

머신러닝을 활용한 하향취업의 결정요인 분석: 자아존중감을 중심으로*

경윤영** 신범철*** 이영석**** 이태현*****

| 목 차 |

I. 서론	IV. 실증분석 결과
II. 선행연구	V. 결론 및 논의
III. 실증분석	

| 논문요약 |

본 연구는 머신러닝의 랜덤포레스트 모델을 사용하여 하향취업 여부를 예측하였다. 청년층의 직업 선택에 관한 정보를 얻고자 청년 패널조사(YP2007) 5차부터 12차 자료를 이용하였다. 또한, 랜덤포레스트 분석에 사용한 예측인자는 취업자의 객관적 정보인 인적특성, 직업속성, 구직활동 경험, 인적자본 변수뿐만 아니라 주관적인 평가가 들어가는 직업선택 시 중요도, 자아존중감의 변수도 포함하여 총 84개의 예측인자를 사용하였다.

주요 결과는 다음과 같다. 첫째, 랜덤포레스트 모델을 사용하였을 때 하향취업 여부의 기준을 교육수준을 나눈 모형1의 정확도, 민감도 및 특이도는 약 94%, 91%, 96%였으며 기술수준으로 나눈 모형2의 정확도, 민감도 및 특이도는 약 93%, 92%, 95%로 측정되어 두 모형 모두 예측 성과가 매우 높게 나타났다. 둘째, 랜덤포레스트 모델을 통해 예측인자의 변수 중요도를 살펴본 결과 모형1,

* 본 논문은 2020년도 고용패널조사 학술대회(2020년 10월 6일)에서 발표한 논문을 수정 보완한 것임.

** 제1저자: 성균관대학교 경제학과 석사 졸업

*** 교신저자: 경기대학교 경제학부 교수

**** 공동저자: 경기연구원 연구원

***** 공동저자: 고려대학교 심리학부 박사과정

모형2 모두 월평균 임금, 종사하는 산업, 회사소재지 등의 직장속성 변수가 주요 변수로 채택되어 해당 변수들이 하향취업에 중요한 영향을 미치는 것으로 나타났으며 주관적 요인인 자아존중감 변수 역시 예측에 중요한 변수 중 하나임을 확인하였다.

이러한 연구 결과는 하향취업 여부를 예측하는 것에 있어 자아존중감이 하나의 중요 변수로 인식되기에 청소년기에 자아존중감 향상을 위한 체계적인 정책 및 교육과정을 마련할 필요가 있음을 시사한다. 또한 자아존중감을 높여야 하는 교육대상뿐만 아니라 부모, 교사 등 주변 인물들에 대한 교육도 함께 진행되어야 함을 의미한다.

▪ 주제어: 하향취업, 자아존중감, 머신러닝, 랜덤포레스트, 예측모형

I. 서론

하향취업은 인적자본의 비효율적 활용과 노동시장의 불안정성을 야기한다는 점에서 중요한 문제이다. 하향취업자의 낮은 직무만족도와 높은 임금손실 규모는 근로의욕을 저하시켜 생산성 저하, 잦은 이직 등으로 이어져 심각한 사회적 비용이 발생하게 된다. 이는 인적자본을 비효율적으로 활용하게 할 뿐만 아니라 노동시장의 불안정성을 높여 사회·경제적으로 부정적인 영향을 미치게 된다(Bender & Heywood 2006; 진선미·공정식 2011; 남성일·전재식 2011; 김종성 외 2012; 김기홍 2018; 오삼일·강달현 2019).

오삼일·강달현(2019)에 따르면 2000년대 이후 하향취업률¹⁾이 지속적으로 상승하고 있으며, 최근에는 대졸 취업자 중 약 30%가 하향취업을 하고 있는 것으로 나타났다. 적정 취업자 대비 하향취업자의 임금손실 규모는 약 36%로 나타나고 있으며, 하향취업자의 적정 취업 전환율은 1년 후 4.6%, 2년 후 8%, 3년 후 11.1%로 나타난다. 즉, 하향취업을 한 취업자들은 손쉽게 적정 취업으로 이행하지 못하는 상황이다.

1) 오삼일·강달현(2019)은 하향취업률을 대졸 취업자 수 대비 하향취업자 수로 정의하였다.

하향취업 문제가 심각해지면서 이와 관련된 다양한 주제로 연구되어 왔다(Bender & Heywood 2006; 진선미·공정식 2011; 남성일·전재식 2011; 김종성 외 2012; 김기홍 2018; 오삼일·강달현 2019). 인구통계학적 변수, 직업 속성 변수, 취업 준비성과 등 객관적으로 측정 가능한 변수들을 포함하여 분석되었다. 특히, 자아존중감은 고용 가능성 또는 취업 준비에 직·간접적으로 영향을 미칠 수 있다는 연구가 진행되어 왔다(Korman 1976; Ellis & Taylor 1983; Thunholm 2004; Salemela-Aro & Nurmi 2007). 그럼에도 불구하고 자아존중감은 그동안 하향취업을 설명하는 변수로서 충분히 고려되지 못하였다.

따라서 본 연구는 청년 패널조사(YP2007) 5차부터 12차 자료를 기반으로 머신러닝의 랜덤포레스트 방법을 활용하여 하향취업 여부를 예측성과를 측정하고 예측인자 중 중요한 변수가 무엇인지 살펴보고자 한다. 또한, 머신러닝은 전통적 회귀분석과 달리 설명변수의 외생성과 오차항의 분포를 전제하지 않기 때문에 변수에 관한 제약에서 상대적으로 자유롭고 주관적 변수들도 고려가 가능하다는 장점을 가지고 있다(최필선·민인식 2018).²⁾ 또한, 랜덤포레스트는 설명변수가 많고 설명변수 간 상호작용이 복잡한 자료에도 적용이 가능하며 예측력과 안정성 또한 높다(Cutler et al. 2007; Strobl et al. 2009). 그렇기에 본 연구는 변수들 간의 상호작용이 존재하는 자아존중감이 라는 주관적 변수의 중요성을 고려하여 머신러닝의 랜덤포레스트 기법을 활용하여 연구를 진행하였다. 본 연구의 구성은 다음과 같다. II장에서는 하향취업의 이론적 배경과 하향취업과 자아존중감에 대한 선행 연구들을 살펴본다. III장에서는 본 연구에서 사용하는 데이터 및 분석방법을 제시한다. IV장에서는 분석 결과를 제시하고 V장에서는 분석 결과를 바탕으로 시사점을 제시하고자 한다.

2) 최필선·민인식(2018)은 랜덤포레스트를 활용하여 대졸자의 취업 예측 모형을 비교 분석하였는데, 주관적 판단 변수를 포함한 모형이 취업 성과와 취업의 질을 더 잘 예측한다고 주장한다.

II. 선행연구

1. 하향취업과 자아존중감

하향취업은 노동자의 교육수준과 직무에서 요구하는 교육수준의 불일치를 의미하는 일자리-학력 불일치(education-job mismatch)는 노동시장의 비효율성을 야기한다는 측면에서 중요하게 인식되어 왔다(김기홍 2018; 오삼일·강달현 2019). 국내의 경우 1990년대 이후 지속적인 학력의 상향 평준화와 그에 따른 고학력자의 과잉공급 현상³⁾이 나타나면서 하향취업 연구가 활발히 진행되었다(오삼일·강달현 2019; 조민수 외 2014; 김기홍 2018). 신선미·손유미(2008)는 하향취업을 노동시장의 구조적 문제를 강조하는 개념으로 보았으며, 남성일·전재식(2011)은 하향취업을 노동 공급자의 측면에서 인적자본의 비효율적인 활용 문제로 접근하였다.

