

청년취업에 영향을 미치는 요인과 시대에 따른 변화과정 탐색: 1998년부터 2020년까지 랜덤포레스트 분석을 중심으로

송승익*, 김정환**, 박세연***

본 연구는 머신러닝의 랜덤포레스트 분석을 활용하여 1998년부터 2020년까지 각 연도별 청년 취업에 영향을 미치는 중요한 요인들을 밝히고, 그 영향요인들의 시대에 따른 변화과정을 파악하는 것을 목적으로 한다. 이를 위해 한국노동패널 1998년(1차년도)부터 2020년(23차년도)까지의 자료를 활용하였으며, 연구 모형에는 ‘개인특성’, ‘사회적배경’, ‘교육 관련’, ‘만족도’, ‘기타’ 요인으로 구성된 총 47개 변수가 투입되었다. 연구 결과를 정리하면 다음과 같다. 첫째, 연도별 변수중요도 분석, 연도별·변수별 중요 영향변수의 빈도 분석, 주요 변수별 변화과정 분석 등으로 시장선호와 선별이론(Market Signaling and Screening Theory), 통계적 차별이론(Statistical Discrimination Theory) 등 노동시장과 관련된 기존의 이론들을 검증하였고 ‘학력(학교)’, ‘직업훈련경험’, ‘혼인상태’, ‘사회경제적지위’ 등과 같이 꾸준히 중요도 상위 변수로 나타난 변수를 밝혔다. 둘째, 연도별 중요도 변수 및 패턴 변화를 통해 시대적 배경의 이해 수준을 높일 수 있었다. 특히, 2020년도 청년 고용률 저하는 코로나-19의 상황적·환경적 요인임을 확인하였다. 셋째, 선행연구와 변수의 중요도가 다르게 나타나거나 ‘사회경제적지위향상가능’, ‘생활만족도’ 등과 같은 새로운 변수가 영향요인으로 제시되어 연구의 지평을 넓힐 수 있었다.

주요용어 : 청년취업, 청년실업, 한국노동패널, 머신러닝, 랜덤포레스트

I. 서 론

최근 청년들이 심상치 않다. 청년들을 가리켜 ‘잃어버린 세대’라고 일컫기도 하고, 우리 사회를 떠받칠 기둥이어야 할 청년이 각종 정책에 있어 배려하고 지원해줘야 할 대상이 되어버렸다. 실제로 구글 트렌드 검색을 통해 ‘청년’이라는 키워드를 검색해보면, 최근 1년간 일주일에 최소 50회 이상 검색되는 등의 꾸준한 관심을 받고 있고, 연관 검색어로는 국민취업지원제도, 청년 면접 수당, 청년취업 장려금 등의 각종 청년취업 정책들이 쏟아져 나온다. 이는 청년들에 대한 관심이 꾸준하면서도 그 이면에는 취업에 대한 불안, 사회적 피해의식, 정책적 보상 등의 여러 의미가 내포되어 있다고 할 수 있다.

* 연세대학교 일반대학원 교육학과 박사과정,

** 고려대학교 일반대학원 통계학과 박사과정,

*** 서강대학교 일반대학원 신문방송학과 석사과정

통계청 2020년 연간 고용동향 자료에 의하면, 2020년 청년실업률은 9.0%로 나타나 지난해 대비 0.1%p가 상승하였으며, 청년층(15-29세)의 고용률 또한 42.2%로 지난해 대비 1.3%p 하락하였다. 이는 15세 이상 전체 고용률이 65.9%라는 점을 고려하면 우리 사회에서의 실업 문제가 청년층에 집중되어 있음을 보여준다(통계청, 2021). 이러한 청년기의 실업 문제를 더욱 심각하게 받아들여야 하는 이유 중에 하나는 청년기가 학교에서 노동시장으로 이행(transition)하는 매우 중요한 시기이기 때문인데, 이러한 청년기의 실업과 미취업 문제는 청년 계층에 불안과 함께 정체성 및 존재감의 상실을 유발하는 등의 심각한 문제로 나타나고 있다.

한편, 다시 생각해보면 이러한 청년취업 문제가 과연 오늘날의 문제인가 하는 의문이 든다. 박정희 정권 이후 고도의 경제성장을 하면서 청년들은 일자리 걱정없이 노동시장에서 열심히 그 역할을 다해주었다. 하지만 1990년대 말 IMF 경제위기로 인한 취업난을 겪으면서 청년취업의 충격이 시작되었다고 할 수 있다. 그런데 1990년대 이후 역대 정부의 꾸준한 관심과 노력에도 불구하고 오늘날까지 청년취업 문제가 왜 여전히 사회적 문제로 남아있는 것인가. 이렇듯 과거부터 지속되어 온 청년취업 문제에 대해 근본적인 해결책과 시사점을 찾기 위해서는 이 문제에 관한 전반적인 점검과 검토가 필요하다고 할 수 있다.

이러한 청년취업에 관한 선행연구들을 살펴보면, 청년취업을 취업성과의 일부로 연구가 이루어지고 있는데, 취업성과는 취업여부, 전공일치취업, 대기업취업, 임금 등으로 다양하게 정의되고 있다(이병식·최정운, 2009). 청년취업에 영향을 미치는 요인으로서는 성별(금재호, 2010; 주휘정·김민석, 2018), 사회경제적 배경(박환보·김성식, 2011; 채창균·김태기, 2009), 학점 등 대학 관련 요인(정태영·이기엽, 2005; 이영민·이수영·임정연, 2014), 외국어능력(이기종·김은주, 2006), 진로 및 직업의식(노경란·박용호·허선주, 2011; 주휘정·김민석, 2018), 수도권 대학 여부(이병식·최정운, 2009) 등에 대해 주로 연구가 이루어졌다. 국외 연구로는 Pascarella & Terenzini(2005)의 연구에서 최근 30년간의 연구를 메타분석한 결과, 전공, 학업성취도, 일 경험, 교과외 활동, 교수와의 상호작용, 우수대학 졸업 여부를 노동시장에서의 성과에 영향을 미친 요인으로 분석하였다.

이렇듯 대다수의 연구들이 취업성가를 주요 변수로 연구를 수행하고 있으나, 위에서 언급한 바와 같이 다양하게 정의되거나 측정되어 일관성이 부족한 한계가 있다. 또한, 청년정책은 보다 넓은 범위에서 시행되고 있는 데에 반해, 선행연구는 대부분 대학졸업자들 위주로 이루어지고 있었으며, 청년취업에 영향을 미칠 것으로 예상되는 다양한 변수들을 매우 제한적으로만 활용하고 있었다. 결국 대부분의 연구에서 연구하고자 하는 목적에 따라 관심있는 변수만을 제한적으로 포함하여 모형을 구성하였기 때문에 결과를 일반화하기 어렵고, 영향을 미칠 수도 있는 요인들을 의도적으로 배제하였다는 문제가 있었다. 또한 연구방법에 있어서도 활용할 수 있는 변수가 제한적인 회귀분석, 구조방정식 분석, 다층분석 등을 적용했다는 특징이 있었는데, 이러한 모형들은 통제변수를 어떻게 포함하느냐에 따라 연구 결과가 완전히 달라질 수 있고, 많은 변수를 투입하게 되면 모형이 불안정하다는 한계가 있다(Yoo & Rho, 2017). 그리고 무엇보다 청년취업에 영향을 미치는 요인들을 분석함에 있어 IMF 직후부터 현재까지 방대한 데이터를 기반으로 거시적인 측면에서 분석한 연구가 없었다는 점과 청년취업에 영향을 미치는 요인이 무엇인지만 밝혔을 뿐, 요인들 간의 상대

적 중요도를 밝히지 못했다는 점이 가장 큰 한계라고 할 수 있다.

따라서 본 연구에서는 이러한 한계들을 고려하여 청년취업에 영향을 미치는 요인들을 IMF 경제위기 직후인 1998년부터 2020년까지 거시적으로 탐색함과 동시에 머신러닝 기법을 통해 기존 선행연구에서 강조했던 변수들뿐만 아니라 기존에 고려되지 않았던 변수들까지 가급적 모든 변수를 포함하여 분석을 시도하고, 각 연도별 영향요인의 상대적 중요도를 확인하고자 한다. IMF 경제위기 직후인 1998년부터 2020년 현재까지 각 연도별 청년취업에 영향을 미치는 요인은 무엇이고, 시대에 따른 영향요인의 변화에 대해 분석할 것이다. 이를 위해 1998년부터 청년취업에 관한 패널데이터가 갖춰져 있는 한국노동패널 자료를 활용할 것이며, 해당 자료 내에서 취업된 상태에서 응답을 요구하는 문항을 제외하고는 거의 모든 문항에 대해 변수화하여 분석을 시도하고자 한다. 이를 통해 그동안 청년취업 문제에 영향을 미쳤던 요인들에 대해 거시적인 측면에서 살펴봄으로써, 시대에 따라 상대적으로 중요한 설명변수가 어떻게 달라졌는지, 또는 시대의 변화에도 불구하고 꾸준히 중요하게 나타난 공통적인 설명변수는 무엇인지를 밝힐 수 있을 것이며, 포스트 코로나-19 시대에 청년취업 문제 해결에 도움이 될만한 정책적 시사점을 도출할 수 있을 것으로 기대한다. 본 연구의 구체적인 연구문제는 다음과 같다.

첫째, 1998년부터 2020년까지 각 연도별 청년취업에 영향을 미치는 요인은 무엇인가?

둘째, 1998년부터 2020년까지 청년취업에 영향을 미치는 요인은 어떻게 변화하였는가?

셋째, 최근(2020년) 청년취업에 영향을 미치는 중요한 요인들은 무엇이며, 코로나-19 이전과 차이가 있는가?

II. 이론적 배경

1. 청년취업

가. 청년의 정의

청년은 사전적으로 신체적으로나 정신적으로 가장 왕성하게 발달하는 힘이 넘치는 시기에 있는 나잇대를 의미한다. 청년에 속하는 나잇대에 대한 기준은 사회 변화에 발맞춰 그 범위가 확장되고 있으며 청년의 범위에 대해 상이한 기준이 제시되고 있다. 유엔(UN)은 지역 간 통계적 일관성을 위해 청년을 만 15세에서 24세 사이의 사람들로 정의한다. 이에 OECD의 청년실업률 조사대상 연령층은 만 15세에서 24세이다. 하지만 우리나라는 일반적으로 15세에서 29세를 청년층으로 말하며 청년실업률 조사대상 역시 만 15세에서 29세로 한다(통계청, 2021). 한편, 2020년 8월 5일부터 시행된 국내 청년기본법에서 말하는 청년이란 만 19세 이상 34세 이하인 사람이다. 다만, 법령과 조례에서 청년에 대한 연령을 다르게 적용하는 경우에는 그에 따를 수 있도록 자율성을 보장하고 있어 청년의 기준이 다양하게 제시되고 있다. 이에 따라 지자체마다 지역의 특성에 맞게 청년의 나이를

다르게 규정하고 있다. 전국 17개 시·도의 청년 나이 기준을 살펴보면, 만 19세에서 39세가 7곳으로 가장 많다. 그 다음으로 만 18세에서 39세 4곳, 만 18세에서 34세 3곳, 만 19세에서 34세 2곳, 만 15세에서 39세 1곳 순이다. 이러한 지자체에서는 인구고령화에 따른 젊은 층의 인구유출을 막고 관련 지원 사업을 펼치기 위해 청년의 연령 기준을 매우 유동적으로 적용하고 있다. 따라서 지자체의 청년 연령 기준은 본 연구의 목적에 부합하지 않으므로 지자체의 청년 연령 기준에 대해서는 본 연구의 고려사항에 포함하지 않았으며, 청년기본법 또한 다른 법령과 조례에서 정하는 연령을 그대로 따르도록 했기 때문에 청년기본법 역시 그 기준이 되기는 어렵다고 보았다.

한편, 청년고용촉진특별법(청년고용법)은 청년 미취업자에게 직업능력개발훈련을 지원하기 위해 만들어진 법률로 청년고용을 촉진하고 지속적인 경제발전과 사회 안정에 이바지함을 목적으로 한다. 여기서 말하는 ‘청년’이란 취업을 원하는 사람으로서 대통령령으로 정하는 나이에 해당하는 사람을 지칭하며 이는 15세 이상 29세 이하인 사람을 말한다. 미취업 청년이 공공과 중소기업체에 고용되어 일할 수 있도록 정부가 금전적인 지원을 할 수 있는 근거가 되는 법률로 만 15세 이상 29세 이하의 청년을 대상으로 한다. 여기서 청년의 기준이 만 15세부터인 이유는 청년을 ‘신체적·정신적으로 한창 성장하거나 무르익은 시기에 있는 사람’으로서 ‘사회경제적활동이 가능한 연령층’을 청년의 시작으로 보고 있기 때문이다. 따라서 본 연구에서는 통계청의 경제활동인구조사 청년층 부가조사와 대졸자직업이동경로조사를 이용한 통계자료와 연계가 가능하고, 대통령령, 조세제한특별법, 청년고용촉진특별법의 청년 연령 기준을 따라 만 15세에서 29세 이하를 청년층으로 정의함으로써 일반적인 청년층 정의와 비교가 가능하도록 함과 동시에 청년고용촉진특별법에 따른 다양한 정부의 취업정책에 중요한 시사점을 확인하고자 한다.

