

빅데이터 이용한 범죄 예방, 영화 ‘マイ너리티 리포트’가 현실로

Article at a Glance

공공 분야 데이터의 특징은 ‘모수(母數)’에 가까운 정보를 획득할 수 있고 이로 인해 자료가 통계적으로 비편향성(非偏向性)을 갖게 된다는 것이다. 따라서 ‘평가’와 ‘예측’ 두 가지 측면 모두에서 오류를 최소화할 수 있는 장점이 있다. 공공 데이터 활용의 선도 분야인 각국/각 지역 범죄예방 시스템에서 도출해낸 데이터 활용 원칙은 다음과 같다.

- 1) 다종의 데이터를 통합해 활용해야 한다.
- 2) 빅데이터에만 의존한 판단을 지양하고 다양한 맥락적 고려를 해야 한다.
- 3) 지속적인 데이터 검토 및 평가를 통해 예측 모형의 정확도를 제고해야 한다.
- 4) 데이터의 ‘공공성’ 자체에 대한 중요성을 염두에 둬야 한다.

편집자주

4차 산업혁명과 인공지능이 화두가 되면서 그 근간이 되는 ‘빅데이터’에 대한 관심이 커졌습니다. 각 기업에 쌓이는 민간 영역의 데이터도 중요하지만 공공 분야 데이터가 더해질 경우 더 큰 가치를 창출할 수 있습니다. 허정우 JCDA파트너즈 대표와 고승연 동아일보 미래전략연구소 공공전략센터장이 공공 분야 빅데이터 활용 전략을 함께 연재합니다.

연재를 시작하며

데이터 분석을 하는 이유는 무엇일까? 데이터 분석을 위해 수많은 방법론을 만들어내고 실험해보는 이유는 무엇일까? 이 두 질문에 대답하는 것이 곧 데이터 분석의 목적에 답하는 것이다.

데이터 분석에는 수많은 이유가 있지만 크게 ‘평가’와 ‘예측’이라는 두 가지 목적으로 구분해 볼 수 있다. 평가는 분석 시점을 기준으로 ‘과거’에 대한 해석을, 예측은 ‘미래’에 대한 추론을 각각 의미한다. 지금까지 다양한 방법론이 양산된 이유 역시 최소한의 변수를 활용해 최대한 많은 정보를 획득해 평가하고 미래에 대한 예측 오차를 최소화하기 위해서였다.

공공 데이터: 독점이 만들어내는 비편향성

데이터 분석에 대한 논의를 잠시 멈추고 이제 잠깐 공공 부문과 민간 부문의 차이에 대해 생

각해보자. 정부를 포함한 공공 분야가 민간과 차이를 보이는 가장 큰 특징은 공공 분야에서 생산되는 데이터의 성격이다.

첫째, 공공 부문은 특정 영역(특정 지역이 될 수도 있고 특정 산업이 될 수도 있음)에서 독과점적인 위치에 있다. 즉, 해당 영역에서 산출되는 정보를 독과점적인 위치에서 확보하고 관리할 수 있는 기반을 가지고 있다. 예를 들어, 서울시는 서울시 대중교통 사용자의 모든 이용정보를 파악하고 있다. 반면 특정 운수회사는 해당 회사 버스 이용자의 시간대별 구성과 같은 일부 정보만을 가지고 있을 뿐이다. 따라서 타사나 다른 대중교통 수단에 대한 이용정보에 대한 확보는 실질적으로 불가능하다. 또 다른 예로 카드사를 생각해볼 수 있다. 카드사는 자사 카드 소비자의 대중교통 이용 정보만을 파악할 수 있기 때문에 이 정보를 기반으로 ‘대중교통’ 분야에서 수행할 수 있는 활동은 제한적일 수밖에 없다. 독과점적인 위치에서의 정보 획득 기반은 데이터 분석의 기초가 되는 데이터를 최대한 모집단과 유사하게 가져갈 수 있어 평가와 예측의 오류를 최소화하는 효과를 얻을 수 있다. 만약 독과점적인 위치에서 확보할 수 있는 데이터가 두 종류 이상인 경우에는 데이터 간 통합을 통해 새로운 가치를 찾아낼 수 있는 기회도 얻을 수 있다. 실제 서울시의 경우 서울 시민의 버스 이용 정보와 지하철 이용 정보, 그리고 이들을 통합해 얻어진 환승 정보를 활용해 서울시내 버스 노선 체계를 수정·보완했다.