국내외의 연구에서 하향취업의 원인을 설명하는 다양한 이론들이 제시되었다(Nauze-Fichet & Tomasini 2002; 김종성 외 2012). 주요 이론으로는 인적자본 이론, 직업탐색이론, 신호 이론, 직무 이동 이론, 직무 경쟁이론 등이 있다(Nauze-Fichet & Tomasini 2002). 박천수(2005)는 직무 경쟁이론을 국내 노동시장의 구조적 현상을 가장 잘 설명해 주는 이론으로 보았다. 이유는 하향취업을 노동시장의 수요와 공급의 불일치에 의한 일시적 현상으로 보는 인적자본 이론(Freeman 1976)과 달리 직무 경쟁이론은 노동시장의 구조적인 측면에서 하향취업을 설명하기 때문이다(Thurow 1975; 이은수 2017).

직무 경쟁이론(Thurow 1975)에 따르면 직무기술은 실제 업무 과정을 통해서만 터득할 수 있으며, 이는 직업훈련과 연관되어 있다고 가정한다. 따라서 고용주는 직업훈련 비용이 적게 드는 근로자를 선호하기에 훈련 가능성 및 직무 적응성을 고려하여 구직자를 채용하게 된다. 또한, 특정 시점에서 일자리의 양과 질은 이미 결정되어 있으며, 구직자들은 정해진 일자리에 줄을 서서 기다리는 것으로 본다. 따라서, 구직자는 이미 정해진 일자리(직무)에 대해서 임금수준, 복지, 지속 가능성, 발전 가능성, 워라벨(Work Life

3) 교육부(2009)에 따르면 대학 진학률은 1980년 27.2%에서 2008년 63.5%로 크게 증가하였음(신선미·공정식 2011; 황남희·정주연 2011)

Balance) 등과 같은 개인의 선호를 바탕으로 일자리의 위계를 정하게 되며, 괜찮은 일자리에 취업하기 위해 다른 구직자들과 경쟁하게 된다. 구직자들은 괜찮은 일자리를 얻기 위해 경쟁자들보다 상대적 역량이 높아야 하므로 많은 사람들이 선호하는 직무일수록 필요로 하는 역량 수준은 높아지게 된다(박천수 2005; 이영석·송선혜 2020). 그렇기에 일자리의 양과 질이 이미 정해져 있는 노동시장에서 괜찮은 일자리를 얻기 위해 구직자들 간의 경쟁이 과열되고, 이는 고학력자의 과잉 공급 현상을 불러일으키며 하향취업을 유발하게 된다.

한편, 자아존중감은 개인의 삶 전반에 큰 영향을 미친다는 관점에서 많은 연구들이 이루어지고 있다(Josephs et al. 1992; Emeler 2001; Bonanno et al. 2002; Baumeister et al. 2003; Howe et al. 2004). 자아존중감(self-esteem)이란 자신이 사랑받을 만한 가치가 있는 소중한 존재이고 어떤 성과를 이루어낼 만한 유능한 사람이라고 믿는 것으로 정의된다(Maslow 1989; Maslow 2012; McLeod 2007). 자아존중감은 사람에 따라 다르지만 일반적으로 성인이 되기 전에 형성되며 만들어지는 과정에서 관계적인 요소와 성격적인 요소가 함께 작용하게 된다. 즉, 자아존중감은 영·유아기에서부터 청소년기 사이에 사회 환경에 긴밀하게 영향을 받아 자신에 대한 믿음을 내재화시키는 과정을 통해 형성된다고 본다(Crocker & Park 2003; Crocker & Park 2004). 또한 한번 형성된 자아존중감은 쉽게 변하지 않아 개인의 특성들 중 안정적인 성향을 보인다(Caspi et al. 1998; Deaux & Snyder 2012).

자아존중감과 관련한 선행연구들을 살펴보면 자아존중감은 부정적 정서, 대인관계, 학업 성취, 문제해결능력, 진로 결정, 사회성 및 도덕성 등과 관계가 있는 것으로 나타났다. 자아존중감이 가지는 이러한 특성들은 구직자가 노동시장으로 이행하는 과정에 직접적 또는 간접적으로 영향을 미칠 수 있다. 먼저, 자아존중감이 낮은 사람에게서 나타나는 부정적인 정서들은 하향취업에 직접적으로 영향을 미칠 수 있다. 자아존중감이 낮은 사람들은 자신의 역량을 저평가하여 이를 충분히 사용하지 못하는 경향이 있다(Baumeister et al. 2003). 또한 자아존중감이 낮은 사람은 강한 방어기제를 보이며 위험 회피 성향을 띠게 된다(Alves-Martins et al. 2002; Di Paula & Campbell 2002). 위험회피적인 성향으로 인해 자신의 역량에 대한 낮은 기대감으로 자신의 역량보다 낮은

역량을 요구하는 직무, 직업을 선택하게 해 하향취업으로 이어지게 된다.

다음으로, 자아존중감이 하향취업에 간접적으로 영향을 미치는 경로이다. 자아존중감은 학업성취, 자기 주도학습력, 학업에 대한 효능감, 학습 동기 향상 등 학업과 관련 요인들(박선영 2003)과 의사소통 능력, 대인관계 및 직장 적응력과 같은 사회성(조현재 2017; 이소영·이경순 2019), 문제해결능력(양혜주 2013), 진로결정(이문수·박연복 2019), 도덕성(이난 2018)과 긍정적인 관련성이 있는 것으로 나타나고 있다. 이러한 요인들은 구직자의 채용을 결정짓는 주요 요인인 졸업 시점, 직무적합성, 출신학교, 직업기초능력, 직장적응력, 도덕성/인성 등(채창균·양정승 2019)에 영향을 미칠 수 있다. 더 나아가 낮은 자아존중감을 가진 사람들은 우울, 긴장(Arndt & Goldenberg 2002), 스트레스에 노출될 확률이 높다. 이러한 불안정한 심리상태는 취업 과정 동안 부정적인 영향을 주어 하향취업으로 이어질 수도 있고(Taylor & Armor 1996), 장기 미취업자가 되는 등 낮은 취업 성과(Harter 1993)를 얻게 될 수도 있다.

2. 머신러닝 랜덤포레스트 기법

머신러닝은 방대한 양의 데이터를 컴퓨터 스스로가 익혀 분석하는 컴퓨터 알고리즘으로 과거의 데이터를 학습시킨 특정 모델을 활용하여 미래를 예측(prediction)할 때 주로 사용된다. 최근 사회과학 분야에서 빅데이터를 활용한 연구가 활발히 진행되면서, 연구에 알맞은 머신러닝 모델의 필요성과 변수 선택의 문제가 중요해졌다(Varian 2014). 국내의 사회과학 분야에서 예측 모형 중 랜덤포레스트를 사용한 연구로는 유진은(2015), 최필선·민인식(2018), 이은정 외(2020) 등이 있다.

전통적인 경제 분석 모델은 모수를 추정하여 가설을 검증하는 방식이며 다변량 분석의 경우 독립변수의 외생성, 오차항 분포등과 같은 제약이 있으며 다중공선성이 일어난 가능성도 배제할 수 없다. 반면 머신러닝 기법은 전체 모집단 자료와 같은 방대한 데이터를 활용할 수 있으며, 변수의 수가 많은 경우에도 모델을 사용할 수 있는 장점이 있다(Varian 2014).

의사결정나무(decision tree)의 경우 마디(node)에서 가지치기(pruning)을 할 때 한 변수만을 사용하기 때문에 모형의 안정성이 떨어지는 단점이 있다.