나. 청년취업

청년취업 문제가 지속되면서 청년에 대한 연구가 다양한 분야에서 꾸준히 이루어지고 있다. 청년취업은 취업성과, 노동시장성과, 선호직장 취업 등 다양하게 정의되어 연구되었다. 정태영과 이기엽의 연구(2005)에서 청년취업은 취업여부와 대기업 취업 여부로 취업 상태를 분류하였고, 박환보의 연구(2011)에서는 청년취업을 노동시장성과의 일부로 보며 질적인 측면을 고려하여 취업여부, 정규직여부, 대기업 정규직여부의 세 가지 변인으로 구성해 분석하였다. 한편, 한성민 외의 연구(2016)에서는 취업형태의 구분 없이 취업여부에 따라 청년취업을 정의하였으며, 권혜진 외의 연구(2010) 역시 근무형태, 일자리 형태, 비임금노동, 질병 휴직 등을 모두 취업으로 간주하기도 하였다.

취업의 사전적 의미는 직업을 가진다는 뜻이나, 통계를 작성할 때의 취업자는 특정한 의미를 갖는다. 통계청은 경제활동인구조사 시 취업자 개념을 국제노동기구(ILO) 기준에 따르고 있다. ILO은 수입을 목적으로 조사대상 주간동안 1시간 이상 일한 사람을 취업자로 보아 근로형태를 가리지 않고 수입을 목적으로 1주 동안 1시간 이상 일했다면 모두 취업자로 정의한다. 더불어, 복수의 활동 상태를 갖는 사람은 취업자, 실업자, 비경제활동 인구 중 반드시 하나의 활동 상태에만 배타적으로 귀속되도록 하는 우선성 규칙(Priority rule)을 적용하도록 한다. 이 규칙으로 국내 거주하는 15세

이상 모든 인구는 빠짐없이 취업자, 실업자, 비경제활동인구 중 하나의 활동 상태를 갖게 된다(통계청, 2021).

따라서 본 연구는 1998년부터 2020년까지 23년에 걸쳐 청년취업에 영향을 미치는 중요한 요인들을 밝히는 탐색적 연구로서, 거시적이고 탐색적 수준에서의 영향변수들을 분석하기 위해 취업에 대하여 보다 넓은 범위에서 분석하고자 한다. 따라서 청년취업을 정의하는데 있어 전공일치취업, 대기업취업, 임금, 만족도, 정규직 여부 등과 취업의 범위를 한정하는 지표가 아닌, 통계자료에서의 ‘취업’의 정의에 따라 청년취업을 넓은 범위에서 정의하고자 한다. 이에 본 연구에서의 취업은 근무형태(정규직/비정규직), 일자리형태(상용직, 일용직, 임시직 등)에 상관없이 근무하는 일자리가 있는 경우 취업으로 간주하며, 청년층에 해당하는 만 15세에서 29세 이하인 사람의 취업 상태를 청년취업이라고 본다.

다. 청년취업 관련 이론

청년취업 관련 연구 중 청년취업의 영향요인에 초점을 맞춘 연구는 IMF 이후 청년 실업 문제가 사회적 이슈로 대두되면서 본격적으로 이뤄졌다. 취업 영향요인은 사회적 배경, 취업준비행동 등과 같이 취업에 영향을 미치는 요인들을 말한다(한성민 외, 2016). 청년취업 영향요인에 대해 이해하기 위해 노동시장에서의 성과와 관련된 이론을 살펴보는 것이 필요하다. 노동시장에서의 청년취업과 관련된 이론으로는 인적자본론, 시장선호와 선별이론, 노동시장분절론, 통계적 차별이론, 지위획득이론, 이행노동시장이론 등을 꼽을 수 있다. 관련 이론을 본 연구의 범주에 따라 구분하여 설명하면 다음과 같다.

먼저, 개인특성에 관한 이론이다. 시장선호와 선별이론(Market Signaling and Screening Theory)은 교육이나 직업훈련으로 개인의 지식과 기술이 길러져 취업 영향요인으로 작동하는 것이 아니라 출신 학교가 어디인지 등 교육수준이 개인의 선천적 능력을 평가하는 기준으로 작동한다고 말한다. 인적자본론은 교육수준에 따라 개인의 능력이 길러져 취업성과의 질이 향상될 것이라고 주장하지만, 선별이론에서의 교육 및 직업훈련은 고용주가 단순히 개인의 능력을 판별하는 수단이라 할 수 있다(Spencer, 1973). 또한, 개인 속성에 따라 노동시장에서의 성과 차이가 발생한다는 통계적 차별이론(Statistical Discrimination Theory)이 있다. 이는 정보의 비대칭성에 기인한 것으로 기업이 과거의 통계적 관찰에 의존해 성별, 연령, 인종 등과 같이 관찰할 수 있는 근로자의 특성에 의거하여 채용, 승진 등을 결정한다는 것이다(Arrow, 1972; Phelps, 1972).

다음으로, 사회경제적 배경에 관한 이론이다. 지위획득이론(Status Attainment Theory)은 부모의 학력, 소득 수준 등 개인의 사회경제적 배경을 주요 취업 영향요인으로 본다. Duncan(1961)은 아버지의 직업 지위가 자녀의 학력에 영향을 미치며 직업 지위도 그대로 세습하게 된다고 주장하였다(김종성, 2013). 지위획득 이론에 기초한 선행연구에서는 부모 학력 및 가구소득을 취업 영향요인 변수로 포함해 취업에 긍정적인 영향을 미치는 것으로 나타난 연구결과를 일관되게 보고하였다(박성재, 2005; 황여정·백병부, 2008; 박환보, 2011). 더불어, 김종성과 이병훈(2014)은 부모의 소득 수

준이 자녀의 취업준비과정에 경제적 도움과 교육에 대한 지원 정도로 이어지는 구조를 갖고 있어 부모의 소득이 자녀의 취업에 중요하게 고려된다고 설명하였다.

마지막으로, 교육훈련에 관한 이론이다. 인적자본론(Human Capital Theory)은 완전경쟁 노동시장을 전제로 개인의 취업 준비에 따라 취업성도가 다르게 나타날 수 있다고 본다. 개인의 취업 준비를 인적자본으로 보며 개인의 경험, 교육 등 개인이 지닌 능력이 축적되면 될수록, 취업시장에서 더 나은 결과를 볼 수 있다는 입장이다(Becker, 1964). 이에 따라 인적자본론의 입장에서 청년취업 영향요인을 본다면, 교육과 훈련에 주요 초점을 맞춘다고 할 수 있다.

그밖에, 노동시장에 존재하는 격차에 따라 취업 영향요인에서 사회구조적인 차별이 발생할 수 있음을 시사하는 노동시장분절론(Segmented Labor Market Theory)이 있으며(김덕호, 2020), 취업 영향요인으로 정부의 정책을 강조하는 이행노동시장이론(Transitional Labor Market Theory)이 있다. 이 이론에서는 정부의 적극적인 취업시장 정책으로 개인의 취업을 방해하는 취업 관련 정보 부족, 훈련 부족 등 위험요인에 관한 관리가 가능하며 더 나은 상태로 나아갈 수 있다고 주장하기도 한다(Schmid, 1998).

2. 청년취업 영향요인

위의 이론을 바탕으로 청년취업에 영향을 미치는 요인들과 관련된 선행연구를 종합해보면, 크게 (1) 개인특성, (2) 사회경제적 배경, (3) 학업 관련 요인, (4) 직업훈련 관련 요인 등으로 분류할 수 있다.

우선, 개인특성의 경우 성별, 연령, 혼인 여부 등이 취업 상태에 영향을 미치는 것으로 나타났다. 성별은 여성보다 남성이 취업에 유리한 것으로 꾸준히 보고되고 있는데(금재호, 2000; 황여정·백병부, 2008; 금재호, 2010; 박환보, 2011; 문영만·홍장표, 2017; 주희정·김민석, 2018; 김덕호, 2020), 이는 통계적 차별이론으로도 일정 부분 이해할 수 있다. 채용에 있어 남성을 더 선호하는 기호가 존재하고 있는 것이다. 연령은 취업 영향요인으로 엇갈리는 결과가 나타나고 있는데, 대체로 연령이 높을수록 임금 수준이나 고용상태가 보다 나아지는 것으로 나타난다(박성재, 2005; 박환보, 2012; 김덕호, 2020).

사회경제적 배경으로는 부모학력, 가구소득 등 가족배경이 취업에 영향을 미치는 요인으로 나타났다. 특히, 가구소득 및 부모의 소득과 자녀의 취업성도에 관한 연구결과는 일관되게 유의미한 것으로 나타나고 있다(박성재, 2005; 황여정·백병부, 2008; 박환보, 2011). 박성재(2005)는 ‘한국직업 능력개발원 졸업자조사 자료’를 기반으로 지방대 졸업자의 취업성도를 분석한 결과, 부모직업, 가구소득, 출신지역 등 개인의 사회경제적 배경이 취업에 유의한 영향을 미치는 것으로 분석했다. 부모의 소득수준은 자녀의 취업에 관한 경제적 지원이 얼마만큼 가능한가와 연결되기 때문에 더욱 취업성도에 정적인 영향을 미치게 된다(김덕호, 2020).

학업 관련 요인으로 학력, 대학유형, 대학소재지, 전공, 학점 등을 꼽을 수 있다. 특히 학력은 일관되게 취업에 영향을 미치는 것으로 보고되어왔다. 학력이 높을수록 취업할 확률이 크며, 동일 학

력 내에서도 대학의 유형, 서열, 소재지, 전공 등 대학 수준이 세분화되어 취업에 영향을 미치는 요인으로 보고되었다. 높은 서열로 평가받는 대학 졸업자는 더 높은 임금수준을 받는 것으로 나타나고(류장수, 2003; 오호영, 2007), 전공에 따라서도 임금수준에 차이가 발생하는 것으로 보고된다(박성재, 2005; 김창환·김형석, 2006; 김덕호, 2020). 황여정·백병부(2008)는 ‘대졸자직업이동경로조사(GOMS)’ 1차년도 자료를 사용하여 대졸 청년층의 노동시장 성과 결정요인을 추정한 결과, 취업 여부에 영향을 미친 변수는 대학소재지, 전공, 학점 등으로 나타났다. 연보라 외(2015) 연구 결과에서도 대학 소재지가 취업성과에 유의미한 영향을 미치는 결과가 보고되었다. 또한, 학점도 취업에 주요한 영향을 미치는 요인으로 나타났다. 정태영·이기업(2005)은 ‘대학 졸업생의 학사 데이터’를 이용하여 학점이 졸업 후 취업에 미치는 영향을 분석한 결과, 학점이 대기업 취업에 매우 중요한 영향을 미쳤다고 밝혔다.

마지막으로 직업훈련 관련 요인은 자격증, 어학능력, 어학연수, 인턴 및 직무 경험, 노동 경험 등을 취업 영향요인으로 말한다. 이영민 외(2014)는 ‘대졸자직업이동경로조사(2010GOMS)’ 자료 기반으로 선호직장 취업요인을 분석한 결과 복수전공, 해외연수, 영어점수, 대학 재학 중 취업 관련 교과목 수강 정도 등 취업 사전준비가 많았던 학생들이 선호직장 취업률이 높은 것으로 나타났다. 문영만·홍장표(2017)도 개인적 속성인 성별을 비롯하여 영어성적, 인턴 및 직무 경험이 취업에 유의미한 영향을 미친다고 밝혔다. 이처럼, 취업을 위한 준비의 일환으로 어학 능력이나 어학연수 경험이나 대학 재학 중의 일자리 경험 및 직업훈련 경험도 취업 가능성을 높이는 것으로 나타났다. 김덕호(2020)는 정부에서 일 경험, 해외취업 등 직무능력향상을 위한 프로그램을 제공하며 구직자의 취업을 지원하는 ‘취업성공패키지 프로그램’ 참여가 취업에 미치는 영향을 알아본 결과, 취업에 긍정적인 영향을 미치는 것으로 나타났다. 이는 이행노동시장이론이 주장한 바와 같이, 시기적절한 취업 관련 정부의 정책 및 교육과정을 강조한다. 취업을 위한 준비 활동으로서 직업훈련 관련 요인의 중요성이 선행연구를 통해 강조되고 있는 가운데, 김성훈(2015)은 대학생의 취업 준비 활동에 영향을 미치는 요인으로 관련 선행연구 결과를 토대로 정리하였다. 그 결과, 대체로 사회경제적 배경 요인으로 꼽히는 가구소득, 부모계급, 부모학력 등이 주요 영향 요인으로 드러났다. 여러 취업 영향요인 내 사회경제적 배경이 직업훈련 관련 요인에 영향을 미치고 있다는 점을 확인할 수 있다. 이러한 선행연구에서 밝힌 청년취업과 관련된 분석 대상과 영향요인을 정리하면 <표 1>과 같다.

선행연구를 검토한 결과, 청년취업에 관한 많은 연구가 이뤄지고 있으나 지금까지의 청년취업 연구에 분명한 한계점이 드러났다. 우선, 대다수의 청년취업 관련 연구들이 취업성과를 주요 변수로 연구를 수행하고 있으나, 위에서 살펴본 바에 같이 해당 변수에 대해 임금 수준, 기업 규모, 직장 만족도 등 상이한 정의로 측정되어 선행연구 간 취업성과 개념에 대한 일관성이 부족하다는 점이 나타났다. 또한, 청년취업에 영향을 미칠 것으로 예상되는 변수들이 매우 제한적이고 한정적으로 활용되고 있음을 확인할 수 있었다. 무엇보다 대다수의 청년 연구가 대졸 청년에 한해 이뤄진 한계점이 있다. 이에 본 연구는 만 15세 이상 29세 이하의 청년층 대상으로, 그들의 취업여부를 기준으로 1998년부터 2020년까지 청년취업에 영향을 미치는 영향요인이 무엇인지 밝히고, 시대별 영향요인 변화를 분석하고자 한다.

