조직 간 자원 보완성을 통한 데이터 2차 사용의 이중 가치 창출과 자산화: 유통·물류 산업의 데이터 마찰과 정착 경로 분기에 관한 단일 사례 연구

Dual Value Creation and Data Assetization in Secondary Data Use through Inter-Organizational Resource Complementarity: A Single-Case Study of Data Friction and Divergent Settlement Paths in the Retail–Logistics Industry

이대성(주저자)
Dae-Sung Lee(First Author)

성균관대학교 경영대학 School of Business at Sungkyunkwan University(*leeds1122@gmail.com*)

국내 대기업 물류 운영사와 빅데이터 전문기업 간 2년에 걸친 데이터 협력을 단일 사례로 고찰하였다. 자원 보완성은 협력의 출발점으로, 상호 부족한 자원인 운영 데이터와 분석 역량의 결합이 복수의 사업 가치를 창출하고, 그 가치가 지속 가능한 데이터 자산으로 정착하며, 그 정착 경로가 조건에 따라 어떻게 분기하는지를 하나의 인과 사슬로 분석하였다. 그리고 패턴 매칭과 인과 추론 테스트로 분석적 일반화를 수행하였다. 분석 결과, 데이터의 비경합성으로 동일한 데이터 자원이 상이한 AI 모델을 통해 이중 가치로 실현됨을 확인하였다. 또한 협력 맥락에서 정식 계약을 기점으로 데이터 자산화의 다단계 제도화가 관찰되었다. 나아가 데이터 마찰을 데이터 분석 차원과 제도적 환경 차원으로 구분하고, 두 차원의 차이가 자산화 정착 경로의 분기와 대응한다는 분석 틀을 제시하였다. 실무적으로는 사업 가치가 확인된 이후에도 제도적 환경 마찰이 해소되지 않으면 자산화 정착이 지연될 수 있음을 시사하였다.

주제어: 자원 보완성, 이중 가치, 데이터 자산화, 제도화, 데이터 2차 사용, 데이터 마찰

This study examines a two-year data collaboration between a major Korean logistics operator and a big-data firm as a single-case study. Resource complementarity serves as the starting point of the collaboration, where combining mutually lacking resources—operational data and analytical capabilities—could generate multiple forms of business value. The study analyzes, within a single causal chain, how these values crystallize into sustainable data assets and how the settlement paths diverge under varying conditions. Pattern matching and causal-inference tests support analytic generalization. The findings confirm that, due to the non-rivalry of data, the same data resource realizes dual value when applied through distinct AI models. Furthermore, within this collaborative context, a multi-stage institutionalization of data assetization was observed, with the formal contract serving as a critical turning point. The study further distinguishes data friction into a data-analytic level and an institutional-environmental level, proposing an analytical framework in which differences across these levels correspond to divergent settlement

최초투고일: 2026. 06. 01 수정일: (1차: 2026. 07. 21) 게재확정일: 2026. 08. 01

Copyright 2026 THE KOREAN ACADEMIC SOCIETY OF BUSINESS ADMINISTRATION
This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0, which permits unrestricted, distribution, and reproduction in any medium, provided the original work is properly cited.

paths for data assetization. From a practical perspective, the findings imply that even after business value is established, the final settlement of data assetization might be delayed if institutional-environmental friction remains unresolved.

Keyword: Resource Complementarity, Dual Value, Data Assetization, Institutionalization, Secondary Data Use, Data Friction

Ⅰ. 서 론

디지털 경제가 전개되면서 데이터는 운영의 부산물을 넘어 전략적 자산으로 인식되고 있다. Xu et al.(2024)은 데이터의 지위가 데이터(data)에서 데이터 자원(data resource)을 거쳐 데이터 자산(data asset)으로 진화한다고 보았다. 먼저 데이터는 수집·관리되어 분석에 활용할 수 있는 상태로 갖추어지는 자원 준비(resource readiness)를 거쳐 데이터 자원이 된다(Xu et al., 2024). 이렇게 준비된 데이터 자원은 실제 활용을 통해 사업 가치를 창출한다(Grover et al., 2018). 그 가치가 일회성에 그치지 않고 지속적으로 실현될 때 데이터는 데이터 자산으로 기능하며, 데이터를 자산으로 성립시키는 근거는 데이터가 창출하는 미래 가치이다(Birch et al., 2021; Xu et al., 2024). 이처럼 데이터 활용으로 창출된 가치가 반복적으로 실현될 수 있는 구조를 갖추어 가는 과정이 데이터 자산화(data assetization)이며, 그 핵심은 데이터를 보유하는 데 그치지 않고 데이터가 반복적으로 활용되어 가치를 창출할 수 있도록 관리·조정하는 체계를 구축하는 데 있다.

국내에서도 디지털 비즈니스에 대한 관심이 빠르게 높아져, 인공지능과 데이터를 중심으로 비즈니스 기회 창출을 위한 다양한 트렌드가 보고되고 있다(김태경 외, 2025). 국내 사회과학·공학 분야에서 인공지능 연구는 양적으로 성장하고 있으며, 윤리와 더불어 데이터 관리·거버넌스의 중요성도 부각되고 있다(이창환 외, 2025). 거버넌스가 부각되는 배경에는 데이터의 지위가 높아지는 것만으로는 가치 실현으로 이어지지 않는다는 인식이 자리한다. 데이터 거버넌스는 데이터를 활용하고 수익화할 수 있는 조직적 역량의 토대를 이루며, 부족한 자원을 보완하는 조직 간 협력을 조정함으로써 사업 가치 실현을 뒷받침한다.

Wixom and Ross(2017)는 데이터 보유 자체가 더 이상 차별점이 아니며, 데이터 수익화(data monetization) 능력이 디지털 경제에서 경쟁우위의 원천이 된다고 지적하였다. 많은 조직이 데이터가 사업 가치를 창출할 잠재력을 인식하면서도, 수익화 전략을 실질적인 재무 성과로 연결하는 데에는 여전히 어려움을 겪고 있다(van der Meulen et al., 2025). 가치 실현 과정을 구체적으로 분석하기 위해서는 데이터 활용이 창출하는 성과의 유형과 위계를 명확히 할 필요가 있다. Grover et al.(2018)의 가치 창출 프레임워크에 근거하여, 연구에서는 데이터 활용이 조직에 가져오는 다양한 성과를 포괄하는 상위 개념으로 데이터의 사업 가치(business value)를 사용하며, 이하 맥락상 명확한 경우 '가치'로 줄여 쓴다. 그 실현 형태는 운영 효율을 높이는 효율화 지향 가치(efficiency-oriented value)와

새로운 기회를 여는 발견 지향 가치(discovery-oriented value)의 두 하위 유형으로 구분된다. 두 하위 가치가 동일한 데이터 자원으로부터 동시에 실현되는 것을 이하 이중 가치(dual value)로 부른다. 이렇게 창출·구분된 가치는 그 자체로 완결되지 않으며, 반복적인 의사결정과 운영 구조에 내재화되어야 조직의 지속적 경쟁우위로 이어질 수 있다. 그 정착의 과정과 조건이 이하 논의의 초점이 된다.

사업 가치는 데이터 자원과 분석 역량 자원이 한 조직 안에 함께 갖춰진 경우에도 쉽게 실현되지 않는다. 더욱이 두 자원이 서로 다른 조직에 분산되어 있을 경우, 데이터 보유와 가치 실현 사이의 간극은 한층 커질 수 있다. 조직 간 데이터 협력은 이 간극을 메우는 중요한 출발점이 될 수 있다. 그러나 기존 연구는 이 협력이 실제로 어떤 메커니즘을 통해 사업 가치 창출과 데이터 자산화로 이어지는지, 그리고 그 과정이 어떤 조건에서 안정적으로 정착되거나 분기되는지를 충분히 설명하지 못하고 있다.

데이터 보유와 가치 실현 사이의 간극이 해소되는 과정을 살펴보기 위해 대기업군에 속한 국내 물류 운영사 A사와 데이터 전문 기업 B사의 빅데이터 통합 협력을 단일 사례로 분석하였다. 협력은 2024년 하반기부터 약 2년에 걸쳐 진행되었다. A사는 일평균 약 170만 건의 물류 운영 데이터와 물류 운영 도메인 지식을 보유하고 있으나 분석 역량과 수요 선행 데이터가 부족하다. 반면 B사는 연간 약 2,000만 명 규모의 유통 구매·결제 데이터, 표준 상품 마스터 체계, 빅데이터 분석 역량을 보유하고 있으나, 자사 보유 데이터만으로는 외부 온라인 시장의 거래를 관찰하기 어렵다. 두 회사는 서로에게 부족한 자원을 보완함으로써 단독으로는 실현하기 어려운 사업 가치를 창출할 수 있는 협력 구조를 형성한다. 협력은 동일한 통합 데이터 기반에서 AI 예측 모델(물류 예측; 효율화 지향 가치)과 분류 모델(온라인 상품 분류; 발견 지향 가치)을 중심으로 전개되었다. 이러한 설정은 데이터 자산화와 AI 활용 논의가 만나는 지점을 조직 간 협력의 현장에서 관찰할 기회를 제공한다.

연구의 목적은 두 회사의 데이터 협력에서 사업 가치가 어떻게 발생하고, 어떻게 지속 가능한 데이터 자산으로 정착하며, 그 정착 경로가 왜 갈라지는지를 규명하는 데 있다. 이를 위해 자원 보완성, 데이터 비경합성, 다단계 제도화, 그리고 데이터 마찰이 어떻게 발생하고 해소되어 하나의 경로를 형성하는지를 분석한다. 자원 보완성은 인과 사슬의 출발점으로, 서로에게 없는 자원인 운영 데이터와 분석 역량의 결합을 성립시킨다. 데이터 비경합성은 결합된 동일한 데이터 자원이 예측 모델과 분류 모델을 거쳐 효율화 지향 가치와 발견 지향 가치를 동시에 실현하고, 그 가치가 양사에 각각 귀속되는 근거가 된다. 다단계 제도화는 발생한 가치가 지속 가능한 데이터 자산으로 정착하는 과정이다. 그 과정에서 데이터의 2차 사용에 따른 데이터 마찰은 데이터 분석 차원과 제도적 환경 차원으로 나뉘어 발생하며, 두 차원에서 마찰이 해소되는 양상에 따라 자산화 정착 경로가 달라진다.

연구 목적은 세 가지 연구 문제(Research Question, 이하 RQ)로 구체화된다. 연구 문제는 가치의 발생(RQ-1), 데이터 자산으로의 정착(RQ-2), 데이터 마찰에 따른 정착 경로의 분기(RQ-3)라는 순차적 흐름으로 구성된다.

연구 문제 RQ-1: 두 회사가 보유한 데이터 자원과 분석 역량의 보완성은 통합 데이터 아키텍처와 예측·분류 모델을 통해 어떻게 작동하며, 동일한 데이터 자원으로부터 어떠한 복수의 사업 가치를 창

출하는가?

연구 문제 RQ-2: 양자 협력에서 발생한 가치는 어떠한 제도화 과정을 거쳐 일회성 프로젝트를 넘어 지속 가능한 데이터 자산으로 정착되는가?

연구 문제 RQ-3: 데이터의 2차 사용에서 발생하는 데이터 분석 차원과 제도적 환경 차원의 데이터 마찰은 가치의 정착 경로를 어떻게 분기시키며, 예측 모델과 분류 모델의 정착 경로는 왜 서로 다르게 나타나는가?

여기서 '가치가 데이터 자산으로 정착한다'는 표현은 가치 자체가 독립적인 자산으로 전환된다는 의미가 아니다. 데이터가 창출한 가치가 반복 가능한 조직 구조에 내재화되어 자산적 지위가 지속성을 획득하는 과정을 뜻하며, 이와 같은 과정적 구성개념(processual construct)의 정의와 근거는 2.4절에서 상술한다.

위 세 가지 연구 문제에 대한 사례 분석 결과, 다음과 같은 양상이 관찰되었다. 조직 간 협력에서 동일한 데이터 자원은 서로 다른 AI 모델을 거쳐 이중의 사업 가치로 실현될 수 있음이 확인되었다. 예측 모델의 효율화 지향 가치는 A사에, 분류 모델의 발견 지향 가치는 B사에 각각 귀속되었다. 창출된 가치는 탐색적 분석, 정식 계약, 공식적 가치 입증, 상시 운영 및 시스템 수정으로 이어지는 다단계 제도화를 거쳐 지속 가능한 데이터 자산으로 정착하는 모습을 보였다. 다만 자산화는 단선적으로 완결되지 않고, 동일한 계약과 통합 데이터 아키텍처를 공유한 두 모델의 정착 경로가 데이터 마찰의 차원별 차이에 따라 분기하는 것으로 나타났다.

이상의 결과는 기존 논의에 세 가지 이론적 기여를 제시한다. 먼저 소유권과 이익 배분을 중심으로 전개되어 온 데이터 비경합성 논의를 보완 관계의 이질적 자원이 결합하는 조직 간 협력으로 확장한다. 그리고 기업 내부 차원에 머물던 데이터 자산화 논의를 정식 계약을 전환점으로 하는 조직 간 다단계 제도화 과정으로 재구성한다. 나아가 상호 연관된 현상으로 다뤄져 온 데이터 마찰을 발생 원천에 따라 구분하는 분석 틀을 제시한다. 실무적으로는 사업 가치가 확인된 이후에도 제도적 환경 차원의 마찰이 해소되지 않으면 자산화 정착이 지연될 수 있음을 시사하며, 마스터 데이터 표준화의 사전 구축과 정제 책임 분배의 사전 설계가 필요함을 보여준다.

연구는 단일 사례 설계에 기반한 명제 검증 연구이며, 검증의 일반화는 통계적 일반화가 아니라 분석적 일반화(analytic generalization)에 해당한다. 검증 대상인 명제는 2장 이론적 배경에서 도출하고, 단일 사례 설계의 근거, 분석 단위, 명제 검증의 인과 추론 절차는 3장에서 다룬다. 명제에 대한 사례 분석 결과는 4장에서 제시하며, 5장에서는 연구 문제별 결과 해석과 대안적 설명을 검토한다.

II. 이론적 배경

2.1 데이터 자원과 빅데이터 분석 역량

자원 보완성(resource complementarity). 자원이 가치를 창출하는 핵심 원리는 서로 다른 자원 사이의 보완성에 있다. Milgrom and Roberts(1995)는 두 자원 요소가 함께 쓰일 때 서로의 효과를 강화하는 관계로 보완성(complementarity)을 규정하고, 전체가 부분의 합보다 커지는 이 구조를 초가법성(supermodularity)의 수학적 형식으로 이론화하

였다. 자원 보완성(resource complementarity)은 보완성 이론을 자원 차원에 적용한 것이다(Harrison et al., 2001). 보완 관계의 자원이 결합될 때 발생하는 시너지는 유사 자원의 결합으로는 도달할 수 없는 고유하고 모방하기 어려운 가치를 만들어낸다. 다만 보완 자원의 존재 자체가 시너지를 보장하지는 않으며, 효과적 통합과 관리가 동반되지 않으면 시너지는 발생하지 않는다(Harrison et al., 2001). 자원 보완성은 가치 창출의 출발점이며, 자원을 결합하는 통합 방식과 그 위에서 작동하는 운영 역량이 가치 실현의 결정적 통로가 된다.

데이터 비경합성(non-rivalry). 비경합성(non-rivalry)은 한 주체의 사용이 다른 주체의 동일 자원 사용을 제약하지 않는 자원의 특성이다. Jones and Tonetti(2020)는 비경합성을 데이터 경제학의 핵심 개념으로 정립하면서, 데이터가 일반적인 비경합적 자원과 어떻게 구별되는지를 규명하였다. 데이터는 여러 기업·용도·분석 모델에 동시에 쓰여도 가치가 소진되지 않으며, 활용 범위가 넓어질수록 가치가 수확체증한다. 비경합성은 데이터가 자원(resource)에서 자산(asset)으로 격상되는 경제적 기반이 된다. 비경합성을 다룬 기존 논의의 초점은 소유권과 이익 배분에 있었다. Jones and Tonetti(2020)는 동일 데이터가 여러 기업에 의해 사용될 때 발생하는 사회적 이익에 주목하면서, 그 이익을 실현하기 위한 데이터 소유권의 귀속과 프라이버시 보호의 균형 및 후생의 관점에서 논의를 전개하였다.

최근 논의는 비경합성의 이면도 함께 지적한다. 데이터 공유가 경쟁우위를 침식할 수 있어 기업이 데이터를 공유(data sharing)하지 않고 비축(data hoarding)하게 되고, 그 결과 비경합 자원이 사회적으로 최적인 수준에 미치지 못하고 좁게 활용되는 딜레마가 제기된다(Wang et al., 2025). 이 딜레마는 데이터가 전략적 대체재(strategic substitutes)로 작동하는 조건, 곧 경쟁자와 데이터를 공유할 때 기업이 보유한 데이터의 상대적 가치가 낮아지는 상황을 전제한다(Liu et al., 2025; Wang et al., 2025). 결국 비경합성의 가치 창출은 데이터 공유 주체들이 경쟁 관계인지 보완 관계인지에 달려 있다고 볼 수 있다.

