

## 데이터 과학에 기반한 납세자 행동 예측모델 도입방안 연구(요약)

과거에 사후 검증을 강화함으로써 성실신고를 유도하려 했던 여러 세정 노력들은 2012 ~ 2014년 세수 부족 시기에 납세자의 반발에 부딪히며 한계를 노정하였다. 이후 ‘자발적 성실신고 지원’, ‘탈세·체납 엄정 대응’ 2가지 큰 방향으로 국세행정의 패러다임이 전환되었고 현재까지 국세행정 운영의 근간이 되고 있다.

변화된 국세행정은 납세자의 자발적인 행동 변화를 유도하여 세정 순응도를 제고하는 데 크게 기여했지만, 한계점도 여전히 존재하고 있다. 따라서, 성실납세 문화를 실질적으로 정착시키고 행정 효율성을 제고하기 위해서는 국세행정의 새로운 운영방식을 모색해 볼 필요가 있다. 현재의 패러다임은 ‘신고 전·후’의 성실도 확보에 중점을 두고 있지만, 그보다 앞서 사업을 개시하고 거래가 이루어지는 단계에서 일상적으로 납세자를 모니터링하고 납세자의 행동을 예측하는 방안을 한층 더 발전시킬 필요가 있다. 이에 민간에서 많이 활용되는 연체 위험도 예측, 고객행동 예측 분석등을 국세행정에 응용할 수 있는 방안을 강구해 보고자 「데이터 과학에 기반한 납세자 행동 예측모델 도입방안 연구」를 연구주제로 선정하였다.

데이터 과학을 국세행정의 관점에서 보면 그것은 다양한 지식을 활용하여 납세자로부터 데이터를 수집하고 분석하여 유용한 정보를 창출함으로써 국세행정의 효율적이고 효과적인 운영을 도모하는 모든 활동을 의미한다고 볼 수 있을 것이다. 한편 납세자 행동은 세금의 신고, 납부와 관련한 납세자의 의사결정과 실제로 이루어진 행위를 의미하며, 납세자 행동은 소득, 재정 상황, 조세에 대한 태도, 정부 정책에 대한 인식, 세법 준수 의식 등 다양한 요인의 영향을 받을 것으로 예상해 볼 수 있다.

미국의 사회보장국(The United States Social Security Administration, SSA)과 의료보장센터(The Centers for Medicare & Medicaid Services, CMS)는 비정상적이고 의심스러운 사기 청구 패턴을 식별하기 위해 예측 분석(Predictive Analytics)을 사용한다. 아울러 미국 시카고 경찰국은 IBM과의 협력을 바탕으로 기계학습 및 예측 분석을 범죄 예방에 활용하고 있으며, 로스엔젤레스 경찰국은 과거 범 죄 및 체포 데이터를 기반

으로 범죄 발생 장소와 미래 범죄자를 예측하고자 노력하고 있다. 또한 미 교육부(The United States Department of Education)는 교육 현장에서 데이터 마이닝과 빅데이터 분석을 이용하여 교육의 질을 향상시키고 있는 등 다양한 방면에서 데이터 과학이 활용되고 있다.

국내에서도 국가정보자원관리원은 경찰청의 입장일지 데이터를 이용하여 여죄 추적 모델을 구현하였고, 서울시 영등포구는 ‘여성 안심 빅데이터 셉테드(CPTED) 협업 플랫폼’을 구축하여 활용하고 있다. 또한 건강보험공단은 2013년부터 ‘국민건강 알람서비스’를 통해 감기, 눈병, 식중독, 천식, 피부염 5개 질병에 대해 4단계로 위험 정도를 제공하고 있고, 고용부는 반복적으로 위법 행위를 하는 사업장을 선별하는데 데이터 분석을 활용하고 있다.