<표 1> 청년취업 영향요인

연구자	분석 대상	영향요인
박성재 (2005)	한국직업능력개발원 졸업자 조사데이터 지방대 졸업생	성별, 연령, 부모직업, 가구소득, 전공, 최종학교, 출신지역 등
정태영·이기엽 (2005)	대학 졸업생의 학사 데이터	학점
황여정·백병부 (2008)	대졸자직업이동경로조사 (GOMS) 1차연도 자료	대학소재지, 전공, 학점, 취업관련 대학의 지원활동에 대한 만족도, 졸업 전 구체적 취업목표 설정 여부, 어학연수 비용, 연령, 가구소득, 재학 중 근무경험 등
박환보(2011)	대졸자직업이동경로조사 (2008GOMS) 자료	성별, 가계소득, 출신고 등
이영민 외 (2014)	대졸자직업이동경로조사 (2010GOMS) 자료	복수전공, 해외연수, 영어점수, 취업 사전준비 등
연보라 외 (2015)	대졸자 직업이동 경로조사(2011)	출신대학 소재지
문영만·홍장표 (2017)	한국고용정보원 '청년패널 데이터'	성별, 영어성적, 인턴 및 직무 경험, 최종학력,
박주상·문영만 (2018)	대졸자직업이동경로조사 (2015GOMS) 자료	대학재학 중 노동 경험
김덕호 (2020)	대졸자직업이동경로조사 (2016GOMS) 자료	연령, 성별, 혼인, 전공, 부모소득, 경제적지원, 청년취업성공패키지 참여경험

자료: 선행연구를 바탕으로 연구자가 정리

III. 연구 방법

1. 분석 자료

본 연구는 한국노동패널(KLIPS: Korean Labor and Income Panel Study) 자료를 활용하였다. 한국노동패널조사는 국내 유일의 노동 관련 패널조사로, 우리나라 노동시장의 현황을 파악하고, 올바른 노동정책의 수립과 실행을 위하여 1998년에 시작되어 1년에 1회씩 전국의 12,000여 가구와 해당 가구에 속한 만15세 이상 가구원을 대상으로 조사를 실시하고 있다. 특히, 조사 대상을 만 15세 이상 가구원을 대상으로 하고 있고, 국내 노동 관련 패널 데이터로는 유일하게 1998년부터 조사를 수행 중이라는 점이 가장 큰 특징인데, 만 15세 이상 29세 이하 청년들을 대상으로 하고 IMF 경제위기 직후인 1998년부터 2020년 현재까지 분석하는 본 연구에 적절하다고 판단하였다(한국노동연구원, 2021).

한국노동패널조사는 개인특성, 사회경제적 배경, 경제활동, 교육 및 직업훈련 등 다양한 범위의 조사를 통해 학문적 연구와 정책적 발전에 기여하는 것을 목적으로 하며, 전국의 12,000여 가구와 그 가구에 속한 23,000여명의 개인을 매년 추적 조사하는 대규모 패널조사이다(한국노동연구원, 2021). 본 연구는 만 15세 이상 29세 이하 청년들을 대상으로 그들의 취업여부에 영향을 미치는 요

인을 밝히기 위해 머신러닝 기법을 적용하여 최대한 많은 설명변수를 투입하여 분석하고자 하기 때문에 1998년(1차년도)부터 2020년(23차년도)까지 전체 패널 데이터 중 가구원을 조사 대상으로 한 개인용 자료를 활용하였다.

또한, 1차년도부터 23차년도까지 각각의 연도별 횡단면 자료를 활용하였으므로, 각 차수별 분석 대상이 모두 상이하였다. 각 횡단면 자료에서 각 변수에 해당하는 문항에 대해 모두 빠짐없이 응답한 가구원만 포함하기 위해 결측이 발생한 행은 완전히 삭제하는(Listwise) 방식으로 처리하였다. 각 연도별 분석 대상은 아래 <표 2>과 같다.

<표 2> 각 연도별 분석 대상

구분	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차
대상	4,129	3,438	3,248	3,045	2,986	3,133	2,774	2,631	2,585	2,540	2,372
12차	13차	14차	15차	16차	17차	18차	19차	20차	21차	22차	23차
2,356	2,452	2,279	2,188	2,120	1,809	1,940	2,070	2,222	3,685	3,099	3,062

자료: 한국노동패널 1차-23차 조사

2. 분석 변수

가. 반응변수(종속변수)

본 연구는 1998년부터 2020년까지 각 연도별 청년취업에 영향을 미치는 중요한 요인들을 밝히고, 그 영향요인들의 시대에 따른 변화과정을 분석하는 데 목적이 있다. 이를 위해 종속변수로 취업여부를 선정하였는데, 취업여부는 1차년도부터 23차년도까지의 KLIPS 조사에서 ‘취업자’ 또는 ‘미취업자’로 응답한 내용을 코딩하여 활용하였다. 이는 정규직 또는 비정규직과 같은 근무형태나, 상용직 또는 일용직 등과 같은 일자리 형태, 사업여부 등에 관계없이 정기적으로 근무하는 일자리가 있을 경우 모두 ‘취업자’로 파악한 변수이다. 본 연구는 1998년부터 2020년까지 23년에 걸쳐 청년취업에 영향을 미치는 중요한 요인들을 밝히는 탐색적 연구로서, 거시적이고 탐색적 수준에서의 영향변수들을 분석하기 위해 취업에 대하여 보다 넓은 잣대를 적용하였다. 이러한 취업에 대한 넓은 범위의 적용은 취업형태의 구분 없이 취업여부를 보는 한성민 외(2016)의 연구나 근무형태, 일자리 형태, 비임금노동, 질병 휴직 등을 모두 취업으로 파악한 권혜진 외(2010)의 연구에서도 비슷한 맥락을 확인할 수 있다.

나. 설명변수(독립변수)

본 연구는 앞서 밝힌 연구 목적을 달성하기 위해 각 연도별 포함할 수 있는 모든 독립변수를 포함하고자 하였다. 1차년도부터 23차년도까지 모두를 포괄하는 변수는 총 150개이며, 이 중 근로시간, 직무만족도, 근로계약기간, 고용안정성, 근로시간, 임금, 사회보험 가입, 노조 등 현재 근로 중인 상태에서만 응답이 가능하거나 종속변수인 '취업여부'와 연관이 없는 변수들만 제외하고 47개의 변수를 선정하였다. 47개의 변수 중 중간에 추가되거나 삭제되는 설문 등으로 인해 연도별 14개에서 25개의 변수를 분석 대상으로 하였다. 분석에 포함한 전체 47개의 설명변수(독립변수)들만 정리하면 성별, 혼인상태 등의 개인특성 변수, 학력, 사회경제적지위 등의 사회적 배경 변수, 직업훈련경험, 자격증보유여부 등의 교육 관련 변수, 가족의 수입, 주거환경 등에 대한 생활만족도 변수, 사회경제적지위 향상 가능성에 대한 변수 등의 기타 변수이며, 세부 내용은 아래 <표 3>와 같다.

<표 3> 분석 대상 설명변수(독립변수)

구분	개인특성(16개)	사회적배경(16개)	교육 관련(8개)	만족도(6개)	기타(1개)
변수명	<ul style="list-style-type: none"> · 성별 · 거주지역 · 출생지(시도) · 만14세성장지(시도) · 동거여부 · 혼인상태 · 혼인상태변화여부 · 현재건강상태 · 현재장애,또는,장애여부 · 종교 · 군복무여부 · 부모생존여부 · 고등학교유형 · 고등학교소재지(국내/외) · 고등학교소재지(시도) · 대학교졸업여부 	<ul style="list-style-type: none"> · 사회적지위 · 사회경제적지위 · 차별경험유무(일자리취업시) · 친척의유무(대학교수) · 친척의유무(중앙일간지.방송) · 친척의유무(국장급이상공무원) · 친척의유무(의사) · 친척의유무(장성급이상군인) · 친척의유무(국회의원) · 친척의유무(대기업임원급) · 친척의유무(관사검사변호사) · 사회보험급여수급여부 · 아버지교육수준(학교) · 어머니학력 · 학력(학교) · 학력(전공계열) 	<ul style="list-style-type: none"> · 영어어학연수경험 · 영어독해능력 · 영어작문능력 · 영어능력향상을위한비용투자 · 토익.토플.토픽시험경험 · 직업훈련경험 · 자격증보유여부 · 직업훈련이외교육 	<ul style="list-style-type: none"> · 만족도(가족의수입) · 만족도(여가활동) · 만족도(주거환경) · 만족도(가족관계) · 만족도(친인척관계) · 만족도(사회적친분관계) 	<ul style="list-style-type: none"> · 사회경제적지위향상가능

자료: 한국노동패널 1차-23차 조사

한편, 패널 조사의 특성상 연도별 설문문항이 조금씩 달라지고, 추가로 패널을 모집하기도 하였는데, 각 연도별 활용한 독립변수의 수는 <표 4>에서, 분석 대상은 <표 2>에서 확인할 수 있다.

<표 4> 각 연도별 설명변수(독립변수)

구분	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차
대상	16	14	23	20	17	17	18	20	21	25	20
12차	13차	14차	15차	16차	17차	18차	19차	20차	21차	22차	23차
21	21	21	21	21	21	21	21	21	20	21	21

자료: 한국노동패널 1차-23차 조사

3. 연구 방법

가. 머신러닝 기법(랜덤포레스트)

수많은 머신러닝 기법 중 랜덤포레스트(Randomforest)는 단어에서도 알 수 있듯이, (의사결정) 나무들이 많이 모여있는 모형이라고 할 수 있다. 이렇듯 랜덤포레스트는 의사결정나무(Decision tree) 분석에서 파생된 분석 기법으로, 수백 개의 의사결정 나무를 생성하여 보다 안정적이고 정확한 예측을 하기 위해 활용된다. 랜덤포레스트의 모태가 되는 의사결정나무에서는 반응변수를 가장 잘 설명하는 설명변수로 가치를 뽑어나가도록 설계되어 있고, 설명변수가 범주형 변수인 경우에는 지니지수(Gini index)를 활용하고 연속형 변수인 경우에는 분산을 이용하여 최적의 설명변수를 찾아내기 때문에 랜덤포레스트 분석에서도 의사결정나무 분석과 마찬가지로 설명변수가 연속형이거나 범주형에 상관없이 활용할 수 있다는 장점이 있다. 하지만 랜덤포레스트에서는 의사결정나무와 같이 최종 결과를 직관적인 그래프로 얻기 어렵다는 단점이 있다(유진은, 2015).

랜덤포레스트는 설명변수가 많을수록 예측력이 높다는 특징이 있는데, 이는 안정적인 모형을 제공하는 장점으로 대표된다. 또한 랜덤포레스트에서는 중요도지수(variable of importance index)를 제시하는데, 반응변수에 대한 설명변수의 중요도(영향력)를 숫자나 그래프로 확인할 수 있도록 해준다(박창이 외, 2013). 이러한 중요도지수는 통계적 유의성이나 표준오차 등을 활용하기 때문에 그 변수의 영향력이 클수록 중요도지수가 크다고 볼 수 있다(Strobl et al., 2009). 따라서 랜덤포레스트 분석을 통해 그 결과값으로 중요도지수를 확인할 수 있고, 이 중요도지수가 클수록 영향력이 크다고 볼 수 있으므로, 설명변수 간의 상대적 영향력(중요도)를 분석할 수 있는 것이다.

이러한 랜덤포레스트의 주요 특징들을 종합해 봤을 때, 다수의 설명변수를 포함할 수 있다는 점, 다수의 설명변수를 포함함에도 불구하고 높은 예측력과 안정적인 모형을 제공한다는 점, 설명변수가 연속형인지 범주형인지에 관계없이 활용할 수 있다는 점, 다수의 설명변수들 간의 상대적인 중요도를 밝혀준다는 점 등에 있어 기존의 회귀분석, 구조방정식, 다층모형 등의 통계방법으로는 해결할 수 없는 다양한 분석을 시도할 수 있다는 점이 가장 큰 특징이라고 할 수 있다. 따라서 1999년부터 2020년까지 청년취업에 영향을 미치는 요인들을 밝히고 그 요인들 간의 상대적인 중요도를

분석함과 더불어 연도별 변화과정을 탐색하는 본 연구와 가장 적합한 연구방법이라고 할 수 있다.

나. 분석 절차

본 연구는 연구 목적을 위해 머신러닝 기법을 활용한 랜덤포레스트(Randomforest) 방법을 통해 분석하였다. 랜덤포레스트를 사용한 이유는, 일반적인 형태의 회귀분석을 통해서 나오는 계수값으로는 설명변수마다 단위가 다르기 때문에, 계수값이 높다고해서 그 변수가 중요하다고 이야기할 수 없다. 하지만 랜덤포레스트는 앞에서 언급한 대로 각 변수를 분할시키면서 모형의 성능향상 여부를 판단하기 때문에 변수중요도를 계산할 수 있다.

