

연구논문

머신러닝 기반의 인과 포레스트 기법을 활용한 처치효과 검증: 교내 동아리활동 참여가 협업능력에 미치는 효과를 중심으로

석유미* · 이진실**

이 연구의 목적은 머신러닝 기반의 인과 포레스트 기법을 활용해 처치효과를 검증하는 방법을 소개하고, 이를 활용하여 교내 동아리활동 참여가 협업능력에 미치는 효과를 분석하는 것이다. 인과 포레스트는 머신러닝을 활용한 인과추론 방법으로, 이 연구에서는 인과 포레스트 기법에 대한 소개와 함께 이를 활용하여 구체적으로 교내 동아리활동이 협업능력에 미치는 효과를 분석하였다. 평균 처치효과 분석결과, 공변인을 통제한 후에도 교내 동아리활동에 참여한 집단에서 참여하지 않은 집단에 비해 협업능력 점수가 통계적으로 유의하게 더 높게 나타나, 동아리활동이 교과수업 위주의 학교교육을 보완하여 청소년기 사회적 발달에 기여할 수 있음을 확인하였다. 성별, 또래관계, 교사관계에 따른 차별적 처치효과를 탐색했을 때에는 통계적으로 유의한 결과는 발견되지 않았다.

주제어: 머신러닝, 인과추론, 인과 포레스트, 교내 동아리활동, 협업능력

* 버지니아 대학교 데이터 사이언스 스쿨 조교수(youmisuk@virginia.edu), 제1저자.

** 서울대학교 교육학과 강사(leejinsil326@gmail.com), 교신저자.

I. 서론

교육학이나 심리학 등 사회과학 분야에서 다양한 처치들에 대한 효과를 무선허당 실험(randomized experiments) 혹은 관찰된 자료(observational data)를 이용해 검증하려는 노력이 이루어지고 있다. 무선허당 실험 디자인을 이용하는 경우, 처치 할당이 연구자에 의해 결정되기 때문에 참여자에 의한 선택편의(selection bias)의 문제가 발생하지 않는다. 그러나 관찰된 자료를 사용하는 경우, 연구자 이외의 다른 요소들로 인해 처치 상태가 결정되기 때문에, 선택편의의 문제가 발생한다. 이러한 선택적 편의를 없애기 위해서 다양한 통계적 방법들이 소개되었으며(Cook et al. 2002), 그중 경향점수 방법이 최근 20년간 많이 사용되고 있다(Austin 2011; Schafer & Kang 2008; Steiner & Cook 2013; 김준엽 외 2008). 경향점수는 개개인이 처치를 받을 조건적 확률을 뜻하는데(Rosenbaum & Rubin 1983), 이분형 처치변수(binary treatment variable)의 경우 일반적으로 로지스틱 모형을 사용하여 경향점수를 추정한다. 구체적으로, 연구자가 사전 지식을 바탕으로 공변인들과 처치변수와의 관계를 직접 모형에서 설정하고, 설정된 모형을 바탕으로 경향점수 및 처치효과를 추정한다.

하지만 때때로 사전 지식만으로 변수들 간의 관계를 파악하기 어려운 경우들이 있다. 예를 들면, (i) 처치변수와 관련된 사전 연구가 많이 진행되지 않은 경우, (ii) 수집한 표본의 특수성으로 인해 변수 간 관계가 사전 지식과 완전히 일치하지 않는 경우, (iii) 변수들이 고차항(higher-order terms)의 관계와 같이 복잡한 양상의 관계일 때 이를 정확하게 예상하기 어려운 경우, (iv) 데이터가 방대하고 변수가 매우 다양하여 연구자가 모든 변수들 간의 관계를 파악하기 어려운 경우 등이 있다. 이러한 상황에서는 머신러닝 기법이 유용할 수 있는데, 머신러닝 방법들에서는 변수들 간의 복잡한 관계를 데이터 기반의 알고리즘을 통해 유연하게 자동적으로 설정한다는 장점이 있다. 연구자의 사전 지식이 충분하지 않은 경우 불충분한 사전 지식으로 설정한 모형 결과에만 의존하는 것보다는, 데이터 기반의 머신러닝 방법들을 사용하면 보다 강건한(robust) 추정치를 산출할 수 있다.

이러한 머신러닝 기법들은 다양한 분야에서 활용되고 있는데, 최근에는 이러한 머신러닝을 인과추론에 접목하고자 하는 시도도 이루어져 왔다. 머신러닝 방법을 활용한 인과추론 방법들이 많이 개발되어 왔고, 최근에 개발된 머신러닝 기반의 인

과추론 방법들은 평균 처치효과(average treatment effect)와 개별 / 조건적 평균 처치효과(conditional average treatment effects)를 추정하는 데 목적이 있다(Athey & Imbens 2016; Dorie et al. 2019; Imai & Ratkovic 2013; Künzel et al. 2019; Suk et al. 2020a). 머신러닝을 활용한 대표적인 인과추론 방법으로는 인과 포레스트(causal forests) (Wager & Athey 2018; Athey et al. 2019), 베이지안 가법 회귀나무 모형 (Bayesian additive regression trees model; BART) (Hill 2011), 표적 최대우도 추정(targeted maximum likelihood estimation; TMLE) (van Der Laan & Rubin 2006) 등이 있으며, 인과 포레스트는 랜덤 포레스트(random forests) (Breiman 2001) 기반의 인과추론 머신러닝 방법이다. 머신러닝 기반의 인과추론 방법들은 유연하게 자동적으로 모형을 설정하기 때문에 연구자의 모형설정 오류로 인해서 발생할 수 있는 편의를 줄일 수 있다. 이러한 장점이 주목받으면서 머신러닝 인과추론 방법을 활용하여 처치효과를 분석하려는 연구들도 최근에 활발하게 진행되고 있다(Suk et al. 2020b; Suk & Kang 2021; Davis & Heller 2017 등). 본 연구에서는 이러한 머신러닝 기반의 인과추론 방법을 국내에 소개하고, 이를 적용하여 교내 동아리활동 참여가 협업능력에 미치는 효과를 추정하고자 한다.

최근 청소년기 전인적 발달을 위해 학업 외 활동의 중요성이 부각되고 있으며, 교육부는 2009 개정 교육과정부터 창의적 체험활동의 하위영역으로 동아리활동을 포함하고 학교에서 이루어지는 동아리활동을 적극적으로 장려하고 있다(교육과학기술부 2009). 동아리활동은 학생들의 자발적인 참여를 통해 이루어지기 때문에 무선허당 실험 연구가 현실적으로 불가능하며, ‘동아리활동에 참여하기로 선택한 학생들’과 ‘동아리활동에 참여하지 않겠다고 선택한 학생들’은 동아리활동을 경험하기 전에 이미 체계적인 차이를 가진다. 선행연구에 의하면, 학생의 사회정서적 특성, 가정환경, 학교 내 환경 등이 동아리활동의 참여에 영향을 미치는 것으로 밝혀져 있다(김나영·황혜영 2019; 김영서·홍세희 2021; 심우정·백선희 2020; 이승렬 2013; 이진실 2021 등). 또한 학생 개인특성 및 부모의 양육태도, 교사와의 관계 등은 협업능력에도 영향을 주는 것으로 알려져 있다(김도희 2020; 이승희·신소영 2018; 전란영 2020; 조영희·백아롱 2021 등). 처치에 참여하기 전에 이미 체계적인 차이가 존재하고 처치변수와 결과변수에 동시에 영향을 미치는 공변인이 존재할 때, 이를 통제하지 않으면 선택편의가 발생하여 동아리활동의 효과를 정확하게 확인할 수 없게 된다.