둘째, (첫 번째 조건을 만족한) 공공 부문의 데이터는 편향돼 있을 가능성이 낮다. 우리나라 건강보험심사평가원은 전 국민의 의료행태 및 전 요양기관의 진료행태에 대한 데이터를 기반으로 업무를 수행한다. 민간 실손보험회사 역시가입자의 의료행태 등에 대한 데이터를 소유하고 있다. 하지만 가입자 다수가 고액 진료 시에 실손보험을 활용하는 행태를 볼 때 그들이 갖고 있는 데이터는 ‘실손보험 가입 필요성을 느끼는 사람’과 ‘고액 진료를 경험한 사람’의 데이터로 편향돼 있을 가능성이 크다. 실손보험 회사가 확보한 데이터로는 가입자의 질환 변화에 대한 예측, 혹은 의료기관의 질적 수준 등에 대한 평가가 제한적으로 이뤄질 수밖에 없는 반면 건강보험심사평가원은 의료기관의 허위 진료 등에 대한 심사, 의료기관의 질적 수준 및 역량 수준에 대한 평가, 심지어 전염병 환자의 이동경로 분석 후 방역 등에 대한 대안 제시 등 다양한 활동을 수행할 수 있다. 데이터 분석을 수행할 때 부분적이거나 편향된 데이터가 평가와 예측에 있어 수많은 오류를 양산할 수 있다는 것을 감안하면 공공분야는 평가와 예측을 위한 데이터 수집 기반을 일정 수준 이상 확보하고 있는 셈이다.

공공 분야 빅데이터, 어떻게 활용되고 있나?

이제 앞선 두 논의를 합쳐보면 빅데이터의 장점을 활용해 공공 부문이 데이터 분석을 수행하고 이 결과를 활용하는 체계가 발생할 수 있음을 유추할 수 있다. 이는 특정 집단이 아닌 다수의 국민들에게 아주 큰 의미를 줄 수 있음을 의미한다. 물론 수집기반이 있는 것과 이를 제대로 수행하고 있는 것은 별개의 문제다. 실제로 데이터를 수집해 ‘의미 있는’ 결과를 산출하는 역량은 공공 부문이 민간보다 떨어진다. 오죽하면 ‘엉터리 통계’를 일컫는 우스갯소리로 ‘공무원 통계’라는 말까지 나오겠는가.

현재 빅데이터를 활용한 대부분의 성공 사례, 즉 과거에 대한 평가 정확도를 높이고 미래의 예측 정확도를 높인 사례는 민간에 집중돼 있다. 지금까지 공공 분야에서 성공적으로 빅데이터가 활용됐다고 평가받는 분야는 크게 의료, 범죄 등의 일부 영역으로 한정돼 있다. 이 영역들은 주요국에서 공공 부문의 빅데이터가 논의된 초기 시점부터 ‘사회문제 해결 및 정책 활용 가능 분야’로서 논의된 영역에 해당한다. [표 1]은 2010년을 전후로 주요국에서 사회문제 해결을 위해 빅데이터가 활용된 사례를 보여주는데 의료와 범죄 관련 영역으로 범위가 치중돼 있는 것을 알 수 있다.

표 1 주요국의 초기 빅데이터 활용 사례

구분	의료	범죄	기타
미국	- 국립보건원 의료체계 개혁	- 태아 대응 프로그램(위험 예측 및 통합 대응) - 범죄 예측 프로그램 - 탈세 방지 프로그램	
영국	- 비만 대책 프로그램 - 전염병 대응 프로그램	- 잠재적 위험관리 프로젝트	
일본			- 지능형 교통 시스템 - 재난 대응 프로그램
싱가포르	- 조류인플루엔자 시뮬레이션	- 해안 안전 시뮬레이션(해상 태리 및 침투 대비)	

• 한국정보화진흥원, ‘빅데이터 시대, 공공부문 빅데이터 추진방향(2012)’, ‘신한국의 데이터기반 국가 미래전략 실행현황과 시사점(2012)’ 재면집

의료와 범죄 관련 영역에서 빅데이터가 빠르게 적용될 수밖에 없었던 이유는 무엇일까?