빅데이터 분석 역량(big data analytics capability, BDAC). 빅데이터 분석 역량(BDAC) 문헌은 데이터 자체가 아니라 데이터를 활용하는 조직 역량이 성과의 직접적 원천이라는 명제를 일관되게 견지해 왔다(Gupta and George, 2016; Wamba et al., 2017). Gupta and George(2016)는 BDAC를 유형·인적·무형 자원으로 이루어진 형성적 구성(formative construct)으로 정의하였다. 세 자원 차원이 함께 결합되어야 BDAC가 발현된다는 점에서, BDAC는 보완성의 논리가 조직 내부에서도 작동하는 다차원 자원으로 볼 수 있다. 이후 연구는 BDAC가 동적·운영 역량을 매개로 경쟁 성과로 전환되고(Mikalef et al., 2020), 그 산출물이 행동으로 전환되어 의사결정·프로세스에 영향을 미치는 사업 가치(business value)를 창출한다는 점(Grover et al., 2018)을 제시하였다.

AI 기반 모델과 사업 가치. 빅데이터 분석 역량이 실제로 구현되는 도구가 인공지능 모델이다. Perifanis and Kitsios(2023)는 양손잡이(ambidexterity) 이론(March, 1991)을 AI 사용 양태에 적용하여, AI의 사업 가치를 일상적 사용(routine use, exploitation)과 혁신적 사용(innovative use, exploration)의

두 축으로 정리하였다. 사례는 이 구도를 인공지능 모델에 적용하여 두 측면에 주목한다. 하나는 일상적 사용에 해당하는 운영 효율 측면(operational efficiency)으로, 이미 수행되는 업무의 정확도와 효율을 끌어올리고 운영 의사결정을 개선하여 비용을 절감하는 효율화 지향 가치다. 다른 하나는 혁신적 사용에 해당하는 새로운 가치 영역을 여는 측면(unlocking new forms of value)으로, 자사 데이터만으로는 관찰할 수 없던 외부 시장과 수요를 가시화하여 사업 판단의 근거를 넓히는 발견 지향 가치다. 두 가치의 구분은 인공지능 모델 유형과 자료 형식의 구조적 차이와 대응한다. 효율화 지향 가치는 구조화·시계열 데이터에 기반해 수요의 선행 신호를 결합함으로써 미래 상태를 추정하는 예측 모델을 통해(Lee et al., 2000), 발견 지향 가치는 비정형 텍스트 데이터를 분석 가능한 범주로 전환하는 분류 모델을 통해(Devlin et al., 2019; Liu et al., 2019) 실현된다. 이와 관련하여 국내에서도 온라인 소비자 행동 데이터를 소상공인의 무형 자산가치로 재정의하고, 이를 Distance-to-Default 기반 구조적 위험 모형에 통합하여 폐업 위험 예측 성능을 향상시킨 연구가 보고되었다(정동영 외, 2026).

2.2 데이터 자산화, 제도화 및 거버넌스

데이터 자산화(data assetization). 데이터 자산화(data assetization)는 데이터를 단순한 운영 기록이 아니라 전략적 자산으로 인식하고 그 가치를 실현 가능한 형태로 전환하는 과정이다. Xu et al.(2024)은 데이터의 지위 변화를 데이터, 데이터 자원, 데이터 자산으로 이어지는 단계적 진화로 정리하면서 자원 준비·역량 구축·데이터 활용의 자산화 프레임워크를 제시하고, 데이터 자산을 실질 가치 또는 잠재 가치(real or potential value)를 지닌 자원으로 정의하였다. Birch et al.(2021)은 자산화를 측정·거버넌스·평가(measurement·governance·valuation)의 실천을 거쳐 자원이 미래 수익 흐름(future revenue streams)을 낳는 자본화된 재산(capitalized property)으로 전환되는 과정으로 규정하였다. 두 논의는 자산의 성립 근거를 공통으로 잠재·미래 가치에 둔다. 곧 데이터를 자산으로 성립시키는 것은 데이터 보유나 소유권 확보 자체가 아니라, 데이터가 창출하는 미래 가치이다.

제도화(institutionalization). 가치가 일회성에 그치지 않고 조직 변동에도 지속되는 자산화가 되려면, 그 가치가 구조로 정착하는 제도화(institutionalization)를 거쳐야 한다. 제도화는 특정 실천이 처음의 정당화 근거나 도입 주체로부터 점차 분리되어, 조직 구성원의 교체에도 살아남는 당연시된 구조로 자리 잡는 과정을 가리킨다(Selznick, 1957; Scott, 2014). Tolbert and Zucker(1996)는 제도화 과정을 습관화(habitualization), 객체화(objectification), 침전화(sedimentation)의 단계적 진행으로 구조화하였다. 습관화는 특정 자극이나 조직 문제에 대응하여 정형화된 문제 해결 행동이 형성되고, 그러한 행동이 특정 상황과 연결되는 단계다. 객체화는 이러한 행동에 일반적이고 공유된 사회적 의미가 부여되어, 조직 의사결정자들 사이에서 가치 있는 해결책으로 인식되는 단계다. 침전화는 그 구조가 역사적 연속성을 획득하고 조직 구성원의 교체를 넘어 지속됨으로써 완전한 제도화에 이르는 단계다. 제도화를 단계적 과정으로 보는 이 관점은, 일회적으로 발현된 가치가 어떤 경로를 거쳐 조직 변동에도 지속되는 자산으로 정착하는가

라는 사례의 문제의식과 직접 맞닿는다.

Tolbert and Zucker(1996)의 제도화 3단계는 이후 연구에서도 제도의 진행 수준을 구분하는 분석 틀로 활용되고 있다. Rageth et al.(2021)은 제도적 견고성(institutional robustness)을 평가하기 위한 이론적 틀에서 제도화 정도(degree of institutionalization)를 핵심 차원으로 설정하고, 이를 전(前)제도화(pre-institutionalization), 준제도화(semi-institutionalization), 완전 제도화(full institutionalization)의 세 수준으로 구분하였다. 이들은 각 수준을 습관화, 객체화, 침전화와 연결하면서, 제도화가 새로운 문제 해결 행동의 형성에서 공유된 의미와 사회적 합의의 확보를 거쳐, 역사적 연속성과 전승 가능성을 갖는 구조로 진전된다고 보았다.

데이터 거버넌스(data governance). 데이터 자산화가 조직 내부의 실천에 머무르지 않고 조직 간 데이터 협력으로 확장될 경우, 제도화는 단순한 조직 내부의 구조화 문제가 아니라 데이터의 소유·접근·활용을 조정하는 거버넌스 문제와 결합된다. 따라서 조직 간 데이터 자산화 과정을 이해하기 위해서는 데이터 거버넌스 관점이 함께 고려될 필요가 있다. Robey et al.(2008)은 조직 간 정보시스템(interorganizational systems, 이하 IOS) 연구가 IOS 채택 요인, IOS의 경제 거래 거버넌스, IOS의 조직적 결과의 세 축에서 전개되어 왔음을 보였다. Jussen et al.(2024)은 53개 조직 간 데이터 공유 사례를 동등 협력(peer-to-peer)과 중개자 기반(data intermediary-based)의 두 그룹으로 분류하였다. 동등 협력에는 양자 간 공유와 서비스 중심 공유가, 중개자 기반에는 마켓플레이스, 트러스트, 플랫폼 및 데이터 스페이스가 해당된다. Abraham et al.(2019)은 데이터 거버넌스를 데이터를 전략적 기업 자산으로 관리하기 위한 교차기능적 프레임워크로 정의하고, 그 메커니즘을 구조적·절차적·관계적 세 축으로 분류하였다.

2.3 데이터 2차 사용과 데이터 마찰

데이터의 2차 사용(secondary use of data). 데이터의 2차 사용은 특정 목적을 위해 생성·수집된 데이터를 그와 다른 후속 목적에 재활용하는 것을 가리킨다(Safran et al., 2007). 조직 간 데이터 협력에서 이루어지는 운영 데이터의 분석 활용은 물론, 사례의 예측 모델이 결합하는 유통사 구매·기상 데이터 역시 원생성 목적과 다른 분석 목적에 활용된다는 점에서 넓은 의미의 2차 사용에 해당한다. 기존 데이터를 2차적으로 활용한 국내 연구를 보면, 음원 플랫폼의 음향 데이터가 장르 전략과 인기도 분석에 활용되었고(이중원&정옥경, 2026), 소비자 위치정보가 대출 상환 데이터와 결합되어 신용 채무불이행 예측을 보완할 수 있는 정보임을 보인 예가 있다(류가루&방영석, 2026).

그러나 2차 사용에서는 데이터를 생성한 측과 재사용하는 측이 분리되며, 생성 측이 생성 맥락과 목적에 관한 암묵지를 보유한다는 데이터 창조자의 우위(data creators' advantage)가 재사용자에게 제약으로 작용한다(Pasquetto et al., 2019). Borgman and Groth(2025)는 이 비대칭을 창조자와 재사용자 사이의 거리(distance)로 개념화하였다. 그 가운데 이 연구가 주목하는 것은 원생성 목적과 재사용 목적의 차이를 다루는 목적 차원의 거리(distance in purposes)다. 목적 거리가 클수록 데이터의 이동 과정에서 조정 비용과 장애를 낳는다. 이 장애를 설명하기 위해 제시된 개념이 데이터 마찰(data

friction)이다.

데이터 마찰(data friction). 데이터 마찰은 데이터가 한 활용 맥락에서 다른 활용 맥락으로 이동할 때 그 이동을 가로막는 저항으로, 데이터의 내재적 속성이 아니라 사람·기관·시스템 사이의 접점에서 발생하는 사회·기술적 현상이다(Edwards, 2010). 후속 연구는 마찰의 발생 원천이 단일하지 않음을 보였다. Bates(2018)는 마찰의 구성 요인을 인프라·관리, 사회문화, 규제의 세 범주로 확장하였고, Aula(2019)는 마찰이 일상적 데이터 실천만이 아니라 거버넌스를 둘러싼 법적 협상에서도 발생함을 보였다. 앞선 논의를 종합하여, 2차 사용에서 발생하는 데이터 마찰을 발생 원천에 따라 데이터 분석 차원(data-analytic level)과 제도적 환경 차원(institutional-environmental level)으로 재구성한다.

데이터 분석 차원. 데이터 분석 차원은 데이터의 정보적 속성과 생성 맥락에서 비롯되는 마찰을 가리킨다. Edwards et al.(2011)은 마찰의 핵심이 데이터 자체보다 데이터를 해석하는 데 필요한 메타데이터에 있음을 보였다. 맥락 정보가 누락되거나 표준화되지 않으면 데이터의 재사용 가능성과 해석 가능성은 크게 제한되며, 이는 데이터 창조자의 우위(Pasquetto et al., 2019)와 목적 차원의 거리(Borgman and Groth, 2025)가 재사용자에게 부과하는 제약과 맞닿아 있다. 목적 거리가 클수록 데이터를 새로운 분석 목적에 맞게 변환·해석하는 조정 비용이 커지며, 분석의 가능성과 성과가 제약된다.

제도적 환경 차원. 제도적 환경 차원은 데이터의 구성적 특성이 법·규제·거버넌스와 결합하여 데이터의 공유와 활용을 제약하는 마찰을 가리킨다. Bates(2018)가 제시한 세 범주 가운데 규제 프레임워크는 마찰이 단순한 기술적 장애가 아니라 행위자들 사이의 권력과 이해관계, 규범이 충돌하며 형성되는 현상임을 보여준다. 법적 이용 제약, 개인정보 보호 의무, 영업비밀과의 결합, 거버넌스 책임의 분배는 어떤 데이터가 누구에게 보이고 누가 활용할 수 있는가를 제약한다. 대표 사례로 핀란드 의료 데이터 개혁에서의 거버넌스 간 이해관계 충돌을 들 수 있으며(Aula, 2019), 데이터 소유권의 모호성은 자산화의 핵심 제도적 장애가 될 수 있다(Qin et al., 2025).

2.4 연구 명제의 도출

이상의 이론적 배경을 토대로, 기존 문헌이 충분히 설명하지 못한 공백에서 연구 명제를 도출한다. 세 명제는 1장에서 제시한 연구 문제, 곧 가치의 발생(RQ-1), 데이터 자산으로의 정착(RQ-2), 데이터 마찰에 따른 정착 경로의 분기(RQ-3)에 각각 대응한다.

먼저 데이터 비경합성 문헌은 비경합성을 주로 소유권과 이익 배분, 또는 경쟁자 간 데이터 공유의 관점에서 논의하여, 상호 보완적인 이질적 자원, 곧 데이터와 분석 역량이 결합하는 협력 관계는 분석 범위 밖에 두었다(P-1). 데이터 자산화 문헌은 데이터가 자원에서 자산으로 진화하는 기업 내부 차원의 단계와 자산화를 성립시키는 측정·거버넌스·평가의 실천을 설명해 왔으나, 복수 기업 간 협력에서 발생한 일회성 가치가 어떠한 다단계 제도화 과정을 거쳐 지속 가능한 데이터 자산으로 정착하는지는 직접 다루지 않았다(P-2). 또한 데이터 마찰 문헌은 분석 차원과 제도적 환경 차원의 마찰을 상호 연관

된 현상으로 다루어, 발생 원천이 서로 다른 두 차원의 마찰이 각기 다르게 발생·해소되고 그 차이가 자산화의 정착 경로를 가른다는 점은 충분히 검토하지 않았다(P-3). 이상의 공백을 검증 가능한 명제로 정리하면 다음과 같다.

명제 P-1(가치 발생): 이질적 자원의 결합과 데이터의 비경합성으로 인해, 동일한 데이터 자원은 서로 다른 AI 모델을 통해 복수의 사업 가치로 동시에 전환될 수 있으며, 그 가치는 협력 조직 간에 차별적으로 귀속될 수 있다.

명제 P-2(가치 정착): 양자 협력에서 획득된 가치는 탐색적 분석, 거버넌스 합의, 공식적 가치 입증, 상시 운영, 시스템 수정의 다단계 제도화를 거쳐 지속 가능한 데이터 자산으로 정착될 수 있다.

명제 P-3(정착 경로): 양자 협력에서 획득된 가치의 정착 경로는 데이터의 2차 사용 과정에서 발생하는 데이터 분석 차원의 데이터 마찰과 제도적 환경 차원의 데이터 마찰의 차이에 따라 분기한다.

명제 P-2와 P-3이 사용하는 '가치의 정착'은 탐색적 분석, 거버넌스 합의, 공식적 가치 입증, 상시 운영 및 시스템 수정으로 이어지는 상정된 인과 연쇄(a presumed set of causal sequences) 전체를 지칭하는 과정적 구성개념(processual construct)이다. Yin(2018)은 설명 구축(explanation building)을 상정된 인과 연쇄를 명시하는 작업으로 설명하며, Langley(1999)는 과정 자료로부터 이론을 구성할 때 정확성과 함께 간명성(parsimony)을 중요한 기준으로 제시한다. '가치의 정착'은 이 복합 과정을 일관되고 간명하게 지칭하기 위한 분석적 축약(analytic shorthand)이며, 데이터가 창출한 가치가 반복 가능한 조직 구조에 내재화되어 자산적 지위가 지속성을 획득하는 과정을 의미한다.

세 명제의 이론적 공백과 검증 논리 간의 대응 관계는 3장의 〈표 3〉에 정리한다.

Ⅲ. 연구 방법

3.1 연구 설계

연구 설계는 Yin(2018)의 단일 사례 연구(single-case study) 모형을 따른다. 사례연구 방법 선택의 근거는 연구질문의 형태, 행동적 사건에 대한 통제 정도 및 현상의 동시대성이다(Benbasat et al., 1987; Yin, 2018). 연구 문제 RQ-1~RQ-3은 가치가 어떻게 발생, 정착 및 분기하는가를 묻는 'how' 형태로 진술되어 있으며, 분석은 인과적 'why' 질문을 함께 함축한다. Yin(2018)에 따르면 'how·why' 질문은 시간에 걸쳐 전개되는 작동 방식과 인과 경로를 다루므로 설명적 사례연구(explanatory case study)에 해당한다. 양사 협력의 전개는 연구자가 통제하거나 조작할 수 없다는 점에서 실험과 구별되고, 사례는 2026년 5월 현재까지 진행 중인 동시대적 현상(역사 연구와의 구별점)이라는 점에서 사례연구 적용 조건이 충족된다.

Yin(2018)은 단일 사례 설계의 근거를 결정적·특이한·일상적·계시적·종단적 사례의 다섯 유형으로 구분하였다. 연구 사례는 다섯 근거 가운데 결정적·계시적·특이한·종단적 사례의 네 가지에 해당하며, 각 근거의 정의와 연구 사례에 대한 적용은 〈표 1〉과 같다.

〈표 1〉 단일 사례 설계의 정당화 근거

사례 유형	정의	연구 사례 적용
결정적 사례 (critical case)	▪ 단일 사례가 유의미한 이론에 대한 결정적 검증(critical test)을 제공하는 경우 (이론이 명제가 참이라고 믿어지는 명확한 조건을 명시한 상황에서, 사례가 그 명제를 확인, 도전, 확장함으로써 지식과 이론 구축에 기여하는 경우)	▪ 사례가 명제 P-1・P-2・P-3에 대한 경험적 근거를 제공하여 분석적 일반화에 기여 ▪ 참조: 〈표 3〉 명제별 인과 추론의 검증 논리
계시적 사례 (revelatory case)	▪ 외부 연구자가 접근하기 어려웠던 현상을 관찰할 기회가 주어진 경우	▪ 외부 연구자가 접근하기 어려운 두 회사 간 이질 자원(데이터・분석 역량)의 결합과 양사 데이터 간 결합의 내부 과정을 관찰할 기회
특이한 사례 (unusual case)	▪ 이론적 규범이나 일상적 발생에서 벗어난 양상	▪ 조직 간 협력을 단위로 동일한 통합 데이터 아키텍처 아래에서 서로 다른 가치의 자산화 정착 경로가 분기하는 드문 조건이 형성
종단적 사례 (longitudinal case)	▪ 단일 사례를 둘 이상의 서로 다른 시점에서 연구하는 경우	▪ 2024년 하반기 비공식 탐색에서 출발하여 2025년 5월 정식 계약, 12월 상시 배치 운영 진입, 2026년 5월 현재 예측 모델의 시스템 수정(완료)에 이르는 2년의 단계적 진화 추적

분석 단위(unit of analysis)는 양사의 데이터 통합 협력 과정이다. 의사결정, 과정, 프로그램 및 일련의 사건 같은 유동적 실체도 분석 단위로 적합하며, 둘 이상의 조직이 결합하는 협력은 공동 프로젝트, 공동 프로그램, 조직 간 협력(interorganizational arrangement)으로 구분된다(Yin, 2018). 조직 간 협력 자체를 분석 단위로 삼으면, 내부 구성 프로젝트를 하위 분석 단위(embedded units of analysis)로 함께 다룰 수 있다. 연구 사례는 데이터 통합 협력을 사례로 삼고 그 안의 예측 모델과 분류 모델을 하위 분석 단위로 설정한다.

