꿀 수 있는 중요한 개념임에는 틀림없다. 아마도 빅 데이터 분석이 활성화 되고 있는 지금이 새로운 패러다임의 등장이라고 생각한다. 의학에서 어떻게 사용될지에 대해서 생각하기 보다는 회계사나 변호사가 하는 일들을 할 수 있을것이라고 생각한다. 의학의 경우, 인간이 직접 할 수 밖에 없는 이유가 특수한 변수가 있는데 그것은 ‘경험’을 통해서만 알고 인간이 가진 통찰력이 필요하다고 생각하기 때문이다. 시간이 더 지나서 미래에는 병이 걸리지 않고 미리 진료, 회복하는 ‘예방 의학’분야가 발달하여 많은 의사가 필요 없을 것이라고 생각한다.

-Questionnaire

Questionnaire

결과적으로 스스로 중요하고 생각하는 질문(약 47%) 외에는 유의미한 상관 관계를 얻지 못했다. 특히, 자신의 업무에서 필요하다고 생각하는지에 대한 질문이 상관 관계를 얻지 못한 것은 예상하지 못한 결과다.(데이터를 직접 입력했기 때문)

우리가 원하는 결과를 얻지 못했다. 그 이유는 실제로 샘플링(Sampling)을 통해 사람들을 모으고, 그 사람들을 대상으로 설문조사를 한 것이 아니라 우리가 임의적으로 약 40명의 데이터를 만들어서 분석을 했기 때문이다. 따라서 이 설문을 통해 이루어진 SPSS 분석은 실제로는 하나도 의미가 있는 값이 나오지 않았다. 하지만 SPSS를 실제로 설치하고, 데이터 분석을 위해 사용할 수 있는 데이터로 바꾸고, 그 데이터와 독립변인을 중심으로 여러가지 회귀 분석을 해보는 과정을 통해서 추후에 데이터를 어떻게 설문을 해야하고, 어떤 식으로 변환해야하는지 등의 기술적인 학습을 할 수 있었다.

