

데이터 마이닝 기법을 이용한 특성화고등학교 교육성과 분석

김진[†] · 용환승^{††}

요 약

본 연구는 초·중등 교육정보 공시를 통하여 전국 특성화고등학교의 학교단위별 자료(2012년도)를 파악한 후 데이터 마이닝 기법을 활용해 교육성과에 대한 분석을 실시하였다. 특성화고등학교의 교육성과는 취업률, 진학률, 기능경기대회 입상성적으로 정하였다. 분석은 크게 두 가지로 나뉘는데 첫 번째, 학교현황에 따라 각 교육성과에 유의한 차이가 있는지 알아보고 두 번째, 교육활동을 포함한 학교의 세부적 사항 중 각 교육성과에 영향을 미치는 요인을 탐색하였다.

주제어 : 데이터마이닝, 교육정보 공시, 특성화고등학교 교육성과

An Analysis of Specialized Vocational High School's Educational outcome using Data Mining Technique

Jin Kim[†] · Hwan-Seung Yong^{††}

ABSTRACT

This study reviewed the all specialized vocational high school across the nation, through the data per school in the Elementary, Middle & High Schools information disclosure system(2012), and carried out an analysis on the specialized vocational high school's educational outcomes in using the data mining technique. The educational outcomes of specialized vocational high schools were defined as the employment rate, the university entrance rate, the aware records in various vocational techniques contests. As the first research question, this study investigate whether there was any significant differences in educational outcomes depending on school's general characteristics. And as the second research question, this study explored the factors influencing on educational outcome.

Keywords : Data Mining, Educational information disclosure system, Vocational High School's Educational outcome

[†] 정 회 원: 이화여자대학교 교육대학원 컴퓨터교육과
^{††} 종신회원: 이화여자대학교 컴퓨터공학과(교신저자)
논문접수: 2014년 7월 21일, 심사완료: 2014년 9월 18일, 게재확정: 2014년 11월 17일

1. 서론

최근 연도별 교육부 주요 업무 계획서를 확인하면 산업체 요구인력과 특성화고등학교 간 밀접한 협력을 위해 다양한 방안을 강구하고 있다[1]. 이는 교육부 단위에서 특성화고등학교의 전문성 강화와 같은 다양한 정책을 통해 고등학교 졸업자의 취업률을 높이고자 하는 대안이라고 볼 수 있다. 따라서 사회문제인 청년실업률 감소를 위한 대안 중 하나로써 특성화고등학교의 성공적인 교육을 위한 다각도의 연구가 지속되어야 할 필요가 있다. 따라서 학교현황에 따른 교육성과 차이를 살펴보고 성공적인 교육적 성취를 이루는데 영향을 주는 교육적 요인을 정확히 파악하는 것이 중요하다. 본 연구를 위해 필요한 분석방법으로 데이터 마이닝을 들 수 있다.

즉, 본 연구의 목적은 데이터 마이닝 기법을 활용하여 학교 단위의 일반적 특성들을 대상으로 교육성과 간의 관계를 파악하고, 교육성과에 영향을 미치는 요인을 탐색함으로써 특성화고등학교를 위한 새롭고 유의한 지식을 도출하는 것이다. 초·중등 교육정보 공시자료를 추출하여 교육성과에 관한 데이터 마이닝 분석을 하였다. 첫 번째는 학교의 일반적 특성에 따라 교육성과에 통계적으로 유의한 차이가 있는지 알아보고, 두 번째는 교육성과에 도달하는데 영향을 미치는 교육적 요인이 무엇인지 탐색하고자 한다.

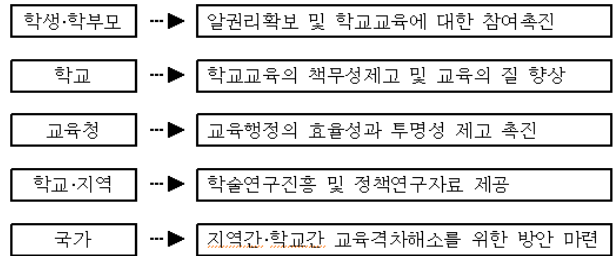
2. 이론적 배경

2.1 초·중등 교육정보 공시

2.1.1 교육정보 공시의 배경과 이해

교육관련기관의 정보 공개에 관한 특례법 시행으로 지난 2008년 12월 1일부터 전국 11,300여개 초·중·고등학교의 주요정보가 학교 홈페이지와 정보공시 사이트인 ‘학교알리미’를 통해 공개되고 있다. 교육관련기관 장은 시행령 제3조(초·중등학교 공시정보의 범위·횟수 및 시기 등) 1항에 근거하여 매년 1회 이상 공시하며, 학교 현장 및 시도교육청, 총괄관리기관에서는 초·중등학교 정보공시 양식 및 지침서에 근거하여 공시항목별 자

료를 입력·관리·운영하고 있다[2].



<그림 1> 학교정보 공시 추진 방향[2]

학교정보 공시 추진 방향은 <그림 1>과 같다. 교육정보화 백서(2013)에 따르면 2013년 기준으로 정보공시 대상 학교는 총 1만 2,003개교이며, 이 중 교육행정정보시스템(NEIS)을 사용하지 않는 학교들은 해당 학교 홈페이지를 통해 학교 정보를 공시하고 있다.

초·중등 교육정보 공시는 교육관련기관이 보유·관리하고 있는 정보를 적극적으로 공개함으로써 국민의 알 권리를 보장하고 학술과 정책연구를 진흥시키며, 아울러 학교 교육에 대한 참여와 교육행정의 효율성 및 투명성을 높이기 위해 도입된 제도이다[3]. 즉, 교육정보 공시는 학교의 자율성이 커짐에 따라 요구되는 학교운영의 투명성을 제고하기 위해 필요하다. 또한 학교단위에서는 정보공시를 통해 책임과 의무를 수행하는 과정을 국민들에게 보여줌으로써 각 학교의 수행에 따른 결과가 투명하고 명확하게 드러나야 한다[4].

2.1.2 교육정보 공시 항목

정보이용자들은 ‘학교알리미’를 통해 각 항목을 확인할 수 있는데 공시항목별로 매년 1회 이상 주요 정보들이 공시되고 있다. 공시항목은 분류 기준 상 ‘학생현황, 교원현황, 교육활동, 교육여건, 예·결산 현황, 학업성취도’와 같이 크게 9가지로 분류되며 지침 기준 상 크게 16가지로 분류된다. 지침기준으로는 ‘학교규칙 등 학교운영에 관한 규정, 교육과정 편성 및 운영, 학년·학급당 학생 수 및 전·출입, 학업중단 등 학생변동, 학교의 학년별 교과별 학습, 학교시설, 직위·자격별 교원현황, 학교 및 법인의 회계, 학교운영위원회, 학교급식, 보건관리·환경관리 및 안전관리, 학교폭력 발생현

황 및 처리, 국가·시·도 학업성취도 평가에 대한 학술적 연구를 위한 기초자료, 학생 입학 상황 및 졸업생의 진로, 초·중등교육법 제63조부터 제65조까지의 시정명령, 교육여건 및 학교운영상태 정보에 관한 사항, 자율공시'가 해당된다[5].