그동안 동아리활동의 효과를 분석한 선행연구들에서는 주로 회귀분석이나 구조

방정식모형 등을 활용해왔는데, 투입된 공변인이 충분하지 않거나 집단 간 사전 동질성을 확보하는 데 한계가 있는 경우가 많았다. 청소년활동 효과성에 대한 실증연구 66편을 검토한 박경미 외(2020)는 청소년활동 효과를 분석한 기존의 선행연구들이 연구영역이나 연구방법에 있어 확장 없이 반복적인 형태로 이루어지고 있다는 점을 지적하기도 하였다. 본 연구에서는 머신러닝 기반의 인과 포레스트 기법을 활용해 처치효과를 검증하는 방법을 소개하고, 이 방법을 적용하여 중학생의 교내 동아리활동 참여가 협업능력 발달에 미치는 효과를 분석하였다. 동아리활동이 협업능력에 미치는 평균 처치효과(average treatment effect; ATE)를 추정함과 동시에, 성별, 또래관계, 교사관계에 따라 동아리활동의 효과가 다르게 나타나는지(heterogeneous treatment effects)도 탐색하였다.

II. 이론적 배경

1. 머신러닝을 활용한 인과추론

1) 머신러닝을 활용한 인과추론 방법

최근 10년 동안 머신러닝 방법을 활용한 인과추론 방법들이 많이 개발되고 있으며, 이 방법들은 평균 처치효과와 조건적 평균 처치효과를 추정하는 데 사용되고 있다. 기존의 모수적 방법과 달리, 머신러닝 방법들은 처치 모형(treatment model)과 결과 모형(outcome model)의 함수적 관계를 매우 유연하게 자동적으로 설정한다. 여기서 처치 모형은 처치변수가 종속변수인 모형(즉, 교내 동아리활동 참여 여부를 종속변수로 하는 모형)을 뜻하고, 결과 모형은 결과변수가 종속변수인 모형(즉, 협업능력을 종속변수로 하는 모형)을 뜻한다. 앞에서 언급했듯이, 대표적인 머신러닝을 활용한 인과추론 방법으로는 인과 포레스트, BART, TMLE가 있다.

일반적으로 인과추론을 위한 머신러닝 방법들은 머신러닝 방법을 이용하여 결과 모형 또는 처치 모형을 설정한다. 결과 모형은 관찰된 공변인과 처치변수의 조건적 종속 평균, $m(\mathbf{x}, z) = E[Y_i | \mathbf{X}_i, Z_i]$, 혹은 공변인만의 조건적 종속 평균 $m(\mathbf{x}) = E[Y_i | \mathbf{X}_i]$ 으로 추정된다. 처치 모형은 공변인의 조건적 처치 평균 $e(\mathbf{x}) = E[Z_i | \mathbf{X}_i]$ 으로 추정된다.¹⁾ 인과추론을 위한 머신러닝 방법들은 각기 다른 학습 알고리즘을

사용하여 결과 모형 혹은 처치 모형을 추정하고 조건적 평균 처치효과를 계산하기 위해서 각기 다른 접근법을 사용한다. 구체적으로, BART는 베이지안 가법 회귀나무 기법을 사용하여 결과 모형을 추정하고, 일반적으로 결과 모형만으로 조건적 평균 처치효과를 계산한다. TMLE는 슈퍼러너(SuperLearner; van der Laan et al. 2007)를 기반으로 한 앙상블(ensemble) 학습 알고리즘을 사용하여 결과 모형 및 처치 모형을 추정하고, 처치 예측값을 이용한 변수를 결과 모형에 추가적인 공변인으로 투입하여 처치효과를 계산한다. 그리고 인과 포레스트는 랜덤 포레스트를 사용하여 결과 모형 및 처치 모형을 추정하고 가중(weighted) 선형 회귀 접근법을 통하여 조건적 평균 처치효과를 계산한다.

2) 인과 포레스트(causal forests)

인과 포레스트는 랜덤 포레스트를 기반으로 결과 모형과 처치 모형을 설정한다. 랜덤 포레스트는 다수의 결정 나무(decision trees)를 결합하여 분류 혹은 평균 예측치를 도출하는 앙상블 머신러닝 방법이다(Breiman 2001; Hastie et al. 2017). 인과 포레스트는 이러한 랜덤 포레스트 알고리즘을 수정하여 결과 및 처치를 예측하고 추정된 예측값을 바탕으로 조건적 평균 처치효과와 평균 처치효과를 계산한다. 구체적으로 조건적 평균 처치효과를 계산하기 위해서, 인과 포레스트는 가중 선형 회귀 접근법을 사용하고 기존 변수들에서 예측된 값을 뺀 잔차 변수들을 이용한다. 예측된 결과 및 처치 추정치는 머신러닝에서 흔하게 사용되는 OOB 리브-원-아웃(out-of-bag leave-one-out) 방법으로 추정된다. 이는 특정 i 를 제외하여 모형을 설정한 후, 추정된 모형으로 특정 i 에 대한 예측치를 계산하는 방법이다. 결과적으로, 인과 포레스트는 다음 식에 따라서 개별 조건적 평균 처치효과 $\tau(\mathbf{x})$ 을 계산한다.

$$\tau(\mathbf{x}) = \frac{\sum_i \alpha_i(\mathbf{x}) (Y_i - \hat{m}^{(-i)}(\mathbf{X})) (Z_i - \hat{e}^{(-i)}(\mathbf{X}))}{\sum_i \alpha_i(\mathbf{x}) (Z_i - \hat{e}^{(-i)}(\mathbf{X}))^2}$$

여기서 $\alpha_i(\mathbf{x})$ 는 특정 i 가 조건적 평균 처치효과 $\tau(\mathbf{x})$ 을 계산하는데 미치는 공헌도를 뜻하며, 0과 1사이의 값을 가진다. 뺀 $(-i)$ 는 리브-원-아웃 방법을 뜻한다.

1) 여기서 Y_i 는 결과변수, Z_i 는 처치변수, X_i 는 공변인을 나타내는 기호로 사용되었고, i 첨자는 개별 분석단위(사례)를 나타내는 기호로 사용되었다.

그리고 $\hat{m}^{(-i)}(\mathbf{X})$, $\hat{e}^{(-i)}(\mathbf{X})$ 는 Wager와 Athey(2018)가 제안한 아니스트한(honest) 랜덤 포레스트 알고리즘을 통해 추정된다. 여기서 ‘아니스티(honesty)’란 인과 포레스트의 중요한 특성으로, 인과 포레스트의 나무(tree) 특성 중 하나인데 나무 모형을 설정하는 데 사용된 자료는 나무의 각 잎(leaf)에서 처치효과를 계산하는 데는 사용되지 않는다는 것을 의미한다. 마지막으로, 평균 처치효과는 개별 조건적 처치효과 추정치를 평균하여 계산할 수 있다(Athey et al. 2019; Wager & Athey 2018).