<SPSS 과정>

성별	나이	직책	기업규모	기업업종	기업형태	당신은빅데이터의이름을알고있나요	당신은어떤환경으로빅데이터를알게되었나요	당신은빅데이터가중요하다고생각하시나요	당신은실제업무에서빅데이터를사용하시나요	당신의업무에서빅데이터가필요하다고생각하시나요	빅데이터가장필요한업무가무엇이라고생각하시나요	빅데이터가장필요한업무가무엇이라고생각하시나요		
1	남	30세 미만	과장(사...)	대기업	통신, IT	B2B	언어가 의미하는 것이 무...	TV	매우 중요하다	사용한다(12번)	매우 필요하다	해설과 연관되...	자료 조사 및 분석	매우 중요하다
2	남	30 - 39 세	팀장(국...)	중소기업	통신, IT	B2C	언어가 의미하는 것이 무...	인터넷	중요하다	사용하지 않는...	보통이다	실제 업무와 ...	매출분석	중요하다
3	여	40 - 49 세	대리	중소기업	기타	기타	어디서 물어본 기억은 있...	TV	보통이다	사용하지 않는...	보통이다	기타	자료 조사 및 분석	필요하지 않
4	남	50세 이상	팀장(사...)	대기업	유통, 서...	B2B	어디서 물어본 기억은 있...	신문, 라디오, 서적	중요하지 않다	사용하지 않지...	매우 필요하지...	기타	매출분석	필요하지 않
5	남	40 - 49 세	차장, 부...	중소기업	금융, 보험	B2C	어떤 의미인지 정확히 알...	인터넷	매우 중요하다	사용한다(12번)	매우 필요하다	실제 업무와 ...	매출분석	매우 중요하
6	여	30세 미만	사원	중소기업	제조	B2C	정확한 프로세스를 알고...	인터넷	매우 중요하다	사용한다(12번)	매우 필요하다	실제 업무와 ...	의사결정	매우 중요하
7	남	40 - 49 세	과장(사...)	대기업	유통, 서...	B2B	처음 물어본다	모름	매우 중요하지 않다	사용하지 않는...	매우 필요하지...	기타	매우 중요하	
8	여	50세 이상	차장, 부...	중소기업	통신, IT	B2B	어디서 물어본 기억은 있...	TV	중요하지 않다	사용하지 않는...	매우 필요하지...	기타	매우 중요하	
9	남	30 - 39 세	대리	대기업	그물(중합)	B2B	정확한 프로세스를 알고...	인터넷	매우 중요하다	사용한다(12번)	필요하다	연론에서 중요	마케팅	매우 중요하
10	남	30세 미만	사원	중소기업	통신, IT	B2C	정확한 프로세스를 알고...	TV	매우 중요하다	사용한다(12번)	매우 필요하다	연론에서 중요	마케팅	중요하다
11	남	30 - 39 세	사원	중소기업	기타	기타	어떤 의미인지 정확히 알...	신문, 라디오, 서적	중요하다	사용한다(12번)	필요하다	연론에서 중요	자료 조사 및 분석	중요하다
12	여	30 - 39 세	대리	중소기업	통신, IT	B2C	정확한 프로세스를 알고...	신문, 라디오, 서적	매우 중요하다	사용한다(12번)	매우 필요하다	연론에서 중요	마케팅	매우 중요하
13	여	40 - 49 세	과장(사...)	대기업	그물(중합)	B2B	어떤 의미인지 정확히 알...	TV	보통이다	사용한다(12번)	보통이다	연론에서 중요	자료 조사 및 분석	보통
14	남	50세 이상	팀장(국...)	대기업	통신, IT	B2B	어디서 물어본 기억은 있...	신문, 라디오, 서적	보통이다	사용하지 않는...	필요하다	연론에서 중요	의사결정	필요하지 않
15	남	30세 미만	대리	중소기업	그물(중합)	B2C	정확한 프로세스를 알고...	TV	중요하다	사용한다(12번)	필요하다	연론에서 중요	자료 조사 및 분석	매우 중요하
16	남	40 - 49 세	차장, 부...	중소기업	유통, 서...	B2B	어떤 의미인지 정확히 알...	TV	보통이다	사용한다(12번)	보통이다	연론에서 중요	매출분석	중요하다
17	남	50세 이상	팀장(국...)	중소기업	그물(중합)	B2C	언어가 의미하는 것이 무...	주변 사람	중요하지 않다	사용하지 않는...	보통이다	기타	마케팅	필요하지 않
18	여	30세 미만	사원	중소기업	통신, IT	B2C	어떤 의미인지 정확히 알...	인터넷	매우 중요하다	사용한다(12번)	필요하다	연론에서 중요	자료 조사 및 분석	매우 중요하
19	여	40 - 49 세	차장, 부...	중소기업	금융, 보험	B2B	언어가 의미하는 것이 무...	인터넷	보통이다	사용하지 않는...	보통이다	기타	마케팅	보통
20	남	30 - 39 세	대리	중소기업	유통, 서...	B2C	어떤 의미인지 정확히 알...	신문, 라디오, 서적	매우 중요하다	사용한다(12번)	매우 필요하다	사람들이 필요	자료 조사 및 분석	매우 중요하

1. 우리가 설문한 데이터(Google 설문지)를 엑셀 데이터로 전환한다.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	성별	String	3	0	1. 성별	{1, male}	None	3	☰ Left	Nominal	Input
2	나이	String	12	0	2. 나이	{1, under 30}	None	3	☰ Left	Nominal	Input
3	직책	String	18	0	3. 직책	{1, Staff}	None	3	☰ Left	Nominal	Input
4	기업규모	String	20	0	4. 기업규모	{1, Conglom}	None	6	☰ Left	Nominal	Input
5	기업업종	String	17	0	5. 기업 업종	{1, manufac}	None	6	☰ Left	Nominal	Input
6	기업형태	String	6	0	6. 기업 형태	{1, B2C}	None	6	☰ Left	Nominal	Input
7	당신은빅데이터의이름을알고있나요	String	85	0	7. 당신은 빅...	{1, use}	None	10	☰ Left	Ordinal	Input
8	당신은어떤환경으로빅데이터를알게되었나요	String	25	0	8. 당신은 어...	{1, TV}	None	10	☰ Left	Nominal	Input
9	당신은빅데이터가중요하다고생각하시나요	String	26	0	9. 당신은 빅...	{1, use}	None	10	☰ Left	Ordinal	Input
10	당신은실제업무에서빅데이터를사용하시나요	String	43	0	10. 당신은 어...	{1, use}	None	8	☰ Left	Nominal	Input
11	자신의업무에서빅데이터가필요하다고생각하시나요	String	56	0	11. 자신의 ...	{1, use}	None	12	☰ Left	Ordinal	Input
12	자신의업무에서필요하다면그이유가무엇인가요	String	59	0	12. 자신의 ...	{1, relation}	None	11	☰ Left	Nominal	Input
13	빅데이터가장필요한업무가무엇이라고생각하시나요	String	24	0	13. 빅 데이...	{1, data ana}	None	13	☰ Left	Nominal	Input
14	빅데이터를위해기업에서투자해야한다고생각하시나요	String	26	0	14. 빅 데이...	{1, use}	None	12	☰ Left	Ordinal	Input