2.2 데이터 마이닝

2.2.1 데이터 마이닝의 정의 및 특징

데이터 마이닝이란 수많은 데이터를 분석하여 그 속에서 특정패턴을 발굴하고 이를 기반으로 예측을 수행하는 일련의 자동화된 프로세스를 의미한다[6]. 즉, 대규모 데이터 저장소에서 유용한 정보를 자동적으로 탐색하는 과정이라고 볼 수 있다[7]. 또한 데이터 마이닝은 미래를 예측하고 데이터를 설명하는 데 도움이 되는 도구로써 데이터 속의 구조적 패턴을 찾아 서술하는 데 관심이 있다(Ian Witten, 2011).

이러한 데이터 마이닝은 유용한 정보 추출이라는 목적 달성을 위해 통계학, 데이터베이스 기술, 패턴인식, 기계학습 등 매우 다양한 학문에 활용하고 있다. 데이터 마이닝이 기존의 통계적 기법과 구별되는 차이라면, 주어진 가설 검정으로 그치는 것이 아닌 데이터로부터 유용한 새로운 가설과 규칙을 이끌어 내는 데 있다고 볼 수 있다(전치혁, 2012).



<그림 2> KDD 과정[7]

데이터로부터 지식발견(KDD: Knowledge Discovery in Database)이라는 용어는 데이터로부터 지식을 뽑아내는 총체적인 과정을 의미하는데 이 과정은 데이터전처리에서 데이터 마이닝 결과의 후처리까지의 변환 과정으로 구성된다. 자세한 과정은 <그림 2>와 같다.

2.2.2 데이터 마이닝의 분석 기법

데이터 마이닝을 사례에 적용하기 위해서는 적

절한 기법을 선택하는 것이 중요하다. 어떠한 기법도 적용 분야나 사례에 관계없이 가장 뛰어난 결과를 제공하지는 못하며, 각 기법의 특성과 장단점 등을 고려하여 가장 적절한 기법을 선정해야만 한다[10]. 이제 본 연구에 필요한 연관분석, 회귀분석, 분류화 및 의사결정트리 기법을 중심으로 각 분석 방법에 대해 자세히 알아보려고 한다.

연관분석이란 큰 데이터 집합들에서 숨겨진 유용한 관계들을 발견하는데 유의한 방법이며, 이러한 관계들은 연관 규칙들(association rules) 또는 빈발 항목들의 집합 형식으로 나타낼 수 있다. 분류규칙과 비슷한 연관규칙은 형식 $X \rightarrow Y$ 의 함축적인 식이다[7][8]. 여기서 X와 Y는 서로 상이한 항목집합, 즉 $X \cap Y = \emptyset$ 이다. 여기서 연관규칙의 강도는 지지도(support)와 신뢰도(confidence)의 향으로 측정될 수 있다. 지지도는 한 규칙이 주어진 데이터 집합에 얼마나 자주 적용할 수 있는지를 결정하고, 반면 신뢰도는 Y에 속한 항목들이 얼마나 빈발하게 X를 포함한 트랜잭션들에 나타나는지를 결정한다[7].

하나의 종속변수가 독립변수들에 의해서 어떻게 설명되는지 혹은 예측되는지를 알아보기 위해 변수들간 관계를 함수식으로 표현하는 통계적 방법을 회귀분석이라 한다. 회귀분석에서는 종속변수의 유형에 따라 두 가지 모형으로 분류할 수 있는데, 종속변수가 연속형이면 일반적인 회귀분석 모형이 되고 이항 또는 순서의 범주형이면 로지스틱 회귀모형이 된다[11]. 선형회귀분석의 장점 중 하나는 모형화 과정에서 필요한 4가지의 방법(전진선택법, 후진선택법, 단계선택법, 전체입력방법)으로 독립변수를 선택하는 작업을 포함시킬 수 있다는 것이다.

분류는 각각의 속성 집합 x 를 미리 정해진 클래스 레이블 중 하나인 y 에 매핑하는 목표함수 f 를 학습하는 작업이다. 분류 기법은 이진(binary) 범주나 명목형(nominal) 범주를 갖는 데이터 집합을 예측하거나 서술하는 것에 가장 적합하다. 이에 속하는 기법들은 의사결정트리 분류기, 규칙기반 분류기, 신경망, 지지도 벡터 기계, 나이브 베이저안 분류기 등이 있다.

의사결정트리(decision tree)는 의사결정규칙(decision rule)을 나무구조로 도표화하여 분류

(classification)와 예측(prediction)을 수행하는 분석 방법이다. 이 방법은 나무구조에 의해서 모형이 표현되기 때문에 해석이 용이하고, 두 개 이상의 변수가 결합하여 목표변수에 어떤 영향을 주는지 쉽게 알 수 있는 장점이 있다. 다만, 의사결정트리에서는 연속형 변수를 비연속적인 값으로 취급하기 때문에 분리의 경계점 근처에서는 예측 오류가 클 가능성이 있다. 일반적으로 의사결정트리분석을 위해서 CHAID(Kass, 1980), CART(Breman et, 1984), C4.5(Quinlan, 1993)과 같은 다양한 알고리즘이 제안되고 있으며 최근에는 이들의 장점을 결합하여 보다 개선된 알고리즘들이 제안되고 있다[11].

2.3 관련연구

2.3.1 학업 관련 데이터 마이닝 연구

먼저 학습자에 구분을 두지 않고 학업과 관련한 데이터 마이닝 연구를 살펴보고자 한다. 특히 연구에 활용된 분석방법 및 알고리즘이 무엇인지 중점적으로 알아보았다.

김완섭(2010)은 대학생들의 교양과목 시험결과에 대한 객관적인 교육성과 합격여부 지표와 그에 대한 요인을 발견하였다. 데이터 마이닝의 분류기법인 SPSS의 C5.0, CART, CHAID와 같은 의사결정트리 알고리즘을 비교, 분석한 결과 교육성과 영향요인으로써 담당교수, 소속대학, 강의기간, 9시 수업 유무, 강의실이 탐색되었다. Paulo Cortez(2008)는 Portugal의 중등학교(secondary school) 학생의 학업능력이 낮다는 점에 착안하여 이진법 분류화와 회귀분석을 이용해 학업수행능력 예측에 관한 분석을 하였다. 2년간의 학생관련 데이터를 대상으로 수학과 모국어성적을 예측하는데 주요 변인을 탐색하였다. 그 결과 결석횟수, 학교선택사항 등 학교관련 변인이 가장 큰 요인으로 나타났고 인구통계학적 변인, 사회적인 활동 변인 순으로 영향을 미친다는 사실을 찾았다.