외적 타당성(external validity)은 사례 연구의 발견이 다룬 사례를 넘어 다른 상황으로 일반화될 수 있는가를 묻는 설계 품질 기준이다(Gibbert et al., 2008; Yin, 2018). 확보 수단은 통계적 일반화가 아니라 분석적 일반화이며, 연구 설계 단계에서 연구 목적의 진술, How/Why 형태의 연구 문제 설정, 적절한 이론적 명제(theoretical propositions)의 식별로 출발한다(Yin, 2018). 1장에서는 연구 목적과 연구 문제 RQ-1・RQ-2・RQ-3을 진술하였고, 연구 문제에 대응하는 명제 P-1・P-2・P-3은 2.4절에서 이론적 공백으로부터 도출하여 외적 타당성 확보의 출발점을 세웠다. 명제는 4장 검증을 거쳐 5.2절 이론적 기여에서 분석적 일반화(이론적 개념의 보강・수정・진전; Yin, 2018)로 발전한다. 설계 품질의 나머지 차원인 구성 타당성과 신뢰성은 3.2절(자료 수집), 내적 타당성은 3.3절(명제 검증의 인과 추론 절차)에서 검토한다.

자료 분석은 Yin(2018)이 권고한 패턴 매칭(pattern matching)과 설명 구축(explanation building)을 결합한다. 패턴 매칭은 명제 P-1・P-2・P-3을 예상 패턴으로 사전 설정하고 사례 자료가 패턴에 어떻게 대응하는지를 확인하는 절차이며 4장에서 수행한다. 설명 구축은 패턴 매칭으로

확인된 현상에 인과 설명을 단계적으로 정교화하는 절차이며, 인과 사슬의 통합과 대안적 설명 검토를 포함하여 5장에서 수행한다.

사례의 하위 분석 단위인 예측·분류 모델은 조직 문제 해결을 위해 설계·구축된 인공물(artifact)이다. 그 유효성 평가는 설계과학 연구(design science research)의 평가 프레임워크인 FEDS(Framework for Evaluation in Design Science)에 근거한다(Venable et al., 2016). 평가의 관점은 실험실·시뮬레이션 등 통제된 조건에서 인공물의 효능(efficacy)을 검증하는 인공적 평가(artificial evaluation)가 아니라, 자연주의적 평가(naturalistic evaluation)다. 자연주의적 평가는 실제 사용자·실제 시스템·실제 문제라는 세 실재 요건을 충족하는 실제 조직 환경에서 인공물의 실효성(effectiveness)을 검증하며, 통계적 유의성이 아니라 실제 조직 맥락에서의 수용·활용·성과로 판단한다.

3.2 자료 수집

자료 수집은 구성 타당성과 신뢰성을 확보하기 위한 절차이다. 구성 타당성은 연구 개념이 적절한 조작적 측정으로 연결되었는가를, 신뢰성은 연구 절차가 반복 가능하며 동일한 결과를 산출할 수 있는가를 가리킨다(Gibbert et al., 2008; Yin, 2018).

양사 협력의 역할 분담에 따라, AI 기반 모델 개발을 위한 자료 수집과 분석은 주로 B사가 수행하였다. A사는 운영 중심 조직으로서 운영 데이터 제공과 예측 모델 결과의 A사 시스템 반영에 협력하였으며, 모델 개발·분석·해석·배치 운영은 B사의 분석 역량과 전담 인력이 담당하였다. 이에 따라 모델 성능과 분석 결과의 기술적 검토는 분석 역량을 보유한 B사가 수행하였고, A사의 검증 참여는 데이터 제공, 운영 적용, 거버넌스 사실 확인, 그리고 분석 결과가 운영 현실에 부합하는지에 대한 타당성 확인에 중심을 두었다. 구성 타당성과 신뢰성을 확보하기 위해, Yin(2018)이 권고한 다중 출처 활용, 증거 사슬 유지 및 정보 제공자 검토 절차를 적용하였다. 익명화와 내부자인 연구자 위치성에서 비롯하는 잔여 한계는 5.4절에서 논의한다.

자료 수집의 원칙은 다중 출처 삼각측량(triangulation)이다. 성격이 다른 여러 자료 출처에서 같은 사실을 교차 확인하여, 단일 출처의 편향이나 결손이 결론을 좌우하지 않도록 한다. 삼각측량은 자료원·조사자·이론·방법의 네 유형으로 구분되며(Patton, 2015; Yin, 2018), 모두 충족해야 하는 요건이 아니라 실제로 존재하는 위협에 대응하는 유형을 선택적으로 적용한다(Yin, 2018). 연구가 통제해야 할 위협은 단일 자료 출처에 의존할 때 발생하는 편향과, 내부자 연구자의 단일 관점에서 비롯하는 해석 편향이다. 이에 대응하여 Yin(2018)이 구성 타당성 확보의 핵심 수단으로 권고한 자료원 삼각측량(data triangulation)을 적용하고, 거버넌스·운영 사실에 대한 양사 확인 절차를 함께 둔다. 자료원 삼각측량은 프로젝트 산출물 문서, 시스템 산출물, 거버넌스 문서, 연구자 직접 관여의 네 출처에서 동일한 사례 사실을 교차 확인한다.

가령 명제 P-1이 다루는 모델 성능은 모델 개발 보고서(프로젝트 산출물)와 모델 학습 로그·결과 데이터셋(시스템 산출물)에서 생성 주체와 형식이 다른 두 출처가 같은 사실을 가리키는지 독립적으로 대조된다. 거버넌스·운영 사실에 대한 양사 확인은 정식 계약 체결, 데이터 제공 범위, 예측 모델 결과의 A사 시스템 반영, 상시 운영 진입 등 양사가 공동 관여한 사실을 대상으로 2025년 양사 실무 합의·보고 절차를 거쳐 A사와 B사 두 주체로부터 받

았다. 모델 성능과 분석 결과는 B사가 기술적으로 검토하였으며, A사 실무 책임자도 B사의 설명을 토대로 분석 결과가 운영 현실에 부합하는지를 확인하였다. 출처 간 해석 차이는 양사 협의 과정에서 조정되었으며, 연구 수행 범위 안에서 실질적 모순은 확인되지 않았다. 네 자료 출처가 명제 P-1·P-2·P-3 가운데 어느 측면을 지지하는지는 〈표 2〉와 같다.

연구 사례 회사와 내부 시스템 명칭에는 익명화 원칙이 적용된다. 양사의 실명은 본문 어디에도 노출되지 않으며, 운영사는 A사, 데이터 전문 기업은 B사로 표기한다. B사가 운영하는 표준 상품 마스터 데이터 체계는 일반 술어인 통합 상품 마스터(Product Category Master, 이하 PCM)로, B사의 빅데이터 적재 환경은 통합 빅데이터 적재 시스템으로, 분석 도구는 표준적 SQL·Python 분석 환경으로 추상화한다. 모델 성능 지표·분석 규모 등 학술적 평가에 필요한 수치는 본문에 공개하고, 기업 운영이 직접 노출되는 비교 수치는 제외하였다. 학술적 가치 평가와 기업 정보 보호의 균형을 추구한 결과이며, 익명화로 인한 검증 한계는 5.4절에서 명시한다(Yin, 2018).

3.3 인과 추론 방법

사례 연구에서 명제는 자료 수집과 분석의 방향을 설정하는 핵심 장치이다(Yin, 2018). 명제 검증에 앞서 그 이론적 위치를 확인한다. Gregor(2006, pp.625-626)에 따르면 인과적 설명과 예측을 모두 제공하는 이론은 설명·예측형(Type IV)으로 분류되며, 검증 가능한 명제(testable propositions)를 통해 예측 차원을 실현한다. 연구의 세 명제는 각각 데이터 비경합성이 이중 가치 발생으로 이어지는 인과 관계(P-1), 다단계 제도화가 자산화 정착으로 이어지는 인과 관계(P-2), 두 차원 데이터 마찰의 비대칭이 정착 경로 분기로 이어지는 인과 관계(P-3)를 진술하며, 2.4절에서 사전 명문화되어 4장

〈표 2〉 자료원별 수집 내역과 관련 명제

출처	구체적 자료	수집 시기	관련 명제
(1) 프로젝트 산출물 문서	모델 개발 보고서 2건, 데이터 사전, 변수 목록, 모델 학습·평가 결과표, 라벨링 회차별 정확도 추이표	2025년 5~11월(공식적 분석, 보고서 완료 2회: 8월/11월)	P-1(가치 발생) P-3(정착 경로)
(2) 시스템 산출물	모델 학습 로그, 예측·분류 결과 데이터셋, 변수 중요도 분석 결과, 카테고리 체계 매핑 데이터셋	2025년 5~11월(예측 모델: 5~8월, 분류 모델: 8~11월)	P-1(가치 발생) P-3(정착 경로)
(3) 거버넌스 문서	정식 계약서·부속 협약, 예측 모델 배치 운영 로그, 분류 모델 정제 방안 검토 문서	2025년 5월 정식 계약, 2026년 5월 현재 예측 모델 시스템 수정 완료 및 분류 모델 정제 방안 검토 중	P-2(가치 정착) P-3(정착 경로)
(4) 연구자 직접 관여	내부자인 연구자 기획·총괄, A사·B사 실무 협의·회의, 예측 모델 결과의 A사 시스템 반영 확인	2024년 9월~2026년 5월 현재	명제 전반 (맥락 이해·해석)

에서 사례 증거에 대조 검증된다. 따라서 세 명제는 모두 설명・예측형(Type IV)에 해당한다.

과정 추적(process tracing)은 인과 관계가 실제 사례 안에서 어떻게 전개되었는지를 사례 내부 증거를 통해 추적하는 사례 내 분석 방법이다(Beach and Pedersen, 2019). 인과 화살표의 양 끝(원인과 결과)에 머무르지 않고 그 사이의 작동 방식을 열어 보이는 데 분석의 초점이 있다. Beach and Pedersen (2019)은 과정 추적을 이론 검증, 이론 구축 및 결과 설명의 세 유형으로 구분하였다. 본 연구는 이 가운데 이론 검증 과정 추적(theory-testing process tracing, 이하 TT-PT)을 채택한다.

이론 검증 과정 추적(theory-testing process tracing)은 사전에 이론화된 인과 관계가 사례 안에서 검증 가능한 관찰 가능한 증거(observable manifestations)를 남겼는지를 점검하여 명제를 검증한다. 명제 P-1・P-2・P-3은 2장의 이론적 배경으로부터 2.4절에서 도출되어 사전 진술된 이론적 명제이며, 4장의 사례 분석은 사후적 명명(post-hoc labeling)이 아니라 명제에 대한 대조 검증 과정에 해당한다.

TT-PT는 IS 분야의 비판적 실재론(critical realism) 사례연구 절차와도 부합한다. Wynn and Williams (2012)는 비판적 실재론 기반 사례연구의 방법론적 원칙을 다섯 단계(사건 명료화, 구조・맥락 명료화, 소급추론, 경험적 입증, 삼각측량 및 다중 방법)로 정리하였다. 다섯 원칙 가운데 4단계 경험적 입증(empirical corroboration)은 식별된 인과 관계가 사례 자료에 의해 지지되며 대안 설명보다 우월한 설명력을 가지는지를 단일 사례 안에서 체계적으로 평가하는 작업이다. 연구의 핵심 과제는 단일 사례 안에서 경험적 입증을 어떻게 체계적으로 수행할 것인가에 있으며, Van Evera(1997), Collier(2011) 및 Mahoney(2012)의 인과 추론 테스트가 그 절차적 도구를 제공한다.

Van Evera(1997)는 단일 사례에서 가설을 검증하는 네 가지 유형의 증거 테스트를 정리하였고, Collier (2011)와 Mahoney(2012)가 이를 필요성(necessity)과 충분성(sufficiency)의 2×2 논리로 정리하였다. 인과 추론 테스트는 다음과 같이 (a) ~(d) 네 형태로 구분되며, 각각 명제 검증에서 서로 다른 역할을 수행한다.

(a) 방향 지시 테스트(straw-in-the-wind test)는 명제와 정합적 방향성을 시사하는 보조 증거로, 통과・실패 어느 쪽도 결정적이지 않다. (b) 통과 테스트(hoop test)는 필요조건을 점검하여, 실패 시 명제를 기각하나 통과만으로 충분 입증은 아니다. (c) 결정적 단서 테스트(smoking gun test)는 다른 가설로 설명하기 어려운 충분 단서로, 통과 시 명제를 강하게 입증하나 실패가 기각을 뜻하지는 않는다. (d) 이중 결정 테스트(doubly decisive test)는 필요성과 충분성을 동시에 충족하여 통과 시 결정적 입증, 실패 시 결정적 기각을 모두 가능하게 한다. 사회과학에서 실제로 통과 가능한 이중 결정 테스트를 구성하기는 매우 어렵다는 것이 일반적 평가이다(Collier, 2011; Mahoney, 2012). 그 대안으로, 통과 테스트(필요성 점검)와 결정적 단서 테스트(충분성 점검)를 함께 통과하는 방식이 사용된다. 두 테스트가 모두 통과하면 이중 결정에 준하는 추론력을 확보하므로, 사회과학의 실제 연구에서는 가장 널리 활용되는 검증 조합으로 간주된다.

단일 사례의 명제 검증은 기존 문헌이 제시한 이론의 적용 한계를 단일 사례 증거로 드러내고 그로부터 더 나은 일반 이론을 추론하는 형식으로 수행된다. 예를 들어, Yin(2018, p.51)이 모범 사례로 제시한 Gross et al.(1971)의 단일 학교 사례연구는

〈표 3〉 명제별 인과 추론의 검증 논리

구분	P-1(가치 발생)	P-2(가치 정착)	P-3(정착 경로)
인과 주장	▪ 데이터의 비경합성으로 인해 동일한 통합 데이터 아키텍처가 두 방향으로 작동하여 효율화 지향 가치(예측 모델)와 발견 지향 가치(분류 모델)가 동시에 발생	▪ 다단계 제도화(탐색적 분석, 정식 계약, 공식적 분석, 상시 배치 운영, 시스템 수정)를 통해 일회성 가치가 지속 가능한 데이터 자산으로 정착	▪ 양사가 획득한 가치의 정착 경로는 데이터 분석 차원과 제도적 환경 차원의 데이터 마찰에 따라 분기
기존 문헌	▪ 데이터 자원의 비경합성 논리 ① 동일 데이터의 여러 기업에 사용은 사회적 이익(수확체증) (Jones and Tonetti, 2020) : 소유권, 프라이버시, 후생의 관점 ② 비경합성의 이면: 전략적 대체재 조건(경쟁관계)에서 데이터 공유 기피 딜레마 (Wang et al., 2025; Liu et al., 2025)	▪ 데이터 자산화 및 제도화 ① 기업 내부 차원의 자산화 단계 (Xu et al., 2024) ② 사용자(users) 자산화·계약 기반 정착 (Birch et al., 2021) ③ 제도화 3단계 (Tolbert and Zucker, 1996; Rageth et al., 2021) ④ 조직 간 데이터 거버넌스 (Robey et al., 2008; Jussen et al., 2024; Abraham et al., 2019)	▪ 데이터 마찰 ① 메타데이터·목적 차원의 거리 관련 분석 차원 마찰(Edwards et al., 2011; Borgman and Groth, 2025) ② 권력·이해관계·규범이 충돌하는 규제 현상(Bates, 2018) ③ 데이터 실천과 제도적 차원의 상호 연관(Aula, 2019)
이론적 확장 (critical case test)	▪ 비경합성을 소유권·이익 배분 또는 경쟁자 간 데이터 공유의 관점에서 논의 → 보완 관계의 조직 간 이질 자원(데이터·분석 역량) 결합 조건으로 적용 범위를 재해석	▪ 자산화를 기업 내부 차원의 단계적 진화 또는 거버넌스 문제로 분석 → 복수 기업 간 협력의 맥락에서 다단계 제도화(탐색적 분석, 정식 계약, 공식적 분석, 상시 운영, 시스템 수정) 과정으로 재구성	▪ 데이터 마찰을 상호 연관된 현상으로 다룸 → 발생 원천에 따라 데이터 분석 차원과 제도적 환경 차원으로 구분 → 발생 원천이 상이하다는 경계 조건에서, 한 차원의 해소가 다른 차원으로 이어지지 않아 두 마찰이 자산화 정착 경로를 분기
사례 증거	4.1절	4.2절	4.3절
	4.4절 〈표 9〉		

기존 문헌에서 제시된 혁신 저해 요인(barriers to innovation)만으로는 혁신 실패를 충분히 설명하기 어렵다는 점을 보여주며, 구현 과정(implementation process)에 주목하는 새로운 설명을 제시하였다. 본 연구는 명제 P-1·P-2·P-3에 대해 통과 테스트와 결정적 단서 테스트를 핵심 도구로 적용하고, 방향 지시 테스트를 보조 증거로 활용한다. 명제별 검증의 출발점은 각 명제가 다루는 기존 문헌의 적용 한계와 본 사례가 제시하는 이론적 확장의 방향이며, 〈표 3〉에 사전 정리한다. 4장의 사례 분석은 〈표 3〉의 사전 정리에 따라 각 테스트의 통과 여부를 점검하고, 명제별 평가 결과를 4장에서 제시할 〈표 9〉에 종합한다.