2. 동아리활동과 협업능력의 관계

동아리는 “공동의 목적과 관심사에 의해 형성, 운영되는 작은 모둠(김정주 외 2003)”이라는 뜻으로, 같은 취미나 특기를 가진 학생들이 모여 자신의 소질과 적성을 발전시키고 사회성과 협동성을 기르는 집단 활동(교육부 외 2013)을 의미한다. 동아리활동은 청소년들이 자유롭게 선택하여 참여하고, 활동이 일정 기간 지속적으로 이루어진다는 특징을 가지고 있다. 또한 학교에서 이루어지는 동아리활동은 쉽게 접근할 수 있으며, 교과활동만으로는 달성하기 어려운 정서적, 사회적 측면의 발달을 보완하는 교육적인 가치를 가진다(전자배 2020).

동아리활동은 청소년정책 및 교육정책에서도 강조되어 왔는데, 1993년 청소년기본법이 시행되고 그 이후 청소년정책기본계획 등이 마련되면서 동아리활동을 포함한 청소년활동이 정책적으로 체계화되었다(관계부처합동 2018). 2009 개정 교육과정에서는 전인적 성장을 위해 교과 외 활동의 일환으로 창의적 체험활동을 신설하였고 세부영역에 동아리활동을 포함하였다(교육과학기술부 2009). 2015 개정 교육과정에 동아리활동을 창의적 체험활동의 하위영역으로 포함하고 있고, 건전한 집단 활동을 통해 의사소통 역량과 공동체 역량 등을 함양하는 것을 목표로 제시하였다(교육부 2015a). 2016년부터는 자유학기제가 전국의 모든 중학교에서 시행되면서 동아리활동 등의 비교과활동이 더욱 강조되고 활성화되고 있다(교육부 2015b).

동아리활동은 여러 구성원들이 모여서 이루어지는 집단활동으로, 사회적 상호작용을 경험하고 공동의 목표를 위해 협력하고 갈등을 해결하는 과정을 배우는 기회가 될 수 있다. OCED DeSeCo(Defining and Selecting key Competencies) 프로젝트에서는 21세기 인재가 갖추어야 할 생애 핵심역량을 도출하였는데, ‘이질적 집단과 상호작용하기’를 핵심역량 중 하나로 선정하고 그 하위요소로 ‘타인과 협력하는 능력(ability to cooperate)’을 제시하였다(OECD 2005). 복잡하고 빠르게 변화하는 사

회적 흐름 속에서 한 사람의 능력으로는 해결하기 어려운 문제들이 많아짐에 따라, 협업능력은 미래 사회에서 필수적인 역량으로 인식되고 있으며(전자배 2020), 국내 청소년 역량지수 개발 연구에서도 핵심역량에 대인관계역량 및 그 하위요인으로 협동능력을 제시하였고(성은모 외 2015), 청소년활동진흥원에서도 청소년활동을 통해 함양하기를 기대하는 6가지 핵심역량 중 하나로 협업능력을 제시하였다(한국청소년활동진흥원 2018).

실증적인 데이터를 통해 동아리활동이 협업능력에 미치는 효과를 분석한 다수의 선행연구들에서는 동아리활동 참여가 협업능력에 긍정적인 영향을 미친다고 보고해 왔다(김한솔·장여옥 2019; 이진실 2021; 전자배 2020 등). 김한솔·장여옥(2019)의 연구에서는 자율동아리활동에 따른 핵심역량의 차이를 일원분산분석을 통해 분석했을 때, 동아리활동에 활발한 집단이 거의 참여하지 않은 집단에 비해 협업능력이 높게 나타났다. 김아름·정혜원(2020)의 연구에서는 구조방정식모형을 적용하여 자율활동, 봉사활동, 진로활동을 측정변인으로 하는 창의적 체험활동이 협업능력에 미치는 효과를 분석하였고 정적인 효과를 미치는 것으로 나타났다. 하지만 이 연구들에서는 동아리활동 및 협업능력에 동시에 영향을 미칠 수 있는 공변인은 통제되지 못하였다는 한계가 있었다. 전자배(2020)의 연구에서는 3단계 회귀분석을 적용해 성별, 학업성취도, 어머니 교육수준, 가정의 경제적 수준을 통제한 후, 교내 동아리활동과 교외 동아리활동이 협업능력에 유의한 긍정적 영향을 미친다는 점을 밝혔다. 해당 연구에서는 일부 개인 및 가정특성을 통제하였지만, 인과적인 효과를 확인하기에는 한계가 있었다. 반면, 황여정(2020)의 연구에서는 다층모형을 적용해 다양한 학생 및 학교수준 변인을 통제한 후 청소년활동의 효과를 분석했을 때 교내 동아리활동이 갈등해결역량에 미치는 효과가 통계적으로 유의하지 않아, 다소 다른 결과를 보이는 선행연구도 존재하였다.

경향점수를 통한 인과추론 기법을 적용하여 보다 엄밀한 방법으로 동아리활동의 효과를 분석한 연구들도 있었는데, 이진실(2021)의 연구에서는 경향점수 역확률가중치 방법을 적용하여 동아리활동 참여가 1년 후의 협업능력에도 긍정적인 영향을 미친다고 밝혔다. 김나영·황혜영(2019)의 연구에서는 결과변수로 협업능력은 아니지만 사회적 자아개념을 결과변수로 하여 동아리활동의 효과를 분석하였는데, 경향점수 매칭 방법을 적용했을 때 교내 동아리활동 참여집단이 비참여집단에 비해 사회적 자아개념이 유의하게 더 높음을 밝혔다. 경향점수 기법을 활용한 이들 연구에서는 분석과정에서 10~20개의 다양한 공변인들을 고려하였고, 동아리활동에 참여

한 집단과 비참여 집단의 경향점수를 유사하게 조정하여 집단 간 동질성을 확보하였다는 점, 그리고 별도의 검정을 통해 공변인 균형을 확인한 후 동아리활동의 효과를 분석하였다는 점에서 방법론적인 의의가 크다. 선행연구에서 활용된 경향점수 방법에서는 선형 회귀모형 또는 로지스틱 회귀모형과 같은 모수적인 방법을 통해 모형을 설정하고 분석해왔다. 이러한 모수적 접근에서는 변수들 간의 특정한 함수적 관계를 미리 가정하고서 함수를 구성하는 모수를 추정하기 때문에, 이러한 함수적 관계(예를 들어, 공변인과 로짓 간의 선형 관계)가 적절할 때에는 효과적인 추정 방법이 될 수 있지만, 그렇지 않을 때에는 데이터 특성에 따라 모형을 유연하게 적용하기에는 제약이 있다. 본 연구에서는 이를 보완할 수 있는 머신러닝 기법을 활용하여 교내 동아리활동이 협업능력에 미치는 효과를 재검증하였다.

III. 연구방법

1. 연구대상 및 데이터

한국아동·청소년패널조사 2018(Korean Children and Youth Panel Survey 2018; KCYPS 2018) 중에서 중1패널 1차년도와 2차년도 자료를 활용하였다. 한국아동·청소년패널조사 2018은 한국청소년정책연구원에서 전국의 초등학교 4학년, 중학교 1학년을 표집하여 2018년부터 2024년까지 매년 반복적으로 추적 조사하도록 설계된 자료이다. 중1패널 1차년도 자료는 중학교 1학년 시기에, 2차년도 자료는 중학교 2학년 시기에 조사된 자료에 해당한다. 중1패널의 조사대상자 수는 2,590명이며, 연구변수에 결측 사례를 제외(listwise)한 2,251명의 자료를 분석에 활용하였다.

