통계의 창

WINDOW OF STATISTICS

2024.
WINTER
VOL.34

생성형AI 시대의 검색 혁명, SearchGPT와 AEO로의 마케팅 패러다임 전환 | 스포츠통계와 행동경제학

데이터 과학과 증거의 피라미드 | 기네스 맥주와 얽힌 이야기「통계와 기네스」

CONTENTS

통계의 창
2024. Winter
Vol.34

발행일 2024년 11월 5일

발행인 김정란

발행처 통계교육원

기획 정명진, 박현지

주소 대전광역시 서구 한밭대로 713(월평동) 통계센터 통계교육원

전화 042-366-6151, 6152

팩스 042-366-6498

이메일 mjjung@korea.kr, hyunjii@korea.kr

디자인 및 인쇄 (주)피그마리온(02-516-3923)

ISSN 2005-1379

1 통계 ISSUE

2 통계 FOCUS

3 통계 EDU

01 경제를 바라보는 도구 통계 지수

홍기훈 | 홍익대학교 경영학과 교수

현대 사회에서 경제는 우리의 일상과 불가분의 관계를 맺고 있다. 경제 활동은 국가의 발전을 좌우하고, 개인의 생활 수준을 결정하는 중요한 요소로 작용한다. 그러나 경제는 복잡한 구조를 가지며, 수많은 변수들이 상호작용하는 거대한 시스템이기 때문에 이러한 경제의 흐름을 정확하게 이해하고 예측하기 위해서는 다양한 경제 지표와 통계 지수들이 필수적이다.

통계 지수는 경제의 다양한 측면을 수치화하여 보여주며, 이를 통해 정책 입안자, 기업가, 투자자, 그리고 일반 국민들이 경제의 상태를 보다 명확하게 이해할 수 있도록 도울 수 있다. 이 컬럼에서는 경제 통계 지수가 무엇인지, 이 지수들이 왜 중요한지 그리고 대표적인 경제 통계 지수에는 무엇이 있는지에 대해 논의해 보려 한다.

경제 통계 지수란?

S&P500, NASDAQ, KOSPI, KOSDAQ, CPI, 빅맥(Big Mac)지수 등 각종 지수(index) 들을 다양한 경로를 통하여 자주 접할 수 있다. 예를 들어 빅맥지수는 맥도날드의 대표적 햄버거이자 그 조리법, 크기, 재료구성이 전 세계적으로 표준화 되어있는 빅맥의 판매가격을 기준으로 하여 각국의 상대적 물가수준과 통화가치를 비교하는 지수를 말한다. 이 지수는 '환율은 각국 통화의 상대적 구매력을 반영한 수준으로 결정된다'는 구매력평가설, '동일 제품의 가치는 세계 어디서나 같다'는 일물일가의 법칙에 기반하여 적정 환율을 산출하는데, 이 환율을 빅맥환율이라고도 한다.

이와 같이 지수는 일관된 기준을 가지고 어떤 현상

이나 사물을 측정하여 이에 대한 판단의 근거로 사용될 수 있도록 만든 계량화 된 통계를 말한다. 사회에는 다양한 경제지표가 국가 경제의 전반적인 상황을 객관적으로 확인하기 위해서 사용되고 있다. 가격 변화 또한 마찬가지로, 일상생활에서는 수많은 상품들이 거래되고 있으며, 상품 가격은 수시로 오르고 내리기 때문에 전반적인 가격 변화를 관찰하는 것은 매우 어렵다. 이러한 문제를 해결하기 위하여 가격의 전반적인 움직임을 표현하는 각종 가격지수를 만들어 발표하고 있다.

경제 통계 지수 왜 중요할까?

가격지수는 가장 보편적으로 화폐의 구매력을 측정하는 수단으로 사용되며 그러므로 경기판단 지표로서도 활용될 수 있다. 또한 가격지수는 각기 다른 시점의 가치를 비교할 수 있게 표준화시키는 과정에서 디플레이터의 기능을 수행할 수 있다. 경제 현상을 분석하다 보면 서로 다른 시점 간의 금액을 비교해야 할 때가 있다. 이때 현재의 금액을 과거 시점의 금액으로 환산해야 하는데 이때 사용되는 것이 가격지수이다. 이러한 가격지표의 유용성 때문에 여러 분야에서 그 분야의 전반적인 상황을 관측, 분석하기 위하여 다양한 지표들을 만들어 사용하고 있다.

조금 더 구체적으로 이야기하면 경제지수는 국가나 지역의 경제 상태를 진단하고 평가하는 데 필수적인 도구이다. 예를 들어, GDP(국내총생산)는 경제 성장률을 나타내며, 실업률은 노동 시장의 건강 상태를 보여준다. 이러한 지수들을 통해 경제의 현재 상태를 파악할 수 있으며, 경기 침체나 확장의 신호를 조기에 포착할 수 있다.

또한 이러한 지수들은 경제 정책 결정의 근거를 제공한다. 정부와 중앙은행은 경제 지수를 바탕으로 경제 정책을 수립한다. 예를 들어, 소비자물가지수(CPI)가 급격히 상승하면 중앙은행은 금리 인상을 고려할 수 있고, 반대로 실업률이 높다면 정부는 경기 부양책을 시행할 필요성을 느낄 수 있다. 경제 지수는 이처럼 정책 결정자들이 경제를 안정시키기 위한 적절한 조치를 취하는 데 중요한 근거를 제공한다. 이런 맥락에서 기업과 투자자들에게 경제 지수는 중요한 참고 자료이다.

예를 들어, 산업생산지수가 상승하면 특정 산업이 성장하고 있음을 의미하므로, 기업은 이에 맞춰 생산을 확대하거나 신규 투자를 고려할 수 있다. 투자자들은 주식, 채권, 부동산 등 다양한 자산의 가치를 평가하고 투자 결정을 내리기 위해 경제 지수를 분석한다.

경제 통계 지수는 어떻게 만들어 질까?

통계지수는 여러 요소의 변동을 하나의 수치로 요약하여 특정 현상을 간결하게 표현하는 방법이다. 그러므로 경제 통계 지수를 만드는 목적은 여러 데이터를 체계적으로 결합해 하나의 값으로 표현하는 것이다. 먼저, 경제 통계 지수를 만들기 위해서는 무엇을 측정할지 명확히 정의해야 한다. 예를 들어, 물가 변동을 나타내는 소비자 물가지수(CPI)나, 주식 시장의 변동을 반영하는 주가지수가 그 예시이다. 측정 대상이 정해지면, 이를 비교할 기준 시점을 설정한다.

일반적으로 과거의 특정 시점을 기준으로 설정하며, 이 시점의 값을 특정 값, 일반적으로 100으로 정하고 다른 시점의 변화를 이에 대비하여 측정한다.

다음으로 중요한 단계는 개별 요소의 가중치를 설정하는 것이다. 가중치는 각 요소의 중요도를 반영하는데, 예를 들어 소비자 물가지수를 작성할 때는 사람들이 많이 소비하는 품목에 더 높은 가중치를 부여할 수 있다. 이러한 가중치 부여를 통해 더욱 현실적인 지수를 만들 수 있다.

방법론이 정해졌다면 이제 기초 데이터가 필요하다. 데이터는 일반적으로 통계청이나 시장 조사 등 여러 출처를 통해 얻을 수 있다. 수집된 데이터를 바탕으로 기준 시점과 비교 시점 간의 변동을 계산하고, 이를 통해 개별 지수를 산출한다. 개별 지수는 기준 시점의 값을 100으로 두고, 비교 시점의 값을 이와 대비해 변동률을 계산하는 방식으로 구한다. 개별 지수가 계산되면, 각 요소에 부여된 가중치를 반영하여 종합 지수를 도출한다. 가중치가 적용된 개별 지수들을 모두 더한 후, 총 가중치로 나누어 최종적인 지수를 얻는다. 이렇게 만들어진 지수는 특정 현상의 변동성을 요약한 수치로, 경제 동향 분석이나 정책 수립 등의 과정에서 활용될 수 있다.

경제 통계 지수의 예

❶ 국내총생산

일단 가장 흔히 접할 수 있는 경제 통계 지수는 국내총생산(GDP: Gross Domestic Product)일 것 같다. GDP는 일정 기간 동안 국가 내에서 생산된 모든 재화와 서비스의 총 가치를 의미하기 때문에 한 국가의 경제 활동 수준을 평가하는 가장 기본적이면서도 중요한 지표로 볼 수 있다.

경제 성장률을 판단할 때, GDP의 변화는 중요한 역할을 한다. 예를 들어, GDP가 상승하면 경제가 성장하고 있음을 나타내며, 이는 일반적으로 생활 수준의 개선과 더불어 일자리 증가를 의미한다. 그러나 GDP는 경제의 모든 측면을 포괄하지 못한다는 점을 알아야 한다. 예를 들어, GDP는 소득 분배의 불균형, 환경 파괴, 사회적 복지 등 질적인 측면을 충분히 반영하지 못하기 때문이다. 그러므로 GDP만을 가지고 경제의 건강 상태를 판단하기보다는, 다른 지수들과 함께 종합적으로 해석할 필요가 있다. 예를 들어, 한 국가의 GDP가 지속적으로 상승하더라도, 소득 불평등이 심화되고 있다면 이는 지속 가능한 경제 성장으로 판단하기 어려울 수 있다.

❷ 소비자 물가지수

또 우리가 아주 많이 접하는 경제 통계 지수에는 소비자물가지수(CPI: Consumer Price Index)가 있다. CPI는 소비자가 일상적으로 구매하는 상품과 서비스의 가격 변동을 측정하는 지수이며 물가 상승률, 즉 인플레이션을 측정하는 데 사용된다. 그러므로 CPI는 경제의 안정성을 평가하는 데 중요한 역할을 한다. CPI가 급격히 상승하면 물가가 빠르게 오르고 있음을 의미하며, 이는 소비자의 구매력을 감소시킬 수 있다. 한국은행이나 미국의 연방준비위원회와 같은 중앙은

행들이 통화 정책을 결정하는데 CPI는 아주 중요한 기준을 제시한다.

❸ 생산자 물가지수

소비자물가지수가 있다면 생산자물가지수(PPI: Producer Price Index)도 있다. PPI는 생산자가 판매하는 상품과 서비스의 가격 변동을 측정하는 지수로, 소비자물가지수(CPI)와 함께 인플레이션을 측정하는 중요한 지표 중 하나이다. PPI는 주로 원자재와 중간재의 가격 변동을 반영하기 때문에, 소비자물가지수보다 경제 전반에 미치는 영향을 더 빨리 포착할 수 있다.

예를 들어, PPI가 상승하면 이는 결국 소비자 가격에도 영향을 미칠 가능성이 크다. 따라서 PPI는 미래의 소비자물가 동향을 예측하는 데 중요한 역할을 한다. PPI는 또한 산업별 가격 동향을 분석하는 데 유용하다. 이는 기업의 비용 구조와 이익률에 직접적인 영향을 미치며, 궁극적으로 기업의 경영 전략과 투자 결정에 중요한 요소로 작용한다. 그러므로 경제 분석가들은 PPI를 통해 산업별 가격 압력과 그에 따른 경제적 영향을 예측할 수 있다.

❹ 실업률

실업률 또한 우리가 자주 접하는 경제 통계 지표 중 하나이다. 실업률은 경제의 건강 상태를 나타내는 중요한 지표로, 노동시장의 활력을 평가하는 데 사용된다. 실업률은 경제활동인구 중에서 일자리를 구하지 못한 사람들의 비율을 의미한다. 실업률이 낮을수록 경제가 활발히 돌아가고 있으며, 많은 사람들이 고용되어 있다는 것을 나타낸다. 반면 실업률이 높아지면 이는 경제 활동이 위축되고 있음을 의미하며, 개인의 생활수준이 악화될 가능성이 크다.

❺ 산업생산지수

산업생산지수도 아주 중요한 경제 지표 중 하나이다. 산업생산지수는 제조업, 광업, 공공서비스 등 주요 산업 부문의 생산 활동 수준을 측정한 지수로, 경제의 공급 측면을 이해하는 데 중요한 역할을 하기 때문이다. 이 지수는 경제의 전반적인 생산 능력과 활동 수준을 파악하는 데 사용되며, 특히 제조업과 같은 핵심 산업의 상태를 분석하는 데 유용하다. 산업생산지수가 상승하면 경제 활동이 활발하다는 신호로 해석할 수 있으며, 이는 GDP 증가로 이어질 가능성이 크다. 반대로 산업생산지수가 하락하면 경제가 둔화되고 있을 가능성이 있다. 일반적으로 산업생산지수는 경제 성장의 선행 지표로 여겨진다. 이는 기업이 미래의 수요를 예상하고 생산을 늘리거나 줄이는 등의 결정을 내리기 때문이다. 그러므로 이 지수는 경제 예측과 관련하여 중요한 의미를 가지며, 특히 경기 변동에 민감한 산업에서 더 중요한 의미를 가질 수 있다.

❻ 경상수지

경상수지는 한 국가가 외국과의 거래에서 벌어들인 수익과 지출을 나타내는 지표이다. 경상수지는 상품 및 서비스 무역, 소득 이동, 그리고 일방적 이전(예: 해외 원조) 등으로 구성되며, 국제 무역과 외환의 흐름을 파악하는 데 아주 중요한 역할을 한다. 경상수지가 흑자를 기록하면 그 국가는 외국에서 더 많은 돈을 벌어들이고 있음을 의미하며, 이는 통화 가치의 상승과 외환보유고의 증가로 이어질 수 있다.

반면 경상수지가 적자를 기록하면 외국으로 돈이 더 많이 빠져나가고 있음을 의미하며, 이는 통화 가치의 하락과 외채 증가로 이어질 수 있다. 경상수지는 경제의 대외 경쟁력을 평가하는 데 중요한 역할을 하며, 국제 금융시장에서 한 국가의 신용도와 투자 매력을 판단하는 데 사용될 수 있다. 또한, 경상수지의 변화는 환율 변동과 국제 무역 정책에 영향을 미칠 수 있기 때문에, 이를 면밀히 분석하고 대응하는 것이 중요하다.

결언

일반적으로 지수의 바람직한 요건은 관심영역을 잘 집약하여 대표할 수 있는 대표성을 가져야 하고, 측정방법과 데이터의 객관성 확보가 가능한지의 측정 가능성이 필요하고, 시간적 변화나 다른 여러 조건들의 변화와 무관하게 비교 가능한 일관성이 있어야 하며, 사용자가 지수를 이해하고 쉽게 분석할 수 있는지에 대한 용이성이 충족되어야 한다.

경제 지수는 경제의 상태를 진단하고, 정책을 수립하며, 미래를 예측하는 데 필수적인 도구이다. 이러한 지수들을 적절히 활용함으로써 정부, 기업, 개인은 더 나은 경제적 결정을 내릴 수 있다. 따라서 경제 지수는 단순한 숫자 이상의 의미를 가지며, 우리가 살아가는 사회의 방향을 결정하는 중요한 역할을 한다.

02

생성형AI 시대의 검색 혁명, SearchGPT와 AEO로의 마케팅 패러다임 전환

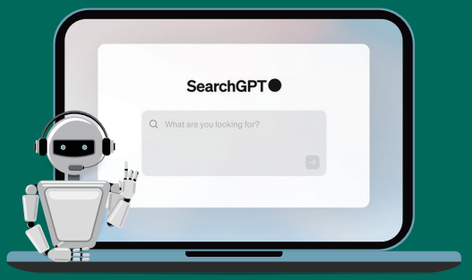
윤성임 | 소셜앤비즈 대표

서론:챗GPT 등장 이후 Search AI 기술로 검색 패러다임 변화

2022년 11월 30일, OpenAI가 출시한 챗GPT(chatGPT)는 인공지능(AI) 기술, 특히 자연어 처리(NLP) 분야에 커다란 변화를 가져왔다. 챗GPT는 대규모 언어 모델(LLM)을 통해 인간처럼 자연스러운 대화(프롬프트)를 주고 받으며 글쓰기, 이미지 생성, 영상 제작 등 다양한 작업을 수행할 수 있게 함으로써 산업 전반에 걸쳐 업무 방식을 혁신적으로 변화시켰다.
특히, 지난 7월 말 챗GPT의 개발사 오픈AI(Open AI)가 검색 AI(Search AI) 기술인 '서치GPT(Search GPT)'의 프로토타입을 공개하면서 정보 검색과 지식 접근 방식의 패러다임이 크게 변화하고 있다.

기존의 키워드 기반 검색 방식에서 벗어나, AI는 사용자의 질문을 이해하고 맥락에 맞는 답변을 제공하는 방향으로 진화하고 있다. 이와 함께 등장한 검색 AI 서비스에는 MS Bing AI, 퍼플렉시티 AI(Perplexity AI), You.com, 네이버 검색 큐(Cue:), 구글 검색 AI 오버뷰, Goover AI 등이 있다. 그중 Perplexity AI는 실시간 데이터와 결합하여 더욱 정확하고 최신의 정보를 제공하며, 기존의 SEO(Search Engine Optimization)와 차별화된 AEO(Answer Engine Optimization)로의 전환을 예고하고 있다.

이는 기존의 단순 키워드 기반 검색에서 사용자의



(출처 : OpenAI)

질문의 맥락을 정확히 이해하고 관련 정보를 종합하여 직접적인 답변을 제공하는 '대화형 검색'으로의 패러다임 전환이 시작된 것이다. 시장에서는 지난 20여년 검색 시장을 독점해 온 구글과의 본격적

인 경쟁을 예상하고 있다.

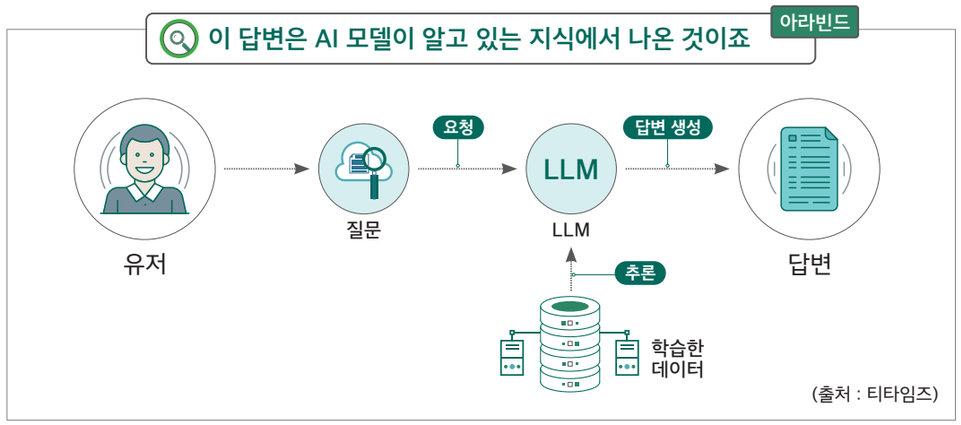
2022년 12월 7일, 챗GPT 출시 일주일 후 Perplexity AI가 서비스를 시작하였다. 이 AI 답변 엔진(Answer Engine)은 구글의 검색 엔진을 대체할 가능성으로 IT 업계의 큰 주목을 받고 있다. 이를 중심으로 검색 AI의 현황을 살펴보고자 한다.

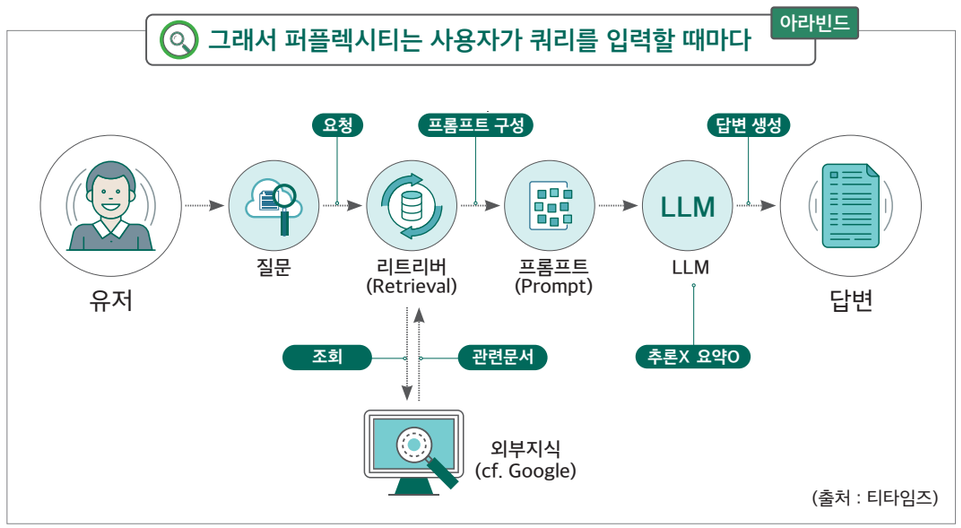
생성형AI와 검색 AI와의 차이점

생성형AI 대표인 챗GPT와 검색 AI 대표인 Perplexity AI는 정보 제공 방식에서 차이가 있다. 지난 9월 한국을 방문한 아라빈드 스리니바스(Aravind Srinivas) 퍼플렉시티 공동창업자 겸 CEO의 티타임즈 인터뷰 내용을 기반한 생성형AI(챗GPT)와 검색 AI(퍼플렉시티)와의 차이점은 다음과 같다.

챗GPT는 대규모 언어 모델을 통해 학습된 데이터에서 답변을 생성하는 방식으로, 사전 학습된 데이터에 기반하여 답변을 제공한다. 그래서 최신 정보 제공에는 한계가 있으며, 할루시네이션(Hallucination)이 발생할 가능성도 존재한다. 따라서 생성된 정보에 대한 정확성 검증이 반드시 필요하다.

반면, Perplexity AI와 같은 검색 AI는 사용자의 질





문에 대해 AI가 리트리벌(Retrieval) 단계를 통해 외부 데이터에서 실시간으로 검색하고, 이 검색된 문서를 바탕으로 프롬프트(Prompt)를 구성하여 LLM에서 추론하지 않고 요약하여 답변을 생성하는 구조이다. 이는 최신 정보를 바탕으로 신뢰성 높은 답변을 제공할 수 있으며, 출처를 명확히 제공하기 때문에 정보의 정확성이 높아진다.

챗GPT와 Perplexity와의 차이점을 표로 정리해보면 다음과 같다.

기존 검색(SEO)과 AI 검색(AEO)으로 새로운 검색 최적화 시대 도래

1990년대 인터넷이 보급되면서 우리는 기존 미디

챗GPT와 퍼플렉시티와의 차이점

특성	생성형 AI(챗GPT)	AI 검색(Perplexity)
데이터 소스	학습한 데이터 기반	외부 지식 및 실시간 데이터
정보 처리	LLM을 통해 직접 생성	Retrieval을 통해 관련 문서 검색 후 요약
프로세스 단계	질문 → LLM → 답변 생성	질문 → Retrieval → Prompt 구성 → LLM → 답변 생성
응답 생성	모델이 알고 있는 지식에서 생성	외부 데이터를 기반으로 정확한 정보와 출처를 제공
정보 신뢰성	모델의 학습 시점에 따라 다름	최신 정보 반영 가능
주 사용층	다양한 범용적 사용자에게 적합, 특히 창의적 콘텐츠 생성을 하는 작가, 디자이너 등 창작자에게도 유리	일반적으로 Perplexity와 같은 AI 검색 엔진은 AI 검색의 실시간 정보 검색과 분석, 정확한 데이터를 필요로 하는 연구자, 학생, 전문가 등이 주 사용자 층

SEO와 AEO의 차이점

구분	SEO(Search Engine Optimization)	AEO(Answer Engine Optimization)
목적	• 검색 엔진에서 웹페이지의 상위 노출	• AI 답변 엔진에서 학습된 최신의 정확하고 유용한 답변 제공
목표	• 웹사이트 트래픽 증가 및 검색 결과 페이지(SERP)에서 상위 노출	• AI 답변 엔진에서 사용자의 검색 의도를 파악하고 맥락에 맞는 최적의 답변 제공
사용자 경험	• 키워드 중심의 질문 • 사용자가 여러 링크를 클릭해 필요한 정보를 찾아야 함	• 자연어 방식의 복잡한 질문 • AI가 질문 맥락을 이해하여 출처 및 맞춤형 답변을 제공 • 후속질문제시
최적화 대상	• 키워드, 링크 구조, 메타데이터, 콘텐츠 품질	• 질문에 대한 명확한 답변, 구문 분석, 자연어 처리 기반의 정확한 정보
주요 기술	• 키워드 연구, 백링크, 메타 태그, 페이지 로딩 속도	• 자연어 처리(NLP), 구조화된 데이터, LLM(대규모 언어 모델), 구문 분석
적용 방식	• 검색 알고리즘이 웹페이지를 크롤링하고 인덱싱	• AI 답변 엔진이 질문을 이해하고 관련 정보 추출
중점 요소	• 키워드 밀도, 페이지 구조, 사용자 경험(UX), 반응형 웹 디자인	• 정확한 답변 제공, 질문과 관련된 구체적이고 신뢰성 있는 정보 제공
최적화 방법	• 블로그, 기사, 제품 설명 페이지 등 다양한 형식	• FAQ 스타일, 구조화된 데이터, 명확한 질문-답변 형식
성공 측정 지표	• 클릭률(CTR), 페이지 체류 시간, 반송률(Bounce Rate)	• 답변의 정확성, 사용자 질문에 대한 AI의 이해도와 만족도
미래 전망	• 검색 엔진 알고리즘의 발전에 따른 지속적 변화	• AI 기반의 개인화된 답변 제공 및 더욱 정교한 질문 분석

어(신문, TV, 라디오, 전화 등) 대신 웹으로 정보를 쉽게 접근할 수 있게 되었다. 폭발적인 속도로 웹에 정보가 늘어나면서 원하는 정보를 정확히 찾는 것이 어려워졌다. 2000년대 PageRank라는 새로운 알고리즘을 기반으로 만들어진 구글 검색 엔진이 등장하여 원하는 정보를 빠르게 검색할 수 있게 되었다. 지난 2006년에는 옥스포드 영어 사전과 메리엄-웹스터 사전에 '구글링'이라는 단어가 추가되기도 하였다.

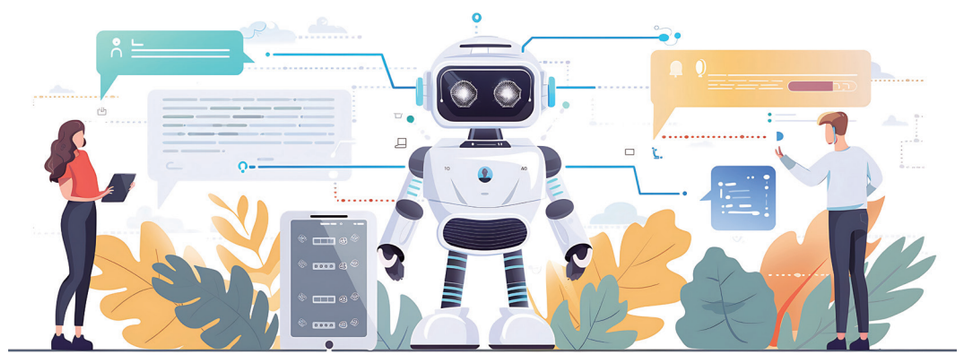
지난 20여년간 인터넷 시대의 '검색'의 대명사가 된 구글 SEO(Search Engine Optimization, 검색 엔진 최적화)는 전통적으로 웹사이트가 검색 엔진 결과 페이지(SERP)에서 상위에 노출되도록 기술적, 콘텐츠적 요소를 최적화하는 것이 핵심이었다.

그러나, 답변 기반 AI 검색이 주도하는 시대에는 AEO(Answer Engine Optimization, 답변 엔진 최적화)가 새로운 최적화 전략으로 떠오르고 있다. AEO는 단순히 검색 엔진에서 상위 노출을 노리는 것이 아니라, AI가 사용자의 질문에 대해 직접적으

로 유용한 답변을 제공할 수 있도록 웹 콘텐츠를 최적화하는 방법을 의미한다.