또한 타나이 미르와이스(2011)는 WEKA의 속성선택(Attribute Selection Method)과 군집분석(Cluster)을 이용해서 학습실패요인에 대한 분석을 하였다. 인구통계학적·학업성취 데이터, 설문조

사 등을 속성으로 요인분석 결과, EM 알고리즘의 정확도가 높다는 사실과 모국어로 강의를 듣고 과제완성률이 80%를 넘는 군집이 높은 성취도를 나타낸다는 결과를 도출했다.

신선미(2014)는 SPSS, WEKA, R을 이용하여 대학별 교육성과 요인 분석과 군집분석을 하였다. 분석방법별 비교를 통한 대학 교육성과의 공통적 영향 요인은 신입생 충원율, 재학생 충원율, 전임교원1인당 학생 수 임을 알 수 있었다.

2.3.2 초·중등 정보공시 데이터 관련연구

다음은 초·중등 정보공시 데이터에 관련된 연구 및 결과이다. 학교별 공시항목이 다양함에 따라 다각도의 연구가 요구되며 그에 따른 연구효과 또한 기대하고 있음을 알 수 있다. 실제 한국교육학술정보원에서는 매년 에듀데이터 활용 학술대회와 포럼을 통해 연구 자료와 논문들을 공유하고 있다.

2012년도 에듀데이터 활용 학술대회에서 천세영, 이성은(2012)은 교육지원청 수준에서 초등학교와 중학교를 대상으로 기초학력 결정요인 분석을 위하여 중다회귀 방법을 사용하였다. 분석 결과 순위별 요인은 방과후학교 지원율, 교사들의 현임교 장기 근무기간, 학생1인당교육비가 학업성취도 보통학력이상 도달비율로 긍정적인 영향을 미치는 것으로 나타났다. 반면 목적사업비와 초빙교사 비율은 보통학력이상 도달비율에 부정적 영향을 보이는 것으로 나타났다.

정병문(2014)은 초·중등 정보공시 자료를 이용해 일반계 고등학교 학교특성(설립유형)에 따른 학업성취도 관련 변인들을 분석하였다. 그 결과 학업성취도 보통학력 이상 도달 비율은 국·공립유형보다 사립유형의 학교 학업성취도가 높았다. 학업성취도에 영향을 미치는 변인들은 학교특성의 총 학생수, 교사 1인당 학생수이고 교육과정특성의 자율학교 지정 여부, 학생 1인당 대출자료수, 동아리활동 참여 비율, 재정특성에서는 학생 1인당 수익자 부담금, 학생 1인당 목적사업비로 나타났다.

또한 이은경(2014)은 초·중등 정보공시 자료 중 특성화고등학교(상업계열) 졸업생 진로 현황, 학

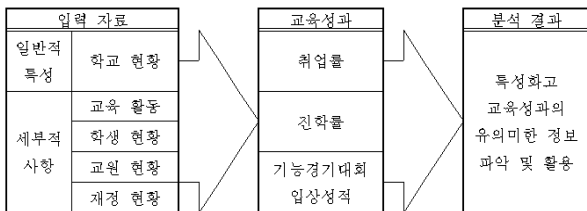
교교육과정 편성·운영 및 평가에 관한 사항, 학교별 교수 및 생활지도 등에 대한 평가항목 및 그 결과, 교과별 학업성취사항을 활용하였다. 특성화고등학교 학생들의 진로선택에 미치는 요인을 분석한 결과 유의한 특징으로 나타난 한 가지는 특성화고등학교 교육과정 편성이 진학에는 영향을 미치지 않았으나, 취업에는 유의한 영향을 미친다는 사실이다. 하지만 전국의 특성화고등학교 전체 데이터를 사용하지 않았기 때문에 일반화하기에 한계가 있을 수 있다.

본 연구는 전국 특성화고등학교의 교육성과에 대해 두 가지의 데이터 마이닝 도구를 이용해 기법별 비교 분석하였다. 따라서 최종적으로 각 도구별 분석 결과를 종합함으로써 특성화고등학교 교육성과에 대한 유의한 지식을 도출하고자 한다.

3. 연구방법

3.1 연구모형

본 논문의 특성화고등학교 교육성과에 대한 연구는 초·중등 교육정보 공시 자료를 크게 두 범주로 구분하여 분석한다.



<그림 3> 연구모형

연구 모형은 <그림 3>과 같으며 학교단위의 일반적인 특성에 따라 취업률, 진학률, 기능경기대회 입상성적을 나타내는 교육성과 지표에 통계적으로 유의한 차이가 있는지 분석하고, 그 외 세부적인 공시 데이터들 중 각 교육성과에 영향을 미치는 요인이 무엇인지 탐색하고자 한다. 교육성과 중 기능경기대회 입상성적은 초·중등정보공시 데이터가 아닌 국제기능올림픽대회 한국위원회에서 추출한 자료임을 밝힌다.

3.2 연구대상 및 변수 설명

<표 1> 최종 분석용 데이터 구분 및 설명

분류	독립변수	종속변수	설명	변수형태
학교현황	지역		1.수도권 2.광역시 3.지방	명목
	설립유형		1.공립, 2.사립	명목
	계열		1.공업 2.상업·정보 3.가사·실업 4.농·수산	명목
	남녀공학형태		1.남녀공학 2.남자, 3.여자	명목
교육활동	1일 평균 도서대출인원			수치
	1학년 공통과정		0=FALSE, 1=TRUE	이진
	교과교실제		0=FALSE, 1=TRUE	이진
	수준별 수업		0=FALSE, 1=TRUE	이진
	영어프로그램		0=FALSE, 1=TRUE	이진
	동아리 활동 비율			수치
	방과후 교과참여율		방과후교과 참여학생수/전교생 수×100	수치
	방과후 특기적성참여율		방과후특기적성 참여학생수/전교생 수×100	수치
	국어보통학력이상 비율		2011년 데이터	수치
	수학보통학력이상 비율		2011년 데이터	수치
	영어보통학력이상 비율		2011년 데이터	수치
			취업률	수치
			진학률	수치
		기능경기대회 입상성적	전국기능경기대회 성적 + 지방기능경기대회 성적	수치
학생현황	전교생 수			수치
	교사1인당 학생 수			수치
	학업 중단율			수치
교원현황	비정규교원 비율			수치
	전문교과교원 비율		전문교과교원 수/전체 교원 수×100	수치
	여교원 비율			수치
재정현황	1인당 평균장학금			수치
	1인당 학교발전기금			수치
	방과후학교 수익자부담금		공시자료로는 1인당 부담금 파악이 어려움	수치

*설명의 공란은 공시자료 그대로 사용한 변수임.

*교육활동의 3개 변수를 제외한 나머지는 2012년도 데이터

본 연구의 대상인 특성화고등학교는 특성화고·마이스터고 포털사이트인 ‘하이파이브’에서 특목고에 속하는 마이스터고등학교를 제외한 고등학교를 말한다. 현재 2014년 기준(교육청 통계) 전국의 특성화고등학교는 475개교, 마이스터고등학교는 35개교이다.