2. 연구변수

1) 처치변수: 교내 동아리활동 참여여부

본 연구의 처치변수는 교내 동아리활동 참여여부이다. 조사시점을 기준으로 지난 1년 동안 교내 공식 동아리에 한 번이라도 참여한 경우에는 참여집단(처치집단)으로, 참여하지 않은 경우에는 비참여집단(통제집단)으로 구분하였다. 처치변수는 2차

년도(중2시점)에 측정된 자료를 활용하였으며, 참여집단에 해당하는 학생은 1,399명으로 전체의 62.2% 정도였고, 비참여집단에 해당하는 학생은 852명이었다.

<표 1> 교내 동아리활동 참여여부(처치변수)에 대한 빈도분석 (n=2,251)

처치집단 구분	사례 수(명)	비율(%)
교내 동아리활동 참여집단	1,399	62.2
교내 동아리활동 비참여집단	852	37.8
전체	2,251	100.0

2) 결과변수: 협업능력

결과변수는 협업능력으로, 2차년도(중2 시점)에 측정된 값을 활용하였다. 협업능력은 ‘나는 그룹활동을 할 때, 친구의 장점이거나 능력을 적극적으로 알려준다’, ‘나는 그룹활동을 할 때 나의 역할에 최선을 다한다’, ‘나는 그룹에서 발생한 문제에 대해 잘잘못을 따지기보다는 해결하려고 노력한다’ 등의 14개 문항을 통해 측정되었고, 4점 리커트(Likert) 척도(1=전혀 그렇지 않다, 2=그렇지 않은 편이다, 3=그런 편이다, 4=매우 그렇다)로 측정되었다. 협업능력은 집단축진(6문항), 팔로워십(3문항), 갈등문제해결(5문항)의 3개의 하위요인으로 구성되어 있으며, 각 하위요인별로 평균 점수를 구한 후 이를 평균내어 협업능력 점수를 생성하였다. 협업능력(결과변수)에 대한 기술통계치는 <표 2>와 같고, 측정된 점수의 cronbach α 값은 0.928로 매우 높은 신뢰도를 보였다.

<표 2> 협업능력(결과변수)에 대한 기술통계 분석 (n=2,251)

결과변수	평균	표준편차	최솟값	최댓값
협업능력	2.97	0.49	1.00	4.00

3) 공변인

공변인으로는 성별, 학업적 특성(학업성취수준, 학업열의), 정서적 특성(자아존중감, 우울), 사회적 특성(사전 협업능력, 사회적 위축, 학업열의), 전년도 경험(전년도 교내 동아리활동 참여여부 및 만족도), 가정의 SES(월평균 가구소득, 부모의 학력

수준), 부모특성(양육태도, 기대교육수준, 부모-자녀 진로관련 대화 정도, 부모-자녀 함께 보내는 시간, 어머니 직업여부), 학교특성(학교별 전년도 교내 동아리활동 참여율, 학교소재지)을 투입하여 통제하였다. 사전 공변인은 처치가 이루어지기 전 시점인 1차년도(중1시점)에 측정된 자료를 활용하였다.

<표 3> 사전 공변인에 대한 변수 설명 및 기술통계 분석(n=2,251)

구분	사전 공변인	변수설명	평균	표준 편차	
개 인 특 성	성별	남학생=0, 여학생=1의 더미변수	0.46	0.50	
	학업적 특성	학업성취수준	전 과목 성적에 대해 ‘1=매우 못함~5=매우 잘함’의 자기보고식 응답	3.44	0.92
		학업열의	‘나는 공부를 하면 뿌듯해진다’ 등 16개 문항의 평균값	2.48	0.54
	정서적 특성	자아존중감	‘나는 내가 장점이 많다고 생각 한다’ 등 10개 문항의 평균값 (역채점 문항은 역코딩)	3.00	0.50
		우울	‘불행하다고 생각하거나 슬퍼하 고 우울해한다’ 등 10개 문항의 평균값	1.85	0.55
	사회적 특성	사전 협업능력	중1 시기에 측정한 협업능력 점 수(결과변수와 동일한 방법으로 점수 생성)	3.02	0.51
		사회적 위축	‘주위에 사람들이 많으면 어색 하다’ 등 5개 문항의 평균값	2.16	0.75
		또래와의 관계	‘친구들과의 관계가 좋다’ 등 황여정(2020)을 참고한 4개 문 항의 평균값	3.20	0.54
		교사와의 관계	‘선생님은 나의 의견을 존중하 시고 자유롭게 말하도록 하신 다’ 등 황여정(2020)을 참고한 4개 문항의 평균값	2.74	0.53
	전년도 경험	전년도 교내 동아리활동_만족도 낮음(=1)	전년도 교내 동아리활동에 참 여하지 않은 집단을 참조집단 으로 하여, 참여했으나 만족도 가 낮은 집단, 참여했고 만족도 도 높은 집단을 각각의 더미변 수로 투입	0.05	0.22
전년도 교내 동아리활동_만족도 높음(=1)		0.60		0.49	

가정 특성	가정의 SES	월평균 가구소득	12개 범주(100만 원 미만, 100만 원 이상 200만 원 미만 등)에 대한 응답을 각 범주의 중간값을 부여하여 변환	525.01	227.79
		부모의 학력수준	아버지 학력과 어머니 학력을 교육연수로 변환한 후 평균값 활용	14.53	2.38
	부모 특성	양육태도_자율성 지지	‘부모님은 나를 믿어주신다’ 등 4개 문항의 평균값	3.35	0.58
		양육태도_애정	‘부모님은 나를 사랑한다는 것을 표현하신다’ 등 4개 문항의 평균값	3.38	0.57
		양육태도_구조제공	‘내가 무언가 하려고 할 때, 부모님이 어떻게 하라고 가르쳐 주신다’ 등 4개 문항의 평균값	3.07	0.54
		기대교육수준_전문대학 졸업(=1)		0.09	0.29
		기대교육수준_4년제 대학 졸업(=1)	자녀의 장래희망 교육수준에 대해 ‘고등학교 졸업’을 참조집단으로 하여, 2~3년제 전문대학 졸업, 4년제 대학교 졸업, 대학원 이상, 잘 모르겠음 집단을 각각의 더미변수로 투입	0.66	0.47
		기대교육수준_대학원 졸업(=1)		0.19	0.39
		기대교육수준_잘 모르겠음(=1)		0.04	0.20
		부모-자녀 진로 관련 대화 정도	아버지, 어머니와 진로관련 대화를 얼마나 자주하는지에 대한 4점 리커트 척도 응답을 평균내어서 활용	3.22	0.74
		주당 부모-자녀 함께 보내는 시간	자녀와 함께 보내는 시간에 대한 응답(주중 시간, 주중 분, 주말 시간, 주말 분)을 주당 시간으로 변환	35.99	17.4
		어머니 직업 여부	직업 있음(일을 한 경우)=1의 더미변수	0.73	0.44
	학교 특성	전년도 교내 동아리활동 참여 비율	학생 개인의 교내 동아리활동 참여 여부를 학교별로 평균내어 생성	20.81	17.51
		학교소재지 도시 여부	읍·면 지역=0, 도시 지역=1의 더미변수	0.86	0.35

3. 자료분석

본 연구에서는 인과 포레스트 기법을 적용하기 위해서 R 프로그램 패키지 grf (Tibshirani et al. 2021)를 사용하였다. 먼저, grf 패키지에 있는 함수 causal_forest에 처치변수, 결과변수, 공변인을 투입하여 개별/조건적 처치효과 추정치를 산출하였다. 그리고 함수 best_linear_projection에 사용하여 개별/조건적 처치효과 추정치를 바탕으로 평균 처치효과와 학생들의 특성(성별, 또래관계, 교사관계)에 따른 차별적 처치효과를 분석하였다. 구체적인 R코드는 연구결과에 함께 제시하였다.