3.4 연구 절차의 종합 검토

3.1~3.3절에 분산된 연구 절차적 선택이 사례 연구의 엄정성(rigor)을 충족하는지를, Dubé and

〈표 4〉 사례 연구의 엄정성 확보를 위한 조치 사항

영역	평가 기준 (Criteria)	연구의 조치 사항	본문 위치
AREA 1 사례연구 설계 (Design of the case study)	연구 목적 (Research purpose)	▪ How/Why 인과 규명을 목적으로 하는 설명적(explanatory) 사례연구로 위치 지음	3.1절
	연구 문제 (Research questions)	▪ RQ-1~RQ-3을 How/Why 형태로 진술하고 명제 P-1~P-3과 대응시킴	1장, 2.4절
	단일·복수 사례 설계 (Single vs. multiple-case design)	▪ 단일 사례 설계 채택, 분석적 일반화를 목표로 명시	3.1절
	사례 선택 (Selection of cases)	▪ 결정적·계시적·특이한·종단적 사례의 네 근거로 단일 사례 선택을 정당화	3.1절 〈표 1〉
	분석 단위 (Unit of analysis)	▪ 조직 간 데이터 통합 협력을 분석 단위로, 예측·분류 모델을 하위 분석 단위(embedded units of analysis)로 설정	3.1절
	연구 맥락 (Research context)	▪ 2024년 하반기~2026년 5월의 종단적(longitudinal) 진행 사례임을 명시, 협력 단계별 시점 기술	3.1절
AREA 2 사례연구 수행 (Conduct of the case study)	자료 수집 과정의 명료화 (Elucidation of the data collection process)	▪ 협력 단계별 자료 수집 경로와 자료-명제 대응 규칙을 사전 명시	3.2절 〈표 2〉
	자료 수집 방법 (Data collection methods)	▪ 프로젝트 산출물, 시스템 산출물, 거버넌스 문서 및 연구자 직접 관여의 네 출처에서 수집	3.2절
	다중 자료 수집 방법의 활용 (Use of multiple data collection methods)	▪ 모델 성능·분류 정확도의 정량 자료와 거버넌스·분석 맥락의 정성 자료를 통합	3.2절
AREA 3 증거 분석 (Analysis of the case study evidence)	자료 분석 과정의 명료화 (Elucidation of the data analysis process)	▪ 명제 P-1~P-3을 예상 패턴으로 사전 설정한 패턴 매칭(pattern matching) 절차 명시와 이론 검증 과정 추적(TT-PT) 기반 인과 추론 테스트로 명제 검증 ▪ 설명 구축(explanation building)과 대안적 설명 검토로 인과 사슬 정교화	4장 5.1절
	분석 기법·도구 (Data analysis techniques and tools)	▪ 코딩·자료 환원(coding): 예측 모델(AutoML)의 파생 변수 생성/ 필터링/학습, 분류 모델(KoRoBERTa)의 비정형 텍스트(송장명) 라벨 코딩/학습 ▪ 시각적 표현(visual displays): 각종 보고서·문서의 집계 결과를 그림과 표로 제시	4장
AREA 4 엄정성 강화 전략 (Strategies for enhancing case study rigor)	인용 (Quotes)	▪ 거버넌스 문서(정식 계약서·협약), 모델 보고서의 원문 진술과 회의록·면담 기록의 진술을 인용하여 익명화 처리 후 사례 증거로 반영	4장
	삼각측량 (Triangulation)	▪ 네 출처(프로젝트 산출물, 시스템 산출물, 거버넌스 문서, 연구자 직접 관여)에 대한 자료원 삼각측량(data triangulation) 적용	3.2절
	참여자 검토 (Subject review)	▪ 거버넌스·운영 사실을 양사 정보 제공자가 교차 확인; 모델 성능·분석 결과는 B사가 기술적으로 검토하고 A사 실무 책임자가 운영 타당성 관점에서 확인	3.2절
	사례연구 프로토콜 (Case study protocol)	▪ 연구 절차의 반복 가능성과 신뢰성을 확보하기 위해, 자료 출처·수집 시기·명제 대응 규칙을 사전에 명시한 사례연구 절차서 작성	3.2절 〈표 2〉
	사례연구 데이터베이스 (Case study database)	▪ 모델 보고서, 학습 로그, 결과 데이터셋, 거버넌스 문서 등 문서·시스템 산출물의 보관	4장

〈표 4〉 사례 연구의 엄정성 확보를 위한 조치 사항(계속)

영역	평가 기준 (Criteria)	연구의 조치 사항	본문 위치
AREA 5 연구의 한계와 관련성 (Limits and relevance of the case study)	방법론적 한계 (Methodological limitations)	▪ 내부자 연구자 위치성(positionality), 익명화에 따른 외부 검증 제약, 단일 사례의 통계적 일반화 한계 명시	5.4절
	연구자에 대한 제언 (Recommendations for researchers)	▪ 문자적·이론적 복제(literal·theoretical replication)와 정착 경로의 종단적 추적을 향후 과제로 제시	5.4절
	실무자에 대한 함의 (Implications for practitioners)	▪ 사업 가치가 확인된 이후에도 제도적 환경 마찰이 해소되지 않으면 자산화가 지연될 수 있다는 시사점을 제시	5.3절

Paré(2003)가 제시한 다섯 영역(사례연구 설계, 수행, 증거 분석, 엄정성 강화 전략, 연구의 한계와 관련성)을 기준으로 검토한다. 다섯 영역은 설계의 정합성에서 자료의 신뢰성, 분석의 투명성, 한계의 명시에 이르는 연구 수행의 전 과정을 포괄한다. 〈표 4〉는 각 영역의 평가 기준과 연구가 취한 조치를 대응시켜, 앞 절들에 흩어진 방법론적 선택이 하나의 평가 체계 안에서 어떻게 맞물리는지를 종합한다.

Ⅳ. 분석 결과

4.1 예측·분류 모델과 이중 가치

A사는 송장·집하·배송 데이터와 물류 운영 도메인 지식을 보유하고 있지만, 분석·모델 개발 역량과 전담 인력, 수요 형성 단계의 구매·결제 데이터, 표준 상품 마스터 체계는 갖추지 못하였다. B사는 연 약 2,000만 명 이용 회원 규모의 유통 구매·결제 데이터, 통합 상품 마스터(PCM), 빅데이터 분석 역량을 보유한다. 반면 B사는 외부 온라인 기반 거래 데이터, 비정형 운영 데이터, 물류 운영 도메인 지식은 갖추지 못하였다. B사의 빅데이터 분석 역량은 BDAC의 세 차원(Gupta and George, 2016)으로 구체화된다. 유형 자원으로는 분석용 정보계, 대규모 데이터를 적재하는 통합 빅데이터 적재 시스템, 표준적 SQL·Python 분석 환경이 갖추어져 있다. 인적 자원으로는 인공지능 분석 전문가 조직이 운영된다. 무형 자원으로는 산출물·코드 공유 체계와 지식 공유 활동이 축적되어 있다. B사는 BDAC의 세 차원을 모두 충족한다. 이와 달리, A사는 BDAC 세 차원이 상대적으로 결핍되어, 빅데이터 분석 역량의 비대칭이 자원 보완성의 한 축을 이룬다.

양사의 협력은 이 보완 구조를 서로 확인하는 데서 출발하였다. A사는 송장 데이터와 추가 필요 데이터를 자사의 자산으로 인식하였고, 양사는 B사가 물동량 예측 모델의 분석 가치를 먼저 증명하면 A사가 송장을 포함한 추가 필요 데이터를 제공·공유한다는 조건으로 협력에 합의하였다. 이 조건에 따라 물동량 예측 모델이 탐색적 분석으로 선행되었고, 분류 모델과 배차량 예측 모델은 정식 계약 이후 공식적 분석 단계에서 착수되었다. 양사의 자원 구조는 〈표 5〉와 같다.

양사 데이터는 예측 모델과 분류 모델 각각의 입

〈표 5〉 조직 간 자원 보완성

구분	A사 (물류 운영사)	B사 (데이터 전문 기업)
보유 자원	▪ 송장·집하·배송 운영 데이터 ▪ 물류 운영 도메인 지식	▪ 유통사 구매·결제 데이터, 상품 마스터(PCM) ▪ 빅데이터 분석 역량(BDAC 세 가지 차원)
결핍 자원	▪ 분석 역량·전담 인력 ▪ 수요 선행 데이터, 표준 마스터 데이터	▪ 외부 온라인 기반 대규모 거래 데이터 ▪ 비정형 운영 데이터, 물류 운영 도메인 지식
결합 효과	▪ 예측 모델을 통한 효율화 지향 가치	▪ 분류 모델을 통한 발견 지향 가치

력 구조로 결합된다. 예측 모델은 A사의 원천 변수(점소·네트워크·집하 물동량 집계)와 시계열·날짜·노선 파생 변수, B사의 결합 변수(유통사 구매·날씨 데이터)로 이루어진 통합 변수 마트를 입력으로 받는다. 날짜·날씨 관련 파생 연산은 D-7~D+7일 구간의 평균·변화, 달력 단위(일·주차·분기·절기·공휴일), 기온·습도·예보 변수를 포함한다. 예측 모델 개발 과정에서 A사 데이터 변수 274개(원천 22개, 파생 252개)에서 B사 데이터 변수 133개를 더한 총 407개 변수로 기초 데이터 집합을 구성하였다.

분류 모델은 송장 데이터(15개 변수의 정형·비정형 혼합)에서 송장명(상품명 필드의 비정형 텍스트)만 입력으로 받아 표준 카테고리로 분류한다. A사 송장 데이터는 일평균 약 170만 건, 월평균 약 4,800만 건 규모로 누적된다. 사례에서 또 하나의 중요한 자산은 데이터 자체가 아니라 B사가 보유한 통합 상품 마스터(PCM)다. PCM의 카테고리는 모든 채널에서 동일 상품에 동일 카테고리가 부여되도록 합의된 표준으로, 외부에서 역설계하기 어렵다. 다채널 유통의 운영 과정에서 누적된 수십만 건 규모의 상품명-카테고리 매핑이 PCM 위에서 1차 학습 라벨 자산으로 기능한다. 이로써 비정형 텍스트 분류에서 가장 큰 비용·시간 장벽인 대규모 라벨링 데이터 구축 부담이 상당히 해소된다. 양사 데이터 결합에서는 식별자 직접 매핑이 어려워, 주소 변수 등을 마스킹한 결합 키를 활용하는 간접 매핑과 지역·시간대·노선·카테고리 같은 상위 집계 단위의 통계적 결합이 병행되었다. PCM은 상품 마스터 표준이자 분류 모델의 학습 라벨로서 데이터 결합과 라벨링의 두 국면에서 매개 역할을 한다.

통합 데이터 아키텍처가 두 방향(예측·분류 모델)으로 작동하여 양사에 서로 다른 가치를 동시에 산출한다. 효율화 지향 가치는 집계·구조화 시계열 데이터를 입력으로 받는 예측 모델로 실현된다. 발견 지향 가치는 비정형 텍스트 데이터를 입력으로 받는 분류 모델로 실현된다.

예측 모델과 효율화 지향 가치. 협력 이전 A사는 물동량 예측을 실무자의 직관과 경험에 의존하였다. 자체 데이터만 쓴 예측은 두 가지 한계를 안고 있었다. 운영 시계열을 학습 가능한 변수로 가공할 분석 역량·인프라가 부재하였으며, 외부 시장의 수요 형성 신호와 기상 정보 같은 보조 자원이 존재하지 않았다. B사의 빅데이터 분석 역량은 A사 운영 데이터를 학습 가능한 시계열 변수(시차·누적·이동평균 등)로 체계적으로 파생한다. 이와 더불어 자사 상품 구매·결제 데이터와 외부 기상 데이터를 보조 신호 변수로 결합하여 통합 변수 마트를 구성한다. B사 구매·결제 데이터가 제공하는 수요 형성

단계의 선행 신호는 Lee et al.(2000)의 수요 선행 신호 명제와 상응한다.

예측 모델은 물동량·배차량 각각의 통합 변수 마트를 입력으로 받으며, 물동량 예측 모델의 최종 변수는 164개, 배차량 예측 모델의 최종 변수는 171개다. 당초 예측 모델의 변수는 A사 변수 274개와 B사 결합 변수 133개를 더한 총 407개로 구성되었다. 물동량·배차량 두 예측 모델은 이 동일한 기초 변수 집합을 공유하며, 목표 변수별 중요도 기반 필터링으로 물동량 모델 변수는 137개, 배차량 모델 변수는 171개로 압축되었다. 이어서 성능 개선을 위한 조정 단계에서 오차 분석을 기반으로 추가 변수의 투입을 시험하였다. 성능 향상이 확인된 물동량 모델은 최종 164개 변수로 확정하고, 오히려 성능 하락이 확인된 배차량 모델은 추가 없이 기존 변수 171개로 최종 확정하였다. 학습에 사용된 데이터의 기간은 2023년 1월 1일부터 2025년 4월 16일까지이다. 전체 데이터는 학습(training) 80%·검증(validation) 10%·테스트(test) 10%로 분할하였으며, 테스트 데이터는 학습에 사용하지 않고 성능 평가에만 사용하였다. 예측 알고리즘은 자동화 머신러닝(AutoML) 도구인 AutoGluon을 사용하여 RandomForest, ExtraTrees, XGBoost, LightGBM 등 후보군을 동일 학습 데이터 위에서 자동 학습·평가·앙상블하였다. 핀볼 손실(pinball loss)이 주 평가지표, 평균절대오차(MAE)가 보조 지표로 사용되었다.

예측 정확도의 산정 방식은 다음과 같다. 본문과 〈표 6〉의 예측 정확도는 1 − (MAE/실측 평균)으로 산출하였다. 이 방식은 통계 지표인 MAE를 현업이 직관적으로 이해할 수 있는 정확도 형태로 변환한 실무적 산정 방식이다. MAPE는 실측치가 0이거나 0에 근접한 관측 단위에서 계산이 불가능하거나 오차율이 과대 계상되어 사용하지 않았다. 예를 들어, 일별 하위 노선 기준으로 접수 건수(물동량)나 배차 대수(배차량)가 0이거나 한 자리인 경우가 다수 존재한다. 특히 배차량은 소규모 배차가 많아, 배차 1대를 예측하였는데 실제 3대인 경우 절대 오차는 2대에 그치지만 건별 오차율은 약 66.7%로 계상되는 왜곡이 발생한다. 이에 따라 오차율을 건별로 계산하여 평균하는 방식이 아니라, 전체 오차의 평균(MAE)을 실측 평균에 대비하는 방식을 채택하였다.

핀볼 손실에서는 과소·과대 추정에 서로 다른 벌점을 부과하는 기준값인 분위값(τ)을 사전에 설정한다. 배차 누락 시 추가 차량을 긴급 동원하는 비용이 차량 유휴 비용보다 크다는 운영 비용 구조를 반영하여, 과소 추정의 벌점에 20%를 가산하는 벌점 비 1.2 대 1을 적용하였다. 두 비용의 상대 비율로 이 기준값을 정하는 일반적 방식에 따라 $\tau = 1.2/(1.2+1) \approx 0.545$가 된다. 그 타당성은 A사 실무진과의 비즈니스 해석과 시뮬레이션 검증으로 확인되었다. A사는 배차 부족이 배송 지연과 긴급 배차 비용으로 직결되는 과소 추정을 비즈니스 관리 대상으로 삼는다. 이 설정에서 핀볼 손실은 예측 부족에서 발생하는 고비용 긴급 배차의 위험을 우선 줄이며, A사의 운영 요구와 부합한다.

평가 결과는 〈표 6〉과 같다. 협력 이전 A사의 수기 예측 정확도는 물동량 70% 미만, 배차량 50% 미만이었다. 공식적 분석에서 전체 평균 정확도는 물동량 예측 88.7%, 배차량 예측 86.3%로 나타나, 기존 수기 예측보다 각각 18.7%p 이상과 36.3%p 이상 개선되었다. 부분 표본에서도 대체로 안정적인 성능을 보였으나, D-1일이 휴일인 구간에서는 배차량 예측 정확도가 79.9%로 낮아져 향후 보완 과제로 확인되었다.

〈표 6〉 예측 모델의 평가 결과

지표		물동량	배차량	평가
예측 정확도(%)	전체	88.7	86.3	▪ 양사 통합 데이터를 활용한 공식적 분석의 성과 지표 ▪ A사의 수기 예측 정확도: 물동량 70% 미만, 배차량 50% 미만 – 물동량 예측 모델 성과의 단계별 비교 (〈표 7〉 참조) ▪ 양사 협력의 직접적 정당성을 제공
	D-1일 ≠ 휴일	89.5	86.7	
	D-1일 = 휴일	86.4	79.9	
핀볼 손실 (Pinball Loss, 평가지표)		202.8	0.094	▪ 과대·과소 예측에 비대칭 가중치를 부여하는 지표 ▪ 운영 비대칭(배차 누락 비용 〉 차량 유휴 비용)을 손실함수에 반영
평균절대오차 (MAE, 보조 지표)		404.5	0.176	▪ 평균 오차 수준을 정량화하는 보조 지표

주) 측정 단위는 물동량이 처리 건수(건), 배차량이 배차 차량 수(대)이며, 예측과 평가의 관측 단위는 일별 하위 노선이다.