랜덤포레스트는 조절하는 주요 파라미터 값으로 트리의 개수인 *ntree*, 각 트리에 사용될 변수의 후보개수인 *mtry*가 있다. *ntree* 파라미터의 경우 값이 클수록 정확도는 개선되지만, 그 값이 커질수록 컴퓨터 연산수행시간이 증가하며 정확도 역시 일정수준 이상에서는 개선이 되지 않는다. *mtry* 파라미터는 크기가 작을수록 각 트리들이 모두 다르게 구성될 수 있으므로 예측력이 향상될 수 있으나(Guyon, I., & A. Elisseeff, 2003), 너무 작은 값을 선택하게 되면 가중치가 작은 변수들이 상위 노드에 위치하게 되면서 불순도가 높아질 가능성이 있다. 따라서 두 개의 파라미터 값에 대해 적절한 값의 선정이 중요한데 *ntree*는 적정값에 대한 기준은 없으나 김정오차가 안정화될 정도라고 인지하는 500으로 설정하였으며, 변수의 후보개수인 *mtry*의 경우 분류트리는 변수의 개수의 제곱근을 사용하였다(Han & Kim, 2019).

한편, 본 연구에서는 랜덤포레스트 패키지로 *randomForest* 패키지(Ver, 4.6-14)를 사용하여 분석하고 시각적 효과를 위해 *randomForestExplainer* 패키지(Ver, 0.10.1)를 추가로 활용하였다. *randomForestExplainer* 패키지를 활용하면 보다 많은 정보를 시각적으로 제공할 수 있는데, 일반적으로 사용되는 *randomForest* 패키지의 *varImpPlot* 함수에서는 y축에서 중요한 변수의 순서와 x축에서 *MeanDecreaseAccuracy*, *MeanDecreaseGini*의 두가지 값이 나오는데 이것은 랜덤포레스트 알고리즘에서 각 의사결정나무의 가지치기를 할 때 평균적으로 정확도가 얼마나 향상되었는지와 노드 불순도를 얼마나 개선하였는지를 보여준다. 하지만 이 값에 대한 절대적인 좋고 나쁨의 기준이 명확하지 않아 값에 대해 직접적인 활용이나 해석이 제한되는 부분이 많은데, *randomForestExplainer* 패키지에서는 중요한 변수의 순서뿐만 아니라 해당변수가 의사결정나무에서 몇 번째 상위노드에 위치하고 있는지 및 몇 개의 의사결정나무에서 상위 노드에 위치하였는지를 시각적으로 제시하여 주므로 해석적으로 보다 직관적인 시각화를 제공한다.

위에서 설명한 것처럼 랜덤포레스트가 변수우선순위를 제공하여 주지만, 이것은 설명변수가 상대적으로 얼마정도로 반응변수에 영향을 주는지, 어떻게 영향을 주는지는 알려주지 않는다. 따라서 구체적으로 설명변수의 효과를 확인하기 위해서는 회귀분석이 필요한데, 여기서는 반응변수가 취업여부로 이진변수이기 때문에 로지스틱 회귀분석을 이용하여 반응변수에 대한 설명변수의 효과를 추가적으로 확인하였다.

데이터 전처리 및 랜덤포레스트, 로지스틱 회귀분석과 관련된 코딩은 모두 R프로그램(ver 4.0.3)

을 활용하였으며 랜덤포레스트 연구모형은 아래와 같다.

- 1) $\hat{f}(x) = 0$ 으로 두고, 데이터 세트의 모든 개별관측치에 대해 잔차(r)를 $r_i = y_i$ 라 둔다
- 2) 총 횟수 $B(B = 500)$ 에 대해 반복수행 한다
 - (1) 출력변수를 y 가 아닌 r 로 두고, $d+1$ 개의 잎(끝마디)를 가진 의사결정나무 \hat{f}^b 를 적합
 - (2) $\hat{f}(x) \leftarrow \hat{f}(x) + \lambda \hat{f}^b(x)$
 - (3) $r_i \leftarrow r_i - \lambda \hat{f}^b(x_i)$
- 3) 최종 모형은 결과는 $\hat{f}(x) = \sum_{b=1}^B \lambda \hat{f}^b(x)$ 가 된다.

또한 로지스틱 회귀분석 관련 연구모형은 아래의 수식과 같다.

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \text{학력}'_j \beta_1 + \text{직업훈련경험}'_k \beta_2 + \dots + \text{종교}'_l \beta_n + \epsilon_i$$

IV. 연구 결과

1. 모형의 예측정확도

랜덤포레스트를 통해 확인된 변수의 중요도는 기본적으로 랜덤포레스트를 통해 구성된 모형이 좋은 성능을 가지고 있다는 것을 바탕으로 해야한다. 즉 설명변수들을 통해 반응변수에 대한 예측력이 높아야 해당 설명변수들이 반응변수들을 잘 설명하고 있다는 것을 보여주므로, 만약 설명변수가 반응변수를 잘 설명하지 못한다면 변수의 중요도 역시 의미가 퇴색될 수 밖에 없다.

본 연구에서는 각 차수별 데이터를 훈련데이터와 검증데이터로 7:3의 비율로 구분, 훈련데이터만을 가지고 랜덤포레스트 모형을 적합한 뒤 검증데이터의 설명변수들을 모형에 넣어 반응변수인 취업상태를 예측, 그 결과를 실제 검증데이터의 취업상태와 비교함으로써 정확도를 계산하였다.

모든 차수에 대한 랜덤포레스트 모형의 검증정확도 평균은 평균 69.39%로 약 70%의 정확도 성능을 보여주었으며, 이를 통해 연구에서 사용된 모형이 설명변수가 반응변수를 충분히 설명할 수 있는 정도라고 판단하였다. 각 차수별 랜덤포레스트 모형 예측정확도는 <표 5>와 같다.

<표 5> 랜덤포레스트 차수별 예측정확도

구분	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차
정확도(%)	75.2	70.7	68.3	75.2	67.2	69.5	75.6	68.7	68.9	66.1	67.4
12차	13차	14차	15차	16차	17차	18차	19차	20차	21차	22차	23차
69.9	57.5	72.1	66.7	70.9	69.8	69.8	67.5	69.4	70.3	71.5	67.8

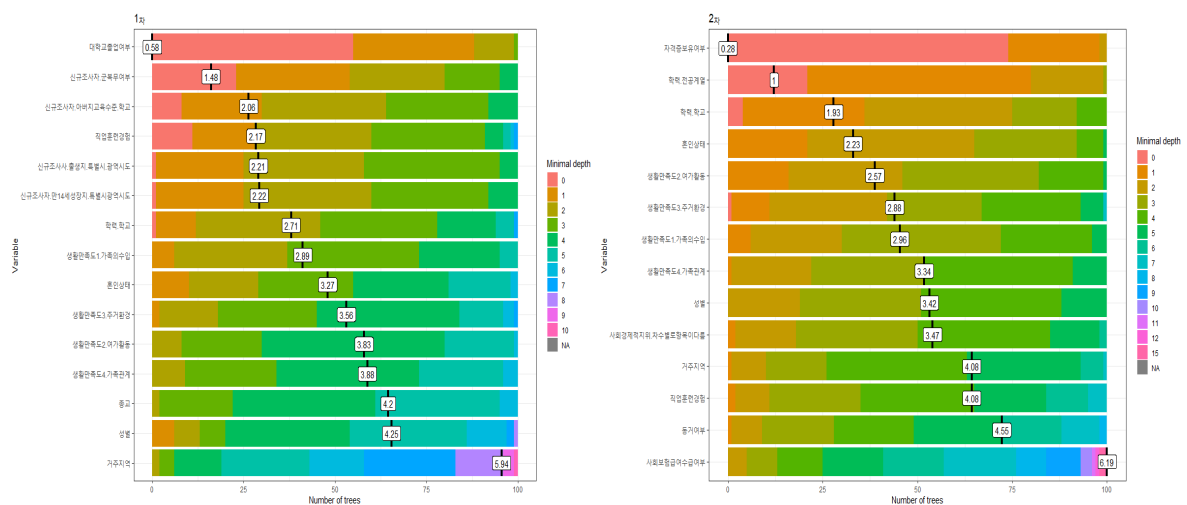
자료: 한국노동패널 1차-23차 조사

2. 연도별 청년취업 영향요인

가. 연도별 중요 영향변수

본 연구에서는 청년취업과 관련된 설명변수들의 랜덤포레스트 분석 결과를 아래 [그림 1]과 같이 확인하였다. 대표적으로 1차년도(1998년)와 2차년도(1999년)의 분석 결과만 포함하였으며, 나머지 연도의 세부 결과는 지면의 한계로 별도로 포함하지 않았다.

[그림 1] 랜덤포레스트 분석 결과(1-2차년도)



자료: 한국노동패널 1차-2차 조사

우선, 각 설명변수들의 중요도지수를 보여주는 랜덤포레스트 결과표를 살펴보면, 색깔로 구분되는 minimal depth는 그 값이 낮을수록 중요한 변수를 나타내는 것으로 해당변수가 상위 노드 0을 기준으로 몇 번째 노드에 위치하고 있는지를 알려준다. 앞서 설명하였듯이 랜덤포레스트는 여러개의 의사결정나무를 앙상블 기법으로 사용하는 알고리즘인데, 의사결정나무에서 데이터의 분류 및 회귀 시 가장 중요한 변수가 상위 노드에 위치하게 된다. 또한 분석결과와 가운데에 표시된 숫자는 minimal depth에 해당하는 값으로 해당 변수가 평균적으로 몇 번째 노드에 위치하였는지를 나타내며, 그 값이 낮을수록 중요한 변수라 할 수 있다. 분석결과와 y축은 위에서 아래의 순으로 취업여부에 영향을 주는 중요한 변수의 순서이며 x축은 트리의 수로 minimal depth의 색상과 함께 각 변수가 몇 개의 트리에서 minimal depth의 값을 가졌는지 정보를 제공한다.

1차년도(1998년)의 결과를 살펴보면, 설명변수 중 ‘대학교졸업여부’의 중요도가 가장 높은 것으로 나타났으며 다음으로 ‘근로여부’, ‘아버지 교육수준’, ‘직업훈련경험’, ‘출생지’, ‘성장지’ 순으로 중요도 지수가 높았다. 도출된 상위 15개의 변수는 대학교졸업여부, 출생지, 성장지 등의 ‘개인특성 변수’와 아버지 교육수준 등 ‘사회적 배경’ 변수, 그리고 직업훈련경험 등의 ‘교육 관련 변수’, 가족

의 수입에 대한 만족도 등의 ‘생활만족도 변수’로 구분할 수 있다.

본 연구에서는 모든 연도의 랜덤포레스트 분석 결과를 단순 나열하기보다, 분석의 용이함과 독자의 가독성 등을 고려하여 각 연도별 결과를 재분류하고 재구조화해서 제시하고자 한다. 먼저, 본 절에서 각 연도별 중요도가 높은 순서대로 상위 5개 변수들만 추려서 분석을 실시하고, 다음 절에서 연도별·변수별 중요 영향변수의 빈도를 중심으로 살펴볼 것이다. 변수중요도가 높은 상위 5개 변수들을 살펴보면, 아래 <표 6>와 같다.

<표 6> 연도별 변수중요도 상위 변수

구분	변 수 명
1차	대학교졸업여부, 군복무여부, 아버지교육수준, 직업훈련경험, 출생지
2차	자격증보유여부, 학력(전공계열, 학교), 혼인상태, 생활만족도(여가활동)
3차	학력(학교), 혼인상태, 생활만족도(사회적친분관계), 성별, 생활만족도(여가활동)
4차	대학교졸업여부, 고등학교유형, 어머니학력, 고등학교소재지, 혼인상태
5차	직업훈련경험, 학력(학교), 혼인상태, 성별, 생활만족도(여가활동)
6차	직업훈련경험, 학력(학교), 생활만족도(가족의수입), 혼인상태, 현재건강상태
7차	차별경험유무, 직업훈련경험, 학력(학교), 현재건강상태, 생활만족도(가족의수입)
8차	직업훈련경험, 학력(학교), 생활만족도(가족의수입), 사회경제적지위, 사회경제적지위향상가능
9차	직업훈련경험, 학력(학교), 혼인상태, 사회경제적지위, 성별
10차	직업훈련경험, 학력(학교), 영어능력향상을위한비용투자, 사회경제적지위, 현재건강상태
11차	학력(학교), 직업훈련경험, 사회경제적지위향상가능, 사회경제적지위, 현재건강상태
12차	학력(학교), 현재건강상태, 직업훈련경험, 혼인상태, 사회경제적지위향상가능
13차	학력(학교), 직업훈련경험, 혼인상태, 사회경제적지위, 현재건강상태
14차	학력(학교), 직업훈련경험, 혼인상태, 사회경제적지위, 현재건강상태
15차	학력(학교), 직업훈련경험, 사회경제적지위향상가능, 혼인상태, 현재건강상태
16차	학력(학교), 직업훈련경험, 현재건강상태, 사회경제적지위, 사회경제적지위향상가능
17차	학력(학교), 직업훈련경험, 현재건강상태, 사회경제적지위, 혼인상태
18차	학력(학교), 직업훈련경험, 현재건강상태, 사회경제적지위향상가능, 생활만족도(사회적친분관계)
19차	학력(학교), 현재건강상태, 직업훈련경험, 사회경제적지위, 사회경제적지위향상가능
20차	학력(학교), 직업훈련경험, 사회경제적지위향상가능, 사회경제적지위, 현재건강상태
21차	학력(학교), 직업훈련경험, 자격증보유여부, 사회경제적지위, 혼인상태
22차	학력(학교), 직업훈련경험, 사회경제적지위, 혼인상태, 사회경제적지위향상가능
23차	학력(학교), 혼인상태, 직업훈련경험, 사회경제적지위, 성별

자료: 한국노동패널 1차~23차 조사

연도별 변수중요도가 높은 변수들 위주로 살펴본 결과, 우선 ‘학력(학교)’, ‘직업훈련경험’, ‘혼인상태’, ‘사회경제적지위’ 등의 변수가 공통적으로 중요하게 나타났다. 이렇게 공통적으로 중요하게 나타난 변수들을 추가적으로 로지스틱 회귀분석을 실시한 결과, 학력이 높을수록, 직업훈련경험이 있을수록, 혼인한 상태일수록, 사회경제적지위가 높을수록 취업상태일 것이라는 결과를 얻을 수 있었다. 이 중 학력, 사회경제적지위와 같은 변수는 성별, 부모직업, 최종학교 등이 박성재의 연구(2005)에서 청년취업에 영향을 미치는 요인으로 밝혀진 바와 일치하는 결과를 나타냈으며, 혼인상태 및 직업훈련경험 변수는 김덕호의 연구(2020)의 결과와 비슷한 결과를 확인할 수 있었다.