IV. 연구결과

1. 인과 포레스트를 활용한 평균 처치효과(average treatment effect) 분석

<표 4> 평균 처치효과 분석결과²⁾

	추정치	표준오차	t통계량	p값
공변인 통제 전	2.891	0.017	172.353***	<0.001
공변인 통제 후	0.069	0.020	3.385***	<0.001

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

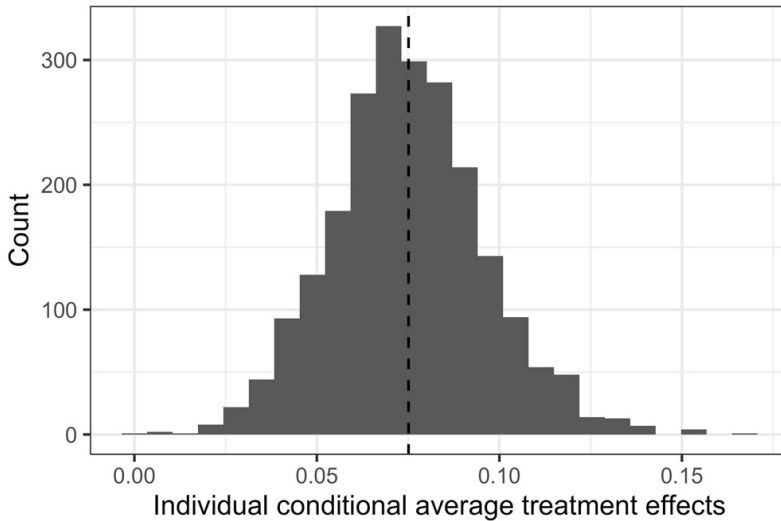
인과 포레스트를 활용하여 학생들의 교내 동아리활동 참여가 협업능력에 미치는 평균처치 효과(ATE)를 분석한 결과,³⁾ 공변인 통제 전 처치집단(교내 동아리활동 참여집단)과 통제집단(교내 동아리활동 비참여집단)의 협업능력 점수 차이는 2.891로

2) 평균 처치효과 추정 시, 개별 처치효과를 추정한 후 Athey & Wager(2019)가 소개한 방법으로 추가적인 보정을 하여 평균 효과를 계산할 수 있다. <표 4>와 <표 5>에 제시한 결과는 추가적인 보정이 적용된 값에 해당한다.

3) 본 연구에서는 동아리활동이라는 처치의 성격이 참여대상이 특정한 조건에 한정되어 있다기보다 청소년들에게 일반적으로 참여를 권장하는 활동이라는 점을 고려하여, 모집단 전체에 대한 평균 처치효과(ATE: Average Treatment Effect)를 기준으로 분석하였다. 인과 포레스트에서 연구자의 관심이나 연구맥락에 따라 처치집단에 대한 평균 처치효과(ATT: Average Treatment effect for the Treated)를 분석하는 것도 가능한데, ATT 분석을 위해서는 R프로그램 grf패키지의 average_treatment_effect(, target.sample='treated') 명령어를 활용할 수 있다.

나타났다. 하지만 인과 포레스트를 통해서 공변인을 통제한 후 분석했을 때 평균 처치효과 추정치는 0.069로 줄어들었으며, 공변인을 통제한 후에도 통계적으로 유의한 긍정적인 효과를 보였다.

앞서 기술하였듯이, 인과 포레스트는 개별 조건적 처치효과 추정치를 산출하고 이들의 평균값으로 평균 처치효과를 구한다. <그림 1>은 개별 처치효과 추정치의 분포를 보여주며 수직 점선은 평균 처치효과 추정치를 나타낸다.



<그림 1> 개별 평균 처치효과 분포

<인과 포레스트를 활용한 평균 처치효과 분석 R코드>

```
# 인과 포레스트 적용
library(grf)
out <- causal_forest(X=공변인, W=처치변수, Y=결과변수)

# 평균 처치효과 분석
ATE <- best_linear_projection(out)
```

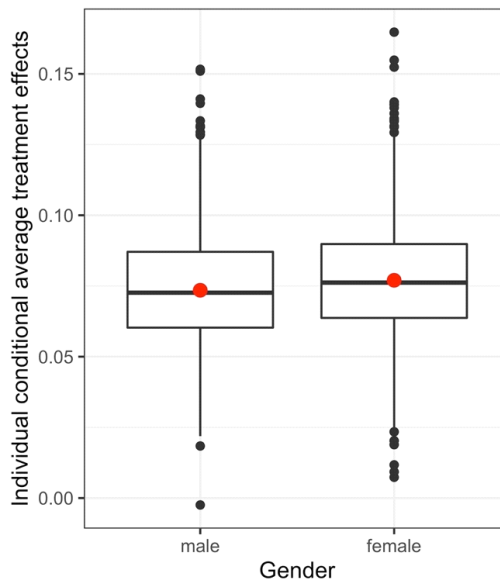
2. 인과 포레스트를 활용한 차별적 처치효과(heterogeneous treatment effects) 분석

학생들의 교내 동아리활동 참여가 협업능력에 미치는 차별적 처치효과를 살펴본 결과, 통계적으로 유의한 차별적 처치효과는 발견되지 않았다. 즉, 성별, 또래와의 관계, 교사와의 관계에 따라 교내 동아리활동의 효과가 통계적으로 유의하게 달라지는 않았다.

<표 5> 성별, 또래와의 관계, 교사와의 관계에 따른 차별적 처치효과 분석결과

	추정치	표준오차	t 통계량	p 값
성별	0.020	0.041	0.476	0.634
또래와의 관계	0.005	0.041	0.115	0.908
교사와의 관계	-0.012	0.045	-0.275	0.784

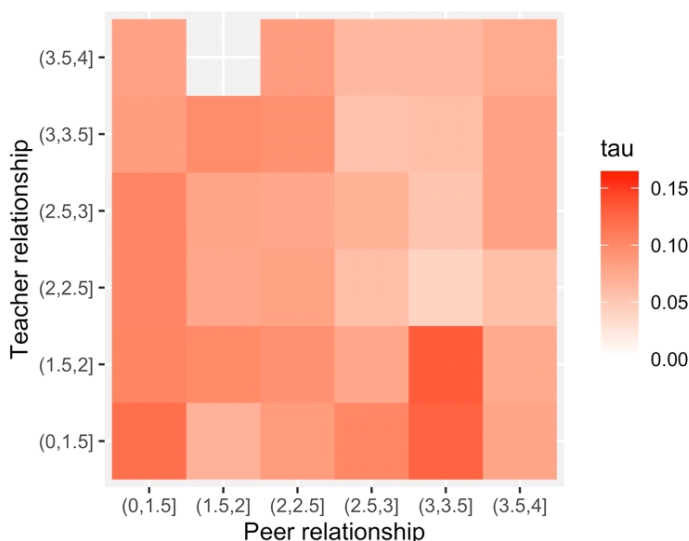
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.



