

머신러닝을 활용한 특성화 고등학교 취업률 예측 모델 개발 연구

연구 책임자 : 정광훈 연구위원(KERIS)
공동 연구자 : 장시준 연구위원(KERIS)
이종현 책임연구원(KERIS)
김찬희 연구원(KERIS)
이운지 교수(대전과학기술대학교)
윤지아 교수(충남대학교)
연구 보조원 : 장휘창 연구원(영남대학교)
과제 기획자 : 김한성 선임연구원(KERIS)
과제 책임자 : 계보경 부장(KERIS)

본 연구에서 제시된 정책적 대안이나 의견 등은 한국교육학술정보원의
공식 의견이 아니라 연구진의 견해를 밝힙니다.

차 례

I. 서 론	1
1. 연구의 필요성 및 목적	1
2. 연구의 내용 및 범위	4
3. 연구 방법	5
II. 배경 문헌 분석	7
1. 특성화고 관련 배경 문헌 분석	7
가. 특성화고 개념과 특성	7
나. 특성화고 현황 및 관련 정책	8
다. 특성화고 취업률 예측 요인	21
라. 소결	34
2. 머신러닝 관련 배경 문헌 분석	36
가. 머신러닝의 개념과 주요 알고리즘	36
나. 머신러닝 기반 교육 분야 연구 사례	44
다. 머신러닝을 위한 도구	47
라. 소결	48
III. 머신러닝 기반 특성화고 취업률 예측 모델 개발	51
1. 개요	51
2. 분석 방법 및 절차	51
가. 연구모형 설계	51
나. 분석 대상 데이터 수집	52
다. 데이터 전처리	53
라. 기술통계 분석	54
마. 예측모형 개발	55
바. 데이터 마이닝	55
3. 분석 결과	56
가. 기술통계 분석	56
나. 예측모형 성능 분석	66
다. 데이터 마이닝 결과 분석	73
4. 소결	79

IV. 특성화고 취업률 예측 서비스 Prototype 설계	84
1. 개요	84
2. 설계 방향 및 원리	84
가. 서비스 기초정보	84
나. 머신러닝 활용 유사 서비스 구축·운영 사례	87
다. 최종 설계 원리	90
3. 서비스 프로토타입(안)	91
가. 구성요소	91
나. 화면구현	93
4. 소결	103
V. 요약 및 결론	104
1. 주요 결과 요약	104
2. 결론 및 논의	106
참고 문헌	109

표 목 차

<표 I-1> 연구 절차 및 방법(요약)	6
<표 II-1> 직업계 고등학교 학교 및 학생 현황	10
<표 II-2> 직업계 고등학교 계열별 주요 현황	10
<표 II-3> 교육 희망사다리 구축 프로젝트 전문계고 학생 취업 지원 강화 세부 사업	14
<표 II-4> 특성화고·마이스터고 관련 지원 사업·제도	17
<표 II-5> 각종 정책별 세부 추진 내용	20
<표 II-6> 특성화고 취업률 영향요인 관련 문헌분석 종합	30
<표 II-7> 특성화고 취업률 예측 요인 도출	35
<표 II-8> 머신러닝 예측모형 성능 평가 관련 개념	37
<표 II-9> WEKA 지원 알고리즘의 주요 기능 및 특징	49
<표 III-1> 고등학교 학교회계 예산 공시 데이터('09년~'18년)	53
<표 III-2> 수집 데이터 전처리 주요 과정	54
<표 III-3> 기술통계량 분석	56
<표 III-4> 시·도별 데이터 빈도분석	58
<표 III-5> 시군코드 빈도분석	59
<표 III-6> 지역 구분 빈도분석	64
<표 III-7> 최근 9년간 시·도별 특성화고 취업률 평균값	65
<표 III-8> 머신러닝 알고리즘별 예측모형 성능분석(하위 25% 기준 값 기준)	66
<표 III-9> 머신러닝 알고리즘별 예측모형 성능분석(중위값 기준)	67
<표 III-10> 머신러닝 알고리즘별 예측모형 성능분석(상위 25% 기준 값 기준)	68
<표 III-11> 머신러닝 알고리즘별, 기준별 ROC area값 총괄표	69
<표 III-12> Random Forest 예측모형 최적화(하위 25% 기준 값 기준)	70
<표 III-13> Random Forest 예측모형 최적화(중위값 기준)	71
<표 III-14> Random Forest 예측모형 최적화(상위 25% 기준 값 기준)	72
<표 III-15> 변인별 영향력 분석(하위 25% 기준 값 기준)	73
<표 III-16> 변인별 영향력 분석(중위값 기준)	74
<표 III-17> 변인별 영향력 분석(상위 25% 기준 값 기준)	75
<표 III-18> 기준 값에 따른 주요 변인별 영향력	76
<표 III-19> 취업률 하위 25% 기준 주요 변인별 평균값 t검정 결과	77

<표 III-20> 취업률 중위값 기준 주요 변인별 평균값 t검정 결과	78
<표 III-21> 취업률 상위 25% 기준 주요 변인별 평균값 t검정 결과	79
<표 III-22> 특성화고 취업률 예측 모델 개발에 필요한 독립변인	79
<표 IV-1> 서비스 프로토타입 구성요소(안)	91
<표 IV-2> 서비스 프로토타입 최종 구성요소(안)	92

그 립 목 차

[그림 II-1] 직업교육 선진화 방안의 전문계고 체제 개편 방안	11
[그림 II-2] 학업·취업 병행 교육 체제 구축 방안의 추진 방향	15
[그림 II-3] 공생발전을 위한 열린 고용사회 구현방안의 추진 과제	16
[그림 II-4] WEKA 메인 GUI 화면	47
[그림 III-1] 연구모형	52
[그림 III-2] K-Fold 교차 유효성 검증 방법론	55
[그림 III-3] 시·도별 데이터 빈도분석 도표	59
[그림 III-4] 지역 구분 코드 빈도분석	64
[그림 III-5] 최근 9년간 시·도별 특성화고 취업률 평균 시각화 도표	65
[그림 III-6] 변인별 영향력 분석결과에 따른 취업률 수준별 처방 방안	82
[그림 IV-1] 데이터 흐름도 예시 (EDS 적재 데이터 활용 시)	86
[그림 IV-2] 데이터 흐름도 예시 (학교 보유데이터 활용 시)	86
[그림 IV-3] 씨밋수학 서비스 화면	87
[그림 IV-4] 닥터매쓰 서비스 화면	88
[그림 IV-5] 산타토익 서비스 화면	89
[그림 IV-6] 서비스 메인화면	93
[그림 IV-7] 서비스 안내·학교검색 - 서비스 목적 및 개요	94
[그림 IV-8] 서비스 안내·학교검색 - 학교선택(검색)	95
[그림 IV-9] 서비스 안내·학교검색 - 학교선택(지도)	96
[그림 IV-10] 서비스 안내·학교검색 - 보유 데이터 업로드	97
[그림 IV-11] 예측결과 - 예측결과 리포팅	98
[그림 IV-12] 예측결과 - 취업률 예측 변인 값	99
[그림 IV-13] 예측결과 - 알고리즘 안내	100
[그림 IV-14] 모의계산 - 목표수준 선택	101
[그림 IV-15] 모의계산 - 변인 값 조정	102

I. 서 론

1. 연구의 필요성 및 목적

특성화고등학교는 특정 분야의 인재 양성을 목적으로 설립된 학교이다(초·중등교육법시행령 제91조, 이하 특성화고). 2019년 기준 총 513개(초·중등교육법시행령 제90조에서 규정한 마이스터고 포함)로 국내 전체 고등학교 2,360개의 21.7%를 차지하며, 학생수를 기준으로 245,085명으로 전체의 17.4%가 재학중이다.

‘학생 개개인의 소질과 적성에 맞는 교육을 통해 우수한 인재를 양성하고 좋은 일자리에 취업할 수 있는 학교’(특성화고·마이스터고 포털, 2020)를 장점으로 내세우는 특성화고는 학생들에 대한 취업 지원이 가장 중요한 교육 활동 중 하나이다. 고졸 기능 인력을 적재적소에 배치하기 위한 취업 지원은 학교 자체의 목표일뿐 아니라 정부 정책에서도 강조되고 있다. 교육부는 2019년 「고졸취업 활성화 방안」을 통해 ‘고졸도 성공할 수 있는 경로 구축은 아이들의 다양한 꿈 실현, 경쟁 중심 교육 완화 및 일자리 문제 해결의 중요한 열쇠’라는 점을 강조하면서 2022년까지 특성화고 취업자 비율 60% 달성을 목표로 제시한 바 있다(교육부, 2019). 특성화고 취업은 그간 중등직업교육에 대한 범정부 차원의 지원과 기업의 노력으로 2010년 19.2%에서 2017년 50.6%로 지속적으로 확대되어 왔다(교육부, 2019). 그러나 최근 ‘17개 교육청 모두 취업률이 하락(평균 9.5%)한 가운데 일부는 20% 이상 큰 폭으로 취업률이 하락(교육부, 2020)’하였다는 점에서 보다 효과적이고 정교한 특성화고 취업 관련 정책을 수립하기 위한 예측모델을 개발하여 활용할 필요가 있다.

특성화고 학생의 취업 관련 선행 연구들은 크게 학생 개인의 특성과 관련된 요인과 학교 차원의 요인으로 범주화하여 취업률 영향 요인을 다루고 있다. 개인특성은 성별, 성적, 자기효능감, 학업중단율 등이 제시되었으며, 학교 차원

에서는 학교 소재지, 환경, 동아리 운영, 방과후학교 참여여부, 진로프로그램 운영 등이 대표적 요인에 포함되어 연구가 이루어져 왔다(김지효, 2018; 노경란, 허선주, 2012;; 이쌍철, 엄문영, 2014; 이지혜, 정철영, 2010; 장현진, 이지혜, 2012; 조규형, 정철영, 2014; 천영민, 오민홍, 2015; 홍성표, 정진철, 2016). 연구 결과는 주로 해당 요인이 취업에 영향을 주는 유의성을 검증하는 형태로 이루어졌다. 개별 요인이 취업률에 영향을 미치는 유의성 검증 결과는 연구 기간, 연구자, 연구에 활용된 데이터의 차이에 따라 다른 결과를 나타내기도 한다. 예를 들어 장현진 외(2012)의 연구에서는 학교계열이 취업률에 영향을 미치는 요인으로 유의성이 있다고 하였으나 이쌍철 외(2014)의 연구에서는 유의성이 없는 것으로 보고하고 있다. 가구소득의 경우 장현진 외(2012)의 연구와 이쌍철 외(2014)의 연구는 부적 영향을 미친다고 보고하고 있으나 보호자 소득으로 표현한 천영민 외(2015)의 연구에서는 정적 영향을 미친다고 보고하고 있는 것을 발견할 수 있다. 이와 같은 연구 결과는 특성화고 취업에 영향을 미치는 특정 요인을 유의성을 통해 개별적으로 설명하는 것은 가능하지만, 개인과 학교 특성을 복합적으로 고려한 종합적인 시각을 보여주는 데는 한계가 있다.

연구 대상이 되는 데이터는 일부 학생 대상의 설문조사(조규형 외, 2014; 홍성표 외, 2016)를 활용하거나 고졸자취업진로조사(HSGOMS), 한국교육고용패널(KEEP)과 같은 패널조사(노경란 외, 2012; 최동선 외, 2013; 이쌍철 외, 2013; 천영민 외, 2015) 또는 시·도교육청의 종단연구(김지효, 2018)를 활용하였다. 데이터의 크기에 차이가 있기는 하지만 대부분 표집을 활용하는 연구이다. 표집 연구는 방법의 효율성에도 불구하고 대표성 있는 표집을 선택하지 못할 경우 연구 결과의 의미가 상실될 수 있으며, 연구를 설계할 때 모집단의 동질성, 표집 크기의 적정성, 표집방법의 일관성 등은 물론 현실적인 시간과 비용도 고려하여야 하는 어려움이 존재한다. 설문조사와 같은 중소 규모 표집 연구의 제약에 대한 해결책의 하나로 대규모 자료를 활용하는 다년간의 패널조사나 종단연구가 지속적으로 늘어나고 있으며, 특성화고 취업 관련 선행연구도 그 추세를 반영하고 있다. 서울시교육청의 종단연구나 한국교육고용패널과 같은 대규모 패널조사의 사례에서 볼 수 있듯이 데이터의 규모 자체로만 본다면

빅데이터를 특징짓는 요소 중 하나인 대규모 데이터를 이미 오래전부터 연구에 활용해 왔다고 할 수 있다. 그러나 표집 규모의 확대는 사회과학 방법론의 근본적인, 전통적인 틀을 그대로 유지하고 있다는 점에서는 차이가 없으며, 그 보다는 데이터의 질적 변화의 측면과 함께 분석 전략을 새로운 방식으로 재구성하여야 한다는 데 주목할 필요가 있다(한신갑, 2015).

빅데이터와 머신러닝 기술의 활용은 사회과학 연구에 대한 새로운 접근으로 전통적인 연구의 한계를 극복할 수 있는 가능성을 보여주고 있다. 머신러닝은 컴퓨터가 머신러닝 알고리즘을 활용하여 데이터를 학습한 후 향후 발생할 일을 예측하는 것으로 최근 교육 분야 연구에서도 국가 수준 학업성취도 등의 자료를 분석하여 교육정책에 활용하는 방안들이 연구되고 있는 등 활용 저변을 넓혀가고 있다(이종현, 2019). 머신러닝은 연구 대상 데이터를 분류, 예측, 군집 알고리즘에 입력하여 예측모형을 개발하는 데 주목적이 있다. 즉, 개별 요인의 유의성을 검증하는 것이 아니라 요인간의 관계와 영향력을 학습하여 종속변인(본 연구의 경우 취업률)을 가장 잘 예측하는 모델 개발을 통해 각 요인간의 관계를 복합적으로 고려하는 종합적인 시각의 연구 결과를 도출해낼 수 있다.

본 연구는 선행연구를 통해 특성화고 취업률에 영향을 미치는 요인을 종합하여 분석한 예측모형을 개발하고 정책적 활용 방안을 모색하는 데 목적이 있다. 구체적으로 2009년부터 2018년까지 10년간 축적된 학교정보공시와 학교회계 공시자료 약 2천만 건의 데이터를 머신러닝 알고리즘으로 분석함으로써 전통적인 방식으로 수행된 선행 연구와 차별화된 연구 분석 방법의 전환을 시도하였다. 또한 연구결과로써의 예측모형 제시를 통해 특성화고 취업률 관련 정책 개선을 위한 시사점을 도출하고자 하였다.

2. 연구의 내용 및 범위

본 연구는 머신러닝 알고리즘을 활용하여 특성화고 취업률 예측모형을 개발하고 이를 활용하기 위한 서비스 프로토타입(prototype) 설계를 통해 특성화고 취업 정책 관련 시사점을 도출하는 것을 목적으로 하고 있다. 연구의 내용은 배경 문헌 분석, 머신러닝 기반 특성화고 취업률 예측 모델 개발, 특성화고 취업률 예측 서비스 프로토타입 설계로 구성되어 있다.

먼저 II장에서는 특성화고 관련 배경 문헌 및 선행연구와 머신러닝에 대해 살펴보았다. 특성화고의 특성과 기본 현황을 살펴보았으며, 고등학교 직업교육 선진화 방안, 학업·취업 병행 교육체제 구축 방안 등 2010년부터 꾸준히 이어져온 특성화고 관련 지원사업, 정책, 제도 등을 살펴보았다. 이어 선행연구 분석을 통해 특성화고 취업률 관련 예측 요인을 고찰하였다. 예측요인은 크게 개인특성과 학교 차원의 요인을 살펴보았으며, 학교 차원은 교육, 지도, 학생 참여 활동 등의 교수학습 요인과 학교환경 요인으로 구분하여 분석하였다. 전통적인 연구 방법을 전환하는 방안으로 머신러닝의 개념과 주요 분석 알고리즘도 살펴보았다. 머신러닝을 기반으로 하는 교육 분야 연구 사례로 학습 성취 예측모형 연구, 취업률 및 교육성과 예측모형 연구 등도 분석하였으며, 연구에 활용할 머신러닝 분석 도구에 대해서도 살펴보았다.

III장에서는 머신러닝 기반의 특성화고 취업률 예측모형을 개발하고 그 성능을 분석하였다. 연구모형 설계와 분석 대상 데이터 수집, 데이터 전처리, 예측모형 개발 등 분석 방법 및 절차를 설명하였으며, 분석 결과를 제시하였다. 분석 결과로 먼저 예측모형 개발과 데이터마이닝 분석을 위해 변인에 대한 기술통계량 분석 결과를 제시하였으며, 13개 머신러닝 알고리즘을 활용하여 개발한 예측모형 제시와 이 예측모형에 대한 성능 분석 결과를 설명하였다. 이어 데이터마이닝 결과 분석을 통해서는 취업률 하위 25% 기준 값, 중위값, 상위 25% 기준 값을 기준으로 변인별 영향력을 분석한 결과를 설명하였으며, 분석 결과를 종합한 시사점을 도출하였다.

IV장에서는 III장에서 개발하고 성능 분석을 거친 예측모형을 교육현장에서

활용하기 위한 서비스 프로토타입을 설계하였다. 프로토타입은 III장에서 개발한 예측모형 결과를 바탕으로 서비스 사용자 범위 및 내용을 설정하였다. 이 과정에서 유사 서비스 구축·운영 사례 분석 결과를 종합하여 프로토타입 설계 방향 및 원리를 도출하였다. 이를 실제 화면으로 나타내기 위해 화면 요소(User Interface: UI)를 설계하였다. 화면 요소에 대한 타당도 확보를 위해 전문가 자문과 검증을 진행하였다.

V장에서는 연구 결과 요약에 토대로 정책적 시사점을 설명하였으며 본 연구가 갖는 한계점과 향후 추가적인 연구 방향을 제시하였다.

3. 연구 방법

본 연구에서는 배경 문헌 및 선행 연구 분석을 통해 살펴본 요인을 토대로 연구모형을 설계하였으며, 머신러닝 분석 절차에 따라 연구 결과인 예측 모형을 도출하였다.

연구의 타당도와 신뢰도를 확보하기 위한 방안도 마련하였다. 연구하고자 하는 내용을 정확하게 규명하는가와 관련되어 있는 구인 타당도(construct validity)를 확보하기 위해서 연구모형에 사용된 예측요인은 선행연구에서 사용된 요인을 매칭하여 구성하였다. 일반화 가능성과 관련된 외적 타당도(external validity) 확보를 위해 머신러닝을 활용하는 교육 연구 사례를 참조하여 분석 절차를 제시하였으며, 이 절차를 적용한 실증적 분석을 통해 유사한 연구에 활용할 수 있음을 확인하고자 하였다. 본 연구에서 신뢰도(reliability)는 분석하고자 하는 내용을 신뢰할 수 있는 자료를 사용하여 정확하게 바라보고 있느냐는 관점에서는 구인 타당도와도 밀접한 관련이 있다. 본 연구에서는 10년간 축적된 공적·개관적 데이터를 활용하여 신뢰도를 확보하고자 하였으며, 연구 결과로 도출된 예측모형에 대해서는 K-Fold 교차 유효성 검증 방법론을 활용한 머신러닝 알고리즘 유효성 검증을 통해 신뢰도를 확보하고자 하였다.

본 연구에서 사용한 절차와 방법은 다음과 같다.

<표 I -1> 연구 절차 및 방법(요약)

연구절차	본 연구에 적용한 방법
배경 문헌 분석	<ul style="list-style-type: none"> ○ 특성화고 특성 및 관련 정책 분석 ○ 특성화고 취업률 예측 요인 관련 선행연구 분석 ○ 머신러닝 관련 배경 문헌 및 연구 사례 분석
연구모형 설계	<ul style="list-style-type: none"> ○ 선행연구 분석을 통해 연구 모형 도출 ○ 종속변인 : 특성화고 취업률 ○ 독립변인 : 학교환경 요인, 교수학습 요인, 개인특성 요인으로 구성, 총 17개 세부요인 설계
분석 데이터 수집	<ul style="list-style-type: none"> ○ 신뢰성, 객관성이 확보된 데이터 수집 : 에듀데이터서비스(EDSS: EduData Service System) 활용 ○ 약 2천만 건의 분석데이터(2009년~2018년, 10년간) -학교정보공시 15,411,240건, 학교회계 공시 5,307,430건
데이터 전처리	<ul style="list-style-type: none"> ○ 분석 대상 요인 항목 추출 등 총 8단계의 데이터 전처리 과정 처리 ○ 최종 3,229행, 21개 열로 데이터 정제 ○ 취업률 하위 25% 기준 값, 중위 값, 하위 25% 기준 값 도출
기술통계 분석	<ul style="list-style-type: none"> ○ 예측모형 개발과 데이터마이닝을 위해 주요 요인별 기초통계량 분석 ○ 시도별, 지역행정구분명 데이터 빈도 분석 ○ 시도별 특성화고 취업률 평균값 분석
예측모형 개발	<ul style="list-style-type: none"> ○ 예측모형 개발을 위한 머신러닝 도구로 WEKA 3.8 활용 ○ 예측모형 개발을 위해 머신러닝 알고리즘 13개 적용 -Bayes Net, Decision Table, IBk, J48, KStar, LMT, Multilayer Perceptron, Naive Bayesian, Part, Random Forest, Simple Logistics, SMO, ZeroR ○ 적용 알고리즘 성능평가를 위해 K-Fold 교차 유효성 검증(cross-validation) -분류 정확도에 대한 accuracy 및 예측 유효성에 대한 ROC area 값 등 도출 -기준 값 변화에 따른 성능분석 및 예측모형 최적화
데이터 마이닝	<ul style="list-style-type: none"> ○ 기준 값에 따른 변인별 영향력 분석 ○ 기준 값 변화에 따른 영향력 순위 도출 및 결과의 타당성 검토를 위한 t검정
프로토타입 개발	<ul style="list-style-type: none"> ○ 유사 서비스 운영 사례 분석 ○ 서비스 프로토타입 설계원리, 구성요소(UI) 도출 ○ 전문가 자문 및 검증 ○ 프로토타입 화면 구현
결과 정리	<ul style="list-style-type: none"> ○ 분석 결과 제시 및 관련 정책에 대한 시사점 기술 ○ 후속 연구를 위한 시사점 제언

II. 배경 문헌 분석

1. 특성화고 관련 배경 문헌 분석

가. 특성화고 개념과 특성

우리나라 고등학교에서는 취업 목적의 직업교육이 처음 시작되었는데, 직업교육이 이루어지는 대표적인 두 중등 직업전문교육기관인 특성화고등학교와 마이스터고등학교(산업수요 맞춤형 고등학교)를 직업계고등학교로 통칭한다.

직업계고등학교는 「초·중등교육법 시행령」에서 규정하고 있는 고등학교의 유형 중 특정 산업 분야의 인재 양성을 주된 목적으로 하는 학교를 통칭하는 행정 용어로(허영준, 오혁제, 김명찬, 2019), 특수목적고에 해당하는 마이스터고, 특성화고 직업계열 및 전문학과를 설치한 일반고가 여기에 해당된다.

특성화고는 ‘소질과 적성 및 능력이 유사한 학생을 대상으로 특정 분야의 인재양성을 목적으로 하는 교육 또는 자연현장실습 등 체험 위주의 교육을 전문적으로 실시하는 고등학교’를 의미한다(초중등교육법 시행령 제91조; 김진원, 모영민(2019)에서 재인용). 특성화고는 1996년 2월 교육개혁위원회의 ‘신교육체제 수립을 위한 교육개혁방안 II’에서 고교단계 직업교육을 다양화하는 방안으로 처음 제안되었으며, 이를 통해 1998년 특성화고의 전신인 실업계 고등학교가 설립되었다. 2010년에는 직업교육 선진화 방안 및 이에 따른 초중등교육법 시행령 개정에 따라 기존의 전문계 고등학교 중 종합고를 제외한 직업교육 분야 특수목적고(마이스터고 제외), 특성화고, 일반 전문계고가 특성화고로 전환되었다(허영준, 김종우, 2013; 김진원, 모영민(2019)에서 재인용).

정부는 이러한 개혁을 통해 진학이 아닌 ‘취업’을 교육 목적으로 하는 고교 유형을 정착시키고자 하였다. 이를 위해 교육과정의 운영이나 학생선발, 교사

임용 방식 등에 있어서 일반고와 차별화하여 현장 중심, 실천 활동 중심의 교육목적을 달성하고자 하였으며, 특성화고 취업역량강화사업 추진 등을 통해 정책적 기틀을 지속적으로 마련하는 등의 시도를 하였다(김진원, 모영민, 2019).

특성화고를 기존 ‘전문계고’에서 명칭만 변경한 것과 달리, 마이스터고는 정부가 직업교육에서의 선도모델을 육성한다는 차원에서 특성화고 가운데 특정 요건을 갖춘 학교를 선정하여 특수목적화한 고등학교라는 점, 재정 지원과 학교 운영 자율성의 정도, 입학 전형, 졸업 이후 진로 등에 있어서도 차이가 있다. 마이스터고는 모든 재학생들이 졸업 후 취업을 희망하는 것에 반하여, 특성화고는 취업 외에도 대학 진학을 희망하는 학생들이 존재한다. 실제 마이스터고 취업률은 꾸준히 90% 이상을 유지하고 있지만, 특성화고 취업률은 2017년 기준 50.8%로(교육부, 2017), 졸업 후 대학에 진학하는 학생의 비중 또한 32.8%임을 알 수 있다(교육부, 한국교육개발원, 2017).

나. 특성화고 현황 및 관련 정책

1) 특성화고 현황

가) 특성화고 및 마이스터고 도입 배경

정부에서는 특성화고의 집중 육성을 위하여 2005년 64개교에서 2008년 170개교로 확대하였다. 이를 위해 학생, 학부모 만족도 제고, 교수·학습 및 현장 실습 개선, 우수 신입생 확보, 산학겸임교사 활용 등의 노력을 진행한 결과, 특성화고를 지정하기 전 신입생 지원율이 113.7%에서 지정한 후에 155.2%로 확대되는 등 직업교육에 대한 부정적인 이미지와 불신을 해소하는 데 일정 부분 기여하였다고 볼 수 있다.

MB 정부는 전문계고 육성을 위해 ‘고교 다양화 300 프로젝트’의 일환으로 「마이스터고 50개교 육성 정책」을 추진하였다. 이는 기존에 발생해왔던 특성화고 문제를 해소할 수 있는 최적의 대안으로 마이스터고 육성에 대한 내용을 포함하고 있었다.

「마이스터고 50개교 육성 정책」을 추진하기 전의 특성화고는 형식적 산학 연계, 군 복무로 인한 기업체에서의 채용 기피 등으로 인하여 불투명한 진로 문제가 꾸준히 제시되어 왔다. 구체적으로 학생은 산발적, 형식적인 산학 연계 등으로 현장 기술 습득에 어려움을 겪었고, 기업은 병역 의무를 마친 전문대학이나 대학 졸업자를 선호하게 되었다. 이로 인해 특성화고 학생은 단기 잡무 위주의 취업을 회피하게 되었고, 그 결과는 자연스레 취업률 저조로 이어졌다. 한편 특성화고 학생에 대한 보수, 근무 조건 등의 열악한 처우는 학생들이 취업 대신 진학을 선택하는 결과를 초래하여 2007년 특성화고 학생의 대학 진학률은 71.5%까지 높아지기도 하였다.

이와 같이 직업계고의 학교 유형으로서 마이스터고와 특성화고는 특성화 분야별 고졸 기능기술 인력의 양성 및 적재적소에 배치하기 위한 취업 지원 교육 활동을 다양한 전략을 활용하여 추진하고 있다.

나) 직업계고 기본 현황

2019년 교육부와 한국교육개발원의 교육기본 통계를 토대로 직업계 고등학교의 기본 현황을 살펴보면 다음과 같다.

2019년을 기준으로 국내 고등학교 수는 총 2,360개이며, 이 중 마이스터고와 특성화고(직업계열) 학교는 총 513개로 전체 고등학교의 21.7%를 차지하고 있다. 학생 수의 경우, 총 1,411,027명의 고등학생 중 17.4%에 해당하는 245,085명이 마이스터고 및 특성화고에 재학 중인 것으로 나타났다. 학급당 평균 학생 수의 경우 마이스터고는 18.6명, 특성화고는 20.2명으로 나타났으며, 교원 1인당 평균 학생 수는 마이스터고가 6.3명으로 특성화고 8.6명에 비해 약 2명 정도 낮은 것으로 나타났다(김성남, 김남희, 2020).

<표 II-1> 직업계 고등학교 학교 및 학생 현황

구분	총계	마이스터고	특성화고(직업계열)
학교 수	513 (100.0%)	47 (9.16%)	466 (90.83%)
학급 수	11,571 (100.0%)	943 (8.15%)	10,628 (91.85%)
학생 수	245,085 (100.0%)	17,754 (7.24%)	227,331 (92.76%)
학급 당 학생 수	19.2	18.6	20.2
교원 1인당 학생 수	7.3	6.3	8.6

※ 주: 교육통계 조사 시, 경남공군항공과학고등학교(마이스터고)는 제외. 김성남, 김남희(2020).
내용 재구성
※ (출처) 한국교육개발원 교육통계서비스 (기준일 : 2019.4.1.)

<표 II-2> 직업계 고등학교 계열별 주요 현황

구분		공업	농생명	상업정보	수산해양	가사실업	소계
마이스터고	학교 수	33	7	2	3	2	47
		70.2%	14.9%	4.3%	6.4%	4.3%	100.0%
	학생 수	14,014	1,622	469	1,059	590	17,754
		78.9%	9.1%	2.6%	6.0%	3.3%	100.0%
	학급당 학생 수	18.9	16.7	19.6	18.4	19.7	18.6
특성화고 (직업계열)	학교 수	194	36	177	7	52	466
		41.6%	7.7%	38.0%	1.5%	11.2%	100.0%
	학생 수	103,978	12,632	88,137	1,499	21,085	227,331
		45.7%	5.6%	38.8%	0.7%	9.3%	100.0%
	학급당 학생 수	20.5	18.6	20.2	15.9	20.9	20.2
총계	학교 수	227	43	179	10	54	513
		44.2%	8.4%	34.9%	1.9%	10.5%	100.0%
	학생 수	117,992	14,254	88,606	2,558	21,675	245,085
		48.1%	5.8%	36.2%	1.0%	8.8%	100.0%
	교원1인당 학생 수	8.5	7.4	8.9	6.4	8.8	8.6

※ (출처) 한국교육개발원 교육통계서비스(기준일 : 2019.4.1.), 김성남, 김남희(2020). 내용 재구성

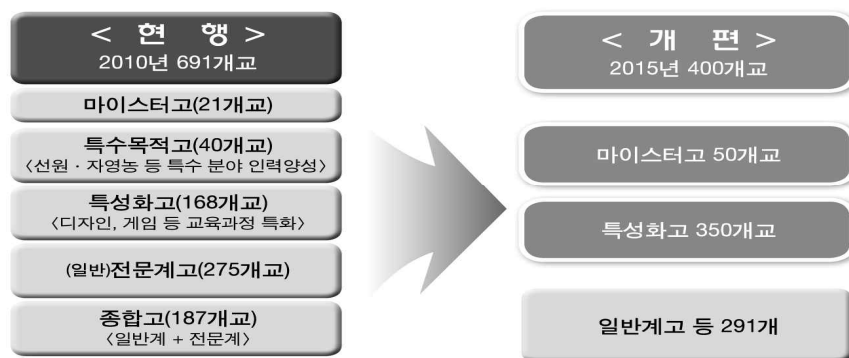
계열별 현황을 구체적으로 살펴보면, 마이스터고의 경우 47개교 중 공업계열 학교가 33개로 70.2%를 차지하고 있었으며, 특성화고의 경우 공업계열 학교가 41.6%, 상업·정보계열 학교가 38.0%로 다른 계열에 비해 높은 비중을 차지하고 있었다. 마이스터고의 경우 농생명 계열에서 학급 당 학생 수, 교원 1인당 학생 수가 가장 적었으며, 특성화고에서는 수산·해운 계열에서 동일한 지표에 대해 가장 적은 수준인 것으로 나타났다.

2) 특성화고 고졸 취업 지원 정책

가) 고등학교 직업교육 선진화 방안¹⁾

전체 전문계고를 ‘분야별 특화된 직업교육 기관으로 개편’, 산업계 정부부처 공동으로 ‘선취업 후진학 여건 조성’을 중점 추진 방향으로 제시하고, 이를 위한 중점 과제로 전문계고 체제 개편 및 지원 방안을 제시하였다.

먼저 전문계고 체제 개편은 [그림 II-1]과 같이 691개의 전체 전문계고를 마이스터고(50개교)와 산학협력형 특성화고(350개교)로 개편하고, 종합고 등 나머지 전문계고의 일반계고 전환(291개교)을 추진하는 방안을 제시하였다.



[그림 II-1] 직업교육 선진화 방안의 전문계고 체제 개편 방안

※ (자료) 국가고용전략회의(2010). 고등학교 직업교육 선진화 방안, p.2.

1) 국가고용전략회의(2010). 고등학교 직업교육 선진화 방안

○ 취업 선도 모델로서 마이스터고를 50개교로 확대 추진

공업계를 포함하여 농업 해양 건설 등 다양한 분야로 마이스터고를 확대하였다. 이를 위해 지식경제부, 농림수산식품부 등 각 부처별 육성 산업 분야와 마이스터고 간 연계, 학생들의 경력개발 지원, 국립 마이스터고(3개교)를 중소기업청으로 이관하여 산업계와 집중적으로 연계하는 방안을 추진하였다. 또한, 전국적 취업 네트워크 구축을 위해 각 분야별 산업별 협회와의 협약 확대, ‘마이스터고 취업 포털 시스템’ 마련 추진하였다.

○ 산학협력형 특성화고로 개편 확대 추진

진학 위주의 특성화고를 산학협력형 특성화고로 확대 및 내실화(350개교)하고, 선취업 후진학으로 유도하였다. 또한, 산업계와 특성화고 간 협력을 확대하고 연계를 강화하였다. 이를 위해 정부 부처 연계형 특성화고는 국토교통부, 보건복지부 등 신규 참여 부처를 확대하여 2015년까지 150개교로 확대하도록 계획하였다. 그리고 교육청 지원형은 지자체, 산업계, 교육청이 공동으로 육성하는 특성화고로 개편하여 2015년까지 200개교로 확대하는 계획을 제시하였다.