사용자 경험 측면에서 SEO는 키워드 중심의 질문에 따라 나타나는 결과물인 여러 링크를 사용자가 직접 하나 하나 클릭해 최종적으로 필요한 정보를 찾고 정리해야 하는 반면, AEO는 사용자의 자연어 방식의 복잡한 질문에 AI가 질문 맥락을 이해하여 출처와 함께 맞춤형 답변을 제공한다는 것이다. 또한 후속 질문도 추가 제시해 줌으로써 더 깊은 대화와 학습을 통해 최적의 답변을 도출하게 도와준다는 것이다.

기술적인 관점에서도 SEO와 AEO와의 차별점은 두드러진다. SEO는 링크 구조, 키워드, 메타데이터 최적화에 중점을 둔다면, AEO는 구문 분석, 자연어 처리 및 답변의 정확성을 중요시한다. 즉, AI가 스스로 콘텐츠의 맥락과 의미를 이해하고 사용자에게 적합한 답을 제공할 수 있도록 데이터를 구조화하는 방식이다.



AEO 콘텐츠 전략

검색이 일상화된 인터넷 세상에서 검색 엔진은 오늘날 우리가 정보를 얻는 주요 도구 중 하나로, 우리의 일상적인 삶과 업무에 엄청난 영향을 미친다. 이를 통해 우리는 필요한 정보를 빠르고 효율적으로 찾으며, 의사결정을 내리고 다양한 문제를 해결한다.
또한, 사용자는 여러 링크를 탐색해가며 원하는 답을 찾는 SEO 방식에서 AI가 최신의 정확한 답변을 신속하게 제공해주는 AEO 방식으로 검색 습관을 크게 전환하고 있다. 이에, 검색 AI 시대 기업이나 개인은 브랜드 가시성(Visibility)을 높일 수 있는 AEO 콘텐츠 최적화가 매우 중요하다.

AI기반 마케팅 패러다임 전환을 리드하는 국내 기업인 'The Core'에서 발표한 내용 등을 기반으로 필자는 AEO 콘텐츠 전략을 다음과 같이 5가지로 제안해본다.

1 자연어 처리(NLP)와 사용자 의도 파악을 결합한 질문-답변 콘텐츠

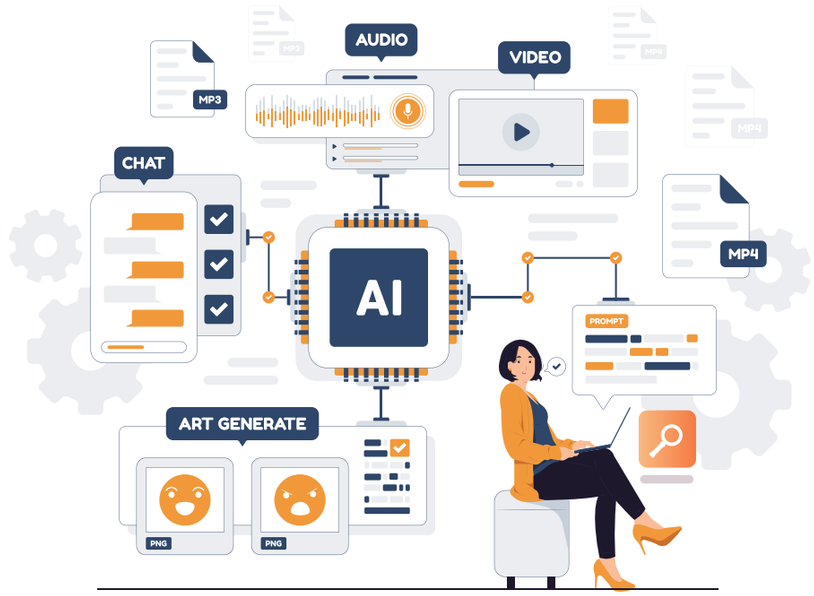
음성 검색이나 복잡한 질문에 대한 답변을 제공할 때, 자연어 처리(NLP) 기술을 활용하여 사용자의 질문을 정확하게 해석하고 의도를 파악한다. 이를 통해 FAQ 스타일의 콘텐츠를 제공하여 검색 엔진이 자연스럽게 콘텐츠를 이해하고 처리할 수 있게 한다.

2 구조화된 데이터를 활용한 간결한 답변 제공 콘텐츠

스키마 마크업과 같은 구조화된 데이터를 사용하여 간결하고 명확한 답변을 제공한다. 예를 들어, Schema.org를 활용한 리치 스니펫 최적화, FAQ, How-to, Q&A 등의 스키마 유형을 적용하여, 검색 엔진이 해당 정보를 더 정확히 이해하고 검색 결과에 반영할 수 있도록 유도한다.

3 신뢰할 수 있는 정보 기반 콘텐츠

권위 있는 웹사이트나 신뢰성 높은 출처에서 인용한 정보를 사용하여 콘텐츠의 신뢰성을 높인다. AI 검색 엔진은 신뢰할 수 있는 정보 소스를 선호하므로, 이러한 신뢰성 높은 콘텐츠는 검색 결과에서 가시성을 높이고 상위에 노출될 가능성을 증가시킨다.



4 키워드를 포함한 꾸준한 최신성 콘텐츠

최신 정보와 키워드를 지속적으로 반영한 콘텐츠는 SEO뿐만 아니라 AEO에서도 중요하다. AI는 최신 질문에 대해 정확한 답을 제공할 수 있도록 지속적으로 콘텐츠를 업데이트해야 하며, 이를 통해 AI가 신뢰할 수 있는 답변을 제공할 수 있는 소스로 인식된다.

5 멀티모달 콘텐츠

텍스트, 이미지, 동영상, 인포그래픽 등 다양한 미디어 형식을 결합한 멀티모달 콘텐츠는 AI가 사용자의 질문에 대해 더 풍부하고 정확한 답변을 제공하는 데 도움을 준다. 예를 들어, 제품 사용법에 대해 설명할 때, 단순 텍스트보다는 동영상 튜토리얼이나 이미지가 포함된 콘텐츠가 더 효과적으로 정보를 전달할 수 있다.

결론

AI 검색 기술의 발전은 정보 검색 방식뿐만 아니라 마케팅에도 큰 영향을 미치고 있다. SEO에서 AEO로의 전환은 검색 결과의 정확성과 사용자 경험을 개선하는 데 기여하며, 마케터들은 AI가 선호하는 최적화된 콘텐츠 전략을 통해 브랜드 가시성을 극대화할 수 있다.

미래의 검색 AI는 더욱 개인화된 검색 경험을 제공하고, AI 마케팅도 자동화와 개인화를 기반으로 발전할 것이다. 따라서, 구조화된 데이터와 최신성을 반영한 콘텐츠, 멀티모달 콘텐츠 전략을 통해 AEO 최적화에 집중하는 것이 중요하다.

최근 들어 통계에서 그 중요성이 점점 커져가고 있는 분야가 있다. 바로 스포츠 분야이다. 스포츠는 속성 상 승자와 패자가 가려지는 분야인데 예전에는 타고난 신체적 특성에 더해 근성과 노력으로 승자가 정해졌다고 하면 오늘날에는 훈련부터 전략, 전술에 이르기까지 데이터를 분석하고 활용함으로써 승리에 한발 더 가까이 갈 수 있다는 게 통설이다. 특히 오늘날의 스포츠 경기는 산업화되고 있으므로 개인이나 팀의 승리가 곧 부와 직결됨에 따라 스포츠통계 분석은 더욱더 발달할 수 밖에 없다.

03 스포츠통계와 행동경제학

정태성 | 한국행동경제학연구소 대표 및 스포츠시그널 대표

스포츠통계의 시작, 야구

일반적으로 스포츠통계란 스포츠에서 발생하는 다양한 데이터를 수집, 분석, 해석하는 학문을 의미한다. 초기에는 주로 선수 개인의 퍼포먼스를 극대화시키기 위해 운동수행 능력과 관련된 데이터를 수집하여 분석하는 생체역학이나 운동역학에서 주로 사용하였으나 이에 더해 기존에 선수나 팀이 보여준 기록을 기반으로 선수나 팀의 능력을 새로운 지표로 측정하고 이를 모델화하여 승패를 예측할 수 있는 영역까지 확산되었다. 이렇듯 스포츠통계의 외연도 확대되고 수준도 높아짐에 따라 처음에는 야구의 '빌 제임스'처럼 좋아서 시작하는 개인부터 시작해서 팀으로 확산되었으며, 스포츠 분야만 전문적으로 분석하는 기업이 생겨났을 뿐만 아니라 최근에는 FIFA와 같은 국제스포츠 기구에서도 전문 통계 지표를 제공하기 시작했다. 심지어 FIFA가 제공하는 통계 지표는 기존 스포츠통계업체보다 한층 세밀해서 볼 점유율을 넘어 새로운 유형인 경합까지 도입했고, 팀의 공격 방향도 왼쪽과 중앙, 오른쪽만 따지던 기존과 달리 5가지로 세분화하였으며, 빌드업, 롱볼, 역습, 압박 등을 인공지능(AI) 알고리즘에 기반한 추적 데이터로 분석해 제공한다.

스포츠통계에 대해 개인보다는 팀, 아마추어보다는 프로구단에서 더 필요로 하는 것은 자명한 사실이다. 왜냐하면 승리가 곧 부와 연결되는 산업의 영역이기 때문이다. 1승을 올릴 때마다 팀의 인기는 늘어나게 되는데 특히 우승권에 있으면 각종 스폰서십이 높은 가격으로 붙고, 열광적인 팬들은 구단에 돈을 쏟기 시작하며 구단의 가치는 더욱더 높아지게 된다. 구단주들은 구단의 가치가 높아져서 투자금액 이상의 돈을 벌 수 있으면 구단을 팔기도 한다. 이러한 일들이 벌어지는 곳이 바로 스포츠산업이기 때문에 1승을 올리기 위해, 그리고 우승을 하기 위해 통계분석에 열을 올릴 수밖에 없다.

이러한 스포츠통계는 어느 영역에서부터 발전했을까? 기본적으로 인기가 있다는 전제 하에 데이터 수집 및 분석이 쉬운 영역부터 발전하기 시작하였는데 그게 바로 메이저리그로 대변되는 야구이다. 1970년대 통계학자 빌 제임스는 타율이나 평

균자책점 등 밖에 제공하지 못하는 야구계에 개탄을 하며 다양한 통계지표들로 구성된 ‘세이버메트릭스’를 제안했다. 이후 ‘야구를 모르는 사람들의 숫자놀음’으로 치부되며 일부 팬들 사이에서만 회자되었던 세이버메트릭스는 거의 30년이 지난 시점인 2000년대 초반에 그 가치를 인정받기 시작했다. ‘머니볼’로 잘 알려진 오클랜드의 ‘빌리 빈’ 단장이 팀을 맡으며 세이버메트릭스에 기반한 출루율로 일대 센세이션을 일으킴으로써 메이저리그에 큰 충격을 주었다. 사실, ‘빌리 빈’ 단장이 초석을 깔아 놓았지만 데이터 야구를 더욱 굳건히 한 것은 어린 나이에 보스턴의 단장이 된 ‘테오 엡스타인’이다. 원래 보스턴에서 ‘빌리 빈’ 단장을 영입하려 했으나 실패한 뒤, 바로 단장으로 임명된 데이터 전문가 ‘테오 엡스타인’은 2004년에 팀 내 유명한 선수들을 전격 트레이드 하며 86년만에 ‘밤비노의 저주’를 깨고 팀을 우승시키는 쾌거를 달성했다.

EPL에서 스포츠통계의 발전과 한계

이에 반해 축구는 유럽에서의 인기에 비해 생각보다 늦게 데이터를 중요시하기 시작했다. 야구는 투수와 타자간 1:1 대결부터 시작하는 반면에 축구는 끊임없이 뛰어다니는 11명의 선수로 구성되어 있다. 지금이야 발달된 기술의 카메라로 촬영하고, 선수들의 웨어러블 조끼를 통해 많은 데이터를 확인할 수 있지만 현재와 같이 기술이 발전하기 전까지는 데이터를 확인하고 수집하는 것 자체가 매우 힘들었다. 축구 분야에서 데이터를 활용한 선구적인 인물은 경제학자 출신인 아스널의 ‘아르헨 벵거’ 감독인데 마땅한 장비가 없던 시절에도 데이터의 중요성을 깨닫고 코치진들과 함께 직접 선수 데이터를 측정하고 다녔다. 최근에 리버풀의 전성시대를 다시 만들고 현역에서 잠시 물러난 ‘클롭’ 감독의 경

우에도 리버풀에서의 성과는 구단에서 데이터 분석을 전격적으로 지원해 준 결과라고 말한다.
이렇듯 스포츠통계는 오늘날 기술적인 발전과 더불어 그 영역을 확대해 나가고 있는데, 보다 더 정확한 분석과 예측을 하려는 많은 사람들이 어쩔 수 없어하는 영역이 있다. 스포츠통계는 선수들의 실력과 성적을 인과관계로 바라보고 분석하는데 성적과 인과관계가 성립하지 않는 단어, 그리고 실력과 반대되는 단어인 ‘운’이다.
마이클모부신의 ‘The Success Equation’에 따르면 프로스포츠 성적에 운이 차지하는 비중이 미국 메이저리그는 약 34%, 영국 프리미어리그는 31%, 미국 풋볼리그(NFL)는 38%라고 한다. 만약 데이터 분석을 통해 각 팀들이 비슷한 실력의 선수들로 구성된다면 운에 따르는 비중은 더욱 높아지게 된다.
그런데, 운이라고 하는 것은 그 개념이 모호하기 때문에 측정되어지는 실력 이외에 측정되어지지 않지만 경기에 영향을 미치는 많은 변수들을 ‘운’이라고 통칭하는 것이 더 정확할 것이다.
그렇다면 측정 되어지지 않은 많은 변수들, 의사결정 과정에 있어서 영향을 끼칠만한 심리적 요인들이 고려되어야 하는데 과연 어떻게 할 수 있을까? 이에 대한 실마리를 찾을 수 있는 분야가 바로 행동경제학이다. 행동경제학은 인간은 합리적으로 행동하는 것 같지만 사실은 여러 가지 요인에 의해 비합리적인 의사결정과 행동을 할 수 있다는 전제 하에 그러한 심리와 행동을 과학적으로 풀어내는 분야이므로 스포츠통계에서 놓치고 있는 부분을 보완할 수 있다.

NBA에서 스포츠통계의 발전

그렇다면 실제로 행동경제학이 스포츠통계가 발전하는데 어떻게 영향을 끼쳤을까? 이에 대해서는 미국 프로농구 NBA팀 필리델피아 세븐티식서스(Philadelphia 76ers)팀의 ‘대릴모리(Daryl Morey)’ 사장의 사례로 파악해보자.
대릴모리는 노스웨스턴대 컴퓨터공학, MIT 슬론 경영대학원 MBA 출신으로 농구와는 담을 쌓은 사람이다. 다만, 컨설팅회사에서 일할 때 NBA 명문 구단 중 하나인 ‘보스턴 셀틱스’의 컨설팅을 담당했는데 탁월한 데이터 분석력으로 2006년 휴스턴 로키츠(Houston Rockets)로 스카우트 된 후, 바로 단장직을 맡았다. 휴스턴 로키츠에 부임한 초기에는 그를 향한 야유와 조롱이 대부분이었다. ‘농구도 모르는’, ‘머릿속으로 분석만 하는’ 등의 조롱이 넘쳐났고, ‘Nerd’(너드: 지능이 뛰어나지만 강박관념에 사로잡혀 있거나 사회성이 떨어지는 사람을 일컫는 말)라는 단어는 그를 표현하는 단어가 되었다.
그럼에도 불구하고 그는 신인선수 드래프트에서 가장 적합한 선수를 뽑아야 했기에 휴스턴에 와서 가장 먼저 한 일은 선수의 미래 성적을 예측하는 통계모델을 만드는 일이었다. 즉, 농구 선수의 특성 중 향후 성공할 수 있는 주요 요인을 찾아내어 각 특성에 가중치를 부여하여 모델을 만드는 일인데, 바로 이 모델을 만드는 데 있어서 가장 중요한 점은 어떤 특성이 중요한지, 그리고 얼만큼의 가중치를 부여하는지이고 이 때 대릴모리에게서 눈여겨봐야 할

점은 바로 행동경제학에서 얘기하는 '편향'을 제거하려 애썼다는 점이다.

구체적으로 얘기하자면 그 전까지는 거들떠 보지 않았던 선수에게 양쪽 부모가 다 있는지, 선수 친인척 중에 NBA 선수가 있는지, 대학 때 다양한 포지션을 맡은 경험이 있는지, 벤치 프레스를 어느 정도까지 할 수 있는지 등 모을 수 있는 많은 데이터를 모았으며, 이를 선수의 성적과 상관관계가 있는지 검증했다. 물론, 대부분 상관관계가 나오지는 않았다. 하지만 결론적으로 그때까지 중요했던 선수들의 득점력, 리바운드, 스틸 등 기존 지표보다는 분당 스틸, 분당 득점 등 효율성 지표에 더 중점을 두었고, 선수들 신장보다는 최대한 팔을 뻗을 수 있는 암리치 혹은 윙스팬이 더 중요하게 생각되었다. 그런데, 농구 선수 선발에 있어서 또 중요한 단계는 트라이아웃이다. 지금 현재 몸 상태가 어느 정도 수준이고 얼마나 능력 있는지를 눈으로 보고 판단하는 단계인데, 이때는 전문가의 능력이 중요하다. 대릴모리는 선수를 관찰할 때 받는 즉각적인 인상을 중심으로 나머지 데이터를 받아들이는 '확증편향'이 발생한다는 점을 알아차렸다. 선수를 두고 즉각적인 견해를 가지면 그 견해를 지지하는 증거만 받아들이는 것이 바로 확증편향으로 대부분 NBA 전문가가 그런 편향을 가지고 있었다. 대표적인 예로 드는 선수는 '린새너티'(Lin + Insanity 광기)'라는 신조어의 주인공인 '제레미 린'이다.

대릴모리가 본 NBA에서 나타난 편향(Bias)

대릴모리가 고안한 선수평가 시스템에서 제리미 린은 '첫 발 스피드' 등에서 최고 수준으로 그 해에 15번째 선수로까지 평가를 높게 하고 있었는데, 스카우트들 사이에서 그는 운동신경이 발달하지 못한 아시아 청년에 불과해서 여기에 대한 확증편향 증거들만 잔뜩 제시되고 있었다. 결국, 대릴모리는 소심하게 그를 지나치게 되었는데 훗날 매우 후회스

러운 장면 중 하나로 회고한다.

대릴모리가 봤던 또 하나의 편향은 소유효과(혹은 부존효과 Endowment Effect)이다. 카일 라우리(Kyle Lowry)라는 평판이 좋은 가드와 드래프트 1차 지명권을 트레이드 해달라는 요청을 받았을 때, 처음 전문가들끼리는 절대 안된다는 의견이 대다수였으나 입장을 바꿔놓고 생각했을 때, 꽤 괜찮은 거래라고 생각되어 진행하게 되었다. 내가 보유하고 있는 무언가에 대해서는 객관적인 가치보다 훨씬 더 높은 가치를 부여하는 '소유효과' 때문에 카일 라우리는 트레이드에 부정적이었으나 반대 입장에서 생각했을 때는 1차 지명권이 훨씬 더 괜찮다는 결론을 내려 결국 트레이드를 진행하게 되었고, 그 결과 현재 NBA 최고 선수 중 하나로 꼽히는 '제임스 하든'을 획득할 수 있었다.

이렇게 2000년대 중반부터 NBA에 데이터 농구를 대입해서 지금은 모든 구단들이 각자에 맞는 데이터 시스템을 구축하고 활용하도록 만든 선구자는 대릴모리라고 볼 수 있는데, 대릴모리는 데이터예측 시스템을 바꾸고 수정하고 신뢰하게 만든 데에는 그가 들었던 행동경제학 수업이 영향을 끼쳤다고 직접 얘기하고 있다.

스포츠통계 마지막 퍼즐, 행동경제

그럼 앞으로 스포츠통계가 발전하는 지향점에 행동경제학은 또 어떤 영향을 미칠까?

우선 행동경제학을 사전적 정의에서 바라보자면 인간이 제한된 합리성에 기반하여 생각하고 행동하는 것을 연구하는 학문이다. 그런 관점에서 스포츠에서의 선수, 감독, 심판 등도 역시 인간인지라 제한된 합리성을 가지고 생각하고 행동함으로써 우리가 예측하는 범위 밖에서의 편향(Bias)을 보일 수 있다. 예를 들어 우리나라는 야구에서 ABS 존을 도입해서 그나마 다행이지만, 투수와 타자와의 대결에서 심판의 편향에 따라 스트라이크와 볼의 선언이 달라질 수 있고, 이 때문에 경기 결과가 달라질 수도 있다.

또한 지금까지의 스포츠통계분석은 기존 기록을 기반으로 다양하게 분석한 결과이기 때문에 실제로 어떠한 상황이나 맥락이 스포츠 경기에 어느 정도의 영향을 끼쳤는지 밝혀내기는 쉽지 않다. 즉, 인과관계에 있어서 독립변수가 종속변수에 영향을 끼쳤는데, 실제로 독립변수에 영향을 끼치는 맥락과 상황, 동기 등을 간과할 수 있다.

예를 들어 야구라는 종목을 들어보면 선수 대부분은 일년에 한번 이상 찾아오는 슬럼프가 있다. 그러면 우리가 그 슬럼프가 언제 어느 정도 오는지 예측할 수 있을까? 그 슬럼프가 신체적인 주기에 따라 나타나는 현상이라면 당연히 누적된 데이터로 분석이 가능하겠지만 외적인 상황에 따라 슬럼프가 온 것이라면 그 상황(자극)이 선수의 경기력에 영향을 끼치는지 아닌지는 RCT와 같은 실험에 따라 인과관계를 유추할 수 있다. 또 다른 예를 들어보자. 온라인 미디어가 발달해 있는 현 상황에서 온라인 상의 개인에 대한 댓글이나 언급의 종류나 정도에 따라 선수의 경기력에 심각한 영향을 끼칠 수도 있다. 이 역시 어떤 선수는 대수롭지 않게 여기는 반면, 어떤 선수는 심적 불안감 때문에 경기에 나서도 제 실력을 발휘할 수 없을 수도 있다. 위의 예처럼 어떠한 자극이나 동기가 선수의 행동을 유발할 수도 있는데 이러한 부분을 찾아낼 수 있는 것이 바로 행동경제학의 이론들이고, RCT라고 하는 무작위 통제실험이라는 기법이다.

앞서 언급하였듯 스포츠 통계학의 발전은 승률을 높이는 데 결정적 기여를 해 왔다. 그리고 이제는 지금까지 와는 다른, 남들이 보지 못하는 통계에 대한 고민과 분석을 더해야 하는 시점이 되었다.

결국 선수와 감독 모두 인간이기에 인간의 심리와 행동을 과학적으로 접근하는 행동경제학이야말로 새로운 변인들을 찾아내고 이에 따라 새로운 예측 모델을 만들어 내기에 가장 적절하지 않을까 싶다.

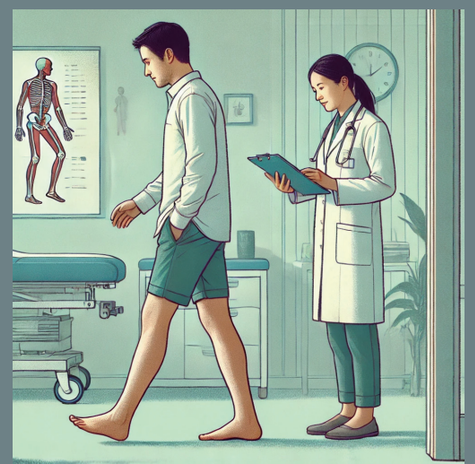
대량의 정량적 데이터 뿐만 아니라 비정형데이터 분석을 주요 학문 도구로 삼는 행동경제학이라는 렌즈를 통해서 우리는 예측의 정확도를 몇 %라도 높일 수 있을 것이다.

04 걷는다를 통계하다

안소연 | 분당서울대학교병원 교수

통찰

대학교에서 확률분포를 배웠다. 단순히 눈에 보이는 사건을 숫자로 표현할 수 있을 뿐만 아니라 사건의 확률로도 설명할 수 있었다. 시각적 심상을 수치적 가능성으로 드러낼 수 있다니. 보이지 않는 현상까지 수치로 표현하고 그 너머에 있는 불확실성은 확률로 설명하는 그 개념에 나는 깊이 매료되었다.
집과 학교는 꽤 멀었다. 수업에 늦을까 봐 종종 뛰었고, 환승 요금을 아끼려고 하염없이 걷기도 했다(환승요금을 당연히 내야 하는 시기가 있었다). 확률분포를 배우고 나니, 매일 발걸음수는 정규분포를 따를 거라는 생각이 들었다. 한동안 평균보다 많이 걸었다면, 고된 하루를 보낸 날에는 택시를 타도 괜찮겠지 하고 합리화할 수 있었다. 평균에 회귀하도록 스스로 평균을 조정하면서, 정규분포로 내 일상을 해석하였다.



걸음수뿐만 아니었다. 보폭이 커지면 걸음수는 줄어드니, 보폭도 확률분포로 표현하고 싶었다. 걷는 속도는 어떠한가. 꼬리에 꼬리를 무는 듯이 정규분포로 표현할 수 있는 현상을 찾아냈다. 그러나 아침 등교와 저녁 하교 사이의 간격은 정규분포로 표현하는데 한계가 있다는 생각이 들었다. 일정한 패턴을 찾기 어려웠기 때문이었다. 이런 의문은 이후 버스를 기다리는 시간 간격이 포아송 분포로 설명될 수 있다는 것을 배우고 나서 풀렸다.
이제는 더 이상 환승 요금을 걱정하지 않는다. 그렇지만 스마트폰에서 측정되는 걸음수는 물론이고 나의 일상 대부분이 계량화되는 시기에 살고 있다. 의료계에서는 인간의 겉으로 보이는 행동뿐만 아니라 보이지 않는 내부 장기의 변화를 비침습적으로 측정하고 인지적 지각과 감정까지도 수치로 표현하고자 노력한다.

정신건강의학과 교수님과 함께 보행 데이터를 분석한 경험이 있다. 비록 장비의 영점 조정 문제로 데이터 표준화가 어려웠지만, 그 과정에서 보행이라는 일상적인 행동을 어떻게 수치화하는지 배웠다. 보행

을 계량한다는 것은 곧 일상 생활의 패턴과 건강 상태를 측정할 수 있는 새로운 창을 열어주는 셈이다.

분할-계량-정복

모두 사람은 나름의 방식으로 세상을 바라본다. 심리학에서는 이를 프레임(frame)이라 부른다.[1] 각자의 프레임은 복잡한 세상을 이해하는 창이다. 예를 들어, 동메달을 딴 선수가 은메달을 딴 선수보다 행복하다는 역설적인 현상을 어떻게 설명할 수 있을까. 동메달을 딴 선수는 노메달과 비교하는 반면, 은메달을 딴 선수는 금메달과 비교하는 프레임을 가지고 있기 때문이다.

나에게는 통계가 내 삶의 프레임이다. 통계는 객관적이면서도 포괄적인 도구로, 복잡한 세상을 이해하는 척도가 된다.
복잡한 현상을 단순한 구성 요소로 분해하여 이해하고 이를 다시 조합하여 전체적인 현상을 설명하는 분할-정복 과정은 과학과 공학 분야에서 널리 사용되는 문제 해결 방식이다. 예를 들어, 물리학에서는 복잡한 운동을 뉴턴의 운동 법칙으로 단순하게 설명하고, 컴퓨터 그래픽스에서는 복잡한 이미지를 픽셀 단위로 분해하여 처리한다. 소프트웨어 개발 시 운영체제를 다양한 모듈로 나누어 개발하고 관리한다.

통계에서도 이런 분할-정복 원칙이 반복된다. 예를 들면, 가설 검정에서 귀무가설과 대립가설이라는 두 가설로 단순히 쪼개버린다. 귀무가설은 차이가 없다는 가정을 전제로 한다. 복잡한 현상을 단순화시켜 차이가 없다고 가정하고 이러한 귀무가설에 집중하여 검정을 실시하여 현상을 이해하려는 시도를 한다. 마치 법정에서는 무죄 추정의 원칙을 따르듯이, 통계학에서는 귀무가설 추정 원칙에 따라 데이터를 살펴보고 '합리적인 의심을 할 여지가 없을 만큼 확신을 가지는 정도의 증명력'('통계적으로 유의한 차이')을 보여야 귀무가설을 기각할 수 있다.