하지만 랜덤포레스트는 앙상블 기법을 사용하여 의사결정나무의 불안정성을 개선시킨 모델로 무작위성을 활용하여 부트스트랩 표본을 생성한다. 즉, 무작위로 의사결정나무를 생성, 이를 종합하여 랜덤포레스트 모델을 만드는 것을 통해 의사결정나무의 분류 정확성을 개선, 데이터를 통해 예측 결과가 달라지는 불안정성을 해소할 수 있다(유진은 2015).

Ⅲ. 실증분석

1. 분석방법

본 연구에서는 실증분석을 위해 청년 패널조사(YP2007)자료를 사용한다. 청년 패널조사는 고용노동부와 한국고용정보원에서 청년층의 직업 선택 및 노동시장 이동에 관한 실태를 파악하고자 실시한 패널조사로서 청년층의 취업 형태, 직업이동 과정, 학교생활, 직업교육훈련, 자격증 등을 파악하고 있다. 또한, 분석방법으로는 머신러닝 기법 중 하나인 랜덤포레스트를 사용하여 하향취업의 가능성을 예측하고 중요도 지수를 통해 하향취업에 영향을 미치는 예측인자를 찾아낸다.

(1) 랜덤포레스트

랜덤포레스트는 분류 분석⁴⁾의 한 종류로 의사결정나무가 발전된 형태의 모델이다. 즉, 랜덤포레스트는 다수의 의사결정나무들로 구성되는 모형이며 이를 이해하기 위해서는 의사결정나무에 대한 이해가 선행되어야 한다(유진은 2015; 최필선·민인식 2018). 의사결정나무는 설명변수의 중요도 기준에 따라 가지(branch)를 뺀어 나가는데, 각 마디(node)에서 가지가 나뉘질 때 하나의 변수만을 사용한다(유진은 2015). 이때 가장 먼저 사용되는 변수는 가지가 나뉘질 때마다 사용되기 때문에 가장 자주 사용되는 중요 변수가 된

4) 분류 분석이란 반응 변수와 설명변수로 구성된 데이터를 학습시키고, 새로운 관측 값이 주어졌을 때 훈련된 모형을 바탕으로 분류 방향을 예측하는 것을 의미한다.

다. 중요 변수를 선택하는 방법으로는 지니계수(Gini index) 또는 정보 획득량 등의 불순도 측정 도구가 있는데(Pal 2005; 민인식·최필선 2018), 랜덤포레스트의 경우 불순도 측정도구로 지니계수(Gini index)⁵⁾를 사용한다.

의사결정나무는 인접한 상위와 하위 노드 변수 간의 상호작용 효과를 설명할 수 있는 장점이 있다. 하지만 비슷한 예측력을 가지고 있는 설명변수들이 여러 개 있을 경우 한 변수만 선택이 되기 때문에 데이터가 유사함에도 전혀 다른 모형이 도출될 수 있다는 한계가 있다(박창이 외 2013; 최필선·민인식 2018; Pal 2005). 반면, 랜덤포레스트의 경우 배깅(bagging, bootstrap aggregating)을 사용하여 설명변수가 많고 그간의 상호작용이 복잡하여도 모형의 예측력이 높고 안정성이 있다. 배깅은 부트스트랩(bootstrap)방법을 적용한 앙상블 기법으로 랜덤포레스트는 학습 단계에서 부트스트랩 기법을 통해 훈련용 데이터를 구성한 후 설명변수를 무작위로 선택하여 의사결정 나무의 생성을 반복하기 때문에 예측 오차를 줄일 수 있다(Bremian 1996; Bremian 2001). 또한 랜덤포레스트에서 제공하는 중요도 지수를 통해 설명변수에 대해 깊이 있는 이해가 가능하다(유진은 2015).

2. 데이터 및 변수

본 연구는 청년 패널조사(YP2007) 5차부터 12차 자료를 통합(pooled)하여 분석을 수행하였다. 청년 패널은 만 15세에서 29세 청년을 대상으로 학교생활, 구직활동, 일자리 경험, 자아존중감, 취업 준비활동 등 다양한 내용을 조사하고 있기에 다양한 설명변수를 사용하는 랜덤포레스트를 사용하는 것에 이점이 있다. 특히 청년 패널조사를 사용하여 하향취업 여부를 주관적 기준으로 나눌 수 있으며 하향취업 상태의 변화를 파악할 수 있다. 본 연구에서는 지난조사 당시 학생 혹은 미취업자들이 해당 조사 때 전문 대학이상을 졸업한 임금근로자들이 동일직장에서 근무한 사람들을 표본으로 선정, 어떤 요인으로 하향취업이 결정되는지 밝히고자 한다.

5) 지니계수(Gini index)는 $\sum_{j \neq i} (f(C_i, T) / \pi) (f(C_j, T) / \pi)$ 이며 $f(C_i, T) / \pi$ 는 클래스 C_i 에 속한 케이스가 선택될 확률이다(Pal 2005).

(1) 종속변수

하향취업과 관련한 연구들이 활발하게 이루어지면서 하향취업을 측정하는 다양한 평가 방법들이 개발되었다(Nauze-Fichet & Tomasini 2002; 오호영 2005; 신선미·손유미 2008; 남성일·전재식 2011; 진선미·공정석 2011; 김종성 외 2012). 하향취업을 측정하는 방법은 직접적 자기평가·간접적 자기평가와 같은 주관적 측정 방법과, 직무분석 방식·통계적 방식인 객관적 측정 방법이 제시되고 있다(오호영 2005; 남성일·전재식 2011).

주관적 측정 방법 중 직접적 자기평가는 직무에서 요구하는 기술·교육·전공 일치도와 자신의 교육 수준의 일치 정도를 질문하는 방식으로 하향취업을 측정한다. 직접적 자기평가는 조사의 용이성이 크지만 직무 만족도가 설문문에 크게 영향을 미쳐 결과를 왜곡할 가능성이 높다. 한편, 간접적 자기평가는 해당 직무를 수행하는 데 필요한 교육수준을 질문하고 이를 응답자의 교육 수준과 비교하여 측정하는 방식이다. 간접적 자기평가 방식은 직무 불만족에 따른 결과 왜곡을 줄일 수는 있지만, 직무 수행에 필요한 교육수준은 여전히 주관적으로 판단하는 한계가 존재한다.

다음으로 객관적 측정 방법 중 직무분석 방식은 직무분석을 통해 직업이 요구하는 교육수준과 실제 교육수준을 비교하여 측정하는 방식이다. 일반적으로 직업사전에서 직업별로 제시하는 일반교육수준(GED)을 사용하여 측정한다. 직무분석 방식은 측정에 있어서 객관성을 부여한다는 점에서 의미가 있지만, 객관적 기준을 측정하는 GED가 수시로 변화하는 환경을 반영하지 못한다는 점에서 한계가 존재한다. 한편, 통계적 방법은 직업별 교육수준 분포를 분석하여 직무에서 요구되는 교육 수준을 정의한 후 응답자의 교육수준과 비교하여 측정한다. 일반적으로 조사의 규모가 큰 경제활동인구조사 또는 임금구조기본통계조사를 이용하여 표준편차를 고려한 ‘직종별 평균 교육연수’를 직무에서 요구하는 학력으로 정의하여 사용한다(남성일·전재식 2011).

본 연구에서는 주관적 측정 방법인 직접적 자기평가 방식을 사용하였다. 직접적 자기평가 방식은 직무 불만족과 같이 외부 상황에 영향을 받아 과대 추정되어 결과를 왜곡할 수 있다는 단점이 있다. 하지만 모든 직무에 필요한 기술, 교육, 자격요건 등을 자세히 파악할 수 없기 때문에 근로자에게 간

접적으로 파악하는 것이 오히려 효과적일 수 있다(진선미·공정식 2011).