2. 우리가 설문한 데이터를 숫자로 나타낸다. 예를 들어, 성별은 각각 1과 2로 바꾸고, 그 변인의 측정종류이기 때문에 Nominal로 설정한다.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role	
1	성별	Numeric	3	0	1. 성별	None	None	3	☰ Right	Nominal	Input
2	나이	Numeric	12	0	2. 나이	None	None	7	☰ Right	Nominal	Input
3	직책	Numeric	18	0	3. 직책	None	None	6	☰ Right	Nominal	Input
4	기업규모	Numeric	20	0	4. 기업규모	None	None	7	☰ Right	Nominal	Input
5	기업업종	Numeric	17	0	5. 기업 업종	None	None	7	☰ Right	Nominal	Input
6	기업형태	Numeric	6	0	6. 기업 형태	None	None	6	☰ Right	Nominal	Input
7	당신은빅데이터의이름을알고있나요	Numeric	85	0	7. 당신은 빅...	None	None	10	☰ Right	Nominal	Input
8	당신은어떤환경으로빅데이터를알게되었나요	Numeric	25	0	8. 당신은 어...	None	None	13	☰ Right	Nominal	Input
9	당신은빅데이터가중요하다고생각하시나요	Numeric	26	0	9. 당신은 빅...	None	None	12	☰ Right	Nominal	Input
10	당신은실제업무에서빅데이터를사용하시나요	Numeric	43	0	10. 당신은 어...	None	None	11	☰ Right	Nominal	Input
11	자신의업무에서빅데이터가필요하다고생각하시나요	Numeric	56	0	11. 자신의 ...	None	None	11	☰ Right	Nominal	Input
12	자신의업무에서필요하다면그이유가무엇인가요	Numeric	59	0	12. 자신의 ...	None	None	10	☰ Right	Nominal	Input
13	빅데이터가장필요한업무가무엇이라고생각하시나요	Numeric	24	0	13. 빅 데이...	None	None	11	☰ Right	Nominal	Input
14	빅데이터를위해기업에서투자해야한다고생각하시나요	Numeric	26	0	14. 빅 데이...	None	None	14	☰ Right	Nominal	Input

3. 우리가 설정한 숫자로(Male-1, Female-2) 바꾸어 준다.

*Untitled3 [DataSet2] - PASW Statistics Data Editor

	Name	Type	Width	Decimals
1	성별	Numeric	3	0
2	나이	Numeric	12	0
3	직책	Numeric	18	0
4	기업규모	Numeric	20	0
5	기업업종	Numeric	17	0
6	기업형태	Numeric	6	0
7	당신은빅데이터의의미를알고있나요	Numeric	85	0
8	당신은어떠한경로로빅데이터를알게되었나요	Numeric	25	0
9	당신은빅데이터가중요하다고생각하시나요	Numeric	26	0
10	자신의업무에서빅데이터가필요하다고생각하시나요	Numeric	56	0
11	자신의업무에필요하다면그이유가무엇인가요	Numeric	59	0
12	빅데이터가가장필요한업무가무엇이라고생각하시나요	Numeric	24	0
13	빅데이터를위해기업에서투자해야한다고생각하시나요	Numeric	26	0



Linear Regression

Dependent: 16. 빅 데이터를 위해 기업에서...

Block 1 of 1

Independent(s): 7. 당신은 빅 데이터의 의미를 알고 있나요, 8. 당신은 어떠한 경로로 빅 데이터를 알게 되었나요, 9. 당신은 빅 데이터가 중요하다고 생각하시나요, 13. 자신의 업무에서 빅 데이터가 필요하다고 생각하시나요, 14. 자신의 업무에서 빅 데이터가 가장 필요한 업무가 무엇이라고 생각하시나요, 15. 빅 데이터가 가...

Method: Enter

Selection Variable: [] Rule: []

Case Labels: []

WLS Weight: []

OK Paste Reset Cancel Help

4. 정리한 데이터를 가지고 SPSS를 통해서 분석한다.