○ 종합고 등 전문계고의 일반계고 전환 추진

직업교육이 어려운 소규모 전문계고를 단계적으로 통폐합하고, 2015년까지 291개교를 일반계고로 전환하였다. 소규모 전문계고의 직업교육 수요를 거점 특성화고로 흡수하고, 영세 사학의 자발적인 해산과 이들의 통폐합 여건을 조성하였으며, 교원의 전공 전환과 관련하여 잉여 전문교과 교원 문제를 적극 해소하기 위해 유관 분야 부전공 및 복수전공 특별 연수 지원 확대하였다.

○ 전문계고 체제 개편 지원

산업계 수요를 반영한 교육과정 운영을 강화하고, 선취업·후진학 체제를 구축하였다. 또한, 재정 지원 및 평가 관리제도를 마련하였다.

산업계 수요를 반영한 교육과정 운영 강화를 위해서 국·영·수 학업 성취도 평가를 대체하는 직업기초능력 평가를 도입하여 직업교육에 맞는 평가 체제를

구축하고, 산업수요 맞춤형 전문 기술 교육을 강화하기 위하여 산업계와 연계한 실무 중심 교육과정 개편을 단행하였으며, 산학겸임교사의 채용을 대대적으로 확대하는 방안을 제시하였다.

선취업·후진학 체제 구축을 위해 재직 경험만으로 대학에 가는 재직자 특별전형 도입을 확대하고, 저소득층 우수 학생 장학 사업 중의 하나로 특성화고 취업자 우선 지원을 고려하였으며, 고용노동부가 진행 중인 근로자 학자금 대부 사업의 적용 대상 또한 특성화고 졸업자까지로 확대하였다. 취업 캠프, 현장 체험 등 특성화고 취업 기능 강화 사업을 2015년까지 연장하여 학교 내 취업 지원 프로그램을 강화하고, 특성화고 내에 산학협력 코디네이터, 취업 지원관 등을 상근하게 하여 학교의 산학 연계 기능 강화 계획을 제시하였다. 또한 특성화고 졸업 예정자를 포함하여 청년 취업 인턴제를 확대 적용하고, 기능 인재 추천제를 시·도 교육청, 지자체로 확대하고, 특성화고 취업 병역 연기 제도도 전 업종으로 확대하였다.

마지막으로 교사 부전공, 복수전공 지원, 취업우수학교 인센티브 등 전문계고 체제 개편에 따른 재정지원과 마이스터고 재인증 시스템, 특성화고 지정해체 제도 마련 등 학교 평가 관리 제도를 마련하도록 하였다.

나) 교육 희망사다리 구축 프로젝트²⁾

‘교육 희망사다리 구축 프로젝트’의 일환으로 ① 전문계 고등학생의 교육비 전액 지원, ② 전문계고 취업 지원의 대폭 확대의 두 가지 방안을 추진하였다(국민경제대책회의, 2010).

① 전문계 고등학생의 교육비 전액 지원

- 지원 대상 : 특성화고 학생 43만 명의 약 60%에 해당하는 26만 3천 명이며, 수업료와 입학금을 포함하여 1인당 연평균 120만 원을 지원함.
- 지원 방식 : 지방 교육청이 지원 대상자 학교로 직접 계좌 이체하여 학비를 충당하게 하고, 개별 학교별로 중복 지원 여부를 체크하도록 함. 이를

2) 교육과학기술부(2011a). 공생발전을 위한 교육 희망사다리 구축 방안.

위해 2011년에 3,159억 원을 지원하며, 교육부와 교육청에서 반씩 분담하여 재원을 마련함.

② 전문계고 취업 지원의 대폭 확대

산업체 현장 연수, 해외 인턴십 지원 사업 등을 병행하여 좋은 일자리 취업 기회 확대를 통한 특성화고 취업 지원을 대폭 강화함(국민경제대책회의, 2010). 특히 현장 연수, 전문대 연계 프로그램 등을 통해 취업 중심 학교로 내실화·정예화를 추진함(<표 III-1> 참조). 이를 위해 2011년에 특성화고 취업 지원을 위해 510억 원을 투입하여 고품격 직업교육 기관으로 개편하는 계획을 발표하였음. 특성화고 학생 중 약 1만 명의 학생들에게 산업체 체험 연수 및 현장 실습 기회를 제공하고, 1천 명의 우수한 인재를 선발하여 해외 인턴십 기회를 부여함으로써 학생들이 산업 현장에 필요한 인재로 성장할 수 있도록 지원하고자 하였음. 또한, 특성화고-전문대학 연계 프로그램(4년제 통합 과정)을 통해 학생들이 중견 전문 인력으로 육성될 수 있는 경로를 마련하고, 특성화고-전문대학으로 구성된 10개 사업단을 선발하여 맞춤형 교육과정 개발, 학자금 지원, 교원 산업체 연수 등을 패키지로 지원하는 방안을 제시하였음.

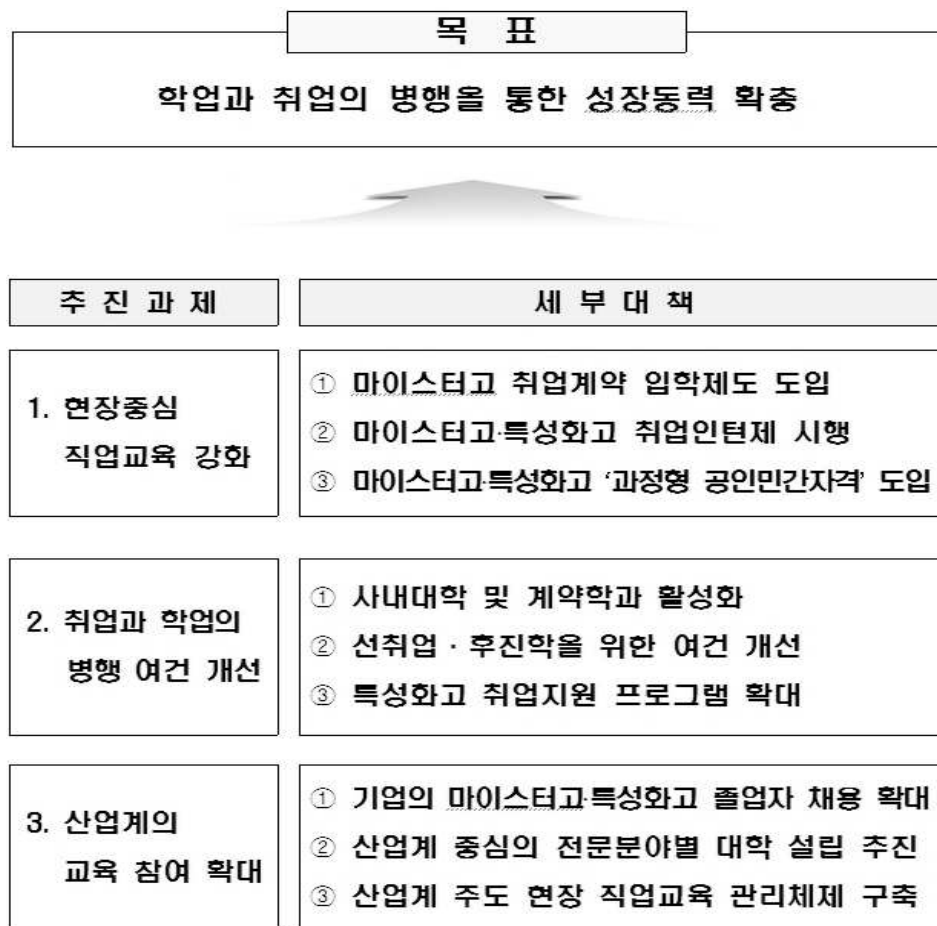
<표 II-3> 교육 희망사다리 구축 프로젝트 전문계고 학생 취업 지원 강화 세부 사업

세부 사업	사업 내용	수혜 규모	2011년 예산안
산업체 현장 연수	• 산업체 체험 연수 및 현장 실습 지원	연간 10,000명	70억원
선진화 지원	• 실습 기자재, 실습동 확충 • 교원의 산업체 연수 • 교육과정 개발 지원	100교 (전체 학교 15%)	200억원
전문계고-전문대 연계 프로그램	• 고교-전문대학 통합과정 운영사업단 지원	10개 사업단	100억원
해외 인턴십 지원	• 해외 기업에서 실습 및 인턴 근무 지원(6개월)	연간 1,000명	140억원

※ (자료) 국민경제대책회의(2010). 2011년 예산(안) 서민 희망 3대 핵심과제, p.7. 양정승(2014) 연구에서 재인용

다) 학업·취업 병행 교육체제 구축 방안³⁾

‘기술력 증진, 생산력 확대, 일자리 복지를 위한 학업 취업 병행 교육 체제 구축방안’은 마이스터고 및 특성화고의 선취업 후진학 체제를 구축하고, 이를 위한 산업체 인센티브 제공 방안들을 주요 내용으로 하고 있다. 이 정책을 통해 정부는 ① 현장 중심 직업교육 강화, ② 취업과 학업의 병행 여건 개선, ③ 산업계의 교육 참여 확대 등의 과제를 추진하였다([그림 II-2] 참조).



[그림 II-2] 학업·취업 병행 교육 체제 구축 방안의 추진 방향

※ (자료) 국가경쟁력강화위원회(2011). 기술력 증진, 생산력 확대, 일자리 복지를 위한 학업 취업 병행 교육

3) 국가경쟁력강화위원회(2011). 기술력 증진, 생산력 확대, 일자리 복지를 위한 학업 취업 병행 교육체제 구축 방안

라) 공생발전을 위한 열린 고용사회 구현 방안

‘공생발전을 위한 열린 고용사회 구현방안(관계부처합동, 2011)’은 학력의 벽을 넘어 더 나은 내일을 꿈꾸는 열린 고용사회를 비전으로 제시하고 있었다.

이 정책을 통해 정부는 ① 직업인으로서 다양한 꿈 키우기 지원, ② 능력에 기초한 열린 채용, ③ 발전의 기회가 개방된 인사관리, ④ 능력 중심의 사회적 여건 조성 등의 과제를 추진하였다([그림 II-3] 참조).

학교	대학만 바라보는 일방적 진로
	1) 직업인으로서 다양한 꿈 키우기 지원 - 체험형 진로지도 체계 구축 - 현장교육을 통한 취업 역량 강화
노동 시장	학력만으로 미래가 결정되는 닫힌 노동시장
	2) 능력에 기초한 열린 채용 - 고졸 취업기회 확대 - 직무능력 중심 채용관행 확산 3) 발전의 기회가 개방된 인사관리 - 임금 및 승진 제도를 성과, 능력 중심으로 개선 - 지속 발전을 위한 일학습 병행 여건 조성
사회	학력 중심
	4) 능력 중심 사회적 여건 조성 - 전문 기능기술인에 대한 제대로 된 대우 - 경력이 학위로 이어지는 체제 구축 추진

[그림 II-3] 공생발전을 위한 열린 고용사회 구현방안의 추진 과제

※ (자료) 관계부처 합동(2011). 공생발전을 위한 열린 고용사회 구현방안.

마) 특성화고·마이스터고 관련 지원 사업·제도

「고등학교 직업교육 선진화 방안」, 「학업·취업 병행 교육체제 구축 방안」의 일환으로 진행된 직업계고 관련 정책 중 특성화고에 관련된 정책을 중심으로 제시하면 다음의 <표 II-4>와 같다.

<표 II-4> 특성화고·마이스터고 관련 지원 사업·제도

구분	목적	내용	비고
특성화고·마이스터고 수업선진화 지원 사업(2011)	창의적인 수업모델 개발 및 보급을 통한 특성화고·마이스터고 수업혁신 및 내실화	<ul style="list-style-type: none"> 수업혁신팀 200팀 육성 우수수업모델 200개 발굴 수업선진화 연수기관 운영 	13억 원
특성화고 글로벌 역량강화사업 (2011)	실무역량과 글로벌 감각을 갖춘 글로벌 기술·기능인재를 양성하여 특성화고·마이스터고 교육의 경쟁력 제고	<ul style="list-style-type: none"> 해외인턴십 등 글로벌 현장학습 운영 중등단계직업교육 글로벌 협력 시범 사업 (ODA) 특성화고 영어 교수 학습자료 개발 및 확산 직업영어 담당 교사 연수 특성화고 영어 수업 모델 선진화 	809억원
산업체 우수강사 채용 지원 사업 (2011)	특성화고 및 마이스터고 학생들의 취업 경쟁력 강화 위한 현장 수요 맞춤형 기술과 역량 전수	<ul style="list-style-type: none"> 산업체 우수강사 330명 지원 취업역량 우수 학교 대상, 12년부터 ‘현장실습지원교사제’ 채택 학교 우선 지원 	100억원
직업기초능력 평가 도입 (2011)	산업체(기업)에 학생의 직업준비도에 대한 신호(signal) 제공	<ul style="list-style-type: none"> 평가 영역: ① 언어 활용, ② 수리 활용, ③ 문제 해결, ④ 조직 적응력 평가 방식: 문제은행 Pool 구축 후 WBT (Web-based Test) 12년: 시범전수평가, 13년: 정규평가 	-
과정형 공익민간자격 제도 도입 (2011)	전문기관에 의해 평가·인증된 산업계 맞춤형으로 개발된 교육과정 이수율 자격과 동일하게 인정	<ul style="list-style-type: none"> 국가직무능력표준(NCS)개발, 교육과정, 민간자격, 인증, 질 관리 등 절차 진행 예정 자격기본법 개정 추진 중 	-
1社1敎 산학협약 지원 (2011)	학교와 기업 산학협력 지원	<ul style="list-style-type: none"> 학교-기업 매칭 1사1교 협약 체결 학교와 기업 산학협력 활동 	-
취업계약 입학제도 및 취업인턴제 세액 공제 (2011)	마이스터고 및 특성화고 재학생 대상 취업계약 입학제도 및 취업인턴제에 소요되는 경비의 일부를 세액공제	<ul style="list-style-type: none"> 취업계약 입학제도 및 취업 인턴제 시행으로 발생하는 경비를 인력개발비로 간주하여 일반 R&D 세액공제대상 (중소기업 25%, 대기업 3~6%)으로 인정 	-
고용창출투자 세액공제제도 개선 (2011)	기업 고용창출형 투자 유도	<ul style="list-style-type: none"> 마이스터고·특성화고 졸업자 채용에 대해서 별도의 강화된 고용창출투자세액 공제한도 우대(1인당 1천 5백만 원 → 2천만 원, 12년) 	기획재정부

구분	목적	내용	비고
공공기관 마이스터고· 특성화고 채용목표제 (2011)	공공기관의 마이스터고·특성화고 졸업생 취업확대 선도역할 실시	<ul style="list-style-type: none"> 2012년부터 공공기관 경영평가 시 고졸 채용 권장 비율(20%)반영 공공기관 고졸자 채용 가이드라인(12.6), 매뉴얼(13.3) 배포 	기획 재정부
산업기능요원 제도 (2011)	특성화고 졸업자의 안정적인 경력개발 및 기업체의 고용유도	<ul style="list-style-type: none"> 2015년까지 연장, 특성화고·마이스터고 졸업생에게만 산업기능요원제도를 적용 	국방부 병무청
특성화고· 마이스터고 취업자 입영연기 (2011)	특성화고 졸업자의 안정적인 경력개발 및 기업체의 고용유도	<ul style="list-style-type: none"> 특성화고·마이스터고 졸업 취업자에 대해 대학생과 동일하게 최대 4년까지 입영을 연기 	국방부 병무청
기능인재 추천제 (2011)	학력차별완화방안에 의거 고졸 출신 우수기능인재 대상으로 특별 채용	<ul style="list-style-type: none"> 특성화고·마이스터고 출신의 우수기능 인재를 대상으로 학교장 추천과 1년 정도의 공직 내 견습기간을 거쳐 특별 채용(안전행정부, 지자체, 시·도교육청) 	-
지역인재 9급 추천채용제 실시·확대 (2016,2017)	특성화고, 마이스터고 졸업자를 일반직 9급으로 선발 (견습근무 6개월 후 임용)	<ul style="list-style-type: none"> 12년 104명→13년 120명→15년 150명 	-
중소기업 취업청년 인턴제 개편 (2011)	청년의 중소기업 장기근속 유인이 강화되도록 인턴 지원금 체계 개편	<ul style="list-style-type: none"> 인턴지원금(기업에 지급) 조정 : 인턴기간 중 매월 약정한 임금의50% 지원(월 80만원 한도) → 3개월 간 임금 50%(60만원 한도) 취업지원금(인턴에 지급) 인상 확대 	고용 노동부
재직자 특별전형 및 계약학과 확대 (2011)	연령별 여건에 맞는 후진학 제도 도입·확대	<ul style="list-style-type: none"> 우수 고졸자 대상 채용조건형 계약학과를 확대 : 13년 33개→17년 70개 최소 선발 기준 : 13년 4%→15년 5.5%로 확대 입학 자격 : 기존 특성화고·마이스터고 졸업자→ 일반고 졸업자로 확대(고등교육법 시행령 개정('4.4.30시행)) 	-
기술전문가 양성사업 (IP meister program)	특성화고·마이스터고 학생 직무발명역량 함양 및 예비기술전문가 양성	<ul style="list-style-type: none"> 특성화고·마이스터고 학생의 문제해결력 및 지식재산 창출역량 함양 창의적 지식근로자로의 성장 지원 	한국 발명 진흥회
국가직무능력 표준(NCS)에 기반한 직업교육 강화	모든 특성화고· 마이스터고에 NCS 기반 교육과정을 적용, 산업현장 요구 직무 중심으로 교육	<ul style="list-style-type: none"> '16년 신입생부터 NCS 학습모듈을 실무과목 교과서로 사용하고, 2, 3학년은 학교여건(실습기자재, 교원 등)에 따라 최대한 실무과목 적용 	교당 ~ 1.1억 ~ 1.7억 원

구분	목적	내용	비고
명장공방 지원 사업	특성화고·마이스터고 학생에게 현장 전문기술을 전수하여 직무능력 제고	<ul style="list-style-type: none"> 우수기술 전수 필요성이 높은 분야(뿌리산업, 전통산업 분야 등)를 중심으로 기술 명장들의 직업교육과정 참여 활성화 지원 	-
특성화고·마이스터고 취업역량 강화 사업	학생 취업역량 강화를 위한 취업선도 특성화고 지원 및 학생 취업역량 제고를 위한 교원 취업 지도 역량 증진	<ul style="list-style-type: none"> 취업마인드 함양 및 진로설계 프로그램 참가, 현장견학체험 및 실습 참여, 직무기초, 인성, 산업안전 교육 취업지도역량 연수 참가, 범교과 취업지도 협력, 직업교육선진화연구대회 등 참가, 산학협력, 추수지도 업무지원 	교육부
산학일체형 도제학교 지원 사업 (2016,2017)	학교(이론기초실습)와 기업(체계적 현장교육훈련)을 오가며 기술·기능인재로 성장	<ul style="list-style-type: none"> 15년 9개교 시범운영, '16년 60개교 시범운영 확대, '17년 198개교 운영으로 산학일체형 도제학교 지속 확대 	교육부
직업계고 학생 비중 확대 사업	22년까지 직업계고 학생 비중 30% 수준 확대 및 취업률 65% 달성	<ul style="list-style-type: none"> 지역, 학교 여건별 직업계고 학생 비중 확대 매력적인 특성화고 전환을 위한 내실화 확대 및 내실화를 위한 행·재정적 지원 규모 확대 및 내실화 추진 관리 	교육부
직업교육 마스터 플랜 수립	민관 협력으로 미래변화에 선제적 대응하는 직업교육 마스터 플랜 수립	<ul style="list-style-type: none"> 체계적 국가인적자원 개발·활용을 위한 범부처 마스터플랜 수립 	교육부
신 산업분야 마이스터고 지정 확대	AI, 빅데이터 등 4차 산업혁명 시대 신 산업분야 마이스터고 지정 확대	<ul style="list-style-type: none"> 4차 산업혁명 시대 유망 분야로 인력 양성 유도 	교육부
직업계고 학점제 도입	현장 수요에 맞는 융합형 교육과정 운영 직업계고 학점제 도입	<ul style="list-style-type: none"> 직업계고 학점제 연구학교 운영('8.23교) 	교육부
학습중심 현장실습 적용	직업계고 학생 학습권과 안전보장을 위한 학습중심 현장실습 적용	<ul style="list-style-type: none"> 조기취업 현장실습 폐지 → 취업 준비과정으로 개편 안전이 담보된 업체를 선정하여 학생 신분으로 현장실습 실시 실무과목과 연계한 학습중심의 교육과정 연계 우수 현장실습 기업 후보군 제공 및 기업참여 인센티브 방안 마련 	교육부

※ (자료) 교육과학기술부(2011b). 특성화고 마이스터고 권역별 설명회 참고자료, 재구성.
관계부처 합동(2014). 학교에서 직장까지 : 일자리 단계별 청년고용 대책, 재구성.
교육부(2014a). 고등교육법 시행령 일부 개정, 재구성.
교육부(2015a). 2015년 마이스터고 사업 추진계획(안), 재구성.
교육부(2016a). 2016년 교육부 업무계획, 재구성.
교육부(2017a). 2017년 교육부 업무계획, 재구성.
교육부(2018a). 2018년 교육부 정부 업무 보고, 재구성.

바) 전문 인재 양성을 위한 직업교육 정책

2013년 2월 정부 인수위원회는 전문인재 양성 및 평생학습 체제 구축 관련 과제로 ‘전문인재 양성을 위한 교육 강화’를 제시하고, 5월에 140개 국정과제 추진 전략과 추진계획을 보완 및 확정하였으며, 이 국정 과제는 고교 졸업 후 대학 진학의 단선적 진로에서 벗어나 꿈과 끼를 펼칠 수 있는 다양한 진로중심 직업교육 기회 제공을 목표로 하였다. 이를 위해 취업중심 고교 직업교육 체제를 강화하고 고졸 재직자의 학위취득 및 경력개발 인프라를 구축하였다.

이와 같은 고교 직업교육 정책은 MB 정부에서 추진한 ‘고졸 취업 및 후진학 활성화 정책’의 지속성을 보장하면서 전문 인재 육성을 위한 인프라 구축 및 교육 내실화를 전제로 하였다.

2016년 10월 교육부에서 제시한 ‘마이스터고 발전 방안’의 추진 방향은 전문인재 양성을 위한 직업교육 경쟁력 강화, 고졸취업 활성화를 통한 능력중심사회 구현이었다. 각 정책별 세부 추진 내용은 <표 II-5>와 같다.

<표 II-5> 각종 정책별 세부 추진 내용

정책명	세부 추진 내용
전문인재 양성을 위한 교육 강화 (2013. 05.)	· ① 범부처 지원 강화, ② 마이스터고 다양화, ③ 현장중심 교육 운영 강화, ④ 직업교육 기회 확대, ⑤ 고졸 취업자의 학위취득 인프라 구축
고졸 취업 및 후진학 활성화 정책 (2014. 02.)	· ① 마이스터고 활성화, ② 고졸 취업의 지속 확대를 위한 기반 강화, ③ 재직자의 후진학 및 경력개발 기회 확대
마이스터고 발전 방안 (2016. 10.)	· ① 산업수요 맞춤형 창의인재 양성을 위한 교육과정 운영체제 구축, ② 졸업생 지속적인 능력개발 및 경력경로 개발, ③ 산학관 협력체계 강화로 안정적 학교운영모델 정립, ④ 운영의 질 관리를 위한 성과관리 체계 강화

사) 현 정부의 국가 직업교육 책임 강화를 위한 직업교육 정책

2017년 7월 새 정부가 발표한 100개 국정과제에는 정부 5대 국정목표인 ‘국민이 주인인 정부’, ‘더불어 잘사는 경제’, ‘내 삶을 책임지는 국가’, ‘고르게 받

전하는 지역’, ‘평화와 번영의 한반도’라는 목표가 포함되어 있다. 이 중 ‘내 삶을 책임지는 국가’라는 목표에서 국가가 책임지는 보육과 교육 전략 내 유아에서 대학까지 교육의 공공성 강화, 교실혁명을 통한 공교육 혁신, 교육의 희망사다리 복원, 고등교육의 질 제고 및 평생 직업교육 혁신, 미래 교육 환경 조성 및 안전한 학교 구현과 관련한 과제로 교육과 관련한 과제가 담겨 있다. 거점국립대 지역강소대학 집중 육성 등 대학의 공공성 경쟁력 강화, 국가 직업교육 책임 강화 및 성인평생학습 활성화를 과제 목표로 제시하였다.

최근 정부는 「고졸취업 활성화 방안(관계부처 합동, 2019a)」에서 학교와 지역사회와의 협력, 지역사회 일자리 취업 지원을 강조하고 있다. 지자체와 교육청 간 취업지원 협의체를 구성하고 지역산업과 연계하여 산학협력 및 맞춤형 일자리 취업 지원, 지자체와 시도교육청 취업연계 우수사례 확산 등 지역사회 일자리 취업지원을 강화하는 방향으로 정책을 추진 중이다.

2010년 이후부터 추진되는 고졸 취업 활성화 정책은 고교 졸업 이후 대학진학이라는 단선적인 진로경로가 고착되어 중소기업의 구인난과 대졸 이상 고학력자의 구직난 해소를 목적으로 하고 있다. 인구감소에 따라 점차 고졸자들은 취업보다 진학을 선호할 가능성이 높다. 지자체는 지역 내 기업의 인력난 특히 고졸인력 확보의 어려움을 해소하기 위해 과거와 달리 고졸인력 확보, 고졸자의 지역 내 기업 취업에 대해 적극적으로 참여하는 모습을 보이고 있다. 이는 지방분권, 지방자치제에서 지역 내 경제발전을 위해 고졸취업에 대한 지자체의 역할이 강조되고 있다. 지금까지 고졸취업 정책이 중앙정부 주도로 추진되어 왔다면 이제는 지방정부가 적극 참여하여 중앙정부와 협력적인 관계를 통해 효과적·효율적으로 고졸취업 정책을 추진할 수 있을 것이다.

다. 특성화고 관련 취업률 예측 요인

직업계 및 특성화 고등학교 학생의 취업 관련 연구들은 학생 개인, 학생의 가정환경과 같은 학생차원, 그리고 학교 특성, 교내 진로 및 취업 관련 프로그램 등과 같은 학교 차원으로 범주화하여 관련 요인들의 취업 가능성에 대한

영향력을 논하고 있다. 본 연구에서는 취업률에 영향을 미치는 학생 관련 변인들을 개인특성 요인으로, 학교 차원 중에서도 교육, 지도, 학생참여 활동을 교수학습 요인으로, 그리고 학교의 전반적 특징을 학교환경 요인으로 분류하여 배경 문헌들을 분석하였다.

1) 개인특성 요인

가) 성별

성별이 취업률 예측요인으로 통계적으로 유의하게 확인된 연구들은 대부분 여학생이 남학생에 비해 취업에 성공할 가능성을 높게 예측하고 있다. 2013 고졸자취업진로조사 데이터를 분석한 연구에서는 여학생의 취업 가능성이 더 높게 나타났다. 남학생의 경우 군 입대로 인한 업무 단절과 그로인한 인력충원에 따른 부담이 문제가 될 수 있는 것이 취업가능성을 낮추는 원인으로 논의되었다(천영민, 오민홍, 2015). 한국교육고용패널 데이터(KEEP) 분석에서 여학생의 취업 확률이 남학생보다 높은 것으로 분석되었다. KEEP 4차 데이터를 활용하여 분석한 연구에서도 역시 여학생이 남학생에 비해 취업선택 가능성이 높게 나타났다(노경란, 허선주, 2012). 고용의 질 측면에서도 여학생이 남학생에 비해 좋은 일자리로 취업하는 사례가 많은 것으로 나타났다(최동선, 이종범, 2013). 특성화고를 졸업한 남학생의 취업률이 상대적으로 낮은 원인으로 남학생의 경우 군 입대로 인한 업무 단절과 그로인한 인력충원에 따른 부담이 문제가 될 수 있다는 논의와 함께 입대전후 고용유지 지원정책 등 국가차원의 지원방안을 과거 연구들에서는 제안하고 있다(천영민, 오민홍, 2015). 최근, 김지효(2018)의 연구에서는 남학생의 취업성고가 여학생에 비해 높은 반대되는 결과가 나타났으며, 그에 따른 여학생의 취업 어려움과 관련하여 차별화된 커리어개발 프로그램과 진로상담을 제안하였다.

나) 학업성적

학교성적이 높은 특성화고 졸업생이 취업성공 가능성이 높은 것으로 나타났다(노경란, 허선주, 2012). 학업성적은 보통 개인의 성실성 및 잠재적 인지능력

을 인증하는 지표로 작용될 수 있기 때문에 특성화고 학생의 학업성적은 학생의 기본적인 생활태도 및 노력을 기울이는 정도를 확인해줄 수 있다. 같은 맥락에서, 학교성적이 높은 학생과 학교생활에 잘 적응하는 학생의 취업성공 가능성이 높게 나타났다(천영민, 오민홍, 2015). 특히, 전공 성적이 높은 특성화고 학생들은 취업으로 진로를 결정하는 것으로 나타난 반면, 국어, 영어, 수학 과목이 높은 학생은 진학을 선택하는 경향을 보였다(오석영, 2012). 홍성표와 정진철(2016)의 특성화고 학생의 취업 및 진학 결정수준과 준비행동 영향요인 분석 연구에서는 일반교과 성적이 높을수록 약한취업선호와 취업준비 미행동 그룹에 속할 확률이 높아지며, 전공교과 학업 성취가 높은 학생은 약한취업선호-취업준비 미행동 그룹에 속할 가능성이 낮게 나타났다. 이러한 결과는 특성화고 학생들 가운데 일반교과 보다 상대적으로 전공교과 흥미가 높은 학생이 취업으로의 긍정적 의사결정과 준비행동을 실시하는 것으로 해석할 수 있다.

다) 학교교육 만족도

특성화고의 경우, 졸업학교에 대한 학생의 전반적 만족도는 취업 성공 가능성에 통계적으로 유의한 예측요인으로, 학교의 만족도가 높을수록 취업할 가능성도 높게 나타났다(천영민, 오민홍, 2015). 이와 상반되게, 전공만족을 개인특성이 아닌 학교특성 변인으로 고려하여 분석한 이쌍철과 주현준(2013)의 연구 결과에서는 전공만족이 높은 학교는 취업률이 낮게 확인되었다. 이러한 결과는 특성화고 학생들의 전공교과에 대한 자신감과 흥미가 취업보다는 대학진학으로의 동기를 높이는 기제로 작용한 것으로 연구 논의에서 설명하고 있다.

라) 학업중단율

학업중단 비율은 학교취업 비율에 통계적으로 영향력이 확인되었다. 이쌍철과 임문영(2014) 연구에서 2009-2013년까지의 에듀데이터 분석 결과, 학생 학업중단 비율이 높아질 경우 특성화고 취업비율이 낮아지는 것으로 분석되었다. 학업중단율은 학생 개인 특성이나 학교 교육과정의 질적 측면이 반영된 매개변인이 될 수 있다. 학업 중단 관련 선행 연구 역시 학업중단은 학교의 노력여

하와 학생 개인의 상황에 의해 발생됨을 확인된 바 있다(Lee, Burkam, 2003).

마) 가정배경

가정의 사회경제적 지위를 나타내는 가구소득 변인이 특성화고 학생의 취업 선택 결정에 영향을 미치는 것으로 나타났다. 즉, 부모의 사회경제적 지위가 상대적으로 낮은 학생은 졸업 후 취업을 선택하는 경우가 많았으며, 그에 반해 부모의 사회경제적 지위가 높을수록 학생의 진학을 위한 투자 성향이 강하게 확인되었다(장현진, 이지혜, 2012). 천영민, 오민홍(2015) 연구에서 특성화고 졸업자의 취업에 미친 영향을 분석한 결과, 보호자의 소득이 높을수록 취업 성공 가능성도 높게 확인되었다. 2013년, 2016년 두 차례 실시된 진로진학 조사 자료 분석 결과에서는 부모의 소득과 최종학력은 특성화고 졸업생의 취업성공에 영향을 미치지 않는 것으로 나타났다. 이처럼 가구의 소득은 각 연구의 데이터에 따라 취업으로의 효과가 일관되지 않은 것으로 나타나 취업률 예측요인으로 불안정한 변인이 될 수 있다. 한편, 부모의 사교육지원정도, 진학정보인식은 취업성공에 부정적 영향, 즉 특성화고 졸업생이 취업보다는 진학으로 진로를 결정하는 영향요인으로 나타났다(김지효, 2018).

바) 자기효능감

특성화고 학생들의 구직효능감이 취업결정에 유의한 영향을 미치는 것으로 나타났다. 이러한 연구 결과는 학생들이 특성화고 졸업 후 양질의 일자리에 성공적으로 취업하여 합리적 대우를 받을 수 있다는 긍정적 인식을 심어주는 교육과 지도가 필요함을 시사한다(조규형, 정철영, 2014). 또한, 취업선호와 취업준비행동 특성 비교 집단을 예측하는 개인특성 요인을 분석한 결과, 진로결정 자기효능감이 높을수록 약한 취업선호 및 진로 미결정, 취업 미행동이 나타나는 그룹에 속할 확률이 낮아지는 것으로 나타났다. 즉, 강한 취업선호와 강한 취업행동을 보이는 집단 학생은 자신의 진로선택에 높은 자신감을 갖는 것을 의미한다(홍성표, 정진철, 2016).

2) 교수학습 요인

가) 동아리 활동

동아리 활동에 참여하는 학생비율이 높을수록 특성화고 취업률 또한 높아지는 것으로 나타났다(이쌍철, 엄문영, 2014). 특히, 동아리 활동에 참여한 학생은 그렇지 않은 학생보다 졸업 직후 취업성공률이 증가하는 것으로 나타났으며, 3년 후 재조사한 결과에서는 동아리 활동 요인의 영향이 상쇄되는 것으로 확인되었다(김지호, 2018). 동아리활동은 학술, 문화예술, 취미활동 등을 넘어 최근 전공 지식 및 기술을 접목시킨 창의적, 융합적 체험활동을 다양하게 구성하여 운영되고 있다. 이에 특성화고 학생들의 취미와 특기 개발은 물론, 전공능력을 향상시키고 실제 진로 영역에서의 실무능력을 발휘할 수 있는 기회를 동아리 활동을 통해 가질 수 있다.