종속변수인 하향취업 여부 변수는 하향취업 혹은 걱정/상향 취업으로 구성된 이항변수이며 설문지에 제시된 교육수준과 기술 수준 두 가지 기준을 사용하여 두 종류의 종속변수를 구성하였다. 청년 패널조사의 응답 질문인 현 직장(일자리)의 업무내용과 수준이 자신의 교육수준(기술 수준)과 비교하여 어떻습니까?의 질문에 ‘(1) 수준이 아주 낮다’, ‘(2) 수준이 낮다’라고 응답한 경우는 0으로 ‘(3) 수준이 적절하다’, ‘(4) 수준이 높다’, ‘(5) 수준이 아주 높다’라고 응답한 경우는 1로 코딩하였다.

현 데이터의 종속 변수 비율로 랜덤포레스트를 실행하면 데이터의 계급 불균형이 발생하여 분류 분석을 통해 집단을 정확하게 분류하기 어렵다. 그렇기에 분류 분석을 진행할 때 랜덤오버샘플링, 랜덤언더샘플링, SMOTE 등의 방법으로 데이터의 계급 불균형을 해소한다. 랜덤오버샘플링은 소수 계급의 표본 수를 무작위로 늘리는 것이고, 랜덤언더샘플링은 다수 계급의 표본수를 무작위로 줄이는 방법이다. SMOTE(Chawla et al. 2002)는 기존의 랜덤오버샘플링에서 표본의 중복으로 인해 생기는 과대적합 문제를 보완하는 방법이다. 소수 계급의 데이터에서 임의의 표본들을 선택하여 표본들 사이에 k 근접 이웃(k nearest neighbors)을 찾아 새로운 표본을 생성한다.

(2) 설명변수

먼저, 본 연구의 랜덤포레스트 분석에서는 총 84개의 설명변수를 사용하였다. 랜덤포레스트는 설명변수가 많고 이들 간의 상호작용이 복잡해도 예측 결과를 얻을 수 있다는 장점을 가진다(유진은 2015; 민인식·최필선 2018). 그 때문에 본 연구에서도 랜덤포레스트를 사용하여 가능한 많은 변수들을 포함시켜 하향취업 여부 예측에 어떤 요인들이 중요하게 작용하는지 살펴보고자 하였다.

랜덤포레스트 분석에 앞서 선행연구들을 바탕으로 인구통계 및 개인 배경 변수, 직업 속성 변수를 포함하여 설명변수를 구성하였다(진선미·손유미 2008; 이찬영 2008; 진선미·민무숙 2010; 황남희·전주연 2011; 조세미 외 2014). 또한 자아존중감을 설명변수로 포함시켜 하향취업 여부 예측에 있어

주요 변수로 선택되는지 살펴보려 하였다. 이와 더불어 직업 선택 시 중요도, 인적자본 변수인 취업 준비 활동, 직업훈련 및 자격과 구직활동 경험 변수(김중성 외 2012)들을 포함시켰다.

<표 1> 랜덤포레스트에 사용된 설명변수

구분	설명변수
인적특성 및 개인배경 (15개)	성별, 생년, 거주지, 혼인상태, 자녀유무, 부채 유무, 최종학력 아버지 직업, 어머니 직업, 아버지 교육수준, 어머니 교육수준, 가구주 여부, 가계비 생활 부담자, 경제적 지원여부, 가구총소득
직업 속성 (19개)	동일직장 여부, 종사자수, 산업코드, 직업코드, 현 직장 소재지, 근속년수, 종사상지위, 정규직여부, 사회보험 부가급여 수혜여부- (국민건강보험, 고용보험, 국민연금, 퇴직금, 상여금, 시간외수당, 유급휴일), 노조유무, 직업만족도, 월평균 임금, 취업자정보
직업선택 시 중요도(13개)	성취, 이타, 개인지향, 경제적 보상, 인정, 신체활동, 직업안정, 다양성, 심신의 안녕, 타인영향, 지적추구, 애국, 자율
인적자본 변수: 취업준비활동 (14개)	학점, 공인영어성적, 영어회화, 제2외국어, 한자능력, 컴퓨터자격증, 직무관련 자격증, 해외경험, 인턴경험, 공모전 경력, 석박사 학위, 봉사경험, 동아리경험, 기타
인적자본 변수: 직업훈련 및 자격(4개)	직업훈련 참여개수, 자격증 누적 취득 개수, 취업사교육경험여부, 해외연수 및 유학 경험 여부
구직활동 경험 (6개)	취업 정보취득경로 1순위, 구직 시 어려운 점, 구직횟수, 최종면접횟수, 구직실패 경험, 입직경로
자이존중감 (13개)	(1) 내 자신이 다른 사람으로 바뀌었으면 하고 생각한다. (2) 가끔 인생의 실패자라는 생각이 든다. (3) 나는 직장을 구하는 데 도움이 되는 정보를 충분히 활용할 수 있다. (4) 내가 하고 있는 일이 잘못되고 있다고 생각되면 빨리 바로 잡을 수 있다. (5) 나는 나에게 어떠한 문제가 생겼을 때, 적절한 해결책을 찾을 수 있다. (6) 나는 남들에게 좋은 인상을 준다. (7) 나에게 업무가 주어지면 그 일을 해결할 수 있는 효과적인 방법을 찾을 수 있다. (8) 나는 아무리 어려운 상황도 잘 극복할 수 있다. (9) 나는 매사에 자신감 있게 임한다. (10) 나는 지금의 내 자신이 싫다. (11) 나는 내 자신이 쓸모없는 존재라고 느낀다. (12) 나는 내가 처리한 일 가운데 무엇이 잘 되었고 무엇이 잘못되었는지 정확히 분석할 수 있다. (13) 나는 여러 사람들 앞에서 이야기하는 것이 두렵지 않다.

IV. 실증분석 결과

1. 예측성과

본 연구는 랜덤포레스트 모델을 활용하여 하향취업 여부를 예측, 그 결과를 비교 분석하였다. 이를 위해 1개의 종속 변수와 84개의 독립 변수를 사용하였고, 학습 및 성과 평가를 위해 전체 표본의 70%는 훈련데이터로 30%는 검증데이터로 무작위 설정하였다. 또한 예측 모형의 성과 지표로는 정확도, 특이도, 민감도 세 가지 지표를 포함하였다. 정확도는 전체 하향취업 여부 자료 중 실제와 예측이 일치하는 비율이다. 특이도는 반응변수가 1(적정/상향취업자)인 자료 중 정분류된 자료의 비율이고 민감도는 반응변수가 0(하향취업자)인 자료 중 정분류된 자료의 비율이다.

<표 3> 정확도, 민감도, 특이도 정의

테스트 결과 (Prediction)	실제 하향취업의 유무 (Reference)		
	하향취업(True)	적정/상향취업(False)	합계
하향취업(True)	TP(True Positive)	FP(False Positive)	TP + FP
적정/상향취업(False)	FN(False Negative)	TN(True Negative)	FN + TN
합계	TP + FN	FP + TN	N= TP + TN + FP + FN

$$\text{정확도(Accuracy)} = (TN + TP) / (TN + TP + FN + FP)$$

$$\text{민감도(Sensitivity)} = TP / (TP + FN)$$

$$\text{특이도(Specificity)} = TN / (TN + FP)$$

2. 랜덤포레스트 분석 결과

랜덤포레스트 모델에서는 Breiman(2001)에 따라 $\text{floor}(\sqrt{n_x})$ 개의 설명변수를 채택하였기에 각 마디에 9개의 설명변수를 선택하였다. 이렇게 무작위

로 선택된 설명변수와 부트스트랩 표본을 이용하여 생성되는 트리의 수를 500개로 정하였다.