또한, 패널 조사가 1년 단위로 이루어지다 보니 연도별 중요도가 높은 변수들이 비슷한 패턴을 보이는 경우가 있었다. 5차년도(2002년)부터 10차년도(2007년)까지는 ‘직업훈련경험’과 ‘학력(학교)’과 같은 변수들이 중요한 패턴을 보였다. 추가적인 로지스틱 회귀분석 결과, 2002년부터 2007년까지 꾸준히 직업훈련경험이 있을수록, 학력이 높을수록 취업한 상태일 것이라는 결과를 얻을 수 있었다. 2000년대 초중반에 해당하는 시기가 학력에 대한 인식이 중요하였으며, 2000년대부터 본격적인 직업훈련을 위해 각지에 직업훈련학교와 직업능력개발원 등을 설립하던 시기였다는 점을 고려했을 때, 해당 시기에 대한 이해를 풍부하게 해주는 결과임을 알 수 있다. 이후 11차년도(2008년)부터 23차년도까지는 ‘학력(학교)’가 가장 높은 변수중요도를 보였으며, ‘혼인상태’, ‘사회경제적지위’, ‘직업훈련경험’, ‘사회경제적지위향상가능’ 변수가 꾸준한 중요도를 나타냈다. 이들 변수 중 11차년도(2008년)부터는 ‘학력(학교)’가 가장 영향력이 큰 요인으로 등장하였다는 점과 직업훈련에 관한 정부의 꾸준한 노력으로 ‘직업훈련경험’이 여전히 취업에 영향력이 크다는 점, 그리고 특히 ‘사회경제적지위향상가능’이라는 변수에 주목할 필요가 있다. 나머지 변수들과는 달리 ‘사회경제적지위향상가능’ 변수는 본 연구에서만 관찰되는 변수로 사회경제적지위 향상이 가능할 것이라는 개인의 의지와 생각이 취업에 영향을 미친다는 것을 의미한다. 이는 사회과학에서 중요하게 여기는 ‘자기효능감’ 변수와 비슷한 맥락을 가진다고 볼 수 있으며, 노동시장에서 그동안 간과했던 변수임을 알 수 있다.

한편, 공통적인 패턴은 없었으나, 특정한 차수에만 나타난 변수들이 있었다. 1차년도(1998년)에서는 ‘대학교졸업여부’, ‘군복무여부’, ‘출생지’, ‘아버지교육수준’이 다른 연도에 비해 중요하게 나타났는데, 이는 대학교를 졸업했는지, 군대를 다녀왔는지, 집안은 어떠한지 등의 학연(學緣), 지연(地緣)과 같은 배경을 중시했던 당시의 시대적 배경으로 이해할 수 있다. 2차년도(1999년)부터는 꾸준한 패턴은 없지만 ‘생활만족도’ 변수가 종종 등장한다. 해당하는 각 연도의 변수간 로지스틱 회귀분석 결과, 사회적 친분관계 등에 관한 생활만족도가 높을수록 취업한 상태라는 결과를 확인할 수 있었다. 이러한 생활만족도에 관한 변수는 선행연구에서는 확인할 수 없는 변수로서 앞으로 중요하게 관심을 가져야 할 변수라고 할 수 있다. 또한 10차년도(2007년)에는 ‘영어능력향상을위한비용투자’ 변수가 등장한다. 해당 설문문항은 10차년도에만 실시되었는데, 앞선 연도에서 중요했던 변수들을 제치고 10차년도에서 영향력이 3번째로 큰 변수로 등장하였다. 앞서 황여정·백병부의 연구(2008)에서 비슷한 맥락의 변수로 ‘어학연수 비용’을 활용한 바 있으나, 본 연구에서 활용된 해당 연도 패널 조사에서도 어학연수 경험이나 비용에 관한 여러 문항이 활용되었는데, ‘어학연수’에 관한 변수보다

는 ‘영어능력향상을위한비용투자’ 변수가 더욱 중요하게 나타났다. 이를 통해 해당 변수가 취업과 관련하여 어학연수 또는 영어능력에 관한 변수 선정에 있어 중요한 의미를 가진다는 것을 알 수 있다.

나. 연도별·변수별 중요 영향변수의 빈도

본 연구에서 수행한 1차년도(1998년)부터 23차년도(2020년)까지의 각 연도별 랜덤포레스트 분석 결과표에 대해 연도와 변수에 따라 중요도가 높은 상위 15개 변수로 등장한 빈도를 정리를 하였으며, 그 결과는 <표 7>과 같다. 이를 통해 빈도가 가장 많은 변수 순으로 정리하면, ‘학력(학교)’, ‘생활만족도(가족의수입, 주거환경)’ 변수가 각각 23회씩으로 가장 많았다. 이들 변수는 총 23회의 분석 모두에서 꾸준히 나타났으며, 그만큼 1차년도(1998년)부터 23차년도(2020년)까지 꾸준히 중요한 변수라고 할 수 있다. 그 다음으로는 ‘생활만족도(여가활동, 가족관계)’, ‘혼인상태’, ‘직업훈련경험’ 변수가 각각 22회씩 나타났으며, ‘생활만족도(사회적친분관계)’ 변수는 21회, ‘성별’ 변수는 20회, ‘생활만족도(친인척관계)’, ‘사회경제적지위’ 변수는 각 19회, ‘현재건강상태’ 변수 16회, ‘사회경제적지위향상가능’, ‘동거여부’ 변수 각 14회, ‘종교’ 변수 13회, ‘거주지역’ 변수는 12회, 그 외 나머지 변수는 5회 이하로 나타났다.

이러한 빈도를 분석한 결과는 연도별 변수중요도 상위 변수만을 분석하였을 때와는 또 다른 결과를 보여주고 있는데, 특이점과 차이점에 기반하여 주요 변수를 중심으로 살펴보면 다음과 같다. 먼저, 선행연구에서 밝혀진 바와 마찬가지로 ‘학력’, ‘성별’, ‘사회경제적지위’, ‘직업훈련경험’의 변수는 꾸준히 중요하게 나타났다. 하지만, 선행연구와 비슷한 변수를 투입하였음에도 랜덤포레스트 결과표에 언급되지 않은 변수들이 있었다. 선행연구에서 중요하게 다루었던 ‘전공’, ‘출신지역’, ‘어학연수’, ‘출신고등학교’의 변수의 경우, <표 7>에서도 알 수 있듯이 본 연구에서도 분석에 포함하였으나, 중요도가 높은 변수에 나타나지 않았다. 이는 선행연구에서 활용되었던 변수들이 중요하지 않다는 것을 의미하는 것은 아니지만, 다른 유사한 변수들과 함께 고려되었을 때 상대적인 중요도가 낮게 나타났기 때문에, 관련된 변수 선정에 있어서 유의해야할 필요성을 찾았다는 데에 의의가 있다고 할 수 있다.

또한, 선행연구에서 다루어지지 않고 간과되었으나, 방대한 데이터를 기반으로 다수의 변수를 투입한 본 연구의 모형을 통해서 새롭게 등장한 변수들이 있었다. 먼저, ‘생활만족도’ 변수이다. 이 변수는 가족의 수입, 주거환경, 여가활동, 가족관계, 사회적친분관계, 친인척관계의 총 6개의 세부항목으로 측정되었는데, 이 6개의 항목 모두 23개년도 분석에서 19회 이상 등장한 중요한 변수이다. 이는 연도별 변수중요도 상위 변수에도 등장하는 변수로 청년취업과의 관계에 있어 중요하게 다루어져야 할 변수라고 할 수 있다. 또한, ‘사회경제적지위향상가능’ 변수 역시 연도별 변수중요도 상위 변수에 등장하는 변수로 빈도 분석에서는 총 23개년도 분석에서 14개년도에서 중요한 변수로 나타났으나, 해당 문항이 8차년도부터 조사된 것을 고려하였을 때 그 중요도는 더욱 크다고 할 수 있다.

한편, 이러한 빈도 분석을 통해 새롭게 등장한 의미있는 변수는 ‘종교’와 ‘거주지역’ 변수이다. 이 두 변수는 흔한 개념임에도 불구하고 선행연구에서도 잘 다루지지 않은 변수로 본 연구에서는 23개 연도 분석에서 각각 13회, 12회 등장하였다. ‘종교’ 변수는 신규조사자만을 대상으로 조사를 실시한 2차년도에서 11차년도를 제외하고는 모든 연도에서 중요한 변수로 나타났다. ‘종교’ 변수에 대한 추가적인 로지스틱 회귀분석 결과 12차수 이후 지속적으로 무교에 비해 불교를 가진 사람이 취업상태이고, 개신교에 비해 무교인 사람이 취업상태라는 결과로 분석되었다. 이는 종교여부, 세부적인 종교 종류와 무교 등의 관계에 대해 추가적인 연구를 통해서 밝혀져야 함을 시사한다.

‘거주지역’ 변수는 ‘서울, 부산, 대구, 대전, 인천, 광주, 울산, 경기, 강원, 충북, 충남, 전북, 전남, 경북, 경남, 제주도, 이북, 외국, 세종’ 총 19개 항목에 대한 문항으로, 여기서는 ‘서울, 인천, 경기’의 3개 지역은 수도권, 나머지 지역은 비수도권으로 분류하여 분석하였다. 로지스틱 회귀분석 결과 대부분의 차수에서 수도권의 취업률이 높은 상태로 확인되었다. 해당 변수 관련하여, 선행연구에서는 ‘출신지역’ 변수를 더 중요하게 다루었는데, ‘출신지역’과 ‘거주지역’ 모두 변수에 포함한 본 연구에서는 두 변수의 상대적 중요도가 현재 응답자가 거주하고 있는 지역이 더 중요하게 나타났다. 이는 청년취업과 관련된 연구에서는 변수중요도에 따라 적절한 변수를 활용할 필요가 있음을 시사한다.

한편, 총 23개 연도에 대한 랜덤포레스트 결과에 대해 각 연도별 1회라도 중요 변수로 등장한 변수만 추리면 총 47개 변수 중 37개 변수가 이에 해당한다. 이를 다시 해당 변수의 성격에 맞게 범주화하여 제시하면 아래 <표 8>과 같다.

<표 8> 중요 변수의 범주화

구분	개인특성(13개)	사회적배경(10개)	교육 관련(7개)	만족도(6개)	기타(1개)
변수명	<ul style="list-style-type: none"> · 성별 · 거주지역 · 출생지(시도) · 만14세성장지(시도) · 동거여부 · 혼인상태 · 혼인상태변화여부 · 현재건강상태 · 종교 · 군복무여부 · 고등학교유형 · 고등학교소재지(시도) · 대학교졸업여부 	<ul style="list-style-type: none"> · 사회적지위 · 사회경제적지위 · 차별경험유무(일자리취업시) · 친척의유무(의사) · 친척의유무(대기업임원급) · 사회보험급여수급여부 · 아버지교육수준(학교) · 어머님학력 · 학력(학교) · 학력(진공계열) 	<ul style="list-style-type: none"> · 영어독해능력 · 영어작문능력 · 영어능력향상을위한비용투자 · 토익,토플,토픽시험경험 · 직업훈련경험 · 자격증보유여부 · 직업훈련이외교육 	<ul style="list-style-type: none"> · 만족도(가족의수입) · 만족도(여가활동) · 만족도(주거환경) · 만족도(가족관계) · 만족도(친인척관계) · 만족도(사회적친분관계) 	<ul style="list-style-type: none"> · 사회경제적지위향상가능

자료: 한국노동패널 1차-23차 조사

<표 7> 연도별·변수별 중요 영향변수의 빈도

구분	1차	2차	3차	4차	5차	6차	7차	8차	9차	10차	11차	12차	13차	14차	15차	16차	17차	18차	19차	20차	21차	22차	23차	계	조사시점
대학교졸여부	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1차부터
군복무여부	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1차부터
아버지교육수준(학교)	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1차부터
직업훈련경험	1	1	-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	22	1차부터
출생지	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1차부터
성장지	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	2차부터
학력(학교)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	23	3차례의
생활만족도(가족의수입)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	23	1/3차례의
혼인상태	1	1	1	1	1	1	1	1	1	-	1	1	1	1	1	1	1	1	1	1	1	1	1	22	6차부터
생활만족도(주거환경)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	23	1차부터
생활만족도(여가활동)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	22	3차부터
생활만족도(가족관계)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	22	6차부터
종교	1	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1	13	1차부터
성별	1	1	1	1	1	1	1	1	1	-	1	1	1	1	-	1	1	1	1	1	1	1	1	20	1차부터
거주지역	1	1	1	-	1	1	1	1	-	-	-	-	-	1	1	1	1	1	-	-	-	1	-	12	1차부터
차격증보유여부	-	1	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-	1	5	1차부터
학력(전공제일)	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	3차부터
사회경제적지위	-	1	1	-	1	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19	3차부터
동거여부	-	1	-	-	-	-	-	1	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1	14	1/4, 6, 7차례의
사회보험금여유금여부	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	8차부터
생활만족도(사회적친분관계)	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	21	8차부터
생활만족도(친인척관계)	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	1	1	1	1	1	19	12차부터
친척의유무(의사)	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	2-23신규만
친척의유무(대기업임원급)	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	2-23신규만
혼인상태변화여부	-	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	2-23신규만
고등학교유형	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	2-23신규만
어머니학력	-	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	2-23신규만
고등학교소재지	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1/2차만
사회적지위	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	3/5차만
현재건강상태	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16	5-23신규만
직업훈련의교육	-	-	-	-	-	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	5	5-23신규만
차별경험유무(새로운일자리)	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	5-23신규만
사회경제적지위향상가능	-	-	-	-	-	-	-	1	1	1	1	1	-	-	1	1	1	1	1	1	1	1	1	14	5-23신규만
영어능력향상유한비용투자	-	-	-	-	-	-	-	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	14	7차만
작문	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	9차만
독해	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	10차만
토익토플테스시험경험	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	10차만