나) 방과후 학교

방과후 학교에 참여한 학생들이 참여하지 않은 학생들에 비해 취업 가능성이 높게 나타났다. 수업 외 산학연계 프로그램, 현장실습, 창업교육 등이 방과후 학교에서 진행되며, 이러한 활동들은 학생들의 직업능력 향상에 영향을 미치는 것으로 확인되었다(천영민, 오민홍, 2015). 특기적성 방과후 학교에 참여율 역시 취업률을 높이는 요인으로 작용하는 것으로 나타났다(이쌍철, 엄문영, 2014). 방과후 학교를 자격증 취득을 위한 강좌를 제공하는 시간으로 활용하고 있는 특성화고가 다수 있는 것을 고려한다면, 방과후 학교에 참여한 학생들은 취업 역량이 강화될 수 있으며 실제 취업 시에도 수료한 교육과정 및 자격증이 유리하게 작용할 수 있다.

다) 기타 선택적 교육활동

학교의 진로교육에 투입된 노력은 곧 높은 취업률로 나타났다. 진로교과 교육, 진로강연, 진로검사, 직업체험 등 특성화고에서 제공하는 선택적 프로그램은 학생들의 직업세계관에 대한 이해의 폭을 넓히는데 긍정적으로 작용하였으며, 이는 곧 학생의 취업선택에 긍정적 효과는 주는 것으로 나타났다(이쌍철,

주현준, 2013). 조규형과 정철영(2014)의 연구에서는 현장실습 프로그램의 학생 참여 비율이 특성화 고등학교 졸업예정자의 취업결정에 긍정적인 영향을 미친 결과를 토대로 실제 취업과 연계한 현장실습지 탐색 등을 효과적인 취업전략으로 제안하였다. 같은 맥락에서 최근 연구에서도 현장직업체험 및 체험기반 진로프로그램에 참여한 학생이 취업을 결정할 가능성이 큰 것으로 나타났다. 그리고 특성화고교 학생들은 산업체견학 및 체험, 교육기관 방문, 취업박람회 등과 같은 진로프로그램 운영 빈도가 높을수록 취업성공 비율이 상당히 증가하는 것으로 확인되었다(김지효, 2018). 천영민과 오민홍(2015)의 연구 결과, 이력서 작성 및 면접 등 구직활동 경험이 있는 학생이 경험이 없는 학생에 비해 상대적으로 취업성공 가능성이 높게 나타났다. 여러 취업 관련 프로그램 중 실제적으로 취업과 연계한 지원일수록 학생들의 취업지원 만족도와 요구도 역시 증가하였다. 그 밖에, 경진대회 및 공모전 수상여부, 자격증취득 여부가 취업에 미치는 영향이 통계적으로 유의하게 확인되었다.

라) 교육활동 예산

특성화고에 투입되는 교육활동 예산은 학생들이 실제 참여할 수 있는 학교 교육활동의 양과 질을 나타내는 지표로 고려될 수 있다. 이영민 외(2016)는 특성화고와 같이 직업교육과정을 운영하는 고교의 NCS(국가직무능력표준, National Competency Standards)기반 교육과정의 표준교육비⁴⁾를 산출한 결과, 실무과목과 기초과목의 통합 표준교육비의 학급당 및 학생당 교육비는 기계 교과군이 가장 많았고, 다음으로 농림·수산·해양 교과군, 동물자원과 금융·경영 교과군 순으로 확인되었다. 산업군에 따른 기준학과별 표준교육비 산출이 상이한 점을 감안한다면 학생당 지원되는 교육활동 예산이 갖는 교육효과는 학교 계열별로 다르게 나타날 수 있다. 또한 정부의 직업계고 비중 확대와 일·학습 병행과 같은 정책 지원의 일환으로 직업계고만을 위한 재정지원 사업을 시행하고 있으며, 이러한 추가 재정은 학생당 교육활동 지원 예산 및 학생당 선택

4) 표준교육비는 일정 규모의 단위학교가 그에 상응하는 인적, 물적, 조건을 확보한 상태에서 교육 목적을 달성하기 위한 정상적인 교육 활동을 수행하는 데 필요한 최저 소요 교육비를 말하며, 산출모형에 따라 교과활동경비, 창의적 체험활동경비, 공동운영경비 등을 포함한다(공은배 외, 2011).

적 교육활동 예산에 있어 고교 간 격차를 만들어낸다. 재정지원을 수혜한 고교는 학생당 선택적 교육활동 예산이 타 고교에 비해 증가하며, 이를 활용하여 현장실습, 체험활동, 취업지원 프로그램 등을 운영하기 때문에 결과적으로 학교의 추가 재정은 학생의 취업성과를 예측하는 변인으로 고려될 수 있다. 실제로 국가 재정지원 사업을 통한 공업계 특성화고등학교 발전 사례 연구(황재호, 김병주, 2014)에서는 투입된 교육활동 예산은 자격증 취득, 취업률 향상과 같은 성과를 높인다고 보고하고 있다.

3) 학교환경 요인

가) 학교계열

직업계 고등학교의 경우, 상업계 고등학교 학생이 공업계에 비해 1.5배 정도 취업 가능성이 높게 나타났다. 이와 같은 통계적 결과는 성별 변수로 통제되었을 때 유의성이 없어졌는데, 상업계열의 경우 대부분 여학생으로 구성되었다는 특성이 통제되었기 때문이다(장현진, 이지혜, 2012). 2011 고졸자 취업진로조사(HSGOMS) 데이터를 활용하여 미진학 특성화고 취업자의 일자리 특성을 분석한 결과에서도 상업계열을 졸업한 학생이 타 계열에 비해 상대적으로 좋은 대기업이나 사무 종사자로 취업하는 비율이 높았다(최동선, 이종범, 2013). 이지혜와 정철영(2010)의 2009년 특성화 고등학교 3학년 학생이 인식하는 취업장벽 수준을 조사한 연구 결과, 계열에 따른 취업장벽 수준 차이는 공업 > 상업관광 > 가사보건 > 농업 순으로 나타났다. 에듀데이터 2009~2015년 5계년 데이터 가운데 특성화고를 포함한 전문계고, 특목고, 마이스터고, 종합고의 자료를 분석한 연구 결과, 학교 계열별 취업비율은 해양계열 > 공업계 > 상업계 > 실업계 > 수산계 > 농림업계 > 가사계 > 종합고 순으로 나타났다. 특히, 공업계열 학교의 취업비율 증가율이 가장 큰 폭으로 확인되었다(이쌍철, 엄문영, 2014). 한편, 한국교육고용패널(KEEP) 1차, 2차, 4차, 5차(2004, 2005, 2007, 2008) 데이터를 분석한 연구 결과에서는 학교 계열이 취업률에 미치는 효과가 통계적으로 유의하지 않은 것으로 나타났다(이쌍철, 주현준, 2013).

나) 학교 소재지

대도시에 위치한 학교가 그렇지 않은 학교보다 취업률이 높게 나타났다(장현진, 이지혜, 2012; 조규형, 정철영, 2014). 장현진과 이지혜(2012)의 특성화고 학생의 취업선택 결정요인을 분석한 연구 결과, 학교 소재지는 학생의 교육환경이 취업이나 진학과 관련한 인식형성에 영향을 미칠 수 있기 때문에 취업선택 시에도 학교 소재지는 취업영향력이 작용될 수 있음을 논하였다. 2013 에듀데이터(EDSS) 분석 결과, 특별·광역시 소재 학교들의 취업률이 가장 높고, 이어 시지역, 읍면지역, 특수지역 순으로 조사되었다. 또한 특별·광역시 학교들의 취업률 증가 폭이 타 지역에 비해 크게 나타났다(이쌍철, 엄문영, 2014). 2013년 특성화고 3학년 재학생을 모집단으로 조사, 분석한 연구에서는 취업결정요인 가운데 학교의 도시 소재 여부 및 수도권 소재 여부가 취업결정에 긍정적으로 영향을 미치는 것을 확인되었다(조규형, 정철영, 2014). 학교의 지리적 특성이 특성화고 학생의 취업결정에 영향을 미친다는 연구결과가 공통적으로 나타났으며, 이는 지역적 특성에 따른 일자리 공급과 지역특성을 반영한 특성화고 교육정책의 수용여부에 의한 차이로 해석될 수 있다.

다) 학비 및 급식비 지원

학비감면 수혜비율이 높은 학교의 취업비율은 높고, 또한 급식지원 학생 비율이 높을수록 학생 취업비율도 높게 나타났다(이쌍철, 엄문영, 2014). 학비감면 수혜비율은 학교차원의 특성이나 개인이 속한 가정의 사회경제적 지위 수준을 대변하는 지표들이라 할 수 있기 때문에 학생차원의 특성과 연결되어 있다. 이와 관련하여 특성화고에 소속된 학생들의 사회경제적 배경 환경이 낮을수록 취업 선택의 가능성이 증가하는 것으로 나타난 선행 연구 결과(이쌍철, 주현준, 2013; 장혜진, 이지혜, 2012)와의 상관성을 고려하여 고교의 학비 및 급식비 지원 비율이 갖는 취업률 예측효과를 추정해볼 수 있다. 한편, 장학금의 경우는 학비감면과는 다른 연관성을 갖는다. 장학금은 학생 개인특성 요인 가운데 성적과 연결된 변인이다. 특성화고의 경우 일반과목 성적과 전공과목 성적이 취업률에 미치는 영향이 다르다는 연구 결과(오석영, 2012)를 고려했을

때 성적을 기준으로 부여된 장학금 혜택이 취업률 예측에 어떻게 작용될 수 있을지 실제 영향력을 분석해볼 필요가 있다.

라) 학교경영방식

김지효(2018)의 특성화고 졸업생 취업에 미치는 영향요인 분석 연구에서는 학교장의 진로교육 정책에 대한 필요성 인식수준이 취업성과를 결정하는 데 통계적으로 유의한 예측 요인으로 확인되었다. 이 연구는 특성화고 취업 예측 요인으로 학교장의 영향력을 잠재적으로 고려해 볼 수 있는 근거를 제시하였다. 그밖에 학교경영방식과 관련하여, 학교운영위원회 등 학부모가 학교 경영 및 의사결정에 활발히 참여한 경우 학교 취업률이 낮아지는 것으로 나타났는데, 이는 학부모의 자녀 학력에 대한 기대수준에 의해 취업보다는 대학 진학 관련 학교활동에 대한 요구가 높아지기 때문이다(이쌍철, 주현준, 2013).

마) 학교시설

학교시설환경 변인은 학생의 취업선택에 통계적으로 유의한 영향을 미치지 않는 것으로 나타났다(장현진, 이지혜, 2012). 학교차원에서의 시설투자는 학교 교육의 만족도에 영향을 미칠 수 있으나, 취업성과에 대한 직접적인 영향력은 확인되지 않았다. 그러나 최근 특성화고는 NCS(국가직무능력표준)기반 교육과정을 운영하면서 실무과목에 필요한 산업체 중심의 설비를 교내에 구축하려 학교시설 예산을 투입하고 있다. 2016년 고교 직업교육기관 표준교육비 산출 연구에 따르면, 표준 설비 재료비에는 공업고 > 농생명고 > 상업정보고 순으로 많이 나타났으며, 2011년에 비해 공업고의 경우 5.37배, 농·생명고 4.98배, 상업·정보고 1.92배 증가하였다. 이에 특성화고의 교육과정 변화에 따른 최근 학교시설에 투입된 예산은 직업교육 성과에 이전과 다른 잠재적 효과를 나타낼 수 있는 가능성이 제기되고 있다(이영민 외, 2016).

<표 II-6> 특성화고 취업률 영향요인 관련 문헌분석 종합

연구	활용데이터	개인특성 요인	교수학습 요인	학교환경 요인
이지혜 & 정철영 (2010)	2009 특성화고 3학년			- 계열(o)
노경란 & 허신주 (2012)	HSGOMS ⁵⁾ 2012	- 성별(o) - 학교성적 - 부모학력 - 가계소득		- 학교만족도
장현진 & 이지혜 (2012)	KEEP ⁶⁾ 4차 2007	- 성별(o) - 가구소득(-)	- 동아리활동 (+)	- 학교계열(o) - 학교소재지(o) - 학교시설환경(x)
오석영 (2012)	2011, 2012 특성화고 진로 이력분석	- 전공성적(+) - 국·영·수 성적(-)		
최동선 & 이종범 (2013)	HSGOMS 2011			- 계열(o)
이쌍철 & 주현준 (2013)	KEEP 1차, 2차, 4차, 5차 (2004, 2005, 2007, 2008)	- 성별(o) - 가구소득(-)	- 진로교육(+)	- 계열(x) - 전공만족(-) - 학부모 학교경영 참여(-)
이쌍철 & 엄문영 (2014)	EDSS ⁷⁾ 2009 ~2013		- 동아리 참여비율(+) - 특기적성 방과 후 참여비율(+)	- 학교소재지(o) - 학업중단율(-) - 학교계열(x) - 학비감면 수혜 비율(+) - 급식지원 학생 비율(+)

연구	활용데이터	개인특성 요인	교수학습 요인	학교환경 요인
조규형 & 정철영 (2014)	2013 특성화고 3학년 재학생	- 자기효능감(+)	- 현장실습프로그램(o)	- 학교소재지(o)
천영민 & 오민홍 (2015)	HSGES 2013	- 성별(o) - 학교성적(+) - 학교생활(+) - 보호자소득(+)	- 방과후 학교 참여(o) - 야학·작상 면접자(o) - 경진대회 및 공모전 수상여부(o) - 자격증취득 여부(o)	- 졸업학교 전반적 만족도(+)
홍성표 & 정진철 (2016)	2014 특성화고 재학생	- 진로결정 자기효능감(+) - 전공교과 학업성취(+)		
김지효 (2018)	2012 서울교육 종단연구 (2015 특성화고 3학년)	- 성별(-o) - 부모소득(x) - 사교육지원정도(-) - 진학정보인식(-)	- 동아리활동(o) - 진로프로그램 운영빈도(o)	- 학교장의 진로교육 정책 필요성 인식(+)

※ (참고) note. (+) 정적영향, (-) 부적영향, (o) 유의성 있음, (x) 유의성 없음.

- 5) 고졸자취업진로조사(HSGOMS: High School Graduates Employment Survey)는 미진학 고등학교 졸업생들을 대상으로 2011년부터 격년 주기로 고졸 청년층에 대한 고용정책 수립을 위한 기초통계 정보를 구축하기 위해 실시되었다.
- 6) 한국교육고용패널(KEEP:)은 2004년부터 같은 표본을 1년 주기로 추적 조사한 패널조사로 한국직업 능력개발원에서 실시하였으며, 특성화고 학생, 학부모, 학교를 대상으로 조사한 자료를 포함하고 있다.
- 7) 에듀데이터(EDSS: EduData Service System)는 유치·초·중·고 및 대학의 교육통계 및 학교정보공시 자료와 연계된 데이터로 중등단계에서 특성화고, 전문계고, 특목고, 마이스터고, 종합고 자료를 포함하고 있다.

4) 해외 직업교육 연구 사례

가) 중국

중국 직업교육은 인적자원 구축을 지원함으로써 경제 및 사회 개발에 기여하고 있다. 중등 직업 학교(secondary vocational school) 졸업생 취업률은 10 년 연속 95% 이상을 유지했으며, 졸업 후 6 개월 후에 직업 대학 졸업생의 취업률은 최근 90%를 초과했다. 취업률 관련 영향 요인으로, 2018년 직업 교육 졸업생의 70%가 학교 근처 지역과 도시에서 일자리를 갖는 것으로 나타났다. 또한, 현대 제조업 및 신흥 산업 인력은 70% 이상이 직업 교육 기관을 졸업한 것으로 나타나 중국 직업학교계열은 최근 취업률이 높은 산업의 영향을 받는 것으로 이해할 수 있다. 중국 직업 교육 졸업생들은 중소기업 클러스터 또는 네트워크 개발을 촉진하고 지역 산업 가치를 상승시키는 등 우수한 성과를 내고 있는 실정이다(Ministry of Education of the People's Republic of China, 2019).

나) 독일

독일의 직업교육은 직업학교와 기업 두 기관으로 이원화되어 시행되고 있다. 두 기관은 실무와 이론을 상호 협력하여 교육하고 있다. 독일의 직업학교의 경우 일반학교와 유사한 진학자격을 수여하는 직업훈련학교와 취업을 할 수 있는 직업 자격(marketable vocational qualification)을 수여하는 직업전문학교, 전문학교, 보건학교 등이 있다. 직업학교의 교과과정은 연방직업훈련연구소(BIBB)와 같은 기관의 중재로 기업과 학교 간의 협력과 현장의 요구 등이 수용되고 반영된다. 이러한 독일의 직업교육과는 상이하게 우리나라 직업교육제도는 공공직업훈련 교육과정 중심으로 편성되어 실제 졸업생이 고용으로 진입할 수 있는 한계를 갖는다. 이처럼 직업교육이 실제 산업체의 인력 수요를 반영하지 못한 현실은 직업계고 학생의 중도탈락률이 높이는 데 영향을 미치고 있다. 독일의 직업교육훈련제도를 분석한 장석인(2006) 연구는 우리나라 직업계고교가 산업현장에서 필요로 하는 내용을 학교교육에서 훈련할 수 있도록 기업의 참여를 강화하고 이러한 직업교육의 운영 특성이 곧 학생의 참여 기업 취업으로 연결될 수 있는 직업교육 시스템을 제안하였다.

다) 스위스

스위스는 독일과 유사한 복선형 학제로 서로 다른 학교체제가 있어 교육 연한이나 교육 내용이 다르고 상호간의 이동이 불가능하다. 직업교육은 후기 중등교육과정의 VET(Vocational Education and Training) 과정에서 수행한다. VET 과정은 다시 상대적으로 전문성이 낮은 직업군을 준비하는 2년 과정과 높은 수준의 전문인력을 훈련하는 3~4년 과정으로 분류된다. 고등교육 단계에서는 전문직업훈련대학(PET)에 진학하여 심화된 지식과 기술을 습득할 수 있다. 스위스는 후기중등직업교육에 진학하는 비율이 절반 정도가 될 정도로 높으며, OECD(2009) 데이터에 의하면 고등교육 진학률이 점차 상승하는 추세를 보인다. 스위스는 고교단계 직업교육에서 도제훈련 제도를 운영하고 있으며, 이는 독일의 이원화 직업교육 시스템과 유사하게 이론교육은 학교에서, 현장훈련은 기업에서 담당하고 있다. 스위스의 경우 기업의 1/3 정도가 도제훈련에 참여하고 있고, 이 가운데 2/3의 기업들이 도제훈련 제도에 참여함으로써 인하여 기업운영에 이득이 된다고 보고하고 있다. 이는 곧 기업과 학교가 노동시장을 이끄는 파트너가 되어 청년 실업 감소에 가시적인 성과를 달성한 것으로 평가된다. 실제 도제훈련 참가자 수와 스위스 GDP 성장률 간에 밀접한 상관관계가 확인되었다(전승환, 2014)

라) 미국

미국의 한 연구는 고등학교를 졸업한 이후, 다시 직업교육에 참여하게 되는 영향 요인을 조사하였다. 이 연구 결과, 고등학교에서의 학업성취도가 낮은 학생, 교육적 포부가 낮은 학생, 그리고 사회 경제적 배경 수준이 낮은 학생이 직업 교육에 참여할 가능성이 더 높은 것으로 나타났다. 이 연구에 분석된 학생 표집은 우리나라 직업계 고등학교 졸업생과는 다소 차이가 있으나, 취업을 적극적으로 준비하는 학생의 특성 요인이 유사한 경향성을 띄고 있다. 미국 직업교육의 차별화된 상황으로 장애가 있는 학생이나 흑인 학생들이 직업 교육의 참여가 높은 것으로 보고되고 있다(Agodini, Uhl, & Novak, 2004).

마) 오스트레일리아(AU)

AU의 한 보고서는 학교교육에서의 직업교육 질에 영향을 주는 요인들을 분석하여 제시하였다. 주요 영향 요인으로 학교차원의 교육예산, 물질적 인프라, 교사의 산업체 경험과 전문성 개발, 그리고 산업체의 연계 참여 등을 언급하였다. 이 보고서는 Australian National Training Authority(ANTA) 편당이 호주 전역의 직업교육을 위한 학교 프로그램에 투자되었음을 소개하며, 향후 학교를 주체로 직업교육을 계획하고 교육과정을 관리, 운영을 지속하기 위한 국가적 자금 지원의 중요성을 제안하였다(Karmel, 2007).

바) 터키

터키의 한 연구는 직업계고교 학생과 일반고교 학생의 취업성과를 비교 분석하였다. 2000년도 기준 터키에서 고등학교를 졸업한 26-35세 사이 남성을 대상으로 조사 분석한 연구 결과 VHS(vocational high school)를 졸업한 학생의 취업률이 약 5% 높게 나타나는 경향을 확인하였으며 통계적으로 유의한 차이가 나타났다. 이 연구는 터키만의 주요한 노동시장의 특징을 고려하고 있다. 터키에서는 일반고등학교를 졸업한 학생이 고등교육으로 진학하지 않은 학생에 비해 취업 역량이 부족하다고 평가됨은 물론이거니와 직업계고교에서와 같이 커리어 중심의 교육이 아닌 폭넓은 교육과정을 배운 학생으로 취급되고 있어 VHS 학생이 실제 노동시장에서 더 나은 평가를 받고 있다(Torun & Tumen, 2017).

라. 소결

특성화고 학생들의 취업 가능성, 취업성과, 취업진로 결정에 영향을 미치는 요인들을 배경 문헌으로부터 분석하고, 이를 토대로 에듀데이터(EDSS)의 관련 변인들과 매칭하여 도출한 특성화고 취업률 예측요인은 <표 II-7>와 같다. 취업률에 영향을 미칠 수 있는 요인은 개인특성, 학교환경, 교수학습으로 구분된다. 첫째, 개인특성 요인으로 학생의 성별, 학업성적, 학업중단율, 학교교육 만

죽도와 같은 학생 특성을 반영한 요인들로 분석되었으며, 이와 관련된 접근 가능한 EDSS 데이터는 남녀 학생 수 비율과 학업중단을 정도로 매칭할 수 있다. 둘째, 교수학습 요인으로 동아리 활동, 방과후 학교, 선택적 교육활동, 교육활동 예산이 취업률 예측 요인으로 분석되었으며, 이는 EDSS 데이터의 동아리 참여비율, 방과후 학교 참여 비율, 선택적 교육활동 예산, 학생당 교육활동 지원 예산과 매칭된다. 셋째, 학교환경 요인으로 학교 소재지, 계열, 경영방식, 시설, 학비감면 및 급식비 지원 비율이 주요 예측요인으로 분석되었으며, 이는 EDSS의 학교특성 데이터 가운데 시도명, 시군명, 지역행정구분명, 학제명, 교장공모제여부, 학교시설 확충 예산, 학비감면액, 장학금, 급식비 지원 정보 데이터와 매칭할 수 있다.

<표 II-7> 특성학교 취업률 예측 요인 도출

구분	배경문헌 요인 분석		EDSS 변수 도출
개인특성 요인	<ul style="list-style-type: none"> · 성별 · 학업중단율/학업성적/학교교육 만족도 	⇒	<ul style="list-style-type: none"> · 남학생 수 비율 · 여학생 수 비율 · 학업중단율
교수학습 요인	<ul style="list-style-type: none"> · 동아리 활동 · 방과후 학교 · 선택적 교육활동 · 교육활동 예산 	⇒	<ul style="list-style-type: none"> · 동아리 참여비율 · 방과후 학교 참여비율 · 학생당 선택적 교육활동 예산 · 학생당 교육활동 지원 예산
학교환경 요인	<ul style="list-style-type: none"> · 학교소재지 · 학교계열 · 학교경영방식 · 학교시설 · 학비감면 수혜비율 · 급식비지원 학생 비율 	⇒	<ul style="list-style-type: none"> · 시도명 · 시군명 · 지역행정구분명 (특별/광역시, 시, 읍/면/특수지역 구분) · 학제명 · 교장공모제여부 · 학생당 학교시설 확충 예산 · 학생당 학비감면액 · 학생당 장학금 · 학생당 급식비 지원액 · 급식비 지원 학생 수 비율

2. 머신러닝 관련 배경 문헌 분석

가. 머신러닝의 개념과 주요 알고리즘

머신러닝(machine learning)은 기계학습이라고도 일컬으며, 인공지능의 한 분야로 데이터를 축적하고 학습하여 미래에 입력된 데이터에 대한 적절한 작업 수행 및 미래 발생할 일들을 예측하는 기술이다(오미애, 2017). 다시 말해 머신러닝은 컴퓨터가 알고리즘을 활용하여 축적된 데이터로부터 패턴을 기계 스스로 학습하여 이후 발생된 일을 예측하는 모델을 생성하고, 모델을 평가, 진화시키는 것을 가능하게 한다. 또한 머신 러닝은 이미지 처리, 영상인식, 음성인식, 웹 검색 분야에 적용되어 앞으로의 일을 예측하고 맞춤형 데이터를 제안하는 핵심기술로 활용되고 있다.

머신러닝 일련의 처리 과정은 표현(representation), 평가(evaluation), 최적화(optimization), 일반화(generalization) 단계를 거친다(오미애, 2017). 표현(representation)은 예측모형으로 입력한 일련의 데이터를 처리하는 방법을 결정한다. 결정된 모형은 평가단계를 통해 목표한 과업이 정확히 수행되었는지를 판단한다. 그리고 평가기준에서 가장 만족도가 높은 조건을 찾아 최적화한다. 표현→평가→최적화 단계를 통해 충분히 학습되어진 예측모형은 일반화되어 새로운 데이터로부터 결과를 예측하는데 활용될 수 있다.

머신러닝 예측모형은 그 성능을 설명할 수 있는 성능 측정과 분석을 위한 몇 가지 개념들의 이해를 바탕으로 유효성을 판단할 수 있다. TP(True Positive: 예측값과 관측값이 모두 참인 경우), FP(False Positive: 예측값은 참이었으나 관측값이 거짓인 경우), FN(False Negative: 예측값은 거짓이나 관측값이 참인 경우), TN(True Negative: 예측값이 거짓이고 관측값도 거짓인 경우) 개념을 기초로 TP rate, FP rate Accuracy, Recall, Precision, F-measure, ROC area의 의미를 해석하여 예측모형의 성능을 분석한다. 머신러닝 성능평가를 위한 개념들의 상세의미와 수식은 <표Ⅱ-8>과 같이 정리한다(Davis & Goadrich, 2006; Fritsch et al., 2013; Kumar & Sahoo, 2012).

<표 II-8> 머신러닝 예측모형 성능 평가 관련 개념

개념	의미	표현식 또는 판단기준
TP rate (진양성률)	관측값이 참인 경우 중 예측값이 참인 비율을 의미	$True\ Positive\ rate = \frac{TP}{(TP+FN)}$
FP rate (위양성률)	관측값이 거짓인 모든 값들 중 예측값이 참인 비율	$False\ Positive\ rate = \frac{FP}{(FP+TN)}$
Accuracy (정확도)	전체 예측값 중 관측값과 정확하게 일치하는 비율	$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}$
Recall (재현율)	관측값이 참인 경우 중 예측값이 참인 비율	$Recall = \frac{TP}{(TP+FN)}$
Precision (정밀도)	예측값이 참인 모든 값들 중 관측값이 참인 비율	$Precision = \frac{TP}{(TP+FP)}$
F-measure	Precision과 Recall의 조화평균을 의미	$F-measure = 2 \times \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}}$ $= 2 \times \frac{Precision \times Recall}{Precision + Recall}$
ROC area	예측결과와 실제결과 간의 예측성능 비교정보로서 측모형 유용성과 절단값 (cut-off value)을 판단	AUC≤0.5 : 유효성 가치 없음 0.5<AUC≤0.7 : 유효성이 낮음 0.7<AUC≤0.9 : 유효성 중간수준 0.9<AUC<1 : 유효성이 높음 AUC=1 : 유효성 완벽

머신러닝 알고리즘은 사람의 조작 없이 스스로 데이터를 통해 학습하고 예측모형의 성능평가를 통해 개선 할 수 있는 프로그램을 구현한다. 머신러닝 알고리즘은 세 가지 유형이 있다. 첫 번째로 지도학습(supervised learning)은 레이블이 지정된 학습 데이터를 사용하여 입력 변수 (X)를 출력 변수 (Y)로 변환하는 매핑 기능을 학습하여 새로운 입력이 주어지면 정확하게 출력을 생성해낼 수 있다. 주로 분류와 회귀 유형의 지도 학습이 많이 쓰인다. 분류(classification)는 입력 데이터를 보고 ‘아프다’와 ‘건강하다’와 같은 레이블을 예측하는 사례와 같이 출력 변수가 범주의 형태로 주어진 샘플의 결과를 예측

하는 데 사용된다. 회귀(regression)는 입력 데이터를 처리하여 강수량, 사람이 키 등의 수치를 예측하는 사례와 같이 출력 변수가 실제 값의 형태 일 때 주어진 샘플의 결과를 예측하는 데 사용된다. 앙상블(ensembling)은 또 다른 유형의 지도학습 알고리즘으로 새로운 샘플에 대해 보다 정확한 예측을 생성하기 위해 여러 머신러닝 모델의 예측을 결합하는 것을 의미한다. 지도학습 알고리즘은 대표적으로 선형 회귀(Linear Regression), 로지스틱 회귀(Logistic Regression), CART, Naïve-Bayes, K-Nearest Neighbors (KNN) 알고리즘이 있다.

두 번째로 비지도학습(unsupervised learning) 모델은 입력 변수 (X)만 있고 해당 출력 변수가 없을 때 사용된다. 즉 레이블이 없는 학습 데이터를 사용하여 데이터의 기본 구조를 모델링한다. 연관(association), 군집(clustering), 차원 축소(dimentionality reduction)가 비지도학습의 유형이다. 연관(association)은 상호 경향이 동시에 발생할 가능성을 예견하는 것으로 시장-바구니 분석(market-basket analysis)으로 많이 활용된다. 예를 들어, 고객이 빵을 구매하는 경우 계란을 구매할 가능성이 80 %라는 것을 발견하기 위해 연관 모델이 사용된다. 군집(clustering)은 서로 유사하도록 샘플을 그룹화 하는데 사용된다. 차원 축소(dimentionality reduction)는 중요한 정보를 계속 전달하면서 데이터 세트의 변수 수를 줄이는 방법으로, 특징 추출 방법(feature extraction methods) 및 특징 선택 방법(feature selection moehods)을 사용하여 수행 한다. 대표적 비지도학습 알고리즘으로 Apriori, K-means, PCA가 있다.

세 번째로, 강화학습(reinforcement learning)은 현재 상태에 따라 최상의 다음 행동을 결정할 수 있게 해주는 알고리즘이다. 강화 알고리즘은 시행착오를 통해 최적의 행동을 학습한다. 예를 들어, 게임 캐릭터가 시행착오를 통해 시간이 지남에 따라 점수를 많이 얻을 수 있는 움직임의 위치와 시기를 결정하는 것과 같다.

각 학습 유형별 몇 가지 대표 머신러닝 알고리즘을 간략하게 소개한다.

1) 지도학습(supervised learning) 알고리즘

가) 회귀(Regression)

기본이 되는 선형회귀 알고리즘은 출력변수(y)를 결정하는 데 사용되는 입력 변수 데이터 세트(X)가 있는 경우, 둘 사이의 관계를 정량화하는 기능을 한다. 즉, 입력변수 (x)와 출력변수 (y) 사이의 관계는 $y = a + bX$ 형식으로 표현하며, 알고리즘은 a(절편)와 b(기울기)의 값을 찾아 데이터 세트의 분포 점들에 가장 근접한 선을 맞춰 오류를 최소화 한다. 통계에서 활용되던 회귀모델은 라쏘(LASSO), 로지스틱(Logistic), 능형(Ridge) 등이 머신러닝에 적용될 수 있다. 라쏘는 선형회귀 유형으로 제곱합의 오차(sum of squared error)를 최소화할 수 있으며, 로지스틱은 종속변수와 독립변수의 관계 측정을 위한 표준회귀의 변형이고, 능형은 높은 상관관계를 갖는 독립변수를 분석하는데 사용된다.

나) 의사결정나무(Decision Tree)

의사결정 나무는 변수들 간의 관계를 나타내는 데이터 구조이다. 노드(node)는 분류를 표현하고 탐색을 구체화시키는 질문을 통해 나무의 형태도 노드들이 뻗어가는 그림이 그려진다. 다양한 머신러닝 알고리즘은 의사결정나무의 형태로 속성을 탐색하는 모형을 띄고 있다. Gradient Boosting과 Radom Forest 알고리즘이 대표적 예이다. Gradient Boosting은 문제해결을 위한 의사결정 알고리즘으로 사용되며, Radom Forest는 특이사항과 이상 징후 패턴을 찾아 데이터를 분류하는 데 사용된다.

그밖에, CART(Classification and Regression Trees)를 직역하면 분류 및 회귀 트리로 의사결정 트리 알고리즘의 한 가지 구현방법이다. 트리구조는 시작 조건이 될 수 있는 루트 노드(root node)와 중간 연결고리가 될 수 있는 인터널 노드(internal node), 출력 값을 나타내는 리프 노드(leaf node)로 구성된다. 의사결정 트리는 입력값을 조건에 따라 분류하여 결과를 예측하는데 사용된다.