	성별	나이	직책	기업규모	기업업종	기업형태	당신은빅데이터의의미를알고있나요	당신은어떠한경로로빅데이터를알게되었나요	당신은빅데이터가중요하다고생각하시나요	당신은실제업무에서빅데이터를사용하시나요	자신의업무에서빅데이터가필요하다고생각하시나요	자신의업무에필요하다면그이유가무엇인가요	빅데이터가가장필요한업무가무엇이라고생각하시나요	빅데이터를위해기업에서투자해야한다고생각하시나요	var
1	1	1	3	1	4	2	3	1	5	1	5	2	1	5	
2	1	2	5	2	4	1	3	3	4	2	3	1	2	4	
3	2	3	2	3	6	3	2	1	3	2	3	5	1	2	
4	1	4	6	1	3	2	2	2	2		1		2	2	
5	1	3	4	2	2	1	4	3	5	1	5	1	2	5	
6	2	1	1	2	1	1	5	3	5	1	5	1	4	5	
7	1	3	3	1	3	2	1	6	1	2	1		5	1	
8	2	4	4	3	4	2	2	1	2	2	1		5	1	
9	1	2	2	1	5	2	5	3	5	1	4	4	2	5	
10	1	1	1	2	4	1	5	1	5	1	5	4	2	4	
11	1	2	1	3	6	3	4	2	4	1	4	4	1	4	
12	2	2	2	2	4	1	5	2	5	1	2	5	4	5	
13	2	3	3	1	5	2	4	1	3	1	3	4	1	3	
14	1	4	5	1	4	2	2	2	3	2	4	4	4	2	
15	2	1	2	2	5	1	5	1	4	1	4	4	1	5	
16	2	3	4	3	3	2	4	1	3	1	3	4	3	4	
17	1	4	5	3	5	2	3	4	2	2	3	5	2	2	
18	2	1	1	2	4	1	4	3	5	1	4	4	1	5	
19	2	3	4	2	2	2	3	3	3	2	3	5	2	3	
20	1	2	2	2	3	1	4	2	5	1	5	3	1	5	

분석할 데이터 결과

<분석 결과>

Model		Coefficients ^a				
		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta			
1	(Constant)	3.063	1.360		2.253	.033
	3. 직책	.007	.064	.009	.104	.918
	4. 기업규모	.111	.127	.075	.874	.390
	5. 기업 업종	-.074	.074	-.079	-.999	.327
	6. 기업 형태	-.089	.149	-.056	-.597	.555
	7. 당신은 빅 데이터의 의미를 알고 있나요?	-.158	.151	-.154	-1.048	.304
	9. 당신은 빅 데이터가 중요하다고 생각하시나요?	.974	.216	.831	4.500	.000
	10. 당신은 실제 업무에서 빅 데이터를 사용하시나요?	-.929	.323	-.369	-2.878	.008
	11. 자신의 업무에서 빅 데이터가 필요하다고 생각하시나요?	-.188	.201	-.169	-.937	.358
	13. 빅 데이터가 가장 필요한 업무가 무엇이라고 생각하시나요?	-.131	.100	-.128	-1.306	.203

a. Dependent Variable: 14. 빅 데이터를 위해 기업에서 투자해야 한다고 생각하시나요?

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.871	.803		2.329	.027
	4. 기업규모	.094	.118	.063	.793	.434
	7. 당신은 빅 데이터의 의미를 알고 있나요?	-.119	.122	-.116	-.977	.336
	9. 당신은 빅 데이터가 중요하다고 생각하시나요?	1.050	.191	.896	5.489	.000
	10. 당신은 실제 업무에서 빅 데이터를 사용하시나요?	-.890	.299	-.353	-2.978	.006
	11. 자신의 업무에서 빅 데이터가 필요하다고 생각하시나요?	-.180	.187	-.162	-.964	.343

a. Dependent Variable: 14. 빅 데이터를 위해 기업에서 투자해야 한다고 생각하시나요?

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.838	.797		2.306	.028
	7. 당신은 빅 데이터의 의미를 알고 있나요?	-.086	.114	-.084	-.753	.457
	9. 당신은 빅 데이터가 중요하다고 생각하시나요?	1.046	.190	.893	5.504	.000
	10. 당신은 실제 업무에서 빅 데이터를 사용하시나요?	-.795	.272	-.316	-2.922	.006
	11. 자신의 업무에서 빅 데이터가 필요하다고 생각하시나요?	-.184	.185	-.166	-.994	.328

a. Dependent Variable: 14. 빅 데이터를 위해 기업에서 투자해야 한다고 생각하시나요?

우리가 가지고 있는 데이터 자체가 유의미하지 않기 때문에 Significant value를 통해서도 Null hypothesis를 부정할 수 없었고, 다양한 독립변인들이 상관관계가 없다는 결과가 나왔다.