AutoGluon이 산출한 최종 예측치는 트리 기반 다섯 모델의 가중 평균으로, 결합 비중은 RandomForest 두 변형이 각각 0.375와 0.25, ExtraTrees 0.1875, XGBoost 0.125, LightGBM 0.0625로 부여되었다. 변수 중요도 분석에서 두 모델 모두 직전 1~3주 시계열 파생 변수가 양의 기여 상위를 차지하였고, 공휴일 관련 변수는 주된 음의 기여를 보였다. 모델 성능의 1차 동력은 운영 시계열 변수의 체계적 파생이며, B사에서 결합된 유통사 매출·포인트, 날씨, 휴일·날짜 변수가 보조 신호로 작동한다.

예측 모델의 효율화 지향 가치(efficiency-oriented value)는 연간 수억 원의 배차 비용 절감에서 발생한다. 데이터 기준일에 누적 데이터를 받아 D+3일(3일 전)의 물동량을 예측하므로, 배차 계획을 3일 앞서 최적화할 수 있고, 예측 실패 시 추가 차량을 즉시 동원하는 고비용 운영도 예방할 수 있다. 모델 도입 전 기준 배차 비용 대비 10~20% 이상의 절감이 추산되어, 양사 협력의 직접적 정당성을 제공한다.

배차량 예측 모델은 별도의 탐색적 분석(PoC)을 거치지 않고 공식적 분석 단계에서 처음 수행되었다. 양사의 협력이 물동량 예측 모델의 가치를 먼저 확인하는 것을 출발 조건으로 삼았고, 배차량 예측은 물동량 예측 결과를 기초로 하는 분석이었기 때문이다. 이에 따라 물동량 예측의 성과가 확인되고 정식 계약이 체결된 이후 배차량 예측 모델의 개발이 시작되었다. 협력 이전 50% 미만이었던 배차량 수기 예측 정확도가 공식적 분석에서 86.3%로 개선된 결과는, 물동량 예측이 정확해지면 배차량 예측도 개선된다는 A사의 운영 진단과 정합적이다.

모델 성능 향상의 원인을 단계별로 확인하기 위해, 협력 이전 수기 예측부터 1·2차 탐색적 분석과 공식적 분석까지 비교 가능한 물동량 예측 모델을 살펴본다(〈표 7〉). 성능 향상의 동인은 알고리즘 정교화가 아니라 데이터 통합에 있다. 모델 선택과 앙상블을 자동화 도구에 맡겼음에도 물동량 예측에서 A사 단독 데이터 모형(1차 PoC 83.1%)보다 높은 정확도(88.7%)가 양사 통합 데이터에서 달성된 사실이 이를 뒷받침한다. 성과의 개선 폭은 단계별 비교로 확인되며, 물동량 예측 모델의 단계별 성과는 〈표 7〉과 같다. 2차 탐색적 분석의 개선(+4.9%p)은 B사 데이터 결합의 효과이고 공식적 분석의 개선

〈표 7〉 물동량 예측 모델 성과의 단계별 비교

구분	모델 투입 변수	분석 모델	물동량 예측률	비고
협력 이전	-	수기 예측	70% 미만	• A사 수기 예측 기준
1차 PoC (24년 11월 수행)	A사 변수 274개 (원천 22개, 파생 252개)	LightGBM	83.1%	• A사 변수 단독 모형 • B사 수행 +13.1%p↑
2차 PoC (25년 1월 수행)	물동량 예측 최종 변수 164개 (A사와 B사의 변수를 결합 후, 중요도 기반 필터링 및 추가 변수 조정)	LightGBM	88.0%	• 데이터 결합 효과 +4.9%p • D+1(1일 전) 예측
공식적 분석 (25년 5~8월 수행)		AutoGluon	88.7%	• 알고리즘 효과 +0.7%p • D+3(3일 전) 예측

(+0.7%p)은 알고리즘 고도화의 효과로, 앞서 제시한 데이터 통합의 기여를 단계별로 수치로 구체화한다. 공식적 분석은 하루 전 예측(D+1)에서 사흘 전 예측(D+3)으로 조건을 강화하면서도 정확도를 개선하였다. 이와 별도로, 1차 탐색적 분석(PoC)에서 시계열 특화 알고리즘(TFT, N-Beats, TSMixer)과 LightGBM을 예측 정확도 기준으로 비교하여 LightGBM으로 선정하였으며, 이는 이론적 우수성의 통계적 검증이 아니라 실무 비교에 따른 선택이다.

이상의 단계별 성과는 통제된 실험이 아니라 실제 운영 환경에서 확인된 것으로, 3장에서 밝힌 자연주의적 평가(Venable et al., 2016)의 관점에서 조직 간 데이터 결합이 만들어 낸 예측 가치의 실무적 증명에 해당한다.

분류 모델과 발견 지향 가치. 분류 모델은 KoRoBERTa(RoBERTa를 한국어 도메인에 적응한 변형)를 기반으로 하며, 분류 가능성 판별(1단계)과 카테고리 부여(2단계)의 2단계로 설계되었다. 알고리즘의 학술적 근거는 트랜스포머 기반 사전학습 언어모델(Devlin et al., 2019; Liu et al., 2019)에 두었으며, 한국어 처리에는 공개 사전학습 모델 KoRoBERTa를 활용하였다. 모델 구현에는 Hugging Face Transformers 라이브러리(Wolf et al., 2020)를 사용하였다.

먼저, 1단계 분류가능성 모델은 송장명의 분류 가능 여부를 판별한다. 송장 약 22.7%는 분석 목적으로 활용하기 어려운 형태로 기록되어 있어, 1단계 모델의 과제는 나머지 분류 가능한 약 77.3%를 가려내는 데 있다. 2025년 8월 접수 송장 기준으로 학습 데이터를 회차별 약 1,000건 단위로 누적 확장하면서, 모델 예측의 불확실성이 높은 표본을 우선 샘플링하는 능동 학습(active learning) 전략으로 모델을 훈련하였다. 회차별 누적 정확도는 1차 89.9%에서 5차 96.2%로 향상된 뒤 6차에서 95.5%로 소폭 하락하였다. 추가 라벨링의 한계 효용이 사라지는 약 5,000건 누적 시점에서 자연 수렴되어 5차 모델이 최종 선정되었다. 5차 분류 가능 판별 모델의 정확도(Accuracy)는 96.2%, 정밀도(Precision)는 94.0%이며, 최종 정분류율 산식에는 후자가 투입된다.

2단계 카테고리 모델은 분류 가능으로 판별된 송장에 PCM 중분류 카테고리를 부여한다. PCM 원본 중분류에서 택배업에 등장하기 어려운 카테고리를 제거하고 불명확한 카테고리를 통합한 결과, 277개 카테고리가 최종 학습 라벨로 활용되었다.

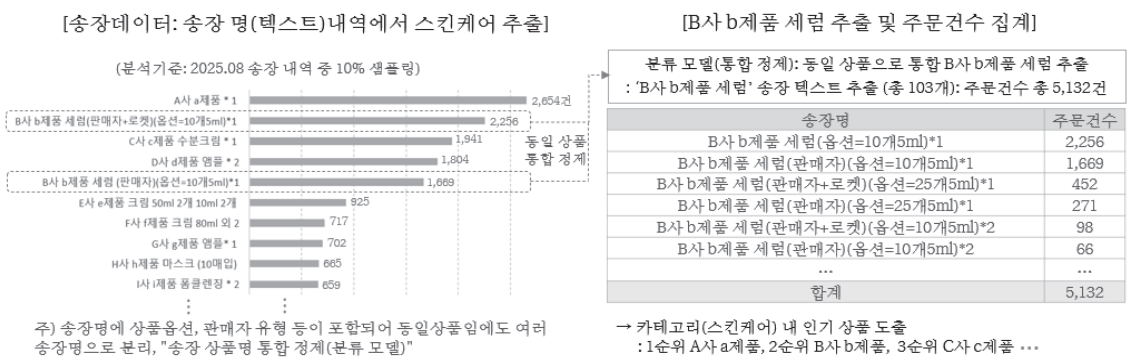
PCM의 상품명 - 카테고리 매핑을 1차 학습 라벨로 사용하였다. 이어 A사 송장명의 도메인 적합성을 보정하기 위해, 카테고리별 대표성이 높은 송장 표본을 정제·선별하여 2차 학습 데이터로 보강하였다. 사전학습 가중치 동결, R-Drop 정칙화, Symmetric Cross-Entropy 손실, 검증 데이터 분리 여부를 조합한 학습 전략 변형 7종을 비교한 결과, 검증 데이터를 별도로 분리하지 않고 가중치 동결에 R-Drop을 결합한 구성이 76.5%로 가장 높아 이를 최종 모델로 선정하였다. 협력 이전 A사는 14개 분류의 자체 송장 카테고리 체계를 운영하였으며, B사의 실무 검토 기준 분류 정확도는 약 66% 수준이었다(분류 불가능 데이터 22.7% 제외 기준). 14개 분류의 해상도로는 상품 구성의 실질적 구분이 어려웠다. 분류 모델은 14개 해상도로는 불가능했던 상품 구성의 구분을 277개 카테고리 해상도에서 76.5% 카테고리 정확도로 달성하여, 외부 시장 관찰에 앞서 A사 내부 데이터의 해상도와 신뢰도를 직접 개선하였다.

두 단계를 결합한 전체 표본 대비 최종 정분류율은 약 55.6%이며, 분류 가능 비율, 분류 가능 판별 정밀도, 카테고리 정확도의 곱으로 산출된다. 산식은 (1 - 22.7%) × 94.0% × 76.5% ≈ 55.6%이다. 다만 전체 정분류율은 발견 지향 가치의 수준을 그대로 반영하지 않는다. 1단계에서 분리되는 약 22.7%는 분석 영역 전반에 무작위로 흩어진 손실이 아니라, 외부 소비 시장 관찰이라는 분석 목적과 관련성이 낮은 거래에 체계적으로 편중되어 있기 때문이다(상세 근거는 4.3절). 시장 관찰의 핵심 거래는 분류 가능 영역에 남으므로, 발견 지향 가치의 평가에는 전체 정분류율(55.6%)과 분류 가능 영역의 카테고리 정확도(76.5%)를 구분하여 함께 보는 것이 적절하다. 전자는 분류 불가 영역까지 모수에 넣은 보수적 하한이고, 후자는 시장 관찰 영역의 실질 성능이다.

분류 결과의 발견 지향 가치(discovery-oriented value)는 외부 온라인 기반 대규모 거래 데이터를 파악할 가능성에서 발생한다. B사는 자사 통합 구매 데이터로 자사 거래 채널의 구매 패턴을 관찰할 수 있으나, 외부 온라인 기반 거래 시장은 직접 관찰하기 어렵다. A사의 송장 데이터는 외부 발송인의 발송 패턴을 포함하여, 분류된 송장 카테고리 분포가 외부 온라인 시장의 수요 분포를 파악하는 통로가 된다. 분석 결과 송장 카테고리는 식품(35%), 생활건강(24%), 패션(18%)의 상위 세 카테고리가 비교적 고른 분포를 보인다. 반면 B사 자사 거래 데이터는 온라인 분포에서 패션(45%)이 단연 높고, 오프라인 분포에서 식품(55%)이 압도적이다. 송장 카테고리 분포는 B사가 소유한 데이터의 어느 분포와도 달라, 자사 데이터만으로는 접근하기 어려운 외부 온라인 시장 전반의 관찰을 제공한다. 새로운 관찰 가능성은 지역별 온라인 상품 현황, 카테고리별 트렌드, 브랜드별 추세로 확장되며, B사 보유 오프라인 데이터와 결합되면 외부 온라인 시장과 오프라인 시장을 함께 조망하는 통합 분석 서비스 사업 기회로 이어진다.

B사는 분류 모델의 결과와 자사 데이터를 연계하여 거래처 현황, 지역별 물동·상품 현황, 터미널별 취급 품목 구성, 화물사고 유형 및 상품 트렌드 등의 활용 주제를 구체화하였다. A사는 해당 분석 결과를 물류 위탁사 등 협업업체에 제공하는 정보서비스로 활용하고자 하며, 이는 기존 고객 접점을 기반으로 서비스를 차별화하고 고객관계를 강화하는 가치에 해당한다.

〈그림 1〉은 발견 지향 가치가 실제 분석 산출물로 구체화되는 양상을 보여준다. 스킨케어 카테고리의 송장명별 주문 건수 상위 항목을 분석한 결과, 동일



〈그림 1〉 송장명 통합 정제 예시(인기 품목 도출)

한 상품이 판매자 유형, 용량 및 구성 옵션의 차이에 따라 100여 개의 서로 다른 송장명으로 분산되어 있어 원자료만으로는 실제 수요를 파악하기 어려웠다. 그러나 송장 상품명을 통합·정제하여 동일 상품의 주문 건수를 합산한 결과, 해당 상품의 실질적인 수요 규모와 카테고리 내 인기 순위가 드러났다.

명제 P-1의 패턴 매칭 결과는 다음과 같다. 명제 P-1은 이질적 자원의 결합과 데이터의 비경합성으로 인해 동일한 데이터 자원이 서로 다른 AI 기반 모델을 통해 복수의 가치로 전환되며, 그 가치가 양사에 차별적으로 귀속될 것이라는 패턴을 예측한다. 예측 모델은 A사 단독 수기 예측보다 높은 물동량 예측 정확도로 효율화 지향 가치를 산출하였고, 분류 모델은 B사 자사 데이터만으로는 관찰하기 어려운 외부 온라인 시장의 카테고리 분포를 가시화하여 발견 지향 가치를 산출하였다. 두 가치는 통합 데이터 기반에서 상호 배제 없이 실현되어, 서로 경쟁하거나 상쇄되지 않고 병존하였다. 효율화 지향 가치는 A사에, 발견 지향 가치는 B사에 귀속되었다. 따라서 관찰된 패턴은 예측된 패턴과 일치한다.

4.2 데이터 자산의 다단계 제도화

자산화의 전개는 Tolbert and Zucker(1996)가 제시한 제도화 3단계, 곧 습관화, 객체화, 침전화의 순차적 진행으로 해석된다. 2장에서 검토한 제도화 과정은 사례의 시간 축에 분석적으로 대응하며, 각 단계의 산출은 다음 단계를 촉발하는 인과 사슬을 이룬다. Rageth et al.(2021) 역시 제도화 정도를 전(前)제도화, 준제도화 및 완전 제도화의 세 수준으로 구분하여 각각 습관화, 객체화 및 침전화와 연결하였다. 본 연구는 데이터 자산화를 복수 기업 간 협력의 맥락에서, 문제 해결 행동이 형성되고 그 행동에 공유된 의미가 부여되며 운영 구조 안에 정착되는 과정으로 파악한다.

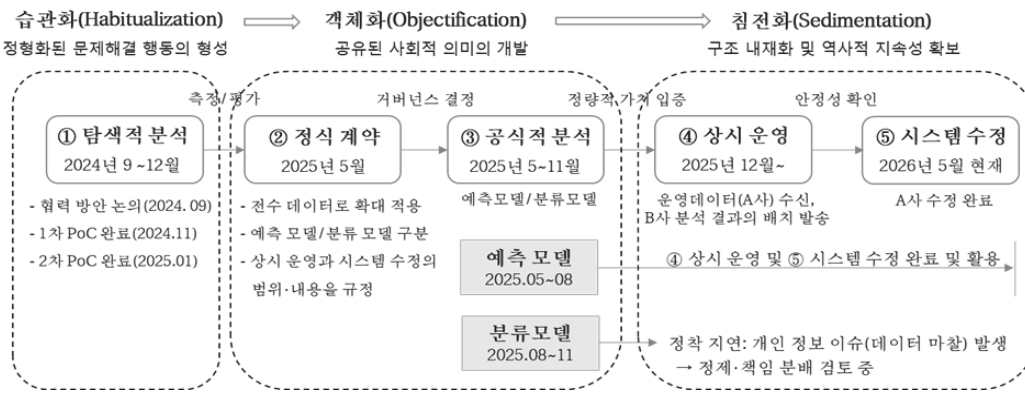
습관화 단계. 습관화는 조직 문제에 대응하여 정형화된 문제 해결 행동이 형성되고 그 행동이 특정 상황과 연결되는 전(前)제도화 단계다. 사례에서 습관화 단계는 2024년 9월 논의되어 2025년 1월까지 수행된 1·2차 탐색적 분석(PoC)으로 나타난다. A사의 물동량 예측 한계에 대응하여 B사의 분석 역량과 A사의 운영 데이터를 결합한 새로운 분석

실천이 형성되었다. 협력의 출발 조건에 따라 탐색적 분석은 물동량 예측 모델에 한정되었으며, 분류 모델과 배차량 예측 모델은 정식 계약 이후 공식적 분석 단계에서 도입되었다. 1차 PoC(2024년 11월 완료)는 파생변수 구성과 데이터 가공이 제한된 초기 상태에서도 물동량 예측률 83.1%를 달성하였다. 이후 양사 실무자 협의를 통해 A사의 운영 맥락과 도메인 지식이 시계열・날짜・노선 파생변수로 반영되면서, 2차 PoC(2025년 1월 완료)에서 물동량 예측률은 88.0%로 향상되었다. 이 향상은 단순한 성능 개선을 넘어, 데이터 원천이 보유한 맥락 지식이 분석 모델에 흡수되며 새로운 분석 실천이 정형화된 문제 해결 행동으로 발전한 과정이다. 정량 성과와 조직 내부의 긍정적 평가는 후속 단계의 공유된 의미 형성과 정식 계약을 촉발하였다.