보행 분석 역시 이러한 분할-통치 원리가 적용되는 대표적인 데이터 영역이다.[2] 보행은 공간과 시간 요소로 나누어 분석할 수 있고, 보행 간격과 보행 속도가 각각의 대표적인 지표로 사용된다. 여기에 양발의 대칭성, 두 발이 모두 땅에 닿는 비율 등 동적 지표를 더해 분석할 수 있다.
흥미롭게도, 보행은 높은 신뢰도로 개인을 특정할 수 있다. 드라마에서 보듯이, 용의자의 보행이 마치 지문처럼 유력한 단서가 되기도 한다. 이는 '법보행'이라는 수사기법이다. 의료에서는 굳이 의료보행이라 명명하지 않더라도 보행은 이미 흔히 활용되는 지표이다. 특히, 정상에서 질환자로 이행되는 과정에서 보행은 중요한 건강 지표로 쓰인다. 신경계 질환을 진료하는 교수님은 환자가 진료실로 들어오는 걸음걸이만 보고도 질병이 진행되는 상태를 가늠할 수 있다고 말씀하였다.

예를 들어, 보폭이 줄어들어 종종걸음 치듯 걷는 현상은 파킨슨병의 전형적인 증상이다. 보폭이 좁아지고 발을 질질 끄는 듯한 모습이 특징적이다. 뇌졸중 환자는 한쪽으로 기울어진 편마비 보행을 보이는 경우가 있다. 치매 환자는 걸음속도가 느려지고 보행 변이성이 증가한다.
또한, 보행 속도와 수명이 관련된다는 연구결과도 꾸준히 보고되고 있다. 건강하고 인지 기능이 좋을수록 빠르게 걸을 수 있으며, 이는 일상에서도 잘 볼 수 있는 사실이라서 쉽게 이해가 된다. 심지어 '걷기만 해도 병이 낫는다'고 하지 않는가.

1) 최인철. (2007). 프레임: 나를 바꾸는 심리학의 지혜, 21세기북스
2) https://en.wikipedia.org/wiki/Gait_analysis
3) Muro-De-La-Herran, A., Garcia-Zapirain, B., & Mendez-Zorrilla, A. (2014). Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications. Sensors, 14(2), 3362-3394
4) 플랜비포유 홈페이지 https://www.planb4u.kr
❖ 그림은 시각적 표현을 돕기 위해 OpenAI의 AI 모델 ChatGPT를 활용하여 제작되었습니다

통합

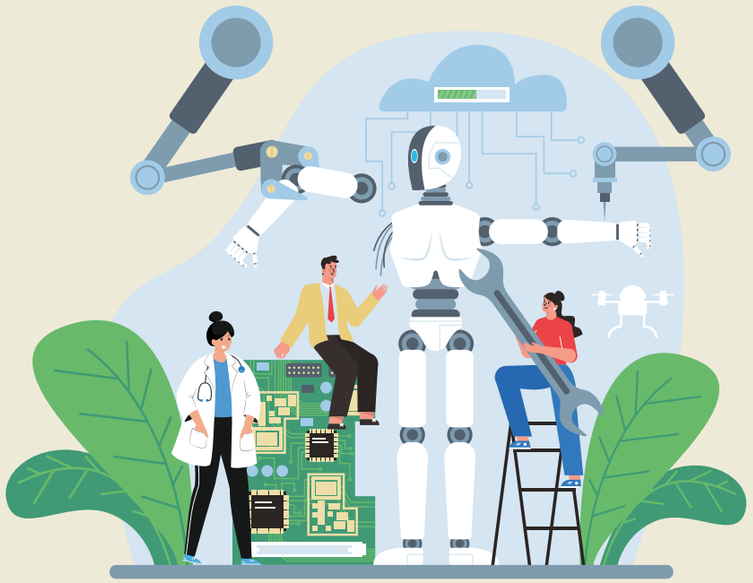
보행 데이터는 크게 가속도 센서와 자이로 센서를 통해 수집된다.[3] 공간에서의 3축 가속도를 측정하고, 회전 운동을 통해 방향 전환을 분석한다. 여기에 압력 또는 각도 센서가 추가되면 다차원 정보를 수집할 수 있다. 측정 장비로 카메라를 이용한 동작 분석 시스템이 가장 정확하지만 비용 역시 으뜸이다. 일반적으로 착용형 센서가 많이 사용된다. 최근에는 스마트폰의 내장 가속도 센서와 자이로 센서를 활용하기도 하고, 스마트 워치, 스마트 밴드, 스마트 슈즈, 심지어 스마트 깔창까지 다채로운 웨어러블 기기가 도입되고 있다.

여기에 더해, 보행 데이터를 실시간으로 분석해 개인의 건강 상태를 모니터링하거나 질병을 예측하려는 연구도 활발하게 이루어지고 있다. 예를 들어, 낙상 위험을 미리 파악하거나 퇴행성 뇌질환 초기단계를 감지하려는 시도 등이 있다. 즉, 보행 분석이 단순한 현상태를 측정하는 단계를 넘어서, 개인 맞춤형 예방과 치료로 나아가기 위한 도구로 활용되는 단계로 이행되고 있다.[4]

걷는다. 이 자연스럽고 일상적인 행위를 어떻게 인간의 생로병사와 엮을 것인가. 결국 우리는 숫자를 통해 세상을 바라보고 그 의미를 확률로 이해한다. 보행처럼 일상적이고 지극히 자연스러운 행동조차 우리가 미처 깨닫지 못했던 수많은 정보를 담고 있었다. 나누고 계량하고 다시 통합하여 통계 모델로 구현할 때 비로소 그 실체를 조금씩 이해할 수 있다. 통계를 통한 통찰은 복잡한 세상을 단순화하여 그 너머의 진실을 향해 걸어가는 여정이 된다.

한 국 표 준 직 업 분 류 현 황 과 최신 노동시장 변화를 반영한 8차 개정

류창진 | 통계청 통계기준과 사무관



개인이 집이나 직장 등에서 생활하는 모습을 측정하고 비교 분석하기 위해서는 개인활동에 대한 조작적 정의와 세부적인 측정기준 마련이 필요하다.

국제연합(UN) 통계위원회에서는 2016년에 개인의 모든 활동을 쪼개고 범주화하여 생활시간조사 행동분류(International Classification of Activities for Time Use Statistics)를 제정하였다. 또한, 국제노동기구(ILO)에서는 재화 생산 및 서비스 제공 활동(‘노동형태’)에 대해 2013년 노동활동분류(Classification of Work Activities) 체계와 2018년 국제노동상태분류(International Classification of Status at Work)를 제정하여 개인의 모든 활동 중 노동 활동에 대한 측정 틀을 마련하였다.

직업분류는 노동 활동 중 ‘임금이나 이윤을 목적으로 하는 노동(employment)’을 측정 대상으로 기업 내에서 개인이 수행하거나 수행해야 할 업무 및 과업(tasks and duties)에 대한 분류기준이며, ILO에서는 1958년에 국제표준직업분류(International Standard Classification of Occupation, ISCO-58)를 제정하였고 우리나라도 1963년에 한국표준직업분류(이하 ‘직업분류’)를 국제기준에 따라 제정하여 현재까지 운영중에 있다.

Intended destination of production	For own final use		For use by others				
Forms of work	Own-use production work		Employment (Work for pay or profit)	Other work*	Volunteer work		
	of services	of goods			in market and non-market units	in households producing	
						goods	services
Relation to 2008 SNA	Within SNA production boundary						
	Inside SNA general production boundary						

*Includes compulsory work performed without pay for others, not covered in the draft resolution.

2013년 ILO 노동활동 분류체계

한국표준직업분류 특징

직업분류에서 다루는 ‘직업’은 계속성, 경제성, 사회성 · 윤리성, 자율성 4가지 속성을 모두 충족해야 한다. 첫째, 직업은 유사성을 갖는 직무를 지속적으로 수행하는 계속성을 가져야 한다. 둘째, 직업은 경제적인 거래 관계가 성립하는 활동을 수행하는 경제성을 가져야 한다. 셋째, 직업은 전통적으로 사회성 · 윤리성을 충족해야 한다. 즉 비윤리적인 영리 행위나 반사회적인 활동을 통한 경제적인 이윤 추구는 직업으로 인정하지 않으며 사회성은 보다 적극적으로 사회 공동체에 기여를 전제조건으로 하고 있다. 넷째, 속박된 상태에서의 제반 활동은 계속성이나 경제성의 여부와 관계없이 직업으로 보지 않는다.

직업분류는 유사한 직무의 집합(‘직업’)을 직능수준과 직능유형에 따라 체계화한 것으로 대분류(분류부호 1자리), 중분류(2자리), 소분류(3자리), 세분류(4자리), 세세분류(5자리) 순으로 계층적으로 구성되어 있다. 여기서 직능수준(skill level)은 업무 수행에 필요한 지식, 경험 등의 난이도 정도이고, 직능유형(skill specialization)은 업무 수행에 필요한 지식, 경험 등의 전문분야를 말한다.

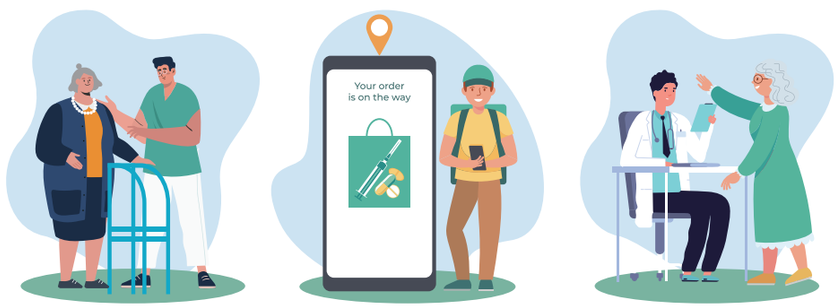
8차 직업분류 체계 현황표

대분류명	(직능수준)	대분류	중분류	소분류	세분류	세세분류
합 계		10	57	167	495	1,270
1 관리자	(제4직능, 제3직능)	1	5	16	25	85
2 전문가 및 관련 종사자	(제4직능, 제3직능)	1	9	48	177	486
3 사무 종사자	(제2직능)	1	7	13	35	63
4 서비스 종사자	(제2직능)	1	5	15	46	91
5 판매 종사자	(제2직능)	1	3	5	16	44
6 농림어업 숙련 종사자	(제2직능)	1	3	5	14	30
7 기능원 및 관련 기능 종사자	(제2직능)	1	9	19	79	197
8 장치·기계 조작 및 조립 종사자	(제2직능)	1	9	30	65	217
9 단순 노무 종사자	(제1직능)	1	6	12	33	52
A 군인	(제2직능 이상)	1	1	4	5	5

특히 직업분류는 ILO 국제기준(ISCO-08)에 따라 직능수준을 고려하여 대분류를 나누고 있어 직능의 구분이 중요하다. 이러한 직능(skill)은 정규교육, 비정규적인 직업훈련과 직업경험 등을 통해 얻어지며, 크게 4개 수준으로 구분되는데 직능수준이 올라갈수록 높은 이해력과 의사소통 능력 등이 요구된다.

직업분류는 경제활동인구조사, 지역별고용조사, 사업체노동력조사 등 108종의 국가통계에서 통계작성 목적이외에도 장·단기 인력수급 정책 수립과 직업 연구를 위한 기초자료로 활용되며, 구인구직 취업알선 정보, 보상액 결정 기준 등 제도적 또는 정책적 목적으로 사용되고 있다.

통계청은 통계법 제22조(표준분류)에 따라 통계작성기관이 동일한 기준을 적용하도록 직업분류를 작성·고시하고 있으며, 국제분류 변경이나 국내 노동상황 변화 등을 반영하기 위해 8차례 개정해 왔다. 이번 8차 직업분류 개정부터 2017년 제정된 통계분류 제·개정 업무처리지침(통계청 훈령)에 따라 5년 주기(연도 끝자리가 4, 9자년)로 적용하고 있다.



8차 개정 절차 및 주요내용

8차 직업분류 개정은 ’17년이후 노동시장 변화와 다양한 개정수요 등을 반영하기 위해 2022년부터 개정을 준비해 왔고 다방면의 개정절차(의견수렴, 기관협의, 개정심의, 개정안 마련, 국가통계위원회 심의)를 거쳐 2024년 7월 1일자로 개정 고시하였고 6개월의 준비기간을 두고 2025년 1월 1일자로 시행된다.

이번 개정의 특징은 ①사회환경 변화로 인한 분류항목 분리·신설, ②성장직업의 분류항목 신설 또는 세분, ③상대적 비중 감소된 직업의 분류항목 통합, ④직업분류 개정수요 반영 및 직업분류 체계 개선 등이다.

첫째, 포스트코로나, 저출산·고령화 등 사회환경 변화에 대응하기 위한 보건 및 돌봄 관련 인력 확대의 영향으로 중분류 보건 전문가 및 관련직, 돌봄 및 보건서비스직 등을 각각 분리·신설하였다. [표1]

보건 전문가 및 관련직은 수의사 항목 분리에 따른 소분류 의사·한의사 및 치과의사의 명칭 변경, 직무 차이에 따른 세분류 약사와 한약사의 항목 분리, 청

[표1] 중분류 보건 전문가 및 관련직, 돌봄 및 보건서비스직 등을 각각 분리·신설

24	보건·사회복지 및 종교 관련직	24 보건 전문가 및 관련직(의사, 수의사, 간호사 등)
		25 사회복지·종교 전문가 및 관련직(보육교사 등)
42	돌봄·보건 및 개인 생활 서비스직	42 돌봄 및 보건 서비스직(아동돌봄 종사자, 요양보호사·간병인 등)
		43 개인 생활 서비스직(미용 관련 서비스 종사자 등)

각능력 재활사(audiologist) 신설에 따른 세분류 언어 및 청각능력 재활사의 분류수준 상향, 환자안전 전담인력 수요에 따른 세분류 환자안전 관리사 항목 신설 등이 있었고, 사회복지·종교 전문가 및 관련직은 유치원교사와 동일한 수준 유지을 위해 소분류 보육교사 수준 상향과 시민사회 활동의 활성화를 위해 시민사회 활동가 명칭 변경 및 수준 상향 등이 있었다.

또한, 돌봄 및 보건서비스직은 아동 및 노인 돌봄의 직무 세분화와 통계생산 가능성을 고려하여 소분류 교사보조 및 아동돌봄 종사자, 요양보호사 및 간병인, 노인 및 장애인 돌봄 종사자 각각을 분류수준 상향·항목 세분화 등이 이루어졌고, 개인생활 서비스직은 반려동물 관련 직업수요를 감안하여 소분류 동물관련 서비스 종사자를 분류수준 상향 및 항목 세분화 등이 있었다.

둘째, 인공지능 등 데이터 활용 확산, 플랫폼 노동 및 신산업 성장 등 노동시장 변화에 맞춰 고용비중이 확대되는 직업분류 항목 신설과 분류수준 상향 등을 통해 통계 활용성을 높였다. [표2]

셋째, 자동화·직무전환 등의 영향으로 노동시장 규모축소에 따라 금형·주조 및 단조원, 제관원 및 판금원, 용접원을 금속성형 관련 기능종사자로 통합, 인쇄필름 출력원 등 세세분류를 인쇄관련 기계조작원으로 통합하여 분류항목을 축소하였다.

넷째, 직업분류 개정 과정에서 1,036개 기관 및 일반이용자를 대상으로 네 차례 의견수렴을 실시하여 제출된 내용(총 62건)을 검토하여 행정사를 사무종사자에서 전문가 및 관련종사자로 분류이동, 네일관리사 분류 상향, 자원봉사 관리원 항목 신설 등 일부 사항을 반영하였다.

[표2] 고용비중이 확대되는 직업분류 항목 신설과 분류수준 상향

직업항목 신설

직업교육훈련 및 평생교육 기관 관리자, 신재생에너지 관련 관리자,
소프트웨어 품질관리 전문가, 디지털 포렌식 전문가,
데이터 시스템 전문가, 실감형 콘텐츠 디자이너, 장애인 직업 상담사,
예술품 및 문화재 감정 전문가, 영상 및 미디어 예술가, 동물보건사,
병원 배식원, 로봇 설치 및 정비원, 전기자동차 조립원,
이차전지 제조 기계 조작원, 늘찬배달원, 정리 수납원 등

분류수준 상향

데이터 전문가(세분류, 4자리)		→	데이터 전문가(소분류, 3자리) 등	
•223	데이터 및 네트워크 관련 전문가		**•224**	**데이터 전문가**
•2231	**데이터 전문가**		•2241	데이터 시스템 전문가
•22311	데이터 설계 및 프로그래머		•22411	데이터 설계 및 프로그래머
•22312	데이터 분석가		•22412	데이터베이스 관리 및 운영자
•22313	데이터 관리 및 운영자		•2242	데이터 분석가
			•22420	데이터 분석가

또한, 직업분류의 체계 개선을 위해, 고용규모가 상대적으로 컸던 중분류 경영 및 회계 관련 사무직은 직무 유사성과 적정 고용규모를 고려하여 중분류 항목을 세분화하였고, 소분류 청소원 및 환경미화원은 직무 배타성과 국제기준(ISCO-08)을 고려하여 청소대상별로 세분류 항목을 재편하였다.

8차 개정에 따른 항목 변화 및 활용 지원

8차 직업분류 개정 경우 기본개념과 대분류 체계는 7차 대비 변함이 없으나, 분류의 현실적합성을 높이기 위해 중분류 5개, 소분류 11개, 세분류 45개, 세세분류 39개 항목을 각각 증가시켰고, 개정유형별로 보면 분류항목 분리·신설 109개, 통합 27개, 이동 54개, 분류범위 변경 28개, 분류명칭 변경 146개 등이 변화되었다.
이번 개정을 통해 반도체·로봇·신재생에너지·전기차 등 산업 육성을 위한 관련 전문가 및 기술공, 인공지능 활용 확대와 관련 데이터 전문가, 개인정보 보호를 위한 정보보안 전문가, 저출산·고령화에 따른 아동돌봄 종사자와 노인 돌봄 종사자 및 간병인, 배달플랫폼 사용 확대에 따른 늘찬배달원 등 최신 직업변화가 반영된 직업분류가 보다 현실성 있는 통계 작성과 각종 정책에 시의성 있게 활용되길 기대한다.

통계청은 ’24년 현재 직업분류를 포함한 통계분류를 39종 제정하여 운영 중이며, 특히 법적 준수의무가 있는 표준분류의 경우 개정시 마다 분류 해설서 및 항목 분류표, 분류항목(신-구, 구-신) 연계표, 색인어 검색 등을 통계분류포털(https://kssc.kostat.go.kr)에서 제공하고 있다. 또한 통계분류 이용자가 개정된 직업분류를 적절히 사용할 수 있도록 e-컨텐츠 개발, 정기 및 부정기 교육 등 활용 지원을 지속할 계획이다.

8차 개정에 따른 신설 및 확대 직업


로봇 설치·정비원


전기자동차 조립원


신재생에너지 관리자

정보보안 전문가


데이터 전문가


요양보호사 및 간병인


늘찬배달원 (퀵서비스 배달원)

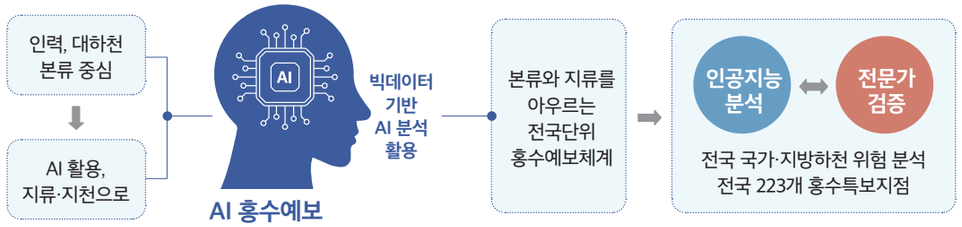

동물보건사

인 공 지 능 기 반 홍수예보 현황 소개

이건행 | 한강홍수통제소 수자원정보센터 기상연구사

기후변화로 인해 서울시 도림천, 포항시 냉천과 같은 중소하천에서 극한 홍수피해가 빈번하게 발생하고 있으며, 기상이변에 따라 갈수록 어려워지는 홍수 예측으로 인해 재해에 대한 사전 대비의 중요성이 커지고 있다. 환경부는 2023년 10월, 이러한 변화에 대응하기 위해 추진해 온 빅데이터 기반의 인공지능 홍수예보 체계를 조기에 구축하고, 이를 활용한 홍수특보 지점을 기존 75개소에서 223개소로 크게 늘린다고 밝혔다. 특히 규모가 작은 지방하천의 홍수특보 지점을 기존 12개소에서 130개소로 대폭 늘려 중소하천의 홍수 피해에 대응하고 있다.

홍수통제소는 환경부의 소속기관으로 전국에 4개의 홍수통제소(한강, 낙동강, 금강, 영산강)가 있다. 수자원정보센터는 4대강 홍수통제소의 공통 업무를 수행하는 부서로서 인공지능 홍수예보 체계 구축 사업을 추진하였다. 이 글에서는 수자원정보센터에서 개발한 인공지능 기반의 홍수예측 기술과 이를 이용하여 환경부 및 홍수통제소에서 운영 중인 인공지능 홍수예보시스템, 그리고 올해부터 달라진 홍수예보 체계들을 소개하였다.



[그림 1] 인공지능을 활용한 홍수예보 개념

홍수예보와 홍수특보

기상청에서 발령하는 호우주의보, 호우경보와 같은 호우특보는 뉴스나 날씨정보에서 자주 접하는데 이에 비해 홍수특보는 낯설 수 있다. 홍수통제소에서 발령하는 홍수특보는 홍수예보의 한 종류로 지금까지 내린 비와 앞으로 내릴 비의 양을 이용하여 앞으로 하천의 수위가 어떻게 변화할 것인지 예측하여, 각 지점마다 설정한 기준 수위에 도달할 것이 예상되면 유관기관과 국민들에게 알려주는 것이다. 참고로 홍수예보에는 홍수특보와 홍수정보가 있으며 좁은 의미로서 홍수정보는 하천에서 관측된 수위를 유관기관의 담당자에게 문자서비스 등을 이용하여 실시간으로 알려주는 것이다.

홍수특보는 홍수주의보와 홍수경보로 나뉘는데, 하천이 견딜 수 있는 유량(계획홍수량)의 50%를 초과할 것이 예상되면 홍수주의보, 70%를 초과할 것이 예상되면 홍수경보를 발령한다.[1] 즉 홍수특보를 발령할 때에는 하천의 수위가 기준 수위를 초과할 것인지 아닌지를 예상하는 과정이 필요한데, 이를 과학적으로 표현하면 예측이 된다. 인공지능 기반의 홍수예보는 이러한 하천 수위 예측 과정에 인공지능 기술을 이용하는 것을 의미한다.

인공지능을 이용한 홍수예측 기술의 도입

기존에는 방대한 자료를 사람이 물리 모형을 이용해 실시간으로 분석을 실시하고 홍수특보 발령여부를 판단하였다. 물리 모형이란 유역과 하천의 물리적인 특성들을 이용하여 하천의 유량을 수학식으로 계산하는 프로그램을 말한다. 하천의 수위를 감시하다가 위험한 하천이 포착되면, 물리 모형을 실행하여 앞으로 기준 수위를 초과할 것인지 아닌지를 예측하고, 초과할 것으로 판단되면 발령문을 작성하여 특보를 발령하는데, 이 과정까지 30분 정도가 소요되었다.

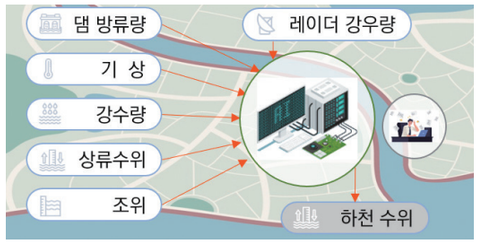
반면 강우량과 댐 방류량, 하천 수위 등의 통계적 상관관계를 10년 치 이상 학습한 인공지능(AI) 홍수예측 모형은 10분마다 자동으로 하천 수위를 계산한다. 자료가 수집된 이후, 컴퓨터를 통해 인공지능 모형이 하천의 수위 예측 결과를 내놓는데 까지 몇 초 밖에 걸리지 않는다.

이와 같이 인공지능 모형에 의해 예측에 필요한 시간이 획기적으로 단축됨에 따라 과거보다 많은 홍수특보 지점을 운용하는 것이 가능해졌다. 물리 모형은 그림 2의 (가)와 같이 하천의 유량에 관여하는 몇몇 요소를 이용하여 하천의 유량을 계산하고 계산된 유량은 하천에서 측정한 유량과 수위와의 관

1) 계획홍수위를 기준으로 하는 경우, 60% 수준은 홍수주의보, 80% 수준은 홍수경보

(가) 물리 모형에 사용되는 입력자료

(나) 인공지능 홍수예측 모형에 사용되는 입력자료

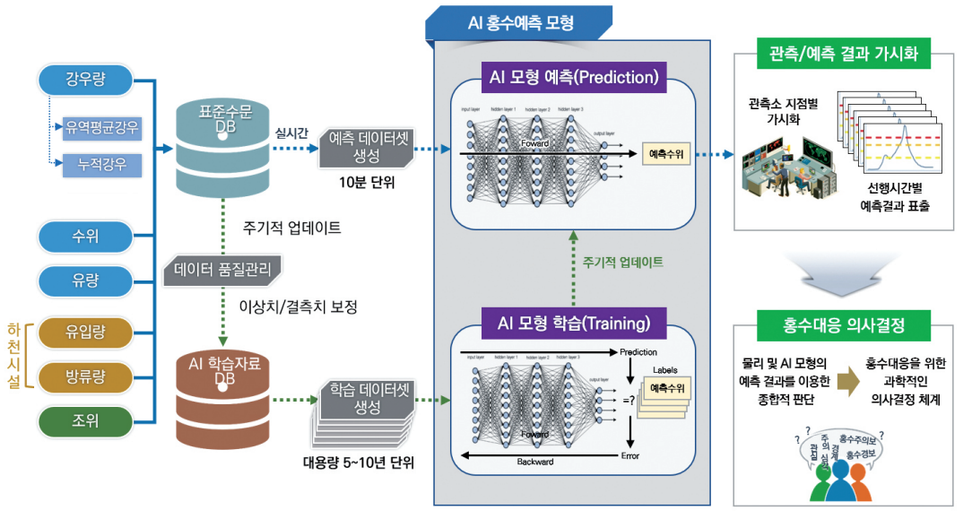


[그림 2] 홍수예측을 위한 물리 모형과 인공지능 모형의 구성

계식을 통해 수위로 변환된다. 반면 인공지능 모형은 그림 2의 (나)와 같이 하천 수위에 영향을 주는 보다 많은 요소를 학습에 이용하고 하천 수위를 직접 예측한다.

인공지능 모형의 학습에는 환경부의 강우량과 수위 자료가 주로 사용되었다. 여기에 한국수자원공사의 댐과 보, 한국수력원자력의 댐, 한국농어촌공사의 저수지 등의 방류량 자료가 추가로 사용되었다. 하구에 위치하고 있어 조위의 영향을 받는 특보지점은 조위 관측소 자료를 포함하여 학습 데이터셋을 구축하였다(그림 3). 학습에 사용된 자료 중, 하천의 수위 변화에 가장 큰 영향을 주는 강우량 자료는 다른 형태로 가공함으로써 학습의 효과를 높일 수 있고, 이에 따라 예측 성능도 향상시킬 수 있다. 일례로, 강우량을 3시간 혹은 6시간 누적 강우량 자료로 변환하여 활용함으로써 땅에 내린 비가 하천으로 흘러들어갈 때까지의 시간과 흘러들어간 물이 하천을 따라 아래로 흐르는 시간을 고려할 수 있도록 하였다. 또한 특정 지점에서 관측된 강우량은 유역의 면적평균강우량으로 바꾸어 학습자료에 추가함으로써 인공지능이 유역의 개념을 학습할 수 있도록 유도하였다.