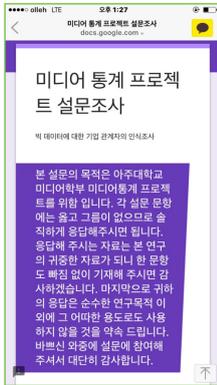


Welcome to the New Paradigm !

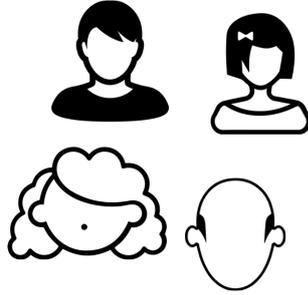
(그림25) 빅 데이터를 중시하는 새로운 패러다임

결론적으로 빅 데이터의 확산은 앞서 주장한 혁신확산이론과 의제 설정이론에 근거하여 설명할 수 있었다. 하지만 확산 이론에서 가장 중요한 확산의 요인 중 상대적 이점이나 적합성으로 인해서 사람들에게 다르게 받아들여진다는 것을 알 수 있었다. 인터뷰를 통해 개개인의 서로 다른 인식과 오피니언 리더의 영향력에 따라 하나의 혁신이 부정적으로 확산될 수도 있다는 사실을 알게 되었다. 하지만 빅 데이터의 경우 많은 오피니언 리더가 긍정적으로 판단하고, 그 중요성에 대해 많이 언급, 노출 함으로서 부정적인 인식보다는 긍정적인 인식이 많다는 것을 확인했다. 또한 의제 설정 이론에 근거하여 언론에서도 빅 데이터라는 의제에 대해 긍정적인 반응이 많았고 그 노출 빈도 수 역시 점점 증가한다는 사실을 확인했다. 이러한 추세와 언론의 긍정적 의제 설정, 많은 오피니언 리더들의 긍정적인 파급력이 기존의 데이터 분석에 대한 인식을 바꾸고, 빅 데이터라는 새로운 패러다임을 형성 시켰다고 결론지을 수 있었다.

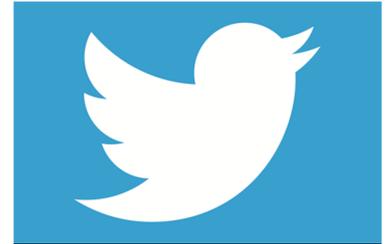
<한계 및 비판>



실제 데이터 측정X



적은 수의 인터뷰



twitter API 사용 실패

‘Big Data’ 유행에 대한 현상에 대한 가설을 검증하기 위해 기업인을 대상으로 하는 설문지를 만들었다. 하지만 현재 다양한 기업에 종사하는 사람들을 샘플링하여 설문조사를 하기엔 한계가 있다. 그 결과, 조사가 제대로 이루어지지 않은 상태로 임의적으로 데이터를 추출하다 보니 SPSS 프로그램을 돌려도 유의미한 상관 관계를 얻지 못했고, 예상치 못한 결과 값이 나왔다. 시간의 제약이나 비용적인 제약이 없다면 실제 실무자들을 대상으로 설문조사를 실시한 후에 제대로 된 노출 정도와 중요도를 파악하여 유의미한 상관 관계를 얻는다면 또 다른 유의미한 결과를 기대할 수 있다.

다음으로, 설문조사 뿐만 아니라 실제로 일반인을 대상으로 인터뷰를 실시하여 빅 데이터에 대한 인식을 확인하였다. 그러나 이 역시 시간과 인터뷰를 해 줄 사람을 충분히 샘플링 하지 못하여 소수의 사람들의 인식정도로 결론은 지어야 했다. 이러한 판단은 성급한 일반화의 오류와 같은 문제를 야기할 수 있기 때문에 유의미한 샘플링을 통해 얻은 더 많은 사람들을 인터뷰하여 정확도를 높여야 한다고 판단하였다.

마지막으로 ‘Big Data’의 확산에 대해서 가설을 설정하고 토의를 해본 결과, 트위터 api 를 사용하여 불 특정 다수의 데이터를 얻으려고 했으나 트위터 api를 사용할 수 있는 조원이 없었기 때문에 사용할 수가 없었다. 트위터 api를 사용하여 분석을 했다면 실제로 인터뷰를 통해 얻은 결과보다 오히려 더욱 정확한 사람들의 인식을 얻을 수 있다고 생각한다. SNS를 이용하는 일반인들을 대상으로 빅데이터에 대한 노출과 중요도 등을 충분한 양의 데이터를 통해 분석한다면 조금 더 유의미한 상관관계를 얻을 수 있다고 판단하기 때문이다.