전국 단위의 475개교 중에서 21개교는 신설학교와 인문계열을 포함하는 일반 고등학교의 형태였기 때문에 공시정보 파악이 어렵거나, 타 학교와의 형평성 문제가 있어 제외하였다. 또한 지방학교 중 주·야간으로 수업이 이루어지는 곳의 데이터들은 평균 파악이 가능하기 때문에 연구 대상에 포함하였다. 즉, 최종 연구대상이 된 전국 454개의 특성화고등학교를 대상으로 각 학교에 해당하는 정보공시자료를 초·중등 교육정보 공시사이트인 ‘학교알리미’에서 검색 및 조회를 통해 시트에 입력하는 방법으로 전처리 과정을 통해 <표 1>과 같이 최종 데이터 셋을 생성하였다.

교육성으로 선택된 취업률, 진학률, 기능경기대회 입상성적은 종속변수로 분석에 사용되며 그 외의 다른 변수들은 독립변수의 역할을 한다. 다만, 수치형인 종속변수는 분석방법에 따라 범주형으로 변환하는 과정이 필요하다. 또한 독립변수들 중에서 ‘교육활동’에 해당하는 항목 중 1학년 공통과정, 교과교실제, 수준별 수업, 영어프로그램은 모두 해당 과정의 실시여부를 나타내는 지표이기 때문에 FALSE(0), TRUE(1)로 표현하여 분석에 포함하였다.

‘취업률’과 ‘진학률’은 2012년도 정보공시 데이터를 백분율 형태 그대로 사용하였다. ‘기능경기대회 입상성적’은 학교별 지방기능경기대회와 전국기능경기대회에서 입상한 통계결과를 활용하였다. 또한 연구에 필요한 2012년도 대회 결과에서 금상, 은상, 동상, 우수상 등 구분하지 않고 모두 합산한 총 입상 개수를 사용하였다.

3.3 분석방법

분석 방법은 도구에 따라 공통된 알고리즘을 쓰거나, 상이할 수도 있다. 분석도구로는 IBM SPSS(Statistical Package for Social Science) Statistics 18과 SPSS modeler 15(구 클레멘타인),

WEKA(Waikato Environment for Knowledge Analysis) 3.6을 사용한다.

<표 2> 분석도구별 분석 방법(알고리즘)

분석내용	분석도구	분석방법/알고리즘	
일반적 특성에 따른 교육성과 차이	SPSS statistics 18	일원배치분산분석	
		T검정	
교육성과 영향 요인 탐색	상관관계	SPSS statistics 18 pearson 상관계수	
	연관규칙	WEKA 3.6 Apriori	
	회귀분석	SPSS statistics 18	중다회귀
		WEKA 3.6	Linear Regression
	분류화 및 의사결정트리	SPSS modeler 15	C5.0
WEKA 3.6		C4.5(J48)	

4. 연구결과

4.1 일반적 특성에 따른 교육성과 차이 분석

특성화고등학교 교육성과 분석을 위한 다양한 변수들 중에서 학교의 일반적 특성(지역, 계열, 설립유형, 남녀공학 형태)에 따라 교육성과에 통계적으로 유의한 차이를 보이는지 알아보려고 한다. 해당 변수의 형태에 따라 크게 두 가지 방법으로 분석을 실시하였다. ‘지역’, ‘계열’, ‘남녀공학 형태’에 따른 교육성과 차이 분석은 세 집단 이상의 평균을 비교하는 방법인 일원분산분석(one-way ANOVA)을 사용한다. ‘설립유형’은 두 집단 간의 비교로써 독립표본 T검정을 사용하여 차이를 분석한다.

4.1.1 ‘지역’에 따른 교육성과 차이

취업률은 수도권이 40.75%, 광역시가 41.41%, 지방이 40.60%를 나타냈으며 진학률은 각각 39.71%, 42.26%, 41.46%의 수치를 보였다. 기능경기대회 입상성적은 각각 2.29개, 4.64개, 3.01개로 나타났다. 분석 결과, 지역(수도권, 광역시, 지방)에 따라 취업률과 진학률은 유의한 차이는 발견하지 못하였다. 기능경기대회 입상성적은 평균 3.2개를 나타내며 모집단 평균의 차이가 발견되어 사후 비교 분석 결과, 광역시가 수도권과 지방보다 높은 성과를 보이며 유의한 차이를 보였다.

4.1.2 '계열'에 따른 교육성과 차이

취업률에서는 공업이 41.13%, 상업·정보가 41.94%, 가사·실업이 40.20%, 농·수산이 35.50%로 나타났으며 진학률은 각각 39.35%, 41.92%, 47.51%, 41.66%를 나타냈다. 기능경기대회 입상성적은 각각 6.03개, 0.58개, 2.31개, 1.78개로 나타났다. 계열(공업, 상업·정보, 가사·실업, 농·수산)별 교육성과의 차이를 분석한 결과 취업률과 진학률은 집단의 분산이 같았으며 유의한 차이를 보이지 않았다. 기능경기대회 입상성적은 모집단 평균의 차이를 보여 사후 비교 분석을 실시한 결과, 공업 계열이 상업·정보, 가사·실업, 농·수산 계열보다 높은 성과를 나타내며 유의한 차이를 보였다.

4.1.3 '설립유형'에 따른 교육성과 차이

취업률에서는 공립이 38.46%, 사립이 44.14%로 나타났으며 진학률은 공립이 40.04%, 사립이 42.66%로 나타났다. 기능경기대회 입상성적은 공립, 사립 각각 3.10개, 2.07개이다. 설립유형(공립, 사립)에 따른 진학률은 사립이 공립보다 높은 유의한 결과를 보였다. Welch-Aspin 검정 시 학교 설립유형에 따른 취업률의 평균 또한 사립이 공립보다 높은 수치를 보이며 통계적으로 유의한 차이가 있었으나, 기능경기대회 입상성적은 공립이 사립보다 높은 유의한 결과를 보였다.

4.1.4 '남녀공학'에 따른 교육성과 차이

취업률에서는 남녀공학 학교가 39.31%, 남학교가 41.99%, 여학교는 46.08%를 나타냈으며 진학률은 각각 41.77%, 38.27%, 40.17%이다. 기능경기대회 입상성적은 각각 3.14개, 8개, 0.9개로 나타났다. 남녀공학 형태(남녀공학, 남학교, 여학교)에 따른 교육성과 분석 시 취업률은 유의한 차이가 있다는 것을 발견할 수 있었고 진학률은 유의한 차이를 나타내지 않았다. 기능경기대회 입상성적은 모집단 평균의 차이를 발견하고 사후비교분석 결과, 남학교가 남녀공학과 여학교보다 높은 성과를 보였다.

4.2 교육성과에 영향을 미치는 요인 탐색

4.2.1 '취업률' 성과 영향 요인

특성화고등학교 교육성과에 영향을 미치는 요인을 탐색하기 위해 학교의 일반적 특성('학교현황')을 제외한 세부적 사항('교육활동', '학생현황', '교원현황', '재정현황')에 속하는 변인들을 독립변수로 하여 회귀분석을 실시하였다.