첫째, 데이터의 구성 때문이다. 아무리 많은 양의 데이터가 있어도 과거와 현재의 데이터 간 형태가 다른 경우 데이터 분석이 불가능해진다. 하지만 의료와 범죄 관련 데이터는 정형화된 형태로 정리된다. 의료는 질병 관련 국제코드로, 범죄는 정형화된 조서 등을 기반으로 관리되기 때문에 타 영역보다 데이터 분석을 위한 장기간의 ‘일관된 방식으로 정리된’ 기초 통계 확보가 용이하다.

둘째, 사건의 반복성 때문이다. 의료와 범죄 영역에서는 유사한 결과를 야기하는 상황이 지속·반복되기 때문에 유사 상황에 대한 데이터가 지속적으로 누적돼 분석할 수 있을 만한 사례 확보가 용이한 편이다.

셋째, 사건 데이터의 차별성이다. 앞의 두 조건을 만족한 공공 데이터는 많을 수 있지만 이들 중 의료와 범죄 영역은 예스(Yes)와 노(No), 좀 더 정확하게 말해 문제 상황과 그렇지 않은 상황이 명확하게 구분된다. 이런 특징으로 인해 다른 데이터보다 제3의 요인이나 정성적 판단이 개입될 여지가 상대적으로 덜한 편이다. 이번 호 아티클에서는 주로 미국 등 선진국을 중심으로 확산되고 있는 공공 데이터 활용 범죄 발생 방지 사례를 다뤄보고자 한다.

1. 공공 데이터와 범죄 발생 방지

‘빅데이터 기반 범죄예방’은 전 세계적으로 확대되고 있는 활용 사례다. 빅데이터를 통해 범죄가 일어날 시간과 장소를 예측하고 이에 대한 정보를 경찰에 전달해 범죄를 사전에 방지하는 것을 ‘예측 치안’이라고 하는데 미국에서는 콤파스(COMPAS), 프리드폴(PredPol), 크라이

스캔(CrimeScan), 크러시(C.R.U.S.H.) 등의 프로그램을 활용해 이를 수행하고 있다. 우리나라도 2016년부터 ‘치안과학기술 연구개발사업’이라는 과제를 통해 빅데이터 기반 범죄 분석 프로그램을 개발하고 있다. 일반적으로 수행하고 있는 ‘빅데이터를 활용한 범죄예방’ 프로그램은 1) 범죄에 대한 예측 2) 범죄자에 대한 예측 3) 범죄자 신원에 대한 예측 4) 피해자에 대한 예측의 네 가지 범주로 이뤄지고 있으며 주로 1)과 2)를 통해 사전 예방 및 대응 활동을 진행하고 있다.

① 범죄에 대한 예측

범죄에 대한 예측은 크게 세 가지 기준을 기반으로 해 ‘어디에서, 어떤 범죄가, 왜’ 많이 발생하는지를 예측한다. 첫째, 범죄는 무작위적으로 분포하지 않고 특정 지역에 집중(이른바 ‘핫스폿’)돼 발생한다. 둘째, 이런 장소적 패턴은 대체로 변화하지 않는다. 셋째, 소수의 중범죄인을 제외한 대부분의 잠재적 범죄자들은 지역 특성과 범죄 기회를 파악하고 적절한 시기와 장소를 선택해 범죄를 저지른다.

범죄 예측의 핵심은 과거에 저지른 범죄기록 기초 데이터를 ‘평가’해 일관된 패턴을 찾아 이를 기반으로 미래에 발생할 수 있는 범죄 발생 가능성을 ‘예측’하는 것이다. 단순하게는 범죄 다발지역 및 시간, 범죄 발생 가능성이 높은 고위험 지역 및 시간에 대한 예측을 수행하지만 회귀분석이나 신경망 분석 등의 분석기법과 조합해 응용할 경우 각 원인 변수에 따른 분석 대상의 상대적 위험 수준까지도 예측할 수 있다.