객체화 단계. 객체화는 습관화 단계에서 형성된 행동에 일반적이고 공유된 사회적 의미가 부여되어, 조직 의사결정자들 사이에서 가치 있는 해결책으로 인식되는 준제도화 단계다. 객체화 단계는 탐색적 분석의 정량 성과가 조직의 공식 의사결정으로 전환되는 과정에서 관찰된다. 1・2차 PoC에서 입증된 물동량 예측 성과는 양사 경영진 보고와 평가로 이어졌고, 양사 데이터 협력의 가치가 조직 내부에서 공식적으로 인정되었다. 공식적 인정은 2025년 5월 정식 계약 체결로 공식화되었다. 정식 계약은 단순한 거래 약정을 넘어, 데이터의 소유・접근・활용을 조정하는 절차적 거버넌스 합의(Abraham et al., 2019)로 기능하였다. 정식 계약을 전후로 분석 기반은 샘플 데이터에서 전수 데이터로 확대되었으며, 이를 활용한 2025년 5~8월 공식적 분석에서 물동량 예측률은 88.7%에 도달하여 효율화 지향 가치를 입증하였다. 또한, 8~11월에는 분류 모델도 병행되어 발견 지향 가치를 입증하였다. 이로써 분석 실천은 특정 담당자의 비공식적 시도를 넘어 양사 의사결정자들이 가치 있는 해결책으로 공유하는 공식 실천으로 객체화되었다.

분류 모델과 배차량 예측 모델은 탐색적 분석 없이 공식적 분석 단계에서 직접 착수되었다. 조직 간의 협력이 B사가 물동량 예측 모델의 분석 가치를 먼저 증명하면 A사가 송장을 포함한 추가 필요 데이터를 제공・공유한다는 조건에서 출발하였기 때문이다. 분류 모델은 송장 분류에 기반한 시장 정보 서비스가 선도 물류기업을 중심으로 동종 시장에 이미 존재하여 실현 가능성과 사업 가치가 외부에서 입증되어 있었고, 배차량 예측 모델은 물동량 예측을 기초로 하는 분석이어서 가치 증명이 물동량 예측에 달려 있었다. 제도화 관점에서 물동량 예측 모델이 습관화, 객체화, 침전화의 전 단계를 순차적으로 거친 반면, 두 모델은 물동량 예측 모델의 성과로 형성된 협력에 대한 양사의 신뢰 위에서 객체화 단계에 직접 진입한 사례로 해석된다. 제도화 3단계는 모델마다 기계적으로 동일하게 적용되는 것이 아니라, 진입 조건의 차이에 따라 차별적으로 전개된다.

침전화 단계. 침전화는 구조가 역사적 연속성을 획득하고 조직 구성원의 교체를 넘어 지속됨으로써 완전한 제도화에 가까워지는 단계다. 침전화 단계는 분석 결과가 일회성 보고를 넘어 운영 인프라에 내재화되는 과정으로 확인된다. 예측 모델은 2025년 12월 상시 배치 운영에 진입하였고, 2026년 5월 현재 예측 모델의 시스템 수정을 통해 운영 인프라에 통합되어 활용되고 있다. 예측 모델은 특정 프로젝트의 산출물에 머무르지 않고 반복적 운영 의사결정과 시스템 구조 안에서 지속적으로 활용된다. 침전화가 기술적 배치의 완료와 운영 구조 내 지속성 확보를



〈그림 2〉 데이터 자산화 진행 과정(다단계 제도화)

뜻한다는 점에서, 예측 모델은 두 조건을 모두 충족하여 침전화 단계에 도달한 것으로 해석된다. 이와 달리, 분류 모델은 공식적 분석에서 가치를 입증하였음에도, 송장 텍스트의 개인정보 이슈에 따른 제도적 환경 차원의 데이터 마찰로 상시 운영 정착이 지연되고 있다. 동일한 통합 데이터 아키텍처 위에서 두 모델의 정착 경로가 분기하는 양상은 〈그림 2〉와 같다.

〈그림 2〉는 자산화 과정이 탐색적 분석(PoC), 정식 계약, 공식적 분석, 상시 운영 및 시스템 수정으로 이어지는 단계적 촉발 사슬로 전개되었음을 보여준다. 자산화는 단선적으로 완결되지 않고, 모델별 마찰 조건에 따라 서로 다른 정착 경로로 분기하였다.

명제 P-2의 패턴 매칭은 다음과 같다. 명제 P-2는 발생한 가치가 복수 기업 간 협력의 맥락에서 다단계 제도화를 거쳐 지속 가능한 자산으로 정착할 것이라는 패턴을 예측한다. 4.2절에서 관찰된 패턴은 Tolbert and Zucker(1996)의 제도화 3단계가 사례의 시간 축에 순차적으로 전개되는 양상으로 나타났다. Rageth et al.(2021)의 제도화 정도 구분에 비추어 보아도, 1・2차 PoC는 새로운 문제 해결 행동이 형성되는 전(前)제도화 수준에, 정식 계약과 공식적 분석은 공유된 의미와 사회적 합의를 획득하는 준제도화 수준에, 상시 배치 운영과 예측 모델의 시스템 수정은 지속성과 역사적 연속성을 확보해 가는 완전 제도화 방향의 진행에 해당한다. 특히 공식적 분석이 정식 계약 이후에 가능했다는 사실은 거버넌스 합의가 자산화의 공식화와 확장에 중요한 전환점으로 작용했음을 보여준다. 물동량과 배차량 예측 모델은 침전화까지 완결하여 다단계 제도화를 통한 자산 정착을 확인하였다. 반면 분류 모델은 객체화 이후 제도적 환경 차원 마찰로 침전화가 유보되었으며, 이 유보는 그 자체로 P-3(정착 경로 분기)을 지지하는 사례의 증거가 된다.

4.3 데이터 마찰과 정착 경로의 분기

데이터 2차 사용에서 발생하는 데이터 마찰은 데이터 분석 차원(data-analytic level)과 제도적 환경 차원(institutional-environmental level)으로 구분된다. 데이터 분석 차원의 마찰은 데이터의 정보적 속성과 생성 맥락에서 비롯되어 분석 모델의 성능과 활용 범위를 제약한다. 제도적 환경 차원의

〈표 8〉 데이터 마찰의 차별화된 정착 경로

구분	예측 모델 (효율화 지향 가치)	분류 모델 (발견 지향 가치)
데이터 유형	집계·구조화 시계열(물동량·배차량)	비정형 텍스트(송장명)
2차 활용 목적	물동량·배차량 예측을 통한 운영 효율화	송장명 분류를 통한 소비·시장 패턴 발견
데이터 분석 차원의 마찰 (분석 성과 제약)	약함	강함 → 원인 규명 후 분리 설계로 해소
	▪ 집계·구조화 데이터로 분석 투입 용이 ▪ 예측률: 물동량 88.7%, 배차량 86.3%	▪ 운영 목적과 분석 목적 불일치 -송장명 체계적 미기재·불완전 기재 등 22.7% ▪ 모델의 2단계 분리 설계 -최종 정분류율 55.6% ≈ 분류 가능 비율 77.3% × 분류 가능 판별 정밀도 94.0% × 카테고리 정확도 76.5%
제도적 환경 차원의 마찰 (상시 활용 제약)	약함	강함
	▪ 출력이 집계 수치 중심 ▪ 개인정보·법규 이슈 낮음	▪ 송장 텍스트에 개인정보 포함 - 이름·연락처·차량번호·현관 비밀번호 등 - 분류 가능 데이터(77.3%)의 약 3% 수준 ▪ 정제 책임 및 관리 방식 결정 필요
정착 경로 (2026.05 현재)	▪ 정착 완료: 예측 모델 시스템 수정 완료 및 활용	▪ 정착 지연: 개인정보 정제·책임 분배 검토 중

마찰은 데이터의 구성적 특성이 법·규제·거버넌스와 결합하여 분석 결과의 상시 활용 가능성을 제약한다. 예측 모델과 분류 모델은 동일한 정식 계약과 통합 데이터 아키텍처를 공유하였으나, 두 차원 마찰의 크기 차이에 따라 서로 다른 정착 경로를 보였다. 두 모델의 마찰과 정착 경로를 대비하면 〈표 8〉과 같다.

예측 모델. 예측 모델은 두 차원의 마찰이 모두 낮았다. 입력은 집계·구조화된 시계열 데이터이고, 출력은 물동량·배차량 예측치라는 수치 데이터다. 운영 데이터와 분석 목적 사이의 정보적 정합성이 양호하여 데이터 분석 차원 마찰이 분석 성과를 크게 제약하지 않았으며, 출력 역시 집계 수치이므로 개인정보나 법규 이슈를 수반하지 않았다. 예측 모델이 가치 입증 직후 상시 배치 운영으로 정착할 수 있었던 것은 이처럼 낮은 마찰 조건에 기인한다.

분류 모델. 분류 모델은 두 차원 모두에서 강한 마찰에 직면하였다. 먼저 데이터 분석 차원의 마찰은 송장명의 정보적 부적합으로 표면화되었다. 4.1절에서 분류 불가로 분리된 약 22.7%의 송장 데이터는 양사 공동 분류 모델 보고서에서 정보량 부족, 데이터 깨짐, 세부 정보 부족, 불필요 세부 사항만 기재의 네 유형으로 정리되었다. 정보량 부족은 'CA20(20EA)', '7215'처럼 의미를 식별하기 어려운 축약·코드형 기재와 '(NULL)' 등 정보 자체의 부재가 대표적이다. 데이터 깨짐은 인코딩 오류로 판독이 불가능한 텍스트이며, 세부 정보 부족은 '13.육류', '잡화, 의류'처럼 카테고리 판별에 필요한 상세가 없는 포괄 기재다. 불필요 세부 사항만 기재는 중요 상품 정보가 누락된 채 '배송 전 미리 연락해 주세요' 등과 같은 배송 메모나 옵션 정보만 기재된 경우다. 위탁자의 주된 관심은 배송이 정확히 도달하는 데 있을 뿐, 송장명을 표준 상품 분류 체계에

맞춰 기록할 유인은 거의 없다. 그 결과, 운영 목적과 분석 목적의 불일치는 송장 데이터를 카테고리 분류 데이터로 재사용하는 시점에 데이터 분석 차원의 마찰로 드러난다. 이 마찰은 Edwards(2010)의 데이터 마찰, Edwards et al.(2011)의 메타데이터 마찰, 그리고 Borgman and Groth(2025)의 목적 차원의 거리(distance in purposes)에 대응한다.

분류 가능 영역이 시장 관찰에 대표성을 가지는 이유는 송장명 생성 방식의 구조적 차이에 있다. 대형 유통 위탁사의 물량은 위탁사가 운영하는 상품 정보 체계에서 생성된 표준화된 상품명이 송장에 그대로 기재되어, 분류 모델이 판별할 수 있는 정보를 안정적으로 담는다. 이와 달리, 분류 불가 영역은 기업 간 거래, 본지점 간 배송, 중소 위탁사·개인 간 거래에서 발생하며, 상품 판별 정보를 담고 있지 않아 분류 목적으로는 활용 자체가 불가능한 데이터다. 즉 분류 불가 영역의 편중은 무작위 표본 손실이 아니라, 표준화된 상품명이 기재되는 거래와 그렇지 않은 거래의 구조적 구분을 반영한다. 이 특성에 근거하여 분류 가능 영역과 불가능 영역을 명시적으로 분리하는 2단계 분류 설계가 도입되었다. 2단계 분리 설계는 분류 가능 영역에서 높은 신뢰도의 카테고리 정보를 확보하고 불가능 영역을 식별된 한계로 분리하여, 분석 차원의 마찰을 상시 운영 정착을 가로막지 않을 수준으로 해소하였다.

분류 모델은 제도적 환경 차원에서도 마찰에 직면하였다. 이 마찰은 송장 텍스트의 구성적 위험, 곧 개인정보 내포 가능성으로 표면화되었다. 수취인 이름·차량번호·휴대전화 번호·현관 비밀번호 등은 원칙적으로 송장의 별도 변수로 관리되어야 하며 송장명 텍스트에는 포함되지 않아야 한다. 그러나 일부 송장에서 이들 정보가 송장명 안에 예외적으로 기입된 형태가 발견되었다. 원생성 시점의 배송 운영 목적에서는 보조 수단으로 기입된 정보였으나, 상시 분석 활용이라는 2차 목적에서는 개인정보 보호 법규에 따른 정제 의무를 수반한다. 송장 텍스트의 원생성 목적과 재사용 목적 사이의 거리는 제도적 환경 차원에서 정제 책임과 관리 책임의 문제로 전환되었다. Bates(2018)의 규제적 측면과 Aula(2019)가 규명한 제도적 환경 차원으로의 마찰 확산이 유통·물류 산업 데이터 맥락에서도 재확인되는 지점이다.

분류 모델의 정착 지연은 상시 운영 준비 단계에서 드러났다. 2025년 11월 양사 공동 보고에서 모델 성능과 결과 활용 가능성이 제시되어 발견 지향 가치가 확인되었고, 그 상태로 분석 과제가 종료되었다. 그러나 이후 상시 배치 운영을 준비하는 과정에서 B사 실무자들은 송장 텍스트 안에 개인정보가 예외적으로 포함될 수 있음을 확인하였다. 2026년 5월 현재 외부 정제 전문 회사에 위탁하여 A사가 송장 텍스트의 개인정보를 정제한 뒤 B사에 제공하는 방안이 양사 간 검토 중이며, 분류 모델의 상시 배치 운영은 보류 상태에 있다. 이 보류는 분류 모델의 분석 가치 자체가 부정되었기 때문이 아니라, 상시 활용을 위한 제도적 조건, 즉 개인정보 정제·이전·관리 책임의 분배가 상시 운영 준비 단계에서 별도 과제로 확인되었기 때문이다. 사례에서 비정형 송장 텍스트는 입력 단계에서 개인정보가 포함될 가능성이 있다는 점에서 제도적 환경 차원의 마찰에 상대적으로 취약한 특성을 보였다.

다만 개인정보 이슈가 분류 모델의 분석 가치 자체를 부정하는 것은 아니다. B사 정보보호 부서의 검토 기준에 따르면, 개인정보가 포함된 송장은 차량번호 등을 폭넓게 포함한 보수적 추정에서도 분류 불가능 데이터 22.7%를 제외한 분석 가능 데이터

(77.3%)의 약 3% 수준으로 확인되었다. 개인정보 포함 송장을 분류 대상에서 모두 제외하는 보수적 시나리오를 적용하면, 최종 정분류율은 77.3% × (1−3%) × 94.0% × 76.5% ≈ 53.9%로 추정되며, 현행 55.6%보다 약 1.7%p 낮은 수준이다. 개인정보에 해당하는 항목이 분석 전처리 단계에서 제외되는 경우를 가정한 시나리오 분석에서도 카테고리 정확도는 76.5%로 나타났으며, 발견 지향 가치가 실무적으로 유지될 수 있음을 시사한다. 단, 분석 데이터는 개인정보보호에 따라 A사에 반납되어 현재 접근이 제한되어 있으므로, 이 수치는 기존 분석 결과에 기초한 시나리오 추정치이며 정제 방안 적용 이후의 실측 재검증은 후속 과제로 남는다. 이러한 점에서 분류 모델의 정착 지연은 분석 가치의 약화가 아니라, 상시 운영을 위한 개인정보 정제 조건이 아직 충족되지 않은 데서 비롯된 것으로 해석할 수 있다.

이상의 분석 결과를 바탕으로 명제 P-3의 예측 패턴과 관찰된 패턴을 대조하면 다음과 같다. 명제 P-3은 정착 경로가 데이터 분석 차원의 마찰과 제도적 환경 차원의 마찰의 차이에 따라 분기할 것이라는 패턴을 예측한다. 예측 모델은 두 차원의 마찰이 모두 낮아 가치 입증 직후 상시 배치 운영으로 정착하였다. 이와 달리, 분류 모델은 분석 차원의 마찰을 원인 규명에 근거한 2단계 분리 설계로 해소하였으나, 제도적 환경 차원의 정제 의무와 책임 분배 문제가 남아 상시 운영이 보류되었다. 예측 모델에서는 데이터 분석 차원과 제도적 환경 차원의 마찰이 모두 낮게 나타났다. 반면 분류 모델에서는 2단계 분리 설계를 통해 데이터 분석 차원의 마찰이 해소되었으나, 제도적 환경 차원의 마찰은 잔존하였다. 이러한 마찰 양상의 차이는 두 모델의 상이한 정착 경로와 체계적으로 대응한다. 따라서 관찰된 사례 패턴은 두 차원의 마찰이 분리되어 작동하며, 그 차이에 따라 정착 경로가 분기한다는 P-3의 예측과 부합한다.