<표 4>의 모형1은 하향취업 여부의 기준을 교육 수준, 모형2는 기술 수준으로 설정한 모형에 대한 예측 결과를 보여준다. 랜덤포레스트로 하향취업 여부를 예측한 결과 모형1의 정확도는 약 93%이고 모형2의 정확도는 약 93%다. 모형1의 민감도는 약 91%로 랜덤포레스트 모형이 실제 하향취업의 경우를 하향취업으로 올바르게 예측한 정도이다. 모형2의 민감도는 약 92%이다. 특이도는 적정/상향 취업자라고 예측한 값 중 실제 적정/상향 취업자의 정도를 나타내며 모형1의 경우 약 96%, 모형2에서는 약 95% 값을 갖는다.

<표 4> 랜덤포레스트 예측성과

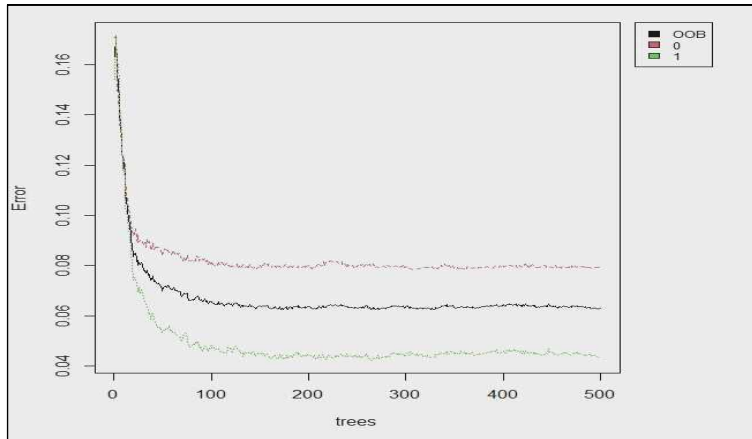
	모형1	모형2
정확도(Accuracy)	0.9353	0.9344
민감도(Sensitivity)	0.9149	0.9199
특이도(Specificity)	0.9589	0.9516

랜덤포레스트의 경우 무작위 결정트리를 어느 정도 생성할 것인지는 연구자의 재량에 달려 있다. 본 연구에서는 최대 500개의 트리를 생성하도록 설정하였고, 500개의 의사결정나무가 반복하여 생성될 때 오차율이 안정적으로 수렴하는지 살펴보고자 하였다.

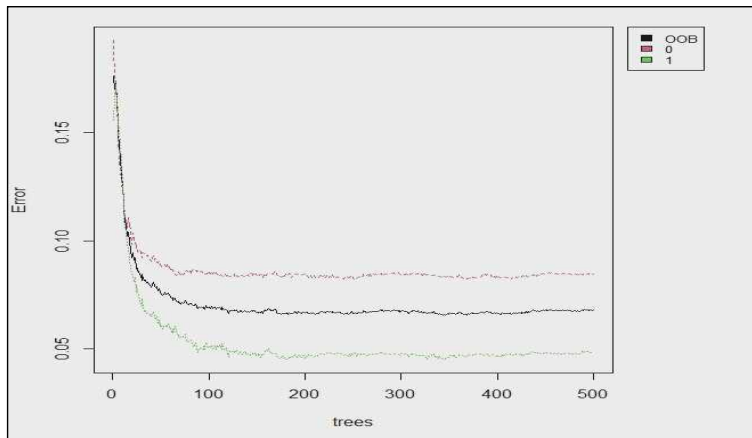
<그림 1>과 <그림 2>의 랜덤포레스트의 트리 개수에 따른 에러율 변화를 살펴보면 약 100개 이상의 의사결정나무가 생성되면 그 이후부터 일정한 값으로 수렴하는 것을 알 수 있다. 즉 본 연구에서 500번의 의사결정나무가 반복되는 랜덤포레스트 모형의 오차율은 안정적으로 수렴하였음을 보였다.

랜덤포레스트의 단점 중 하나는 랜덤포레스트 안에서 수많은 의사결정나무를 생성하기에 예측성과를 내는 반응 변수와 설명변수 간의 영향 관계를 파악하기 어렵다는 점이다. 이러한 문제를 해결하고자 예측성과에 대한 설명변수의 중요도를 나타내기 위해 두 가지 방법을 사용하고 있다.

<그림 1> 의사결정나무 생성에 따른 교육 수준 기준 하향취업의 에러율 변화



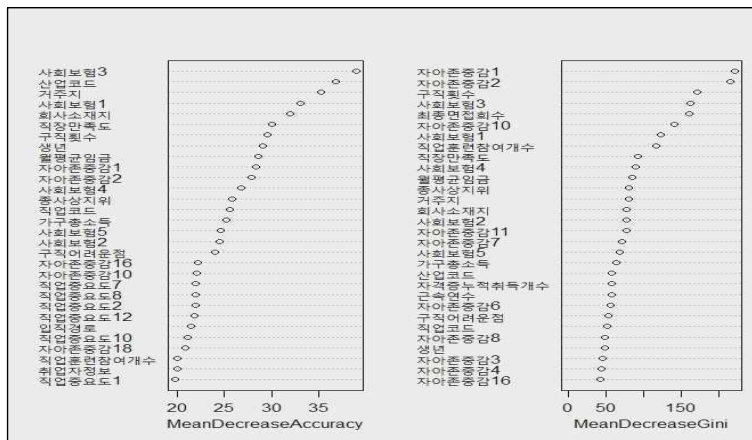
<그림 2> 의사결정나무 생성에 따른 직무수준 기준 하향취업의 에러율 변화



첫째, MDG(Mean Decrease Gini)는 랜덤포레스트를 이루는 각각의 의사 결정나무의 가지가 생성될 때 선택되는 변수들로 인한 GI 지수(Gini Impurity Index)의 감소량의 평균값을 나타낸다. MDG 값이 높다는 것은 선택된 예측 인자를 가지고 분류를 진행하면 불순도를 감소시키는 방향으로 분류가 진행됨을 의미한다. 또한, 불순도가 감소했다는 것은 분류 분석이 잘 진행되었다는 것과 같은 맥락이기 때문에 예측 인자의 MDG 값이 높은 변수가 중요 변

수인 것이다. 두 번째, MDA(Mean Decrease Accuracy)는 의사결정나무의 정확도를 구한 것과 시행된 의사결정나무에서 선택된 설명변수를 재조합(permutation)하여 의사결정나무를 재구축했을 때 구한 정확도의 차이를 변수별로 평균값을 구한 것이다. 그렇기에 분류의 정확도를 높이는 것에 큰 영향을 준 설명변수를 제거했을 때 의사결정나무의 정확도의 감소량은 커진다. 따라서 MDA의 값이 커질수록 변수의 중요도가 높아진다(김태진 외 2018; 최필선·민인식 2018).