학습 데이터셋을 구성하고 모형을 모두 구축한 이



[그림 3] 인공지능 홍수예측 모형의 구성 및 예측 흐름도

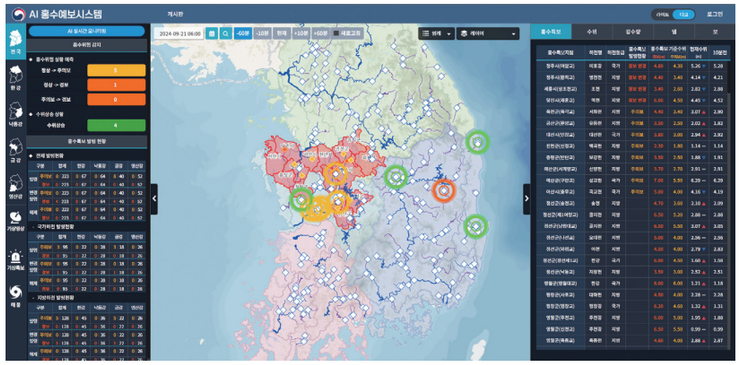
후, 특보지점의 특성을 잘 파악하고 있는 각 홍수통제소 홍수예보관의 의견을 통해 학습 데이터셋을 수정하는 과정을 거쳤다. 홍수예보관의 의견에 따라 일부 지점은 불필요한 학습자료를 제거하여 학습효과를 높였으며, 일부 지점은 새로운 자료를 학습 데이터셋에 추가하여 예측 성능을 향상시켰다.

학습에 사용한 인공지능 기술은 머신러닝(Machine Learning)의 지도학습과 딥러닝(Deep Learning)의 LSTM(Long Short-Term Memory)이다. LSTM은 RNN(Recurrent Neural Network)의 장기의존성, 즉 시계열 자료가 증가함에 따라 점차 과거 정보를 활용하기 어려워지는 문제를 해결하기 위해 제안된 모델이다. 학습이 완료된 인공지능 홍수예측 모형은 향후 6시간까지의 예측결과를 제시한다(그림 3).

인공지능 홍수예측 모형은 하천 수위 상승 가능성을 1차적으로 판단하여 홍수예보관이 직관적으로 확인할 수 있도록 홍수예보시스템에 알람으로 알려준다(그림 4). 이를 감지한 유역별 홍수통제소의 홍수예보관은 물리 모형을 활용하여 인공지능이 예측한 결과의 적정성을 확인한다.

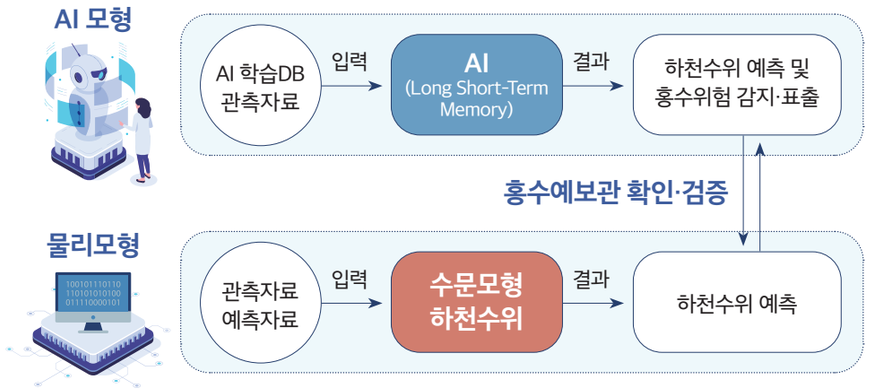
즉, 그림 5와 같이 인공지능 홍수예측 모형은 기존의 물리 모형을 대체하는 것이 아니라 홍수특보 지점의 수위를 효율적으로 감시하고 예측하기 위한

(가) 하천 상승 관측 및 예측에 대한 알람



(나) 현 수위 현황 및 인공지능 홍수예측 상황 모니터링

[그림 4] 인공지능 홍수예보시스템 화면 예(2024.9.21.)

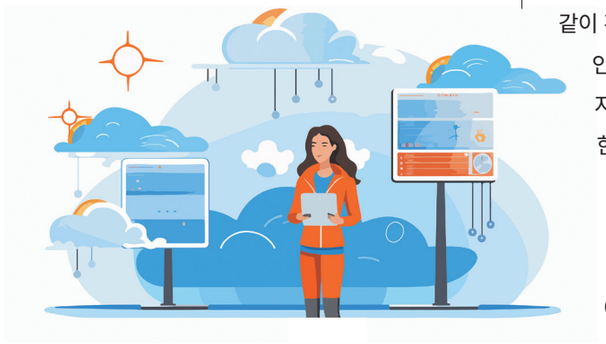


[그림 5] 홍수특보 발령시 인공지능 모형과 물리 모형 교차 검증 과정

수단으로써 추가로 도입되었다고 할 수 있다. 홍수예보관은 하나의 예측 모형에 의존하는 것이 아니라 두 모형의 결과를 비교하여 예측 결과가 유효한지의 여부를 판단하고 홍수특보 발령 여부를 결정하는 것이다.

홍수예보 체계의 전환

금년부터의 홍수예보는 지금까지 설명한 홍수예측 방법의 추가 외에 많은 개선 사항이 있었다. 먼저, 홍수특보의 전달 방법을 다양화하여 전달체계를 강화하였다. 기본적으로 일반 국민과 관계 기관에는 문자서비스를 통해 홍수특보를 전달하고, 재난안전통신망을 이용하여 지방자치단체 및 공공기관과 실시간 소통을 가능하도록 하였다. 특히, 지자체 부단체장에게 홍수특보 발령 사항을 음성메시지로 통보하고 수신을 확인받는 체계를 도입하였다. 부단체장 지휘 아래 실질적인 조치가 빠르게 이루어지도록 하기 위해서다.



국민들이 어디서든 쉽고 정확하게 홍수 위험을 확인할 수 있도록 대국민 알림 문자를 개선하였다. 작년까지는 홍수특보가 발령될 경우 대국민 알림 문자를 글자로만 제공하였으나, 올해부터는 문자에 웹페이지로 연결되는 링크 주소가 추가되었다. 그림 6과 같이 문자를 받은 국민들이 링크 주소를 누르면 안내페이지가 나타나며, 여기서 '내가 침수우려지역 안에 있는지 확인하기'를 누르면, 스마트폰 위치정보(GPS)를 이용하여 그림과 같이 내 위치를 중심으로 홍수주의보 발령 지점의 위치와 침수우려지역을 같이 확인할 수 있다. 또한, '전체 침수우려지역 확인하기'를 누르면 홍수주의보가 발령된 대상 지역 전체의 침수우려지역을 확인할 수 있다.

한편, 그동안의 홍수 위험 정보는 주로 문자로 전달되다 보니, 운전자의 경우 운전 중에 재난방송을 듣지 않은 이상 이를 쉽게 확인하는 방법이 없었다. 이에 환경부는 운전자에게 홍수경보 등 위험 상황을 실시간으로

[그림 6] 홍수특보 발령 시 침수우려지역 확인 서비스

알릴 수 있도록 과학기술정통부 및 국민들이 주로 이용하는 6개 내비게이션 업계(카카오내비, 티맵, 네이버지도, 현대차, 아틀란, 아이나비)와 협업하여 올해 7월부터 본격적으로 서비스를 시작하였다.

이제 운전자들은 운전 중에도 홍수경보 발령 시, 해당 지점 인근에 진입하면 내비게이션 화면과 음성안내를 통해 직접 위험 상황을 인지하고 진입 전 속도를 줄이는 등 주의를 기울일 수 있게 된다. 인공지능을 이용하여 홍수를 예측하여 내비게이션까지 연동하는 이 서비스는 행정안전부 정부혁신 왕중왕전 '미래를 대비하는 정부' 분야 14개 우수사례 중 1개로 선정되기도 하였다.

이 외에 홍수정보시스템(https://n.flood.go.kr)을 운영하여 전국의 하천 상황과 홍수특보 상황을 누구나 확인할 수 있도록 하고, 홍수위험지도정보시스템(https://www.floodmap.go.kr)을 클라우드에서 운영함으로써 안정적인 홍수위험지역 정보를 제공하도록 도모하였다.

인공지능 홍수예보 체계의 미래

여름철 자연재난 대책기간(5월 15일~10월 15일)이 거의 끝나가고 있다. 금년 홍수특보 발령 횟수는 과거 10년 평균 34회에서 금년 170회로 약 5배 증가한 것으로 나타났다. 올해 처음 지정된 중소하천의 홍수특보 지점에서 더욱 많은 특보가 발령되었다. 하천이 작으면 강우의 변화에 하천의 수위가 민감하게 반응하므로 예측의 정확도와 선행시간을 확보하는데에 어려움이 있다. 이는 물리 모형과 인공지능 모형 모두 마찬가지이다. 인공지능 홍수예측 모형을 효율적으로 활용하고 정확도를 높이기 위해서는 학습 데이터셋 재구성 및 재학습하고 새로운 이벤트가 발생하면 새로이 학습시키는 등 홍수특보 지점의 특성을 가장 잘 파악하고 있는 홍수예보관의 지속적인 관리가 필요하다.

또한 인공지능 홍수예측 모형은 한 개의 홍수특보 지점에서의 학습자료의 조합을 통해 매우 다양해질 수 있으므로 모형의 관리, 학습, 배포 등을 효과적으로 관리하기 위한 MLOps(Machine Learning Operations)를 도입을 추진 중에 있다. 인공지능 홍수예측 모형의 활용된 학습자료들은 지금도 계속 관측, 생산되고 있기 때문에 예측 정확도는 지속적으로 개선될 것이며, 이에 따라 국민들이 체감할 수 있는 신속하고 정확한 홍수정보를 제공할 수 있을 것으로 기대한다.

기네스 맥주와 얽힌 이야기

통계와 기네스북

최용석 | 부산대학교 통계학과 교수

Can I have one Pint of Guiness ?

필자는 스코틀랜드와 잉글랜드에서 각각 살면서 선술집 펍(Pub)에서 어둡고 진하고 고소한 향이 깊은 흑맥주인 기네스 맥주를 주문하곤 했다. 기네스 맥주는 피시엔칩스(Fish and Chips)와는 아주 잘 어울린다. 도수는 4.2%로 1 파인트(Pint, 570ml) 잔의 상단에 거품이 구름처럼 덮여있어 몽환적 분위기를 자아내기도 한다. 실제 기네스(Guinnes)라는 이름으로 통일되어 불리기 전에는 포터(porter), 스타우트(stout), 스타우트 포터(stouer porter)로 알려져 있었다 한다. 특히, 기네스 잔이나 광고의 황금색 하프 문양은 중세 시대 스코틀랜드 중심으로 살았던 조상 켈틱(Celtic)의 것으로 매우 고풍스럽다. 실제로 스코틀랜드 위쪽 하이랜드(Highland)를 포함하여 특히 스카이섬(Isle of Skye)에 가보면 도로 표지판의 그 당시 조상들이 사용하던(스코틀랜드 켈트어로 불리기도 하는) 게일어(Gaelic)가 왠지 이질적으로 느껴진다. 그러나 옛 조상들의 언어와 전통을 이렇게 지키려는 모습은 경이롭다.

여기서 영국의 역사와 문화를 논하자는 것은 아니다. 처음 시작한 기네스 맥주와 관련된 두 가지 이야기를 하고자 한다. 첫 번째 이야기는 통계에 대한 것으로 1759년 영국 잉글랜드 후손 아서 기네스 경은 아일랜드 더블린에 양조회사를 설립하였다.
이 회사는 1833년까지 영국에서 가장 큰 규모로 운영되었고 1886년 런던에도 양조장을 설립하였다. 기네스는 영국 양조 무역에서 유일한 위치를 차지하게 되었다고 한다. 다만 기네스 자사가 직판하기 위한 술집이 없었던 까닭에 항상 맛을 개선하고 판매 촉진을 위한 아이디어를 필요로 하였다. 그러던 중 1899년 더블린에 위치한 기네스 양조장에서 아일랜드 수학자인 고셋(William Sealey Gosset, 1876-1937)이 근무하게 된다. 그는 옥스포드 대학 뉴 컬리지에서 화학과 수학을 배웠고 취업 후 통계 지식을 양조와 농업(보리 개량) 모두에 적용하면서 실제적인 연구를 거듭했다. 그는 술의 맛과 질을 위한 적정 효모 투입량을 관리하기 위하여 소표본에 적합한 검정을 발견하였다. 더군다나 이를 위한 분포를 발견하였다.

고셋의 Student's t-test

여기서 잠깐 기네스의 어떤 연구 환경과 주제가 그를 통계학자로 만들어 주었는지를 살펴보자. 기네스는 새로운 접근 방식을 지지하면서 양조를 과학적으로 만드는 프로젝트를 착수했다. 즉, 기네스는 맥주의 생명으로 첨가물과 방부제가 없고 저온 살균도 하지 않음에도 불구하고 그 맛을 유지하도록 다양한 관점에서의 연구가 필요했다. 이를 위해 옥스퍼드나 케임브리지 대학교를 처음 졸업한 우수한 화학자를 양조자로 채용하여 브루어로 불렀고 최고 경영진으로서 직책을 맡겼다고 한다.

실제 고셋은 이 당시 기네스에 채용된 우수한 화학자 인재 6명 중 한 명으로 1899년 화학 분야 1급 학위를 받은 화학자였다. 이전까지의 기네스는 양조의 검은 마법(all the black magic of brewing)과 정성적 기준의 전통적 관행을 따랐다. 수확량, 수분, 보리, 옥수수의 크기, 질감에 대한 오래된 정성적 평가가 중요했다고 한다. 이에 반해 새롭게 구성된 연구진은 건조 조건, 재배 환경, 거름을 고려한 홉과 보리의 양조 품질을 향상 시킨 것에 관심을 두었다. 드디어 1902년 보리의 맥아 품질은 질소 함량에 따라 다르다는 사실에 자료를 축적했으나 분석, 실험, 측정을 통해 얻어진 다양하면서 제각각인 결과값에 혼란스러워 하였다. 1907년에는 동료들이 토양 및 날씨 문제, 보리 분류 및 식별, 유전학과 보리 사육을 연구 하는 동안 토지의 다양함, 강우를 포함한 씨에 주목하였다. 그러나 실제 맥주 맛의 자료에 영향을 미치는 원인 규명과 방법이 필요했고 이와 관련된 수치적 문제에 부딪힌 그들은 이를 고셋에게 가져갔고 이로 인하여 그는 기네스에서 전체 통계 분석을 담당했다고 여겨진다.

William Sealy Gosset

Ronald Aylmer Fisher

기네스 맥주 : 1파인트

고셋은 성품이 온화하고 수학 문제 해결에 관심이 많았고 더군다나 옥스퍼드에서 수학을 좀 한 그는 동료들의 이야기를 듣고 그들의 우려를 매우 빠르게 파악하고 항상

최선을 다해 아이디어를 도출하곤 했다고 한다.

이들의 문제 중 가장 문제가 되었던 점은 실험 자료를 해석하는 데 있어 자료의 수가 매우 작다는 점이었다. 예를 들어, 고셋은 소표본 보리재배 실험에서 4개 표본으로 4곳의 밭에서 수확한 생산량의 평균과 표준편차의 추정값에서 발생한 오차의 한계 문제를 당대 통계학 분야의 거장인 피어슨(Karl Pearson, 1857-1936)에게 문의했으나 답을 들을 수가 없었다. 여기서 실제 고셋이 기네스 연구소에서 동료들과 고민한 문제를 정리하여 The Probable Error of the Mean(평균의 가능한 오차)라는 제목으로 1908년 저널 Biometrika에 기고하였다. 그러나 이때 자사 제품과 관련된 연구 내용의 발표를 금지하는 회사와의 협의로 자신의 이름 대신 Student(학생)를 사용하였다. 그래서 Student's t-test라는 검정방법을 소개한 것이다. 그가 제시한 것은 크게 두 가지인데 첫 번째로 임의 자료가 특정한 분포를 따르는지를 관찰하는 것으로 이를 위해서 피어슨이 이미 1900년에 개발한 카이제곱검정을 활용하는 것이었다. 더 구체적으로 t-검정에 적용한 소표본이 t-분포를 따르는지가 더 궁금한 것이었다. 실제 고셋이 보여준 자료를 활용하여 기초 통계학 개념으로 정리 요약하려 한다.

특정한 분포의 확률밀도함수 $f(x)$를 통해 0에서부터 1.7까지와 그 이상의 값에 대해 계산된 기대도수 E_j와 관찰도수 O_j의 두 종류(I, II)에 대한 분할표가 다음과 같이 주어져 있다. 여기서 확률밀도함수는 다음과 같다.

$$f(x) = \frac{16 \times 750}{\sqrt{2\pi\sigma^2}} x^2 e^{-\frac{2x^2}{\sigma}}$$

기대 도수	1.5	10.5	27	45.5	64.5	78.5	87	88	81.5	71	58	45	33	23	15	9.5	5.5	7
관찰도수 I	3	14.5	24.5	37.5	107	67	73	77	77.5	64	52.5	49.5	35	28	12.5	9	11.5	7
관찰도수 II	2	14	27.5	51	64.5	91	94.5	68.5	65.5	73	48.5	40.5	42.5	20	22.5	12	5	7.5

관찰도수가 특정한 분포를 만족한다는 귀무가설에 대해 적합성검정(test of goodness of fit)으로 다음과 같은 카이제곱통계량을 활용한다.

$$X^2 = \sum_{j=1}^{c} \frac{(\text{관찰도수} - \text{기대도수})^2}{\text{기대도수}} = \sum_{j=1}^{c} \frac{(O_j - E_j)^2}{E_j}$$

여기서 c는 자료의 수이고 근사적으로 자유도 df=c-1인 카이제곱분포를 따른다. 실제 R의 함수 chisq.test()를 활용하면 관찰도수 I의 경우 $X^2 = 48.096$이고 유의확률 p-값은 0.000으로 유의수준 5%에서 귀무가설을 기각하게 된다. 즉, 특정한 분포를 만족하지 않음을 보여준다. 관찰도수 II의 경우 $X^2 = 21.811$이고 유의확률 p-값 0.192는 유의수준 5%에서 귀무가설을 채택하게 되어 자료가 특정한 분포에 잘 적합 되었다고 볼 수 있다.

고셋과 피셔의 만남

두 번째로 고셋이 인용한 소표본 실험인데 종자로 사용되는 밀이나 옥수수 등의 수확량은 씨앗의 강도에 영향을 받는다는 것을 보여주는 사례이다. 사례에서 부드러운 씨앗이 옥수수와 짚 모두에서 더 높은 수확량을 생산한다는 사실을 보여주기 위해 고셋은 t-검정을 활용하고 있다. 두 집단의 평균 검정을 고려하면 되는데 고셋은 차이에 대한 평균과 표준편차에 대한 개념으로 설명하고 있다. 여기서는 짚의 생산에 대한 자료를 활용하기로 하고 첫 번째 사례와 같이 R을 활용하여 현대적으로 설명하고자 하며 실제 고셋의 계산에 오류가 있어 이를 수정하여 설명한다.

실제 표에서 부드러운 씨와 단단한 씨의 수확량 차이의 평균 $\bar{x} = 1.482$과 표준편차 $s = 1.405$로부터 그의 z계산은 $z = \bar{x}/s = 1.482/1.405 = 1.055$이다. 오늘날 t-검정의 통계량 $t = \frac{\bar{x}}{s/\sqrt{n}}$에 따르면 R의 패키지 BSDA의 함수 z.test()에서 제공하는 검정통계량 값 $z = 2.583$로부터 자료의 수 6을 고려한 $2.583/\sqrt{6} = 1.055$로부터 고셋이 정의한 z를 얻을 수 있다.

기네스의 동료 연구자들로부터 통계 문제에 대한 동기부여를 받고 문제 해결에 고민했던 고셋의 노력은 좋은 평가를 받아 1922년 그는 양조장에서 통계 컨설턴트가 되었고 1934년까지 통계부서를 운영하였다. 통계학자로서 어쩌면 요즘 기네스의 풍부한 맛이 고셋의 Student's t-검정에 바탕을 둔 연구 덕분이라고 단정 짓고 싶다.
소표본에서 고셋의 연구에 대한 중요성을 찾아내고 수리적 근거를 제시한 또 한 사람의 학자로 피셔(Sir Ronald Aylmer Fisher, 1890-1962)가 있다. 그는 고셋보다 14살 아래로 영국 런던 출생의 수리통계학자와 우생학자로 현대 통계학의 창시자로 알려져 있다. 1925년 피셔는 통계방법에서 가장 영향력 있는 교재를 집필하면서 서문에서 고셋 Student를 기리며 그의 연구가 두 평균값을 비교하고 회귀계수의 정확한 추정에 활용되는 해결책을 제공한다고 기술하며 이로 인해 고셋은 유명해지기 시작했다고 한다.

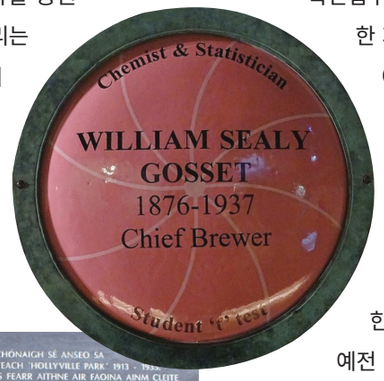
(단위: g)

연도	1899		1900		1901		통계량		
토양의 흙	가벼움	무거움	가벼움	무거움	가벼움	무거움	평균	표준편차	z
부드러운 씨	12.81	12.97	22.22	20.21	13.97	22.57	17.442		
단단한 씨	10.71	12.48	21.61	20.26	11.71	18.96	15.960		
차이	2.10	0.39	0.58	-0.05	2.26	3.61	1.482	1.405	1.055

이러한 피셔는 대표본 활용과 카이제곱분포의 자유도 문제를 포함하여 여러 문제에 대한 논쟁을 피어슨과 주고 받은 것은 유명하다. 아무튼 이렇게 도움을 준 피셔와는 실험계획법(experimental design) 문제로 논쟁을 벌이곤 했는데 고셋이 먼저 세상 떠난 이유도 있지만 실험에서 임의성(randomness)이 완성도가 높다고 평가를 받은 피셔의 이론이 더욱더 자리매김했다.

첫 번째 이야기를 고셋의 말년 삶으로 마무리 짓고자 한다. 1934년 고셋은 자동차 사고로 3개월 동안 병간호를 받는 동안 통계학에 전념할 수 있었지만 남은 여생동안 절름발이가 되었다. 1935년 말 그는 런던에 있는 새 기네스 양조장을 운영하기 위해 아일랜드로 떠났다. 고된 양조장 사업의 일에도 불구하고 그는 계속 통계학 논문을 발표하였다.

여기서 그를 기리는 동판 두 개를 소개한다. 하나는 기네스에서 고셋을 기리며 원형으로 기네스를 숙성시키는 저장통 뚜껑의 모양으로 제작하여 회사 내에 설치하였다. 또 다른 하나는 75주기인 2012년 10월 16일, 아일랜드 더블린 블랙록 카운티 홀리 파크의 세인트 패트릭 보이즈 내셔널 초등학교에 고셋이 기네스 브루어리에 근무하던 1913년부터 1935년 사이에 홀리빌 공원에 살았다는 사실을 알리는 명판으로 아일랜드 통계협회와 지역 당국인 던라오헤어 래스다운 카운티 의회가 합작하여 설립한 것이다.



기네스북과 기네스 맥주

두 번째 이야기는 기네스북에 대한 것으로 1952년 이 기네스 양조장 경영 이사인 휴 비비 경은 아일랜드 위스포드에서 사냥파티를 하는 중 골든 플로비가 세상에서 가장 빠른 새 인지에 대한 논쟁에 휘말린다. 그러나 그는 주위 도서관의 어떠한 참고서에서도 그 답을 찾을 수가 없었다고 한다. 또한 영국의 선술집인 펍에서는 이런 종류의 논쟁은 대다수 사람들 사이에 자주 일어나며 이에 대한 해답을 줄 만한 책의 필요성을 느낀 것이다.

요즘 세상에는 이런 경우 구글이나 우리의 경우 네이버로부터 답을 손쉽게 빠르게 해결한다. 그러나 그 이전에는 사람들 사이의 많은 논쟁의 끝은 늘 내기로 이어졌고 이러한 풍경은 우리에겐 그리 오래되지 않았다.

마침 기네스 파크 로얄 양조장 부양조자로 기록 경신의 달인인 치리스 체터웨이는 휴경의 아이디어를 듣고 기록 책을 출판하기에 적합한 사람들을 추천하였다. 그들은 그가 스포츠 행사를 통해 만난 노리스와 로스 맥휘터(McWhirter) 쌍둥이 형제였다. 특히, 옥스포드에서 단거리경주 대표선수이면서 진상확인업무 대리점을 경영한 그들이 조사하고 수집한 자료에 의해서 1955년 8월 27일 198쪽에 달하는 기네스북이 탄생했다. 어찌보면 기네스북에는 자연과 우주, 건축물, 인간의 업적, 예술과 오락, 스포츠, 정치와 사회, 비즈니스, 인간세계 및 기타 각 분야 걸친 세계 기록을 기록하고 있으니 이 또한 기술(記述)통계의 한 영역이라 여겨진다.

예전 17세기에 영국에서는 존 그란트(1620-1674)가 런던 시의 사망표를 작성하여 출생하는 「남녀 수는 남자 14에 대하여 여자 13의 비율로서 거의 같으며, 남녀 양성은 수적으로 거의 균형을 이루고 있다는 사실과 사망률은 일정한 비례나 비율로

죽는 것이 아니다. 100명의 출생자 중 6세까지 36명이 사망하며, 이후 10년이 지날 때마다 24, 15, 9, 6, 4, 3, 2, 1로 감소한다. 이 사망률로부터 역으로 생존자를 계산하면, 6세까지 64명, 16세에는 40명, 이후 10년마다 25, 16, 10, 6, 3, 1명이 되어, 86세까지 생존자는 없다」라는 사실을 보고서에 정리하였다. 이를 통해 그는 인구통계의 선구자로 정치 산술(political arithmetic)의 창시가 되었다. 그에게 영향을 받아 비슷한 조사보고서를 펴낸 여러 사람들이 등장했으며 혜성의 이름을 제공한 헬리(Edmond Halley. 1656-1742)도 그들 중 한 사람이었다. 이런 조상들의 통계에 대한 철학이 영국에서 기네스북 탄생의 원천이라고 여겨진다. 난 개인적으로 1996년 발간한 기네스북 40년사를 소장하고 있다.

이제 기네스 맥주와 얽힌 통계와 기네스북 이야기를 마무리한다. 내가 거처하는 부산의 경우 기네스 생맥주를 즐기는 곳이 내 경우 세 군데 정도 있다. 피시엔칩스를 같이 즐길 수 있는 아일리쉬 펍, 먹태를 안주로 하는 동네 맥주집, 기본 안주로 먹어야 하는 호텔인데 조만간에 기네스 맥주와 관련된 이야기를 떠올리며 이것의 깊은 맛을 음미해 보고 싶다.