국세행정의 관점에서 보면 3가지 유형의 납세자 행동 예측이 필요할 것으로 생각된다. 우선 세금 확보를 위한 행정에서는 정상적인 궤도에서 이탈하여 세금을 확보하는 데 어려움을 겪게 만드는 불성실 사업자를 예측하고 관리하는 데 납세자 행동 예측의 의의가 있을 것이다. 더불어 납세자 지원을 위한 국세행정에 있어서는 세금 신고를 손쉽게 마칠 수 있도록 도와주고, 세금 납부의 어려움이 있는 사업자를 신속하게 찾아내는 데 납세자 행동 예측이 의미를 가질 수 있을 것으로 생각한다. 그리고 민원 발급을 위한 행정에서는 민원인이 얼마나 많이 집중될지, 민원인이 얼마나 기다려야 할지 예측함으로써 민원 서비스의 품질 제고에 납세자 행동 예측이 필요할 것으로 보인다.

불성실 사업자를 예측하는 데 있어서는 다항 로지스틱 회귀분석 모델을 구현하면 임의의 사업자가 고액·상습채납자일 확률이 얼마일지, 탈루 혐의자일 확률이 얼마일지 예측해 볼 수 있다. 불성실 사업자일 확률이 일정한 수준 이상인 사업자들을 선별하여 세원 관리 또는 세무조사를 담당하는 직원들에게 제공한다면, 과거 경험 많은 직원의 직관에 의존해야 했던 불성실 사업자 관리 방식을 과학적이고 시스템적인 방식으로 전환하는 중요한 계기가 될 수 있을 것이다. 또한 다양한 이상점 탐지 기법을 활용하여 불성실 사업자를 탐지할 수 있다. 사업자가 스스로 통제할 수 없는 요인들을 이용하여 군집을 형성하거나 모형을 적합시킨

다음, 군집 또는 추정된 모형에서 동떨어져 있는 사업자가 존재하는지 확인하면 불성실 사업자를 탐지해 낼 수 있을 것이다.

납세자 지원의 국세행정에 있어서는 행렬분해와 유사도 측정을 이용하면 신고서 작성 과정에서 납세자의 신고 오류를 탐지해 낼 수 있을 것이다. 다른 납세자들의 신고 내용과 비교하여 유사도가 적은 경우 납세자에게 경고를 하고 어떤 항목에 문제가 있는지 알려줌으로써 납세자가 어떤 항목을 과대하게 또는 과소하게 작성했는지 탐지할 수 있다. 그리고 행렬분해를 이용하면 잘못 기재된 값의 근사값도 알려줄 수 있다. 또한 생존분석을 활용하면 경제적 어려움을 겪는 지역과 업종을 실시간으로 탐지할 수 있으므로, 시의적절한 세정지원 안내를 해 줄 수도 있다. 사업자 등록 및 폐업 신고 자료를 이용하여 사업 영위 기간에 대해 생존분석을 실시한다면 어떤 지역 또는 업종의 생존률이 낮은지 탐지할 수 있을 것으로 기대된다. 예컨대 예년에 비해 생존률이 낮아진 지역 또는 업종이 있다면 업황이 좋지 않은 것으로 볼 수 있을 것이다. 아울러 범주형 자료 분석을 사용하면 장려금 홍보 매체별로 정확하게 효과를 평가할 수 있고, 홍보 대상자별로 효과적인 홍보 매체를 분석할 수도 있다.

민원 발급 행정에 있어서는 마코프 체인에 기반한 M/M/s 대기열 모델을 사용하면, 평균 대기시간, 대기해야 할 확률, 대기열에 있는 평균 고객 수 등을 예상할 수 있다. 간단한 모델로 이를 예상할 수 있다면 경험과 직관에 의존하지 않고도 추가 인력의 투입 또는 인근 세무서 분산 안내 등을 통해 원활하게 민원을 처리할 수 있는 방안을 강구해 볼 수 있을 것이다.

데이터 과학은 많은 교육과 훈련이 필요하기 때문에, 기존의 인력을 교육하여 나날이 발전하는 데이터 과학을 뒤쫓아 가기에는 큰 무리가 따른다. 다행히 각광을 받는 분야이기에 많은 사람들이 도전하고 있다. 따라서 국세청, 그리고 넓게는 정부 조직 전체에 걸쳐 데이터 분석 관련 민간 전문 인력을 적극적으로 채용할 수 있을 것이다. 그리고 이를 주도할 전담 조직을 별도로 구성할 필요가 있다.