자료: 한국노동패널 1차-23차 조사

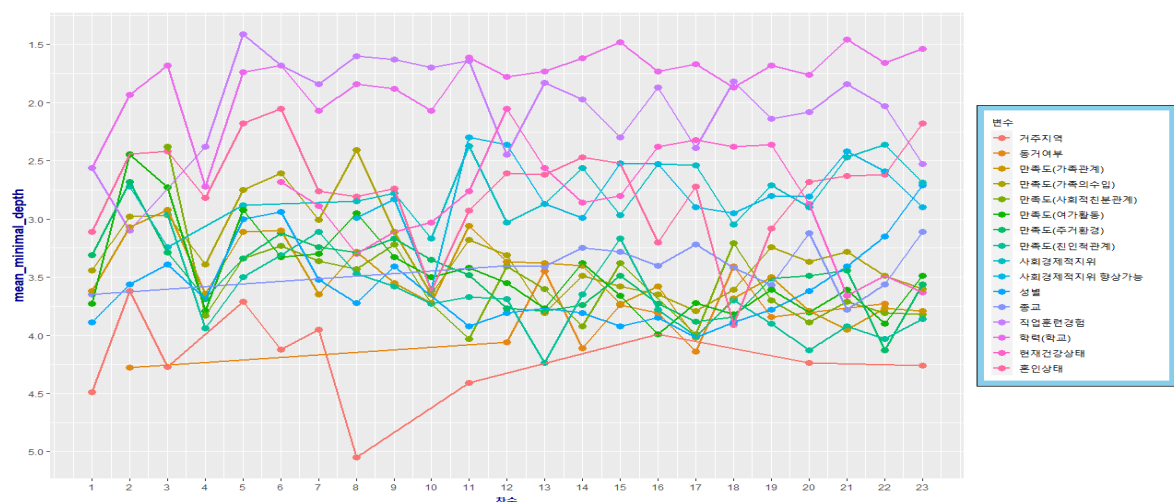
위 <표 3>에서 언급한 전체 분석 대상인 47개 변수에서 <표 7>의 중요 변수로 선정되지 않은 변수는 ‘현재장애또는장애여부’, ‘부모생존여부’, ‘고등학교소재지(국/내외)’, ‘친척의유무(대학교수, 중앙일간지방송, 국장급이상공무원, 장성급이상군인, 국회의원, 판사/검사/변호사)’, ‘어학연수경험’으로 총 10개 변수가 제외되었다.

3. 청년취업 영향요인의 변화과정

가. 주요 변수별 변화과정

청년취업에 영향을 미치는 주요 요인들의 연도별 변화과정을 살펴보기 위해 본 연구에서는 연도별 흐름과 변수의 중요도를 나타내는 지표를 반영하여 R프로그램의 ggplot2 및 ggthemes 패키지를 통해 그래프로 결과값을 출력하여 분석에 활용하였다. 위 <표 8>에서 1차년도부터 23차년도까지 빈도의 합계가 5회 이상인 변수 18개를 선정하였으며, 해당 변수는 ‘학력(학교)’, ‘생활만족도(가족의수입, 주거환경, 여가활동, 가족관계, 사회적친분관계, 친인척관계)’, ‘혼인상태’, ‘직업훈련경험’, ‘성별’, ‘사회경제적지위’, ‘현재건강상태’, ‘사회경제적지위향상가능’, ‘동거여부’, ‘종교’, ‘거주지역’이다. 우선, 해당 변수 전체를 포함한 결과는 아래 [그림 2]와 같다. 가로축은 각 조사연도를 나타내고 세로축은 앞서 설명한 minimal depth 값을 나타낸다. 하지만 그래프의 특성상 그래프 상위에 중요도가 높은 변수를 위치시키기 위하여 minimal depth 값을 역순으로 배열하였다.

[그림 2] 주요 변수별 변화과정



자료: 한국노동패널 1차-23차 조사

우선, 전체 영향변수 중에서 ‘학력(학교)’와 ‘직업훈련경험’ 변수의 중요도가 꾸준히 높게 나타났다. 이는 앞서 연도별 변수중요도 상위 5개 변수들만 비교한 결과나 연도별 중요 영향변수의 빈도

를 분석한 결과에서도 알 수 있었던 결과지만, 다른 변수들과 함께 연도별 흐름에 따른 그래프로 나타내니 세부적인 차이점을 명확히 확인할 수 있었다. 앞선 결과에서는 해당 변수가 꾸준히 중요하였으며, 연도별 중요변수로 등장한 빈도가 많았다 정도로 정리할 수 있었으나, 이번 결과를 통해서 해당 변수 간의 연도별 중요도의 차이를 확인할 수 있었다. 앞선 결과에서는, ‘학력(학교)’, ‘직업훈련경험’, ‘혼인상태’, ‘사회경제적지위’ 등의 변수가 공통적으로 상위 우선순위로 나타났는데, 위 [그림 2] 그래프를 살펴보면, ‘학력(학교)’와 ‘직업훈련경험’이 다른 변수들에 비해 높은 중요도를 보이며 꾸준히 상위 노드를 차지한 반면, ‘사회경제적지위’는 전체 변수 중에서 상대적으로 중간보다 약간 높은 수준의 중요도를 나타냈다. 이를 통해 영향변수 간의 우선순위에 있어서 각 변수 간의 중요도 차이가 존재하기 때문에 변수의 minimal depth 값도 함께 고려해야 함을 알 수 있었다.

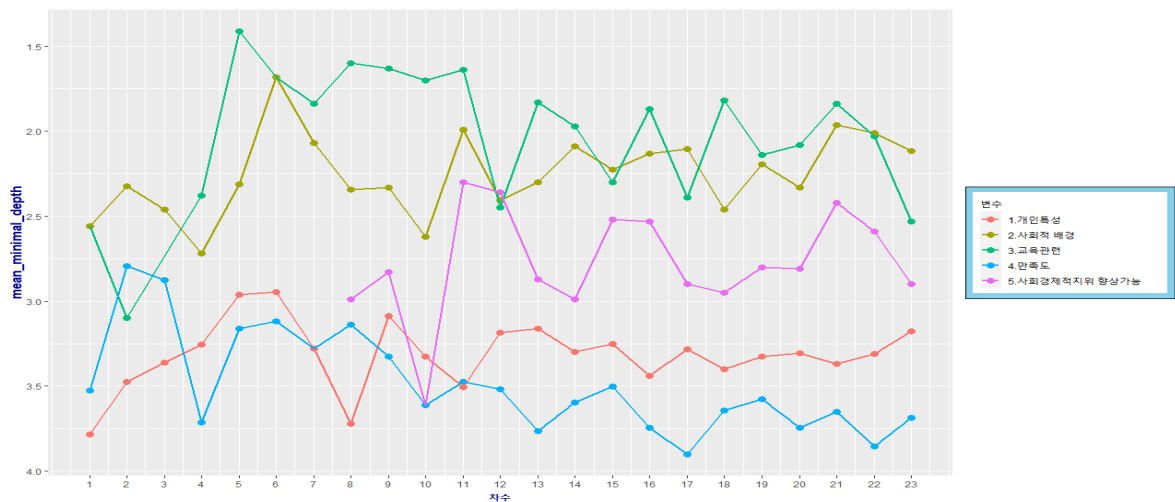
또한, 중요 변수로 등장한 빈도는 비슷한 변수들이 [그림 2]의 그래프 상에서는 확연한 차이를 보이는 변수도 있었다. ‘성별’, ‘사회경제적지위’, ‘사회경제적지위향상가능’, ‘혼인상태’ 변수는 높은 빈도로 중요하게 나타난 변수이지만, 이 중에서도 다른 변수들에 비해 ‘성별’이 상대적인 중요도가 낮게 나타남을 알 수 있다. ‘사회경제적지위’와 ‘사회경제적지위향상가능’ 변수는 서로 비슷한 중요도와 패턴을 보이고 있는데, 이를 통해 사회경제적지위의 향상에 대한 인식과 의지의 중요성을 다시 한 번 확인할 수 있었다.

한편, ‘만족도’에 관한 변수나 ‘종교’, ‘동거여부’ 등의 변수들은 비슷한 수준의 중요도와 패턴을 보이고 있었으며, ‘거주지역’ 변수는 전체 변수들 중에서 가장 낮은 중요도를 보였다. 이렇듯, 변수들 간의 상대적인 중요도를 확인함으로써 연구목적에 따라 해당 연구에서 활용해야 할 변수들을 적절히 판단할 수 있으며, 중요도에 따라 연구의 범위를 풍부하게 확장할 수 있음을 시사한다고 볼 수 있다.

나. 범주에 따른 변화과정

앞서 청년취업에 영향을 미치는 변수들의 연도별 변화과정 분석에 활용한 총 16개에 달하는 변수들을 <표 8>의 5개의 범주에 따른 연도별 흐름을 그래프로 나타내면 아래 [그림 3]과 같다. 해당 그래프는 각 범주에 해당하는 변수의 minimal depth 값을 평균하여 나타냈으며, 해당 연도에 조사되지 않았으면 해당하는 점에서 제외하였다.

[그림 3] 범주에 따른 변화과정



자료: 한국노동패널 1차-23차 조사

위 그림에서 알 수 있듯이, 각 변수들을 범주에 포함하여 비교하니 전혀 다른 결과가 도출되었다. 중요도가 높은 순으로 ‘교육 관련’, ‘사회적 배경’, ‘사회경제적지위향상가능’, ‘개인특성’, ‘만족도’ 순으로 나타났다. 앞선 결과들에서 높은 우선순위와 중요도 빈도를 보였던 ‘개인특성’ 변수들이 해당 변수들의 중요도 값을 합산하고 평균하여 비교함에 따라 다른 범주들에 비해 낮게 나타난 것이다. 물론, 해당 범주에 단일 변수만 포함된 범주도 있지만, ‘학력(학교)’ 변수가 꾸준히 높게 나타났던 결과를 고려하면 다소 특이한 결과라고 할 수 있다. 또한, 모든 범주를 놓고 비교하였을 때 교육 관련 변수인 ‘직업훈련경험’ 변수가 가장 중요도가 높게 나타난 점을 주목할 필요가 있다. 이미 가지고 있거나 노력해서 바꾸기 어려운 변수인 ‘개인특성’과 ‘사회적 배경’보다 ‘직업훈련경험’에 대한 중요도가 높다는 것은 청년들의 취업에 있어 ‘직업훈련경험’ 등의 노력이 더욱 중요하게 작용한다는 긍정적인 사회적 의미가 있다고 할 수 있다.

다. 변수별 시대에 따른 주요 특징

한편, 각 변수의 해당하는 범주 내에서의 분포는 어떠한지, 그리고 각각의 변수별 시대에 따른 주요 특징을 세부적으로 살펴보기 위하여, 각 범주에 속하는 변수들에 대하여 범주별로 구분하여 [그림 4]와 같이 나타내 보았다. 결국 [그림 2]의 전체 분포를 각 범주에 따라 [그림 4]로 나누어 분석해본 것이다.

[그림 4] 주요 변수별 변화과정



자료: 한국노동패널 1차-23차 조사

먼저, 개인특성 범주에서 각 변수별 분포를 살펴보면, ‘거주지역’, ‘동거여부’, ‘종교’ 변수는 중요도에 큰 변화없이 꾸준히 나타나고, 별다른 패턴을 보이지 않았으며, ‘성별’, ‘혼인상태’ 변수는 18차년도 이후 지속적으로 중요도가 증가함을 확인할 수 있었다. 또한, ‘현재건강상태’ 변수는 6차년도에 조사를 시작하는데 이어 12차년도에 정점을 찍고 별다른 패턴을 보이지 않다가 19차년도 이후

꾸준한 감소 패턴을 보였다.

다음으로, 사회적 배경 범주에서 ‘학력’ 변수가 4차년도(2001년)에 급격히 중요도가 감소함을 확인할 수 있다. 이는 2001년 당시 학력 인플레이션이 극심했던 상황과 정부에서 해당 시기부터 본격적으로 학력 구분없이 공공기관 신규 임용 경력을 인정했던 사회적 상황을 통해 짐작해볼 수 있다. 또한, IMF 경제위기 이후 기업들이 인건비 절감 등을 목적으로 비정규직 채용 비중을 늘리면서 대졸 이상 고학력에 적합한 일자리가 상당한 규모로 상실된 것도 원인 중에 하나라고 볼 수 있다(정영근 외, 2011). 한편, ‘학력’ 변수의 중요도는 최근 블라인드 면접 확대 등 학력 배제 노력과 풍토에도 불구하고 1차년도에 비해 23차년도 현재 시점에 오히려 증가하였음을 확인할 수 있다.