4.4 인과 추론 종합 평가

4.1절부터 4.3절까지의 패턴 매칭은 명제 P-1·P-2·P-3이 예측한 패턴이 사례에서 관찰된 패턴과 일치함을 확인하였다. 패턴 일치는 인과 작동의 필요조건일 뿐 충분조건은 아니다. 관찰된 상관이 명제가 진술하는 인과 관계의 실제 작동을 보장하지는 않기 때문이다. 3.3절에서 채택한 이론 검증 과정 추적 절차에 따라, 각 명제의 인과 추론 강도를 Van Evera(1997), Collier(2011) 및 Mahoney(2012)의 테스트 체계로 평가한다. 평가는 명제별로 통과 테스트(hoop test, 필요조건 점검), 결정적 단서 테스트(smoking gun test, 충분조건 점검) 및 방향 지시 테스트(straw-in-the-wind test, 보조 증거)의 순서로 수행된다. 명제별로 기존 문헌의 적용 한계와 사례가 제시하는 새로운 설명의 방향은 〈표 3〉에 사전 정리하였으며, 인과 추론 테스트의 종합 결과는 〈표 9〉와 같다.

종합하면, 각 명제에 대해 통과 테스트와 결정적 단서 테스트에 부합하는 사례 증거가 확인되었다. 명제 P-1은 두 가치가 동일한 통합 데이터 아키텍처에서 발생한다는 필요조건을 충족하였다. 또한 A사 단독 수기 예측보다 높은 물동량 및 배차량 예측 성과가 확보되고, B사 자체 데이터만으로는 관찰하기 어려운 외부 온라인 시장의 카테고리 분포가 통합 데이터에서 관찰되었다는 점에서 반사실 비교의 강한 단서를 제시하였다. 명제 P-2는 탐색적 분석 이후 정식 계약, 공식적 분석, 상시 운영 및 시스템 수정으로 이어지는 다단계 제도화의 시간적 순서를

〈표 9〉 명제별 인과 추론 결과

명제		P-1(가치 발생)	P-2(가치 정착)	P-3(정착 경로)
인과 주장		▪ 데이터의 비경합성으로 인해 동일한 통합 데이터 아키텍처가 두 방향으로 작동하여 효율화 지향 가치(예측 모델)와 발견 지향 가치(분류 모델)가 동시에 발생	▪ 다단계 제도화(탐색적 분석, 정식 계약, 공식적 분석, 상시 배치 운영, 시스템 수정)를 통해 일회성 가치가 지속 가능한 데이터 자산으로 정착	▪ 양사가 획득한 가치의 정착 경로는 데이터 분석 차원과 제도적 환경 차원의 데이터 마찰에 따라 분기
사례 증거	통과 테스트 (hoop test, 필요조건)	▪ 효율화 지향 가치(예측 모델)와 발견 지향 가치(분류 모델)가 같은 통합 데이터 아키텍처로 작동 ▪ 동일 통합 데이터에서 이중 가치가 함께 관찰되어 P-1 필요조건과 부합	▪ 다단계 시간적 순서 관찰 (제도화 3단계; 습관화, 객체화, 침전화) ▪ 정식 계약 이후 공식적 분석, 상시 운영 및 시스템 수정이 순차적으로 진행되어 P-2 필요조건과 부합	▪ 두 모델의 자산화 정착 경로 분기 (예측 모델 정착 완료, 분류 모델 정착 지연) ▪ 예측 모델과 분류 모델이 데이터 마찰에 따라 상이한 정착 경로가 관찰되어 P-3의 필요조건과 부합
	결정적 단서 테스트 (smoking gun test, 충분조건)	▪ (예측 모델) A사 단독 수기 예측(물동량 70% 미만·배차량 50% 미만)보다 높은 물동량 모델 88.7%·배차량 모델 86.3%가 통합 데이터에서 확보 (반사실 비교를 통한 강한 지지 단서) ▪ (분류 모델) A사 자체 분류 체계(14개·66%)와 B사 자사 데이터만으로는 관찰이 어려운 외부 온라인 시장 카테고리 분포를 통합 데이터의 277개 분류(76.5%)로 관찰 (반사실 비교를 통한 강한 지지 단서)	▪ 정식 계약(거버넌스 합의)이 공식적 분석과 상시 운영의 범위와 내용을 규정 ▪ 정식 계약 없이는 공식적 분석과 후속 상시 운영을 실행하기 어려움 → 본 사례에서 단순 학습·역량 축적 설명을 약화 (대안적 설명: 경쟁 가설로서 학습·역량 축적 시각)	▪ 조직 간 협력을 단위로 동일 통합 데이터 아키텍처를 통해 서로 다른 가치를 산출하는 두 모델의 자산화 분기가 두 차원 마찰의 차이와 체계적으로 대응 ▪ 특이한 사례 근거를 직접 활용하는 강한 단서이며, 데이터 마찰은 동일 조건하에서 두 모델의 분기를 다른 가설보다 정합적으로 설명
	방향 지시 테스트 (straw-in-the-wind test, 보조증거)	▪ 예측 모델: A사 운영 시계열 데이터 + B사 시계열 파생·휴일·시장 변수와 상품 구매 데이터 ▪ 분류 모델: A사 송장 데이터 + B사 PCM(상품카테고리)와 학습 라벨·정제 기법	▪ Rageth et al.(2021)의 제도화 정도(전제도화·준제도화·완전 제도화)와 본 사례 단계의 정합 대응 ▪ Xu et al.(2024)의 자산화 프레임워크(자원 준비·역량 구축·데이터 활용) 세 단계와 본 사례 전개의 정합 대응	▪ 데이터 분석 차원 마찰: 분류 모델의 분류 불가 송장데이터가 무작위가 아니라 대표성 낮은 거래에 체계적으로 편중 ▪ 제도적 환경 차원 마찰: 분류 모델의 정착 지연이 개인정보 이슈로 인한 데이터 마찰과 직접 대응
본문 위치		4.1절	4.2절	4.3절

주) 결정적 단서 테스트에서 검토한 주요 대안적 설명은 P-1의 단독 자원 설명, P-2의 단순 학습·역량 축적 설명, P-3의 모델 성능 차이 설명이다(5.1절 참조).

확인하였다. 아울러 정식 계약이 공식적 분석과 상시 운영의 범위와 내용을 규정하였다는 점은 거버넌스 합의 중심의 단계적 자산화 설명을 강하게 지지하며, 자산화 진행을 단순한 학습·역량 축적의 결과로만 보는 대안적 설명을 약화시킨다. 명제 P-3은 두 모델의 상이한 정착 경로가 관찰되어 필요조건과 부합하였다. 또한 동일한 계약과 통합 데이터 아키텍처를 공유한 조건에서 두 모델의 경로 차이가

데이터 분석 차원 및 제도적 환경 차원의 마찰 차이와 체계적으로 대응하였다는 점은 데이터 마찰 설명을 지지하는 강한 단서를 제공하며, 관찰된 경로 분기를 다른 가설보다 더 정합적으로 설명한다. 이상의 검증을 종합하면, 세 명제는 단일 사례 안에서 통과 테스트와 결정적 단서 테스트를 결합한 인과 추론을 뒷받침하는 근거를 제시한 것으로 평가된다. 세 명제가 제시한 인과 관계가 어떻게 하나의 작동 방식으로 통합되는지, 그리고 본 연구의 설명이 대안적 설명보다 사례 패턴을 더 정합적으로 설명하는지는 5장에서 논의한다.

Ⅴ. 결 론

5.1 결과 해석

1장에서 제기한 연구 문제에 상응하여 2장에서 명제를 도출하였다. 이렇게 도출된 명제는 4장에서 패턴 매칭과 인과 추론 테스트를 거쳐 사례 증거로 지지되었다. 이 절에서는 앞서 제기한 연구 문제, 곧 가치의 발생(RQ-1), 데이터 자산으로의 정착(RQ-2), 데이터 마찰에 따른 정착 경로의 분기(RQ-3)에 대하여 연구 결과를 해석한다. 각 해석에는 해당 결과와 경쟁하는 대안적 설명의 검토를 함께 포함한다(Bennett and Checkel, 2015; Yin, 2018). 동등귀결(equifinality) 가능성을 고려하여 검토한 대안 설명은 RQ-1의 단독 자원 설명, RQ-2의 단순 학습·역량 축적 설명, RQ-3의 모델 성능 차이 설명이다.

이중 가치의 발생(RQ-1). 자원 보완성은 서로에게 없는 자원인 A사의 운영 데이터와 B사의 분석 역량을 하나의 통합 데이터 아키텍처로 결합시켰다. 데이터의 비경합성으로 인해 통합된 동일 데이터는 소진 없이 두 방향으로 활용되어, 예측 모델을 거친 효율화 지향 가치와 분류 모델을 거친 발견 지향 가치로 전환되었다. 두 가치는 서로 경쟁하거나 상쇄되지 않고 병존하며 A사와 B사에 각각 귀속되었다. 이중 가치가 어느 한쪽 자원이 아니라 통합 데이터 아키텍처에서 발생했다는 점에서, 대안적 설명으로서 단독 자원 설명보다 자원 보완성과 비경합성의 결합이 이중 가치의 발생을 더 잘 설명한다.

가치의 자산화 정착(RQ-2). 탐색적 분석에서 확인된 가치는 그 자체로는 일회성 성과에 머문다. 사례에서 가치가 지속 가능한 자산으로 정착한 것은 각 단계의 산출이 다음 단계를 촉발하는 다단계 제도화가 작동했기 때문이다. 탐색적 분석의 성과는 정식 계약이라는 거버넌스 합의를 이끌었고, 계약은 공식적 분석과 상시 운영의 범위와 내용을 규정하였으며, 예측 모델은 시스템 수정을 거쳐 반복적 운영 의사결정 구조에 내재화되었다. 이 전개는 습관화·객체화·침전화의 제도화 3단계(Tolbert and Zucker, 1996)와 정합적으로 대응한다. 공식적 분석과 상시 운영이 정식 계약 이후에야 가능했다는 점에서, 단순한 학습·역량 축적 설명보다 거버넌스 합의를 전환점으로 하는 다단계 제도화가 가치의 정착 과정을 보다 잘 설명한다.

정착 경로의 분기(RQ-3). 동일한 계약과 통합 데이터 아키텍처를 공유한 두 모델의 경로가 갈라진 것은 데이터 마찰의 발생 원천이 달랐기 때문이다. 예측 모델은 집계·구조화 데이터를 입력받아 집계 수치를 출력하므로 두 차원의 마찰이 모두 낮았고,

가치 입증 직후 상시 배치 운영으로 정착하였다. 분류 모델은 데이터 분석 차원의 마찰을 2단계 분리 설계로 해소하였으나, 송장 텍스트의 개인정보에서 비롯된 제도적 환경 차원의 마찰이 남아 정착이 보류되었다. 분류 모델의 발견 지향 가치가 공식적 분석에서 입증되었고 개인정보 제외 시나리오에서도 유지되었다는 점에서, 모델 성능 차이 설명보다 두 차원 데이터 마찰의 차이가 두 모델의 경로 분기를 더 잘 설명한다.

연구 문제별 결과 해석은 하나의 인과 사슬로 통합된다. 인과 사슬은 연구 문제(How/Why)에서 출발하여 사례 분석을 거치며 다듬어진 상정된 인과적 연쇄(a presumed set of causal sequences)이며, 설명 구축(explanation building)으로 정교화되었다(Yin, 2018). 곧 이중 가치 창출의 산물은 자산화의 대상이 되고, 자산화의 정착 경로는 두 차원의 마찰 조건에 따라 달라질 수 있다. 조직 간 자원 보완성, 제도화, 데이터 마찰이 결합된 설명은 위에서 언급된 대안적 설명보다 사례 패턴을 더 정합적으로 설명한다.

덧붙여 연구 수행상의 대안(craft rivals) 가운데 연구자 편향(investigator bias)과 종속변수 기반 사례 선택(selection on the dependent variable; King et al., 1994)은 실세계 대안과 달리 연구 설계와 수행 과정에서 발생할 수 있는 설명 위협이다(Yin, 2018). 연구자 편향은 3장에서 문서 증거 우선, 자료원 삼각측량, 거버넌스·운영 사실에 대한 양사 확인으로 완화하였다. 종속변수 기반 사례 선택은 두 지향 가치(효율화 지향 가치와 발견 지향 가치)의 구분과 두 데이터 마찰 차원의 틀을 2장 이론적 배경에서 사전에 정의하고 4장에서 예측 모델과 분류 모델의 분석으로 검증함으로써, 사후 명명(post-hoc labeling)이 아님을 보였다. 두 유형의 잔여 한계와 후속 과제는 5.4절에서 다룬다.

5.2 이론적 기여

데이터 비경합성 문헌(Jones and Tonetti, 2020)은 동일 데이터가 여러 기업에 의해 동일 용도로 동시 사용될 때 발생하는 수확체증 효과를 다루었으나, 보완 관계의 이질적 자원이 결합되는 양자 협력 맥락은 분석 단위 바깥에 있었다. 분석 결과는 이론의 범위를 보완 관계의 이질 자원이 결합되는 양자 협력 맥락으로 확장한다. 동일한 통합 데이터 아키텍처 위에서 비경합성이 서로 다른 용도로의 분기 사용, 곧 효율화 지향 가치와 발견 지향 가치의 동시 발생에서 작동하며, 결합 시너지가 제로섬으로 분배되지 않고, 효율화 지향 가치와 발견 지향 가치가 협력 양사에 각각 동시 귀속되는 양상이 확인되었다. 이는 데이터가 전략적 대체재로 작동할 때 공유 기피 유인이 발생한다는 우려(Wang et al., 2025)와 달리, 자원이 보완적이고 용도가 분기하는 조건에서는 비경합성이 동시 가치 창출로 이어짐을 보여준다.

데이터 자산화 문헌(Birch et al., 2021; Xu et al., 2024)은 자산화를 사회기술적 실천이나 단계적 진화로 정리하였다. 그러나 두 논의는 주로 기업 내부 조직의 진화를 다루었고, 조직 간 협력의 시간축에서 어떤 사건이 진화를 촉발하며 어떤 산물로 종착하는지는 직접 다루지 않았다. 조직 간 협력을 분석 단위로 삼은 결과, 자산화는 복수 기업 간 협력의 맥락에서 탐색적 분석, 거버넌스 합의(정식 계약), 공식적 분석, 상시 운영, 시스템 수정으로 이어지는 다단계 제도화 과정으로 나타났다.

정식 계약이라는 절차적 거버넌스 합의는 탐색적 분석 단계에서 확인된 가치를 공식적 분석과 상시

운영으로 연결한 자산화의 제도적 전환점이었다. 공식적 분석은 학습·역량 축적의 자연스러운 연장만으로 실행된 것이 아니라 정식 계약 이후에 가능했으며, 분석의 확장 범위는 학습 진척이 아니라 계약 내용에 의해 한정되었다. 자산화의 다단계 제도화는 학습·역량 축적 가설보다는 거버넌스 합의를 중심으로 전개된다. 자산화의 종착점 역시 분석 성과의 확인이 아니라 운영 인프라로의 물질적 통합이며, 사례에서는 예측 모델의 시스템 수정 완료가 이에 해당한다.

데이터 마찰은 데이터 분석 차원(Edwards, 2010)과 제도적 환경 차원(Bates, 2018; Aula, 2019)에 걸쳐 상호 연관된 현상으로 다뤄져 왔다. 학술적 기여는 산업 적용과 분석 틀 재구성의 두 측면에 있다. 산업 적용 측면에서는 의료·과학 영역에서 발전해 온 데이터 마찰 분석 틀을 유통·물류 산업의 운영 데이터에 적용하여, 두 차원 마찰을 한 분석 안에서 구체적으로 관찰하였다. 분석 틀 재구성 측면에서는 상호 연관된 현상으로 다뤄져 온 데이터 마찰을 발생 원천에 따라 데이터 분석 차원(정보적 속성·생성 맥락)과 제도적 환경 차원(법·규제·거버넌스)으로 구분하였다. 나아가 마찰의 발생 원천이 상이하다는 경계 조건에서 두 차원의 마찰이 분리되어 작동하며, 동일한 데이터 아키텍처를 공유하는 두 모델의 정착 경로가 서로 다르게 나타나는 양상을 확인하였다. 분류 모델에서는 데이터 분석 차원의 마찰을 2단계 분리 설계로 완화한 뒤에도 개인정보 보호라는 제도적 환경 차원의 마찰이 남아 정착 보류와 밀접하게 관련되었다. 이는 한 차원의 마찰 해소가 다른 차원의 해소로 자동적으로 이어지지 않을 가능성을 사례 수준에서 시사한다.

5.3 실무적 함의

데이터 통합에 앞서 마스터 데이터 표준화가 선행되어야 한다. 데이터 거버넌스 문헌은 데이터 표준을 조직 간 상호운용성을 확보하는 절차적 거버넌스 메커니즘으로 진술해 왔다(Abraham et al., 2019). 특히 마스터 데이터 관리는 단순한 기술 도입이 아니라, 데이터 소유자, 데이터 정의, 법규 제약, 부서 간 이해 조정 등 여러 이슈가 함께 작동하는 복잡한 조직 과제이다(Vilminko-Heikkinen and Pekkola, 2017). B사가 보유한 통합 상품 마스터(PCM)는 분류 모델의 라벨링 비용을 상당히 줄인 핵심 자원이었다. 데이터 통합은 프로젝트 단위로 시도할 수 있으나 표준 마스터 체계는 누적적 설계와 구축을 요구한다. 마스터 데이터 체계의 설계와 구축은 개별 데이터 통합 프로젝트와 분리하여, 지속 가능한 자산으로 인식하고 사전에 추진할 필요가 있다.