그렇다면 주요국들이 범죄를 예측하는 이유는 무엇일까? 몇 가지 이유가 있겠지만 가장 큰 이유는 범죄율 증가에 비해 쫓아오지 못하는 경찰력 부족, 즉 절대인원 부족과 신규 인원의 경험 부족에 따른 역량 부족을 해소하기 위함이다. 구체적으로는 잠재적 범죄자들이 범죄기회를 파악하고 범행을 저지를 것으로 예상되는 시간·장소에 순찰을 강화하고 경찰력이 대비케 함으로써 범죄를 예방하는 것이다.

실제로 범죄 예측은 경찰력 부족 문제를 일정 수준 해소할 수 있는 것으로 나타났다. 미국 캘리포니아 산타크루즈는 인구가 약 10%가량 늘어난 반면 경찰 인원은 약 20%가량 감소하면서 범죄가 25% 이상 증가했다. 이에 효율적으로 범죄에 대응하기 위해 ‘핫스폿’을 예측해 특정 시간대에 범죄가 증가할 것으로 예상되는 지점을 파악하고 이에 대한 정보를 각 경찰과 공유해 순찰방식을 개선했다. 그 결과 시범운영 첫 달에 재산피해 범죄 27% 감소 및 핫스폿에서 7명 검거라는 성과를 거뒀다. 그뿐만 아니라 신입 경찰들에게 핫스폿 정보를 제공함으로써(핫스폿에 대한 정보는 스마트폰 등을 통해 실시간 제공) 범죄에 대한 사전 대응 역량 및 직무수행 능력 향상이라는 결과도 가져왔다. 유사 사례로 미국 로스앤젤레스는 범죄 예측을 통해 범죄율이 12% 감소했다. 동 기간 내 타 지역에선 범죄가 0.4% 증가했다는 사실과 비교해보면 범죄 예측이 경찰력 확대에 효과적인 수단이 될 것임을 알 수 있다.

뉴욕경찰국(NYPD)은 여기에서 한 단계 더 나아가 실시간 감시·대응시스템을 연계한 범죄감시 시스템(DAS·Domain Awareness System)을 테러리스트 및 범인 추적에 활용하고 있다. 이 시스템은 범죄 예측을 기반으로 도시 내 CCTV, 자동차 번호 감지기 등 각종 센서에서 취합

된 정보와 국세청·톨게이트·911녹음파일 등 20여 개의 정보를 통합해 의심스러운 정황이 발생할 경우 일선 경찰관에 관련 정보를 제공함으로써 대응 시간을 최소화하는 데 기여하고 있다. 이는 결과적으로 경찰력을 확대하는 성과로 이어졌다.

② 범죄자에 대한 예측

장소에 대한 예측 차원을 넘어 ‘누가, 언제, 어떤 범죄를 저지를 것인지’를 예측하는 것이다. 영화 ‘마이너리티 리포트’에서 ‘3명의 예지자’가 하던 예언을 ‘빅데이터와 컴퓨팅 파워’가 대신 할 뿐 추구하는 바는 크게 다르지 않다. 물론 아직까지 범죄자에 대한 예측은 크게 높지 않다. 가장 큰 이유는 범죄를 유발할 수 있는 요인에 대한 확정이 어렵고, 또 이와 연관된 광범위한 데이터의 입력과 활용이 쉽지 않기 때문이다. 그뿐만 아니라 영화 ‘마이너리티 리포트’에 나오는 오류인 ‘실제 발생하지 않은, 발생이 확정되지 않은 사건’에 대한 사전 조치에 따르는 인권침해 등 윤리적 문제와 ‘무죄추정의 원칙 위배’라는 법적 문제도 야기한다. 이러한 이유로 ‘범죄자에 대한 예측’은 초범 자체를 막기보다는 재범 가능성성이 높은 범죄자들의 미래 범죄 위험 가능성을 평가하고 이를 예방하기 위해 활용되는 추세다.