다) Naïve Bayes

Bayes는 이미 발생한 사건들로 인하여 한 사건이 발생할 확률을 계산하는

알고리즘이다. 즉 모든 수가 독립적임을 가정하고 아래 표현식과 같이 사전 지식(d)을 고려할 때 가설(h)이 참일 확률을 알려준다.

$$P(h | d) = (P(d | h) P(h)) / P(d)$$

$P(h | d)$ = 사후 확률, 데이터 d가 주어지면 가설 h가 참일 확률

$P(d | h)$ = 가능성, 가설 h가 참인 경우 데이터 d의 확률

$P(h)$ = (데이터에 관계없이) 가설 h가 참일 확률

$P(d)$ = (가설에 관계없이) 데이터 확률

라) 뉴럴네트워크(Neural networks)

뉴럴네트워크 알고리즘은 뇌의 신경망으로부터 아이디어를 착안하여 입력, 처리(히든), 출력 노드로 구성된다. 이 알고리즘은 각 신경망에 가중치를 부여하고, 반복 작업을 통해 목표점에 도달할 때까지 출력에 의해 가중치가 조절된다. 예를 들어 설정된 이미지 없이 반복 학습을 통해 패턴을 발견하도록 설계하는 시스템에 사용된다. 이런 신경망 알고리즘은 음성인식, 개체인식, 이미지 복원, 부정탐지 등에 적용된다.

마) 서포트 벡터 머신(SVM)

SVM(Support Vector Machine)은 레이블이 있는 트레이닝 데이터를 사용하고 옵티멀 하이퍼플레인(hyperplane)으로 결과를 출력하는 머신러닝 알고리즘이다. 하이퍼플레인 차원(dimension)을 하나 내린 부분 공간으로 평면에서의 선과 같다. SVM은 입력되는 특징들이 적을 때 사용된다.

바) k-NN

k-Nearest Neighbors 알고리즘은 데이터 세트를 학습 세트와 테스트 세트로 나누지 않고 전체 데이터 세트를 머신러닝의 트레이닝 세트로 사용한다. 새 데이터 인스턴스에 대한 결과가 필요한 경우 kNN 알고리즘은 전체 데이터 세트

를 통해 새 인스턴스에 대한 k-근접(nearest) 인스턴스 또는 새 레코드와 가장 유사한 k개의 인스턴스를 찾아 다음 평균을 출력한다. k 값은 사용자가 지정한다. 인스턴스 간의 유사성은 Euclidean과 Hamming distance 측정값으로 분석한다. K-NN은 데이터의 분포에 대해서 제한적인 정보만 있을 때 주로 사용된다.

2) 비지도학습(unsupervised learning) 알고리즘

가) Apriori

Apriori 알고리즘은 자주 추출되는 데이터 세트에서 연관된 규칙을 생성하는데 사용된다. 데이터베이스에서 자주 동시에 일어나는 제품들의 같이 조합을 확인하는 market basket analysis로 잘 알려져 있다. 예를 들어 특정 상품을 구매하는 성향이 다른 상품의 구매 성향을 파악하는 데 그에 대한 연결 규칙을 $X \rightarrow Y$ 로 표시한다.

나) 클러스터링 기술(Clustering Techniques)

데이터 샘플에 존재하는 그룹을 찾기 위해 사용되는 기술로 특정한 기준을 바탕으로 변수들이 그룹으로 분류된다. K-means는 유사한 데이터를 군집으로 그룹화하는 반복 알고리즘으로, k 군집의 중심을 계산하고 중심점과 데이터 점 사이의 거리가 가장 작은 군집을 찾아 data point를 지정해준다. K-means는 가장 폭 넓게 사용하는 자율학습 알고리즘이다. 클러스터링의 데이터 혼합 밀도를 극대화할 수 있는 EM 알고리즘도 있다.

다) 주성분 분석(PCA)

PCA (Principal Component Analysis)는 변수의 수를 줄임으로써 데이터를 쉽게 탐색하고 시각화하는 데 사용된다. 데이터의 최대 분산을 새로운 좌표계로 캡처하여 선형 관계의 교차 등을 시각적으로 분석할 수 있다.

라) 커널 밀도 추정(KDE)

Kernel Density Estimation은 확률분포 또는 데이터의 밀도를 추정하여 확률 변수들 사이의 관계를 측정한다. KDE는 한정된 데이터 샘플만으로 추론이 이루어질 때 데이터를 매끄럽게 할 수 있으며, 일반적으로 리스크 관리와 재무 모델링을 위한 분석에 적용된다.

마) 비부정 행렬 인수분해(NMF)

Nonnegative Matrix Factorization은 패턴 인식에 유용하여 머신러닝을 통해 유전자 발현 분석과 소셜 네트워크 분석 분야에 사용된다. 클러스터링 또는 분류 도구로도 사용되며 K-means 클러스터링과 유사하다.

바) 특이값 분해(SVD)

Singular Value Decomposition은 어떤 변수가 가장 중요하고 어떤 것이 제거될 수 있는지 결정하는 알고리즘으로 성능개선을 위해 중복된 데이터를 제거하는데 도움이 된다. 예를 들어 습도와 비 올 확률과 같이 상관관계가 높은 두 개의 변수가 이미 존재한다면, 다른 데이터를 추가하지 않는다. 이처럼 SVD는 어떤 변수가 모델에서 유지되어야 하는지 결정하는데 사용된다.

사) 자기 조직화 지도(SOM)

Self Organizing Map은 패턴 인식 프로세스를 갖는 자율 뉴럴네트워크 모델이다. 시각화된 센서를 대뇌피질에 매핑시키는 방식과 같은 추상화된 수학적 모델로 두뇌가 어떻게 작동하는지에 대한 이해를 바탕으로 머신러닝의 패턴인식에 응용되었다.

3) 강화학습(reinforcement learning) 알고리즘

가) Bagging with Random Forests

Bagging 첫 단계는 Bootstrap sampling 방법으로 데이터 세트로부터 여러 모델을 생성한다. 두 번째 단계는 생성된 다른 트레이닝 세트에서 동일한 알

고리즘을 사용하여 여러 모델을 작성하다. Bagging에서 Random Forests를 사용하는 것은 의사결정 트리 방법과 차별화하여 최상의 분할을 구성하기 위해 임의의 기능을 선택하게 된다. 이는 의사결정 트리에서 유사 구조나 상호 연관된 예측 결과가 나오는 경우를 줄이고자 함이다.

나) Boosting with AdaBoost

Boosting은 여러 가지 약한 분류기(classifier)에서 강력한 분류기를 만드는 일반적인 앙상블 방법이다. Boosting은 훈련 데이터에서 모델을 만든 다음 첫 번째 모델의 오류를 수정하는 두 번째 모델을 만들어내고, 훈련 세트가 완벽하게 예측되거나 최대 개수의 모델이 추가 될 때까지 모델을 추가한다. Adaboost는 Adaptive Boosting의 약자로 의사결정을 수행하는 과정에서 이전 분류를 위해 개발된 알고리즘이다.

4) 엔트로피와 정보획득 지수

가) 엔트로피(entropy)

엔트로피는 정보의 확신도 또는 확률분포가 갖는 정보량을 수치화한 것이다. 엔트로피는 확률밀도가 특정 데이터에 몰려있으면 작아지고, 고루 퍼져 있으면 커진다. 즉, 특정 값이 나올 확률이 높아지고 나머지 확률이 낮아지면 엔트로피는 작아지고, 반대로 여러 데이터가 나올 확률이 유사한 경우 엔트로피가 높아진다. 의사결정트리 모델에서는 엔트로피를 순도(purity)를 결정하는 역할을 한다. 지니 계수(Gini Impurity Measure)라 불순도 측정 공식에 엔트로피가 사용되며, 불순도는 분류된 범주에 다양한 속성을 지닌 개체들이 얼마나 포함되어 있는가를 말해준다.

머신러닝에서는 엔트로피를 정보획득 지수로도 활용한다. 분석 데이터의 불확실성을 정량적으로 표현한 것으로 엔트로피를 근거로 획득할 수 있는 정보의 양을 알 수 있다. 엔트로피와 정보획득 IG(information gain) 지수 계산식은 아래와 같다(홍석미, 안종일, 정태충, 2004; 이형일, 2009). 엔트로피가 증가하면 불확실성이 증가하고 그로인하여 획득할 수 있는 정보의 양이 감소한다.

$$Entropy(S) = - \sum_{i=1}^n p_i \log_2(p_i)$$

※ S : 변인, n : S 내 클래스(class) 수, C_i : i 번째 클래스($i = 1, 2, \dots, n$)
 $p_i: \frac{|C_i|}{|S|}$, $|S|$ 는 변인에 해당하는 표본의 개수

$$IG(X, Y) = Entropy(Y) - \sum_{i=1}^k p_i \times Entropy(C_i, Y)$$

※ Y : 반응변수, X : 설명변수, k : X 내 클래스(class) 수,
 C_i : i 번째 클래스($i = 1, 2, \dots, k$), $p_i: \frac{|C_i|}{|X|}$,
 $Entropy(C_i, Y)$: C_i 에 해당하는 Y 의 엔트로피

나. 머신러닝 기반 교육 분야 연구 사례

최근 교육 분야에서는 특정 종속변인을 예측하는 모형을 탐색하고 비교·검증하기 위하여 머신러닝 분석기법 및 머신러닝 분석 도구를 활용하고 있다. 최근 머신러닝을 교육 분야 연구에 적용한 사례로는 학습 성취 예측모형을 탐색하거나 교육성과 특히 취업률 예측모형을 제안하고 검증하는 연구들에서 주로 찾아볼 수 있다. 그 외 교육과 관련된 사회적 문제를 머신러닝 분석 방법을 적용하여 원인을 탐색하고 해법을 논의한 연구도 있다.

1) 학습 성취 예측모형 연구 적용 사례

머신러닝을 학습 성취 분석에 활용한 사례로, 조현국(2018) 연구는 이러닝 강의에 참여한 학습자 정보를 토대로 시험, 과제, 출석 성적의 중요도를 분석하였다. 이 연구에서는 k-근접 이웃 알고리즘, 서포트 벡터 머신, 의사결정 나무, 랜덤 포레스트, 그래디언트 부스팅, 인공 신경망, 여섯 가지 알고리즘을 사용한 예측 모형별 변수의 중요도 순을 비교 제시하였다. 그리고 중요도가 낮은 요인들이 효과적인 학습을 발현시키는 개선점을 논하였다. 대학생의 학업성취 예측과 관련한 또 다른 연구 사례로, 동영상 기반 학습 환경에서 대학생의 행

동로그가 학업성취에 미치는 영향을 분석한 이정은 외(2020) 연구에서는 머신러닝 분석 도구 R을 활용하였다. 학습자 행동로그 분류를 독립변인으로 하여 종속변인 학업성취의 예측 모형을 K-근접 이웃(K-Nearest Neighbors; KNN), 인공신경망(Artificial Neural Network), 서포트 벡터 머신(Support Vector Machine; SVM), 랜덤 포레스트(RandomForest) 기법을 적용하여 검증하였다. 중학생의 국가수준 기초학력결과에 영향을 미치는 예측요인을 찾기 위해 머신러닝을 활용한 연구 사례도 있다. 이종현(2020)은 에듀데이터 서비스 시스템에서 제공하는 데이터를 기반으로 학생, 교사, 학교, 기관 요인을 독립변인으로, 수학 기초학력 미달 비율을 종속 변인으로 연구모형을 설계하여 WEKA 3.8 버전을 활용하여 데이터 마이닝 분석을 실시하였다. 데이터 마이닝은 대규모 데이터로부터 통계적 패턴을 찾는 방법으로 이 연구에서는 기초학력 미달 비율이 높은 학교와 낮은 학교 간의 변인별 영향력을 탐색하는데 적용되었다.

2) 취업률 및 교육성과 예측모형 연구 적용 사례

머신러닝을 적용하여 청년 취업에 영향을 미치는 요인을 탐색적으로 분석한 한성민 외(2016)의 연구에서는 2012년 노동패널 데이터를 활용하여 개인특성, 가족환경, 취업준비 구분 변수들 가운데 취업결정 요인을 알아내기 위해 데이터 마이닝 분석 방법을 활용하였다. 이 연구에서는 머신러닝 기법으로 의사결정나무 분석과 신경망 분석 방법을 혼합하여 사용하였다. 의사결정나무 분석 방법은 변수 간 교호작용 파악이 용이하다는 장점을 활용하여 청년층의 주요한 취업결정 요인을 분석하는데 사용되었으며, 신경망 분석 방법은 여러 독립 변수들의 상호결합이 종속변수에 대해 높은 예측력을 갖는다는 장점을 활용하여 취업결정 요인의 예측식을 도출하여 영향력을 설명하는데 사용되었다(한성민, 문상호, 이숙중, 2016). 머신러닝을 활용하여 취업률 예측 모델을 제안한 이동훈(2019)의 연구는 의사결정나무, 랜덤포레스트, 인공신경망 알고리즘을 사용한 모델 간 성능을 비교 분석하였다. 이 연구는 2016년 대졸자 직업 이동경로 조사 데이터를 토대로 대졸자의 고용 정책 수립과 취업 지원 프로그램 개발과 같은 최근 사회 문제의 해결방안을 머신러닝 분석 모델을 통해 제안하고자 한

연구로 의의가 있다. 머신러닝을 활용하여 대졸자의 취업 여부와 취업 질을 예측하는 모형을 제한한 연구 사례도 있다. 최필선과 민인식(2018)의 연구는 대학 졸업자가 노동시장 진입과 이동에 영향을 미칠 수 있는 변수들을 선정하여 취업과 미취업, 그리고 정규직 취업과 비정규직 취업을 예측하는 모형을 제시하였다. 이 연구는 청년실업률과 같은 사회적 이슈를 예측할 수 있는 변인으로 대학에서의 교육 훈련 프로그램과 교육성과 및 스펙 변수들을 고려하고 있다. 그밖에 특성화고등학교 교육성과에 영향을 미치는 요인 탐색을 위해 초중등 교육정보 고시 자료를 활용하여 데이터 마이닝 분석을 수행한 연구 사례도 찾아볼 수 있다. 김진(2014) 연구는 교육성과 분석도구로 WEKA 3.6을 활용하였으며, Apriori(연관규칙), Linear Regression(회귀분석), 분류화 및 의사결정트리 알고리즘을 사용하여 학교현황, 교육활동, 학생현황, 교원현황, 재정현황이 취업률, 진학률, 입상성적에 미치는 영향력을 예측하는 연구 모형을 검증하였다.

3) 기타 교육 분야 연구 적용 사례

사교육과 같은 교육문제 관련 연구에도 머신러닝 분석 기법이 적용된 사례를 찾아볼 수 있다. 고등학생들이 사교육에 참여하는 예측 변수를 탐색한 김영식 외(2019) 연구에서는 표본으로 제한된 데이터 분석의 일반화 가능성을 높이기 위해 R 분석 도구의 머신러닝 기법을 활용하였다. 이 연구에 사용된 머신러닝 분석 방법은, 랜덤 포레스트, 나이브 베이즈, 서포트 벡터 머신, 인공신경망으로 사교육 참여 관련 중요 변수 10개에 대한 예측 분석 결과를 분석기법별로 비교하여 제시하고 있다.

다. 머신러닝을 위한 도구

머신러닝을 위한 대표적 도구로 WEKA는 뉴질랜드 와이카토 대학교 (University of Waikato)가 개발한 JAVA기반 플랫폼이다. WEKA란 Waikato Environment for Knowledge Analysis의 줄임으로, 뉴질랜드 섬에서만 볼 수 있는 Mecca 새 이름의 라임을 따라 붙여진 이름으로 알려졌다. [그림 II-4]은 WEKA 3.8의 메인 화면이다.



[그림 II-4] WEKA 메인 GUI 화면

WEKA는 주로 데이터 마이닝에 목적을 두고 알고리즘을 집대성하여 지속적으로 버전을 업그레이드 하고 있으며, 버전과 운용시스템을 고려하여 대표 웹사이트(<https://www.cs.waikato.ac.nz/ml/weka>)에서 프로그램을 서비스하고 사용매뉴얼을 제공하고 있다. 현재 WEKA 3.8은 안정적인 버전으로 이전버전과의 호환성을 유지한 상태로 버그나 기능 업그레이드가 관리되고 있으며, WEKA 3.9는 개발버전으로 호환성에 구애받지 않고 새로운 기능의 패키지를 개발, 추가하는 실험단계의 버전이다.

WEKA 3.8 GUI 인터페이스는 데이터 탐색을 위한 Explorer, 성능비교를 위한 Experimenter, 모델링 시뮬레이션을 위한 Knowledge Flow, 테스트 명령어 인식기 Simple CLI가 있다. 그리고 WEKA Workbench는 머신러닝 알고리즘과 데이터 전처리 도구로 입력 데이터 준비, 학습구조(learning scheme)의 통계적 평가, 입력 데이터 및 학습 결과 시각화 등 실험 데이터 마이닝 프로세스를 지원한다. 또한 다양하고 광범위한 알고리즘이 공통의 인터페이스를 통해 접근할 수 있도록 지원하여 사용자가 다양한 방법의 분석을 비교하여 현재 문제에 가장 적합한 모델을 식별할 수 있도록 돕는다(Frank, Hall, & Witten, 2016).

WEKA를 활용하는 방법으로는 주요하게 세 가지를 꼽을 수 있다. 첫째, 학습 방법을 데이터 세트에 적용하고 그 결과 분석 통하여 데이터를 학습하게 하는 방법이 있다. 둘째, 학습된 모델을 사용하여 새로운 사례에 대한 예측을 만들어내는 방법이다. 셋째, 다른 여러 학습을 적용하고 실행결과를 비교하여 하나의 예측을 선택하는 방법이다. 이러한 데이터 세트에 학습 알고리즘을 적용하는 방법을 WEKA에서는 프로그램 코드를 직접 작성하지 않아도 쉽게 대화식 인터페이스와 메뉴에서 원하는 학습방법을 선택하여 적용할 수 있도록 지원한다. WEKA는 학습 구조를 실제 실행시키는 가장 핵심의 기능 이외에도, 필터(filter)와 같은 데이터 전처리 기능, 정의된 변수를 조정(property sheet)하고 객체를 수정(object editor)하는 메뉴, 데이터 시각화 기능을 포함한다(Frank, Hall, & Witten, 2016).

라. 소결

본 연구에서는 머신러닝을 적용하여 예측모형을 탐색하는 데 WEKA 3.8을 활용하였다. 축적된 데이터 변인들 간의 관계를 학습하여 예측모형을 만드는데 WEKA에서 지원하는 13개 머신러닝 알고리즘을 적용하였으며 알고리즘의 주요 기능 및 특징은 <표Ⅱ-9>에 설명한다.

<표 II-9> WEKA 지원 알고리즘의 주요 기능 및 특징⁸⁾

알고리즘	주요기능 및 특징
Bayes Net	<ul style="list-style-type: none"> • 학습한 변인 간의 확률 관계를 시각화하여 표현이 가능하며, 노드와 화살표로 예측모델의 내부 구조를 보여준다. • 결과가 이해하기 쉬운 형태로 제공되어 불확실성이 높은 데이터 세트의 이해나 예측모형 개발에 유용하다. • 텍스트 마이닝, 자연어 처리, 음성인식, 신호처리, 생물정보학, 오류제어 코드, 의료진단, 일기예보 등 다양한 분야에 적용된다.
Decision Table	<ul style="list-style-type: none"> • 의사결정트리나 신경망과 같이 예측에 사용되는 모델로 상위 테이블의 각 항목이 추가 속성 쌍의 값으로 분류되어 다른 테이블을 구성하는 계층구조 테이블을 형성한다. • 다양한 속성을 포함한 모델을 시각화하여 이해를 쉽게 돕는 특성이 있다(Becker, 1998).
IBk	<ul style="list-style-type: none"> • k-Nearest-Neighbor(kNN) 알고리즘과 동일하며 분류 대상의 데이터가 이미 분류된 데이터와의 거리에서 가장 가까운 k개 분류에서 가장 출현빈도가 높은 범주로 분류시킨다. • 데이터 분포에 대한 사전 지식이 거의 없거나 전혀 없는 경우에도 적용할 수 있는 가장 단순한 분류기법이다. • 금융 분야 예측, 의료분야 위험요소 식별 등에 활용된다.
J48	<ul style="list-style-type: none"> • 의사결정트리(decision tree) 알고리즘으로 노드(node)와 가지(link)로 데이터 분할 과정을 규칙에 의거하여 나무구조와 같이 보여준다. • 예측과정이 나무구조에 의한 추론 규칙에 의해서 표현되기 때문에 다른 머신러닝 알고리즘에 비해서 연구자가 예측 과정을 쉽게 이해하고 설명할 수 있다. • 특정 요인을 규명하거나 의료분야와 같이 의사결정 가이드라인 관리 등에 사용된다.
KStar	<ul style="list-style-type: none"> • 가까운 이웃을 구현하는 kNN 알고리즘 방식과 유사하나 데이터 간 거리를 엔트로피로 측정하는 차이점이 있다. • KStar 알고리즘은 엔트로피 측정을 활용하여 분류값이 변하는 확률을 계산한다. • 의학 분야 및 정보보호 분야에서 활용된다.
LMT	<ul style="list-style-type: none"> • Logistic Model Trees는 leaf에 회귀함수(regression function)가 있는 회귀 트리의 형태이다. • C4.5, CART와 비교했을 때 보다 정확한 예측이 가능하다 보고된 바 있다(Landwehr, Hall and Frank, 2004). • 이진 및 다중 클래스 대상 변수, 그리고 숫자 및 명목 속성과 결측값 처리가 가능하다(Landwehr, Hall, & Frank, 2005).
Multilayer Perceptron	<ul style="list-style-type: none"> • 인공신경망 기법의 한 종류로 독립변인과 종속변인 간 잠재적 비선형 관계를 탐색하는 알고리즘이다.

8) 「이종현(2020). 머신러닝을 활용한 중학교 수학 기초 미달 비율 예측모형 탐색. 영남대학교 대학원 박사학위논문」에 수록된 머신러닝의 주요 알고리즘 설명을 참조하여 종합, 정리하였음.

알고리즘	주요기능 및 특징
(MLP)	<ul style="list-style-type: none"> 신경망 기법은 데이터마이닝 분야에서 폭넓게 활용되는 분류 기법 중 하나로 인간의 뇌세포인 뉴런이 정보를 전달하는 과정에서 다른 뉴런과 교류하고 의사결정을 내리는 방식을 반영한 알고리즘이다.
Naive Bayesian	<ul style="list-style-type: none"> 분류값 속성인 class에 독립된 속성을 연결한 단순한 분석 구조로 처리 소요시간이 빠르고 효율적이다. 규모가 큰 데이터 처리 분야에 유용하며 텍스트 마이닝, 메일이나 문서 분류 등에 실제 많이 적용되고 있다.
Part (Partial Decision Tree)	<ul style="list-style-type: none"> 일련의 규칙에 따라 학습 데이터를 삭제하는 RIPPER 알고리즘과 학습데이터를 분류하는 규칙 생성 의사결정을 위해 C4.5 알고리즘을 결합하였다. 최적화되지 않은 의사결정트리를 먼저 생성한 다음 일련의 규칙을 생성하는 방식이다. 단일 규칙을 생성하기 위해 부분적으로 의사결정트리를 생성하여 경쟁시킨 다음 가장 적용범위가 큰 규칙만을 남기고 나머지는 폐기한다. 게임, 공장 자동화, 전문가시스템 등에 적용된다.
Random Forest	<ul style="list-style-type: none"> 의사결정트리 과정 중 적합(overfitting)으로 인한 오류를 없애고 학습데이터로부터 생성한 분류 가운데 가장 신뢰도가 높은 안정적인 예측모형을 만들어낸다. 독립변수가 많고 상대적으로 표본 수가 적은 경우 예측력이 높다고 평가되고 있으며 대용량 분석 시 속도나 정확도 측면에서도 우수하게 평가되고 있다. 대용량 데이터를 분석하는 바이오 정보학 분야에서 최근 많이 활용되고 있다.
Simple Logistics	<ul style="list-style-type: none"> 선형 로지스틱 회귀 모형(logistic regression)에 기반을 둔 알고리즘으로 독립변수와 종속변수의 선형적 예측에 사용되어 일차함수의 형태로 표현된다. 선형 로지스틱 회귀 알고리즘은 분류 값이 두 가지인 경우 활용된다.
SMO	<ul style="list-style-type: none"> Sequential Minimal Optimization은 서포트 벡터 머신 패턴 분류 방식으로 두개의 범주로부터 최적의 결정면(decision surface)을 찾아 데이터를 분류한다. 서포트 벡터 머신은 서로 다른 범주에 속하는 데이터들 간의 거리를 최대화 할 수 있는 분류기를 구현하는 것을 목적으로 사용된다. 문자, 음성, 문서, 이미지의 패턴 분석이 필요한 분야에 활용된다.
ZeroR	<ul style="list-style-type: none"> 종속변인과 무관하게 분류가 용이한 변인의 특징을 기준으로 데이터를 예측한다. 종속변인과 다른 분류 방법 간의 성능 기준치를 결정하는데 활용된다.

Ⅲ. 머신러닝 기반 특성화고 취업률 예측 모델 개발

1. 개요

본 연구의 목적 달성을 위해 머신러닝 기반의 특성화고 취업률 예측 모델을 개발하고 그 성능을 분석하였다. 예측 모델 개발은 연구모형 설계, 분석 대상 데이터 수집, 데이터 전처리, 기술통계 분석, 예측모형 개발, 데이터 마이닝의 과정을 통해 진행하였고, 성능 분석은 예측모형별 예측 유효성 검증을 통해 진행하였다. 소결에서는 연구결과를 종합적으로 분석하고 본 연구의 한계점과 향후 연구 방향을 제시하였다.

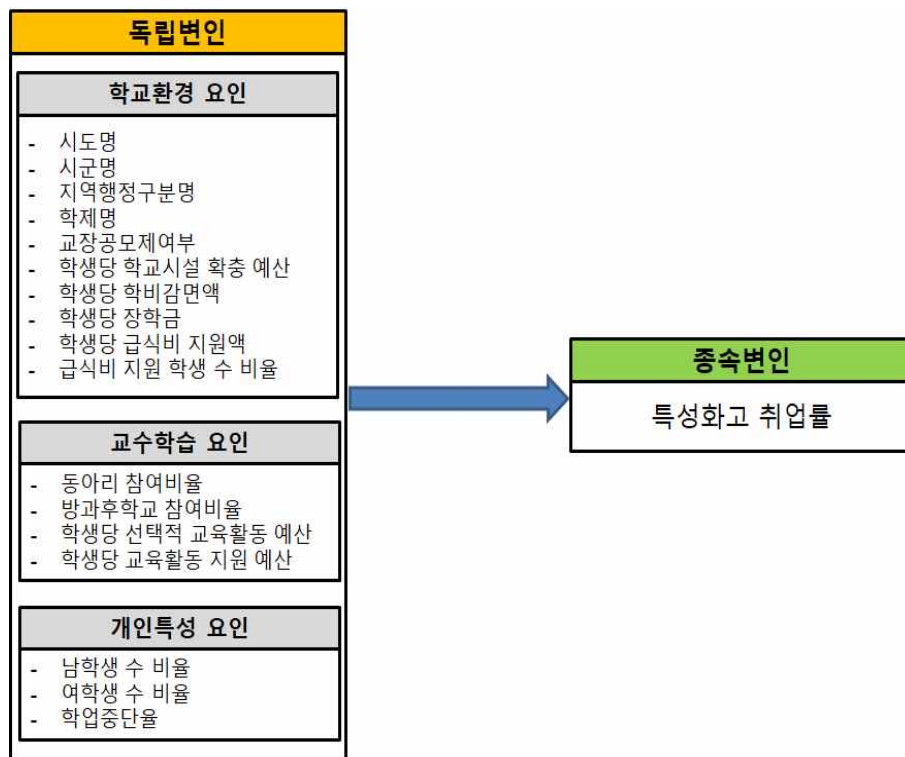
2. 분석 방법 및 절차

머신러닝 기반의 특성화고 취업률 예측 모델 개발을 위해 6개 과정을 진행하였다. 첫 번째 과정은 독립변인과 종속변인 도출 및 연구모형 설계이고, 두 번째 과정은 설계된 연구모형에 적합한 분석 데이터의 수집이었다. 세 번째 과정은 연구목적 달성을 위해 데이터를 전처리 하는 것이었다. 네 번째 과정은 기술통계 분석을 통해 특성화고 취업률 수준 판단을 위한 기준을 도출하는 것이었고, 다섯 번째는 머신러닝 알고리즘을 활용해 예측모델을 개발하는 것이었다. 마지막 여섯 번째 과정은 머신러닝 알고리즘 기반의 데이터 마이닝 기술을 활용하여 독립변인이 종속변인에 미치는 영향력을 측정하는 것이었다.

가. 연구모형 설계

특성화고 관련 배경 문헌 분석 결과를 토대로 다음 [그림 Ⅲ-1]과 같은 연구모형을 도출하였다. 독립변인은 학교환경 요인, 교수학습 요인, 개인특성 요인

으로 구분하여 총 17개 세부요인으로 구성하였다. 학교환경 요인에는 학제명, 학생당 학교시설 확충 예산 등 10개 세부요인이 포함되었고, 교수학습 요인에는 동아리 참여비율, 방과후학교 참여비율 등 4개 세부요인이 포함되었다. 개인특성 요인에는 남, 여학생 수 비율과 학업중단율이 포함되었다. 독립변인에 의해 영향을 받는 종속변인은 특성화고 취업률로 구성하였다.



[그림 III-1] 연구모형

나. 분석 대상 데이터 수집

연구모형을 개발하고 검증하기 위해 학교 수준의 분석 대상 데이터를 수집하였다. 데이터 수집을 위해 교육부와 한국교육학술정보원이 운영하는 에듀데이터서비스(EduData Service System: EDSS)을 활용하였다. 에듀데이터서비스

는 교육관련기관의 정보공개에 관한 특례법과 동법 시행령에 근거하여 교육부, 시·도교육청, 교육유관기관 등에 축적된 교육 관련 데이터를 수집·연계·가공하는 체제로 연구자나 기관이 학술연구 목적으로 자료를 요청한 경우 소정의 심사 절차를 거쳐 자료를 제공하고 있다. 구체적으로 에듀데이터서비스는 초·중·고 교육데이터로 학교정보공시, 초·중·고 교육통계, 대학수학능력시험자료, 국가학업성취도평가자료, 특수교육통계 등 7개 분야, 고·등·교육통계로 대학정보공시, 고·등·교육통계, 취업통계, 평생교육통계 등 11개 분야, 총 18개 분야의 데이터 셋(data set)을 제공하고 있다.

본 연구의 분석 대상 데이터 수집을 위해 에듀데이터서비스를 이용하였고, 그 결과 2009년부터 2018년까지 10년간의 고등학교 학교정보공시 데이터와 학교회계 예산 공시 데이터 2천여만 건을 확보하였다. 구체적으로 학교정보공시 데이터 15,411,240건(17,355행, 888열)과 학교회계 공시자료 5,307,430건을 확보하였다. 학교회계 공시자료는 다음 <표 III-1>과 같이 연도별로 규모가 조금씩 달랐다.

<표 III-1> 고등학교 학교회계 예산 공시 데이터('09년~'18년)

연도 구분	행(row)	열(column)	연도 구분	행(row)	열(column)
2009	1,716	259	2014	1,739	214
2010	1,739	498	2015	1,739	224
2011	1,739	500	2016	1,739	228
2012	1,739	466	2017	1,736	228
2013	1,739	212	2018	1,730	228

다. 데이터 전처리

수집된 전체 데이터 2천여만 건을 대상으로 머신러닝 분석이 가능하도록 전처리(preprocess) 과정을 진행하였다. 전처리 과정은 마이크로소프트의 엑셀 프로그램을 활용하였다. 총 8단계의 전처리 과정을 거쳐 분석대상 최종 데이터

67,809건(3,229개 행, 57개 열)을 최종 추출하였다. 전처리 과정의 주요 내용과 데이터 규모 변화 과정은 다음 <표 III-2>와 같다.

<표 III-2> 수집 데이터 전처리 주요 과정

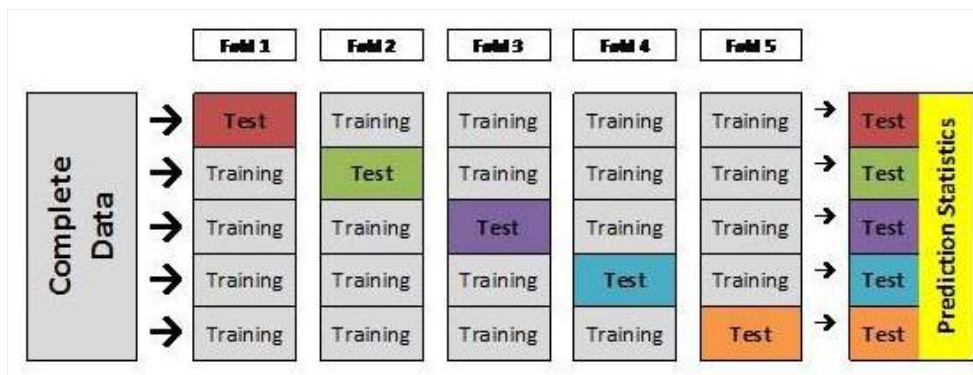
단계구분	주요 내용	데이터 규모 변동 내역	비고
1단계	학교정보공시 데이터를 기본으로 전처리 시작	17,355개 행, 888개 열	
2단계	특성화고 및 마이스터 고등학교 데이터 추출	5,047개 행, 888개 열	
3단계	분석대상 요인 항목 추출	5,047개 행, 54개 열	
4단계	학교시설 확충 예산, 선택적 교육활동 예산, 교육활동 지원 예산 결합이 가능한 데이터 항목 추출	4,504개 행, 54개 열	2009년 학교회계 공시자료 대상 제외
5단계	학교시설 확충 예산, 선택적 교육활동 예산, 교육활동 지원 예산 결합	4,504개 행, 57개 열	
6단계	결측치 제외	3,872개 행, 57개 열	
7단계	마이스터 고등학교 데이터 제외	3,229개 행, 57개 열	
8단계	데이터 항목 정제	3,229개 행, 21개 열	

라. 기술통계 분석

예측모형 개발과 데이터 마이닝 분석을 위해 특성화고 취업률 수준을 판단할 수 있는 지표의 도출이 선행되어야 하였다. 이를 위해 IBM SPSS 프로그램과 마이크로소프트의 엑셀 프로그램을 활용하여 전처리된 데이터를 분석하였다. 구체적으로 특성화고 취업률 하위 25% 기준 값, 중위 값, 상위 25% 기준 값을 도출하였고, 도출된 값을 기준으로 특성화고 취업률 수준을 구분하였다.