(참고자료) •기네스북(1996). 96' The Guinness Book of Records, 동아출판사.
•최용석 History of Statistics: https://yschoi.pusan.ac.kr/yschoi/28169/subview.do
•Joan Fisher Box, J.F.(1987). Guinness, Gosset, Fisher, and Small Samples, Statistical Science, 2,(1), 45-52.
•Pearson, E. S. and Kendall, S. M.(1978). Studies in the History of Statistics and probability, Vol. I and II, Charles Griffin and Co Ltd.
•Stigler, S. M.(1986). The History of Statistics, Havard University Press.
•Student(1908). The Probable Error of a Mean, Biometrika, 6(1), 1-25.
•Plaque image: https://openplaques.org/plaques/41123
•Plaque image: http://blogs.roosevelt.edu/sziliak/w-s-gosset-aka-student

4세대 CCTV 시대로의 전환

김건우 | 한국전자통신연구원 인공지능융합보안연구실 책임연구원

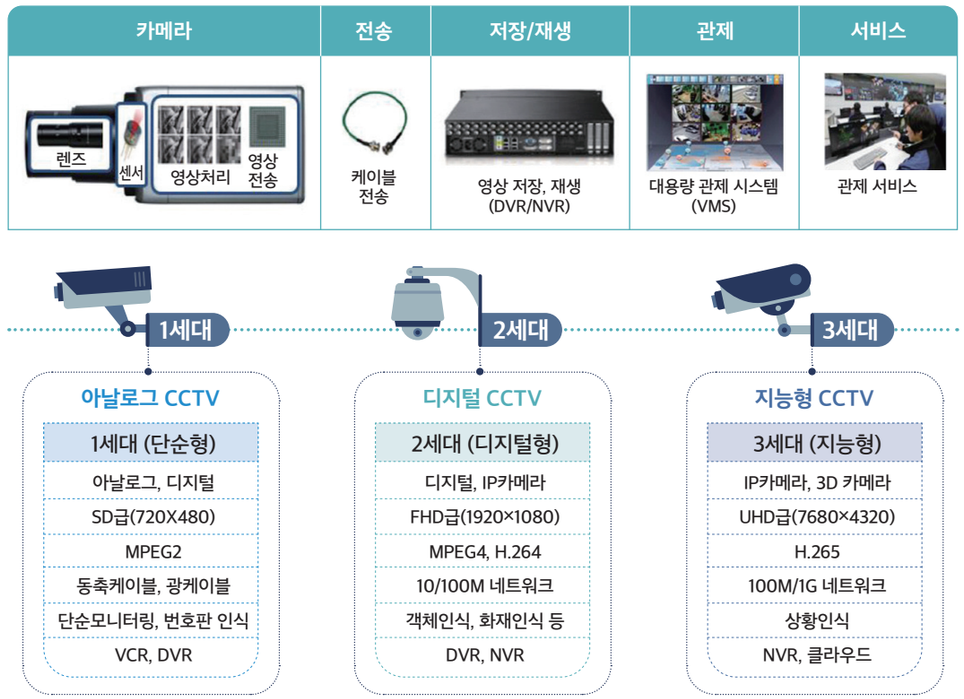
고도화된 현대 사회에서 점차 은밀, 흉폭해지는 범죄, 사고 등 각종 안전현안을 해결하는 치안의 최전방에서 CCTV의 역할과 중요성이 증가하고 있음은 누구나 인지하고 있는 사실이다. 범죄 현장에 출동한 경찰이 주위에 설치된 CCTV 영상을 우선 확보하는 것만 보더라도 영상정보는 현장의 상황을 가장 정확하고 충실하게 표현하는 매개체로, 범죄 현장이 CCTV 사각지대이거나 유의미한 영상 확보에 실패해서 초동 수사에 난항을 겪고 있다는 언론 기사도 자주 접할 수 있다. 이렇듯 CCTV는 이미 우리 사회 곳곳의 안전을 책임지는 안전 파수꾼으로서, 통계청에 따르면 2022년을 기준으로 전국에 설치된 총 CCTV 대수는 1,960만대로, 공공 기관에서 160만대 이상을 운용하고 있으며, 매년 설치 대수가 급격히 증가하고 있는 점을 감안하면 현재는 2,000만대를 훌쩍 넘을 것으로 추측된다.

CCTV의 세대별 진화

CCTV는 현재 3세대까지 발전된 것으로 평가받고 있는데, 초기 1세대 아날로그 CCTV는 외부와 물리적으로 연결이 차단된 폐쇄형 네트워크 환경에서 아날로그 형태의 SD급 영상 데이터가 수집, 저장 및 단순 관제되었다. 2세대 디지털 CCTV는 좀 더 정확하고 효율적인 데이터 관리를 위해서 영상 데이터를 디지털 포맷으로 전환하여 저장, 관리하였으며, IP 주소를 통해서 인터넷과 연결됨으로써 공유, 협업 등 가용성·효율성 측면에서 괄목할만한 성장이 이루어졌다.

현재 진행중인 3세대 지능형 CCTV는 CCTV, 관제/분석 서버 등에 인공지능을 탑재함으로써 기존 단순 모니터링, 검색 위주의 제한적 안전서비스에서 벗어나 시스템 스스로 관심 객체와 상황 등을 인식,

[그림 1] CCTV의 세대별 발전 과정



추적, 검색함으로써 보다 효율적인 관제와 즉각적인 대응이 가능해졌다. 지능형 CCTV 초기에는 제약적 환경에서 객체 검출, 식별, 추적과 침입 등 단순 위험상황 인식이 주였다면 최근에는 다중 CCTV 연계형 추적, 폭행, 군중 이상상황 감지 등 비제약적 환경에서 행위 인식, 협업, 은닉 정보의 생성/추론, 고속화 등에 관한 연구가 활발히 진행되고 있다.
그러면 과연 제 4세대 CCTV는 어떠한 패러다임으로 발전할 것인가?

누구나 접근 가능한 개방형 CCTV

온-디바이스 AI는 현재 글로벌 IT 트렌드를 주도하는 키워드로, 외부 네트워크 연결없이 스마트 디바이스가 보유한 지능을 통해서 정보를 수집, 처리, 분석, 추론하는 업무를 자율적으로 수행하는 것을 의미한다. CCTV에도 이러한 온-디바이스 AI 개념인 엣지 AI CCTV 기술에 대해서 많은 연구가 오래 전부터 수행되었고 객체 탐지 등 일부 AI 기능이 탑재된 상용 CCTV 제품이 이미 널리 설치, 운용되고 있다.
엣지 AI CCTV 기술이 서버에 집중된 부하 분산을 통한 실시간 고속 처리, 프라이버시 보호, 현장 맞춤형 AI 서비스 등의 장점이 있는 반면, 저사양 CCTV HW의 연산 능력 한계로 인한 고성능 AI 기술 탑재 불가, 주기적인 AI 모델 관리, 학습 및 업데이트 어려움 등 현실적인 단점으로 인하여 서버기반 AI CCTV 기술에 비해서 기대만큼의 성장 속도를 보여주지 못하는 것도 엄연한 현실이다. 나날이 새로운 AI 모델이 공개되고 괄목할만한 성능 향상이 이루어지고 있지만, 그 결과물이 엣지 AI CCTV에 제대로 반영되지 못하는 주요 원인 중의 하나는 CCTV 디바이스 플랫폼 자체의 '폐쇄성' 때문일 것이다.

대부분의 상용 CCTV 디바이스는 리눅스 운영체제를 기반으로 모든 기능은 CCTV 제조 시에 제조사에 의해서 개발, 탑재, 판매하고 있어, 사용자나 서비스 제공자가 새로운 기능을 추가하기 위해서는 제조사의 협조를 받아서 개발하거나, 해당 기능이 탑재된 AI CCTV를 신규 구매하는 수밖에 없다. 즉, 현재의 폐쇄형 CCTV 플랫폼상에서는 CCTV 제조사의 협조 없이는 자율적으로 CCTV에 새로운 기능을 탑재하거나 운용, 시험할 수 없다. CCTV 내부 시스템에 대한 외부 접근이 엄격히 제한되기 때문에 보안적인 측면에서는 장점이 있을 수 있으나, 사용자의 가용성 측면에서는 CCTV 제조사가 제공하는 기능에만 국한되고 AI 기술력에 종속되는 뚜렷한 한계가 있다. 즉, 사용자는 CCTV 디바이스 제조 시 탑재된 AI 기능만 사용할 수 있으며, 성능마저도 CCTV 제조사의 기술력에 의존하게 되는데, 영세한 대부분 국내 CCTV 제조 업체의 현실을 감안하면 온-디바이스 AI라는 글로벌 트렌드를 선도적으로 추구하기에 많은 어려움이 있다.

개방형 CCTV의 필요성

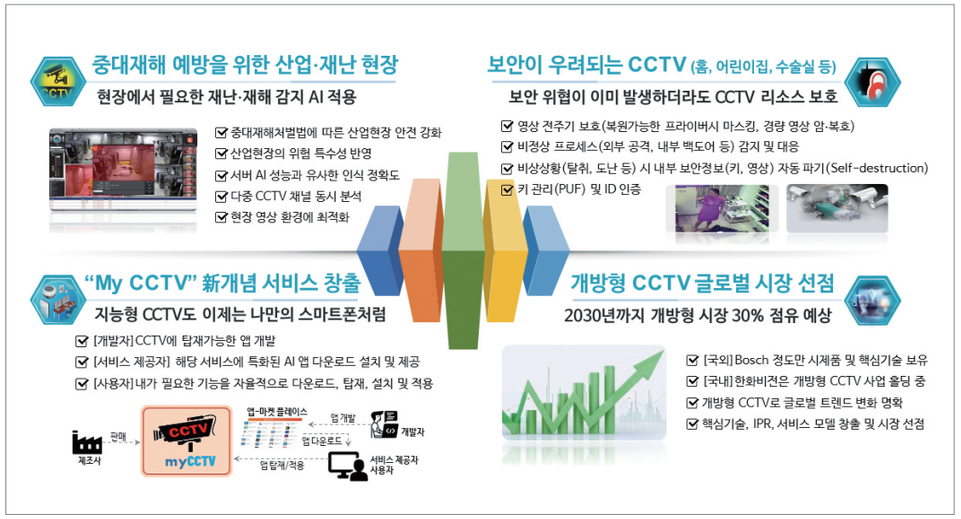
개방형 영상보안플랫폼을 위한 OSSA 사실 표준

따라서, Bosch, 한화비전, Milestone, Pelco, VivoTek 등 글로벌 영상업체들이 주도하여 IoT 디바이스 자체적으로 안전/보안 기능을 탑재, 운용하고 다른 디바이스와 협업할 수 있는 통합 표준 프레임워크인 OSSA(Open Security & Safety Alliance, 현재는 Onvif 표준 단체와 통합) 사실 표준을 제정하였다.

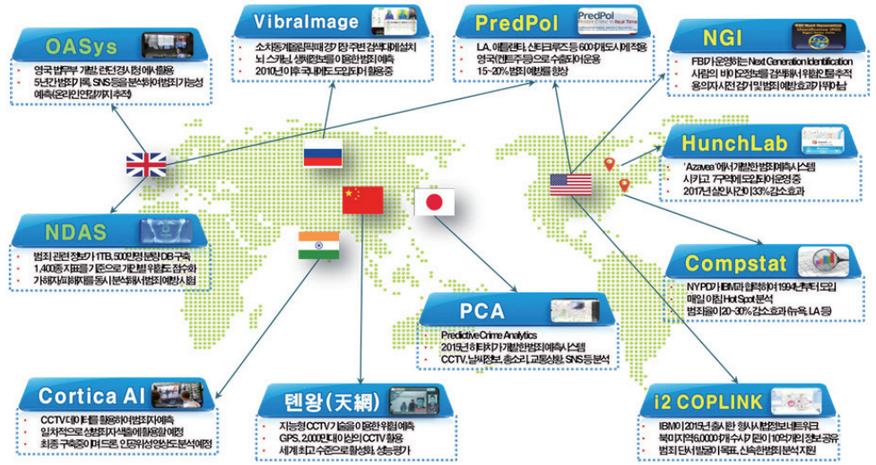
OSSA 표준은 CCTV 등 IoT 디바이스에 안드로이드 운영체제를 탑재하고, 다양한 AI/보안 기능의 자유로운 개발, 탑재, 연동 등이 가능한 표준 프레임워크를 정의함으로써 '개방형 CCTV' 新 생태계를 조성한다.

OSSA 표준 참여 업체

즉, 개방형 CCTV 생태계에서 CCTV 사용자는 자신이 원하는 고성능 AI/보안 기능을 앱-스토어 등 클라우드에서 구매, 자신의 CCTV에 설치해서 자유롭게 운용할 수 있고, 개발자는 다양한 AI/보안 기술을 개방형 CCTV 표준을 준용하는 앱 형태로 개발, 등록, 판매한다. 또한, 서비스 제공자는 CCTV가 설치되는 도메인에 최적화된 AI CCTV 솔루션을 제공하며 지속적이고 용이한 AI 성능 유지와 관리가 보장된다.
즉, 누구나 사용하는 스마트폰과 같이 CCTV에도 나만의 기능을 자율적으로 탑재, 운용, 관리할 수 있는 'My CCTV' 서비스 시대를 기대해도 좋을 듯 하다. 전 세계의 많은 AI 개발자들이 개방형 CCTV에



개방형 CCTV 활용 방안



글로벌 범죄예측 서비스 현황

탑재될 수 있는 초고성능의 모델을 지속적으로 등록하고 사용자/서비스 제공자는 자신의 요구사항에 적합한 AI 모델을 선택함으로써, 서버 기반 AI CCTV 성능에 버금가는 현장 맞춤형 서비스를 제공받을 수 있다. 신뢰할 수 있는 My CCTV 서비스 확산을 위해서는 CCTV 스스로 영상과 시스템을 외부 불법 접근, 악성 코드 등으로부터 보호하고 사용자의 사생활을 보장할 수 있는 내재화된 자율 보호(Security by Design), 능동적 방어 체계의 구축 또한 필수 기능이다.

위험징후를 선제적으로 감지·예측하는 CCTV

지금까지는 AI CCTV 기술이 주로 현재 발생하고 있는 실시간 위험상황(침입, 쓰러짐, 폭행, 화재 등)을 감지하고 메타데이터를 생성·검색하는 등 사후 해결에 초점을 맞추고 있다면, 앞으로는 가까운 미래의 범죄를 사전에 예측, 예방하는 예측 치안(Predictive Policing) 기술에 대한 연구가 더욱 활발히 진행될 것으로 보인다. 이미 미국, 영국, 일본

등 선진국에서는 잠재적인 범죄 활동을 식별하기 위해 수학적, 예측적 분석 기법을 법 집행에 활용하고 있으며, 분석 대상 또한 범죄통계정보에서 SNS 정보, 날씨·교통 등 도시환경 정보, 영상 정보, 총 소리 등 소스의 스펙트럼이 지역·사회적 특성에 맞게 다양화되고 있는 추세이다.

범죄 예측을 위한 기본 통계 모델

범죄 징후를 사전에 감지, 예측하는 방식은 인근 반복 모델(Near-Repeat Model)과 위험 영역 모델(Risk Terrain Model)을 기반으로 한다.

인근 반복 모델은 한 지점에서 범죄가 발생하면 일정 기간 내에 인근 지역에서 동일한 유형의 범죄가 반복적으로 발생할 가능성이 크다는 것으로, 예를 들면 주거침입절도의 경우 한번 침입을 당한 주거지가 재차 피해를 당할 가능성이 4~12배 높고 주거침입절도범 중 76%가 한번 범행한 집을 2~5번까지 재차 침입한다는 통계적 결과가 있다. 즉, 이 모델을 통해서 핫-스팟(Hot-Spot), 핫-타임(Hot-Time)을 설정하고 범죄가 시·공간적으로 집중되는 현상을 실시간 파악하여 사전 대응할 수 있다. 누가 범죄를 저지르는 것을 아는 것보다 언제 어디에서 범죄가 발생하는 지를 아는 것이 중요하다는 인식이 도출된다.

위험 영역 모델은 공간이 내포한 범죄 위험 요인을 통해서 범죄 위험도를 설명하는 것으로, 예를 들면 주거침입절도의 경우, 피해 여부, 전과자의 주거지, 주요 도로 인접 여부, 16~24세 남성의 공간적 집중도, 아파트와 숙박업소의 위치 등이 요인으로 작용할 수 있다. 또한, 성범죄는 유동인구가 많은 유흥업

범죄 예측을 위한 기본 모델

소 밀집 지역, 방화와 절도는 주거지역 등에서 많이 발생하며, GIS 분석 기법이 활용되기도 한다.

전 세계적으로 이미 다양한 범죄예측시스템이 운용되고 있는데, 미국은 각 도시별로 PredPol, CompStat, HunchLab, 영국 경시청에서 사용하는 OASys, NDAS, 일본은 히타치가 개발한 PCA, 그 외 중국, 인도 등에서도 CCTV 영상, 위성 영상과 결합한 범죄 징후 감지 기술이 연구, 서비스로 운용되고 있다.

과거와 현재를 아우르는 다차원 범죄 예측 기술

기존 범죄예측 기술은 과거 발생한 범죄통계정보에 기반해서 가까운 미래의 범죄 유형별 발생 가능성을 확률적으로 측정하는 방식으로 비용 대비 높은 치안 효과를 거두고는 있지만, 현재 발생하는 상황이 전혀 반영되지 못하는 단점이 있다. 미래 상황은 현재 발생하는 실시간 상황의 연장선에 있으며,

다차원 범죄예측 기술

현재는 미래에 전개될 상황과 가장 밀접한 관련이 있기 때문이다. 즉, 현재 발생하는 상황과 과거 동일 지역/시간대에 발생했던 범죄 패턴과의 유사도를 분석하여 범죄 징후를 감지하는 다차원 범죄 예측 기술이 개발되었다.

다차원 범죄예측 기술은 현재 발생하는 상황을 정확하게 인식하기 위한 CCTV 영상 데이터와 과거 범죄통계정보가 융합되어 범죄 발생 가능성이라는 확률적 추론 결과를 도출하는 방식으로 기존 방식보다 정확한 예측이 가능할 뿐 아니라 현장 상황을 자동 식별, 추적할 수 있어 즉각적인 범죄징후 상황 파악과 대응이 가능하다는 장점이 있다.

범죄 예측서비스의 실효성은 이미 미국 등 선진국의 사례에서 검증되어 70% 수준의 범죄 예측 성능, 총기, 살인 등 강력범죄의 20~30% 감소 효과가 있다는 보고가 있다. 단순 범죄 통계정보 기반의 단순 범죄 예측만으로도 이러한 강력범죄 예방효과가 있다면, AI CCTV와 융합된 다차원 범죄 예측서비스는 현장의 구체적인 상황을 고려하는 맞춤형 범죄 예방 효과를 거둘 수 있을 것으로 기대된다.

AI CCTV가 중심이 되는 미래형 첨단치안시스템

국내 CCTV 산업은 과거 글로벌 DVR 시장을 반짝 선점했을 뿐, 최근에는 미국과 중국에 뒤처지는 AI 기술력, 중국의 대규모 물량 중심의 저가공세, AI 반도체 핵심부품 경쟁력 저하 등 전반적으로 영세한

산업 규모를 벗어나지 못하고 있다. 다행히 공공분야를 필두로 Re-id, 군중 이상상황 인식, 접근 제어 등 실증서비스 적용이 추진 중이며, 민간분야에도 다양한 AI CCTV 기술이 확산되고 있는 추세이다. 더 나아가 글로벌 CCTV 시장 경쟁력 확보를 위해서 원천기술력 확보는 물론 AI CCTV 산업의 패러다임을 정확하게 예측하여 선제·집중 공략하는 Breakthrough R&D 전략이 필요하다.

필자가 주장하는 글로벌 CCTV 시장을 선점할 수 있는 키워드는 <개방>과 <예측>이다. 누구나 최고 수준의 AI CCTV 기술을 사용할 수 있고 능동적으로 참여할 수 있는 My CCTV 생태계, 개인의 사생활이 보호되고 신뢰할 수 있는 CCTV 서비스 체계, 시민들이 안전을 체감할 수 있는 예측치안 서비스를 통해서 미래형 첨단치안시스템이 실현될 수 있으며 그 중심에는 AI CCTV가 있다.

생성형 AI 시대 통계 데이터 융합
‘관리’와 ‘활용’으로 가치 확보해야

박재현 | IT DAILY 기자

들어가며

바야흐로 데이터의 시대다. 주위에 보이는 네트워크 통신이 가능한 전자기기는 모두 데이터를 발생·누적시킨다. 요즘에는 단순히 데이터 생산보다 유의미한 정보를 추출할 수 있도록 수집, 분석, 해석하는 과정을 거쳐 통계화 된다. 통계화된 데이터는 특정 집단의 정책 결정, 연구 개발, 비즈니스 전략 등 여러 분야에 적용되면서 의사결정의 신뢰성과 근거가 되곤 한다.

하지만 단순히 통계 데이터를 융합하는 것만큼 중요한 것이 있다. 바로 잘 융합된 통계 데이터들의 현행화와 변조에 대응할 수 있는 ‘관리’ 체계를 구비하는 것과 사용자들이 손쉽게 사용할 수 있도록 접근성을 확보한 ‘활용’이다. 생성형 AI 시대에 ‘데이터 거버넌스 기반 데이터 관리’와 ‘접근성을 갖춘 데이터 활용’ 등 2가지 측면에서

통계 데이터 융합의 가치를 더할 수 있는 방안을 모색해 본다.

생성형 AI와 통계 데이터 융합

1. 중요성

통계 데이터는 다양한 형태의 데이터를 수집, 분석, 해석해 유의미한 정보를 얻을 수 있는 데이터로 정의할 수 있다. 우리 주변에서도 쉽게 찾아볼 수 있다. TV 경제 채널에서, 혹은 정부 발표 자료에서, 글로벌 조사기관의 보고서들에서 잘 나타난다. 환율 통계 데이터는 한국은행이 발표하는 물가지수 데이터에서, 금리 통계 데이터는 금융당국이 발표하는 보고서 등 다양하게 발표되고 있다.

여러 통계 데이터를 활용해 원하는 분석 정보를 추출하고 얻은 정보는 정책 입안, 연구 개발 방향, 비즈니스 전략 수립 등 의사결정을 지원하는 도구로 사용될 수 있다. 통계 데이터는 기술 통계와 추론통계로 구분되며, 각각 방법론에 따라 데이터를 처리하고 활용하는 방식도 상이하다.

최근에는 생성형 인공지능(AI) 바람이 전 산업군을 휩쓸며 통계 데이터의 중요성이 부각되고 있다. 생성형 AI의 중심에 있는 AI 모델은 실제 지식이 아닌, 통계적으로 훈련된 데이터에서 패턴을 예측하는 방식으로 작동한다. AI가 학습할 수 있는 신뢰성 높은 데이터의 양이 무수히 많아야 생성형 AI의 완성도가 높다는 의미다.

그렇다면 가장 완성도 높은 데이터는 무엇일까. 데이터의 완성도를 평가하는 여러 잣대가 있지만, 단순하게 생각하면 데이터에서 유의미한 정보들을 추출할 수 있도록 여러 차례 정제된 과정을 거친 통계 데이터일 것이다.

또한 이렇게 정제된 통계 데이터를 다양한 출처의 데이터와 결합해 포괄적이고 유의미한 정보를 도출하는 ‘융합’의 과정을 거칠 경우 데이터 분석 정확성을 향상할 수 있고, 새로운 인사이트를 발굴할 수 있으며, 다양한 분야로의 활용 가능성도 넓힐 수 있다.

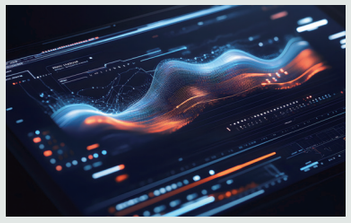
데이터 분석을 위한 4대 프로세스

한 글로벌 데이터 기업 관계자는 본 기자와의 미팅에서 “통계 데이터 융합은 정확한 예측과 분석 그리고 효율적인 자원 활용, 혁신 촉진 등 3가지 측면에서 강점이 있다. 다양한 출처의 데이터를 결합해 정확한 예측 모델을 구축할 수 있고 정책결정이나 비즈니스 전략 수립에 지대한 영향을 줄 것이다. 또 데이터 융합을 통해 중복된 정보를 제거하고 필요한 정보를 효율적으로 추출할 수 있다. 마지막으로 새로운 인사이트를 제공함으로써 연구 개발 및 산업 혁신을 촉진할 수도 있다”고 강조하기도 했다.

그렇다면 생성형 AI 시대에 통계 데이터 융합은 어떠한 가치가 있을까. 이 해답을 챗GPT에 물어봤다. 그 결과 '생성형 AI와 통계 데이터 융합은 생성형 AI 모델로 하여금 기존 데이터 분석뿐만 아니라 새로운 데이터를 생성할 수 있는 능력을 고도화할 수 있게 한다. 또 다양한 분야에서 통계 데이터를 효과적으로 활용할 수 있도록 할 수 있다'는 답을 내놨다. 쉽게 말해 생성형 AI는 통계 데이터를 보다 풍부하게 활용할 수 있는 도구로, 데이터 분석의 정확성과 효율성을 높일 수 있도록 기여할 수 있다는 것이다. 또 통계 데이터를 생산하는 여러 산업군의 데이터 기반 의사결정을 지원하고, 새로운 인사이트를 제공할 수 있다는 것이다. 일례로 금융 산업에서는 다양한 통계 데이터를 생성형 AI로 분석하고 추가 움직임을 예측함으로써 리스크 관리와 거래 전략 발굴에 도움을 얻을 수 있다.

2. 관리

통계 데이터를 비롯해 모든 데이터는 유의미한 데이터 분석(빅데이터)을 위한 재료로 사용된다. 정확하고 신뢰할 수 있는 결과를 얻기 위해서는 재료 관리가 중요하다. 이 과정을 IT 업계에서는 데이터를 사용하는 기관 및 기업, 조직이 정한 규정인 '데이터 거버넌스(Data Governance) 기반 데이터 관리'라는 단어로 사용한다.
정확한 의미는 데이터의 생성부터 사용에 이르기까지 모든 단계를 체계적으로 '관리'해 데이터의 품질과 보안을 보장하는 것이라는 의미다. 데이터에 대한 중요성이 확대될수록 체계적인 데이터 거버넌스 기반 데이터 관리 역시 부각될 것으로 예상된다.



일반적인 데이터 거버넌스 구현 절차는 크게 데이터 표준화, 메타데이터 관리, 데이터 품질관리, 데이터 계보관리, 데이터 카탈로그 관리 등의 순서로 진행되는 경우가 많다. 먼저 여러 곳에서 수집된 데이터의 형식이나 범위 등을 기준에 맞게 일치시키는 '데이터 표준화'를 수행해 데이터의 규격을 완성하고, 이렇게 완성된 데이터의 메타정보를 토대로 메타데이터를 관리한다. 이후 데이터에 대한 품질을 관리하면서 데이터 거버넌스 체계를 잡아가곤 한다.
글로벌 컨설팅 기업 맥킨지가 공개한 보고서에 따르면, 글로벌 상위 2,000개 기업 중 70% 이상이 최근 2년 사이 최신 기술을 활용한 새로운 데이터 아키텍처를 도입했거나 가까운 시일 내에 도입하려는 로드맵을 가지고 있는 것으로 나타났다. 하지만 이들 중 약 50%는 한 가지로 통합되지 않은 데이터 모델을 활용하고 있으며, 대다수는 자사의 데이터 중 25% 이하만을 단일한 데이터 플랫폼에 통합하고 있는 것으로 나타났다. 이는 기업이 보유한 데이터 중 75% 이상은 통합되지 않은 개별 데이터 저장소에 보관되고 있으며, 데이터 사일로(Silo)화가 상당한 수준으로 일어나고 있다는 것을 의미한다.

결국 아무리 데이터를 잘 정제하고 활용할 수 있는 환경이 마련됐더라도, 데이터 거버넌스 기반 '관리' 체계가 미흡하다면 데이터의 품질과 신뢰성은 하락하게 될 것이다. 통계 데이터 융합 역시 이와 다르지 않다. 잘 정제된 통계 데이터를 타 데이터와 융합해 유의미한 정보를 도출하기 위해선 융합 과정이 끝난 데이터를 특정 기업 및 기관이 정한 데이터 규정에 따라 관리해야 데이터의 신뢰성을 확보할 수 있다는 것이다.
글로벌 데이터 기업 관계자는 "통계 데이터를 타 데이터와 융합했다면, 이후 데이터 거버넌스 기반 관리 전략을 수립해야 한다. 일반적으로 데이터 통합 플랫폼을 사용해 분산된 데이터를 한 곳에서 관리한다. 이때 데이터 중복을 제거(클렌징)한 후 고품질의 데이터를 유지하는 작업이 필요하다"며 "특히 융합된 통계 데이터의 품질을 유지하고 개선하기 위해 통계 데이터에 대한 품질관리도 병행해야 한다"고 조언했다.