교정시설에 수용된 인원을 관리하는 데에는 상당한 수준의 비용이 투입되는 것으로 알려져 있다. 우리나라의 경우 2015년 기준으로 1인당 약 2500만 원의 비용이 필요하며 교정시설 신규 건설 등의 어려움으로 과밀 수용의 문제도 동시에 발생하고 있다. 미국 역시 이러한 문제(OECD 발표 자료에 따르면 미국 교정시설 수용률은 103.9%로 13위에 해당)를 해소하기 위해 교정시설에서 지역사회로의 환원(가석방 및 보호관찰 등)을 시행한다. 교정당국은 대상자에 대한 주기적 관찰을 통해 이들의 재범 가능성을 최소화할 필요가 있었는데 구체적으로 예산 부담을 최소한으로 유지한 상황에서(교정시설 수용보다는 확연하게 낮은 비용을 사용하면서) 재범 가능성 역시 최소화하는 감시전략 수립이 필요했다. 이를 위해 미국 필라델피아의 성인 보호관찰가석방국(APPD·Adult Probation and Parole Department)은 ‘랜덤 포레스트(Random Forest)’라는 기법을 활용해 2년 내 재범 가능성을 3가지(없음, 경범죄, 중범죄) 위험군으로 구분하는 방식을 개발했다. 재범확률 추정을 위해 범죄경력, 과거 수감기록, 마지막 중범죄 후 경과 기간 등 12개 대표변수를 산출하고, 이를 활용해 가석방이나 보호관찰 대상자의 2년 내 재범 가능성을 예측하는 것이다. APPD는 이렇게 추정된 위험군을 기반으로 과거 획일적 감시전략을 위험군별 감시전략으로 전환했다. 그 결과 고위험군을 관리하는 요원에게는 적은 수의 대상자를 할당하고, 저위험군을 관리하는 요원에게는 상대적으로 많은 수를 할당함으로써 보다 개선된 감시 체제를 운영할 수 있게 됐다.

영국 역시 범죄자 평가 시스템(OAsys·Offender Assessment System)이라는 재범 예측 프로그램을 활용하고 있는데 폭력 범죄의 경우 재범예측 확률은 75% 수준에 이른다. 이보다 좀 더 적극적으로 범죄자를 사전 예측하는 프로그램으로는 런던 경찰국(MPS)에서 운용하고 있는 쟁단 범죄 예방 프로그램(Complete Analytics Pilot Program to Fight Gang Crime)이 있다. 구체적인 방법론은 아직 공개되지 않았지만 5년 내 범죄 전력이 있는 런던 내 쟁단의 범죄 기록, SNS 게시물에서의 특정 발언 등을 수집해 대상자가 강력범죄를 저지를 가능성을 예측하고 있으며, 일정 수준 이상의 효과를 거두고 있다고 한다. 물론 이 프로그램은 인권침해 등의 가능성으로 인해 일부 시민단체들로부터 공격을 받고 있는 것도 사실이다.

③ 범죄자 신원에 대한 예측

이미 발생한 범죄가 누구의 소행인지를 분석하는 활동으로 검거를 위한 분석 활동에 가깝다. 구체적으로 범죄 장소 및 범죄자를 프로파일링하고 향후 발생할 범행의 시간 및 장소 등을 예측하는 것이다. 범죄자 신원에 대한 예측은 단지 검거를 위한 분석 활동에만 사용되지 않고 프로파일링을 기반으로 범죄 가능성은 미연에 방지하기 위해 활용되기도 한다. 미국 플로리다의 경우 100만 명 이상 소년범의 범죄기록을 기반으로 프로파일링을 실시하고 맞춤형 교화프로그램을 개발해 소년범들이 이를 이수케 함으로써 재범 가능성을 최소화하고 있다. 실제로 이렇게 개발된 맞춤형 교화프로그램은 시행 2년여 만에 소년범죄를 15%가량 줄이는 데 기여했다. 그뿐만 아니라 프로파일링을 지리학, 수사심리학 등과 연계해 지역별 경찰력 자원배분의 최적화(범죄예방의 첫 번째 범주인 ‘범죄에 대한 예측’)를 위해서도 활용하고 있다.