마. 예측모형 개발

본 연구의 예측모형 개발을 위해 뉴질랜드 와이카토(waikato) 대학의 weka 3.8 버전을 활용하였다. 전처리 완료된 데이터를 분석하기 위해 weka가 제공하는 머신러닝 알고리즘 중 13개를 활용하였다. 13개 머신러닝 알고리즘은 Bayes Net, Decision Table, IBk, J48, KStar, LMT, Multilayer Perceptron, Naive Bayesian, Part, Random Forest, Simple Logistics, SMO, ZeroR이었다. 머신러닝 알고리즘 성능평가는 K-Fold 교차 유효성 검증(cross-validation) 방법론을 활용하였다. K-Fold 방법론은 가장 널리 활용되는 예측 유효성 검증 방식 중 하나로서 관측 데이터를 동일한 크기의 K개로 분할하여, 분할된 1개 영역과 나머지 K-1 영역을 학습과 평가에 반복적으로 활용하는 방법이다 ((Hastie et al., 2013).



[그림 III-2] K-Fold 교차 유효성 검증 방법론

바. 데이터 마이닝

독립변인이 종속변인에 영향을 미치는 정도를 측정하기 위해, 특성화고 취업률을 기준으로 정보획득(information gain: IG) 지수를 측정하였다. 이 과정은 weka의 데이터 마이닝 알고리즘에 전처리 완료된 67,809건의 데이터를 입력하여 결과를 도출하였다.

3. 분석 결과

가. 기술통계 분석

1) 주요 변인별 기술통계량 분석

예측모형 개발과 데이터 마이닝 분석을 위해 변인에 대한 기술통계량 분석을 진행하였다. 자세한 내용은 다음 <표 III-3>과 같다. 주요 결과로는 남학생 수 비율 평균 54.77%, 학업중단을 평균 3.66%, 동아리 참여비율 평균 92.39%, 방과후학교 참여비율 평균 45.18%, 급식비 지원 학생 수 비율 평균 45.6%, 졸업생 취업률 평균 40.72%였다. 한편 예측모형 개발 및 데이터 마이닝을 위한 졸업생 취업률 기준값은 하위 25%는 27.69%, 중위값은 41.13%, 상위값은 52.94%였다

<표 III-3> 기술통계량 분석

구분	n	최소값	최대값	평균	표준편차	왜도		첨도	
	통계량	통계량	통계량	통계량	통계량	통계량	표준오차	통계량	표준오차
지역 행정구분명*	3,229	0	2	1.04	.83	-.08	.04	-1.55	.09
학제명**	3,229	51	58	52.88	1.52	1.75	.04	2.59	.09
교장공모제 여부***	3,229	0	1	.09	.29	2.88	.04	6.28	.09
남학생 수 비율 (단위: %)	3,229	0	100	54.77	35.2	-.34	.04	-1.18	.09
여학생 수 비율 (단위: %)	3,229	0	100	45.23	35.2	.34	.04	-1.18	.09
학업중단을 (단위: %)	3,229	0	47.37	3.66	3.53	2.68	.04	13.35	.09
학생당 장학금 (단위: 원)	3,229	0	3,030,211	266,566.4	281,260.52	1.55	.04	4.64	.09

<표 III-3> 기술통계량 분석

구분	n	최소값	최대값	평균	표준편차	왜도		첨도	
	통계량	통계량	통계량	통계량	통계량	통계량	표준 오차	통계량	표준 오차
학생당 학비 감면액 (단위: 원)	3,229	0	2,678,505	543,950.69	433,409.37	1.09	.04	.48	.09
동아리 참여비율 (단위: %)	3,229	0	1,180.54	92.39	58.11	2.31	.04	37.88	.09
방과후 학교 참여비율 (단위: %)	3,229	0	718.92	45.18	52.7	3.53	.04	26.16	.09
학생당 급식비 지원액 (단위: 원)	3,229	0	31,294,607,143	58,424,357.86	552,128,318.36	56.13	.04	3,176.4	.09
급식비 지원 학생 수 비율 (단위: %)	3,229	0	225.25	45.60	26.68	1.58	.04	3.03	.09
학생당 선택적 교육활동 예산 (단위: 원)	3,229	0	744,034,091	12,446,951.16	27,360,826.35	12.26	.04	248.1	.09
학생당 교육활동 지원 예산 (단위: 원)	3,229	0	700,000,000	14,323,064.28	28,547,823.81	10.01	.04	168.04	.09
학생당 학교시설 확충 예산 (단위: 원)	3,229	0	907,267,081	4,136,788.84	31,185,568.84	16.93	.04	358.56	.09
졸업생 취업률 (단위: %)	3,229	0	91.89	40.72	17.75	.01	.04	-.52	.09
유효수 (목록별)	3,229	-							

*(지역 구분 값) (0)읍지역, 면지역, 특수지역, (1)시, (2)특별/광역시

** 학제명 구분 값) (51)농림업고, (52)공업고, (53)상업고, (54)수산고, (55)해양고, (56)가사고, (57)실업고, (58)종합고

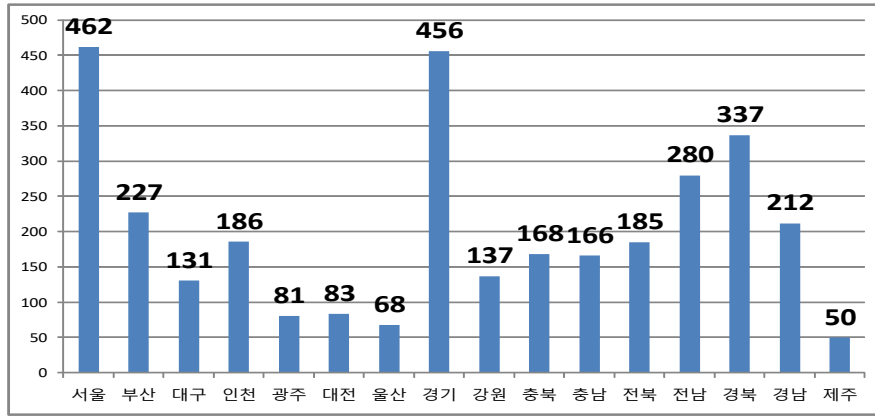
*** (교장공모제여부 구분 값) (0)해당없음, (1)해당있음

2) 시·도별 데이터 빈도분석

전처리 완료된 데이터를 시·도별로 구분하여 빈도 분석한 결과는 다음 <표 III-4>, [그림 III-3]와 같다. 서울 소재 학교의 비율이 14.3%로 가장 높았고, 서울, 경기(14.1%), 경북(10.4%) 순으로 빈도수가 높았다. 서울, 경기, 경북, 전남(8.7%), 부산(7%)의 빈도수 합은 54.5%로 5개 시·도의 데이터가 과반수를 차지하였다. 아울러 세종의 경우, 유효 데이터가 존재하지 않았다. 이는 세종특별자치시의 경우 2020년 3월 관내 처음으로 세종장영실고등학교가 개교하였기 때문이다.

<표 III-4> 시·도별 데이터 빈도분석

시·도 코드					
구분	빈도	퍼센트	유효 퍼센트	누적퍼센트	
유효	서울	462	14.3	14.3	14.3
	부산	227	7.0	7.0	21.3
	대구	131	4.1	4.1	25.4
	인천	186	5.8	5.8	31.2
	광주	81	2.5	2.5	33.7
	대전	83	2.6	2.6	36.2
	울산	68	2.1	2.1	38.3
	경기	456	14.1	14.1	52.5
	강원	137	4.2	4.2	56.7
	충북	168	5.2	5.2	61.9
	충남	166	5.1	5.1	67.0
	전북	185	5.7	5.7	72.8
	전남	280	8.7	8.7	81.4
	경북	337	10.4	10.4	91.9
	경남	212	6.6	6.6	98.5
	제주	50	1.5	1.5	100.0
	합계	3,229	100.0	100.0	



[그림 III-3] 시·도별 데이터 빈도분석 도표

앞서 시·도별 데이터 빈도분석결과를 시·군별로 세분화하여 빈도 분석한 결과는 다음 <표 III-5>와 같다.

<표 III-5> 시군코드 빈도분석

시군코드					
구분	빈도	퍼센트	유효 퍼센트	누적퍼센트	
서울	종로구	27	.8	.8	.8
	중구	27	.8	.8	1.7
	용산구	9	.3	.3	2.0
	성동구	27	.8	.8	2.8
	동대문구	18	.6	.6	3.3
	중랑구	18	.6	.6	3.9
	성북구	14	.4	.4	4.3
	강북구	9	.3	.3	4.6
	도봉구	18	.6	.6	5.2
	노원구	27	.8	.8	6.0
	은평구	45	1.4	1.4	7.4
	마포구	25	.8	.8	8.2
	양천구	18	.6	.6	8.7
	강서구	45	1.4	1.4	10.1
	구로구	36	1.1	1.1	11.2
	금천구	9	.3	.3	11.5
	영등포구	9	.3	.3	11.8
	동작구	9	.3	.3	12.1
	관악구	27	.8	.8	12.9
	서초구	9	.3	.3	13.2
	강남구	18	.6	.6	13.8
	송파구	9	.3	.3	14.0
	강동구	9	.3	.3	14.3

시군코드					
구분	빈도	퍼센트	유효 퍼센트	누적퍼센트	
부산	중구	9	.3	.3	14.6
	서구	20	.6	.6	15.2
	동구	9	.3	.3	15.5
	영도구	9	.3	.3	15.8
	부산진구	45	1.4	1.4	17.2
	동래구	9	.3	.3	17.4
	남구	36	1.1	1.1	18.6
	해운대구	18	.6	.6	19.1
	사하구	27	.8	.8	19.9
	금정구	9	.3	.3	20.2
	강서구	9	.3	.3	20.5
	연제구	9	.3	.3	20.8
	수영구	9	.3	.3	21.1
	사상구	9	.3	.3	21.3
대구	동구	31	1.0	1.0	22.3
	서구	8	.2	.2	22.5
	남구	27	.8	.8	23.4
	북구	4	.1	.1	23.5
	수성구	16	.5	.5	24.0
	달서구	35	1.1	1.1	25.1
	달성군	10	.3	.3	25.4
	중구	27	.8	.8	26.2
인천	동구	18	.6	.6	26.8
	남구	43	1.3	1.3	28.1
	연수구	27	.8	.8	29.0
	남동구	9	.3	.3	29.2
	부평구	25	.8	.8	30.0
	계양구	9	.3	.3	30.3
	서구	20	.6	.6	30.9
광주	강화군	8	.2	.2	31.2
	남구	27	.8	.8	32.0
	북구	36	1.1	1.1	33.1
	광산구	18	.6	.6	33.7
대전	동구	17	.5	.5	34.2
	중구	27	.8	.8	35.0
	유성구	30	.9	.9	36.0
울산	대덕구	9	.3	.3	36.2
	중구	9	.3	.3	36.5
	남구	18	.6	.6	37.1
	동구	14	.4	.4	37.5
	울주군	27	.8	.8	38.3

시군코드					
구분	빈도	퍼센트	유효 퍼센트	누적퍼센트	
경기	수원시	53	1.6	1.6	40.0
	성남시	17	.5	.5	40.5
	부천시	18	.6	.6	41.1
	안양시	52	1.6	1.6	42.7
	안산시	54	1.7	1.7	44.3
	용인시	17	.5	.5	44.9
	군포시	9	.3	.3	45.2
	시흥시	34	1.1	1.1	46.2
	평택시	27	.8	.8	47.0
	화성시	15	.5	.5	47.5
	오산시	9	.3	.3	47.8
	광명시	9	.3	.3	48.1
	광주시	8	.2	.2	48.3
	하남시	18	.6	.6	48.9
	김포시	6	.2	.2	49.1
	이천시	15	.5	.5	49.5
	안성시	9	.3	.3	49.8
	여주군	6	.2	.2	50.0
	고양시	27	.8	.8	50.8
	남양주시	9	.3	.3	51.1
	의정부시	9	.3	.3	51.4
	파주시	15	.5	.5	51.8
	동두천시	9	.3	.3	52.1
	양주시	9	.3	.3	52.4
	포천시	2	.1	.1	52.5
강원	춘천시	15	.5	.5	52.9
	원주시	17	.5	.5	53.5
	강릉시	18	.6	.6	54.0
	동해시	9	.3	.3	54.3
	태백시	18	.6	.6	54.8
	삼척시	9	.3	.3	55.1
	홍천군	8	.2	.2	55.4
	영월군	9	.3	.3	55.7
	정선군	9	.3	.3	55.9
	화천군	9	.3	.3	56.2
충북	고성군	16	.5	.5	56.7
	청주시	67	2.1	2.1	58.8
	충주시	18	.6	.6	59.3
	제천시	18	.6	.6	59.9
	청원군	5	.2	.2	60.0
	보은군	16	.5	.5	60.5
	영동군	17	.5	.5	61.1
	진천군	9	.3	.3	61.4
	단양군	9	.3	.3	61.6
	증평군	9	.3	.3	61.9

시군코드					
구분		빈도	퍼센트	유효 퍼센트	누적퍼센트
충남	천안시	33	1.0	1.0	62.9
	공주시	18	.6	.6	63.5
	보령시	9	.3	.3	63.8
	아산시	5	.2	.2	63.9
	서산시	9	.3	.3	64.2
	논산시	25	.8	.8	65.0
	금산군	9	.3	.3	65.3
	부여군	23	.7	.7	66.0
	서천군	9	.3	.3	66.2
	홍성군	9	.3	.3	66.5
	예산군	9	.3	.3	66.8
	당진군	8	.2	.2	67.0
전북	전주시	27	.8	.8	67.9
	군산시	15	.5	.5	68.3
	익산시	9	.3	.3	68.6
	정읍시	27	.8	.8	69.5
	남원시	17	.5	.5	70.0
	김제시	15	.5	.5	70.5
	완주군	9	.3	.3	70.7
	진안군	15	.5	.5	71.2
	장수군	9	.3	.3	71.5
	임실군	8	.2	.2	71.7
전남	고창군	9	.3	.3	72.0
	부안군	25	.8	.8	72.8
	목포시	36	1.1	1.1	73.9
	여수시	24	.7	.7	74.6
	순천시	40	1.2	1.2	75.9
	나주시	35	1.1	1.1	77.0
	광양시	9	.3	.3	77.2
	구례군	8	.2	.2	77.5
	고흥군	14	.4	.4	77.9
	보성군	18	.6	.6	78.5
	화순군	9	.3	.3	78.8
	장흥군	8	.2	.2	79.0
	강진군	9	.3	.3	79.3
	해남군	2	.1	.1	79.3
	영암군	18	.6	.6	79.9
	함평군	8	.2	.2	80.1
	영광군	17	.5	.5	80.7
	장성군	8	.2	.2	80.9
	진도군	9	.3	.3	81.2
	신안군	8	.2	.2	81.4
경북	포항시	32	1.0	1.0	82.4
	경주시	70	2.2	2.2	84.6
	김천시	27	.8	.8	85.4
	안동시	17	.5	.5	86.0

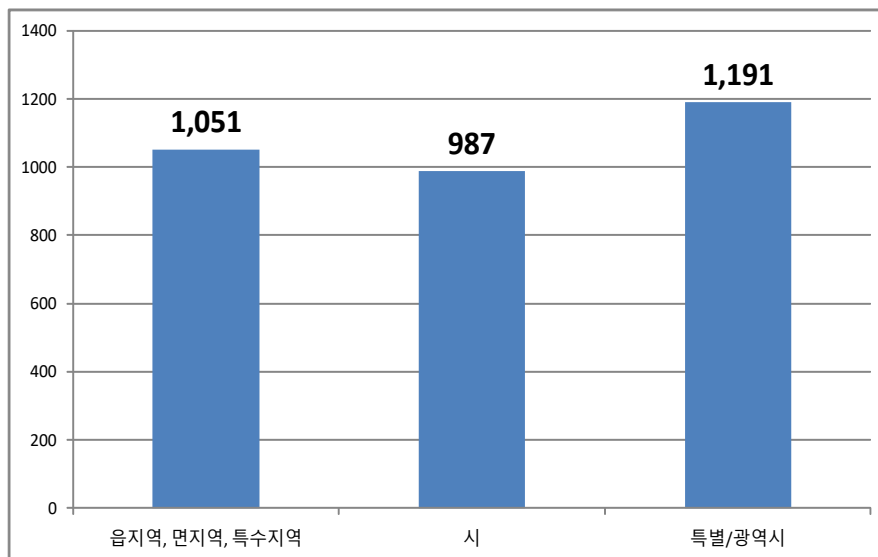
시군코드					
구분	빈도	퍼센트	유효 퍼센트	누적퍼센트	
	구미시	9	.3	.3	86.2
	영주시	18	.6	.6	86.8
	영천시	17	.5	.5	87.3
	상주시	17	.5	.5	87.9
	문경시	18	.6	.6	88.4
	경산시	18	.6	.6	89.0
	군위군	7	.2	.2	89.2
	의성군	25	.8	.8	90.0
	청송군	9	.3	.3	90.2
	영덕군	9	.3	.3	90.5
	청도군	9	.3	.3	90.8
	고령군	9	.3	.3	91.1
	봉화군	17	.5	.5	91.6
	울릉군	9	.3	.3	91.9
경남	창원시	38	1.2	1.2	93.1
	마산시	1	.0	.0	93.1
	진주시	27	.8	.8	93.9
	진해시	1	.0	.0	94.0
	사천시	14	.4	.4	94.4
	김해시	33	1.0	1.0	95.4
	밀양시	18	.6	.6	96.0
	거제시	9	.3	.3	96.3
	의령군	9	.3	.3	96.5
	함안군	6	.2	.2	96.7
	창녕군	18	.6	.6	97.3
	고성군	8	.2	.2	97.5
	산청군	8	.2	.2	97.8
	함양군	9	.3	.3	98.0
	거창군	6	.2	.2	98.2
	합천군	7	.2	.2	98.5
제주	제주시	34	1.1	1.1	99.5
	서귀포시	16	.5	.5	100.0
합계		3,229	100.0	100.0	

3) 지역행정구분명 데이터 빈도분석

전처리 완료된 데이터를 특별/광역시, 시, 읍/면/특수지역으로 구분하여 빈도 분석한 결과는 다음 <표 III-6>과 [그림 III-4]와 같다. 지역행정구분에 따른 비중은 특별/광역시 36.9%, 읍/면/특수지역 32.5%, 시 30.6% 순이었다.

<표 III-6> 지역 구분 빈도분석

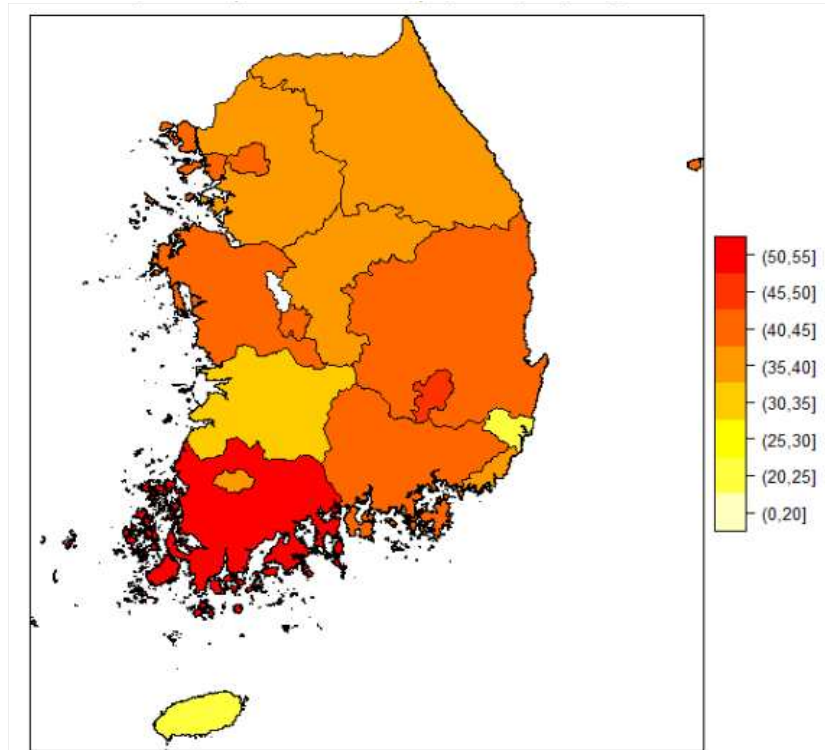
지역 구분	빈도	퍼센트	유효 비율	누적 비율
읍지역, 면지역, 특수지역	1,051	32.5	32.5	32.5
시	987	30.6	30.6	63.1
특별/광역시	1,191	36.9	36.9	100.0
합계	3,229	100.0	100.0	



[그림 III-4] 지역 구분 코드 빈도분석

4) 시·도별 특성화고 취업률 평균값 분석

한편 [그림 III-5]은 2010년부터 2018년까지 시·도별 특성화고 취업률 평균 분석결과를 시각화한 도표로서, 이를 살펴보면 전남, 대구 순으로 취업률이 높았음을 알 수 있다.



[그림 III-5] 최근 9년간 시·도별 특성화고 취업률 평균 시각화 도표

<표 III-7> 최근 9년간 시·도별 특성화고 취업률 평균값

지역	취업률 평균	지역	취업률 평균
서울	41.29	충북	38.23
대전	44.77	충남	44.73
대구	45.06	경남	44.95
광주	38.53	경북	42.81
부산	37.62	전남	52.54
울산	24.01	전북	32.89
경기	36.01	제주	24.79
인천	44.21	세종	0.00
강원	36.60	.	.

※ (참고) 세종의 경우 2020년 3월 관내 처음으로 세종장영실고등학교 개교하여 유효 데이터 없음

나. 예측모형 성능 분석

1) 기준 값 변화에 따른 예측모형 성능분석 결과

앞서 설명한대로 예측모형 개발은 Bayes Net 등 13개 알고리즘을 활용하여 개발하였고, 개발된 예측모형의 유효성 검증은 K-Fold 교차 유효성 검증(cross-validation) 방법론을 활용하였다. 예측모형별 유효성 검증은 특성화고 취업률 하위 25% 기준 값, 중위값, 상위 25% 기준 값을 활용해서 진행하였다.

첫 번째, 특성화고 취업률 하위 25% 기준 값을 활용해 개발한 13개 예측모형 유효성 검증 결과는 <표 III-8>과 같다. 예측 유효성이 가장 높은 예측모형은 Random Forest 알고리즘에 의해 개발된 것이었으며, ROC area(=AUC) 값이 0.845로 유효성이 높은 수준(AUC>0.9)에 근접하는 결과를 나타냈다. Bayes Net, Decision Table, IBk, J48, LMT, Multilayer Perceptron, Naive Bayesian, Part, Simple Logistics 알고리즘을 활용해 개발된 예측모형도 중간수준(AUC>0.7)의 예측 유효성을 나타냈다.

<표 III-8> 머신러닝 알고리즘별 예측모형 성능분석(하위 25% 기준 값 기준)

예측모형 산출 알고리즘	TP rate	FP rate	accu- racy	recall	precision	F- measure	ROC area
Bayes Net	0.714	0.323	0.714	0.714	0.762	0.729	0.748
Decision Table	0.771	0.510	0.771	0.771	0.748	0.750	0.749
IBk	0.784	0.381	0.784	0.784	0.779	0.781	0.705
J48	0.775	0.437	0.775	0.775	0.763	0.767	0.715
KStar	0.253	0.252	0.253	0.253	0.646	0.109	0.502
LMT	0.779	0.491	0.779	0.779	0.758	0.760	0.739
Multilayer Perceptron	0.778	0.453	0.778	0.778	0.762	0.766	0.742
Naive Bayesian	0.495	0.301	0.495	0.495	0.718	0.514	0.734
Part	0.767	0.582	0.767	0.767	0.737	0.729	0.750
Random Forest	0.755	0.245	0.755	0.755	0.755	0.755	0.845
Simple Logistics	0.760	0.643	0.760	0.760	0.725	0.704	0.715
SMO	0.750	0.750	0.750	0.750	(산출안됨)	(산출안됨)	0.500
ZeroR	0.750	0.750	0.750	0.750	(산출안됨)	(산출안됨)	0.499

두 번째, 특성화고 취업률 중위값을 활용해 개발한 13개 예측모형 유효성 검증 결과는 <표 III-9>과 같다. 예측 유효성이 가장 높은 예측모형은 Random Forest 알고리즘에 의해 개발된 것이었으며, ROC area(=AUC) 값이 0.848로 유효성이 높은 수준(AUC>0.9)에 근접하는 결과를 나타냈다. Bayes Net, Decision Table, IBk, J48, LMT, Multilayer Perceptron, Part, Simple Logistics 알고리즘을 활용해 개발된 예측모형도 중간수준(AUC>0.7)의 예측 유효성을 나타냈다.

<표 III-9> 머신러닝 알고리즘별 예측모형 성능분석(중위값 기준)

예측모형 산출 알고리즘	TP rate	FP rate	accu- racy	recall	precision	F- measure	ROC area
Bayes Net	0.689	0.311	0.689	0.689	0.691	0.688	0.759
Decision Table	0.700	0.300	0.700	0.700	0.700	0.700	0.774
IBk	0.723	0.277	0.723	0.723	0.723	0.722	0.722
J48	0.729	0.271	0.729	0.729	0.729	0.729	0.740
KStar	0.500	0.500	0.500	0.500	0.500	0.336	0.502
LMT	0.731	0.269	0.731	0.731	0.731	0.731	0.793
Multilayer Perceptron	0.682	0.318	0.682	0.682	0.682	0.682	0.739
Naive Bayesian	0.528	0.472	0.528	0.528	0.572	0.443	0.687
Part	0.695	0.305	0.695	0.695	0.695	0.695	0.771
Random Forest	0.763	0.237	0.763	0.763	0.763	0.763	0.848
Simple Logistics	0.665	0.335	0.665	0.665	0.666	0.664	0.709
SMO	0.660	0.340	0.660	0.660	0.660	0.660	0.660
ZeroR	0.500	0.500	0.500	0.500	(산출 안됨)	(산출 안됨)	0.498

세 번째, 특성화고 취업률 상위 25% 기준 값을 활용해 개발한 13개 예측모형 유효성 검증 결과는 <표 III-10>와 같다. 예측 유효성이 가장 높은 예측모형은 Random Forest 알고리즘에 의해 개발된 것이었으며, ROC area(=AUC) 값이 0.864로 유효성이 높은 수준(AUC>0.9)에 근접하는 결과를 나타냈다. Bayes Net, Decision Table, IBk, J48, LMT, Multilayer Perceptron, Part 알고리즘을 활용해 개발된 예측모형도 중간수준(AUC>0.7)의 예측 유효성을 나타냈다.

<표 III-10> 머신러닝 알고리즘별 예측모형 성능분석(상위 25% 기준 값 기준)

예측모형 산출 알고리즘	TP rate	FP rate	accu- racy	recall	precision	F- measure	ROC area
Bayes Net	0.702	0.279	0.702	0.702	0.774	0.721	0.792
Decision Table	0.794	0.467	0.794	0.794	0.777	0.776	0.797
IBk	0.790	0.343	0.790	0.790	0.792	0.791	0.726
J48	0.814	0.360	0.814	0.814	0.806	0.809	0.705
KStar	0.749	0.750	0.750	0.750	0.562	0.643	0.502
LMT	0.812	0.386	0.812	0.812	0.802	0.804	0.767
Multilayer Perceptron	0.762	0.460	0.762	0.762	0.749	0.754	0.772
Naive Bayesian	0.651	0.339	0.651	0.651	0.736	0.673	0.688
Part	0.797	0.478	0.797	0.797	0.781	0.777	0.794
Random Forest	0.842	0.367	0.842	0.842	0.835	0.831	0.864
Simple Logistics	0.746	0.718	0.746	0.746	0.673	0.661	0.600
SMO	0.750	0.750	0.750	0.750	(산출 안됨)	(산출 안됨)	0.500
ZeroR	0.750	0.750	0.750	0.750	(산출 안됨)	(산출 안됨)	0.498

상기 세 가지 기준 값을 활용해 개발한 총 39개 예측모형의 유효성 검증결과를 종합하면 <표 III-11>과 같다. 예측 유효성이 가장 높은 예측모형은 Random Forest 알고리즘에 의해 개발된 것이었으며, 각 기준 값에 대응하여 ROC area 값이 0.845, 0.848, 0.864의 결과를 나타냈다. 이는 세 가지 기준 값에 대해 모두 유효성이 높은 수준(AUC>0.9)에 근접하는 결과라 할 수 있다. Bayes Net, Decision Table, IBk, J48, LMT, Multilayer Perceptron, Part 알고리즘을 활용해 개발된 예측모형도 세 가지 기준 값에 대응하여 중간수준(AUC>0.7)의 예측 유효성을 나타냈다.

<표 III-11> 머신러닝 알고리즘별, 기준별 ROC area값 총괄표

예측모형 산출 알고리즘	하위 25% 기준	중위값 기준	상위 25% 기준
Bayes Net	0.748	0.759	0.792
Decision Table	0.749	0.774	0.797
IBk	0.705	0.722	0.726
J48	0.715	0.740	0.705
KStar	0.502	0.502	0.502
LMT	0.739	0.793	0.767
Multilayer Perceptron	0.742	0.739	0.772
Naive Bayesian	0.734	0.687	0.688
Part	0.750	0.771	0.794
Random Forest	0.845	0.848	0.864
Simple Logistics	0.715	0.709	0.600
SMO	0.500	0.660	0.500
ZeroR	0.499	0.498	0.498

2) 예측모형 최적화

앞서 진행한 기준 값 변화에 따른 예측모형 성능분석 결과를 검토하여 가장 예측 유효성이 높았던 예측 모형을 최적화하고 결과를 분석하였다. 예측모형 산출 알고리즘은 Random Forest 알고리즘이었으며, 하위 25% 기준 값, 중위 값, 상위 25% 기준 값에 따라 요인 구성을 달리하며 예측 유효성을 측정한 결과는 다음 <표 III-12>, <표 III-13>, <표 III-14>과 같다.

먼저 <표 III-12>을 살펴보면, 학생당 선택적 교육활동 예산을 제외하였을 경우 ROC area 값이 0.849로 가장 높았으나, 제외하기 전 결과인 0.845와 비교했을 때 약 0.5% 정도의 성능 향상만 가능하였다.

<표 III-12> Random Forest 예측모형 최적화(하위 25% 기준 값 기준)

제외 변인	TP rate	FP rate	accuracy	recall	precision	F-measure	ROC area
시도명	0.807	0.439	0.807	0.807	0.793	0.792	0.844
시군명	0.801	0.451	0.801	0.801	0.786	0.785	0.837
지역 행정구분명	0.811	0.424	0.811	0.811	0.798	0.797	0.848
학제명	0.809	0.437	0.809	0.809	0.796	0.793	0.842
교장공모제여부	0.810	0.429	0.810	0.810	0.798	0.796	0.848
남학생 수 비율	0.808	0.434	0.808	0.808	0.795	0.793	0.846
여학생 수 비율	0.810	0.431	0.810	0.810	0.797	0.796	0.846
학업중단율	0.814	0.413	0.814	0.814	0.802	0.801	0.848
학생당 장학금	0.805	0.449	0.805	0.805	0.791	0.788	0.846
학생당 학비감면액	0.813	0.415	0.801	0.813	0.801	0.800	0.848
동아리 참여비율	0.803	0.441	0.803	0.803	0.788	0.788	0.846
방과후학교 참여비율	0.809	0.429	0.809	0.809	0.795	0.795	0.844
학생당 급식비 지원액	0.809	0.429	0.809	0.809	0.795	0.795	0.844
급식비 지원 학생 수 비율	0.798	0.457	0.798	0.798	0.782	0.782	0.844
학생당 선택적 교육활동 예산	0.812	0.432	0.812	0.812	0.799	0.797	0.849
학생당 교육활동 지원 예산	0.810	0.432	0.810	0.810	0.797	0.795	0.848
학생당 학교시설 확충 예산	0.810	0.431	0.810	0.810	0.798	0.796	0.846

다음으로 <표 III-13>를 살펴보면, 남학생 수 비율, 동아리 참여비율, 학생당 선택적 교육활동 예산을 제외하였을 경우 ROC area 값이 0.849로 가장 높았으나, 제외하기 전 결과인 0.848과 비교했을 때 역시 약 0.1% 정도의 성능 향상만 가능하였다.

<표 III-13> Random Forest 예측모형 최적화(중위값 기준)

제외 변인	TP rate	FP rate	accu- racy	recall	precision	F- measure	ROC area
시도명	0.736	0.264	0.736	0.736	0.737	0.736	0.825
시군명	0.750	0.250	0.750	0.750	0.750	0.750	0.835
지역행정구분명	0.759	0.241	0.759	0.759	0.759	0.759	0.846
학제명	0.758	0.242	0.758	0.758	0.758	0.758	0.843
교장공모제여부	0.760	0.240	0.760	0.760	0.760	0.760	0.846
남학생 수 비율	0.761	0.239	0.761	0.761	0.761	0.761	0.849
여학생 수 비율	0.759	0.241	0.759	0.759	0.759	0.759	0.847
학업중단율	0.761	0.239	0.761	0.761	0.761	0.761	0.846
학생당 장학금	0.760	0.240	0.760	0.760	0.760	0.760	0.849
학생당 학비감면액	0.761	0.239	0.761	0.761	0.761	0.761	0.848
동아리 참여비율	0.761	0.239	0.761	0.761	0.761	0.761	0.849
방과후학교 참여비율	0.768	0.232	0.768	0.768	0.768	0.768	0.847
학생당 급식비 지원액	0.757	0.243	0.757	0.757	0.757	0.757	0.846
급식비 지원 학생 수 비율	0.767	0.233	0.767	0.767	0.767	0.767	0.845
학생당 선택적 교육활동 예산	0.762	0.238	0.762	0.762	0.762	0.762	0.849
학생당 교육활동 지원 예산	0.763	0.237	0.763	0.763	0.763	0.763	0.846
학생당 학교시설 확충 예산	0.756	0.244	0.756	0.756	0.756	0.756	0.845
남학생 수 비율, 동아리 참여비율	0.763	0.237	0.763	0.763	0.763	0.763	0.848
남학생 수 비율, 학생당 선택적 교육활동 예산	0.753	0.247	0.753	0.753	0.753	0.753	0.846
동아리 참여비율, 학생당 선택적 교육활동 예산	0.759	0.241	0.759	0.759	0.760	0.759	0.847
남학생 수 비율, 동아리 참여비율, 학생당 선택적 교육활동 예산	0.755	0.245	0.755	0.755	0.755	0.755	0.845

마지막으로 <표 III-14>을 살펴보면, 학생당 학비감면액을 제외하였을 경우 ROC area 값이 0.871로 가장 높았으나, 제외하기 전 결과인 0.864와 비교했을 때 역시 약 0.8% 정도의 성능 향상을 하는데 그쳤다.