3. 접근성 갖춘 활용 체계

"궁극적으로 모든 데이터는 '활용'될 때 존재가치가 있다." 이 말은 데이터와 관련한 명언 중 데이터 본연의 가치를 가장 잘 꿰뚫는 문장이다. 데이터를 수집하고 저장하며 관리하는 일련의 모든 과정이 '활용'이라는 절차를 위해 존재한다는 의미다. 통계 데이터를 잘 활용하기 위해서는 어떠한 것이 가장 중요할까. 그 해답은 바로 '접근성을 갖춘 활용 체계를 구비하는 것'이다. 통계 데이터나 기타 데이터를 바라보는 마지막은 결국 사람이다. 사람들이 통계 데이터가 내포한 유의미한 정보에 얼마나 쉽고, 빠르게, 간편하게 접근하는지가 잘 활용하는지에 대한 여부를 결정짓는다.
생성형 AI 이전 통계 데이터는 유의미한 정보를 내포한 데이터로, 최종적으로 사람들에게 직관적이고 직접적으

로 표나 그래프, 문장으로 제공됐다. 결국 통계 데이터를 보는 사람들이 직접 유의미한 정보를 추출해야 했다. 가령 증권의 경우 여러 지표들이 담긴 통계 데이터를 분석가들이 본인들이 보유한 지식과 노하우, 정보, 데이터를 결합해 유의미한 분석 결과를 보고서로 제공하곤 한다.

하지만 최근 종착지가 생성형 AI 인터페이스로 변화했다. 사람들은 생성형 AI에 다양한 통계 데이터를 요청하기도 하고 데이터 학습을 추가하거나 새로운 데이터와 융합해 유의미한 정보를 얻기도 한다. 생성형 AI로 인해 사용자가 통계 데이터로부터 유의미한 정보에 보다 쉽게 접근할 수 있게 된 것이다. 단순 표·그래프 형태의 통계 데이터가 담긴 보고서가 아닌 UI 인터페이스가 통계 데이터를 요약하고 분석하고 유의미한 정보를 추출해 주기까지 하는 형태로 진화한 것이다.

그렇다면 '접근성' 측면에서 사용자가 이용하는 생성형 AI가 통계 데이터에 더 빠르게 접근하게 만들기 위해서는 기술적으로 필요한 것은 무엇일까. 챗GPT에 물음을 던지면, △데이터 아키텍처 및 플랫폼 강화 △클라우드 기반 솔루션 활용 △자동화된 데이터 관리 프로세스 △자연어 기반 데이터 탐색 △보안 및 개인정보보호 강화 등 5가지 해답을 내놓는다. 물론 상당히 좋은 방법이다. 여러 답안들을 한 문장으로 정의하면 생성형 AI가 사용자 질의에 대한 해답이 어느 데이터에 있는지 빠르게 찾아갈 수 있는 이정표를 만드는 것이다. 위 5가지 해답 모두 이 답에 귀결된다.

최근 생성형 AI가 데이터를 찾는 길을 쉽게 표기하는 기술인 '데이터 카탈로그'도 부상하고 있다. 데이터 카탈로그는 데이터 자산에 대한 수집-검색-공유-활용을 위한 데이터 접근성과 가시성을 제공하는 백과사전으로 정의할 수 있다. 특정 데이터가 어디에 있는지 어느 데이터베이스(DB)에 있는지 표기된 일종의 안내서인 것이다.

쉽게 말해 사용자가 원하는 데이터를 생성형 AI가 쉽게 찾을 수 있도록 비즈니스 용어, 데이터 소유자, 민감 정보 여부, 데이터 탐색 및 필터링, 접근 권한 관리, 마스킹/암호화 등 데이터 접근에 대한 정보부터 업데이트 주기 및 시점, 업데이트 건수 및 이상치 등 데이터 품질, 레코드 크기와 키(Key) 정의, 레코드 간 관계 등 데이터 구조 정의, 데이터 흐름, 변화 등 데이터 가시성에 대한 정보까지 다양한 기술 요소들이 담겨있다. 이러한 데이터 카탈로그는 검색증강생성(RAG)을 위해서도 핵심적인 역할을 수행하기도 한다.

만일 기업이 생성형 AI를 설치하고 직원들이 데이터를 잘 활용하게 만들고 싶다면, '접근성'에 초점을 맞춘 활용 체계를 잘 구현해야 할 것이다.

마치며

거시적인 관점에서 통계 데이터를 타 데이터와 융합해 유의미한 정보를 추출하는 '통계 데이터 융합'의 가치 있게 더하는 방법을 현직 IT 전문지 기자의 시선에서 바라봤다. '데이터 거버넌스 기반 데이터 관리'를 통해 일관성 있고 신뢰할 수 있도록 융합된 통계 데이터 관리 체계를 마련하고, 생성형 AI가 융합된 통계 데이터를 더 빠르고, 쉽게 접근할 수 있도록 '접근성 갖춘 활용 체계'를 구비한다는 점을 강조했다.

최근 데이터 드리븐 기업 및 조직(Data Driven Enterprise and Group)이라는 단어가 IT 업계 및 국내 기업들 사이에서 들불처럼 번지고 있다. 아마 한 기업을 운영하는 대표, 그룹을 이끄는 리더들은 한 번쯤 들어봤을 것이다. 단어의 뜻은 데이터를 기반으로 의사결정을 내리는 기업 및 조직을 의미한다. 이를 위해 데이터 수집과 분석도 중요하지만, 구성원이 데이터를 이해하고 활용할 수 있는 역량인 데이터 리터러시(Data Literacy)를 갖추고 적시에 데이터에 접근할 수 있는 데이터 관리 체계와 AI 기술을 활용해 데이터를 분석하고 유의미한 정보 추출을 자동화하는 것이 요구된다.

통계 데이터를 융·결합해 비즈니스를 창출하기 위해서는 데이터 사전 준비, 데이터 결합키 생성 및 정보전달, 데이터 결합, 추가 처리 및 반출심사 요청 등 개인정보보호를 비롯해 법적 규제, 기술적 절차 등 준수해야 할 요소들도 많고 기술적인 방안이나 통계 데이터 융합 사례, 쉽고 재밌는 요소들을 글로 전개하는 것도 의미가 있을 것이다.

다소 딱딱하고 기술적일 수 있지만, 필자는 데이터는 '활용될 때' 본연의 가치가 있다는 점을 알리고 싶다. 이를 위해 더 잘 활용할 수 있는 방안, 그리고 활용하기 전과 후 융합 통계 데이터를 관리하는 방안 등을 서술하는 것이 융합된 통계 데이터의 가치를 돋보이게 할 수 있다고 생각한다.

만인이 데이터 생산자 시대로의
진화와 데이터 문해력

이은경 | 전북대학교 과학학과 교수

"나에게 아주 긴 지렛대와 이를 지탱할 수 있는 받침대만 있다면, 나는 지구도 들어 올리겠다"
지렛대의 원리를 발견한 고대 그리스의 수학자이자 물리학자, 아르키메데스의 말이라고 전해진다. 현상 밑에 있는 법칙을 알면 겉으로 불가능해 보이는 문제도 원칙적으로 해결할 수 있다는 메시지를 전달할 때 자주 인용되는 말이다. 한 데이터 과학자는 이 말의 2022년 버전으로 다음과 같이 말했다.
"회사의 데이터를 달라, 그럼 문제를 풀어주겠다."

시대별로 중요한 문해력은 다르다

사회 구성원으로서 삶을 살아갈 때 꼭 필요한 최소한의 정보를 얻고 이용할 수 있는 기본 능력 중 하나가 해당 분야의 문해력(literacy) 또는 소양이다. 만일 데이터를 주면 어떤 문제든지 풀 수 있는 사회라면 데이터 문해력 또는 데이터 소양은 사회생활을 위한 기본 능력 중 하나가 되어야 한다. 데이터 과학 전문가가 될 사람, 일상 업무와 생활에서 데이터를 다루고 데이터 관련 정보와 해석을 이용할 사람, 본인이 데이터를 활용한 업무를 직접 하지는 않지만 각종 미디어에서 데이터 관련 정보를 접할 사람까지 필요에 맞게 단계적인 교육이 필요하다. 데이터 문해력의 범위를 어디까지로 할 것인지는 분명하지 않지만 쉽게는 문자 문해력과 비슷한 정도로 이해할 수 있을 것이다.
문자 문해력은 문자 해독 능력 즉, 읽고 쓰는 능력과 나아가 글을 읽고 그 뜻을 추론할 수 있는 독해 능력까지 포

함한다. 데이터 문해력은 데이터의 특성에 대한 이해, 데이터를 분석하고 그 결과를 읽는 능력, 자신이 생산하는 데이터를 관리할 수 있는 능력에서 시작하여 데이터를 기본으로 문제를 해결하는 기술적인 활용능력과 전문 영역으로 이어진다.
사회의 주요 정보가 어떤 형태로 제공되는가에 따라 기초 소양으로서 필요한 문해력의 종류가 달라진다. 문자 문해력이 교육의 기본이 된 것은 인쇄술이 발전하고 도시화되면서 많은 필수 정보가 문자로 제공되었기 때문이었다. 그에따라 공적인 보편 교육 제도가 도입되기 전에 이미 교회에서 운영하는 주일학교 등을 통해 평범한 가정의 아이들에게 읽고 쓰기를 가르치기 시작했다. 산업혁명기 이후에는 사회를 끌고 가는 새로운 힘으로서, 생산현장에서의 문제를 해결하는 강력한 수단으로서 과학이 중요성을 가지게 되었다. 과학 문해력이 중요해진 것이다. 근대국가에서 도입한 보편 초중등 교육에 과학 교과목이 포함되기 시작했고, 과학 문해력이 강조되었다. 아동과 성인을 위한 대중용 과학서적이 출간되었고 과학 강연의 인기가 높았다. 영국 화학자 험프리 데이비의 과학 강연에는 런던의 상류층 부인들이 참여했고, 마이클 패러데이의 크리스마스 과학 강연은 아동, 청소년들에게 특히 큰 인기를 끌었다.

20세기를 통해 과학 문해력을 강조하는 움직임은 계속 되었다. 특히 20세기에는 과학기술이 국가 경쟁력의 핵심 요소 중 하나가 되었기 때문에 세계 각국의 정부는 정규교육을 보완하는 각종 대중 과학 프로그램을 학교 안팎에서 실시했다. 같은 배경에서 과학 문해력 증진을 위한 프로그램의 접근성을 높이기 위해 과학관 등의 전문 기관을 확대하고 대중매체를 적극 활용했다. 뿐만 아니라 과학과 예술, 과학과 문화를 접목하는 융합적인 프로그램을 개발하여 과학의 여러 특성과 잠재성을 다양하게 제공하려는 노력을 기울였다.

데이터의 진화와 문해력

정보화 사회의 성숙, 디지털 기술의 발전과 함께 우리는 데이터가 폭증하는 시대를 살고 있다. 자연스럽게 과학 문해력 다음으로 데이터 문해력이 중요하게 되었다. 데이터 문해력을 구성하는 내용은 데이터의 내용, 형식, 데이터를 다루는 기법과 기술의 발전, 데이터 활용 목적의 변화 등에 따라 변화해왔다. 데이터 문해력 하면 제일 먼저 떠올리는 것은 많은 숫자로 된 데이터를 통계처리한 결과를 이해하는 능력일 것이다. 데이터는 사실, 개념, 사건 등을 나타내는 정보의 단위이고, 오랫동안 숫자로 표시되었다. 가장 원시적인 형태의 데이터는 단순 기록, 또는 자료에 가까웠다. 부유층이나 권력층의 재산, 세금, 군수 물자 기록 또는 교구의 인구기록 등이었다.

근대과학은 정보의 단위로서의 개념을 가진 데이터를 수집, 생산하고 활용하는 방법을 개발함으로써 객관적 기초로서 데이터의 가치를 확립했다. 천문학자들은 천체 현상을 정밀하게 관측한 데이터를 축적하고 이를 과학 원리에 맞추어 분석함으로써 천체의 운동 법칙을 밝히려고 노력했다. 코페르니쿠스가 제안한 태양중심설은 당시의 종교적 우주관은 물론 일상의 경험과도 잘 맞지 않았지만, 과학자들이 정밀한 천체관측 데이터를 신뢰한다면 받아들일 수 밖에 없는 결론이었다.
이후 물리학, 화학 등에서 실험을 방법론으로 도입하면서 일반 관찰이 아니라 정교하게 설계된 실험으로부터 데이터를 생산하게 되었다. 실험 데이터는 새로운 법칙을 발견하는 양적 기초가 되거나 과학 가설을 검증하는 양적

근거가 되었다. 일부 현상에서 과학자들이 통계 기법을 수용하자 이전에 양적으로 다루기 어렵던 영역에서도 연구 성과를 낼 수 있었다. 예를 들어 고전역학의 방법으로는 3개 이상의 물체들 사이에 힘이 작용할 때의 운동방정식을 수학적으로 풀기 매우 어렵다. 대안으로서 통계학 발전 성과를 받아들인 과학자들은 불규칙해 보이는, 엄청난 수의 공기 분자들의 운동을 다루는 통계역학의 법칙을 찾아낼 수 있었다. 또다른 예로서 멘델은 완두콩의 유전 형질을 구분하고 여러 대를 거쳐 재배한 결과를 통계적으로 다루어 당시까지 질적으로만 이해하던 유전 현상을 양적으로 규명한 멘델의 유전법칙을 발견했다.

사회과학은 자연과학의 방법론, 즉 양적 데이터에 기반하여 복잡하고 주관적으로 보이는 사회 현상의 기저에 깔린 법칙과 구조를 밝혀낼 수 있게 되었다. 경제학자들은 축적된 정보기록을 데이터로 간주하고 이를 통계적으로 분석하여 경제법칙을 확립했다. 그 선구자 중 한 명인 아담 스미스가 '근대경제학의 아버지'로 불리는 이유다. 데이터의 통계 분석은 이후 개인의 행위, 사회 관계, 나아가 인간 심리에 대한 연구로 확대되었다. 사회과학자들은 수집된 데이터 분석에서 더 나아가 마치 과학자들이 실험을 통해 관심있는 현상에 대한 데이터를 생산하듯 설문조사, 사회실험을 통해 데이터를 생산하기 시작했다.
특히 설문조사 방법론이 정립된 후에는 국가, 사회, 지역, 조직, 특정 인구집단 등으로 세분화하여 사회적, 인식적 특성을 파악할 수 있게 되었다. 1,000명, 또는 10,000명 등 전체 인구 대비 극히 작은 수의 사람들에게 설문조사를 한 결과를 통해 전체 여론을 파악하고 그 결과를 신뢰하게 된 것은 다 양적 방법론에 대한 신뢰에서 비롯되었다.

데이터를 통해 사회현상을 이해하기 위해서 사람들은 숫자를 사회현상, 가치평가, 관계의 특성 등으로 전환하여 인식할 수 있는 능력을 가져야 한다. 예를 들어 어떤 정당에 대한 지지도가 30%라는 설문결과 또는 한 사회의 평균 수명이 80세라는 인구통계 분석을 접했을 때, 숫자 30과 80은 사람들의 머릿속에서 정치적 입장에 대한 호불호, 노령인구의 건강과 보건복지 정책 등 추상적인 내용으로 번역되어야 한다. 숫자가 같아도 번역결과의 함의는 사람마다 다를 수 있다. 과학자, 사회과학자, 통계 분석가들은 분석 결과의 의미를 더 쉽고 분명하게 제시하기 위해, 표나 분석 결과를 시각화한 각종 그래프와 다이어그램을 도입했다. 시각화된 자료는 데이터 소비자의 이해를 돕는 기능을 하면서 동시에 그래프 축과 간격, 비율 등을 통해 해당 결과 생산자의 해석과 의도를 전달하기도 한다. 그래서 의도하지 않았다 하더라도 단순화에 따른 데이터 누락, 과장, 오해의 소지가 있다는 점을 데이터 생산자와 소비자 모두 이해할 필요가 있다.

이와 같은 데이터의 종류, 수집 또는 생산방식과 그에 따른 데이터의 특성, 분석 결과의 해석 또는 의미 파악 과정에 대한 이해가 데이터 문해력의 기본이 되어야 한다. 그 기초 위에서 데이터를 분석하는 기술과 기법을 익히고 문제를 푸는 연습이 이루어져야 한다. 그런데 현실에서는 '문제풀이'에 더 집중하는 것 같다. 우리는 초중등 교육과정에서 기초 교육으로서 수학을 12년간 공부한다. '수포자'라는 말이 따로 있을 정도로 수학을 어려워하고 대학 진학 또는 사회 진출 이후 직무 관련성이 없으면 수학에서 배운 것을 실제 접하거나 써먹을 기회가 많지 않다. 가장 두드러진 예외가 확률과 통계다. 데이터의 시대에 살고 있기 때문이다.

학교 교육에서는 입시 때문에 확률과 통계 과목 시간에 '계산'하는 데 집중하는 경향이 있고, 오히려 위에서 말한 데이터와 통계의 특성, 기능, 그리고 결과를 나타내는 방식에서 나타날 수 있는 여러 해석 및 오해의 가능성에 대해서는 거의 다루지 못한다. 사실 비전공자에게 필요한 데이터 문해력은 수학능력시험의 수리영역이 아니라 언어, 사회 영역의 문항을 해결하거나 언론 기사를 읽고 이해하는 데 필요한 능력에 더 가깝다.

데이터 생산자의 문해력

데이터 문해력과 관련하여 새로운 이슈는 데이터 생산과 관리 문제다. 이전의 데이터는 인구센서스, 경제통계, 또는 설문조사처럼 특정한 목적과 의도에 따라 관련자들에 의해 수집되거나 생산되었다.
반면 이제는 대중 일반이 일상에서 매일 생산하는 수많은 정보들을 의미있는 데이터로 활용할 수 있는 기술과 분석기법이 개발되었다. SNS에 개인이 올린 사진이나 글, 각자의 필요에 따라서 이루어지는 검색의 키워드, 각종 기관이나 웹사이트에 가입할 때 입력한 신상정보 등은 원래 특정 목적을 위한 데이터로 생산되지는 않았다. 그러나 정보가 디지털화되어 쉽게 이전이 가능해졌고, 데이터마이닝 등의 기법 덕분에 이러한 '의도없이 생산된 정보'가 데이터로서 활용가능해졌다. 단순히 활용가능해진 정도가 아니라 용도가 무궁무진한 강력한 데이터가 되었다. 그 결과 중 하나는 일반 대중이 일상에서 데이터를 생산하는 역할을 하게 되었다는 점이다.

개인들이 특별한 의도없이 매일 생산한 정보들을 종합하여 데이터로서 활용하여 분석하면 해당 개인에 대해 많은 것을 파악할 수 있다. 나아가 이러한 정보들을 종합하면 개인을 넘어 집단, 사회에 대해 많은 것을 파악할 수 있다. 이 과정을 이해하고 자신이 생산하는 정보를 데이터로서 관리해야 한다는 인식과 이를 위한 행위는 21세기 빅데이터 시대의 데이터 문해력의 새로운 요소가 되었다. 우리의 데이터 문해력 교육에는 이러한 내용이 충분히 반영되어야 한다. 코딩 교육만으로 충족되기 어려운 문제다.

통계로 바라보는 세상이야기

신동헌 | 도서출판 지일북스 대표

오늘 당신의 하루는 어떠셨나요?

바쁜 현대 사회에서 시간을 절약하려는 노력은 다양한 방식으로 나타나고 있습니다. '분초사회'(Time-Efficient Society)는 시간 효율성을 극도로 높이려는 트렌드 속에서 모든 사람들이 분초를 다투며 살게 됐다는 의미를 담은 단어라고 합니다. 트렌드모니터에서 발표한 「시간 절약 서비스 관련 U&A 조사」에 따르면, 응답자의 82.4%가 '시간'을 현대 사회에서 가장 중요한 자원으로 꼽았습니다. 이를 통해 시간이 현대인들에게 얼마나 큰 의미를 지니는지 한번 더 확인할 수 있었는데요. 이러한 분초사회에서는 빠르게 합리적인 결정을 내리기 위해 자신의 가치관과 일치하는 인물이나 콘텐츠의 추천을 바탕으로 제품을 구매하는 소비 트렌드인 '디토소비'(Ditto Consumption)가 나타난다고 합니다.

우리 국민들이 느끼는 주관적 웰빙은?

한국행정연구원의 「2023년 사회통합실태조사」 결과에 따르면, 국민 삶의 만족도는 0~10점 척도 응답의 평균값으로, 주관적 웰빙의 인지적 측면을 측정하는 요소 중 하나인데요, 6.4점으로 전년대비 0.1점 소폭 감소하였으나, 평균 6점 이상으로 보통(5점)보다 높은 수준을 유지하였습니다. 연령대별로는 40-49세의 삶의 만족도가 6.6점으로 가장 높고, 60세 이상이 6.2점으로 가장 낮았으며, 성별로는 남녀 모두 6.4점으로 차이가 없었습니다. 행복을 나타내는 긍정정서는 주관적 웰빙의 정서적 측면을 측정하기 위한 지표인데요. 조사 결과에 따르면, 2023년 긍정정서는 6.7점으로 전년도 수준을 유지한 것으로 나타난 반면, 부정정서(걱정, 우울)는 2023년 3.1점으로 최근 3년 동안 지속적으로 하락하는 추세를 보였습니다.

'나 혼자 산다' 1인 가구 역대 최대

최근 통계청이 발표한 「2023년 인구주택총조사 결과(등록센서스 방식)」에 따르면, 2023년 1인 가구는 일반가구의 35.5%인 783만 가구로 전년대비 4.4%(33만 가구)증가하였습니다. 1인 가구의 증가율은 2020년 8.1%로 고점을 찍은 이후 감소 추세이지만, 1인 가구의 비중은 역대 최고 수준이었습니다. 연령대별 1인 가구 비율을 보면 20대 이하가 18.6%로 가장 높고 60대와 30대가 각각 17.3%로 그 뒤를 이었는데요. 성별로 보면 남자는 30대(21.8%)가, 여자는 60대의 비중(18.6%)이 가장 높게 나타났습니다. 고령인구는 950만 명으로 전년대비 5.0%(45만 명) 증가한 것으로 나타났는데, 연령별로 보면 고령인구 중 65~74세가 전체 고령인구의 58.2%로 가장 많았으며, 75~84세가 31.3%, 85세 이상이 10.5%를 차지했습니다.

갓생 : MZ세대의 새로운 라이프 스타일

2023년 한국방송광고진흥공사가 전국의 만 20~59세 남녀 2천 명을 대상으로 조사한 결과에 따르면 '갓생'이라는 신조어는 전체 조사대상자의 절반인 50.5%가 알고 있으며, '들어본 적은 있지만 무슨 뜻인지 모른다'는 사람이 33.6%, '들어본 적 없고 무슨 뜻인지 잘 모른다'는 사람은 16.0%로 나타났습니다. '갓생'을 살기 위해 중요하다고 생각하는 요소로는 재테크 및 업무 관련 공부 등 '자기개발'이 41.6%로 가장 높았으며, 저축이나 투자 등의 '재테크'가 37.3%, '주기적인 운동'이 37.2%로 그 뒤를 이었습니다. 갓생을 살기 위해 금액을 지출하거나 투자할 의향이 있냐는 질문에는 전체 응답자 중 72.9%가 '그렇다'라고 답했고, 투자 의향자의 51.0%가 10만 원 이상을 지출할 의사가 있다고 응답했습니다.

2023 인구주택총조사 '인구 부문'

통계청은 지난 7월 29일「2023년 인구주택총조사 결과」를 발표하였습니다. 인구총조사는 1925년에, 주택총조사는 1960년에 처음 실시되었는데요. 2010년까지는 5년 주기로 현장조사 방식의 인구주택총조사를 실시하였으나, 2015년 기준부터 행정자료 기반의 등록센서스 방식을 도입했습니다. 조사한 결과에 따르면, 우리나라의 총인구는 5,177만 명으로 전년 대비 8만 명(0.2%) 증가한 것으로 나타났는데요. 이 중 내국인은 4,984만 명(96.3%)으로 전년대비 10만 명 감소하였으며, 외국인은 194만 명(3.7%)으로 전년대비 18만 명 증가하였습니다. 내국인은 2021년 이후 감소세가 계속되고 있지만, 외국인은 코로나19로 2020년부터 2년 연속 감소하다 2022년 이후 증가세로 바뀌었다는 점을 알 수 있었습니다.

2023 인구주택총조사 '주택 부문'

통계청이 발표한 「2023 인구주택총조사 결과」에 따르면 2023년 11월 1일 기준 우리나라의 총주택 수는 1,955만 호로 전년대비 2.0%(39만 호) 증가하였습니다. 이 중 공동주택은 1,547만 호로 전체 주택의 79.2%를 차지하였으며, 단독주택은 386만 호(19.8%), 비주거용 건물내주택은 21만 호(1.1%)였습니다. 우리나라 주택 중 가장 많은 비중을 차지하는 주택은 바로 '아파트'였는데요. 총 주택의 64.6%인 1,263만 호였습니다. 다음으로 많은 주택은 '단독주택'으로 총 주택의 19.8%인 386만 호였고, 다음으로 많은 주택은 연립/다세대주택(총주택의 14.5%, 284만 호)이었습니다. 특히 빈집도 많이 늘었는데요, 2023년 11월 1일 기준 전국에 미거주 주택(빈집)은 154만 호로 전년대비 5.7%(8만 호) 증가하였습니다.

커피, 하루에 몇 잔이 적당할까?

통계청 「서비스업조사」에 따르면, 2022년 기준 국내 커피 전문점 사업체 수는 10만 729개로 전년대비 4.5% 늘었고, 매출액은 14.7% 증가한 15조 5천억 원으로 나타났습니다. 식품의약품안전처 「식품 등의 생산실적」에 따르면, 2022년 기준 국내 커피 시장 규모는 약 3조 1,717억 원으로, 액상 커피 판매 비중이 35.6%로 가장 높았으며, 볶은 커피(32.6%), 조제 커피(24.8%), 인스턴트 커피(7.0%) 순으로 나타났습니다. 현재 식약처에서는 카페인 1일 섭취 권고량을 성인 400mg, 임산부 300mg, 어린이와 청소년은 체중 1kg당 2.5mg 이하로 정해두고 있는데요. 시중에 판매하는 기본 사이즈 아메리카노 한 잔의 카페인 함량이 150mg 내외라는 것을 생각하면 2~3잔 정도가 적당하다고 합니다.

국내 미술시장의 새로운 활력소는?

젊은 세대를 중심으로 한 온라인 중심의 생활과 투자에 대한 관심이 증가하면서 아트슈머(art+consumer)라는 새로운 신조어가 등장했습니다. 국립현대미술관에 따르면, 2023년 상반기 방문객 가운데 20·30세대의 비중이 63%에 달했다고 합니다. 특히 젊은 수집가들의 등장은 국내 미술시장에 활기를 불어넣고 아트테크의 확산에도 많은 영향을 미치고 있습니다. 예술경영지원센터의 「한국 미술시장 결산 및 전망」에 따르면, 미술시장 거래액은 2021년 7,563억원으로 전년 3,849억원 대비 96.5% 증가하면서 상승세를 보이다가 2022년 8,066억으로 정점을 찍었습니다. 2023년 미술시장 거래액은 6,695억 원(전년대비 17%)으로 다소 감소하였는데요. 글로벌 경기 침체의 영향으로 보입니다.

불청객 태풍에 대처하는 우리의 자세

2002년 루사, 2003년 매미, 2010년 곤파스, 2022년 힌남노 등 역대 우리나라에 가공할 위력을 남긴 태풍은 모두 늦여름에서 초가을 사이에 찾아와 엄청난 인명과 재산피해를 남겼는데요. 행정안전부가 발표한 「2022년 재해연보」에 따르면 2022년 자연재난 피해액은 5,927억 원이었고, 호우 피해 3,326억 원(56.1%), 태풍 2,440억 원(41.2%) 순으로 나타났습니다. 2022년에는 자연재해 피해액이 전년보다 약 9배 증가했을 정도로 피해가 컸고, 그중 3건은 특별재난지역으로 선포됐는데, 9월 6일 남부를 관통해 총 12명의 사망·실종자를 냈던 '힌남노'도 있었습니다. 기상청 「태풍발생통계」를 보면, 지난 30년간 북서태평양에서 태풍은 연평균 25.1개가 발생하고 이 중 약 3.4개가 우리나라에 영향을 주었습니다.

은퇴자 70%가 계속 일하려는 이유는?