④ 피해자에 대한 예측

세 번째 범주인 ‘범죄자 신원 예측’과 유사한 기법을 활용해 범죄 발생 가능성이 높은 지역에서 피해를 당할 가능성이 높은 피해자를 예측하는 것이다. 기본적으로 앞서 언급했던 세 가지 범주의 예측을 기반으로 수행하는데 범죄에 대한 예측을 피해자 관점에서 재해석하는 것으로 이해하면 된다. 일반적으로 범죄 예측 후 가능성이 높은 지역으로 순찰 시, 피해 가능성이 높은 집단을 파악하고 해당 집단을 관찰할 수 있는 장소에서 잠복을 수행해 범죄 발생 직전 피해자를 구제하고 범인을 검거하는 목적으로 사용된다.

2. 공공 데이터 구축을 통한 범죄 예방 시스템과 시사점

그렇다면 빅데이터를 활용한 범죄 예방 시스템 구축·운영의 시사점은 무엇일까?

첫째, 데이터 간 통합의 중요성이다. 여기서의 통합은 범죄와 관련해 경찰이 보유하고 있는 데이터들을 단지 묶어놓는 것만을 의미하지는 않는다. 경찰 보유 데이터와 타 부문이 가지고 있는 각종 데이터를 통합함으로써 기존 데이터만 활용한 예측보다 정확도를 향상시키고 지리학 데이터와의 결합을 통해 빅데이터 분석에서 가장 난해하다는 시각화까지도 해결하고 있다.

이러한 데이터 통합은 경찰 입장에서는 통합이지만 경찰에 데이터를 제공하는 기관 입장에서는 일종의 데이터 개방이 일어나고 있음을 의미한다. 공공 부문의 부문 간 데이터 개방과 통합으로 새로운 가치가 발굴되고 있는 것이다. 흥미로운 점은 아직까지 어떤 국가도 통합을 위한 통일된 규칙은 제시하지 못한 반면 통합의 순서와 방향은 대체로 일치한다는 것이다. 데이터 통합은 대체로 과거의 범죄 데이터를 데이터베이스화하는 것을 1단계로, 여기에 신고·제보 데이터를 결합하는 2단계, 타 공공 부문이 확보하고 있는 인적사항 관련 데이터를 결합하는 3단계, 센서·CCTV 등 지역에 설치된(혹은 민간이 보유하고 있는) 인프라에서 확보되는 실시간 정보까지도 통합하는 4단계의 구조를 보이고 있다. 1~2단계가 경찰 내부의 과거 및 실시간 데이터를 통합하는 것이라면 3~4단계는 외부의 과거 및 실시간 데이터까지도 통합하는 단계

라 볼 수 있다. 실시간 데이터의 빠른 입력과 이에 따른 신속한 처리 및 전송의 적시성 때문에 과거 데이터보다 통합되는 시점이 후순위로 나타나고, 개인정보 공유 등의 문제 해소가 필요하기 때문에 내부 데이터보다 외부 데이터의 통합이 후순위로 나타나는 경향을 보인다.

둘째, 빅데이터에만 의존한 판단은 하지 않는다는 것이다. 즉 계량화된 데이터에 기반한 예측만으로 판단하는 것은 지양한다는 얘기다. 여기에는 빅데이터를 활용한 범죄 예방 시스템 구축이 논의되고 시작됐던 시점에 비정형 데이터 활용 수준이 높지 못해 경찰들이 가지고 있던 ‘노하우’라는 비정형 데이터를 연계할 수 없었던 것이 하나의 원인으로 작용한다. 그뿐만 아니라 어떤 예측모형을 만들 때에는 그러한 예측을 하게 된 이유, 즉 이론적 근거를 기반으로 만들어야 하는데 범죄는 유발요인에 대한 명확한 규명이 쉽지 않아 모형화에 한계가 있다는 것도 중요한 이유였다. 결국 컴퓨터가 하는 예측은 말 그대로 분석일 뿐이지 평가가 아니므로 최종적인 판단은 사람이 수행해야 한다는 것이다.