<표 III-14> Random Forest 예측모형 최적화(상위 25% 기준 값 기준)

제외 변인	TP rate	FP rate	accuracy	recall	precision	F-measure	ROC area
시도명	0.839	0.381	0.839	0.839	0.832	0.827	0.865
시군명	0.836	0.394	0.836	0.836	0.829	0.823	0.855
지역행정구분명	0.844	0.368	0.844	0.844	0.838	0.833	0.866
학제명	0.840	0.376	0.840	0.840	0.832	0.828	0.863
교장공모제여부	0.845	0.366	0.845	0.845	0.839	0.834	0.865
남학생 수 비율	0.842	0.367	0.842	0.842	0.835	0.832	0.865
여학생 수 비율	0.840	0.372	0.840	0.840	0.833	0.829	0.866
학업중단율	0.841	0.366	0.841	0.841	0.833	0.830	0.865
학생당 장학금	0.844	0.365	0.844	0.844	0.837	0.833	0.865
학생당 학비감면액	0.841	0.362	0.841	0.841	0.834	0.831	0.871
동아리 참여비율	0.840	0.369	0.840	0.840	0.832	0.829	0.864
방과후 학교 참여비율	0.849	0.354	0.849	0.849	0.843	0.839	0.866
학생당 급식비 지원액	0.845	0.356	0.845	0.845	0.838	0.835	0.869
급식비 지원 학생 수 비율	0.841	0.365	0.841	0.841	0.833	0.830	0.866
학생당 선택적 교육활동 예산	0.841	0.373	0.841	0.841	0.834	0.830	0.868
학생당 교육활동 지원 예산	0.841	0.368	0.841	0.841	0.834	0.831	0.866
학생당 학교시설 확충 예산	0.839	0.375	0.839	0.839	0.832	0.828	0.868

다. 데이터 마이닝 결과 분석

1) 기준 값에 따른 변인별 영향력 분석 결과

특성화고 취업률 하위 25% 기준 값, 중위값, 상위 25% 기준 값을 기준으로 변인별 영향력을 분석하였다. 이를 위해 머신러닝 알고리즘을 활용하여 데이터 마이닝을 진행하였다. 각 기준 값에 대응하는 결과는 다음 <표 III-15>, <표 III-16>, <표 III-17>과 같다.

먼저 <표 III-15>를 살펴보면, 특성화고 취업률 하위 25% 기준 값을 초과하는 학교와 25% 기준 값 이하인 학교 간 차이를 발생시키는 주요 변인의 영향력은 학생당 교육활동 지원 예산, 학생당 선택적 교육활동 예산, 동아리 참여 비율, 학생당 장학금, 시군명 순으로 높았다.

<표 III-15> 변인별 영향력 분석(하위 25% 기준 값 기준)

연번	변인 구분	정보획득 지수 값
1	학생당 교육활동 지원 예산	0.07129
2	학생당 선택적 교육활동 예산	0.07112
3	동아리 참여비율	0.06442
4	학생당 장학금	0.04419
5	시군명	0.04391
6	학생당 학교시설 확충 예산	0.04283
7	시도명	0.04085
8	급식비 지원학생 수 비율	0.03894
9	학생당 급식비 지원액	0.03177
10	학생당 학비감면액	0.02950
11	학업중단율	0.01017
12	남학생 수 비율	0.00856
13	여학생 수 비율	0.00856
14	학제명	0.00794
15	방과후학교 참여비율	0.00577
16	교장공모제여부	0
17	지역행정구분명	0

다음으로 <표 Ⅲ-16>를 살펴보면, 특성화고 취업률 중위값을 초과하는 학교와 중위값 이하의 학교 간 차이를 발생시키는 주요 변인의 영향력은 학생당 교육활동 지원 예산, 학생당 선택적 교육활동 예산, 동아리 참여비율, 학생당 학교시설 확충 예산, 학생당 장학금 순으로 높았다.

<표 Ⅲ-16> 변인별 영향력 분석(중위값 기준)

연번	변인 구분	정보획득 지수 값
1	학생당 교육활동 지원 예산	0.1080
2	학생당 선택적 교육활동 예산	0.1019
3	동아리 참여비율	0.0893
4	학생당 학교시설 확충 예산	0.0545
5	학생당 장학금	0.0498
6	시군명	0.0495
7	학생당 학비감면액	0.0419
8	시도명	0.0340
9	급식비 지원 학생 수 비율	0.0339
10	학업중단율	0.0313
11	학생당 급식비 지원액	0.0296
12	남학생 수 비율	0.0157
13	여학생 수 비율	0.0157
14	학제명	0.0128
15	교장공모제여부	0
16	지역행정구분명	0
17	방과후학교 참여비율	0

마지막으로 <표 Ⅲ-17>을 살펴보면, 특성화고 취업률 상위 25% 기준 값을 초과하는 학교와 상위 25% 기준 값 이하의 학교 간 차이를 발생시키는 주요 변인의 영향력은 학생당 교육활동 지원 예산, 학생당 선택적 교육활동 예산, 동아리 참여비율, 학생당 학교시설 확충 예산, 시도명 순으로 높았다.

<표 III-17> 변인별 영향력 분석(상위 25% 기준 값 기준)

연번	변인 구분	정보획득 지수 값
1	학생당 교육활동 지원 예산	0.09753
2	학생당 선택적 교육활동 예산	0.09659
3	동아리 참여비율	0.08212
4	학생당 학교시설 확충 예산	0.06258
5	시도명	0.05324
6	시군명	0.05046
7	학생당 장학금	0.02971
8	학생당 급식비 지원액	0.02648
9	학생당 학비감면액	0.02488
10	학업중단율	0.02237
11	학제명	0.01399
12	급식비 지원 학생 수 비율	0.01172
13	여학생 수 비율	0.00534
14	남학생 수 비율	0.00534
15	지역행정구분명	0.00219
16	교장공모제여부	0
17	방과후학교 참여비율	0

상기 세 가지 기준 값에 따른 영향력 상위(5위 까지) 변인을 종합적으로 분석하면 다음 <표 III-18>과 같다. 기준 값에 따른 변인별 영향력 순위를 살펴보면, 1, 2, 3위는 학생당 교육활동 지원 예산, 학생당 선택적 교육활동 예산, 동아리 참여비율로 동일하였다. 하위 25% 기준 값의 경우, 학생당 장학금과 시군명이 각각 차순위를 차지하였다. 중위값의 경우는 학생당 학교시설 확충 예산, 학생당 장학금이 각각 차순위를 차지하였다. 상위 25% 기준의 경우는 학교당 학교시설 확충 예산, 시도명이 각각 차순위를 차지하였다.

<표 III-18> 기준 값에 따른 주요 변인별 영향력

변인별 영향력 순위	하위 25% 기준 값	중위값 기준	상위 25% 기준 값
1위	학생당 교육활동 지원 예산	학생당 교육활동 지원 예산	학생당 교육활동 지원 예산
2위	학생당 선택적 교육활동 예산	학생당 선택적 교육활동 예산	학생당 선택적 교육활동 예산
3위	동아리 참여비율	동아리 참여비율	동아리 참여비율
4위	학생당 장학금	학생당 학교시설 확충 예산	학생당 학교시설 확충 예산
5위	시군명	학생당 장학금	시도명

2) 영향력 상위 변인에 대한 t검정 결과

데이터 마이닝 결과 기준 값 변화에 따른 영향력 순위를 도출하였다. 도출된 결과의 타당성 검토하기 위해 기준 값을 기준으로 주요 변인별 t검정을 실시하였다. t검정은 하위 25% 기준 값, 중위값, 상위 25% 기준 값에 따라 각각 진행하였으며, 결과는 다음 <표 III-19>, <표 III-20>, <표 III-21>과 같다.

첫 번째로 <표 III-19>을 살펴보면, 취업률 하위 25% 기준 값으로 구분하였을 경우 영향력이 높은 주요 변인 중에서 t검정 결과 유의한 변인은 학생당 교육활동 지원 예산, 학생당 선택적 교육활동 예산, 동아리 참여비율이었다. 구체적으로 학생당 교육활동 지원 평균 예산은 하위 25% 기준 값 이하인 경우 7,946,243원이었고, 하위 25% 기준 값 초과인 경우 16,451,305원이었다. 각 경우의 차이는 학생당 8,505,062원이었다. 학생당 선택적 교육활동 평균 예산은 하위 25% 기준 값 이하인 경우 8,715,486원이었고, 하위 25% 기준 값 초과인 경우 13,692,314원이었다. 각 경우의 차이는 학생당 4,976,828원이었다. 동아리 평균 참여비율은 하위 25% 기준 값 이하인 경우 67.44%, 하위 25% 기준 값 초과인 경우 100.71%였다. 각 경우의 차이는 33.27%p였다.

<표 III-19> 취업률 하위 25% 기준 주요 변인별 평균값 t검정 결과

변인 구분	취업률 하위 25% 이하(n=808)	취업률 하위 25% 초과(n=2,421)	t(p)
학생당 교육활동 지원 예산***	7,946,243원	16,451,305원	9.58(0.00)*
학생당 선택적 교육활동 예산***	8,715,486원	13,692,314원	4.32(0.00)*
동아리 참여비율**	67.44%	100.71%	14.55(0.00)**
학생당 장학금	279,306원	262,315원	-1.64(0.10)*

※ 원 단위는 소숫점 이하 첫째자리에서 반올림, 퍼센트는 소숫점 이하 셋째자리에서 반올림

※ t(p)는 양측검정

※ 시군명은 시군구를 코드화한 값으로서 평균값 t검정에서는 제외함

* F-검정(단측) 결과 $p < 0.05$ 로 등분산 아님

** F-검정(단측) 결과 $p \geq 0.05$ 로 등분산

*** t검정이 $p < 0.05$ 로 유의미함

두 번째로 <표 III-20>를 살펴보면, 취업률 중위값 기준 값으로 구분하였을 경우 영향력이 높은 주요 변인 중에서 t검정 결과 유의한 변인은 학생당 교육활동 지원 예산, 학생당 선택적 교육활동 예산, 동아리 참여비율, 학생당 장학금이었다. 학생당 교육활동 지원 평균 예산은 중위값 이하인 경우 10,715,216원이었고, 중위값 초과인 경우 17,933,147원이었다. 각 경우의 차이는 학생당 7,217,931원이었다. 학생당 선택적 교육활동 평균 예산은 중위값 이하인 경우 9,554,510원이었고, 중위값 초과인 경우 15,341,184원이었다. 각 경우의 차이는 학생당 5,786,674원이었다. 동아리 평균 참여비율은 중위값 이하인 경우 70.05%, 중위값 초과인 경우 109.73%였다. 각 경우의 차이는 39.68%p였다. 학생당 평균 장학금은 중위값 이하인 경우 282,505원이었고, 중위값 초과인 경우 250,618원이었다. 각 경우의 차이는 31,887원이었다.

<표 III-20> 취업률 중위값 기준 주요 변인별 평균값 t검정 결과

변인 구분	취업률 중위값 이하(n=1,615)	취업률 중위값 초과(n=1,614)	t(p)
학생당 교육활동 지원 예산***	10,715,216원	17,933,147원	7.24(0.00)**
학생당 선택적 교육활동 예산***	9,554,510원	15,341,184원	6.04(0.00)*
동아리 참여비율***	75.05%	109.73%	17.76(0.00)**
학생당 학교시설 확충 예산	3,370,124원	4,903,929원	1.40(0.16)*
학생당 장학금***	282,505원	250,618원	-3.23(0.00)*

※ 원 단위는 소숫점 이하 첫째자리에서 반올림, 퍼센트는 소숫점 이하 셋째자리에서 반올림

※ t(p)는 양측검정

※ 시군명은 시군구를 코드화한 값으로서 평균값 t검정에서는 제외함

* F-검정(단측) 결과 $p < 0.05$ 로 등분산 아님

** F-검정(단측) 결과 $p \geq 0.05$ 로 등분산

*** t검정이 $p < 0.05$ 로 유의미함

세 번째로 <표 III-21>를 살펴보면, 취업률 상위 25% 기준 값으로 구분하였을 경우 영향력이 높은 주요 변인 중에서 t검정 결과 유효한 변인은 학생당 교육활동 지원 예산, 학생당 선택적 교육활동 예산, 동아리 참여비율, 학생당 학교시설 확충 예산이었다. 구체적으로 학생당 교육활동 지원 평균 예산은 상위 25% 기준 값 이하인 경우 12,071,170원이었고, 상위 25% 기준 값 초과인 경우 21,081,538원이었다. 각 경우의 차이는 학생당 9,010,368원이었다. 학생당 선택적 교육활동 평균 예산은 상위 25% 기준 값 이하인 경우 10,834,440원이었고, 상위 25% 기준 값 초과인 경우 17,286,483원이었다. 각 경우의 차이는 학생당 6,452,043원이었다. 동아리 평균 참여비율은 상위 25% 기준 값 이하인 경우 82.9%, 상위 25% 기준 값 초과인 경우 120.86%였다. 각 경우의 차이는 37.96%p였다. 학생당 평균 학교시설 확충 예산은 상위 25% 기준 값 이하인 경우 3,296,372원이었고, 상위 25% 기준 값 초과인 경우 6,659,081원이었다. 각 경우의 차이는 3,362,709원이었다.

<표 III-21> 취업률 상위 25% 기준 주요 변인별 평균값 t검정 결과

변인 구분	취업률 상위 25% 이하(n=2,422)	취업률 상위 25% 초과(n=807)	t(p)
학생당 교육활동 지원 예산***	12,071,170원	21,081,538원	7.84(0.00)**
학생당 선택적 교육활동 예산***	10,834,440원	17,286,483원	4.71(0.00)*
동아리 참여비율***	82.90%	120.86%	16.28(0.00)*
학생당 학교시설 확충 예산***	3,296,372원	6,659,081원	2.12(0.03)*

※ 원 단위는 소숫점 이하 첫째자리에서 반올림, 퍼센트는 소숫점 이하 셋째자리에서 반올림

※ t(p)는 양측검정

※ 시도명은 특별/광역시와 도를 코드화한 값으로서 평균값 t검정에서는 제외함

* F-검정(단측) 결과 $p < 0.05$ 로 등분산 아님

** F-검정(단측) 결과 $p \geq 0.05$ 로 등분산

*** t검정이 $p < 0.05$ 로 유의미함

4. 소결

본 연구의 목적을 달성하기 위해 머신러닝 기반의 특성화고 취업률 예측 모델을 개발하였다. 특성화고 취업률 예측 모델을 개발하기 위해 독립변인은 <표 III-22>과 같이 구성하였고, 변인별 데이터 확보는 교육부와 한국교육학술정보원이 운영하는 EDSS(에듀데이터서비스)를 활용하였다.

<표 III-22> 특성화고 취업률 예측 모델 개발에 필요한 독립변인

구분	요인 명	
학교환경 요인 (10개)	- 시도명	- 시군명
	- 지역행정구분명	- 학제명
	- 교장공모제여부	- 학생당 학교시설 확충 예산
	- 학생당 학비감면액	- 학생당 장학금
	- 학생당 급식비 지원액	- 급식비 지원 학생 수 비율
교수학습 요인 (4개)	- 동아리 참여비율	- 방과후 학교 참여비율
	- 학생당 선택적 교육활동 예산	- 학생당 교육활동 지원 예산
개인특성 요인 (3개)	- 남학생 수 비율	- 여학생 수 비율
	- 학업중단율	

개발된 예측모델은 예측 유효성 성능평가를 진행하였는데, 평가 결과를 분석해보면 특성화고 취업률 예측 모델 개발을 위해 17개 변인이 모든 유효함을 확인할 수 있었다.

예측 유효성 성능평가 결과를 구체적으로 살펴보면, 특성화고 취업률 기준을 하위 25% 기준 값, 중위값, 상위 25% 기준 값으로 구분하여 예측모형을 개발한 결과, 하위 25% 기준 값을 활용한 경우 Bayes Net, Decision Table, IBk, J48, LMT, Multilayer Perceptron, Naive Bayesian, Part, Random Forest, Simple Logistics의 10개 머신러닝 알고리즘으로부터 도출된 예측모형이 유효하였다. 중위값을 활용한 경우는 Bayes Net, Decision Table, IBk, J48, LMT, Multilayer Perceptron, Part, Random Forest, Simple Logistics의 8개 머신러닝 알고리즘으로부터 도출된 예측모형이 유효하였다. 상위 25% 기준 값을 활용한 경우는 Bayes Net, Decision Table, IBk, J48, LMT, Multilayer Perceptron, Part, Random Forest의 8개 머신러닝 알고리즘으로부터 도출된 예측모형이 유효하였다. 종합적으로 예측 유효성을 판단하면 Random Forest가 세 가지 모든 경우에서 가장 우수한 성능 결과를 보여주었다.

한편 머신러닝 기반의 데이터 마이닝 기술을 활용해 특성화고 취업률에 영향을 미치는 변인별 영향력을 분석한 결과, 특성화고 취업률 하위 25% 기준으로는 학생당 교육활동 지원 예산, 학생당 선택적 교육활동 예산, 동아리 참여비율, 학생당 장학금, 시군명 순으로 영향력이 높음을 확인하였다. 중위값 기준으로는 학생당 교육활동 지원 예산, 학생당 선택적 교육활동 예산, 동아리 참여비율, 학생당 학교시설 확충 예산, 학생당 장학금 순으로 영향력이 높음을 확인하였다. 또한 상위 25% 기준으로는 학생당 교육활동 지원 예산, 학생당 선택적 교육활동 예산, 동아리 참여비율, 학생당 학교시설 확충 예산, 시도명 순으로 영향력이 높음을 확인할 수 있었다.

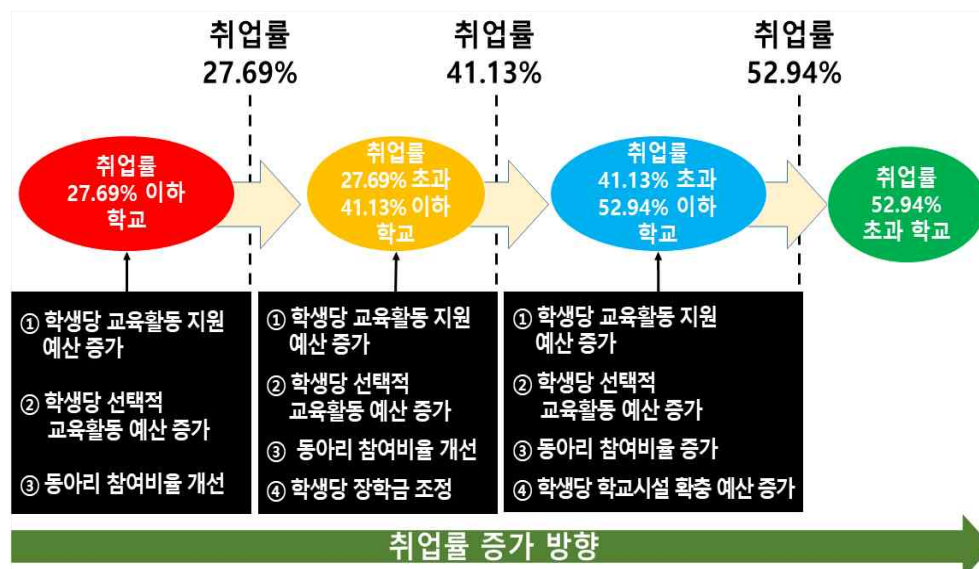
기준 값 변화에 따른 주요 영향요인을 t검정을 통해 분석한 결과는 다음과 같다. 첫 번째, 특성화고 취업률 하위 25% 기준으로 살펴보았을 때, 학생당 교육활동 지원 평균 예산은 하위 25% 기준 값 이하인 경우 7,946,243원이었고, 하위 25% 기준 값 초과인 경우 16,451,305원이었다. 각 경우의 차이는 학생당

8,505,062원이었다. 학생당 선택적 교육활동 평균 예산은 하위 25% 기준 값 이하인 경우 8,715,486원이었고, 하위 25% 기준 값 초과인 경우 13,692,314원이었다. 각 경우의 차이는 학생당 4,976,828원이었다. 동아리 평균 참여비율은 하위 25% 기준 값 이하인 경우 67.44%, 하위 25% 기준 값 초과인 경우 100.71%였다. 각 경우의 차이는 33.27%p였다. 이와 같은 결과는 현장실습 프로그램이 많을수록 취업률이 높아진다는 선행연구 결과(김강호, 2009; 김지경, 김종성, 2014; 안재영, 이병욱, 2012; 이덕현 외, 2014; 조규형, 정철영, 2014), 동아리 활동 학생 참여비율이 높을수록 취업률이 높아진다는 선행연구 결과(이쌍철, 엄문영, 2014)와 일치하였다.

두 번째, 특성화고 취업률 중위값 기준으로 살펴보았을 때, 학생당 교육활동 지원 평균 예산은 중위값 이하인 경우 10,715,216원이었고, 중위값 초과인 경우 17,933,147원이었다. 각 경우의 차이는 학생당 7,217,931원이었다. 학생당 선택적 교육활동 평균 예산은 중위값 이하인 경우 9,554,510원이었고, 중위값 초과인 경우 15,341,184원이었다. 각 경우의 차이는 학생당 5,786,674원이었다. 동아리 평균 참여비율은 중위값 이하인 경우 70.05%, 중위값 초과인 경우 109.73%였다. 각 경우의 차이는 39.68%p였다. 학생당 평균 장학금은 중위값 이하인 경우 282,505원이었고, 중위값 초과인 경우 250,618원이었다. 각 경우의 차이는 31,887원이었다. 이와 같은 결과도 현장실습 프로그램이 많을수록 취업률이 높아진다는 선행연구 결과(김강호, 2009; 김지경, 김종성, 2014; 안재영, 이병욱, 2012; 이덕현 외, 2014; 조규형, 정철영, 2014), 동아리 활동 학생 참여비율이 높을수록 취업률이 높아진다는 선행연구 결과(이쌍철, 엄문영, 2014)와 일치하였다. 그러나 학비감면 수혜 학생 비율이 높을수록 취업률이 높아진다는 선행연구 결과(이쌍철, 엄문영, 2014)와는 일치하지 않았으나, 학업성취도가 낮을수록 취업할 가능성이 높다는 선행연구 결과(김경근, 1999; 김경근, 변수용, 2006; 김기현, 2004; 김기현, 방하남, 2005; 변수용, 김경근, 2008; 채창균, 2006)와는 일치하였다.

세 번째, 특성화고 취업률 상위 25% 기준으로 살펴보았을 때, 학생당 교육활동 지원 평균 예산은 상위 25% 기준 값 이하인 경우 12,071,170원이었고, 상

위 25% 기준 값 초과인 경우 21,081,538원이었다. 각 경우의 차이는 학생당 9,010,368원이었다. 학생당 선택적 교육활동 평균 예산은 상위 25% 기준 값 이하인 경우 10,834,440원이었고, 상위 25% 기준 값 초과인 경우 17,286,483원이었다. 각 경우의 차이는 학생당 6,452,043원이었다. 동아리 평균 참여비율은 상위 25% 기준 값 이하인 경우 82.9%, 상위 25% 기준 값 초과인 경우 120.86%였다. 각 경우의 차이는 37.96%p였다. 학생당 평균 학교시설 확충 예산은 상위 25% 기준 값 이하인 경우 3,296,372원이었고, 상위 25% 기준 값 초과인 경우 6,659,081원이었다. 각 경우의 차이는 3,362,709원이었다. 이와 같은 결과도 현장실습 프로그램이 많을수록 취업률이 높아진다는 선행연구 결과(김강호, 2009; 김지경, 김종성, 2014; 안재영, 이병욱, 2012; 이덕현 외, 2014; 조규형, 정철영, 2014), 동아리 활동 학생 참여비율이 높을수록 취업률이 높아진다는 선행연구 결과(이쌍철, 엄문영, 2014)와 일치하였다. 아울러 학교시설이 좋을수록 취업에 유리하다는 선행연구 결과(이만기, 2014)와도 일치하였다.



[그림 III-6] 변인별 영향력 분석결과에 따른 취업률 수준별 처방 방안

앞서 설명한 데이터 마이닝 기술을 활용한 변인별 영향력 분석결과를 종합하여 시사점을 도출하면 [그림 III-6]와 같이 나타낼 수 있다. 즉, 취업률 27.69%(하위 25%)이하로 예측되는 학교에는 학생당 교육활동 지원 예산 및 학생당 선택적 교육활동 예산을 지원하고 동아리 참여비율의 개선을 권고할 필요가 있다.

또 취업률 27.69%(하위 25%) 초과 취업률 41.13%(중위값) 이하로 예상되는 학교에는 학생당 교육활동 지원 예산 및 학생당 선택적 교육활동 예산을 지원하고 동아리 참여비율의 개선과 학생당 장학금의 조정을 권고할 필요가 있다. 취업률 41.13%(중위값) 초과 취업률 52.94%(상위 25%) 이하로 예측되는 학교에는 학생당 교육활동 지원 예산 및 학생당 선택적 교육활동 예산, 학생당 학교시설 확충 예산을 지원하고, 동아리 참여비율의 개선을 권고할 필요가 있다.

본 연구는 학교 수준에서 취업률에 영향을 미치는 요인을 측정할 수 있는 데이터를 수집하여 예측모형을 만들고, 요인별 영향력을 분석하였다. 따라서 본 연구의 결과를 학생 개인 차원에서 처방적으로 적용하는 데에는 무리가 있다. 향후 이와 같은 한계점을 극복하기 위하여 학생 수준에서 데이터를 확보하여 계량적인 분석을 실시할 필요가 있다. 또한 본 연구는 2010년부터 2018년까지 9년간의 고등학교 학교정보공시 데이터와 학교회계 예산 공시 데이터를 분석대상으로 설정하였다. 향후 공개대상 데이터가 확대될 경우 예측 유효성 제고를 위해 추가 연구가 요구된다.

한편 본 연구는 계량적인 데이터를 토대로 예측모형을 개발하고 변인별 영향력을 분석한 만큼 연구결과를 질적인 측면의 분석에 활용하는데 한계가 있다. 예를 들어, 본 연구의 결과를 근거로 학교의 예산을 어떻게 효율적으로 집행할 수 있는지에 대한 판단이나 동아리 운영 효과성을 개선하기 위한 질적인 측면의 제언으로 활용하기에는 어려움이 따른다.

IV. 특성화고 취업률 예측 서비스 프로토타입 설계

1. 개요

본 장에서는 앞서 개발된 특성화고 취업률 예측모형과 요인별 영향력 정보를 활용하여 취업률 예측 서비스 프로토타입을 설계하고자 하였다. 이를 위해 먼저 본 서비스 사용자를 설정하였으며, III장을 통해 개발된 취업률 예측모형을 활용하여 서비스 범위를 설정하였다. 프로토타입 설계를 위해 AI 및 머신러닝을 활용한 교육 분야 유사 서비스 구축 사례를 탐색하여 서비스 설계 방향 및 세부요소를 도출하였다. 이후 교육 분야 인터페이스 전문가의 검증을 받아 특성화고 취업률 예측을 위한 최종 프로토타입(안)을 설계하였다. 서비스 프로토타입은 파워포인트를 활용하여 시각화 하였다.

2. 설계 방향 및 원리

가. 서비스 기초정보

1) 서비스 사용자

본 서비스 대상은 소속 시도 및 학교의 당해연도 취업률, 그리고 취업률 향상을 위해 현재 시점에 처방이 가능한 부분이 무엇인지 궁금해 하는 특성화고 학교 관계자와 시도교육청 담당자이다. 본 서비스 사용자는 AI, 머신러닝 관련 이해도가 낮고 배경지식이 없음을 가정한다.

2) 서비스 내용 및 범위

본 서비스에서 제공하고자 하는 내용은 III장에서 특성화고 취업률 예측 모델 개발을 통해 확인된 주요 결과이다. 사전 설정한 독립변인의 검증 결과 독

립변인 17개 모두 취업률 예측에 유효한 것으로 나타났다. 취업률 수준은 3개 구간(하위 25%, 중위, 상위 25%)으로 제시되었으며, 구간에 따라 각 요인의 영향력은 다르게 나타났다. 이를 활용, III장에서는 취업률 수준에 따라 취업률을 높이기 위한 처방방안도 제시하고 있다. 따라서 본 서비스에서 제공할 수 있는 주요 내용은 ① 개별 학교의 예측 취업률 구간정보, ② 취업률 예측에 유효한 개별 학교의 변인 값, ③ 취업률 예측 구간에 따른 요인별 영향력 등이다. 본 서비스에서 개별 학생, 혹은 시·도 단위 수준의 예측 정보는 제공하지 않는다. 또한 본 서비스 사용자는 모든 학교의 취업률을 예측해볼 수 있다.

3) 서비스 방법

본 서비스는 교육정보통계시스템 내 빅데이터 분석 지원 서비스의 취업지원 서비스로 구축·운영될 수 있다. 이 경우 취업률 예측에 필요한 데이터 수집은 DB 내에 수집·저장되어 있는 학교정보공시, 학교회계 데이터를 불러오거나 학교가 보유하고 있는 데이터를 직접 입력하는 형태로 이루어질 수 있다.

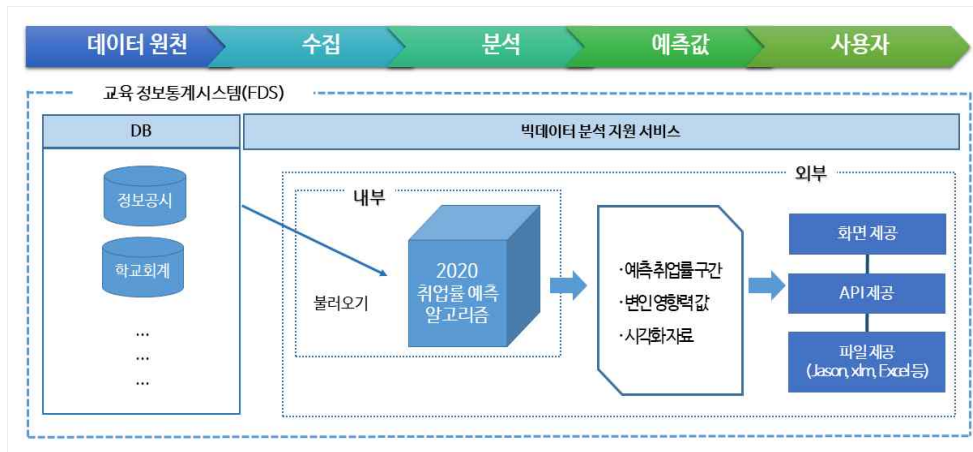
DB 내 적재되어 있는 데이터를 활용할 경우, 사용자는 학교명만 검색하면 해당 학교의 예측을 위한 변인 값을 쉽게 불러올 수 있다. 보유 데이터를 활용할 경우 사용자는 제공되는 엑셀파일을 다운받아 양식에 따라 예측에 필요한 변인 값을 입력하고, 파일형태로 업로드 할 수 있다. 본 서비스는 예측 시점에서 가장 최신(전년도) 데이터를 활용하여야 한다.

데이터 수집이 끝나면 서비스 내부의 취업률 예측 알고리즘을 통해 당해 특성화 고등학교의 취업률 예측값이 분석된다. 사용자는 서비스 화면 단에서 분석결과를 도표, 차트 등과 같은 시각화된 형태로 제공받으며 그 결과를 파일, API 등으로도 제공받을 수 있다. 그에 따른 데이터 흐름은 [그림IV-1], [그림IV-2]와 같이 표현된다.

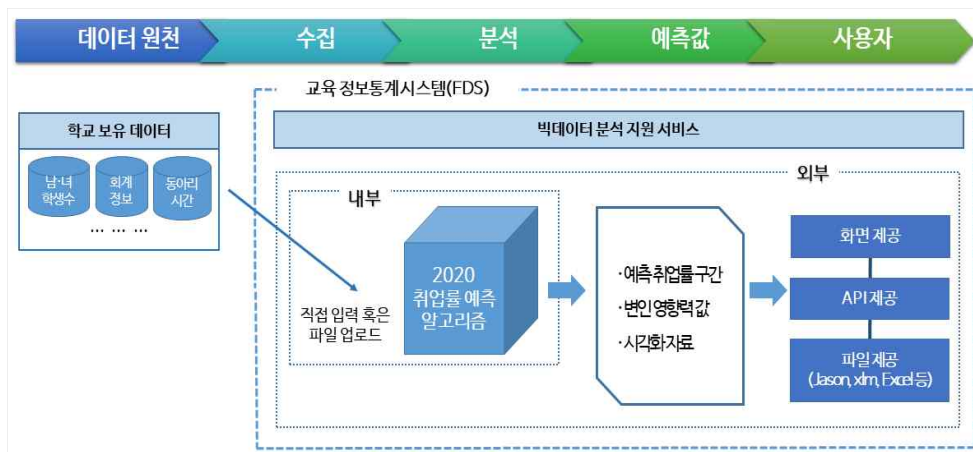
본 연구에서 활용된 예측요인의 실 데이터 값은 학교정보공시제도를 통해 수집되어, 교육정보통계시스템(EDS, Edu data service)⁹⁾에 적재 된다. 따라서

9) 교육정보통계시스템: 학교행정, 학교회계, 학교정보공시 등의 원천 시스템으로부터 데이터를 수집·저장하고 있는 통계 서비스로, 교육부와 한국교육학술정보원이 함께 운영하는 서비스

당해 연도의 예측 취업률은 전년도 학교정보공시(1차)¹⁰⁾가 이루어지는 4월 이후 제공된다. 머신러닝 모델은 매년 현행화 되며, 취업률에 영향을 미치는 주요 변인과 예측력이 높은 알고리즘도 매년 달라질 수 있다.