통계청 「2024년 5월 경제활동인구조사 고령층 부가조사 결과」에 따르면, 고령층의 경제활동참가율은 60.6%로, 전년동월대비 0.4%p 올라 역대 최고치를 기록하였으며, 고령층 취업자 역시 943만 6천 명으로 전년동월대비 31만 6천 명 증가하였습니다. 고령층 취업자 중 보건·사회·복지업의 취업자(121만 명) 비중이 12.8%로 가장 높았으며, 농림어업(116만 6천 명, 12.4%), 제조업(114만 9천 명, 12.2%) 등이 그 뒤를 이었습니다. 반면에 예술·스포츠·여가(10만 8천 명, 1.1%)와 금융·보험업(17만 9천 명, 1.9%)은 상대적으로 고령층 취업자 비중이 낮았습니다. 고령층들이 장래에 일하기를 희망하는 이유는 '생활비에 보탬'(55.0%)이 절반 이상을 차지했고, 일하는 즐거움(35.8%), 무료해서(4.2%) 등이 뒤를 이었습니다.

2024년 청년 취업자 줄고 실업자 늘어

통계청 「2024년 5월 경제활동인구조사 청년층 부가조사 결과」에 따르면, 15세에서 29세까지의 청년층 인구는 전년보다 24만 3천 명 감소한 817만 3천 명이며, 이 중 경제활동참가율은 50.3%로 전년동월대비 0.2%p 하락했고, 청년층 취업자는 383만 2천 명으로 전년동월대비 17만 3천 명 줄었습니다. 청년들이 졸업 후 첫 일자리에 취업하는 데에는 전년동월대비 1.1개월 증가한 평균 11.5개월이 걸렸고, 첫 직장을 구하는데 3년 이상 걸린 경우도 9.7%에 달했습니다. 졸업 후 첫 일자리가 현재 직장인 경우는 34.3%로 나타났는데, 첫 일자리를 그만둔 이유는 '근로여건 불만족'이 45.5%로 가장 많았으며, '임시적, 계절적 일의 완료·계약기간 끝남'이 15.6%, '개인·가족적 이유'가 15.3%로 그 뒤를 이었습니다.

2023년 온라인 쇼핑 총거래액 229조 원

통계청이 발표한 「온라인 쇼핑 동향」 결과에 따르면, 2023년 온라인 쇼핑 총거래액은 228조 8,607억 원으로 전년 대비 8.4% 증가했습니다. 상품군별 온라인 쇼핑 거래액 구성비를 살펴보면 음·식료품이 13.1%로 가장 큰 비중을 차지했고, 음식 서비스(11.5%), 여행 및 교통서비스(10.5%) 등이 그 뒤를 이었습니다. 음·식료품과 생활용품의 거래액이 증가한 것은 바로 쿠*이나 마켓** 같은 유통업체들의 성장이 있었기 때문입니다. 이는 모바일 쇼핑에도 비슷한데요, 2023년 온라인 쇼핑 거래액 중 모바일 쇼핑 거래액은 169조 320억 원으로 전년 대비 7.0% 증가한 것으로 나타났는데요. 상품군별 모바일 쇼핑 거래액 비중은 음식 서비스(15.4%), 음·식료품(13.2%), 여행 및 교통서비스(9.5%) 등의 순으로 높았습니다.

2023년 합계출산율 역대 최저치

통계청의 「2023년 출생 통계」에 따르면, 2023년 출생아 수는 23만 명으로, 전년보다 1만 9천 2백명 감소하였는데요. 30년 전인 1993년 출생아 수 72만 명과 비교하면 무려 67.9%나 감소하였습니다. 특히 여자 1명이 평생 낳을 것으로 예상되는 '합계출산율'은 0.72명으로 출생 통계 작성(1970년) 이래 최저치를 기록했는데, 1993년(1.65명)에 비해 0.93명이나 감소했습니다. 2023년 모(母)의 평균 출산연령은 33.6세로 전년대비 0.1세 상승한 것으로 35세 이상 고령 산모 비중이 증가했는데요, 첫째아는 33.0세, 둘째아는 34.4세, 셋째아는 35.6세, 넷째아는 36.5세였습니다. 첫째·둘째·셋째아 출산 모의 평균 연령 역시 전년대비 0.1~0.2세 상승했고, 35세 이상의 고령 산모 비중도 전년대비 0.6%p 증가한 36.3%였습니다.

내 손안의 작은 통계지도, SGIS!

SGIS는 Statistical Geographic Information Service의 약자인데요. 이는 통계정보와 지리정보를 융합한 통계지리정보서비스로 통계청에서는 2008년부터 이 서비스를 제공하고 있습니다. 통계주제도는 국민의 관심사와 관련된 6개 주제의 통계를 지도상에서 확인할 수 있는 서비스이며, 살고 싶은 우리 동네는 원하는 지역과 라이프스타일을 고려한 사용자 맞춤형 주거지역 추천 서비스이고, 생활권역 통계지도는 사용자가 선택한 특정 관심시설을 기준으로 일정시간 내 도달 가능한 생활권역의 통계정보를 제공하고 있습니다. 통계지리정보시스템의 개발지원센터에서도 지도 API, 데이터 API, 모바일 SDK* 등의 무료 Open API를 제공하여 다양한 서비스 활용 및 창조적이고 다양한 어플리케이션 개발을 지원하고 있습니다.

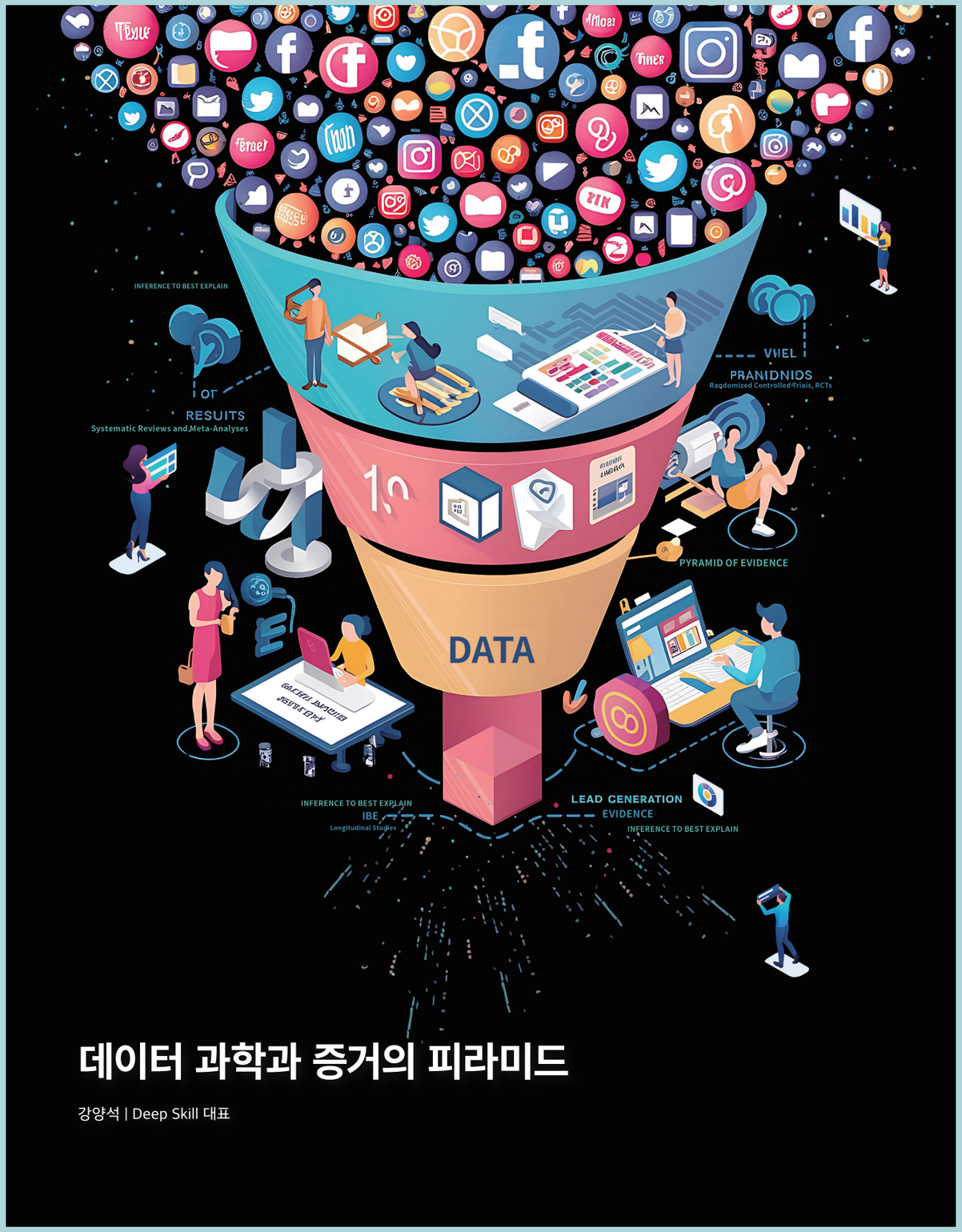
* Software Development Kit

국가통계포털 KOSIS 톺아보기!

국가통계포털은 KOrean Statistical Information Service의 약자로 국가승인통계를 제공하고 있습니다. KOSIS에서는 2023년 12월말 기준, 400여 개 통계작성기관에서 생산하는 1,400여 종의 국가승인통계를 수록하고 있습니다. 국내통계에서는 인구·경제·사회·환경 등 30개 분야에 걸쳐 주요 국내 통계자료를 제공하는데, 대표적으로 경제총조사, 인구주택총조사, 소비자물가조사 등이 있습니다. 국제통계에서는 국제경제 및 사회의 흐름을 파악할 수 있는 주요 국제지표 및 통계자료가 국제지구별로 분류되어 있어 쉽게 찾아볼 수 있습니다. 끝으로 북한 통계에서는 국/내외에 있는 북한 관련 통계정보를 체계적으로 수집하여 서비스하는데, 분단 이후 출생인구, 경제성장률 등 다양한 주제의 통계들을 제공하고 있습니다.

"문화재, 국가유산으로 불러주세요"

2024년 5월 17일, '문화재'라는 명칭이 '국가유산'으로 바뀌었다는 사실 알고 계신가요? 1972년 제정된 유네스코(UNESCO)의 '세계 문화 및 자연유산 보호에 관한 협약'에 따라 이미 많은 국가에서 유산 개념을 사용하고 있는데, 우리나라도 문화유산 보존에 집중하는 과거 회귀형에서 과거와 현재, 미래를 아우르는 '국가유산'으로 변경되었고, 분류체계 역시 유네스코 유산체계에 부합하는 문화유산, 자연유산, 무형유산으로 재정립되었습니다. 국가유산청이 발표한 2023년 국가유산관리 현황에 따르면 2023년 우리나라의 전체 국가유산 수는 총 15,281건으로 전년대비 1.3% 증가하였고, 이 중 국가지정·등록문화유산은 5,316건으로 전체 국가유산 중 34.8%, 시·도문화유산은 9,965건으로 65.2%를 차지하였습니다.



데이터 과학과 증거의 피라미드

강양석 | Deep Skill 대표

데이터 과학 진리 여정

과학, 철학, 종교의 공통점은 무엇일까? 이 세 영역은 인간이 추구하고 경험하는 영역 중 '진리(변하지 않는 교훈)'을 추구한다는 공통점을 가지고 있다. 이때 이 셋은 오묘하게 협업의 관계를 가지고 있는데 과학과 종교는 의심을 기반으로, 대신 종교는 믿음을 기반으로 한다. 또한, 철학과 종교는 형이상학을 주로, 대신 과학은 형이하학을 주로 다룬다. 오묘한 협업 관계이다. 또한 각각이 진리를 추구하는 방식이 다른데, 종교는 계시와 기도를 통해, 종교는 사유와 논리, 그리고 과학은 추론과 실험을 통해 도달한다.

여기서 주지할 사실은 과학이 추론을 통해 진리를 추구한다는 점인데, 추론은 사실을 기반으로 개연을 다루는 행위이다. 개연은 사실스러움을 말한다. 이런 과학의 특성에 대해 과학 철학자들은 '과학은 IBE(Inference to best Explain)'이라고 말한다. 즉, 현재의 사실로 그나마 가장 그럴싸게 설명한 중간 산출물이라는 뜻이다. 그런데 데이터 과학은 당연히 이런 과학의 특성을 그대로 이어 받았다고 볼 수 있다. 다만, 그 대상과 재료가 데이터를 기반으로 한다는 점만 다를 뿐이다. 즉, 데이터 과학은 데이터로 최선을 다해 설명한 잠정적 주장이다. 이런 데이터 과학의 특성을 이해할 때 우리가 주목해야 할 단어가 있는데, 그건 바로 증거(evidence)라는 개념이다. 데이터 과학은 증거를 만드는 행위라고도 바꿔 말할수 있는데, 증거라는 개념과 대비되어 이해하면 좋은 용어가 증명이다.

증명은 강도가 없지만, 증거는 강도가 있다. 증명은 하면 하고, 말면 마는 것이지 중간 단계가 없다. 하지만, 증거는 더 좋은 증거와 더 약한 증거가 존재한다. 그러므로, 우리가 데이터 과학에서 나온 모든 결론은 증거의 지위를 갖는다는 것을 이해할 필요가 있다.

그럼 우린 자연스럽게 어떤 데이터 과학적 증거가 더 높은 설명력을 가질까에 궁금증을 가질 필요가 있다. 데이터를 썼다고 해서 모두가 같은 증거능력을 갖는 게 아니기 때문이다. 이를 잘 표현한 구성이 바로 증거의 피라미드(pyramid of evidence)라는 개념이다. 피라미드의 하위층일수록 증거능력이 낮고 높아질수록 증거능력 또한 높아지는 구조이다. 우리가 이 피라미드를 이해하는 것이 중요한 이유는 우리는 결국 좋은 의사결정을 하기 위해서 데이터 과학을 하기 때문이다. 한데, 여러 종류의 데이터 과학적 결론 사이에서 우열을 가르는 식견이 없다면 우리가 그 수많은 분석 기법을 배워야 할 이유가 무색해지기 때문이다. 자 그럼, 데이터가 진리가 되기 필요한 필수 개념, 상

관, 인과관계, 증거의 강도 등에 대해 알아보자.

인과추론의 중요성과 증거의 피라미드

오늘날 데이터 기반 사회에서, 우리는 데이터를 분석하고 그 결과에 따라 다양한 의사 결정을 내리는 시대에 살고 있다. 그러나 단순히 상관관계를 파악하는 것만으로는 충분하지 않으며, 이를 넘어서서 인과적 사고를 바탕으로 데이터를 분석할 필요가 있다. 인과적 사고는 데이터에서 나타나는 현상의 원인과 결과를 명확하게 규명하는 사고방식이다. 이 글에서는 인과적 사고와 이를 효과적으로 이해하기 위한 증거의 피라미드 개념에 대해 다루고자 한다.

1. 인과적 사고란 무엇인가?

인과적 사고는 상관관계를 넘어서서 두 변수 사이의 원인과 결과를 이해하는 사고이다. 상관관계는 단순히 두 변수 사이에 통계적인 연관성을 보여주는 것이며, 이것이 반드시 인과 관계를 의미하지는 않는다. 예를 들어, 아이스크림 판매량과 익사 사고 사이에 상관관계가 있다고 해서 아이스크림이 익사 사고를 유발한다고 말할 수는 없다. 여름철에는 날씨가 더워져서 아이스크림 판매량이 늘고, 동시에 사람들이 물놀이를 많이 하기 때문에 익사 사고가 늘어나는 것이다. 이 경우 날씨라는 제3의 변수가 존재하는 것이며, 상관관계는 단순한 우연일 수 있다. 따라서 상관관계를 발견한 이후에, 그것이 실제로 인과 관계인지 아닌지를 파악하는 과정이 필수적이다.

이러한 과정을 통해서만 우리는 올바른 결론을 도출할 수 있으며, 데이터에 기반한 신뢰할 만한 의

사 결정을 내릴 수 있는 것이다. 인과적 사고를 통해 우리는 단순히 데이터의 패턴을 읽는 것이 아니라, 그 패턴의 배경에 숨겨진 메커니즘을 이해하게 되는 것이다.

2. 인과 추론의 필요성 : 산업 현장에서의 교훈

인과 추론이 얼마나 중요한지에 대한 실전적인 예시를 통해 그 필요성을 설명하고자 한다. 과거 글로벌 소프트웨어 회사에서 데이터 과학 팀을 설립하고 운영할 당시, 우리는 대규모 데이터를 분석하여 새로운 수익 모델을 찾고자 했다. 당시 목표는 무료로 제공하던 소프트웨어를 유료 서비스로 전환하여 수익을 창출하는 것이었다. 수백만 명의 사용자 데이터를 기반으로 사용자 행동을 분석하고, 이를 바탕으로 유료화를 위한 다양한 모델을 제시했다.

그러나 결과는 실패로 돌아갔다. 사용자는 유료화에 반응하지 않았고, 우리는 새로운 모델을 도입한 후에도 기대한 수익을 창출하지 못했다. 그 이유는 우리가 데이터를 분석할 때, 상관관계에만 집중했기 때문이다. 사용자가 특정 기능을 많이 사용한다고 해서 그것이 유료화에 적합하다고 판단한 것이 오류였다.

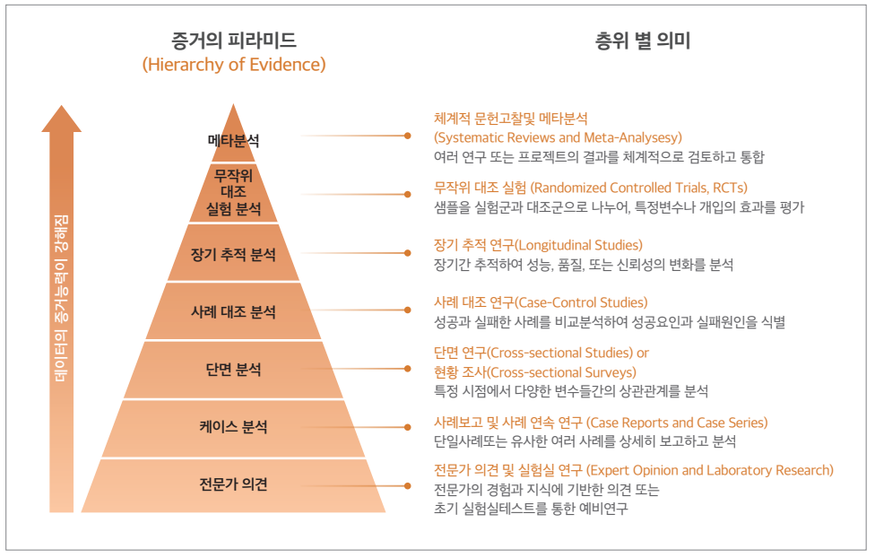
실제로 그 기능이 사용자의 요구를 충족시키는 데 얼마나 중요한지, 그리고 그 기능에 대한 사용자의 실제 필요가 무엇인지를 충분히 이해하지 못했던 것이다. 이 경험을 통해 깨달은 것은 데이터를 해석하는 과정에서 상관관계 이상의 인과적 사고가 필요하다는 것이다. 단순히 어떤 패턴을 발견하는 것만으로는 충분하지 않으며, 그 패턴이 발생한 원인을 찾아야 한다. 이것이 바로 인과 추론의 본질이며, 데이터를 제대로 이해하고 해석하는 데 필수적인 요소이다.

3. 증거의 피라미드 : 데이터 의사 결정의 구조적 접근

이제 인과 추론의 본질을 이해하는 데 도움이 되는 중요한 개념인 증거의 피라미드에 대해 설명하고자 한다. 증거의 피라미드는 데이터 분석에서 나타나는 증거의 신뢰도를 단계적으로 나눈 개념이다. 이는 주로 의료계에서 많이 사용되는 개념이지만, 데이터 과학이나 경영 의사 결정에서도 매우 유용하게 활용될 수 있다. 증거의 피라미드는 의사 결정자가 어떤 데이터를 기반으로 결정을 내릴 때, 그 데이터가 얼마나 신뢰할 수 있는지를 판단하는 기준을 제공해 준다.

3.1. 일반 피라미드 설명

증거의 피라미드란, 말 그대로 증거의 신뢰성과 강도를 피라미드 형태로 계층화한 개념이다. 피라미드의 가장 하단에는 신뢰도가 낮은 증거가 위치하고, 상단으로 갈수록 신뢰도가 높아진다. 이를 통해 데이터의 신뢰성을 평가하고, 분석 결과를 어떻게 해석할지 결정할 수 있다.



증거의 피라미드(출처:DeepSkill)

❶ 전문가의 의견 | 피라미드의 가장 하단에 위치한 것은 전문가의 의견이다. 이는 특정 데이터를 기반으로 한 분석보다는 직관이나 경험을 바탕으로 한 결론을 의미한다. 전문가의 의견은 경험에 기반하고 있기 때문에 유용할 수 있지만, 과학적 근거가 부족할 수 있다는 점에서 가장 낮은 신뢰도를 가진다. 예를 들어, 산업 현장에서의 경영진이나 현장 전문가가 특정한 경험에 의거해 결론을 내릴 수 있지만, 이는 객관적 데이터를 뒷받침하는 증거로서는 약할 수 있다.

❷ 케이스 분석 및 벤치마킹 | 그다음으로는 케이스 분석이다. 이는 특정 상황이나 사례를 분석하여 유사한 결론을 도출하는 방법이다. 케이스 분석은 일종의 '사례 연구'로 볼 수 있으며, 일반적으로는 주어진 상황에서 성공 사례를 분석하거나 이를 벤치마킹하여 적용한다. 하지만 개별 사례를 기반으로 하기 때문에, 이를 전체에 일반화하기는 어렵다.

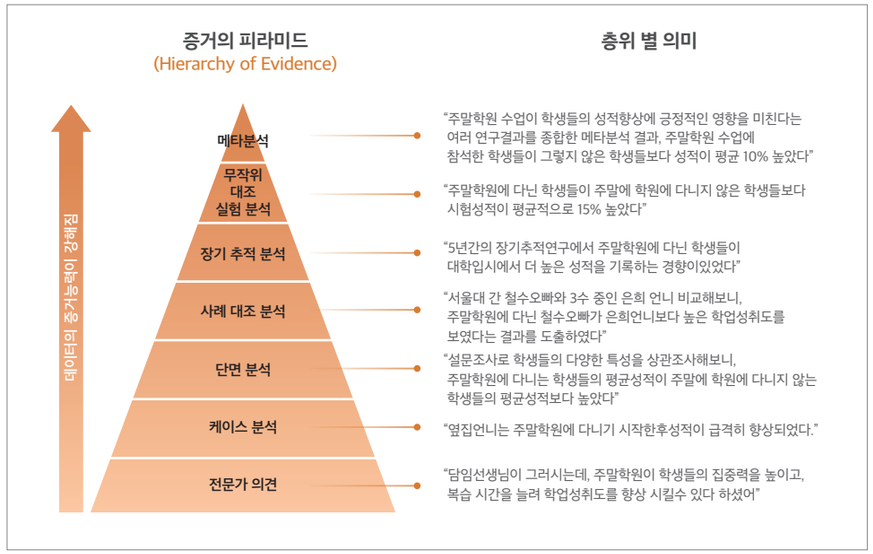
❸ 단면 분석 | 단면 분석은 특정 시점을 기준으로 다양한 변수 간의 상관 관계를 분석하는 방법이다. 단면 분석을 통해서는 여러 변수 사이의 관계를 파악할 수 있지만, 이것이 인과 관계를 명확히 규명하는 데는 한계가 있다. 따라서 이를 근거로 한 의사 결정은 인과적 관계가 아닌 상관 관계에 의존할 수 있다.

❹ 비교 연구 | 피라미드 상에서 비교 연구는 두 개 이상의 사례를 비교하여 원인과 결과를 분석하는 방식이다. 예를 들어, 성공한 기업과 실패한 기업의 경영 전략을 비교하거나, 두 집단 간의 데이터 분석을 통해 인과적 관계를 추론하는 것이다. 이 단계에서부터는 인과 관계를 밝히는 노력이 시작된다.

❺ RCT(무작위 대조 실험) | 증거의 피라미드에서 가장 신뢰할 수 있는 증거는 무작위 대조 실험(RCT)에서 나온다. RCT는 두 집단을 무작위로 나누어 한 집단에만 개입을 하고, 다른 집단에는 개입하지 않음으로써 그 결과를 비교하는 실험이다. 이를 통해 변수 간의 인과 관계를 명확하게 밝힐 수 있는 가장 강력한 방법이다.

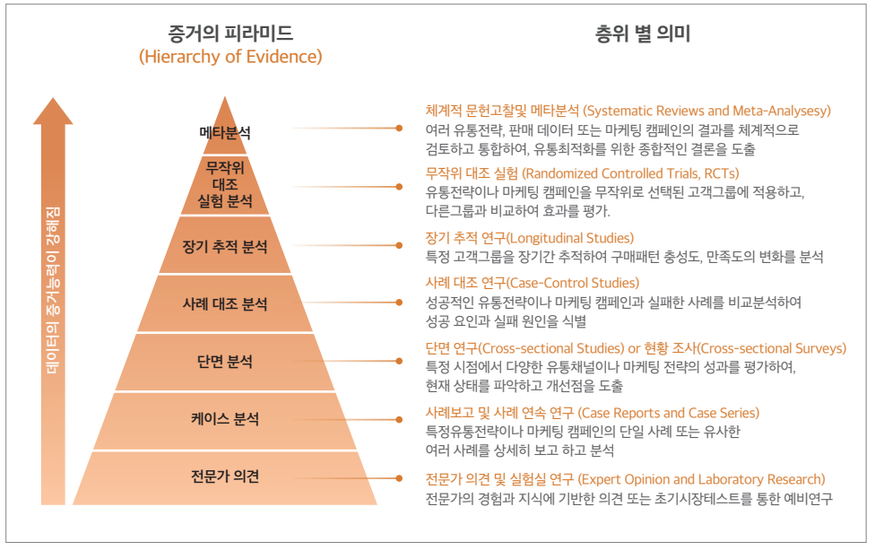
3.2. 산업별 피라미드 설명

증거의 피라미드는 의료뿐만 아니라 다양한 산업 분야에서도 활용될 수 있다. 각 산업별로 증거의 피라미드를 어떻게 적용할 수 있는지 살펴보자.



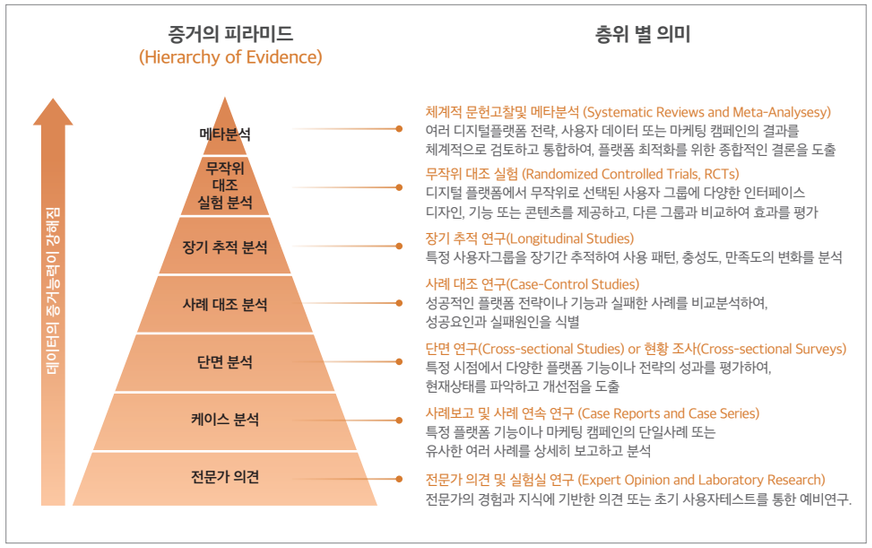
증거의 피라미드 - "딸아, 주말에도 학원에 가야지"(출처:DeepSkill)

❶ 일상적인 소재 | 우선 증거의 피라미드를 기준으로 딸아이에게 왜 주말에 학원에 가야하는지를 설명하면 위와 같다. 이런 일상적이고 직관적인 이해를 바탕으로 분석의 증거능력을 설명할 수 있어야 한다. 잘 살펴보면, 우리가 자녀들을 설득할 때 가장 많이 사용하는 '옆집 오빠 이론'은 전체 층위에서 매우 설득력이 낮다는 것을 알수 있다.