운영 데이터의 분석 재활용에서 데이터 정제 책임의 거버넌스 분배가 핵심 병목이다. 산업 현장에서 데이터의 2차 사용은 개인정보 정제·익명화 같은 제도적 환경 차원의 마찰을 빈번하게 일으키며, 정제 책임을 누가 어떻게 분담하는가에 대한 결정이 자산화 정착의 속도에 영향을 미친다. 데이터 거버넌스 문헌은 조직 간 데이터 협력의 핵심 메커니즘으로 데이터 공유 협약과 데이터 교환 표준의 사전 합의를 진술해 왔다(Abraham et al., 2019). 이를 고려하면, 조직 간 협력이 창출한 가치를 지속 가능한 자산으로 정착시키려면 제도적 환경 차원의 데이터 마찰을 사전에 식별하고, 데이터 원천에서의 정제 책임을 누가 맡을지에 대한 분배 표준을 함께 설계해야 한다. 사전 설계는 데이터 협력에서 가치가 확인된 이후에도 상시 운영으로 정착하지 못하는

위험을 완화하는 하나의 주요한 대응 방안이 될 수 있다.

한국어 비정형 텍스트 분석의 산업 적용에서 평가 기준의 재설정이 요구된다. 분석 결과는 한국어 트랜스포머와 표준 마스터 라벨의 결합을 한국어 비정형 텍스트 분류의 산업 적용 사례로 제시한다. 동시에 분석 불가 영역을 모수에 포함하여 계산한 전체 정분류율이 발견 지향 가치의 수준을 과소평가할 수 있다는 점도 확인되었다. 비정형 텍스트 분석에서 데이터 분석 차원의 데이터 마찰이 체계적으로 편중되어 발생할 때, 전체 기준 지표는 분석 가치의 실현 수준을 과소 진단할 우려가 있다. 비정형 텍스트 분석에서는 전체 정분류율과 분류 가능 영역의 정확도를 구분하는 2단계 평가를 고려할 필요가 있다.

5.4 한계 및 후속 연구

완화되었지만 잔여 한계로 남는 내부자 연구자 위치성은 산출물, 시스템 로그, 내부 거버넌스 문서에 대한 접근 가능성과 맥락 이해의 깊이를 확보하는 강점이 있는 반면, 해석 편향의 가능성을 동반한다. 문서 증거 우선, 자료원 삼각측량, 거버넌스·운영 사실에 대한 양사 확인으로 위협을 완화하였으며, 추가적 보완은 후속 연구의 과제로 남는다. 후속 연구에서는 제3자 연구자에 의한 독립적 검증을 통해 내부자 위치성에서 비롯되는 한계를 보완할 필요가 있다.

또 다른 잔여 한계인 종속변수 기반 사례 선택은, 가치 창출과 자산화가 관찰된 사례를 결과에 근거해 골랐다는 점에서 생기는 위협이다. 연구 설계의 부주의라기보다 조직 간 데이터 결합이라는 현상 자체의 특성에서 비롯된 제약이다. 조직 간 데이터 협력은 접근이 어렵고 과정이 비공개로 진행되는 경우가 많아, 가치 창출에 이르지 못한 협력은 현상 자체가 식별되지 않는다. 양사 협력이 실제로 전개된 계시적 사례(revelatory case)는 일반적으로 접근하기 어려운 데이터 협력의 전개 과정과 명제의 작동 양상을 관찰할 수 있는 기회를 제공하였다. 사후 명명(post-hoc labeling)의 위협은 이론적 배경과 사례 분석을 통한 인과적 검토로 완화하였으나, 단일 사례라는 본질적 제약은 남는다. 후속 연구는 다른 조직 간 데이터 통합 사례를 문자적 복제(literal replication, 비슷한 결과를 예측)와 이론적 복제(theoretical replication, 대조 결과를 예측)로 비교하여 명제를 재검토할 필요가 있다.

단일 사례의 통계적 일반화 제약도 남는다. 분석적 일반화를 목표로 삼았기 때문에 통계적 일반화는 범위 밖이다. 명제 P-1~P-3은 모집단으로의 빈도 분포 추론이 아니라 다른 사례로의 이론적 적용을 통해 검토될 명제다. 설명 범위는 유통·물류 산업의 조직 간 데이터 협력과 동일 데이터 아키텍처 환경, 특히 운영 기능과 분석 기능이 조직 간에 분화된 조건을 중심으로 도출되었다. 다른 산업이나 데이터 거버넌스 구조로의 일반화에는 추가 사례 검토가 필요하다.

분류 모델 정착이 진행 중이라는 종단적 한계도 있다. 예측 모델은 상시 운영을 거쳐 2026년 5월 시스템 수정을 완료하였으나 분류 모델은 데이터 마찰로 정착이 지연되어, 두 경로를 단일 시점에서 동시에 완결적으로 제시하기는 어렵다. 두 모델의 정착 경로를 종단적으로 추적하는 후속 연구가 요구된다.

분류 불가 영역의 편중과 관련해서도 정량적 검증의 한계가 남는다. 편중은 정량 분포가 아니라 유형 구분과 실제 송장명 사례, 발생 메커니즘의 서술로 제시하였는데, 거래 유형별 정량 집계가 분석 당시 연구 범위 밖이었고 이후 데이터가 개인정보 보호를

위해 A사에 반납되어 재수행도 어려웠기 때문이다. 이에 대한 실증적 검증은 데이터 접근이 확보되면 보완할 과제로 남는다.

한계에도 불구하고, 본 연구는 자원 보완 관계의 양사가 운영 데이터와 분석 역량을 결합하여 서로 다른 가치를 창출하고 지속 가능한 자산으로 정착시키는 과정을 약 2년의 종단 단일 사례로 심층 분석하였다. 정착 경로가 두 차원의 데이터 마찰에 따라 분기하는 작동 방식을 함께 규명한 점에서 이론적·실무적 기여가 있다.

참고문헌

김태경, 심선영, 안현철, 이정(2025), "2025 대한민국 디지털 비즈니스 트렌드 인식조사: 학계와 산업계의 다양한 목소리를 들어보다," **경영정보학연구**, 제27권 1호, pp.355-377.

(Kim, T. K., Shim, S. Y., Ahn, H. C., and Lee, J. (2025), "2025 Korea Digital Business Trend Study: Listening to Voices from Academia and Industry," *Information Systems Review*, 27(1), pp.355-377.)

류가루, 방영석(2026), "도시 내 이동이 신용을 드러낸다: 지리적 위치 정보 기반 신용 채무불이행 예측," **경영학연구**, 제55권 1호, pp.285-311.

(Liu, J. and Bang, Y.(2026), "Your Movement in a City Reveals Your Credit: Credit Default Prediction Based on Geolocation Information," *Korean Management Review*, 55(1), pp.285-311.)

이중원, 정옥경(2026), "플랫폼 음향 데이터 기반 K-pop 장르 비정형성과 인기도의 U자형 관계: 아티스트 지위의 조절효과를 중심으로," **경영학연구**, 제55권 1호, pp.501-523.

(Lee, J. W. and Jung, O. G.(2026), "The U-shaped Relationship between K-pop Genre Atypicality and Popularity Based on Platform Audio Data: Focusing on the Moderating Effect of Artist Status," *Korean Management Review*, 55(1), pp.501-523.)

이창환, 이성호, 이종원(2025), "국내 AI 기술 연구 동향 분석," **문화기술의 융합**, 제11권 4호, pp.657-669.

(Lee, C. H., Lee, S. H., and Lee, J. W.(2025), "Analysis of AI Technology Research Trends in South Korea," *The Journal of the Convergence on Culture Technology*, 11(4), pp.657-669.)

정동영, 배경한, 김민솔, 김지영(2026), "소상공인 폐점 예측: Distance-to-Default 모형과 소비자 행동 데이터의 통합적 접근," **경영학연구**, 제55권 3호, pp.1149-1170.

(Jeong, D., Bae, K., Kim, M., and Kim, A. J. (2026), "Predicting SME Store Closures: An Integrated Approach Using a Distance-to-Default Model and Consumer Behavioral Data," *Korean Management Review*, 55(3), pp.1149-1170.)

Abraham, R., Schneider, J., and vom Brocke, J. (2019), "Data Governance: A Conceptual Framework, Structured Review, and Research Agenda," *International Journal of Information Management*, 49, pp.424-438.

Aula, V.(2019), "Institutions, Infrastructures, and Data Friction: Reforming Secondary Use of Health Data in Finland," *Big Data & Society*, 6(2), pp.1-13.

Bates, J.(2018), "The Politics of Data Friction," *Journal of Documentation*, 74(2), pp.412-429.

Beach, D. and Pedersen, R. B.(2019), *Process-Tracing Methods: Foundations and Guidelines*,

2nd ed., University of Michigan Press, Ann Arbor.

Benbasat, I., Goldstein, D. K., and Mead, M. (1987), "The Case Research Strategy in Studies of Information Systems," *MIS Quarterly*, 11(3), pp.369-386.

Bennett, A. and Checkel, J. T.(eds.) (2015), *Process Tracing: From Metaphor to Analytic Tool*, Cambridge University Press, Cambridge.

Birch, K., Cochrane, D. T., and Ward, C.(2021), "Data as Asset? The Measurement, Governance, and Valuation of Digital Personal Data by Big Tech," *Big Data & Society*, 8(1), pp.1-15.

Borgman, C. L. and Groth, P. T.(2025), "From Data Creator to Data Reuser: Distance Matters," *Harvard Data Science Review*, 7(2). https://doi.org/10.1162/99608f92.35d32cfc

Collier, D.(2011), "Understanding Process Tracing," *PS: Political Science & Politics*, 44(4), pp.823-830.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K.(2019), "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Minneapolis, Minnesota.

Dubé, L. and Paré, G.(2003), "Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations," *MIS Quarterly*, 27(4), pp.597-636.

Edwards, P. N.(2010), *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*, MIT Press, Cambridge, MA.

Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., and Borgman, C. L.(2011), "Science Friction: Data, Metadata, and Collaboration," *Social Studies of Science*, 41(5), pp.667-690.

Gibbert, M., Ruigrok, W., and Wicki, B.(2008), "What Passes as a Rigorous Case Study?," *Strategic Management Journal*, 29(13), pp.1465-1474.

Gregor, S.(2006), "The Nature of Theory in Information Systems," *MIS Quarterly*, 30(3), pp.611-642.

Gross, N., Giacquinta, J. B., and Bernstein, M.(1971), *Implementing Organizational Innovations: A Sociological Analysis of Planned Educational Change*, Basic Books, New York.

Grover, V., Chiang, R. H. L., Liang, T.-P., and Zhang, D.(2018), "Creating Strategic Business Value from Big Data Analytics: A Research Framework," *Journal of Management Information Systems*, 35(2), pp.388-423.

Gupta, M. and George, J. F.(2016), "Toward the Development of a Big Data Analytics Capability," *Information & Management*, 53(8), pp.1049-1064.

Harrison, J. S., Hitt, M. A., Hoskisson, R. E., and Ireland, R. D.(2001), "Resource Complementarity in Business Combinations: Extending the Logic to Organizational Alliances," *Journal of Management*, 27(6), pp.679-690.

Jones, C. I. and Tonetti, C.(2020), "Nonrivalry and the Economics of Data," *American Economic Review*, 110(9), pp.2819-2858.

Jussen, I., Fassnacht, M., Schweihoff, J. C., and Möller, F.(2024), "Reaching for the Stars: Exploring Value Constellations in Inter-

Organizational Data Sharing," in Proceedings of the 32nd European Conference on Information Systems (ECIS), Paphos, Cyprus.

King, G., Keohane, R. O., and Verba, S.(1994), *Designing Social Inquiry: Scientific Inference in Qualitative Research*, Princeton University Press, Princeton, NJ.

Langley, A.(1999), "Strategies for Theorizing from Process Data," *Academy of Management Review*, 24(4), pp.691-710.

Lee, H. L., So, K. C., and Tang, C. S.(2000), "The Value of Information Sharing in a Two-Level Supply Chain," *Management Science*, 46(5), pp.626-643.

Liu, E., Ma, S., and Veldkamp, L.(2025), "Data Sales and Data Dilution," *Journal of Financial Economics*, 169, 104053. https://doi.org/10.1016/j.jfineco.2025.104053

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V.(2019), "RoBERTa: A Robustly Optimized BERT Pretraining Approach," *arXiv e-prints*, arXiv:1907.11692.

Mahoney, J.(2012), "The Logic of Process Tracing Tests in the Social Sciences," *Sociological Methods & Research*, 41(4), pp.570-597.

March, J. G.(1991), "Exploration and Exploitation in Organizational Learning," *Organization Science*, 2(1), pp.71-87.

Mikalef, P., Krogstie, J., Pappas, I. O., and Pavlou, P.(2020), "Exploring the Relationship between Big Data Analytics Capability and Competitive Performance: The Mediating Roles of Dynamic and Operational Capabilities," *Information & Management*, 57(2), 103169.

Milgrom, P. and Roberts, J.(1995), "Complementarities and Fit: Strategy, Structure, and Organizational Change in Manufacturing," *Journal of Accounting and Economics*, 19(2-3), pp.179-208.

Pasquetto, I. V., Borgman, C. L., and Wofford, M. F.(2019), "Uses and Reuses of Scientific Data: The Data Creators' Advantage," *Harvard Data Science Review*, 1(2). https://doi.org/10.1162/99608f92.fc14bf2d

Patton, M. Q.(2015), *Qualitative Research and Evaluation Methods: Integrating Theory and Practice*, 4th ed., SAGE Publications, Thousand Oaks, CA.

Perifanis, N.-A. and Kitsios, F.(2023), "Investigating the Influence of Artificial Intelligence on Business Value in the Digital Era of Strategy: A Literature Review," *Information*, 14(2), 85, pp.1-42.

Qin, H., Liu, Y., and Hu, Y.(2025), "Commercial Value of Data Resource: An Overview of Current State of Research and Future Prospects," *SAGE Open*, 15(4), pp.1-21.

Rageth, L., Caves, K. M., and Renold, U.(2021), "Operationalizing Institutions: A Theoretical Framework and Methodological Approach for Assessing the Robustness of Social Institutions," *International Review of Sociology*, 31(3), pp.507-535.

Robey, D., Im, G., and Wareham, J. D.(2008), "Theoretical Foundations of Empirical Research on Interorganizational Systems: Assessing Past Contributions and Guiding Future Directions," *Journal of the Association for Information Systems*, 9(9), pp.497-518.

Safran, C., Bloomrosen, M., Hammond, W. E., Labkoff, S., Markel-Fox, S., Tang, P. C., and Detmer, D. E.(2007), "Toward a National Framework for the Secondary Use of Health

Data: An American Medical Informatics Association White Paper," *Journal of the American Medical Informatics Association*, 14(1), pp.1-9.

Scott, W. R.(2014), *Institutions and Organizations: Ideas, Interests, and Identities*, 4th ed., SAGE Publications, Thousand Oaks, CA.

Selznick, P.(1957), *Leadership in Administration: A Sociological Interpretation*, Harper and Row, New York.

Tolbert, P. S. and Zucker, L. G.(1996), "The Institutionalization of Institutional Theory," in Clegg, S. R., Hardy, C., and Nord, W. R.(eds.), *Handbook of Organization Studies*, SAGE Publications, London, pp.175-190.

Van Evera, S.(1997), *Guide to Methods for Students of Political Science*, Cornell University Press, Ithaca, NY.

van der Meulen, N., Wixom, B. H., and Beath, C. M.(2025), "Maximizing Returns from Data Monetization Strategies," MIT CISR Research Briefing, No. XXV-2, MIT Center for Information Systems Research, https://cisr.mit.edu/publication/2025_0201_DataMonetizationImpact_vanderMeulenWixomBeath, retrieved May 2026.

Venable, J., Pries-Heje, J., and Baskerville, R.(2016), "FEDS: A Framework for Evaluation in Design Science Research," *European Journal of Information Systems*, 25(1), pp.77-89.

Vilminko-Heikkinen, R. and Pekkola, S.(2017), "Master Data Management and Its Organizational Implementation: An Ethnographical Study within the Public Sector," *Journal of Enterprise Information Management*, 30(3), pp.454-475.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R., and Childe, S. J.(2017), "Big Data Analytics and Firm Performance: Effects of Dynamic Capabilities," *Journal of Business Research*, 70, pp.356-365.

Wang, K., Li, J., Lin, S.-B., and Wang, Y.(2025), "Non-Rival Data as Rival Products: An Encapsulation-Forging Approach for Data Synthesis," *arXiv e-prints*, arXiv:2511.06610.

Wixom, B. H. and Ross, J. W.(2017), "How to Monetize Your Data," *MIT Sloan Management Review*, 58(3).

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. M.(2020), "Transformers: State-of-the-Art Natural Language Processing," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pp.38-45, Online.

Wynn, D., Jr. and Williams, C. K.(2012), "Principles for Conducting Critical Realist Case Study Research in Information Systems," *MIS Quarterly*, 36(3), pp.787-810.

Xu, T., Shi, H., Shi, Y., and You, J.(2024), "From Data to Data Asset: Conceptual Evolution and Strategic Imperatives in the Digital Economy Era," *Asia Pacific Journal of Innovation and Entrepreneurship*, 18(1), pp.2-20.

Yin, R. K.(2018), *Case Study Research and Applications: Design and Methods*, 6th ed., SAGE Publications, Thousand Oaks, CA.

• 저자 이대성은 성균관대학교 경영학과 초빙교수를 역임하였으며, 대기업군 기업에서 데이터 담당 임원을 거쳐 현재는 자문으로 활동하고 있다. 성균관대학교에서 경영정보 전공으로 박사 학위를 취득하였다. 주요 관심 분야는 지능형 의사결정지원시스템, 빅데이터 분석, Predictive AI 등이다. Computers in Human Behavior, Online Information Review 등 국내외 학술지에 논문을 발표하였다.