[그림 IV-1] 데이터 흐름도 예시 (EDS 적재 데이터 활용 시)

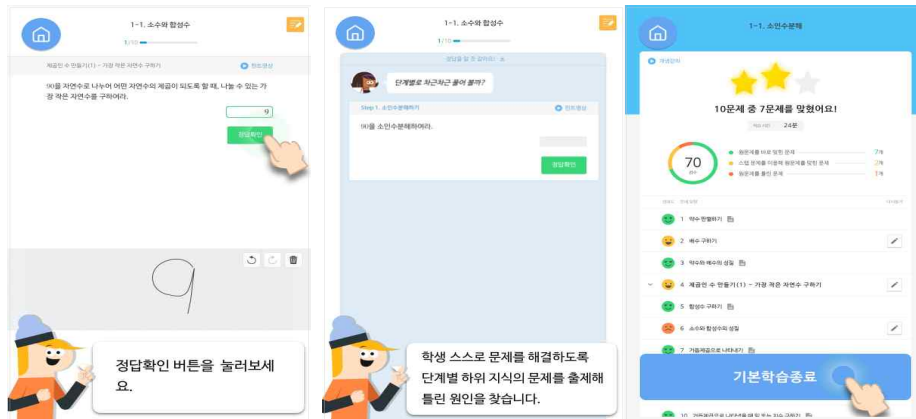


[그림 IV-2] 데이터 흐름도 예시 (학교 보유데이터 활용 시)

10) 학교정보공시: 정기 정보공시는 1년에 3번(4월, 5월, 9월) 이루어짐. (출처) 학교알리미 홈페이지 (<https://www.schoolinfo.go.kr/Main.do>)

나. 머신러닝 활용 유사 서비스 구축·운영 사례

1) 씨밋 수학



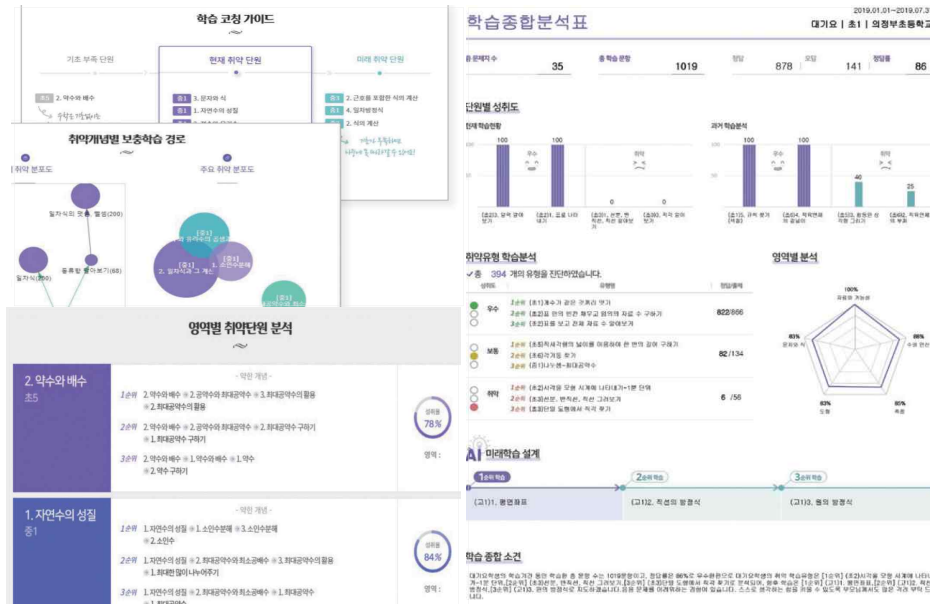
[그림 IV-3] 씨밋수학 서비스 화면

※ (이미지 출처) 씨밋수학 공식 사이트 내 '씨밋수학' 체험하기 화면 캡처 자료
<https://summit.daekyo.com/V2/math.aspx>에서 검색

씨밋수학은 AI 알고리즘을 활용한 수학 학습 플랫폼으로 대교 눈높이에서 운영한다. 총 3가지 알고리즘을 활용, 학습자가 취약한 부분의 원인을 정확히 분석하고 학습자 실력에 따라 차후 학습량과 학습기간을 예측하여 맞춤학습을 처방한다. 더불어 자기주도학습을 위한 공부역량과 성장리포트를 제공한다. 관련된 자료는 시각화 되어 제공되며, 학습자 수준과 그에 따른 처방 내용이 종합적으로 요약되어 코멘트로도 제공된다.

2) 닥터매쓰

닥터매쓰는 천재교육에서 운영하는 수학 솔루션으로, 인공지능(학습 분석을 위한 지식맵 노드 가중치 적용 알고리즘)을 활용하여 학습자의 취약점을 분석하여 제공한다. 학습자의 기초 부족 단원과 현재 취약단원을 분석하고 미래 취약단원을 예측한다. 이를 바탕으로 취약개념별 보충학습 경로, 미래학습 설계, 수준별 문제를 처방하고 학습 종합 소견을 제시한다. 관련 정보는 막대, 방사형 그래프로 시각화 되어 보고서의 형태로 제공된다.

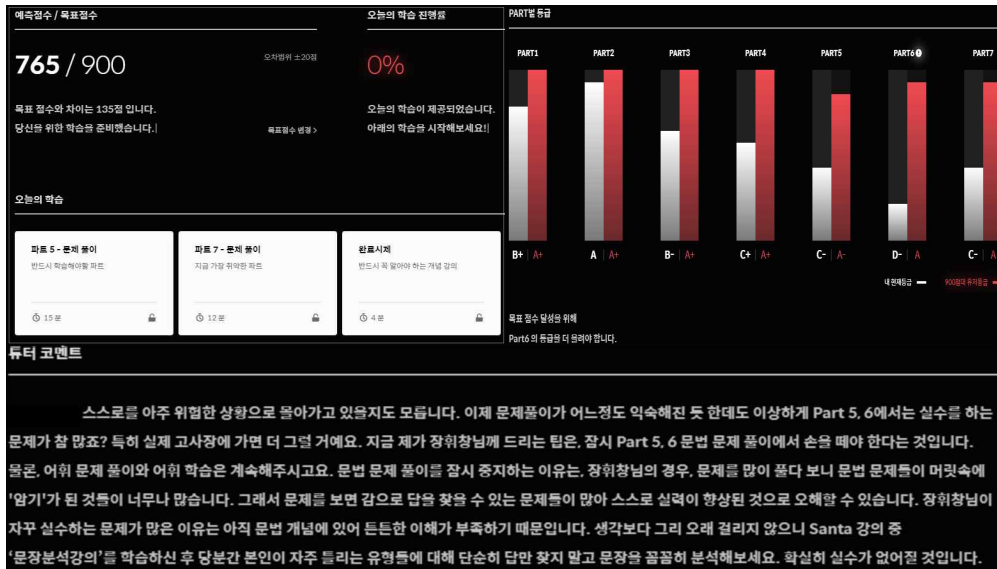


[그림 IV-4] 닥터매쓰 서비스 화면

※ (이미지 출처) 닥터매쓰(2019). AI 수학전문의 닥터매쓰 소개자료
<https://www.drmath.co.kr/>에서 검색

3) 산타토익

산타토익은 토익시험 준비를 위한 개별 맞춤형 학습문제를 어플리케이션 형태로 제공하는 토익 학습 서비스이다. 산타토익은 학습을 시작하기에 앞서, 진단평가를 시행하며 딥러닝과 인공지능망, 평가 최소화 모델을 활용하여 현재 학습자의 토익점수와 위치를 예측한다. 목표하는 토익점수를 위해 취약점을 발견하고, 어떤 부분을 더 중점적으로 학습해야하는지를 판단하여 그에 따른 맞춤형 문제와 강의를 처방한다. 또한 인공지능 튜터는 종합적 평가결과와 점수향상을 위한 전략을 코멘트 형태로 제공한다. 예측되는 모든 점수는 추이 그래프와 막대그래프로 시각화 되어 제공된다.



[그림 IV-5] 산타토익 서비스 화면

※ (이미지 출처) '산타토익' 모바일 어플 실제 화면 캡처자료

썬잇수학, 닥터매쓰, 썬타토익 서비스 사례를 바탕으로 AI 알고리즘을 활용한 서비스가 주요하게 제공하는 요소를 분석해 보았다. 세 가지 서비스의 공통점은 6가지로 요약된다.

첫째, 진단검사 데이터 혹은 이전에 축적된 데이터를 활용하여 현재 학습자의 수준을 분석하거나 예측하여 그 값을 제공한다. 둘째, 현재 수준 대비 목표점수 및 목표하는 수준에 달성하기 위해 보완해야하는 '취약점 종류'와 노력해야하는 '정도(gap) 수준'을 제공한다. 셋째, 주요 분석 값을 그래프(방사형, 막대 등)를 활용, 시각화하여 제공한다. 넷째, 현재 수준에 적합하고 취약점을 중점적으로 개선할 수 있는 맞춤형 문제, 학습경로, 학습전략을 처방한다. 다섯째, 실력과 점수 향상을 위한 종합적 분석결과를 코멘트 형태로 제공한다. 여섯째, 각 서비스가 개발·활용·적용하고 있는 알고리즘 및 예측모델 정보를 간략하게 안내한다.

다. 최종 설계 원리

Ⅲ장 연구결과를 고려하여 설정한 서비스 대상(target audience)·서비스 내용 및 범위·서비스 방법, AI·머신러닝 알고리즘을 활용한 관련 서비스 사례를 종합적으로 분석하여 도출한 취업률 예측 프로토타입의 설계 원리 및 방향은 다음과 같다.

- **(설계원리①)** 본 서비스의 사용자는 AI 머신러닝 알고리즘에 대한 배경지식이 없는 특성화고 관련 업무 담당자이다. 희망하는 학교의 정보를 DB로부터 서비스 단으로 쉽게 불러올 수 있으며, 학교가 보유하고 있는 데이터를 업로드 할 수도 있다.
- **(설계원리②)** 예측값은 시각화하여 차트, 그래프 등을 활용하여 함께 제공한다. 유사 서비스 사례를 분석한 결과 학습 데이터를 통해 분석된 내용은 주로 막대그래프와 방사형 그래프를 활용하여 시각화 되었다.
- **(설계원리③)** 목표값을 설정할 수 있도록 하며, 해당 목표값을 달성하기 위해 필요한 처방의 종류와 처방 수준을 파악할 수 있는 정보를 제공한다. 산타토익의 경우 사용자의 토익 예측점수만 제공할 뿐만 아니라 사용자로 하여금 목표 점수를 입력하게 하고 해당 내용을 바탕으로 그를 달성하기 위해 중점적으로 학습해야하는 부분이 무엇인지 제공한다.
- **(설계원리④)** 예측결과 및 처방내용을 바탕으로 예측 결과 향상을 위한 전략에 대한 종합적 코멘트를 제공한다. 개별 서비스는 예측점수 및 처방내용을 간략하게 기술하고, 종합적인 코멘트 기능을 모두 포함하고 있다.
- **(설계원리⑤)** 머신러닝 알고리즘 및 예측모델링에 대한 정보를 제공한다. 본 서비스 타겟 사용자는 개별학교 및 시도교육청 담당자로 머신러닝 및 예측모델링을 다소 낯설게 느낄 수 있다. 관련 서비스는 각 서비스가 활용하고 있는 AI알고리즘에 대해 상세 정보를 제공하는 것은 아니지만, 서비스 개요에 관련 내용을 간단하게 설명하고, 이를 활용하여 어떤 서비스를 제공하고 있는지를 설명한다. 본 서비스 설계 시 이를 반영하여 사용자의 이해도를 높일 수 있다.

3. 서비스 프로토타입(안)

가. 구성요소

설계 방향 및 원리를 바탕으로, 연구진은 서비스 구성요소와 그를 실제 서비스 화면으로 구현하기 위한 서비스 구성요소 <표 IV-1>과 같이 도출하였다. 그리고 그 역할과 기능에 따라 서비스 메뉴명을 설정하였다.

<표 IV-1> 서비스 프로토타입 구성요소(안)

설계원리	요소	구현형태	메뉴명
①	· 학교검색 창 혹은 선택 가능한 대시보드	검색창·차트	예측 결과
②	· 취업률 예측 위한 개별 학교 주요 변인 값	문구	
②	· 예측결과를 시각화(구간·방사형 그래프) 자료	그래프	
④	· 예측결과에 대한 간략한 분석결과	문구	
③	· 학교에서 조절 가능한 변인만 조정할 수 있게 함	버튼	시물 레이션
③	· 그래프를 현재수준-목표수준 차원으로 제공	그래프	
⑤	· 영향력 높은 변인 값만 조절 가능한 이유 제공	문구	

위 내용을 중심으로 교육 분야 인터페이스 전문가 및 전문가 시스템(expert system) 관련 전문가의 검증 및 자문을 진행하였다. 전문가의 공통적인 자문 결과는 아래와 같다.

첫째, 메인페이지 삽입이 필요하다. 서비스 접속 시 사용자가 가장 처음 만나게 되는 메인 페이지 필요하며 서비스 목적과 개요가 안내 제공되어야 한다.

둘째, 사용자가 취업률을 예측하고 싶은 학교를 쉽게 찾을 수 있도록 검색 기능을 제공하여야 한다. 학교 이름을 모르는 경우도 있으므로 대시보드를 통해 지역별 지도에서 학교를 선택하는 것도 고려해볼 수 있다.

셋째, 서비스 전반에 예측값에 오차가 있을 수 있다는 문구를 작성하여 사용자에게 인지도시킬 필요가 있다. 문구 예시는 ‘해당 내용은 예측일 뿐이며, 실제와 차이가 있을 수 있습니다’ 등의 내용이 될 수 있다.

넷째, 예측결과 제공 시 학교에서 조절 가능하며, 처방 가능한 변인만 시각화 하여 사용자에게 제공한다. 또한 여러 변인 중, 영향력이 높거나 관련성 있는 변인은 사용자가 보기에 쉽도록 가까이 묶어 제공할 필요가 있다.

다섯째, 예측결과 제공 시 관련 요인의 백분위, 평균, 표준편차를 함께 제공한다. 이를 통해 사용자가 타 학교와 본인 학교의 비교가 가능하도록 한다.

여섯째, ‘시뮬레이션’ 기능은 사용자가 목표 취업률을 모의로 설정하거나 취업률 예측에 활용된 변인 값을 조절하는데 목적이 있는 기능이다. 따라서 ‘시뮬레이션’ 대신 ‘모의계산’으로 표현하는 것이 더 적절이 더 적절하다.

일곱째, 대시보드 형태 뿐 아니라 보고서 형태로 리포팅 할 수 있는 기능을 제공해야 한다. 따라서 보고서 인쇄 기능을 추가할 필요가 있다.

자문결과를 바탕으로 수정된 예측 서비스 요소 및 구현방안은 <표 IV-2>과 같다. 아래 표는 사용자가 서비스에 접속하였을 때 경험하게 되는 순으로 나열하였다.

<표 IV-2> 서비스 프로토타입 최종 구성요소(안)

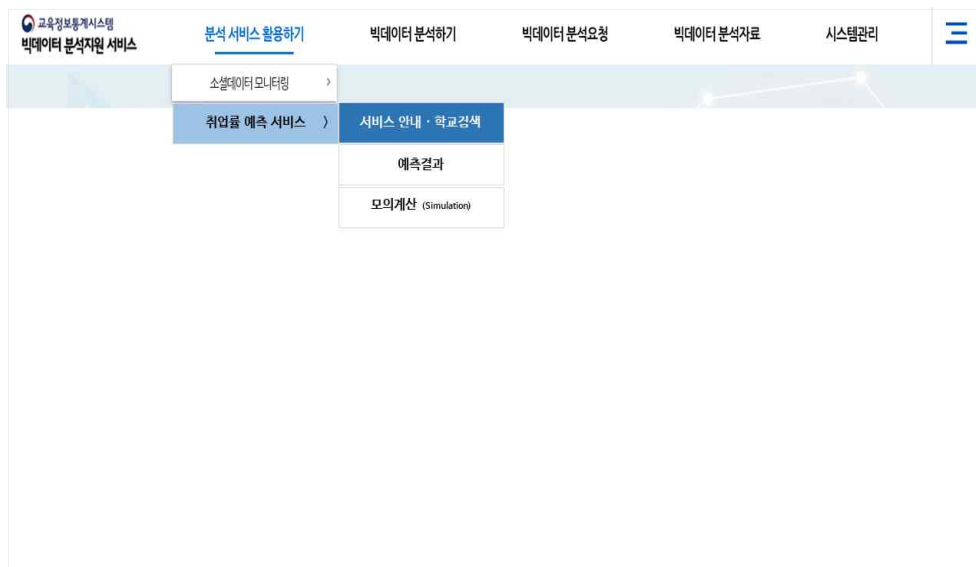
설계원리	요소	구현형태	메뉴명
⑤	· 서비스 목적 및 개요	문구	서비스 안내
①	· 학교검색 및 선택 가능한 대시보드	차트	학교검색
②	· 취업률 예측 위한 개별 학교 주요 변인 값	문구	예측 결과
②	· 관련 변인 값의 백분위, 평균, 편차	문구	
②	· 예측결과를 시각화(구간·방사형 그래프) 자료	그래프	
④	· 예측결과에 대한 간략한 분석결과 리포팅	문구	
⑤	· 예측에 활용된 알고리즘에 대한 안내	문구	
-	· 보고서 인쇄	-	
③	· 영향력 높은 변인(학교에서 조절 가능한 변인) 조정할 수 있게 함	버튼	모의 계산
③	· 학교별 취업률 구간 그래프를 현재수준-목표수준 차원으로 제공	그래프	
⑤	· 모의계산 값 오차 범위 안내	문구	
⑤	· 특정 변인 값만 조절 가능한 이유 제공	문구	

나. 화면구현

앞서 도출된 프로토타입 최종 구성요소와 메뉴 흐름에 따라 실제 화면을 구현하였다. 본 화면은 행정기관 및 공공기관의 웹사이트 표준안인 ‘전자정부 웹사이트 UI·UX(행정안전부, 2019)’와 ‘행정공공기관 웹사이트 구축운영 가이드(행정안전부, 2019)’의 내용을 참고하여 설계하였다. 여기서 설계된 프로토타입 화면은 교육정보통계시스템(EDS) 빅데이터 분석지원 서비스의 특성화고 취업지원 서비스로 구축·활용될 수 있다.

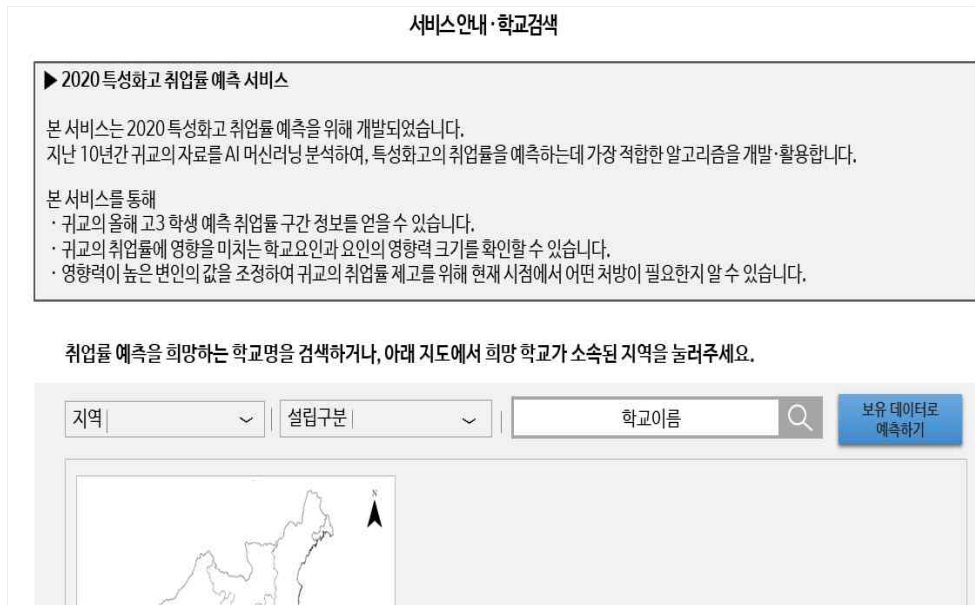
1) 메인화면 및 서비스 안내·학교검색

서비스에 사용자가 접속하게 되면 가장 먼저 만나게 되는 메뉴로 예측 서비스에 대한 간략한 안내를 받을 수 있으며, 취업률 예측을 희망하는 학교를 검색하거나 클릭 할 수 있다.



[그림 IV-6] 서비스 메인화면

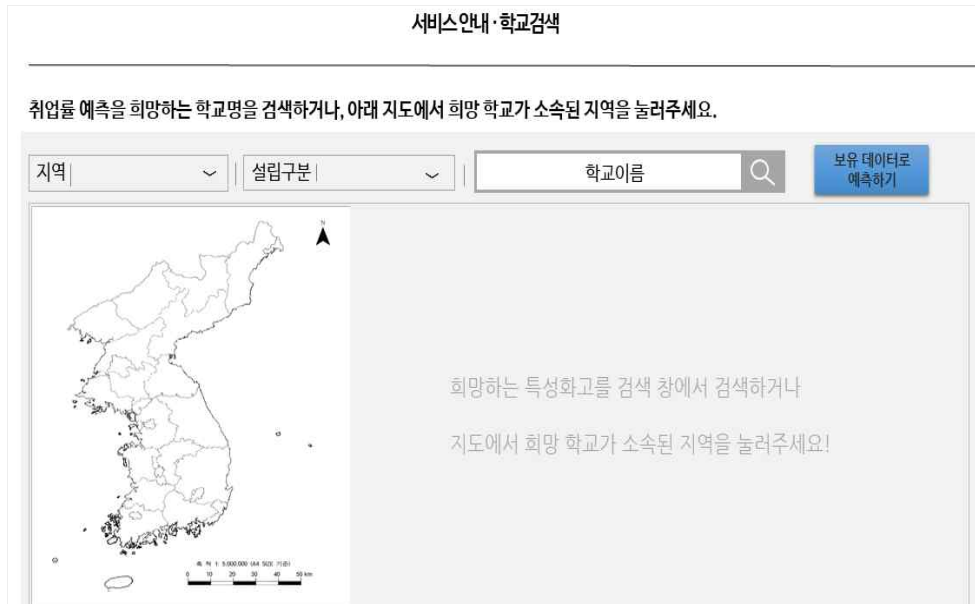
[그림 IV-6]은 서비스 메인화면의 첫 페이지다. 본 서비스의 메뉴는 크게 [서비스 안내·학교검색]-[예측결과]-[모의계산]으로 구성된다. 사용자는 첫 화면에서 본인이 희망하는 메뉴를 선택할 수 있다.



[그림 IV-7] 서비스 안내·학교검색 - 서비스 목적 및 개요

※ (지도 이미지 출처) 국토지리정보원 어린이지도여행 백지도 다운로드 서비스
<https://www.ngii.go.kr/child/content.do?sq=149>에서 검색

[그림 IV-7]은 [서비스 안내·학교검색]을 클릭하게 되는 경우 가장 먼저 만나게 되는 화면이다. 사용자는 본 페이지를 통해 서비스의 개발 목적을 알 수 있으며, 서비스를 활용함으로써 어떤 정보를 얻을 수 있는지를 파악할 수 있다. 설계원리⑤에 따라 설계되었으며, ‘서비스 목적 및 개요’ 요소를 문구로 나타내었다.



[그림 IV-8] 서비스 안내·학교검색 - 학교선택(검색)

※ (지도 이미지 출처) 국토지리정보원 어린이지도여행 백지도 다운로드 서비스
<https://www.ngii.go.kr/child/content.do?sq=149>에서 검색

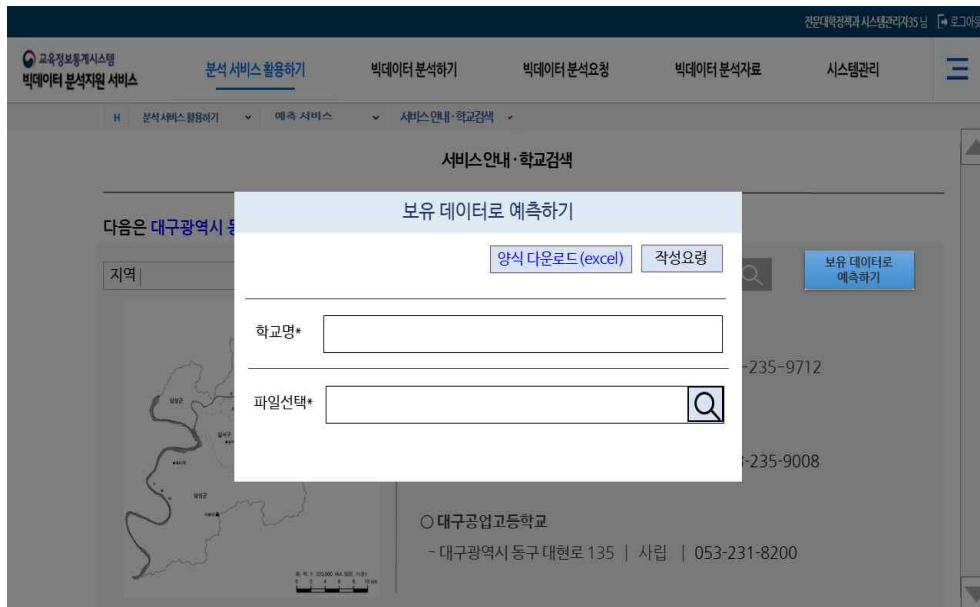
[그림 IV-8]은 취업률 예측을 희망하는 학교를 검색하거나, 지도에서 선택할 수 있는 화면이다. 사용자는 지역과 설립구분을 선택하고, 검색창을 활용하여 학교를 검색할 수 있으며, 화면 왼쪽에 있는 전국 지도를 클릭하여 학교를 선택할 수 있다. 설계원리①에 따라 설계되었으며, ‘학교검색 및 선택 가능한 대시보드’ 요소를 검색창과 차트(지도)로 나타내었다.



[그림 IV-9] 서비스 안내·학교검색 - 학교선택(지도)

※ (지도 이미지 출처) 국토지리정보원 어린이지도여행 백지도 다운로드 서비스
<https://www.ngii.go.kr/child/content.do?sq=149>에서 검색

[그림 IV-9]는 지도에서 시·도를 선택하였을 경우 산출되는 화면이다. 사용자가 시·도를 선택할 경우 해당 지역에 포함되어 있는 특성화고의 목록과 학교에 대한 간략한 정보가 정렬된다. 학교명을 클릭할 경우 그 다음 메뉴로 넘어갈 수 있다. 설계원리①에 따라 설계되었으며, ‘학교검색 및 선택 가능한 대시보드’ 요소를 검색창과 차트(지도)로 나타내었다.

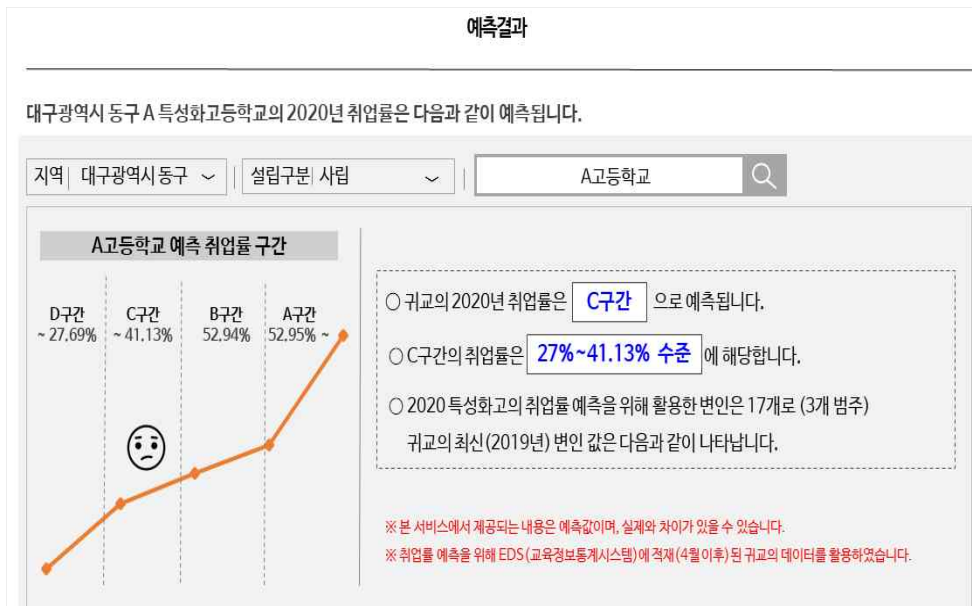


[그림 IV-10] 서비스 안내·학교검색 - 보유 데이터 업로드

[그림 IV-10]은 학교가 검색되지 않거나, DB 내 적재되어 있는 데이터 외 학교가 보유하고 있는 최신 데이터를 활용하여 예측하고 싶은 경우, [보유데이터로 예측하기] 기능을 이용하여 데이터를 업로드 할 수 있다. [보유 데이터로 예측하기] 버튼을 누를 경우 팝업창이 뜨게 된다. 사용자는 [양식다운로드] 버튼을 클릭하면 엑셀파일을 다운로드 받을 수 있으며, 해당 양식 및 작성요령에 따라 예측에 필요한 데이터를 입력할 수 있다. 이후 학교명을 수기로 입력한 후, 작성 완료한 엑셀파일을 선택하여 업로드 하면 된다. 설계원리①에 따라 설계되었다.

2) 예측결과

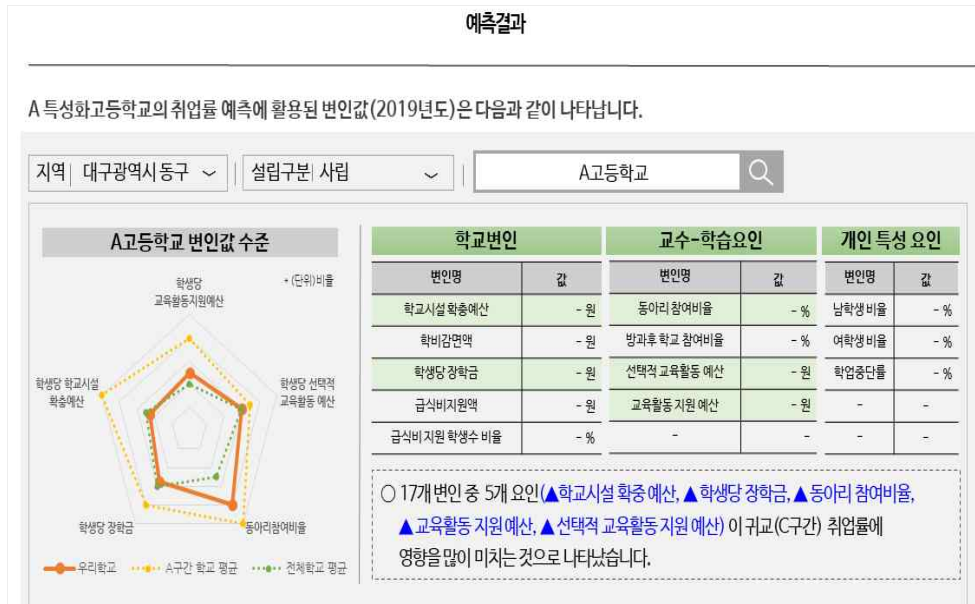
특성화 고등학교의 취업률 예측 결과를 보여주는 메뉴로, 해당 학교의 취업률 예측에 필요한 변인 값과 그의 분석결과를 바탕으로 한 예측 취업률 구간을 확인할 수 있다.



[그림 IV-11] 예측결과 - 예측결과 리포팅

[그림 IV-11]은 희망하는 학교의 취업률 예측결과를 보여주는 화면이다. 사용자가 학교 검색 혹은 보유 데이터를 업로드 하게 되면, 서비스 내부에 탑재되어 있는 취업률 예측모델(예측 분류기)을 통해 해당 특성화고등학교의 예측 취업률 구간(4개)이 판정된다.

판정결과는 왼쪽 그래프를 통해 이모티콘으로 표시되며, 오른쪽에는 그에 대한 간략한 분석결과 리포팅이 이루어진다. 리포팅 내용 아래에 본 예측에 활용된 변인값의 출처, 예측값에 오차가 있을 수 있다는 문구를 삽입하여 사용자에게 안내한다. 설계원리②, ④에 따라 설계 되었으며, ‘예측결과 시각화’, ‘예측결과 리포팅’ 요소를 그래프와 문구로 나타내었다.



[그림 IV-12] 예측결과 - 취업률 예측 변인 값

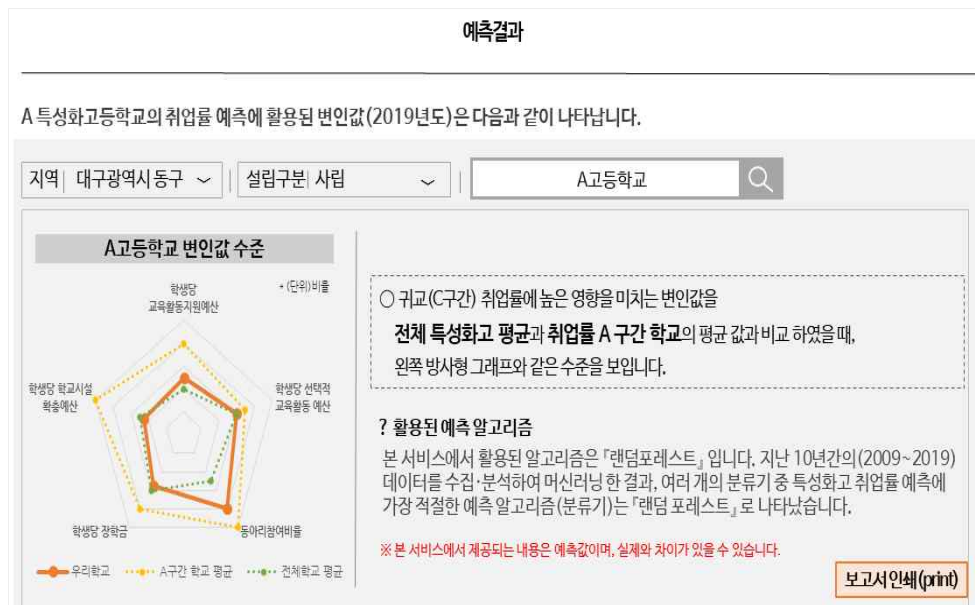
[그림 IV-12] 화면을 통해 사용자는 취업률 예측에 활용된 해당 학교의 실제 변인 값을 표로 확인할 수 있다. 학교를 검색한 경우 DB에서 불러온 데이터 값을 보여주며, 보유데이터를 입력하였을 경우 사용자가 입력한 값을 그대로 나타낸다.