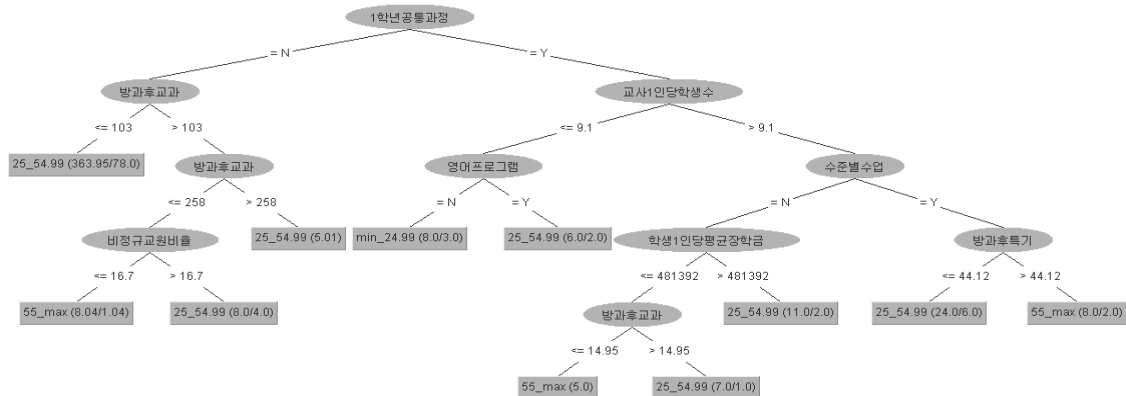
1) 회귀분석

먼저 SPSS를 이용하여 취업률을 종속변수로 다중회귀분석(단계선택법)을 실시한 결과, 취업률 교육성과에 영향을 주는 변수들은 총 3가지로 나타났다. 그 중 모형 1에서 나타난 '수준별 수업'이 가장 영향력이 높았으며 수준별 수업을 실시할수록 취업률이 통계적으로 유의하게 높아지는 것을 알 수 있다. 그 다음으로 영향력을 미치는 변수는 '방과후특기적성 참여율'과 '1학년 공통과정'의 순으로 나타났다.

이어서 도구별 비교를 통해 공통적 요인을 탐색하는 방법으로 WEKA 프로그램의 Linear Regression을 실시하였다. 분석 결과 취업률 성과에는 '수준별 수업'(3.3849)와 '1학년 공통과정'(2.4587) 순으로 영향을 미치고 있었다. 회귀식의 교차 검증은 5 fold cross validation일 때 약 47%의 교차검증 결과를 보여주었다.

2) 분류화 및 의사결정트리

분석에 앞서 교육성과의 3가지 연속형 변수(취업률, 진학률, 기능경기대회 입상성적)들을 범주형 형태로 변환해야 할 필요가 있다. 따라서 취업률에 해당하는 수치형 데이터들을 최상위 성과(55%-max), 보통 성과(25%-54.99%), 최하위 성과(min-24.99%)와 같이 범주형 변수로 변환하여 사용하였다. 주요규칙은 <표 3>과 같으며 분석결과, 취업률 성과의 요인은 '1학년 공통과정'이 가장 영향력 있는 변인으로 나타났다. SPSS를 이용한 분석 결과의 주요 요인들을 중심으로 분석한 의사결정트리는 <그림 4>와 같이 나타났고 교차타당성 확인 결과 타당도는 평균 71.1%, 표준오차 1.1%를 보였다. <그림 5>는 WEKA의 분석결과로써 Number of leaves = 11, Size of tree = 21의 의사결정트리를 확인할 수 있다.



<그림 5> 취업률 성과 의사결정트리(WEKA)



<그림 4> 취업률 성과 의사결정트리 일부(SPSS)

<표 3> 취업률 성과 주요 의사결정규칙

도구	IF	THEN	신뢰도 및 타당도
SPSS	1학년 공통과정=Y & 영어프로그램=Y & 학업중단율≤1.5 & 여교원 비율>56	최상위 성과 (55%-max)	0.833
WEKA	1학년 공통과정=Y & 교사 1인당 학생수>9.1 & 수준별 수업=Y & 방과후특기적성 참여율>44.12		74.15%

3) 취업률 성과 영향 요인 탐색 결과

회귀분석과 분류화를 이용한 취업률 교육성과에 영향을 미치는 요인을 탐색 한 결과를 종합해보면 <표 4>와 같다. 각 분석방법을 통해 가장 영향력이 높은 순서대로 순위를 부여하는 방식으로 발견된 영향 요인들을 나열하고 그 안에서 공통된 항목을 찾았다. 즉, 특성화고등학교 취업률

에 영향을 미치고 있는 주요 요인은 ‘수준별 수업’과 ‘1학년 공통과정’으로 모두 ‘교육활동’과 관련된 변인임을 알 수 있다. 또한 분류 분석에서 공통적으로 발견된 요인은 ‘방과후교과·특기적성 참여율’, ‘영어프로그램’으로 모두 ‘교육활동’과 관련된 변인임을 확인하였다.

<표 4> 취업률 성과 영향 요인 탐색 결과

순위	회귀분석		분류화	
	SPSS	WEKA	SPSS	WEKA
1	수준별 수업	수준별 수업	1학년 공통과정	1학년 공통과정
2	방과후특기 참여율	1학년 공통과정	1일평균 도서대출인원	교사1인당 학생수
3	1학년 공통과정	진학률(-)	방과후특기 참여율	방과후교과 참여율
4		학업중단율 (-)	영어프로그램	수준별 수업
5		교사1인당 학생수	방과후교과 참여율	영어프로그램
6		전문교과교원 비율(-)	수준별 수업	방과후특기 참여율
7			기능경기대회 입상성적	

4.2.2 ‘진학률’ 성과 영향 요인

1) 회귀분석

먼저 SPSS를 이용하여 진학률을 종속변수로 다중회귀분석(단계선택법)을 실시한 결과, 진학률 교육성과에 영향을 미치는 변수들은 총 6가지로 나타났다. 그 중 모형 1에 나타난 ‘방과후학교 수익자부담금’이 가장 영향력이 높았으며 방과후학교 수익자부담금이 높을수록 진학률 또한 통계적으로 유의하게 높아지고 있음을 확인했다. 그 다음으로는 ‘전문교과교원 비율(-)’, ‘방과후특기적성 참여율(-)’, ‘수준별 수업(-)’, ‘학업중단율(-)’, ‘1학년 공통과정(-)’순으로 나타났는데, 이들은 모두 음(-)의 베타(β)값을 나타내어 부정적인 영향을 미치고 있었다.