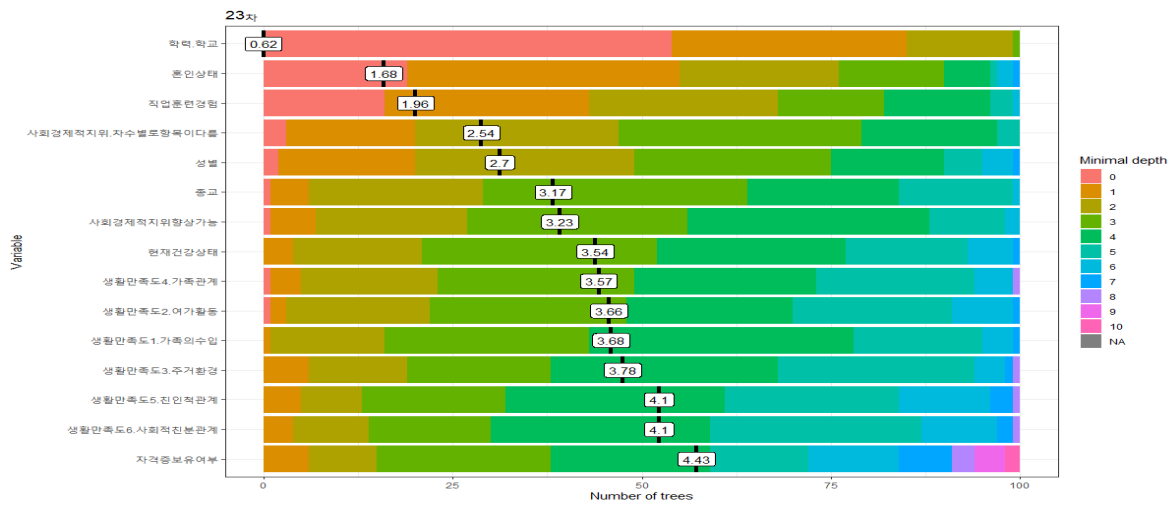
또한, 생활만족도 관련 6개의 변수의 연도별 흐름을 살펴보면, 변수들 상호간의 중요도에 있어서 ‘가족의수입’에 대한 만족도의 상대적 중요도가 다른 만족도 변수에 비해 전반적으로 높게 나타난 가운데, 각 변수들의 연도별 패턴이 비교적 유사하게 나타났다. 한편, 교육 관련 변수인 ‘직업훈련 경험’ 변수에서는 별다른 패턴을 확인하지 못하였다.

마지막으로, ‘사회경제적지위향상가능’ 변수에서는 특징적인 패턴으로 10차년도(2007년)에 그 중요도가 급격히 감소하였다는 점을 확인할 수 있었다. 해당 연도인 2007년 취업시장에서는 대졸자 정규직 취업률이 48.7%라는 조사결과가 있었는데, 이는 대졸자 2명 중 1명이 제대로 된 일자리를 찾지 못했다는 것을 의미하고, 이러한 현실을 반영하여 직장을 구하지 못하고 비정규직을 전전하는 20대를 가리켜 ‘88만원 세대’라는 신조어가 생겨나기도 하였다(세계일보, 2007). 이를 통해 해당 시기에 청년들의 좌절감이 극대화되었다는 사실을 확인할 수 있었는데, 이는 당시의 열악한 취업 시장 상황과 청년들의 좌절감이 팽배한 사회적 분위기로 인해 사회경제적지위 향상 가능성에 대한 인식에 부정적으로 작용함으로써 결국 영향변수의 중요도가 다른 연도에 비해 낮게 나타난 것으로 볼 수 있다.

4. 코로나-19와 청년취업

본 연구에서는 가장 최신의 데이터인 2020년도 자료를 바탕으로 2019년말 코로나-19 사태가 발발한 이후에 청년취업에 영향을 미치는 요인이 어떻게 변화하였는지 추가적으로 분석하였다. 통계지표로 살펴보면, 2020년 청년 고용률은 42.2%로 전년 대비 1.3%p 줄었는데, 이는 2010년 이후 가장 큰 폭으로 감소한 수치이다(통계청, 2021). 이때, 2020년도 청년취업에 영향을 미치는 요인들을 변수중요도에 따라 분석한 결과는 아래 [그림 5]와 같다. 변수중요도가 높은 순서에 따라 ‘학력(학교)’, ‘혼인상태’, ‘직업훈련경험’, ‘사회경제적지위’, ‘성별’, ‘종교’, ‘사회경제적지위향상가능’, ‘현재건강상태’, ‘생활만족도’, ‘자격증보유여부’ 순으로 나타났다.

[그림 5] 랜덤포레스트 분석 결과(23차년도)



자료: 한국노동패널 23차 조사

2020년도에 청년 고용률이 가장 큰 폭으로 줄어들었음에도 불구하고, 청년취업에 영향을 미친 주요 변수들은 타 연도와 큰 차이 없이 비슷한 패턴을 보였다. 이는 코로나-19로 인한 경기침체로 신규채용이 줄어들어 고용률이 낮아지는 등의 외부요인에 의한 결과로 볼 수 있다. 따라서 본 연구의 결과에서 2020년도 청년취업에 영향을 미치는 요인이 타 연도와 큰 차이가 없다는 점은 다른 외부 요인이 작용했다는 점을 반증하는 것을 의미하며, 그것은 코로나-19 상황적·환경적 요인이라고 할 수 있다. 그리고 나아가 이러한 상황적·환경적 요인이 변화하게 되면 청년취업이 다시 원래의 상태로 돌아올 것이라고 희망적으로 볼 수 있을 것이다.

V. 논의 및 결론

본 연구는 IMF 경제위기 직후인 1998년부터 코로나-19 대유행 시기인 2020년까지 총 23년에 걸쳐 방대한 데이터를 바탕으로 각 연도별 청년취업에 영향을 미치는 요인들을 상대적 중요도 관점에서 밝히고, 해당 요인들의 연도별 변화과정을 탐색하였다. 특히, 본 연구에서는 머신러닝 기법 중 하나인 랜덤포레스트(Randomforest) 기법을 활용하여 기존의 회귀분석, 구조방정식, 다층분석 등의 연구방법으로는 제한되는 범주형과 연속형 변수가 혼재되어 있는 총 47개의 변수를 투입하여 청년취업에 영향을 미치는 요인들 간의 상대적 중요도를 확인하였다. 주요 연구결과를 바탕으로 연구의 시사점과 함께 종합적으로 서술하면 다음과 같다.

첫째, 시장선호와 선별이론(Market Signaling and Screening Theory), 통계적 차별이론(Statistical Discrimination Theory) 등 노동시장과 관련된 기존의 이론들을 검증하였다. 우선, 연도별 변수중요도 상위 변수에서 개인특성 변수인 ‘혼인상태’, ‘성별’ 등을 확인할 수 있었고, 연도별·

변수별로 살펴본 중요 영향변수의 빈도 분석에서도 ‘혼인상태’, ‘성별’, ‘현재건강상태’, ‘종교’, ‘동거 여부’ 등의 개인특성 변수들을 확인할 수 있었는데, 이를 통해 개인의 특성에 따라 노동시장에서의 성과 차이가 발생한다는 통계적 차별이론을 지지하는 결과를 확인할 수 있었다. 또한, 연도별 변수 중요도 분석 및 연도별·변수별 중요 영향변수의 빈도 분석에서 공통적으로 ‘학력’ 변수와 ‘사회경제적지위’ 변수에 대한 중요도를 확인할 수 있었는데, 이를 통해 노동시장에서 개인의 교육수준을 바탕으로 개인의 능력을 평가하는 기준으로 작동한다는 시장선호와 선별이론과 사회경제적 배경을 주요 취업요인으로 보는 지위획득이론을 확인할 수 있었다. 그밖에도 각 분야에서 높은 중요도를 보인 ‘직업훈련경험’ 변수를 통해 개인의 경험, 교육 등이 축적됨에 따라 취업시장에서 더 나은 결과를 얻을 수 있다는 인적자본론의 내용을 확인할 수 있었다.

둘째, 전체 조사연도 23년에 걸쳐 꾸준히 중요하게 나타났던 변수들을 밝힘으로써 청년취업 정책의 시사점을 확인하였다. 앞서 연도별 변수중요도 분석, 연도별·변수별 중요 영향변수의 빈도 분석, 주요 변수별 변화과정 분석 등 다양한 관점에서 실시한 분석에서 꾸준히 공통적으로 중요한 영향을 미친 요인은 ‘학력’ 변수였다. 현재 노동시장에서는 학벌주의를 타파하고 학력으로 개인의 능력이 평가받는 사회적 관행을 배제하기 위해 블라인드 채용 등이 널리 퍼져있음에도 불구하고, 학력은 청년취업에 있어 여전히 중요한 변수였다. 실제 블라인드 채용을 실시하는 공공기관과 기업에서 상위권 대학 출신 합격자 비율이 이전과 비슷하다는 한국조세재정연구원의 연구결과(한국조세재정연구원, 2021)에서도 알 수 있듯이, 학력에 대해 단순히 배제할 것이 아니라 새로운 관점에서의 접근이 필요하다는 것을 시사한다고 볼 수 있다. 앞서 설명한 시장선호와 선별이론에서도 알 수 있듯이, 학력은 꾸준히 취업에 중요한 영향을 미쳐왔고 앞으로도 그럴 가능성이 높다는 사실을 알 수 있다. 앞으로 단순히 취업에 학력을 배제하기보다는 어떤 방식으로 형평성에 맞게 학력이라는 변수를 취업 평가요소에 반영해야 할지를 고민해야 할 필요가 있을 것이다.

셋째, 특정 연도에 특히 중요하게 나타난 변수나 중요도 패턴이 급격하게 변화한 패턴을 통해 당시의 시대적 배경을 풍부하게 이해할 수 있었다. 대표적으로, 1차년도(1998년)에서 ‘대학교졸업여부’, ‘군복무여부’, ‘출생지’, ‘아버지교육수준’ 변수가 다른 연도에 비해 중요하게 나타났는데, 이를 통해 대학교를 졸업했는지, 군대를 다녀왔는지, 집안은 어떠한지 등과 같은 학연(學緣), 지연(地緣)을 중시했던 1998년 당시의 시대적 배경을 이해할 수 있는 변수임을 확인할 수 있었다. 또한, ‘학력’ 변수가 4차년도(2001년)에 급격히 중요도가 감소했다가 다시 점차 증가했음을 확인할 수 있었는데, 이를 통해 당시 IMF 경제위기 이후 기업들이 인건비 절감 등을 목적으로 비정규직 채용 비중을 늘리면서 대졸 이상 고학력에 적합한 일자리가 상당한 규모로 상실된 상황으로 이해할 수 있었다(정영근 외, 2011). 그밖에, ‘사회경제적지위향상가능’ 변수의 특징적인 패턴으로 10차년도(2007년)에 그 중요도가 급격히 감소하였다는 점을 확인할 수 있었는데, 이를 통해 당시의 열악한 취업 시장 상황과 청년들의 좌절감이 팽배한 사회적 분위기로 인해 사회경제적지위 향상 가능성에 대한 인식에 부정적으로 작용하였음을 알 수 있었다.

넷째, 선행연구에서는 중요했으나, 본 연구에서는 중요하게 나타나지 않은 변수들을 통해 향후 연구의 방향성을 제시하였다. 박성재(2005), 황여정·백병부(2008) 등의 선행연구들과 비슷한 변수를

투입하였음에도 랜덤포레스트 분석 결과, 선행연구에서 중요하게 다루었던 ‘전공’, ‘출신지역’, ‘어학연수’, ‘출신고등학교’의 변수들은 본 연구에서는 중요도가 높은 변수에 포함되지 않았다. 이는 선행연구에서 활용되었던 변수들이 중요하지 않다는 것을 의미하는 것은 아니지만, 다른 유사한 변수들과 함께 고려되었을 때 상대적인 중요도가 낮게 나타났다는 것을 의미하며, 이는 관련된 변수 선정에 있어서 유의해야할 필요성을 찾았다는 데에 의의가 있다고 할 수 있다.

다섯째, 선행연구에서는 관심이 없었으나, 본 연구에서는 중요하게 나타난 변수들을 통해 연구의 지평을 확장하였다. 특히, ‘사회경제적지위향상가능’ 변수는 본 연구에서만 관찰되는 변수로 사회경제적지위 향상이 가능할 것이라는 개인의 의지와 생각이 취업에 영향을 미친다는 것을 의미한다. 이는 이하영·이수영(2016)의 연구에서처럼 자신이 노력한 만큼 보장받을 수 있다는 믿음이 긍정적인 동기부여로 이어지고, 이는 다시 사회발전의 원동력으로 작동하는 등의 선순환이 이루어질 수 있다고 볼 수 있다. 따라서 사회경제적지위 향상 가능성에 대한 변수는 노동시장에서 그동안 간과했던 변수임과 동시에 사회적 선순환을 이끄는 중요한 변수임을 시사한다고 할 수 있다. 또한, 가족의 수입이나 여가생활 등에 관한 ‘생활만족도’ 변수가 청년취업에 중요한 영향요인임은 선행연구를 통해서 밝혀진 적이 없었다. 생활만족도가 높을수록 조직몰입에 긍정적인 영향을 미친다는 백승규·이영면(2010)의 연구에서처럼 청년취업에도 중요한 영향요인임이 밝혀진 것이다.