<그림 2> 성별에 따른 처치효과 분포

성별에 따른 처치효과 분포를 그래프로 살펴보면 <그림 2>와 같다. <그림 2>에서는 학생들의 성별에 따른 처치효과 분포를 보여주고 빨간 점은 평균값을 나타낸다. 성별에 따라서 처치효과 분포에 큰 차이가 나타나지 않았고 중앙값 및 평균값도 유사하게 나타났다.

또래와의 관계 및 교사와의 관계에 따른 처치효과 분포도 <그림 3>와 같은 히트맵(Heatmap)을 통해 살펴보고, 효과적인 시각화를 위해서 또래와의 관계, 교사와의 관계 변수는 6개 수준으로(1-4점으로 측정된 점수를 0.5점 단위로 구분) 구분하여 제시하였다. <그림 3>에는 또래와의 관계, 교사와의 관계에 따른 각 조건에서의 처치효과 평균값(τ)의 크기가 클수록 더 짙은 색으로 표현되어 있다.⁴⁾ 또래와의 관계 및 교사와의 관계에 따라서 처치효과가 뚜렷하게 다른 패턴을 보이지는 않음을 확인할 수 있다.



<그림 3> 또래와의 관계 및 교사와의 관계에 따른 처치효과 히트맵

4) 또래와의 관계 점수가 1.5점 초과 2점 이하((1.5, 2])이면서 교사와의 관계 점수가 3.5점 초과 4점 이하((3.5, 4])에 속하는 학생은 데이터상에 없었기 때문에, 해당 조건의 처치효과 평균값은 계산되지 않았고 <그림 3>에는 회색으로 표현되어 있다.

<인과 포레스트를 활용한 차별적 처치효과 분석 R코드>

차별적 처치효과 분석

CATE <- best_linear_projection(out, A=차별적 효과 탐색을 위한 공변인(들))

V. 결론 및 논의

이 연구는 사회과학 조사연구에서 처치효과 검증에 활용할 수 있는 머신러닝 기반의 인과 포레스트 기법을 소개하고, 이를 실제 데이터에 적용하여 교내 동아리활동이 협업능력에 미치는 효과를 분석하였다. Rubin(1974)의 인과추론 관점에서 교내 동아리활동이 협업능력에 미치는 효과란 ‘각 학생이 교내 동아리활동에 참여했을 때 보이게 될 협업능력에 대한 잠재적 결과와, 참여하지 않았을 때 보이게 될 협업능력에 대한 잠재적인 결과의 차이($Y_i(Z_i = 1) - Y_i(Z_i = 0)$)’로 정의될 수 있다. 또한 평균 처치효과는 ‘학생들이 교내 동아리활동에 참여했을 때 보이게 될 평균적인 협업능력과 참여하지 않았을 때 보이게 될 평균적인 협업능력의 차이($E(Y_i(Z_i = 1) - Y_i(Z_i = 0))$)’로 정의된다. 본 연구에서는 교내 동아리활동의 평균 처치효과를 추정하였고, 또한 개별 처치효과 추정치를 활용해 이질적인 효과도 탐색하였다.

분석의 결과 교내 동아리활동이 협업능력에 미치는 평균 처치효과(average treatment effect)가 통계적으로 유의하게 나타났다. 즉 교내 동아리활동에 참여했을 때에 참여하지 않았을 때보다 협업능력이 더 높게 기대되었고, 그 차이가 통계적으로 유의하였다. 이러한 결과는 기존에 동아리활동의 효과를 분석한 선행연구들(김아름·정혜원 2020; 김한솔·장여옥 2019; 이진실 2021; 전자배 2020)과 일치하는 결과로 볼 수 있다. 교내 동아리활동이 교과교육 중심의 학교교육을 보완하여 청소년 시기의 사회적 발달에 기여할 수 있다는 점에서 고무적인 결과이다. 효과크기 관점에서 살펴봤을 때에는, Cohen(1977)이 제시한 표준화된 평균차이(d)를 기준으로 효과크기 d 에 대한 추정치가 약 0.141 정도로⁹⁾ 작은 효과크기(small effect size=0.2)보

다 작은 값을 보였다. 청소년의 전인적 발달을 위해 정책적으로 동아리활동 참여를 독려하고 있는데, 참여 자체를 권장하는 것을 넘어서서 동아리활동에서의 경험이 협업능력과 같은 사회적 발달에 실제적인 효과를 가질 수 있도록 교육적 지원방안을 지속적으로 고민할 필요성이 있다고 볼 수 있다.

또한 다음으로 성별, 또래관계, 교사관계에 따라 교내 동아리활동이 협업능력에 미치는 효과가 다르게 나타나는지(heterogeneous treatment effects) 분석했을 때, 통계적으로 유의한 결과는 발견되지 않았다. 동아리활동이 공동체 의식에 미치는 영향이 성별에 따라 다르게 나타나지 않은 선행연구(김경미 2020)와 일부 일치하는 결과로 볼 수 있다. 동아리활동의 차별적 효과에 대해서는 아직까지 연구가 많이 이루어지지 않았는데, 본 연구에서도 데이터의 한계로 인해 교사 및 학교 변인들을 더 다양하게 투입하여 탐색하지는 못하였다는 점이 있다. 추후 동아리활동이 협업능력에 미치는 효과를 더 극대화할 수 있는 요인들에 대한 연구들이 지속적으로 이루어진다면, 실증적인 분석결과를 바탕으로 실제적인 시사점을 제공할 수 있을 것이라 기대된다. 또한 이러한 맥락에서 평균 처치효과뿐만 아니라 이질적인 처치효과에 대한 추정도 가능한 인과 포레스트를 활용한 기법이 실제 연구에서 유용하게 활용될 수 있을 것이다.

본 연구에서는 머신러닝 기반의 인과 포레스트 기법을 활용하여 동아리활동의 효과를 분석하였는데, 로지스틱 모형의 모수기반 접근을 통해 경향점수를 추정하여 선형 회귀분석을 활용해 처치효과를 분석한 결과는 부록의 <별표 1>, <별표 2>에 제시하였다. 결과를 비교했을 때, 경향점수 방법과 인과 포레스트 방법의 평균 처치효과 추정치는 매우 유사한 값을 보였다. 모의실험을 통해 인과 포레스트 방법과 경향점수 방법을 비교한 Suk et al.(2020a)은 연구자가 불충분한 사전 지식으로 공변인 간의 상호작용 항과 같은 고차항을 모형에 설정하지 못했을 경우 인과 포레스트가 더 정확한 효과 추정치를 산출한다는 것을 보여주었다. 본 연구의 경험적 결과만으로 두 방법의 효과성을 비교하기에는 한계가 있지만, 경향점수 방법에서 연구자가 설정한 함수적 관계가 적절하다면 인과 포레스트와 경향점수 방법이 유사한 결과를 보일 것임을 생각할 수 있다. 반면, 처치 맥락에 대한 사전 지식이 풍부하지 않거나 공변인들 간의 복잡한 관계가 예상될 경우에는 인과 포레스트 방법을 보다 적극적으로 활용할 필요가 있다. 머신러닝 기반의 인과추론에서도 분석에 투입할

5) 효과크기 d (표준화된 평균차이)에 대한 추정치는 두 집단의 결과변수 차이에 대한 추정치인 0.069를 결과변수 표준편차인 0.49로 나누어주어 계산되었다.