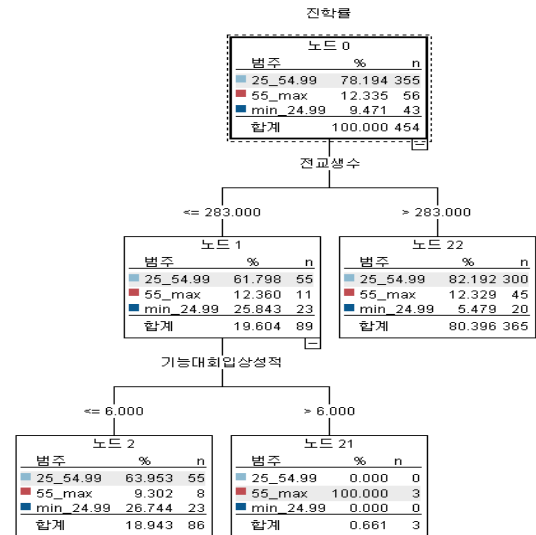
분석 도구별 비교를 위해 WEKA 프로그램의 Linear Regression을 이용하여 분석한 결과, ‘교사 1인당 학생수’(1.0155)가 가장 높은 영향을 주는 것으로 나타났다. ‘학업중단율’(-0.3714)과 ‘전문교과교원 비율’(-0.2215)이 그 뒤를 이었다. 단, ‘학업중단율’과 ‘전문교과교원 비율’은 음의 관계를 나타내며 진학률에 영향을 주고 있다. 회귀식의 교차검증은 5 fold cross validation일 때 약 60%의 결과를 보여주었다.

2) 분류화 및 의사결정트리

SPSS modeler와 WEKA를 이용하여 교육성과인 진학률을 목표변수로 지정하고 분석한 결과이다. 각 성과별 범주(최상위 성과, 보통 성과, 최하위 성과)에서 신뢰도가 상위인 규칙을 정리하며 가장 높은 영향력을 미치는 요인을 탐색하였다. 주요규칙은 <표 5>와 같으며 분석결과, 진학률 성과의 요인은 ‘1학년 공통과정’과 ‘전교생 수’가 가장 영향력 있는 변인으로 나타났다. SPSS를 이용한 분석 결과의 주요 요인들을 중심으로 분석한 의사결정트리는 <그림 6>과 같이 나타났고 교차타당성 확인 결과 타당도는 평균 75.3%, 표준오차 1.3%를 보였다. <그림 7>은 WEKA의 분석 결과로써 Number of leaves = 9, Size of tree = 17의 의사결정트리를 확인할 수 있다.

<표 5> 진학률 성과 주요 의사결정규칙

도구	IF	THEN	신뢰도 및 타당도
SPSS	1학년 공통과정=N & 영어프로그램=N & 국어 보통학력 이상 비율≤53.4 & 교사 1인당 학생수>14.3 & 전문교과교원 비율≤58.9 &방과후수익자부담금>510,000	최상위 성과 (55%)	0.889
WEKA	전교생 수>283 & 1학년 공통과정=N & 전문교과교원 비율≤46.3 &방과후수익자부담금>49,967,800	- max)	79.27%

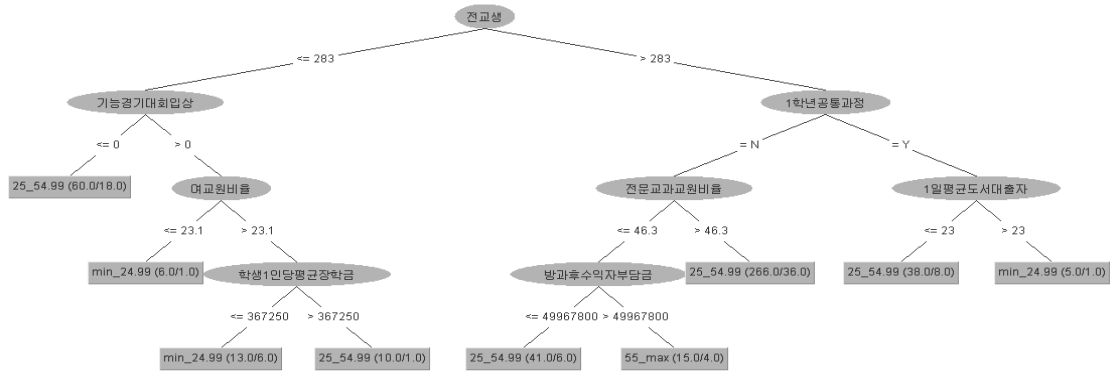


<그림 6> 진학률 성과 의사결정트리 일부(SPSS)

3) 진학률 성과 영향 요인 탐색 결과

회귀분석과 분류화를 이용한 진학률 교육성과에 영향을 미치는 요인을 탐색 한 결과를 종합해 보면 <표 6>과 같다. 각 분석방법을 통해 가장 영향력이 높은 순서대로 순위를 부여하는 방식으로 발견된 영향 요인들을 나열하고 그 안에서 공통된 항목을 찾았다.

‘교육활동’ 지표 중에서는 ‘1학년 공통과정(-)’과 ‘교원현황’ 지표 중에서는 ‘전문교과교원 비율(-)’이 가장 큰 영향력을 주는 변인으로 발견되었다. 그 외 회귀분석에서는 ‘학업중단율(-)’, 분류 분석에서는 ‘전교생 수’, ‘여교원 비율’이 각각 공통된 요인으로 발견되었다.



<그림 7> 전학률 성과 의사결정트리(WEKA)

<표 6> 전학률 성과 영향 요인 탐색 결과

순위	회귀분석		분류화	
	SPSS	WEKA	SPSS	WEKA
1	방과후학교 수익자부담금	교사1인당 학생수	1학년 공통과정	전교생 수
2	전문교과교원 비율(-)	취업률(-)	수준별 수업	1학년 공통과정
3	방과후특기 참여율(-)	학업 중단율(-)	전교생 수	기능경기대회 입상성적
4	수준별 수업(-)	전문교과교원 비율(-)	영어프로그램	전문교과교원 비율
5	학업 중단율(-)	기능경기대회 입상성적	방과후교과 참여율	1일 평균 도서대출인원
6	1학년 공통과정(-)	비정규교원 비율	여교원 비율	여교원 비율
7		수학보통학력 이상 비율(-)	교사1인당 학생수	
8		영어보통학력 이상 비율		

교원 비율'(0.1175)과 '국어보통학력이상 비율'(0.0335) 순으로 기능경기대회 입상성적에 영향력을 미치는 것을 확인할 수 있다. 또한 '교사 1인당 학생 수'(-0.2978)와 '여교원 비율'(-0.0866)은 음(-)의 영향을 주고 있었다. 회귀식은 5 fold cross validation일 때 약 50%의 교차 검증도를 보여주었다.

2) 분류화 및 의사결정트리

SPSS modeler와 WEKA를 이용하여 교육성공인 기능경기대회 입상성적을 목표변수로 지정하고 분석한 결과이다. 각 성과별 범주(최상위 성과, 보통 성과, 최하위 성과)에서 신뢰도가 상위인 규칙을 정리하며 가장 높은 영향력을 미치는 요인을 탐색하였다. 주요규칙은 <표 7>과 같으며 분석결과, '1학년 공통과정'과 '전교생 수'가 가장 영향력 있는 변인으로 나타났다.