예측에 활용된 변인 중, 영향력이 높게 나타난 5개 변인 값 칸의 색을 달리 하여 보여준다. 실제 예측에는 17개의 변인값이 활용되나, 서비스 화면에 ‘시도명’, ‘시군명’, ‘지역행정구분명’, ‘교장공모제여부’는 나타내지 않는다.

사용자는 영향력이 높게 나타난 변인 값을 방사형 그래프로 시각화 한 결과를 확인할 수 있다. 또 방사형 그래프를 통해 ‘전체 특성화 고등학교’와 ‘취업률 A구간’ 학교의 5개 변인(영향력 높은 변인)값 평균을 함께 확인할 수 있다. 이를 통해 본인 학교와 타 학교 간 취업률 영향 변인 값의 수준을 비교할 수 있다. 표와 그래프로 제공된 예측결과는 사용자가 이해할 수 있도록 화면 오른 쪽 공간을 활용하여 간략하게 리포팅 된다.

취업률 구간에 따라 영향력 높은 변인 값이 다르므로, 실제 서비스 구현 시

예측 취업률 구간에 따라 그래프에 활용되는 요인 값과 분석결과 리포팅 내용은 달라진다. 본 화면은 설계원리②, ④에 따라 설계 되었으며, ‘예측결과 시각화’, ‘예측결과 리포팅’ 요소를 그래프, 문구, 포로 나타내었다.

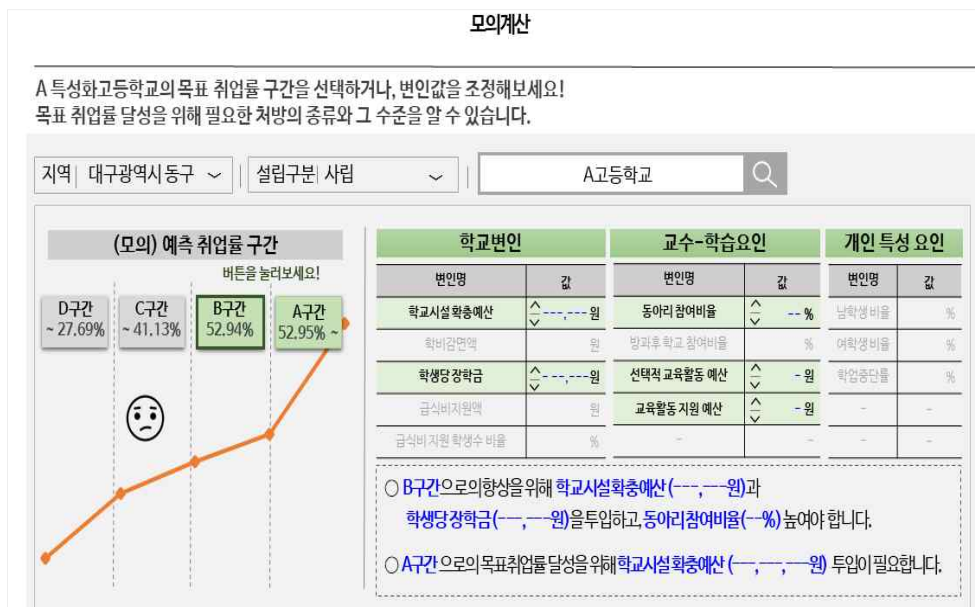


[그림 IV-13] 예측결과 - 알고리즘 안내

[그림 IV-13]는 예측에 활용된 알고리즘을 안내하는 화면이다. 주요 분석 결과 아래에 예측에 활용된 알고리즘에 대한 간략한 정보를 제공한다. [예측결과]를 통해 제공된 모든 리포팅 내용과 분석결과는 보고서 형태로 인쇄 할 수 있도록 하며, 보고서 인쇄 버튼을 제일 하단에 배치하였다. 설계원리⑤에 따라 설계되었으며, ‘예측에 활용된 알고리즘 안내’ 요소를 문구를 통해 나타내었다.

3) 모의계산

본 메뉴를 통해 사용자는 학교 취업률 예측 변인 값(현재)을 모의로 조정할 경우, 예측 취업률 구간이 어떻게 달라지는 지 확인할 수 있다. 또한 목표하는 취업률을 위해 필요한 처방의 종류와 수준도 확인할 수 있다. 모의계산은 목표하는 취업률 구간을 설정하거나, 변인값을 임의로 조정하는 형태 2가지 방법으로 진행할 수 있다.

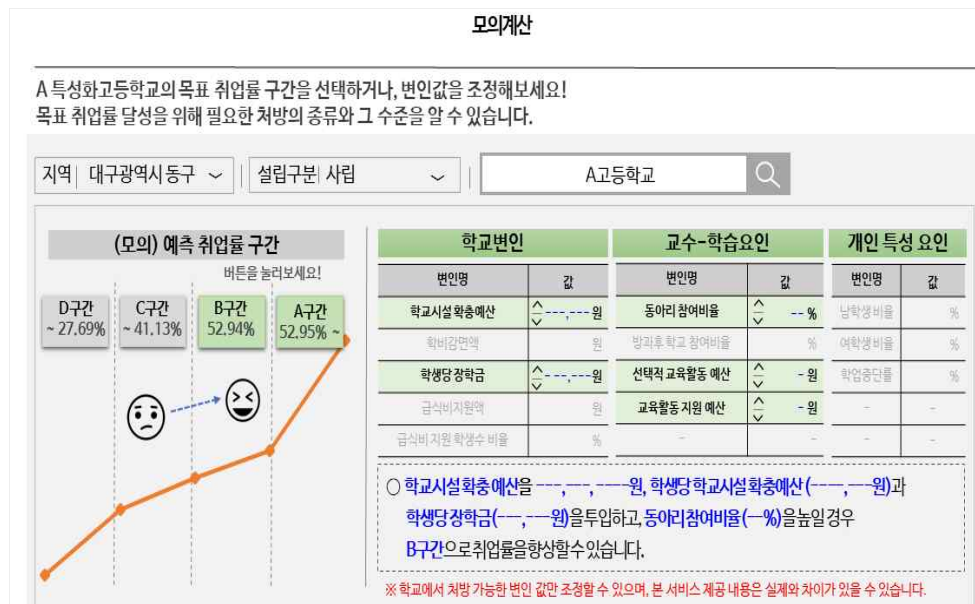


[그림 IV-14] 모의계산 - 목표수준 선택

[그림 IV-14]를 통해 사용자는 현재 예측되는 취업률 수준보다 높은 수준의 취업률 구간 버튼(목표수준)을 선택할 수 있다. 상위 구간 버튼을 누르기 전까지 [예측결과]에서 활용된 학교 변인 값이 표출되며, [예측결과]에서 예측된 취업률 구간보다 상위 구간 버튼만 활성화 된다. 버튼을 누를 경우 목표수준으로의 향상을 위해 투입되어야 하는 변인 종류를 오른쪽 표(셀)에서 확인할 수 있다. 목표 취업률에 따라 개선(투입)이 필요한 변인은 표에서 색을 달리한다.

그래프와 표 하단에 이에 대한 내용을 간략하게 리포팅 하며, 이때 목표수준

달성을 위해 투입해야하는 변인 값 수준을 계산하여 함께 제공한다. 설계원리 ③,⑤에 따라 설계되었으며, ‘목표수준 제공’ 요소를 그래프, 표, 문구를 통해 나타내었다.



[그림 IV-15] 모의계산 - 변인 값 조정

[그림 IV-15]를 통해 사용자는 표를 통해 리포팅 되고 있는 변인 값을 콤보 박스 상하 버튼을 이용하여 조정할 수 있다. 단, 학교에서 처방 가능한 변인 값만 조정 가능하도록 하며 그 외 변인 (ex.성별)은 비활성화 한다.

사용자가 조정하여 상위 구간의 취업률 수준에 해당하는 변인 값 수준을 만족할 경우, 왼쪽 그래프의 이모티콘이 상위 구간으로 옮겨가게 된다. 조정된 변인 값 종류, 조정 수준, 그 결과로 향상될 취업률 구간을 화면 하단에 리포팅 해 준다. 아울러 특정 변인만 조정할 수 있는 이유, 예측값에 오차가 있을 수 있다는 안내 문구를 화면 제일 아래에 삽입한다.

설계원리③,⑤에 따라 설계되었으며, ‘변인값 제공’, ‘모의계산 오차범위 안내’, ‘특정 변인 값만 조절 가능한 이유’ 요소를 버튼, 표, 문구를 통해 나타내었다.

4. 소결

본 장에서는 앞장에서 개발된 예측모형과 요인별 영향력 정보를 활용하여 특성화고등학교의 취업률을 예측할 수 있는 서비스의 프로토타입을 설계하였다. 프로토타입 설계를 위해 먼저 서비스 기초정보를 구체화 하였다. 서비스 주요 사용자, 서비스 내용 및 범위, 서비스 방법을 설정하였으며 서비스 흐름도를 도출하였다. 또한 교육 분야에서 기 구축·운영하고 있는 서비스를 탐색하고, AI 및 머신러닝을 활용하는 서비스들이 공통적으로 제공하고 있는 요소가 무엇인지 분석하였다. 서비스 기초정보와 관련 서비스 사례 분석을 바탕으로 특성화 고등학교 취업률 예측 서비스 프로토타입 개발을 위한 설계원리(5개)를 도출하였으며, 전문가 자문을 통해 화면 구성요소 및 UI를 설계하였다.

서비스 프로토타입은 총 3가지 메뉴(서비스 안내·학교검색 - 예측결과 - 모의계산)로 구성하였으며, PPT를 활용하여 실제 화면예시를 제공하였다. ‘서비스 안내·학교 검색’에서는 취업률 예측을 희망하는 학교를 검색하거나 클릭하여, 예측에 필요한 학교 정보와 데이터를 불러올 수 있다. ‘예측결과’ 메뉴에서는 서비스 내부에 탑재된 알고리즘을 통해 분석된, 학교의 예측 취업률을 확인할 수 있다. 또한 취업률 예측에 활용된 예측 요인 값을 확인하고, 사용자가 예측한 학교와 타 학교의 변인값 수준을 비교할 수 있다. 모의계산 서비스에서는 취업률 예측에 활용된 학교의 변인 값을 모의로 조정하면서, 예측되는 취업률 구간의 변화를 확인할 수 있다. 이를 통해 사용자는 목표 취업률을 위해 필요한 처방 종류와 수준을 확인할 수 있다.

본 장을 통해서 개발된 프로토타입은 관련 서비스 분석, 전문가 검토 등을 통해 개발된 것으로 서비스로 활용되기 위해서는 실제 사용자(학교 교사, 교육청 담당자 등)의 테스트 및 검증 과정이 필요하다.

V. 요약 및 결론

1. 주요 결과 요약

특성화고의 가장 중요한 교육 활동 중 하나는 학생들에 대한 취업 지원이다. 특성화고 학생의 취업에 대한 관심은 정부 정책을 통해서도 지속적으로 강조되어 왔다. 최근 정부는 2022년까지 특성화고 취업자 비율을 60%까지 높이려 하는 목표를 제시하는(교육부, 2019) 등 취업률 향상을 위한 정책적 노력을 기울이고 있으나, 전반적인 취업률 하락에 따른 효과적인 정책 마련의 필요성도 제기되고 있다(교육부, 2020). 본 연구는 머신러닝 기술을 활용한 특성화고 취업률 예측 모델 개발을 통해 효과적인 정책 방향 모색에 기여하기 위한 목적으로 추진되었으며, 연구의 주요 결과는 다음과 같다.

취업률 예측 요인을 도출하기 위해 특성화고 학생들의 취업 가능성, 취업 성과, 취업 진로 결정 등에 영향을 미치는 요인들을 배경 문헌을 통해 분석한 결과, 개인특성, 학교환경, 교수학습 관련 연구 요인들을 도출하였다. 도출된 요인은 본 연구에 활용한 EDSS 데이터와 매칭하여 연구 변수로 선정하였다. 개인특성 요인에서는 남녀 학생수 비율과 학업중단율을 변수로 도출하였으며, 교수학습 요인에서는 동아리 참여비율, 방과후 학교 참여비율, 학생당 선택적 교육활동 예산 등을 변수로 도출하였다. 학교환경 요인에서는 시도 및 시군명, 지역행정구역명, 학제명, 학생당 학교시설 확충 예산, 학생당 학비감면액 및 장학금, 학생당 급식비 지원액 등을 연구를 위한 변수로 도출하였다.

본 연구에서는 머신러닝을 적용하여 예측모형을 탐색하는 분석도구로 WEKA 3.8을 활용하였다. 앞서 도출한 변수간의 관계를 학습하여 예측모형을 만들기 위해 Bayes Net, Decision Table, IBk, Random Forest, SMO

등 WEKA에서 지원하는 13개의 머신러닝 알고리즘을 적용하였으며, 알고리즘의 성능 평가는 K-Fold 교차 유효성 검증 방법을 활용하였다. 특성화고 취업률 기준을 하위 25% 기준 값, 중위값, 상위 25% 기준 값으로 구분하여 예측모형을 개발하였으며, 종합적으로 예측 유효성을 판단하면 Random Forest가 세 가지 모든 경우에서 가장 우수한 성능 결과를 보여 주었다.

머신러닝 기반의 데이터 마이닝 기술을 활용해 특성화고 취업률에 영향을 미치는 변인별 영향력을 분석한 결과, 하위 25% 기준으로는 학생당 교육활동 지원 예산, 학생당 선택적 교육활동 예산, 동아리 참여비율, 학생당 장학금, 시군명 순으로 영향력이 높음을 확인하였다. 중위값 기준으로는 학생당 교육활동 지원 예산, 학생당 선택적 교육활동 예산, 동아리 참여비율, 학생당 학교시설 확충 예산, 학생당 장학금 순으로 영향력이 높음을 확인하였다. 또한 상위 25% 기준으로는 학생당 교육활동 지원 예산, 학생당 선택적 교육활동 예산, 동아리 참여비율, 학생당 학교시설 확충 예산, 시도명 순으로 영향력이 높음을 확인할 수 있었다.

기준 값 변화에 따른 주요 영향요인을 t검정을 통해 분석한 결과에서는 현장실습 프로그램이 많거나 동아리 활동 학생 참여비율이 높을수록 취업률이 높아진다는 선행연구 결과와 일치하거나 유사한 경향을 보이는 것을 확인하였다.

본 연구를 통해 개발된 특성화고 취업률 예측모형과 요인별 영향력 정보를 활용하여 실제 개별 특성화고 취업률을 쉽게 예측해 볼 수 있는 서비스 프로토타입을 설계하였다. 설계된 프로토타입은 개별 학교의 예측 취업률 구간정보, 취업률 예측에 유용한 개별 학교의 변인 값, 구간별 영향력 요인 등을 확인할 수 있도록 구성하여 향후 실제 서비스 개발에 활용할 수 있도록 하였다.

2. 결론 및 논의

본 연구는 학교 수준에서 취업률에 영향을 미치는 변인을 측정할 수 있는 데이터를 수집하여 예측모형을 만들고, 요인별 영향력을 분석하였다. 변인별 영향력 분석 결과, 학생당 교육활동 지원 예산 및 학생당 선택적 교육활동 예산과 동아리 참여비율의 영향력을 확인하였으며, 학교별 취업률 수준에 따라 차이가 있기는 하지만 장학금, 학교시설 확충 등도 취업률에 영향을 미치는 것으로 확인하였다.

예측모형의 변인 가운데 학교 차원에서 변화와 노력을 투입할 경우 현재의 예측 취업률 보다 향상된 결과를 기대할 수 있다. 일반적으로 학생들의 개인특성요인을 변화시키기란 어렵다. 반면, 학교환경요인 가운데서도 학교시설 확충 예산, 학생당 장학금, 그리고 교수학습요인 가운데 동아리 참여비율, 선택적 교육활동 예산, 교육활동 지원 예산은 학교 운영 기간 중 구성원의 합의와 노력을 통해 향상시킬 수 있는 변인이다. 특히, 본 연구에서 이들 요인은 데이터 마이닝 분석 결과 특성화고 취업률에 미치는 영향력이 상당 수준 확인되었다. 이에, 본 연구에서는 특성화고 취업률 예측모형을 바탕으로 하는 서비스에 다섯 가지 변인들의 변화에 따른 향상 가능한 예측 취업률을 시뮬레이션 할 수 있도록 한 프로토타입을 제안하였다. 특성화고 취업률 예측 서비스를 활용할 학교급 관리자는 학교환경요인과 교수학습요인의 조정 가능한 변인들의 영향력을 바탕으로 학생 취업률 향상을 위한 투입 계획을 고려할 수 있을 것이다.

먼저, 학교환경요인의 학교시설과 관련하여 최근 NCS기반 교육과정을 운영하면서 실무과목에 필요한 현장중심 설비 구축은 직업교육의 성과에 중요한 투자로 논의되고 있다(이영민 외, 2016). 이러한 최근의 논의는 취업률 중위값 및 상위 25% 구간에서 학생당 학교시설 확충 예산이 유의한 영향 변인으로 확인된 본 연구의 결과와도 연결시켜 볼 수 있다. 다양한 국가차원의 특성화고 지원 사업은 학교시설 혁신을 위한 예산 활용을 적극 허용하고 있는 바, 취업률을 높이기 위한 학교차원의 노력으로 학교시설 확충예산 확보를 위해 노력을 기울여 볼 수 있을 것이다. 다음으로 학교환경요인 가운데 학생당 장학금의

경우, 취업률 하위 25%와 중위값 기준에서 부적 영향력이 본 연구에서 확인되었다. 선행 연구 분석 결과 학생 장학금의 경우 학비감면 혜택 보다는 성적 우수자에게 제공되는 혜택으로 특성화고 학생의 성적과 연결된 변인이다. 예산변인이 취업률에 직접적 영향 보다는 잠재적 매개 변인으로 작용할 수 있음을 고려해 볼 때, 장학금은 우수한 성적을 받은 학생이 취업보다는 진학을 선택하는 등의 관련 선행 연구 결과를 바탕으로 이해해볼 수 있다(오석영, 2012). 그 밖에 장학금 예산과 관련된 학교단위 특성을 고려하여 간접적 취업률 예측 요인을 추정해볼 수 있을 것이며, 취업률이 중위값 이하에 해당되는 특성화고의 경우 변동가능 예산을 직접적 효과를 낼 수 있는 용도로 사용하는 것을 검토해볼 수 있을 것이다.

교수학습요인은 취업률 상승을 위해 학교에서 추가 투입이 가능한 변인들로 구성된다. 본 연구에서는 동아리 참여비율, 선택적 교육활동 예산, 교육활동 지원 예산이 취업률에 영향력이 높은 주요 변인으로 확인되었다. 즉, 취업률 하위 25%, 중위값, 상위 25% 기준 전반에 걸쳐 취업률이 높은 구간에서 투입된 참여비율과 예산이 많은 것으로 나타났다. 이의 결과는 학생들의 전공 지식 및 기술과 관련된 동아리 운영 활성화와 현장직업체험 및 현장실습의 기회를 늘릴 수 있는 선택적 교육활동은 특성화고 학생들의 진로 영역의 역량 향상과 취업 연계에 도움이 된다는 선행연구들의 결과와 같은 관점에서 해석된다(이쌍철, 엄문영, 2014; 조규형, 정철영, 2014). 특성화고에 투입되는 교육활동 예산은 학생들이 실제 참여할 수 있는 학교교육활동의 양과 질을 나타내는 지표가 된다(이영민, 2016). 정부의 정책적 특성화고 지원 사업은 학생당 교육활동 지원 예산 및 선택적 교육활동 예산의 추가 재정을 확보하는 기회가 될 수 있다. 또한 취업률 예측 서비스를 통해 학교단위에서 실제적으로 투입될 수 있는 변인들을 조정하여 시뮬레이션할 수 있도록 지원한다면 특성화고에서 당해 또는 미래 취업률 향상을 위한 계획을 세우는 데 도움이 될 수 있을 것이다.

그러나 학교 수준의 데이터를 분석하여 도출한 본 연구의 결과를 학생 개인 차원에서 처방적으로 적용하는 데에는 무리가 있다. 향후 이와 같은 한계점을 극복하기 위하여 학생 수준에서 데이터를 확보하여 계량적인 분석을 실시

할 필요가 있다. 또한 본 연구는 2010년부터 2018년까지 9년간의 고등학교 학교정보공시 데이터와 학교회계 예산 데이터를 분석 대상으로 활용하였다. 향후 공개 대상 데이터가 확대될 경우 예측 유효성 제고를 위해 추가 연구가 추진될 필요가 있으며, 계량 데이터 기반의 예측모형에서 확인한 변인별 영향력을 분석을 질적인 측면에서 보완할 필요도 있다. 즉, 변인에 영향을 미친 지역의 구직 환경이나 경제적 동향에 따른 연도별 전체 취업률의 변화 등과 같은 맥락적 상황이나 학교 교육 과정의 특성 등을 종합적으로 고려하는 후속 연구의 필요성이 제기된다.

본 연구 결과를 토대로 하는 취업률 예측 서비스는 프로토타입 설계에 한정하였기 때문에 서비스 자체의 효과성이나 교육적 기여도를 평가하기는 어렵다. 또한 본 연구 결과를 토대로 개별 학교의 취업률 예측 정보만 제공하는 것으로 구성하였으며, 개별 학생 혹은 시·도 단위 수준의 예측 정보는 설계 대상에서 제외하였다. 따라서 향후 실제 서비스 구축 또는 시범 개발·적용을 거쳐 효과를 검증할 필요가 있으며, 후속 연구에서 제안한 개별 학생 단위의 데이터를 분석한 예측모형을 적용하여 학교 수준의 정책적 대안 제시뿐 아니라 학생 개개인의 취업에 도움을 줄 수 있는 서비스로의 발전을 모색할 필요가 있다.

참고문헌

- 관계부처 합동 (2011). 공생발전을 위한 열린 고용사회 구현방안.
- 교육과학기술부 (2011a). 공생발전을 위한 교육 희망사다리 구축 방안.
- 교육과학기술부 (2011b). 특성화고 마이스터고 권역별 설명회 참고자료, 재구성.
- 교육부 (2014a). 고등교육법 시행령 일부개정, 재구성. 세종: 교육부.
- 교육부 (2015a). 2015년 마이스터고 사업 추진계획(안), 재구성. 세종: 교육부.
- 교육부 (2016a). 2016년 교육부 업무계획, 재구성. 세종: 교육부.
- 교육부 (2017). 직업계고 졸업자 취업률 17년 만에 50% 넘어 . 교육부 보도자료(11.20).
- 교육부 (2017a). 2017년 교육부 업무계획, 재구성. 세종: 교육부.
- 교육부 (2018a). 2018년 교육부 정부업무보고, 재구성. 세종: 교육부.
- 교육부(2019). 고졸로 성공할 수 있는 사회를 위한 고졸취업 활성화 방안. 세종: 교육부.
- 교육부(2020). 2020 직업계고 지원 및 취업 활성화 방안. 세종: 교육부.
- 교육부 · 한국교육개발원 (2017). 『교육통계 주요지표 포켓북. (통계자료 SM 2017-01)』. 진천: 한국교육개발원.
- 국가경쟁력강화위원회 (2011). 기술력 증진, 생산력 확대, 일자리 복지를 위한 학업 취업 병행 교육체제 구축 방안.
- 국가고용전략회의 (2010). 고등학교 직업교육 선진화 방안.
- 국민경제대책회의 (2010). 2011년 예산(안) 서민희망 3대 핵심과제.
- 김강호 (2009). 전문계 고등학교 진로지도 활동이 학생의 진로성숙에 미치는 영향. 진로교육연구, 28(4), 183-200.
- 김경근 (1999). 교육 성취에 대한 가족 구성의 영향. 교육사회학연구, 9(3), 1-23.
- 김경근, 변수용 (2006). 한국 사회에서의 상급학교 진학 선택 결정 요인. 교육사회학연구, 16(4), 1-27.
- 김기현 (2004). 가족 배경이 교육단계별 진학에 미치는 영향. 한국사회학, 38(5), 109-142.

- 김기현, 방하남 (2005). 고등 교육 진학에 있어 가족배경의 영향과 성별 격차: 한국과 일본의 경우. 한국사회학, 39(5), 119-151.
- 김성남, 김남희 (2020). 코로나19 발생에 따른 직업계고 온라인 개학 대응 동향 및 주요 이슈. 서울: 한국직업능력개발원.
- 김영식, 김훈호 (2019). 머신러닝 기법을 활용한 사교육 참여 예측 모형 탐색. 교육재정경제연구, 28(3), 29-52.
- 김지경, 김종성 (2014). ‘선취업-후진학’ 정책 도입까지 지난 10년 간 고졸 비진학 청소년의 취업 결정요인 분석. 청소년복지연구, 16(2), 255-272.
- 김지효 (2018). 특성화고 졸업생의 취업에 미치는 영향요인 분석: 연도별 비교를 중심으로. 취업진로연구, 8(3), 1-20.
- 김진 (2014). 데이터 마이닝 기법을 이용한 특성화고등학교 교육성과 분석. 컴퓨터교육학회, 17(6), 21-33.
- 김진원, 모영민(2019). 특성화고와 마이스터고 학생의 진로성숙도 및 영향 요인 비교 분석 연구. 직업능력개발연구, 22(1), 37-73.
- 노경란, 허선주 (2012). 특성화고 졸업 청년층의 취업결정요인. 직업능력개발연구, 15(2), 25-49.
- 닥터매쓰(2019). AI 수학전문의 닥터 매쓰 소개자료. <https://www.drmath.co.kr/>에서 검색.
- 박대륜, 유원진, 장준혁, 유인환, 배영권, 김우열, 안중민 (2020). 머신러닝 플랫폼을 활용한 소프트웨어 교수-학습 모형 개발. 정보교육학회논문지, 24(1), 49-57.
- 변수용, 김경근 (2008). 한국 고등학교의 교육결과에 대한 문화자본의 영향. 교육사회학연구, 18(2), 53-82.
- 안재영, 이병욱 (2012). 현장 실습 요인이 특성화 고교생의 취업 진로 결정에 미치는 영향. 한국기술교육학회지, 12(1), 27-47.
- 양정승, 김유미 (2014). 마이스터고 정책의 초기 노동시장 효과 분석. 한국노동경제논집, 37(3), 75-99.
- 오미애 (2017). 기계학습(Machine Learning)기반 사회보장 빅데이터 분석 및 예측모형 연구. 한국보건사회연구원.
- 오석영 (2012). 특성화고 학생의 진로결정요인 분석: 서울지역 진로결정 집단별 비교. 직업교육연구, 31(3), 135-151.

- 이덕현, 박지혜, 진유림, 장원섭 (2014). 진로지원과 진로탐색행동이 고졸 청년층의 취업성가에 미치는 영향분석. 산업교육연구, 28(2), 1-20.
- 이동훈 (2019). 기계학습 분류 기법을 활용한 대졸자 취업 예측 모델 연구. 단국대학교 석사학위논문.
- 이만기 (2014). 특성화고 고졸청년층의 진로준비 및 진로성숙이 취업에 미치는 영향 연구. 한국산학기술학회논문지, 15(8), 4961-4971
- 이쌍철, 엄문영 (2014). 특성화고 취업의 특징과 취업에 영향을 주는 학교 특징 분석. 한국교육, 41(3), 89-117.
- 이쌍철, 엄문영 (2014). 특성화고 취업의 특징과 취업에 영향을 주는 학교 특징 분석. 한국교육, 41(3), 89-117.
- 이쌍철, 주현준 (2013). 특성화고 졸업생 취업에 대한 학교 효과 분석: HGLM 모형을 활용하여. 직업능력개발연구, 16(3), 59-82.
- 이정은, 김다솜, 조일현 (2020). 동영상 기반 학습 환경에서 머신러닝을 활용한 행동로그의 학업성취 예측 모형 탐색. 한국컴퓨터교육학회, 23(2), 53-64.
- 이종현 (2020). 머신러닝을 활용한 중학교 수학 기초학력 미달 비율 예측모형 탐색. 영남대학교 대학원 박사학위 논문.
- 이지혜, 정철영 (2010). 특성화 고등학교 학생의 취업장벽과 관련 변인. 농업교육과 인적자원개발, 42(1), 25-47.
- 장석인 (2006). 독일의 이원화 직업교육훈련제도의 실태와 우리나라에의 시사점. 유럽연구, 23, 309-335.
- 장현진, 이지혜 (2012). 특성화고 학생의 취업선택 결정요인에 관한 다층분석. 직업능력개발연구, 15(2), 51-74.
- 전승환 (2014). 스위스의 직업교육훈련 체제. 글로벌리포트.
- 조규형, 정철영 (2014). 특성화고등학교 졸업예정자의 취업 결정 요인. 농업교육과 인적자원개발, 46(1), 73-102.
- 조규형, 정철영 (2014). 특성화고등학교 졸업예정자의 취업결정요인. 농업교육과 인적자원개발, 46(1), 73-102.
- 조현국 (2018). 머신 러닝을 활용한 이러닝 학습 환경에서의 학습자 성취 예측 모형 탐색. 학습자중심교과교육학회지, 18(21), 553-572.
- 채창균 (2006). 실업계고교 졸업생의 진로 선택. 한국교육고용패널 2차년도 학술대회.
- 천영민, 오민홍 (2015). 특성화고 졸업자의 재학 중 취업노력 및 취업지원이 실

- 제 취업에 미치는 영향. 질서경제저널, 18(2), 39-60.
- 최동선, 이종범 (2013). 특성화고 졸업 취업자의 좋은 일자리로의 이행에 영향을 미치는 요인. 직업교육연구, 32(1), 1-21.
- 최필선, 민인식 (2018). 머신러닝 기법을 이용한 대졸자 취업예측 모형. 직업능력개발연구, 21(1), 31-54.
- 한성민, 문상호, 이숙중 (2016). 청년취업 결정요인에서 본 청년 취업대책 개선책 연구. 정책분석평가학회보, 26(2), 187-212.
- 한신갑(2015). 빅데이터와 사회과학하기: 자료기반의 변화와 분석전략의 재구상. 한국사회학, 49(2), 161-192.
- 행정안전부(2019a). 전자정부 웹사이트 UI·UX.
- 행정안전부(2019b). 행정공공기관 웹사이트 구축운영 가이드.
- 허영준 오혁제 김명찬(2019). 직업계고 전환학기 운영 방안. 한국직업능력개발원.
- 허영준·김종우(2013). 정부부처, 지자체, 산업체 연계 특성화고 지원 확대 방안 연구. 한국직업능력개발원.
- 홍성표, 정진철 (2016). 특성화고등학교 학생의 취업·진학결정수준과 준비행동에 따른 진로결정 유형화 및 영향요인 분석. 직업교육연구, 35(5), 1-30.
- Agodini, R. Uhl, S. & Novak, T. (2004). Factors That Influence Participation in Secondary Vocational Education. Mathematica Policy Research, Inc.
- Becker, B. G. (1998). Proceedings of the 1998 IEEE Symposium on Information Visualization, 102-105.
- Davis, J., & Goadrich, M. (2006). The Relationship Between Precision-Recall and ROC Curves. Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, 2006.
- Frank, E., Hall, M. A., & Witten, I. H. (2016). The WEKA Workbench-Online Appendix for “Data Mining: Practical Machine Learning Tools and Techniques”(4th ed.). organ Kaufmann.
- Fritsch, J., Kuehnl, T., and Geiger, A. (2013). A new performance measure and evaluation benchmark for road detection algorithms. In International Conference on Intelligent Transportation Systems (ITSC).

- Hastie, T., Tibshirani, R. & Friedman, J. (2013). The Elements of Statistical Learning: Data Mining, Inference and Prediction. *Springer Series in Statistics*, 3rd. Springer.
- Karmel, T. (2007). Vocational education and training in Australian schools. *Australia Education Research*, 34, 101 - 117.
- Kohav, R. (1995). The Power of Decision Tables. In: 8th European Conference on Machine Learning, 174-189.
- Kumar, Y., & Sahoo, G. (2012). Analysis of Parametric & Non Parametric Classifiers for Classification Technique using WEKA. *I.J. Information Technology and Computer Science*, 2012, 7, 43-49.
- Landwehr, L., Hall, M. Frank, E. (2005). Logistic Model Trees. *Machine Learning*. 95(1-2), 161-205.
- Landwehr, N. Hall, M. A. & Frank, E. (2005). Logistic Model Trees. *Machine Learning*. 95(1-2):161-205.
- Lee, V. E., & Burkam, D. T. (2003). Dropping out of high school: The role of school organization and structure. *American Educational Research Journal*, 40(2), 353-393.
- Ministry of Education of the People's Republic of China (2019). Employment rates of vocational education. Retrieved from http://en.moe.gov.cn/features/VocationalEdc/figures/201905/t20190531_383836.html
- Shaw, R. (2019). The 10 Best Machine Learning Algorithms for Data Science Beginners. Retrieved from <https://www.dataquest.io/blog/top-10-machine-learning-algorithms-for-beginners/>
- Torun, H., & Tumen, S. (2017). Do Vocational High School Graduates Have Better Employment Outcomes than General High School Graduates? *International Journal of Manpower*, 40(8), 1364-1388.
- WEKA (2020). WEKA The workbench for machine learning. Retrieved from <https://www.cs.waikato.ac.nz/ml/weka>

연구보고 RR 2020-11

머신러닝을 활용한 특성화 고등학교
취업률 예측모델 개발 연구

발 행 2020년 12월 31일
발행인 박 혜 자
발행처 한국교육학술정보원
(www.keris.or.kr)
주 소 ☎41061 대구 동구 동내로64
전화: (053)714-0114
등 록 제22-1584호(1999년 7월 3일)
ISBN 979-11-6555-114-8(95370)

내용의 무단 복제를 금함. <비매품>



이 저작물은 “공공누리”

출처표시 조건에 따라 이용할 수 있습니다.

* 한국교육학술정보원 : www.keris.or.kr