증거의 피라미드- 유통업 기준(출처:DeepSkill)

❷ 유통업 | 유통업에서는 고객의 구매 패턴, 매출 데이터 등을 분석하여 전략적 결정을 내린다. 이때, 매출 데이터는 단순히 상관관계를 보여줄 뿐이다. 예를 들어, 특정 상품의 매출이 증가했다고 해서 그것이 마케팅 캠페인 덕분이라고 단정할 수는 없다. 고객의 실제 구매 동기, 경쟁사의 가격 정책, 계절적 요인 등 다양한 변수를 고려해야 한다. 유통업에서 증거의 피라미드는 고객 설문조사(전문가 의견)에서 시작해, 비교 연구(경쟁사와의 비교) 또는 RCT(프로모션 효과 분석)까지 나아갈 수 있다.



증거의 피라미드- 인터넷 플랫폼 서비스업 기준(출처:DeepSkill)

❸ IT 산업 | IT 산업에서는 사용자의 행동 데이터를 기반으로 의사 결정을 내린다. 예를 들어, 소프트웨어 사용 빈도가 높은 기능이 있다고 해서 그것을 유료화하는 전략이 항상 성공적이지는 않다. 증거의 피라미드를 통해 사용자의 실제 요구와 행동 패턴을 심층적으로 분석해야 한다. IT 산업에서는 사용자 행동 분석(단면 분석)에서 시작해, 기능 개선 실험(RCT)까지 진행하는 것이 효과적이다.

4. 데이터 분석과 인과적 사고의 균형

데이터 분석에서 중요한 것은 단순히 많은 양의 데이터를 수집하는 것이 아니다. 오히려 중요한 것은 그 데이터를 어떻게 해석하고, 그 데이터를 통해 무엇을 배우는가이다. 많은 경우, 우리는 데이터를 해석할 때 상관관계에 지나치게 집중하는 경향이 있다. 하지만 상관관계는 그저 데이터 간의 패턴을 보여줄 뿐이며, 그것이 반드시 원인과 결과를 의미하지는 않는다. 이때 필요한 것이 바로 인과적 사고이다. 인과적 사고는 데이터를 보다 깊이 있게 이해하게 해주며, 이를 바탕으로 더 나은 결정을 내릴 수 있게 한다. 예를 들어, 마케팅 데이터를 분석할 때, 단순히 매출이 증가한 것만으로는 그 원인을 알 수 없다. 매출이 증가한 이유가 무엇인지, 그리고 그 이유가 지속 가능한 것인지 분석하기 위해서는 인과적 사고가 필요하다.

결론적으로, 인과 추론은 단순한 상관관계를 넘어선 데이터 해석의 핵심이다. 증거의 피라미드는 이를 명확하게 이해할 수 있는 유용한 틀을 제공하며, 데이터의 신뢰도를 계층화해 의사 결정의 정확성을 높인다. 각 산업에서 피라미드를 적용하면, 데이터 분석에서 단순한 패턴 인식을 넘어 심층적인 원인과 결과를 파악할 수 있다. 이는 궁극적으로 더 나은 전략적 결정을 가능하게 하며, 지속 가능한 성과를 창출하는 데 기여하는 중요한 도구이다. 즉, 지나치게 분석 기법 중심으로 데이터 과학을 이해하기보다는, 증거 능력의 격차 관점을 이해하고 그에 맞는 의사결정 강도의 수위를 정하는 식견이 우선이다.

스몰 데이터로 소비자의 트렌드 분석하기

구자룡 | 밸류바인 대표

소비자 트렌드에 대한 인식 차이

일반 소비자에게 트렌드는 주로 현재 유행하는 것, 많은 사람들이 관심을 가지고 있는 것을 의미한다. 예를 들어, 특정 스타일의 옷이 유행하거나 새로운 기술 제품이 인기를 끄는 현상을 트렌드라고 인식한다. 소비자들은 주로 자신의 일상생활에서 직접 체감할 수 있는 현상들을 트렌드로 받아들인다. SNS에서 화제가 되는 주제, TV 프로그램에서 자주 다루는 내용, 주변 사람들 사이에서 유행하는 제품 등이 여기에 해당한다.

반면, 마케터나 기획자가 생각하는 트렌드는 더 넓은 범위와 깊이를 가진다. 이들에게 트렌드는 단순히 현재의 유행을 넘어서 미래의 변화 방향을 예측할 수 있는 징후나 패턴을 의미한다. 마케터와 기획자들은 현재의 인기 현상뿐만 아니라 그 배경에 있는 사회, 경제, 문화적 요인들을 함께 고려한다. 또한 이러한 요인들이 어떻게 발전하고 변화할지를 예측하여 향후 시장의 움직임을 파악하고자 한다. 한마디로 트렌드는 '예측'인 것이다.

이와 같이 트렌드에 대한 인식의 차이는 관점과 목적의 차이에서 비롯된다. 소비자는 주로 현재의 현상과 개인적인 경험을 바탕으로 트렌드를 인식하는 반면, 마케터와 기획자는 더 넓은 맥락에서 현재와 미래를 연결 지어 트렌드를 분석한다. 이러한 차이를 이해하는 것은 비즈니스 전략 수립에 매우 중요하다. 소비자의 트렌드 인식을 이해하면서도, 더 깊고 넓은 관점에서 트렌드를 분석하고 예측할 때, 비즈니스는 현재의 성공과 미래의 성장을 동시에 추구할 수 있다.

스몰 데이터로 트렌드 분석을 할 수 있는 방법들

트렌드 분석은 다양한 방법을 통해 이루어진다. 각 방법은 고유의 장단점을 가지고 있으며, 대부분의 경우 여러 방법을 복합적으로 활용하여 보다 정확하고 포괄적인 분석 결과를 얻고자 한다. 트렌드 분석은 빅데이터 분석, 소비자 설문조사, 미디어 모니터링, 텍스트 마이닝, 웹 크롤링 등 빅데이터를 기반으로 하는 분석 방법과 심층 인터뷰, 참여 관찰, 포토 보이스, 네트노그래피, 민족지학적 시장 조사, 포토 콜라주 등의 스몰 데이터를 기반으로 하는 분석 방법이 있다.

빅데이터가 주목받는 시대지만, 스몰 데이터를 활용한 트렌드 분석 역시 여전히 중요하고 유효한 방법이다. 스몰 데이터는 규모는 작지만 깊이 있고 맥락이 풍부한 데이터를 의미한다. 이는 특히 질적 연구 방법과 밀접하게 연관되어 있으며, 소비자의 행동과 심리를 깊이 있게 이해하는 데 도움이 된다. 여기서는 스몰 데이터를 활용한 트렌드 분석 방법들에 대해 구체적으로 살펴본다(심층 인터뷰와 참여 관찰 방법 등은 2024년 여름호 참조).

1 포토 보이스

참가자들에게 특정 주제와 관련된 사진을 찍게 하고 그에 대해 이야기하게 하는 방법이다. 이를 통해 참가자들의 시각적 경험과 그 의미를 이해할 수 있다. 예를 들어, 10명의 젊은 소비자에게 '패션'이라는 주제로 일주일 동안 매일 사진을 찍게 하여 패션 트렌드를 분석할 수 있다.

2 네트노그래피(Netnography)

온라인 커뮤니티나 소셜 미디어에서 특정 그룹의 상호작용을 깊이 있게 관찰하고 분석하는 방법이다. 네트워크 기반 참여 관찰이라고 할 수 있다. 예를 들어, 특정 브랜드의 팬 커뮤니티를 3개월 동안 관찰하여 브랜드에 대한 소비자 인식 트렌드를 파악할 수 있다.

3 민족지학적 시장 조사

특정 집단의 문화와 생활 방식을 깊이 있게 관찰하고 이해하는 방법이다. 예를 들어, 연구자가 한 달 동안 특정 동네에 거주하며 그 지역의 소비문화를 관찰하여 지역 특화 트렌드를 발견할 수 있다.

4 포토 콜라주

비주얼 데이터를 수집, 분석, 그리고 해석해 트렌드를 파악하는 방법이다. 주로 패션, 디자인, 마케팅에서 많이 활용되며, 이미지를 통해 소비자 선호도와 트렌드를 직관적으로 파악할 수 있다.

이러한 스몰 데이터 분석 방법들은 대규모의 정량적 데이터로는 파악하기 어려운 미묘한 변화나 새로운 트렌드의 조짐을 발견하는 데 유용하다. 또한, 소비자의 행동 뒤에 숨어 있는 동기와 감정을 이해하는 데 도움이 된다. 특히 새로운 시장을 개척하거나 혁신적인 제품을 개발할 때 중요한 인사이트를 제공할 수 있다.

포토 콜라주를 이용한 트렌드 분석

포토 콜라주(Photo Collage, 서로 다른 이미지의 결합을 통해 새로운 의미를 도출)는 비주얼 데이터를 수집 및 분석하고, 트렌드 변화의 방향을 감지하는 데 유용한 방법이다. 트렌드 분석에 포토 콜라주를 활용하는 방법은 다음과 같다.

먼저, 분석하려는 트렌드의 목적을 명확히 정의해야 한다. 예를 들어, 패션 트렌드, 인테리어 디자인 트렌드, 라이프 트렌드, 혹은 마케팅 캠페인에 대한 고객 반응 등을 분석할 수 있다.

둘째, 트렌드를 반영하는 이미지를 다양한 출처에서 수집한다. 여기에는 소셜 미디어, 잡지, 광고, 웹사이트, 그리고 개인이 소유한 스마트폰 등이 포함될 수 있다. 수집할 이미지는 트렌드를 잘 대표하는 요소들, 예를 들어 색상, 패턴, 스타일, 제품 등을 포함해야 한다.

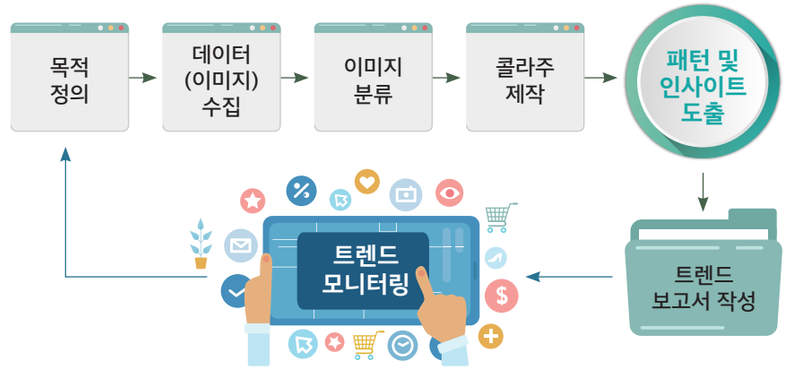
셋째, 수집한 이미지를 주제별로 분류한다. 예를 들어, 패션 트렌드를 분석한다면 색상, 소재, 스타일별로 이미지를 그룹화 할 수 있다. 마케팅 트렌드 분석에서는 브랜드별, 표적 소비자별로 이미지를 분류할 수 있다.

넷째, 주제별로 분류된 이미지를 콜라주로 만든다. 각 콜라주는 특정 트렌드를 시각적으로 표현할 수 있으며, 여러 이미지를 결합함으로써 트렌드의 다양한 측면을 한눈에 파악할 수 있다. 이 과정에서 캔바(Canva) 같은 디자인 도구를 사용하여 이미지를 배치하고, 트렌드별로 정리된 콜라주를 만들어 시각적으로 보기 좋게 구성할 수 있다.

다섯째, 콜라주를 통해 반복적으로 나타나는 패턴을 확인하고 인사이트를 도출한다. 예를 들어, 여러 이미지에서 동일한 색상, 소재, 혹은 디자인 요소가 자주 등장하는지 분석한다. 트렌드의 방향성을 이해하고, 이 정보가 향후 전략에 어떻게 적용될 수 있을지 판단한다.

여섯째, 포토 콜라주로 발견한 패턴을 바탕으로 트렌드 보고서를 작성한다. 주요 트렌드, 예측되는 변화, 그리고 이러한 트렌드가 시장이나 소비자 행동에 어떤 영향을 미칠지에 대한 분석이 포함될 수 있다.

마지막으로 트렌드 모니터링을 통해 지속적으로 이미지를 업데이트하고 콜라주를 수정하며 변화하는 트렌드를 추적한다. 현재의 트렌드를 보다 정확하게 예측하고 대응할 수 있다.



[그림1] 포토 콜라주를 이용한 트렌드 분석 프로세스

예를 들어, 스마트폰의 사진을 이용해 나만의 관점으로 트렌드 분석을 해보자. 스마트폰은 개인 소유물이라 나만 갖고 있는 데이터가 있는데, 바로 사진이다. 스마트폰의 사진(갤러리) 앱에는 직접 찍은 사진들이 있다. 이를 이용한 트렌드 분석을 하게 되면 그동안 인지하지 못한 트렌드 분석이 가능하다. 즉, 내가 트렌드 분석의 대상자임과 동시에 나만의 마이크로트렌드 분석을 할 수 있다.

제시한 포토 콜라주를 이용한 트렌드 분석 프로세스에 따라 목적을 정하고 이미지를 수집하면 된다. 이미지는 나의 스마트폰에 있다. 많은 사진들 중에서 유사한 이미지를 분류한다. 유사한 이미지를 중심으로 여러 장의 콜라주를 만들면 된다. 이때 반복적으로 나타나는 패턴을 찾고 그 패턴의 의

미를 도출하면 된다. 이렇게 찾은 트렌드로 '세컨드 하우스에서의 로컬라이프'가 있다고 하자. (세컨드하우스(Second House)는 주로 생활하는 집 이외에 보유한 다른 주택으로 여가나 휴식을 즐기기 위한 별장 또는 도시 거주자가 주말, 휴일에 쉬기 위해 근교나 지방에 마련한 '두 번째 집'을 의미)

그런데 이 트렌드는 내가 찾은 트렌드로 대단히 주관적인 결과물이다. 구체화 또는 일반화하기 위해서는 시장 조사가 필요할 수도 있다. 키워드 검색을 하고 관련 분야 전문가나 얼리어답터의 의견을 듣고 2차 자료를 확인해야 한다. 일종의 객관화 작업이다. 아직은 미세한 변화이기 때문에 수치화된 자료가 많지 않을 수도 있다. 만약 개인이 아니라 회사에서 트렌드 분석을 한다면, 주변의 동료들이나 고객들의 데이터도 수집하여 분석한 후 모든 분석 결과물을 통합적으로 분석해야 한다. 이러한 과정에 대해 보고서를 작성하여 공유하거나 의사결정자를 설득해야 한다.

[그림2] 세컨드하우스에서의 로컬라이프 포토 콜라주

이제 이 트렌드를 바탕으로 비즈니스에 어떻게 활용할지 방법을 찾아야 한다. 내가 찾은 트렌드가 메가트렌드가 될 것 같다면 사업화 단계로 넘어가야 한다. 사업화에 성공한 이후로도 새로운 것을 보게 되면 무조건 사진으로 찍어두는 것이 중요하다. 지금 당장 트렌드인지 아닌지 생각할 필요는 없다. 일정 시간이 지난 후 새로운 사업을 구상하거나 지금의 트렌드에서 새로운 가치를 만들고 싶을 때, 첫 번째 활동부터 다시 시작하면 된다. 이를 습관적으로 반복하면 된다. 트렌드 모니터링이 필요한 이유다.

스몰 데이터에 의한 성공적인 트렌드 분석 사례

스몰 데이터를 활용한 트렌드 분석은 여러 기업들에 의해 성공적으로 수행되어 왔다. 이러한 분석은 새로운 제품 개발, 마케팅 전략 수립, 고객 경험 개선 등 다양한 영역에서 유의미한 성과를 거두었다. 몇 가지 주목할 만한 사례들이다.

1 에어비앤비의 사용자 경험 연구

에어비앤비는 호스트와 게스트 양쪽의 경험을 깊이 있게 이해하기 위해 실제 사용자들의 집을 방문하여 그들의 경험을 관찰하고 인터뷰를 진행했다. 이를 통해 얻은 인사이트를 바탕으로 플랫폼의 사용자 경험을 지속적으로 개선했다.

2 자라의 점원 피드백 시스템

패스트 패션 브랜드 자라는 매장 점원들의 일일 피드백을 중요한 트렌드 분석 도구로 활용한다. 점원들은 매일 고객들의 반응, 요구사항, 선호도 등을 본사에 보고한다. 이러한 스몰 데이터를 바탕으로 자라는 빠르게 새로운 디자인을 출시하거나 기존 제품을 수정한다.

3 P&G의 현장 관찰 연구

P&G는 새로운 청소용품 개발을 위해 연구원들을 직접 소비자의 집에 파견하여 청소하는 모습을 관찰했다. 이 과정에서 많은 사람들이 빗자루로 쓰레기를 모은 후 손으로 직접 쓰레기를 집어 버린다는 사실을 발견했다. 이러한 인사이트를 바탕으로 P&G는 쓰레받기가 달린 빗자루인 '스위퍼(Swiffer)'를 개발했다.

스몰 데이터 분석은 단순히 표면적인 트렌드를 파악하는 것을 넘어, 소비자의 숨겨진 니즈와 행동 패턴을 발견하고 이를 제품 개발과 마케팅 전략에 반영할 수 있게 해준다. 그러나 이러한 성공 사례들은 단순히 스몰 데이터만을 활용한 결과가 아니라는 점을 유의해야 한다. 대부분의 기업들은 스몰 데이터 분석과 빅데이터 분석을 상호 보완적으로 활용하여 더욱 정확하고 유용한 인사이트를 도출하고 있다. 예를 들어, 스몰 데이터 분석을 통해 발견한 새로운 트렌드의 가능성을 빅데이터 분석을 통해 검증하거나, 빅데이터 분석 결과의 원인을 스몰 데이터 분석을 통해 심층적으로 탐구하는 방식이다.

소비자 트렌드 분석을 통한 새로운 가치 창출

트렌드 분석은 단순히 현재의 현상을 파악하는 것이 아니라, 그 배경에 있는 원인과 향후 발전 방향을 예측하는 것까지 포함한다. 따라서 다양한 방법을 통해 수집된 데이터와 정보를 종합적으로 해석하고, 이를 바탕으로 의미 있는 인사이트를 도출하는 능력이 중요하다.

또한, 트렌드 분석은 일회성 작업이 아닌 지속적인 프로세스로 이해해야 한다. 시장 환경과 소비자의 니즈는 계속해서 변화하기 때문에, 트렌드 분석 역시 지속적으로 이루어져야 한다. 이를 통해 변화의 조짐을 빠르게 포착하고 적시에 대응할 수 있다.
스몰 데이터를 활용한 트렌드 분석은 빅데이터 분석이 놓칠 수 있는 미세한 변화와 맥락을 포착할 수 있는 강력한 도구다. 이는 특히 소비자의 잠재적 니즈를 발견하고, 새로운 시장 기회를 포착하는 데 큰 도움이 될 수 있다. 따라서 기업들은 빅데이터와 스몰 데이터를 균형 있게 활용하는 통합적 접근법을 통해 보다 정확하고 깊이 있는 트렌드 분석을 수행할 수 있으며, 이를 통해 새로운 가치를 창출할 수 있다.

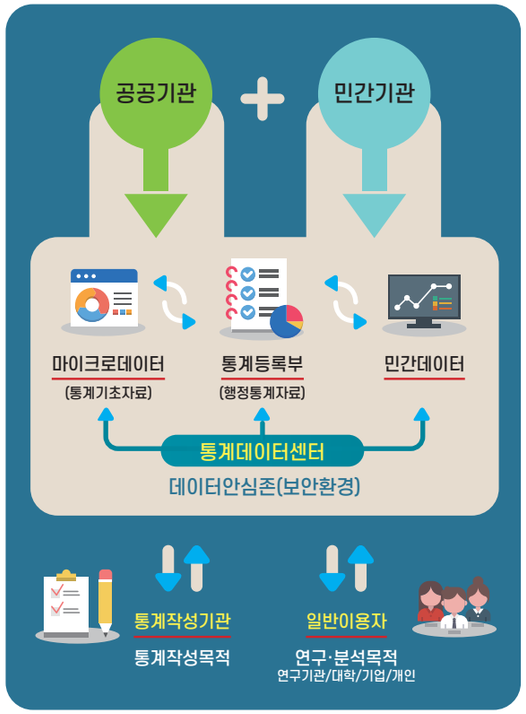
SDC
Statistics
Data Center

행정통계자료와 민간자료를 한곳에!

통계데이터센터 서비스

통계데이터센터가 새로운 서비스로
정보화 사회를 선도합니다.

행정자료를 수집하여 가공한 행정통계자료(통계등록부),
통계청이 제공하는 승인된 통계기초자료(마이크로데이터) 등
통계자료뿐만 아니라 민간자료까지 한 곳에서 분석이 가능한 통계데이터센터(SDC)



1 분석 플랫폼 제공 서비스
- 분석시스템 · 통계패키지 제공
- 통계자료(통계등록부 · 통계기초자료) 및 민간자료, 이용자 반입자료 연계 · 분석

2 전문가 분석지원 서비스
- 분석 경험이 없는 이용자를 위한 데이터 분석 지원
- 센터 이용 상담 및 데이터 분석 자문

3 주문형 분석서비스
- 시간 및 거리상 센터 방문이 어렵거나 직접 자료분석을 하기 힘든 이용자를 위한 서비스
- 센터 이용자료를 활용하여 연계 · 분석 후 이용자가 원하는 형태로 결과를 제공

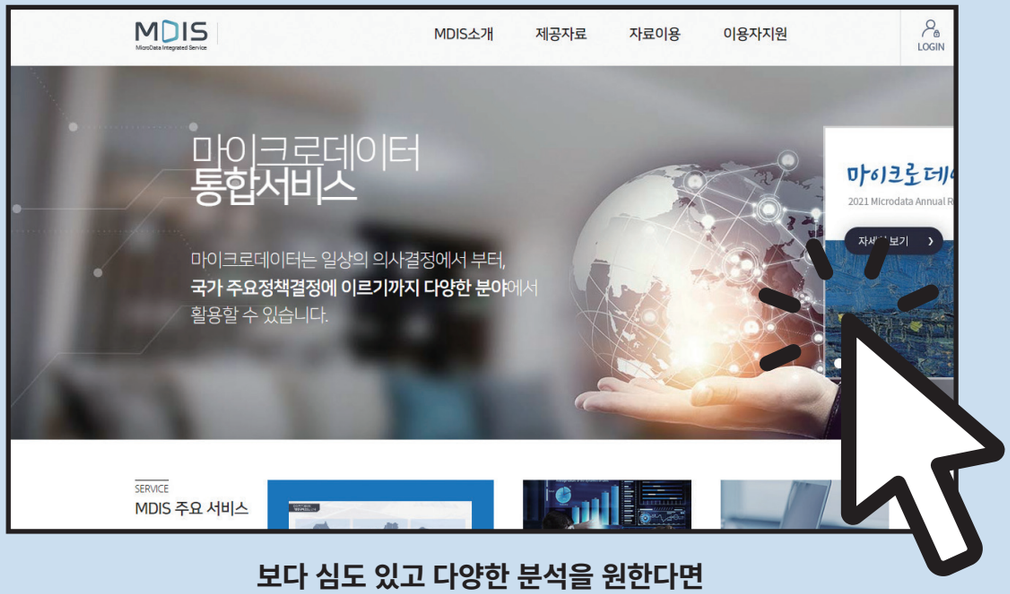
4 명부 서비스
- 분석센터로 방문하여 자료분석 및 표본설계를 통해 데이터 반출

5 이용자 교육 서비스
- 이용자 교육 홈페이지 운영
- 통계분석 프로그램 및 분석사례 교육
- 매년 통계데이터 활용대회 개최

※ RDC 제공자료도 이용 가능합니다.

통계청, 정부부처, 지방자치단체, 연구기관 등

모든 기관의 마이크로데이터를 한 곳으로



보다 심도 있고 다양한 분석을 원한다면
지금 바로 MDIS를 클릭해 보세요.

■ 서비스 소개 (2023년 5월 기준)

가. 서비스명 : 마이크로데이터통합서비스(MDIS, mdis.kostat.go.kr)
나. 제공 통계 수 : 21개 주제별 총 357종 통계 제공
(통계청 50종 및 통계작성기관 307종)
다. 제공 형태 : 마이크로데이터(통계에 따라 사람, 사업체, 가구 기반 자료)

기준		주요 통계
통계청	인구 · 가구	경제활동인구조사, 가계동향조사, 국내인구이동통계, 사망원인통계, 가계금융복지조사, 지역별고용조사, 인구주택총조사, 인구동향조사, 생활시간조사, 사회조사 외 8종
	사업체·농어가	전국사업체조사, 광업제조업조사, 농가경제조사, 기업활동조사, 농림어업총조사, 농산물생산비조사, 경제총조사, 어가경제조사, 운수업조사 외 14종
	행정통계	귀농어귀촌인통계, 영리법인기업체행정통계, 신혼부부통계, 주택소유통계, 중장년층행정통계, 퇴직연금통계, 일자리행정통계, 기업생멸행정통계, 육아휴직통계
통계작성기관		전국다문화가족실태조사, 가족실태조사, 자동차주행거리통계, 직종별사업체노동력조사, 보육실태조사, 기상관측통계, 국민여가활동조사, 외래관광객조사, 한부모가족실태조사, 청소년종합실태조사 외 297종

■ 서비스 내용

가. 구분 : 자료의 민감성 정도에 따라
공공용, 인가용으로 구분 운영

나. 수수료
- 무료 : 공공용 자료
- 인가용 : 선택제 수수료 부과

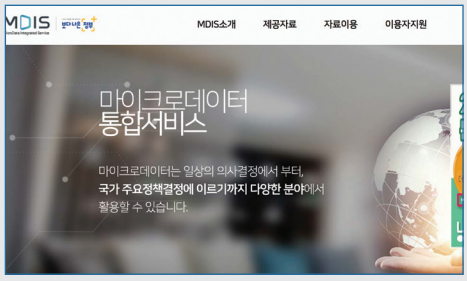
다. 서비스 방법
- 추출 · 다운로드 : MDIS 포털에서 직접 무료 다운로드
- 원격접근서비스 : 승인 후 이용자가 집·사무실 등에서 통계청 서버 접속 후 활용
- 이용센터 : 승인 후 지정된 장소를 방문·활용

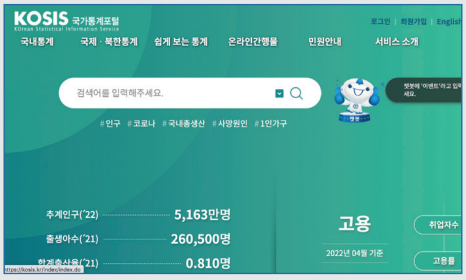
■ 문의

- 연락처 : 재단법인 한국통계진흥원
- 전화 : (02) 512-0167 FAX :(02) 515-0240
- 주소 : (우) 06097
서울특별시 강남구 선릉로 612, 6층
- E-mail : MDIS@stat.or.kr

통계청에서 국가통계를 활용하세요!

통계청은 통계개발·활용·교육에 필요한 모든 정보와 도움을 제공합니다.
다양한 국가통계정보 제공 사이트를 활용하세요.





통계교육원

sti.kostat.go.kr

국내 유일의 국가통계교육 전문기관

통계 작성 및 활용 전문통계과정,
기관맞춤형과정, e-러닝 과정

통계데이터센터

data.kostat.go.kr

행정통계자료와 민간자료를 한곳에

행정통계자료(통계등록부), 민간자료의
연계·융합이 가능한 데이터 플랫폼

MDIS

mdis.kostat.go.kr

원하는 자료를 직접 분석 및 요청

온라인으로 추출/다운로드 선택 시
공공용 마이크로데이터를 무료로 분석 활용 가능

KOSIS

kosis.kr

국가통계 쉽게 찾기

국내, 국제, 북한의 주요 통계를
한 곳에 모아 알기 쉽게 분류해 제공

SGIS

sgis.kostat.go.kr

지도 위 통계정보 살펴보기

인구, 가구, 주택, 사업체 통계 등 각종 통계를
지도(GIS) 위에서 한눈에 파악

통계청
통계교육원