<표 7> 기능경기대회 입상성적 성과 주요 의사결정규칙

도구	IF	THEN	신뢰도 및 타당도
SPSS	1학년 공통과정=N & 1일 평균 도서대출인원>11.6 & 방과후특기적성 참여율>27.87 & 전교생 수>1,024 & 여교원 비율≤41.4	최상위 성과 (11 - max)	0.909
WEKA	전교생 수>1024 & 여교원 비율≤41.4		57.88%

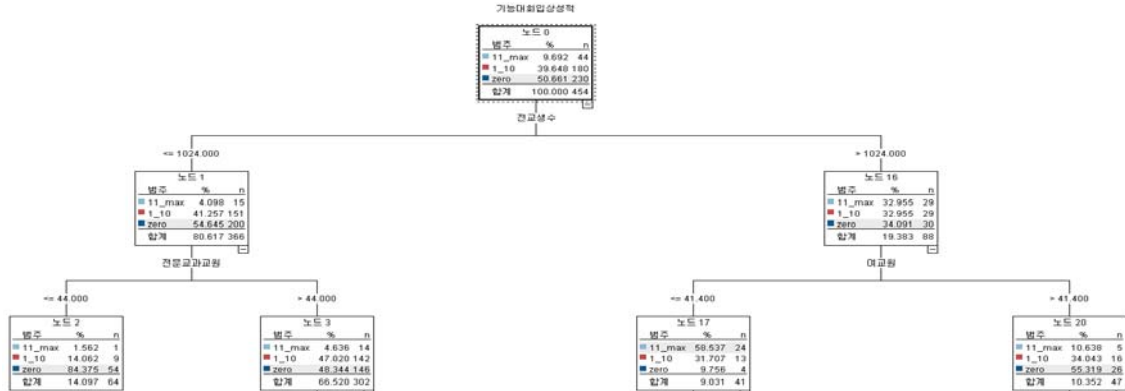
SPSS를 이용한 분석 결과의 주요 요인들을 중심으로 분석한 의사결정트리는 <그림 8>과 같이 나타났고 교차타당성 확인 결과 타당도는 평균 63.4%, 표준오차 1.7%를 보였다. <그림 9>는 WEKA의 분석결과로써 Number of leaves = 8, Size of tree = 15의 의사결정트리를 확인할 수 있다.

4.2.3 '기능경기대회 입상성적' 영향 요인

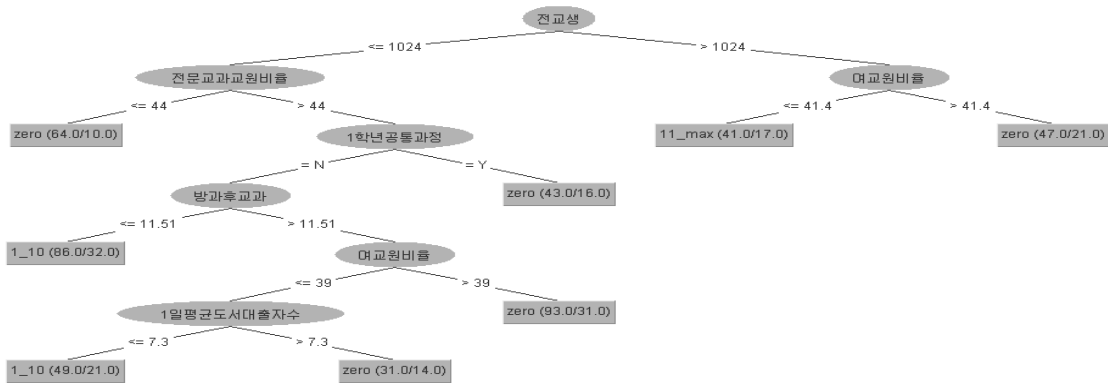
1) 회귀분석

먼저 SPSS를 이용하여 기능경기대회 입상성적을 종속변수로 다중회귀분석(단계선택법)을 실시한 결과, 영향을 미치는 변수들은 총 3가지로 나타났다. 그 중 모형 1에 나타난 '전교생 수'가 가장 영향력이 높았으며 그 다음으로는 '여교원 비율(-)', '전문교과교원 비율' 순으로 나타났는데, '여교원 비율'만은 부정적인 영향을 미치고 있었다.

도구별 비교를 위해 WEKA 프로그램의 Linear Regression을 이용하여 분석한 결과, 큰 수치는 아니지만 '교사1인당 학생 수'(-0.2978), '전문교과



<그림 8>기능경기대회 입상성적 성과 의사결정트리 일부(SPSS)



<그림 9>기능경기대회 입상성적 성과 의사결정트리(WEKA)

3) 기능경기대회 입상성적 성과 영향 요인 탐색 결과

<표 8> 기능경기대회 입상성적 성과 영향 요인 탐색 결과

순위	회귀분석		분류화	
	SPSS	WEKA	SPSS	WEKA
1	전교생 수	교사1인당 학생 수(-)	1학년 공통과정	전교생 수
2	여교원 비율 (-)	전문교과 교원 비율	전교생 수	여교원 비율
3	전문교과 교원 비율	여교원 비율 (-)	1일 평균 도서대출인원	전문교과 교원 비율
4		국어보통학력 이상 비율	방과후특기 참여율	1학년 공통과정
5		진학률	수준별 수업	
6		영어보통학력 이상 비율(-)	여교원 비율	
7		전교생 수	전문교과 교원 비율	

회귀분석과 분류화를 이용한 기능경기대회 입상성적 교육성과에 영향을 미치는 요인을 탐색한 결과를 종합해보면 <표 8>과 같다. 각 분석방

법을 통해 가장 영향력이 높은 순서대로 순위를 부여하는 방식으로 발견된 영향 요인들을 나열하고 그 안에서 공통된 항목을 찾았다. 공통적인 영향 요인은 ‘학생현황’ 지표 중 ‘전교생 수’, ‘교원현황’ 지표 중 ‘전문교과교원 비율’과 ‘여교원 비율 (-)’로 총 세 가지가 발견되었다.

4.2.4 교육성과에 영향을 미치는 요인 탐색 결과

분석 도구별 알고리즘을 적용하여 공통적으로 도출된 특성화고등학교 교육성과 영향 요인 탐색 결과를 종합해보면 <표 9>와 같다. 취업률 교육성과에 영향을 미치는 요인들은 모두 ‘교육활동’ 항목인 ‘1학년 공통과정’과 ‘수준별 수업’으로 나타났다. 진학률에는 ‘교육활동’ 항목인 ‘1학년 공통과정(-)’과 ‘교원현황’ 항목인 ‘전문교과교원 비율(-)’이 영향력 있는 요인으로 탐색되었다. 마지막으로 기능경기대회 입상성적에 가장 높은 영향을 미치는 요인들은 ‘학생현황’ 항목인 ‘전교생 수’와 ‘교원현황’ 항목인 ‘전문교과교원 비율’, ‘여교원 비율

(-)로 발견되었다.