공변인들을 선택하는 과정에서 연구자의 이론적 지식은 여전히 중요한 역할을 한다. 하지만 변수들 간의 관계를 설정하는 과정에서 연구자가 사전 지식이 충분하지 않을 때 머신러닝 기반 인과추론 방법은 이를 보완하여 연구자의 모형설정 오류로 인한 편의를 줄여줄 수 있다는 장점이 있다. 변수들의 관계가 데이터상에서 보다 복잡한 양상을 가지고 있어 선형함수와 같은 간단한 함수적 관계로 표현되지 않을 때, 머신러닝 기반의 인과추론 기법의 유용성이 더욱 커진다고 볼 수 있다.

이 연구는 국내 연구자들에게 아직까지 생소한 인과 포레스트 기법을 소개하고, 교내 동아리활동이 협업능력에 미치는 효과를 분석하는 구체적인 연구 상황에 이 기법을 적용해 처치효과를 추정했다는 점에서 의의가 있다. 인과 포레스트 기법은 인간을 대상으로 하는 무선헌당미 어려운 연구 상황에서 처치의 인과적인 효과를 추정하는 데 유용하게 활용될 수 있으며, 데이터 기반의 실증 분석을 통해 정책적 시사점을 얻는 데에도 활용될 수 있을 것이다. 본 연구의 한계점으로는 데이터 내 변수들의 한계로 인해 다양한 학교수준 변수들을 고려하지는 못하였다는 점이 있다. 또한 본 연구에서는 통계적으로 유의한 차별적 효과를 발견하지는 못하였고 데이터상의 한계로 인해 차별적 효과를 보다 다양하게 탐색하지는 못하였는데, 추후 연구에서는 학교특성, 동아리활동 유형, 학교급, 참여 시점 등에 따른 차별적 효과를 보다 심층적으로 분석한다면 학교 현장에 구체적인 시사점을 제공할 수 있을 것으로 생각된다. 또한 학교 내에 학생들이 내재되어 있는 데이터 구조(nested/multi-level/clustered data structure)를 분석에 반영하지는 못했는데, 최근에는 인과 포레스트 기법을 다층으로 확장하려는 시도들이 방법론 연구에서 이루어지고 있으므로(Suk et al. 2020a; Suk & Kang 2021, in press) 추후 연구에서는 이를 반영한 방법을 고려해 볼 수 있을 것으로 생각된다.

참고문헌

- 관계부처 합동. 2018. 제6차 청소년정책기본계획(2018~2022).
http://www.mogef.go.kr/mp/pcd/mp_pcd_s001d.do;jsessionid=I75x4nZhCUdOWZGlZytEboqS.mogefl1?mid=plc502&bbtSn=704748 (2021.10.8. 인출).
 교육과학기술부. 2009. 《초·중·고 창의적 체험활동 교육과정 해설》. 교육과학기술부.

- 교육부. 2015a. 《창의적 체험활동 교육과정(안전한 생활 포함)》. 교육부.
- 교육부. 2015d. ‘학생의 꿈과 끼를 키워 행복교육을 실현하는 중학교 자유학기제 시행 계획(안)’. (보도자료 2015. 11. 25).
- 교육부·충청남도교육청·한국청소년동아리연맹. 2013. 《동아리활동 길라잡이: 중·고 통합용》. 교육부.
- 김경미. 2020. “참여하는 청소년은 행복한가? 사회참여 활동유형과 성별에 따른 공동체 의식의 매개효과.” 《사회과학연구》 46(1): 141-164.
- 김나영·황혜영. 2019. “경향점수를 활용한 교내 동아리 활동 참여가 고등학생의 학업 및 사회적 자아개념에 미치는 영향 분석.” 《교육혁신연구》 29(3): 291-308.
- 김도희. 2020. “아동이 지각한 부모양육태도와 또래관계의 관계에서 자아존중감과 협동심의 직렬다중매개효과.” 《Journal of Digital Convergence》 18(5): 485-495.
- 김아름·정혜원. 2020. “청소년의 창의적 체험활동 참여가 협업능력에 미치는 영향.” 《학습자중심교과교육연구》 20(16): 691-708.
- 김영서·홍세희. 2021. “고등학생의 청소년활동 참여에 따른 잠재프로파일 유형과 공동체 의식 및 주관적 행복감의 차이.” 《조사연구》 22(1): 91-125.
- 김정주·김용대·성기원. 2003. 《지역사회에서 청소년 동아리 활동 실태와 지원 방안 연구》. 한국청소년개발원.
- 김준엽·정혜경·Seltzer, M.H. 2008. “Drawing Causal Inferences Using Propensity Score Methods in Educational Research.” 《교육평가연구》 21(3): 219-242.
- 김한솔·장여옥. 2019. “자율동아리활동 참여특성에 따른 청소년활동 핵심역량 차이분석.” 《한국청소년활동연구》 5(3): 1-22.
- 박경미·박한솔·최연정. 2020. “청소년활동의 효과에 대한 유형화 분석.” 《청소년활동연구》 6(4): 95-121.
- 심우정·백선희. 2020. “중학생의 동아리 활동 참여 수준 및 참여 동아리 유형 관련 요인.” 《아시아교육연구》 21(2): 433-462.
- 이승렬. 2013. “청소년 동아리 조직유형별 활동 참여의 결정요인 분석.” 《청소년학연구》 20(10): 407-431.
- 이승희·신소영. 2018. “의사소통 및 협업능력 관련 영향요인 간 관계 구조분석.” 《학습자중심교과교육연구》 18(20): 203-222.
- 이진실. 2021. “매개효과 분석에서 오차 간 상관을 고려한 경향점수 역확률가중치 활용방법.” 《Journal of the Korean Data Analysis Society》 23(2): 971-984.
- 전란영. 2020. “부모의 자율성 지지와 구조제공이 아동의 그림을 매개로 동의 협동심에 미치는 영향.” 《학습자중심교과교육연구》 20(11): 117-139.
- 전자배. 2020. “청소년의 동아리활동이 협업능력에 미치는 영향: 자아유능감의 매개를 중심으로.” 《한국청소년연구》 31(1): 101-127.

- 조영희·백아롱. 2021. “중학생의 자아존중감과 협동심 간의 관계에서 부모 및 교사관계의 매개효과: 성별 차이를 중심으로” 《교육학연구》 59(5): 185-213.
- 한국청소년활동진흥원. 2018. “청소년활동을 통해 키울 수 있는 핵심역량은?” (보도자료 2018.12.27).
- 황여정. 2020. “청소년활동 참여 경험이 중학생의 갈등해결역량에 미치는 영향.” 《학습자중심교과교육연구》 20(11): 1097-1125.