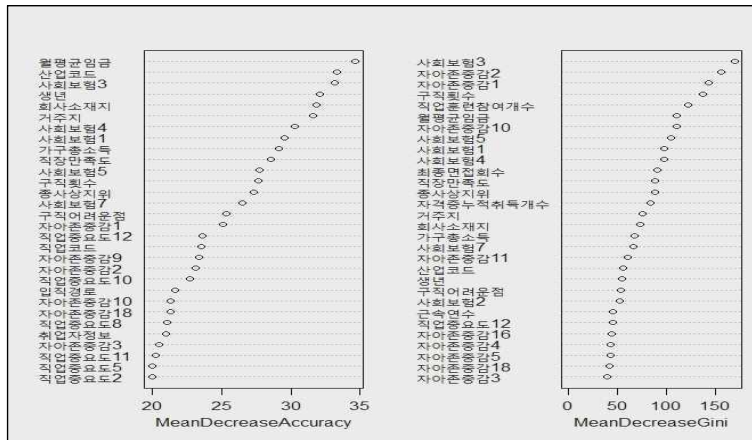
<그림 3> 교육 수준 기준 변수 중요도



<그림 3>은 교육수준을 기준으로 하향취업 여부를 예측한 모형에서 MDA와 MDG 기준 중요성 지수가 높은 변수 30개를 보여준다. MDA 기준에서 살펴보면 직장 속성인 국민연금(사회보험3), 건강보험(사회보험1), 산업코드, 회사소재지, 직장만족도 등의 변수가 중요한 것으로 나타났다. 그다음 주요 변인은 거주지 및 가구총소득과 같은 가족 및 가구 특성이었다. 구직활동 경험 변수에서는 구직 횟수가 주요 변인으로 뽑혔다. 자아존중감 변수들 중에서는 내 자신이 다른 사람으로 바뀌었으면 하고 생각한다(자아존중감1), 가끔 인생의 실패자라는 생각이 든다(자아존중감2), 나는 내가 처리한 일 가운데 무엇이 잘 되었고 무엇이 잘못 되었는지 정확히 분석할 수 있다(자아존중감16)과 같은 변수들이 주요 변수로 선택되었음을 알 수 있었다.

MDG 기준에서는 월평균임금, 종사상지위, 회사소재지와 같은 직장속성의 변수가 중요한 것으로 나타났으며, 내 자신이 다른 사람으로 바뀌었으면 하고 생각 한다(자아존중감1), 가끔 인생의 실패자라는 생각이 든다(자아존중감2),나는 지금의 내 자신이 싫다(자아존중감10)와 같은 자아존중감 변수들이 변수 중요도의 상위권에 있음을 확인하였다. 또한 구직횟수, 직업훈련참여개수와같은 구직활동 경험 변수와 국민연금(사회보험3), 건강보험(사회보험1), 직장만족도와 같은 혜택이 중요 변수임을 확인할 수 있었다.

<그림 4> 기술 수준 기준 변수 중요도



<그림 4>는 기술 수준을 기준으로 하향취업 여부를 예측한 모형에서 MDA와 MDG 기준 중요성 지수가 높은 30개의 변수를 보여준다. 모형2에서도 월평균임금, 사회보험3과 같은 직장 속성이 주요 변수로 채택 되었으며, 거주지, 가구총소득과 같은 인적 속성 변수들이 주요 변수로 나타났다. 또한 자아존중감1, 자아존중감2와 같이 자아존중감 변수들도 주요 변수로 채택되었다. MDA기준에서 살펴보면 월평균 임금, 산업속성, 국민연금(사회보험3), 회사소재지와 같은 직장속성 변수가 주요변수로 채택되었고 거주지, 가구총소득과 같은 개인 배경변수가 주요 변수로 채택되었다. 구직횟수와 같은 취업 노력변수도 주요 변수로 채택되었다. 자아존중감 변수의 경우 내 자신이 다른 사람으로 바뀌었으면 하고 생각 한다(자아존중감1), 나는 매사에 자신

감 있게 임한다(자아존중감9), 가끔 인생의 실패자라는 생각이 든다(자아존중감2) 등이 주요 변수로 선택되었다. MDG 기준에서도 직업 속성인 국민연금(사회보험3), 월평균 임금, 퇴직금(사회보험5)과 같은 변수가 주요 변수들 중 상위에 있음을 확인하였다. 자아존중감 변수도 상위에 올라와 있었는데 가끔 인생의 실패자라는 생각이 든다(자아존중감2), 내 자신이 다른 사람으로 바뀌었으면 하고 생각 한다(자아존중감1), 나는 지금의 내 자신이 싫다(자아존중감10) 등의 변수들이 중요도 순위가 높은 것으로 나타났다. 또한 구직 횟수, 직업훈련참여개수와 같은 취업 노력 변수가 주요 변수로 선택되었다.

<그림 3>과 <그림 4>를 통해 하향취업에 영향을 주는 주요 변수들을 살펴보았다. 이를 통해 기존 선행연구에서 하향취업에 결정요인에 주로 사용된 직업 속성 변수, 개인배경 변수뿐 아니라 자아존중감 변수들과 취업준비도 변수들이 중요 변수로 선택되었음을 확인하였다.

이상의 실증분석 결과를 요약하면, 우선 랜덤포레스트 분석에서는 인적 특성, 개인 배경, 직업 속성, 구직활동 경험, 자아 존중감 등 다양한 변수가 포함된 84개의 예측인자를 사용하여 분석을 진행하였다. 그 결과 교육수준 기준 하향취업 모형에서는 정확도가 93.53%, 민감도가 91.49%, 특이도가 95.89%로 나타났으며, 기술 수준 기준 하향취업 모형에서는 정확도가 93.44%, 민감도가 91.99%, 특이도가 95.16%로 나타났다. 다음으로 예측 요인들의 중요도 지수를 살펴본 결과 개인배경(거주지, 생년, 거주지, 가구 총 소득 등), 직장 속성(산업, 직장에서 제공되는 사회보험, 월 평균임금, 현 직장 소재지 등)과 취업을 위한 노력 변수(구직 횟수), 자아존중감 변수의 중요도 지수가 높게 나타났다.

V. 결론 및 논의

본 연구에서는 하향취업의 원인과 과정에 대해 직무 경쟁이론을 바탕으로 노동시장의 구조적인 측면에서 분석하였다. 일자리의 수와 질이 정해져 있는 상황에서 구직자들은 괜찮은 직장을 얻기 위해 경쟁시장에서 보다 우위를 선점할 수 있도록 상대적 역량을 축적하고자 노력한다. 그럼에도 불구하고

하고 상대적으로 역량이 부족한 구직자는 하향취업을 선택한다. 이러한 관점에서 기존의 하향취업 결정요인 연구에서는 연령, 성별과 같은 인구통계 변수와 거주지역, 가구 소득과 같은 개인 배경 변수, 교육 수준, 학점, 대학진공 등 인적자본변수를 주로 사용하였다. 하지만 장기간의 구직자 경쟁은 구직자 집단의 역량을 상향평준화하면서 이는 어학점수, 자격증, 직업훈련 여부 등으로 대표되는 구직자의 역량이 노동시장에서 현시되는 신호(signalling) 효과를 유발하지 못하였다. 이러한 조건에서 자아존중감은 하향취업 여부에 중요한 요인으로 작용할 수 있다.

머신러닝의 랜덤포레스트 모델을 사용하여 하향취업 여부를 예측하였다. 예측 분석과 함께 총 84개의 예측인자를 사용하여 변수의 중요도를 살펴보았다. 랜덤포레스트를 사용한 예측 분석 결과 직장 속성 변수뿐 아니라 자아존중감 또한 랜덤포레스트의 변수 중요도 지수가 높게 나타나 하향취업 여부 예측에 중요 요인임을 확인하였지만 구직자 역량은 예측에 중요 요인으로 작용하지 못하는 것을 확인하였다.

이러한 예측 결과는 자아존중감이 하향취업 여부를 결정짓는 하나의 가능성일 수 있기에 자아존중감을 향상시킬 수 있는 교육 또는 지원 방안을 마련할 필요성을 시사한다. 특히, 교육 및 지원 방안을 수립하는 과정에서 자아존중감이 가지는 영·유아기부터 청소년기 사이에 형성되고 쉽게 변하지 않는다는 특징과 자아존중감이 형성되는 유년기 시절부터 주변 환경에 큰 영향을 받는다는 특징을 고려해야 한다. 자아존중감이 형성되는 특정 시기에 집중적인 교육 및 지원이 필요하며, 직접적인 교육 대상뿐만 아니라 부모, 교사 등 주변 인물들을 대상으로 한 교육도 함께 진행되어야 한다.