<표 9> 특성화고 교육성과 영향요인 탐색결과

성과 분석	취업률	진학률	기능경기대회 입상성적
회귀 분석	수준별 수업 1학년 공통과정	1학년공통과정 (-) 전문교과교원비율 (-) 학업중단율(-)	전교생 수 전문교과교원비율 여교원 비율(-)
분류화	1학년 공통과정 수준별 수업 영어프로그램 방과후교과참여율 방과후특기참여율	1학년 공통과정 전문교과교원비율 전교생 수 여교원 비율	전교생 수 전문교과교원비율 여교원 비율 1학년 공통과정

5. 결론 및 제언

지금까지 특성화고등학교의 일반적 특성에 따른 교육성과 차이와 교육성과에 영향을 미치는 요인을 발견하기 위해 도구별 분석을 통하여 결과를 도출하였다.

첫째, 취업률은 사립형태와 여학교에서 높은 성과를 나타내며 통계적으로 유의한 차이를 보였다. 취업률 성과에 영향을 미치는 요인을 탐색한 결과, ‘1학년 공통과정’과 ‘수준별 수업’으로 나타났다. 특성화고등학교 입학 시 ‘1학년 공통과정’이 긍정적인 요인이라는 점을 통해 학생들의 진로 선택에도 긍정적인 효과를 줄 수 있다는 추측이 가능하다.

둘째, 진학률에 있어서 사립이 공립형태보다 유의한 차이를 보이며 다소 높게 나타났다. 진학률 성과에 영향을 미치는 요인을 탐색한 결과, ‘교육활동’에 해당하는 ‘1학년 공통과정’과 ‘교원현황’에 해당하는 ‘전문교과교원 비율’이 주요 요인으로 부정적 영향을 미치고 있음이 발견되었다. 특성화고등학교에서는 진학률 또한 중요한 교육성과로 판단하는데, 진학률에 긍정적인 ‘교육활동’ 관련 요인이 발견되지 않은 점에 주목할 필요가 있다.

셋째, 기능경기대회 입상성적은 광역시, 공업계열과 공립, 남학교 형태에서 유의한 차이를 보이며 높은 성과를 나타냈다. 기능경기대회 입상성적에 영향을 주는 요인을 탐색한 결과, ‘학생현황’에 해당하는 ‘전교생 수’와 ‘교원현황’에 해당하는

‘전문교과교원 비율’이 긍정적인 요인으로 발견되었다. 이어 ‘교원현황’에 해당하는 ‘여교원 비율’이 부정적 영향을 미치는 요인으로 탐색되었다. 기능경기대회에서는 보다 전문성 있는 실습형태의 문제들을 다루기 때문에 전문교과에 대한 정확한 이해와 실무 위주의 심화학습이 필요하다는 점을 알 수 있다. 따라서 교육성과를 높이기 위해서는 무엇보다 수준 높은 전문교과교원을 배치하고 반드시 해당 전공과목의 교원을 교과담당교사로 임용해야 할 것이다. 대회 지도경험이 전무하거나 전문지식이 부족한 교사들을 위한 교육프로그램을 마련하는 방안도 필요하다.

본 연구는 초·중등 정보공시 데이터만을 분석한 결과이다. 추후 보다 정확한 특성화고등학교 교육 성과에 대한 연구를 위해서는 학생과 교사들의 개별적인 데이터를 파악하여 추가하거나, 마이스터 고등학교를 포함한 계열별 비교 분석이 필요하다.

참고 문헌

- [1] 교육부 (2014). 2014 교육부 업무보고.
- [2] 이용효 (2009). 초·중등학교 정보공시 분석 및 개선방안 연구. 한국교육학술정보원.
- [3] 교육부·한국교육학술정보원 (2013). 2013 교육정보화 백서. PM 2013-3
- [4] 정제영·강태훈·류성창 (2013). 정부 3.0 추진에 따른 교육정보공시제 관련 이슈 분석. 한국교육학술정보원. RM 2013-8
- [5] www.schoolinfo.go.kr
- [6] www.scribd.com/doc/46940944/
- [7] Tan. P. N. et al. (2007). 데이터 마이닝. 융환승 외 6인 역. 서울: 인피니티북스
- [8] Witten I. H. et al. (2011). *Data Mining*. MA: Morgan Kaufmann publishers
- [9] 전치혁 (2012). 데이터 마이닝 기법과 응용. 서울: 한나래.
- [10] 이선아 (2013). 데이터 마이닝을 활용한 부정거래 탐지에 관한 연구. 석사학위 논문, 서울시립대학교

- [11] 김연형 · 김재훈 · 이석원 (2007). **고객관계 관리와 데이터 마이닝**(개정판). 서울: 교우사.
- [12] 김완섭 (2010). 의사결정트리를 이용한 교육성과 요인에 관한 연구. 한국공학교육학회 논문지, 13, 51-53.
- [13] Cortez P., & Silva A. (2008). *Using data mining to predict secondary school student performance*. 5-12. In Eurosis.
- [14] 타나이 미르와이스 (2011). *Cluster analysis for student's failures*. 석사학위 논문, 대구대학교
- [15] 신선미 (2014). **정보공시 자료의 데이터 마이닝을 통한 교육/연구성과 영향요인 및 대학의 군집분석**. 석사학위 논문, 고려대학교
- [16] 천세영, 이성은 (2012). **교육지원청 수준에서 기초학력 결정요인 분석연구**. 에듀데이터 활용 학술대회, 단위학교 성과 제고를 위한 교육여건 개선 방안 탐색, 11-42. 12월 21일 서울: 한국교육학술정보원 KERIS홀.
- [17] 정병문 (2014). **일반계 고등학교의 학교특성에 따른 학업성취도 관련 변인 분석 - 설립유형별 비교를 중심으로-**. 석사학위 논문, 경상대학교
- [18] 이은경 (2014). **특성화고등학교 학생들의 진로선택에 영향을 미치는 요인에 관한 연구-상업계열을 중심으로-**. 석사학위 논문, 경희대학교
- [19] <http://skill.hrdkorea.or.kr/>
- [20] www.highfive.co.kr



김진

2003 동국대학교
전자계산원 정보처리과
2014 이화여자대학교 교육대학원
컴퓨터교육과(교육학석사)

2014~현재 세명컴퓨터고등학교 재직 중
관심분야: 컴퓨터교육, 데이터 마이닝
E-Mail: kimj0704@sen.go.kr



용환승

1983 서울대학교
컴퓨터공학과(학사)
1985 서울대학교
컴퓨터공학과(공학석사)

1994 서울대학교 컴퓨터공학과(공학박사)
1995~현재 이화여자대학교 컴퓨터공학과 교수
관심분야: 컴퓨터교육, 빅데이터 관리, 데이터 마이닝
E-Mail: hsyong@ewha.ac.kr