- Athey, S., and Imbens, G. 2016. “Recursive Partitioning for Heterogeneous Causal Effects.” *Proceedings of the National Academy of Sciences* 113(27): 7353- 7360.
- Athey, S., J. Tibshirani, and S. Wager. 2019. “Generalized Random Forests.” *The Annals of Statistics* 47(2): 1148-1178.
- Athey, S. and S. Wager. 2019. “Estimating Treatment Effects with Causal Forests: An Application.” *Observational Studies* 5(2): 37-51.
- Austin, P.C. 2011. “An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies.” *Multivariate Behavioral Research* 46: 399-424.
- Cohen, J. 1977. *Statistical Power Analysis for the Behavioral Sciences*(2nd ed.). NY: Academic Press.
- Cook, T.D., D.T. Campbell, and W. Shadish. 2002. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.
- Davis, J. and S.B. Heller. 2017. “Using Causal Forests to Predict Treatment Heterogeneity: An Application to Summer Jobs.” *American Economic Review* 107(5): 546-550.
- Dorie, V., J. Hill, U. Shalit, M. Scott, and D. Cervone. 2019. “Automated versus Do-it-yourself Methods for Causal Inference: Lessons Learned from a Data Analysis Competition.” *Statistical Science* 34(1): 43-68.
- Hastie, T., R. Tibshirani, and J. Friedman. 2017. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Springer.
- Hill, J.L. 2011. “Bayesian Nonparametric Modeling for Causal Inference.” *Journal of Computational and Graphical Statistics* 20(1): 217-240.
- Imai, K. and M. Ratkovic. 2013. “Estimating Treatment Effect Heterogeneity in Randomized Program Evaluation.” *The Annals of Applied Statistics* 7(1): 443-470.
- Künzel, S.R., J.S. Sekhon, P.J. Bickel, and B. Yu. 2019. “Metalearners for Estimating Heterogeneous Treatment Effects Using Machine Learning.” *Proceedings of the National Academy of Sciences* 116(10): 4156-4165.
- Rosenbaum, P.R. and D.B. Rubin. 1983. “The Central Role of the Propensity Score in

- Observational Studies for Causal Effects.” *Biometrika* 70: 41-55.
- Rubin, D.B. 1974. “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies.” *Journal of Educational Psychology* 66: 688-701.
- Steiner, P.M. and D. Cook. 2013. “Matching and Propensity Scores.” in T. Little (ed.), *The Oxford Handbook of Quantitative Methods*: 236-258. New York, NY: Oxford University Press.
- Schafer, J.L. and J. Kang. 2008. “Average Causal Effects from Nonrandomized Studies: A Practical Guide and Simulated Example.” *Psychological Methods* 13(4): 279-313.
- Suk, Y. and H. Kang. 2021. “Robust Machine Learning for Treatment Effects in Multilevel Observational Studies under Cluster-level Unmeasured Confounding.” *Psychometrika* 0(0): 1-34.
- Suk, Y. and H. Kang. (in press). “Tuning Random Forests for Causal Inference under Cluster-level Unmeasured Confounding.” *Multivariate Behavioral Research*.
- Suk, Y., H. Kang, and J.-S. Kim. 2020a. “Random Forests Approach for Causal Inference with Clustered Observational Data.” *Multivariate Behavioral Research* 0(0): 1-24.
- Suk, Y., J.-S. Kim, and H. Kang. 2020b. “Hybridizing Machine Learning Methods and Finite Mixture Models for Estimating Heterogeneous Treatment Effects in Latent Classes.” *Journal of Educational and Behavioral Statistics* 46(3): 323-347.
- Tibshirani, J., S. Athey, E. Sverdrup, and S. Wager. 2021. “grf: Generalized Random Forests. R package version 2.0.2.”
<https://CRAN.R-project.org/package=grf>
- van der Laan, M.J. and D. Rubin. 2006. “Targeted Maximum Likelihood Learning.” *The International Journal of Biostatistics* 2(1): 1-38.
- Wager, S. and S. Athey. 2018. “Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests.” *Journal of the American Statistical Association* 113(523): 1228-1242.

Evaluating the Effects of School Club Activities on Collaborative Competency Using Random Forests for Causal Inference

Youmi Suk

(School of Data Science, University of Virginia)

Jinsil Lee

(Seoul National University)

Recently, there has been a growing interest in using machine learning methods for causal inference to estimate the average treatment effect and the conditional average treatment effects. The purposes of this paper are to introduce machine learning methods for causal inference and to investigate the effects of school club activities on students' collaborative competency using one particular machine learning method based on random forests, called 'casual forests'. Specifically, this paper uses causal forests to estimate the average effect and heterogeneous effects of school club activities on students' collaborative competency. As a result, we found that the average effect of school club participation is significantly positive, and this finding highlights the importance of school club activities as a complement to subject-centered curriculum to enhance students' social development. Our additional analysis found that there are no significant, heterogeneous treatment effects depending on gender, peer relationship, and teacher-student relationship.

Key words: machine learning, causal inference, school club activities, collaborative competency

부록

<별표 1> 경향점수 추정을 위한 로지스틱 회귀모형(treatment model) 분석결과

	추정치	표준오차	Wald통계량	p값
절편	-4.466	0.855	-5.221***	<0.001
성별	0.107	0.105	1.023	0.306
사전 협업능력	0.120	0.130	0.919	0.358
학업성취수준	0.123	0.066	1.862	0.063
학업열의	0.098	0.126	0.779	0.436
자아존중감	0.071	0.148	0.481	0.630
우울	0.085	0.129	0.659	0.510
사회적 위축	-0.003	0.083	-0.038	0.970
또래와의 관계	0.087	0.112	0.779	0.436
교사와의 관계	0.154	0.112	1.372	0.170
전년도 동아리활동_만족도 낮음 더미	0.789	0.244	3.242**	0.001
전년도 동아리활동_만족도 높음 더미	0.910	0.106	8.571***	<0.001
월평균 가구소득	0.000	0.000	0.235	0.814
부모의 학력수준	0.019	0.024	0.787	0.431
양육태도_자율성 지지	-0.071	0.133	-0.534	0.594
양육태도_애정	0.020	0.135	0.152	0.880
양육태도_구조제공	-0.144	0.118	-1.219	0.223
기대교육수준_전문대학 졸업 더미	-0.034	0.401	-0.084	0.933
기대교육수준_4년제 대학 졸업 더미	0.162	0.376	0.430	0.667
기대교육수준_대학원 졸업 더미	0.274	0.397	0.690	0.490
기대교육수준_잘 모르겠음 더미	0.336	0.451	0.744	0.457
부모-자녀 진로관련 대화정도	-0.058	0.073	-0.794	0.427
주당 부모-자녀 함께 보내는 시간	-0.005	0.003	-1.690	0.091
어머니 직업여부 더미	-0.246	0.118	-2.081*	0.037
학교소재지 도시 더미	0.014	0.147	0.099	0.921
전년도 교내 동아리활동 참여비율	0.049	0.003	17.816***	<0.001

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

<별표 2> 경향점수 역확률가중치 및 선형 회귀분석을 활용한 평균 처치효과 추정모형
(outcome model) 분석결과

	추정치	표준오차	t통계량	p값
절편	2.919	0.017	170.92***	<0.001
동아리활동 참여 더미	0.069	0.022	3.19**	0.001

$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.