한편, 본 연구에서 새롭게 등장한 또 다른 의미있는 변수는 ‘종교’와 ‘거주지역’ 변수이다. 이 두 변수는 흔한 개념임에도 불구하고 선행연구에서도 잘 다루지지 않은 변수로 본 연구에서는 23개 연도 분석에서 각각 13회, 12회 중요도가 높은 변수로 등장하였다. 이를 통해 ‘종교’를 가지고 있다는 사실이 취업여부에 중요한 영향을 미쳤음을 알 수 있었다. 또한, ‘거주지역’ 변수와 관련하여, 선행연구에서는 ‘출신지역’ 변수에 대해 주로 다루었는데, ‘출신지역’과 ‘거주지역’ 변수를 모두 포함한 본 연구에서는 두 변수의 상대적 중요도가 현재 응답자가 거주하고 있는 지역이 더 높게 나타났다. 이는 청년취업과 관련된 연구에 있어 비슷하게 여겨지는 변수라 하더라도, 변수중요도에 따라 적절한 변수를 활용할 필요가 있음을 후속연구에 시사한다고 할 수 있다.

여섯째, 2020년도 청년 고용률 저하가 코로나-19의 상황적·환경적 요인임을 확인하였다. 통계청 분석 결과, 2020년도에 청년 고용률이 가장 큰 폭으로 줄어들었음에도 불구하고, 본 연구에서는 청년취업에 영향을 미친 주요 변수들이 타 연도와 큰 차이 없이 비슷한 패턴을 보였다. 이러한 2020년도 청년취업에 영향을 미치는 요인이 타 연도와 큰 차이가 없다는 점은 다른 외부 요인이 작용했다는 점을 반증하는 것을 의미하며, 그것은 코로나-19로 인한 경기침체로 신규채용이 줄어들어 고용률이 낮아지는 등의 상황적·환경적 요인이라고 할 수 있다. 그리고 나아가 이러한 상황적·환경적 요인이 변화하게 되면 고용률이 다시 회복할 수 있을 것이라고 희망적으로 볼 수 있을 것이다.

일곱째, 본 연구의 결과에서 ‘직업훈련경험’이 청년취업에 있어 꾸준히 중요하게 나타난 점 등을 통해 각종 직업훈련에 관한 정책적 시사점을 확인할 수 있다. 본 연구에서는 직업훈련에 관한 경험의 유무가 청년취업 여부에 중요한 영향요인으로 확인되었는데, 이는 정부에서 청년들을 대상으로 시행 중인 일학습 병행, 산업계 주도 청년 맞춤형 훈련, 중소기업탐방 프로그램 등 청년들을

대상으로 한 다양한 직업훈련 정책에 긍정적인 시사점을 준다고 할 수 있다.

본 연구의 기여는 다음과 같다. 첫째, 방법론적 기여이다. 랜덤포레스트는 반응변수나 설명변수의 분포에 대해 가정을 하지 않는 비모수적 방법으로, 설명변수가 많고 설명변수 간의 관계가 복잡한 자료에 적합하고 예측력도 높다(Strobl et al., 2009). 본 연구에서는 범주형과 연속형 자료가 혼재되어 있는 전체 47개의 설명변수에 대해 분석을 실시하였는데, 이는 기존의 로지스틱 회귀모형과 같은 분석방법으로는 분석이 불가능하였다. 또한, 설명변수들 간의 상대적인 중요도를 확인함으로써 연구결과를 보다 풍부하게 활용할 수 있는 가능성을 열어주었다고 할 수 있다. 둘째, 기존 연구를 확장하였다는 점이다. 기존의 선행연구에서는 단순히 변수들 간의 관계를 밝히고, 영향의 정도가 유의한지, 구조적 관계는 어떠한지 등에 관심이 있었다면, 본 연구에서는 그러한 관계를 확장하여 변수들 간의 상대적 중요도를 확인하였다. 또한, 기존의 연구에서는 필요한 변수만을 선별해서 분석을 수행했다면, 본 연구에서는 방대한 데이터와 수많은 설명변수들 중에서 반응변수와의 관계를 가장 잘 예측해주는 변수들을 확인할 수 있었다. 이를 통해 해당 변수들에 대해 추가적인 분석을 수행할 수 있는 기반이 될 수 있는 등 연구의 확장성을 보여주었다고 할 수 있다. 셋째, 거시적 관점에서의 정책적 시사점을 제공한다는 점이다. 본 연구 결과에서처럼 방대한 데이터를 바탕으로 한 머신러닝 기법을 통해 지도자와 정책입안자들에게 다양한 시사점을 제공할 수 있다. 본 연구는 수십 년 간의 데이터를 기반으로 최대한 많은 변수를 포함하였고, 연구대상과 연구범위도 상대적으로 넓게 설정되어 보다 거시적인 관점에서 시사점을 제공할 수 있다.

한편, 본 연구의 한계는 다음과 같다. 첫째, 연구대상을 한정하지 못하였다. 연구대상을 정규직 또는 비정규직, 대졸자 또는 비대졸자 등으로 한정하여 분석을 수행하였다면, 보다 세부적이고 현재 실시 중인 정책 특화적인 시사점을 도출할 수 있을 것이지만, 본 연구의 목적은 청년취업에 영향을 미치는 요인들을 탐색적으로 밝히는 데에 있기 때문에 연구목적에 따라 연구대상을 한정하지 않았다. 향후 추가적인 연구에서는 연구대상을 한정하고 분석을 진행한다면 보다 세부적인 결론을 도출할 수 있을 것이다. 둘째, 정부의 다양한 노동시장 정책 변수를 활용하지 못하였다. 이행노동시장이론(Transitional labor market theory)에서 정부의 적극적 노동시장 정책들의 중요성을 밝힌 것처럼 정부의 노동시장 정책과 청년취업에 관한 영향력을 확인하지 못하였다는 점이 본 연구의 한계라고 할 수 있다. 이는 분석자료의 한계로서 본 연구의 분석자료인 KLIPS는 정책 변수로 직업훈련경험 변수만이 존재하기 때문에, 향후 노동시장 정책 변수가 포함된 새로운 데이터를 활용한다면 더욱 풍부한 결과를 도출할 수 있을 것이다. 셋째, 분석 시기의 한계이다. 본 연구는 IMF 경제위기 직후인 1998년부터의 데이터를 활용하였으나, 경제위기 이전의 자료와 비교 분석이 제한되었다. 후속 연구에서는 경제위기 이전 취업이 호황이었을 당시와 경제위기 이후의 영향 요인이 어떻게 달라졌는지 분석을 시도한다면 의미있는 시사점을 찾을 수 있을 것이라 생각한다.

참고문헌

- 김덕호(2020). 「청년 취업성공패키지 프로그램이 노동시장 성과에 미치는 영향」. 『노동정책연구』, 29-63.
- 김성훈(2015). 「대학생의 학업 외 취업 준비의 영향 요인과 취업 성과」. 『산업노동연구』, 21(3), 117-118.
- 김중성(2013). 「청년층 노동시장 이행의 계층화에 관한 연구: 사회계층별 부모의 개입 전략을 중심으로」. 중앙대학교 박사학위논문.
- 김중성·이병훈(2014). 「부모의 사회계층이 자녀의 노동시장 성과에 미치는 효과」. 『동향과 전망』, 90, 296-330.
- 김창환·김형석(2006). 「대학졸업자의 전공계열별 직업노동시장 성과: 이공계 위기의 노동시장 원인을 중심으로」. 『한국인구학』, 29(3), 1-27.
- 권혜진·권도희·안지혜(2010). 「대학졸업생 취업여부에 대한 영향요인 탐색: 대학 내·외 교육경험을 중심으로」. 『미래교육연구』, 23(2), 33-57.
- 금재호(2000). 「비정규직 함정은 존재하는가?」. 『노동동향분석』, 13(4), 37-53.
- 금재호(2010). 「여성의 경제적 지위는 향상되었는가? 성별 임금격차를 중심으로」. 『노동리뷰』, 2010(9), 48-69.
- 노경란·박용호·허선주(2011). 「진로개발프로그램 참여 경험이 대졸 청년의 능력개발 노력과 주관적 취업 만족에 미치는 영향」. 『고용직업능력개발연구』, 14(2), 55-81.
- 류장수(2003). 「지방대학 졸업생의 노동시장 이행 실태와 성과 분석: 수도권 대학 졸업생과의 비교」. 『산업노동연구』, 9(1), 171-196.
- 문영만·홍장표(2017). 「청년층의 노동시장 격차 및 지역인재 유출요인」. 『지역사회연구』, 25(2), 165-187.
- 박성재 (2005). 「지방대 졸업생의 노동이동과 노동시장 성과」. 『노동정책연구』, 5(4), 65-99
- 박창이·김용대·김진석·송중우·최호식(2013). 『R을 이용한 데이터마이닝(개정판)』, 서울: 교우사.
- 박환보(2011). 「대졸자 취업에 미치는 개인배경과 대학 특성의 영향」. 『직업능력개발연구』, 14(3), 1-25.
- 박환보·김성식(2011). 「개인배경, 취업준비능력, 대학서열유형이 대졸자의 노동시장 성과에 미치는 영향 분석」. 『교육사회학연구』, 21(3), 77-98.
- 백승규·이영면(2010). 「직무만족과 조직몰입의 관계에서 생활만족이 미치는 조절효과」. 『인적자원관리연구』, 17(4), 349-371.
- 연보라·이승진·장희원(2015). 「출신대학 소재지가 노동시장 성과에 미치는 영향: 첫 일자리를 중심으로」. 『지역능력개발연구』, 18(2), 1-24.
- 오호영(2007). 「대학서열과 노동시장 성과: 지방대생 임금차별을 중심으로」. 『노동경제논집』, 30(2), 87-118.

- 유진은(2015). 「랜덤 포레스트: 의사결정나무의 대안으로서의 데이터 마이닝 기법」. 『교육평가연구』, 28(2), 427-448.
- 이기종·김은주(2006). 「취업 및 미취업을 결정하는 요인 탐색: 교육적 관점에서 최종학교 생활을 중심으로」. 『교육논집』, 26(1), 17-37.
- 이병식·최정윤(2009). 「대학졸업자의 취업성가에 대한 개인 및 대학특성 영향요인 탐색: 수도권과 비수도권 대학 간 차이 분석」. 『한국교육』, 36(1), 191-215.
- 이영민·이수영·임정연(2014). 「대학생 선호직장의 취업 결정요인 분석: 취업준비행동의 역할을 중심으로」. 『사회과학연구』, 53(1), 337-374.
- 이하영·이수영(2016). 「사회계층이 국민의 희망 인식에 미치는 영향에 관한 연구: 계층과 정부신뢰의 매개효과를 중심으로」. 『한국정책학회보』, 25(3), 325-356.
- 정영근·장민수·김혜정(2011). 「대졸 고학력 노동시장 미스매치 연구」. 『질서경제저널』, 14(3), 19-39.
- 정태영·이기업(2005). 「대졸자 취업결정요인에 관한 연구: 학점이 취업확률에 미치는 영향을 중심으로」. 『Korea Business Review』, 8(2), 159-184.
- 주희정·김민석(2018). 「청년층 사회경제 활동에 따른 잠재계층 분류 및 영향요인 분석」. 『직업능력개발연구』, 21(2), 89-117.
- 채창균·김태기(2009). 「대졸 청년층의 취업 성과 결정 요인 분석」. 『직업교육연구』, 28(2), 89-107.
- 통계청(2021). 2020년 12월 및 연간 고용동향.
- 통계청(2021). 통계설명자료 : 경제활동인구조사.
- 한국노동연구원(2021). 1-23차 유저가이드.
- 한국조세재정연구원(2021). 『공공기관 채용정책에 대한 연구: 블라인드 채용제도의 도입효과 분석을 중심으로』.
- 한성민·문상호·이숙중(2016). 「청년취업 결정요인에서 본 청년 취업대책 개선책 연구」. 『정책분석평가학회보』, 26(2), 187-212.
- 황여정·백병부(2008). 「대졸 청년층의 노동시장 성과 결정요인」. 『직업능력개발연구』, 11(2), 1-23.
- Arrow, K.(1972). "Models of Job Discrimination", in Anthony, H. P.(eds.), *Racial Discrimination in Economic Life*, MA: Lexington Books.
- Becker, G. (1964). *Human Capital*. NBER. New York.
- Pascarella, E. T. & Terenzini, P. E. (2005). *How College Affects Students: A Third Decade of Research*. San Francisco: Jossey-Bass.
- Phelps, E. S.(1972). "The Statistical Theory of Racism and Sexism", *The American Economic Review*, Vol.62 No.4, pp. 659~661.

- Schmid, G. (1998). "Transitional labour markets : a new European employment strategy." SSOAR Discussion Paper.
- Spencer, A. M. (1973). "Job market signaling." *The Quarterly Journal of Economics*, 87(3), 355-374.
- Strobl, C., Malley, J., & Tutz, G. (2009). "An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests". *Psychological Methods*, 14, 323-348.
- Yoo & Rho. (2017). "TIMSS 2015 Korean Student, Teacher, and School Predictor Exploration and Identification via Random Forests", *The SNU Journal of Education Research*, 26(4), 43-61.
- Guyon, I., & A. Elisseeff. (2003). "An Introduction to Variable and Feature Selection". *Journal of Machine Learning Research*, Vol.3(Mar, 2003), 1157-1182.
- Han & Kim. (2019). "On the optimal size of candidate feature set in random forest". *Applied Sciences*, 9(5):898, 2019.
- 세계일보(2007). 월별로 살펴본 2007년 취업 뉴스. 세계일보(2007.12.26.). <https://www.segye.com/newsView/20071226000303>