마지막으로 본 연구에서 자아존중감이 노동시장 이행에 미치는 간접 경로에 관한 분석이 면밀하게 이루어지지 못하였고 자아존중감과 하향취업 사이를 잇는 다양한 조절 변수 및 매개변수 제시가 다소 미흡하였다. 따라서 향후 연구에서는 자아존중감이 노동시장 이행에 영향을 미치는 간접 경로와 자아존중감과 하향취업 관계에서 조절변수 및 매개변수의 과에 관한 연구가 이루어져야 할 것이다.

| 참고문헌 |

- 김준영·윤정혜 (2010). “하향 취업의 지속성: 청년층의 하향 취업 경험이 비하향 취업으로의 이행에 미치는 효과 분석.” 『산업노동연구』. 제16권, 1호, pp. 173-200.
- 김중성·이병훈·신재열 (2012). “청년층 구직활동과 하향취업.” 『노동정책연구』. 제12권, 2호, pp. 51-73.
- 김정우 (2015). “청년패널 자료를 활용한 청년층의 노조가입 결정요인 및 임금효과 패널분석.” 『산업노동연구』. 제21권, 2호, pp. 1-34.
- 김기홍 (2018). “고학력 청년 신규 취업자의 하향취업.” 『KLI 패널브리프』. 한국노동연구원, pp. 1-9.
- 김태진·홍정식·전윤수·박종률·안태욱 (2018). “랜덤포레스트를 이용한 모기업의 하향 거래처 기업의 분류 자동차 부품산업의 가치사슬을 중심으로.” 『한국전자거래학회지』. 제23권, 1호, pp. 1-22.
- 남성일·전재식 (2011). “하향취업이 직장이동 성향에 미치는 효과.” 『노동정책연구』. 제11권, 3호, pp. 25-51.
- 문선영 (2018). “불균형 자료에서 랜덤포레스트에 기반한 분류 방법의 성능비교.” 고려대학교 석사학위논문.
- 박천수 (2005). “우리나라 노동시장의 구조적 요인과 대응방안.” 『2005 한국의 고용구조』. 한국고용정보원, pp. 1-21.
- 송경애 (2006). “자아 존중감 및 자기결정 동기가 자기 주도적 학습에 미치는 영향.” 『한국유아교육·보육복지연구』. 제10권, 3호, pp. 177-202.
- 신선미·손유미 (2008). “대졸 청년층 하향취업의 결정요인.” 『직업능력개발연구』. 제11권, 1호, pp. 1-21.
- 신선미·민무숙 (2010). “대졸 청년층 하향취업자의 하향취업 상태 변화 요인.” 『직업능력개발연구』. 제13권, 2호, pp. 47-66.
- 양혜주 (2013). “전문대 간호학성의 자아존중감, 자기효능감 및 문제해결능력에 관한 연구.” 『사고개발』. 제9권, 3호, pp. 103-118.
- 오삼일·강달현 (2019). “하향취업의 현황과 특징.” 『BOK이슈노트』. 한국은행.
- 유진은 (2015). “랜덤포레스트: 의사결정나무의 대안으로서의 데이터 마이닝 기법.” 『교육평가연구』. 제28권, 2호, pp. 427-448.
- 이난 (2018). “대학생의 자아존중감, 사회성, 도덕성, 시민성 및 인성의 구조적 관계 분석.” 『교양교육연구』. 제12권, 6호, pp. 267-286.
- 이동훈 (2019). “기계학습 분류 기법을 활용한 대졸자 취업 예측 모델 연구.” 단국대

학교 석사학위논문.

- 이문수·박연복 (2019). “대졸 초기 취업자의 직장만족도, 자아존중감의 변화 양상과 영향 요인 탐색.” 『학습자중심교과교육연구』, 제19권, 19호, pp. 759-783.
- 이은정·조희숙·송영수 (2020). “랜덤 포레스트를 활용한 대졸 신입사원 조기이직 예측 결정요인 탐색.” 『기업교육과인재연구』, 제22권, 1호, pp. 163-193
- 이소영·이경순(2019). “대학생이 지각한 외모만족도와 대인관계 유능성의 관계에서 자아존중감의 매개효과.” 『교육문화연구』, 제25권, 5호, pp. 227-249.
- 이영식·송선혜 (2020). “고졸 취업자의 지역이동 결정요인.” 『노동정책연구』, 제20권, 4호, pp. 93-116.
- 이찬영 (2008). “청년층 임금근로자의 하향취업 행태.” 『직업능력개발연구』, 제11권, 3호, pp. 49-69.
- 장영은·성미영 (2017). “초등학교 아동이 지각한 6학년 학업성취에 대한 4학년 학업 성취의 예측: 5학년 자아존중감 및 학습행동조절의 순차적 매개효과.” 『한국보육지원학회지』, 제13권, 2호, pp. 21-37.
- 조민수·황광훈·오세미 (2014). “청년층의 하향 취업 현황과 이동 분석.” 『고용이슈』, 한국고용정보원, pp. 1-31.
- 조현재 (2017). “대학생의 학습된무기력, 진로적응성, 자아존중감의 구조관계 분석.” 『미래청소년학회지』, 제14권, 4호, pp. 93-109.
- 진선미·공정석 (2011). “대졸자의 하향취업이 노동시장이행에 미치는 영향에 관한 연구.” 『청소년복지연구』, 제13권, 1호, pp. 95-120.
- 채창균·김태기 (2009). “대졸 청년층의 취업 성과 결정 요인 분석.” 『직업교육연구』, 제28권, 2호, pp. 89-107.
- 채창균·양정승 (2019). “한국의 대기업은 어떤 청년을 선호하는가?” 『직업능력개발연구』, 제22권, 3호, pp. 99-127.
- 최필선·민인식 (2018). “머신러닝 기법을 이용한 대졸자 취업예측 모형.” 『직업능력개발연구』, 제21권, 1호, pp. 31-54.
- 황남희·정주연 (2011). “대졸청년층의 하향취업: 연령집단 및 하향취업 측정방법에 따른 비교.” 『정책분석평가학회보』, 제21권, 2호, pp. 271-292.
- Abernathy, T. J., L. Massad and L. Romano-Dwyer (1995). “The relationship between smoking and self-esteem.” *Adolescence*, Vol. 30, No. 120, pp. 899-907.
- Arndt, J. and J. L. Goldenberg (2002). “From threat to sweat: The role of physiological arousal in the motivation to maintain self-esteem.” In A. Tesser, D.A. Stapel, & J.V.Wood (Eds.), *Self and Motivation: Emerging Psychological*