셋째, 상호작용을 통한 예측모형의 정교화를 추구한다는 것이다. 범죄 유발요인에 대한 명확한 규명 없이 과거 데이터만으로 미래를 예측한다는 것은 상당한 위험을 내포하고 있다. 편향적 예측, 재범우려가 높은 위험도가 낮은 인물로 분류하는 예측모형의 오류(2종 오류) 등이 있을 수 있다. 또 범죄가 발생할 것으로 예측된 장소에 경찰력이 사전 대비를 할 경우, 해당 장소에서 범죄를 저지르려 했던 잠재적 죄인은 범행을 멈추기도 하지만 시간과 장소를 바꿔 범행을 재차 시도할 수도 있다. 이 경우 재차 시도하는 범죄에 대한 추가적인 예측은 기존 데이터만으로는 한계가 있을 수 있다. 따라서 빅데이터를 이용한 범죄 예측 모형은 장기간의 데이터를 일관적으로 수집해 입력하는 방식보다는 지속적으로 검토하는 방식을 활용하는 경향을 보인다. 그리고 모형의 개선뿐 아니라 실제 치안활동 운영 방식에 대한 개선도 동시에 수행하고 그 결과를 실시간으로 입력해 예측 치안의 방향성을 다시 검토하는 등 ‘모형-활동-평가’ 간 상호작용’을 통해 모형의 정확도를 개선하고 있다.

넷째, 혁신을 수행하는 기관 자체의 ‘효율성 개선’ 외에 ‘추가적인 가치(고객 편의 증가 혹은 비용 절감)’도 제공해야 한다. 범죄 예측은 기존보다 절감된 예산하에서 범죄 발생률을 억제하는 비용 효율적 특성 외에 사회 전반에서 ‘삶의 질 제고’라는 고객 편의 증가를 수반한다. 특히 일반 시민들의 삶의 질에는 범죄율이라는 통계값보다 ‘자신이 범죄의 피해자가 될 수 있다는 두려움’이 더 영향을 미친다. 예측 치안을 통한 취약 가능 지역에의 역량 집중과 이를 기반으로 수행하는 다양한 치안 활동 운영 전략들은 범죄 외에도 무질서나 두려움 감소에도 효과적인 것으로 알려져 있다. 실제로 범죄 예측을 치안 운영 전략에 활용하는 많은 기관들은 ‘비용 절감이나 범죄 해결뿐만 아니라 지역사회와의 전반적인 협력을 통한 삶의 질 향상 기여’를 운영 목표로 하고 있다.

결론

현재 각국이 사용하고 있는 범죄 예측 시스템이 가진 한계는 ‘우발적 범죄’에 대한 예측이 실질적으로 어렵다는 것이다. 그 이유는 범죄 예측이 ‘핫스폿을 중심으로 범죄패턴이 존재한

다’는 이론적 기반을 근거로 하기 때문이다. 실제 예측 가능한 범죄도 이런 이론을 기반으로 한 범죄들(절도, 폭력 등)로만 국한돼 있다. 이런 한계는 범죄 예측의 이론적 배경이 획기적으로 개선돼 범죄 유발 요인에 대한 명확한 규명이 이뤄지기 전까지 계속될 것이다. ‘상관관계를 규명하는 데는 도움이 될 수 있으나 인과관계 규명에는 한계가 있다’는 빅데이터 분석의 한계가 범죄 영역에서도 동일하게 발생하는 것이다. 따라서 당분간 범죄 예측 시스템은 예측 오차를 줄이고 상관관계 분석을 통한 위험 시그널을 경찰에 제공해 경찰 업무 성과를 개선하는 목적으로 사용될 가능성이 높다. 다음 아티클에서는 교통 분야 공공 데이터의 현황과 활용 전략 및 사례에 대해 소개하겠다.

허정우 JCDA파트너즈 대표 lukehur@jcda.co.kr

필자는 연세대에서 경제학 석사 학위를 받았다. 공군사관학교(경제경영학과) 전임 강사를 거쳐 공공 분야 및 교육·의료 분야 전문 컨설팅업체 갤렙앤컴퍼니에서 매니저로 약 10년간 일했다. 2017년 JCDA파트너즈를 설립해 현재 국내 공공기관 및 대학, 병원 등의 혁신 및 프로세스 개선을 돋고 있다.