- Perspectives*, Washington, DC: American Psychological Association, pp. 43-70.
- Alves-Martins, M., F. Peixoto, M. Gouveia-Pereira, V. Amaral and I. Pedro (2002). "Self-esteem and academic achievement among adolescents." *Educational psychology*. Vol. 22. No. 1, pp. 51-62.
- Bachman, J. G. and P. M. O'Malley (1977). "Self-esteem in young men: A longitudinal analysis of the impact of educational and occupational attainment." *Journal of personality and social psychology*. Vol. 35. No. 6, pp. 365-80.
- Baumeister, R. F., K. Dale and K. L. Sommer (1998). "Freudian defense mechanisms and empirical findings in modern social psychology: Reaction formation, projection, displacement, undoing, isolation, sublimation, and denial." *Journal of Personality*. Vol. 66, pp. 1081-1124.
- Baumeister, R. F., J. D. Campbell, J. I. Krueger and K. D. Vohs (2003). "Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles?" *Psychological science in the public interest*. Vol. 4. No. 1, pp. 1-44.
- Bender, K. A. and J. S. Heywood (2009). "Educational mismatch among Ph.Ds: determinants and consequences." *In Science and engineering careers in the United States: An analysis of markets and employment*. Chicago: University of Chicago Press, pp. 229-255.
- Bonanno, G. A., N. P. Field, A. Kovacevic and S. Kaltman (2002). "Self-enhancement as a buffer against extreme adversity: Civil war in Bosnia and traumatic loss in the United States." *Personality and Social Psychology Bulletin*. Vol. 28. No. 2, pp. 184-196.
- Breiman, L. (1996). "Bagging predictors." *Machine Learning*. Vol. 24, pp. 123-140.
- _____ (2001). "Random forests." *Machine Learning*. Vol. 45, pp. 5-32.
- Caspi, A., B. R. E. Wright, T. E. Moffitt and P. A. Silva (1998). "Early failure in the labor market: Childhood and adolescent predictors of unemployment in the transition to adulthood." *American sociological review*. Vol. 63. No. 3, pp. 424-451.
- Chawla, N. V., K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer (2002). "SMOTE: synthetic minority over-sampling technique." *Journal of artificial intelligence research*. Vol. 16, pp. 321-357.
- Covington, M. V. (1989). "Self-esteem and failure in school: Analysis and policy

- implications." *The social importance of self-esteem*. Berkeley: University of California Press, pp. 72-124.
- Crocker, J. and L. E. Park (2003). "Seeking self-esteem: Construction, maintenance, and protection of self-worth." Tangney (Eds.). *Handbook of self and identity*. New York: Guilford Press, pp. 291-313.
- _____ (2004). "The costly pursuit of self-esteem." *Psychological bulletin*. Vol. 130. No. 3, pp. 392-414.
- Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson and J. J. Lawler (2007). "Random forests for classification in ecology." *Ecology*. Vol. 88. No. 11, pp. 2783-2792.
- Demo, D. H. and R. C. Savin-Williams (1983). "Early adolescent self-esteem as a function of social class: Rosenberg and Pearlin revisited." *American journal of sociology*. Vol. 88. No. 4, pp. 763-774.
- Di Paula, A. and J. D. Campbell (2002). "Self-esteem and persistence in the face of failure." *Journal of personality and social psychology*. Vol. 83. No. 3, pp. 711-724.
- Emler, N. (2001). *Self esteem: The costs and causes of low self worth*. York: York Publishing Services.
- Freeman, R. (1976). *The overeducated American*. New York: Academic Press.
- Friedman, H. L. (1989). "The health of adolescents: beliefs and behaviour." *Social Science and Medicine*. Vol. 29. No. 3, pp. 309-315.
- Harter, S. (1993). "Causes and consequences of low self-esteem in children and adolescents. In Self-esteem." Baumeister, R. F. (eds). *Self-Esteem. The Plenum Series in Social / Clinical Psychology*. Boston: Springer, pp. 87-116.
- Howe, G. W., M. L. Levy and R. D. Caplan (2004). "Job loss and depressive symptoms in couples: common stressors, stress transmission, or relationship disruption?" *Journal of Family Psychology*. Vol. 18. No. 4, pp. 639-50.
- Huyse-Gaytandjieva, A., W. Groot, M. Pavlova and C. Joling (2015). "Low self-esteem predicts future unemployment." *Journal of applied economics*. Vol. 18. No. 2, pp. 325-346.
- Josephs, R. A., R. P. Larrick, C. M. Steele and R. E. Nisbett (1992). "Protecting the self from the negative consequences of risky decisions." *Journal of*

- personality and social psychology*. Vol. 62. No. 1, pp. 26-37.
- Lecky, P. (1969). *Self-Consistency: A Theory of Personality*. Garden City: Doubleday and Island Press.
- Maslow, A. H. (1943). "A theory of human motivation." *Psychological Review* Vol. 50. No. 4, pp. 370-396.
- Nauze-Fichet, E., M. Tomasini, S. Hanchane and E. Verdier (2002). "Diplôme et insertion sur le marché du travail: approches socioprofessionnelle et salariale du déclassement." *Économie et statistique*. Vol. 354. No. 1, pp. 21-48.
- Pal, M. (2005). "Random forest classifier for remote sensing classification." *International Journal of Remote Sensing*. Vol. 26. No. 1, pp. 217-222.
- Salmela-Aro, K and J. Nurmi (2007). "Self-esteem during university studies predicts career characteristics 10 years later." *Journal of Vocational Behavior*. Vol. 70. No. 3, pp. 463-477.
- Strobl, C., J. Malley and G. Tutz (2009). "An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests." *Psychological Methods*. Vol. 14, pp. 323-348.
- Taylor, S. E. and D. A. Armor (1996). "Positive illusions and coping with adversity." *Journal of personality*. Vol. 64. No. 4, pp. 873-898.
- Thunholm, P. (2004). "Decision-making style: habit, style or both?" *Personality and individual differences*. Vol. 36. No. 4, pp. 931-944.
- Thurow, L. C. (1972). "Education and economic equality." *The public interest*. Vol. 28, pp. 28-66.

| 논문투고일 : 2022년 11월 09일 |

| 논문심사일 : 2023년 03월 08일 |

| 게재확정일 : 2023년 03월 15일 |

| ABSTRACT |

**Unraveling the Relationship Between Self-Esteem
and Underemployment:
A Machine Learning Approach**

Yun-Young Kyung

(Dept. of Economics, Sungkyunkwan University)

Beom-Cheol Shin

(Dept. of Economics, Kyonggi University)

Young-Seok Lee

(Gyeonggi Research Institute)

Tae-Hyun Lee

(Dept. of Psychology, Korea University)

This paper uses the job competition theory to explore causes and procedures of underemployment by adapting structural aspects of the labor market. The paper empirically investigates determinants of underemployment by employing random forests. It pays a special attention to self-esteem as a key variable to determine whether or not the labor force status is that of underemployment. For the empirical analyses, a dataset of Youth Panel 2007 (YP2007) is utilized, ranging from the 5th to 12th dataset. Methods and results of the analyses are as follows.

Random forests take 84 independent variables into account including individual characteristics, personal background, types of jobs, job-search experience, and self-esteem. As a result, a model 1 for underemployment based on education is created, with 94% accuracy, 92% sensitivity, and 96% specificity. A model 2 for underemployment based on technology levels is also created with 93% accuracy, 92% sensitivity, and 95%

specificity. Both models find that there is high variable importance such as lists of industrial occupations, business locations, monthly average income, along with self-esteem variables. To conclude, the analysis of random forests finds that self-esteem is predictive in determining underemployment.

This paper considers the main determinant of underemployment with particular interest as self-esteem, which is an integral, albeit overlooked explanatory variable. As the result found that high self-esteem plays a key role to appropriate job transitions, methods for education in improving self-esteem are needed. Timely education is incredibly important. Self-esteem is shaped in the teenage years and from young and preschool periods. Once formed, self-esteem tends to be resistant to change. Also, the process of molding and forming self-esteem is influenced by environments around the person. Therefore, targets for education should be broadened into - not to mention young children - surrounding people including parents and teachers. Further research is needed to explore the specific path of how levels of self-esteem affects preparation procedures for jobs and closely analyze how levels of self-esteem influence career outcomes.

- Key words: Self-Esteem, Underemployment, Machine Learning, Random Forests, Predictive Model