

# 데이터 마이닝을 적용한 실업계 고등학생의 가출 예측모형

## A Forecast Model on Vocational High School for Runaway Students Using Data Mining

중앙대학교 아동복지학과

부 교수 이주리\*

Dept. of Child Development, Chung-Ang Univ.

Associate Professor : Lee, Ju-Rhee

---

### <Abstract>

This study was conducted to evaluate a forecast model designed to describe a vocational high school for runaway students. The study included 2000 adolescents from the KEEP(Korean Education and Employment Panel). A Data mining decision tree model revealed that: (1) Suicide ideation was a risk factor for running away among smokers. (2) High self-evaluation was a risk factor for running away among individuals that smoked and had no suicide ideation. (3) Drinking was predicted as a risk factor for no smokers, while family life dissatisfaction was predicted as a risk factor among non-smokers that drank. (4) Negative relationship with mother was predicted as a risk factor among non-drinking non-smoking.

---

▲주요어(Key Words) : 유아 성학대(young child sexual abuse), 성학대 자기보호 교육(sexual abuse self-protection education), 부모 교육 프로그램(parent education program), 성학대 신화와 고정관념(sexual abuse myths and stereotypes)

### I. 서론

청소년의 가출은 일반적으로 시대와 문화를 막론하고 나타나는 발달단계상의 문제이다. 성인이 되지 않은 청소년들의 가출이 특히 위험한 이유는, 가출이 독립을 이루는 계기가 되기보다 안전하지 못한 환경에 노출됨으로써 유흥업소나 범죄의 희생양이 되고 원조교제나 폭력집단의 가입 등 2차 비행으로 이어질 가능성이 크기 때문이다. 선행연구들에서는 중학생보다 고등학생이, 고등학생들 중에서도 학년이 올라갈수록 가출 경험이 더 많은 것으로 나타나는데(김경희·김지수, 2007; Plass & Hotaling, 1995; Sanchez et al., 2006), 특히 실업계 고등학생들에 대한 관심의 필요성이 제기된다. 입시 위주의 학업 중심 사회인 우리나라에서 그들은 소외의 대상이 되기

쉬울 뿐 아니라, 실제로 인문계 고등학생들보다 더 많은 가출을 경험하고 있는 것으로 나타나기 때문이다(김경희 등, 2007; 이종화 등, 2006; 하순인, 1997). 따라서, 본 연구에서는 실업계 고등학생들을 대상으로 가출에 중요한 원인으로 작용하게 되는 변인들이 무엇인지 살펴봄으로써 그들의 가출에 대한 예방책을 마련하고자 한다.

가출의 원인은 단일변인이라기보다 다양한 변인들의 상호작용 속에서 다면적인 영향을 받는 것으로 보인다. 대부분의 선행연구들도 하나의 요인으로 가출을 설명하기보다 가정이나 학교 또는 개인적 요인 등을 통합적으로 제시하고 있다(권윤아·김득성, 1999; 김경희·김지수, 2007; 배문조·전귀연, 2002; 이종화 등, 2006; 정혜경·권혜진, 2001; 정혜경·안옥희, 2001; De Man et al., 1993; Khong, 2009; Sanchez et al., 2006).

가정요인들은 크게 두 범주로 나누어 살펴볼 수 있는데,

---

\* 교신저자 : 이주리 (E-mail : julie@cau.ac.kr)

가정의 경제수준 혹은 가족 형태 등과 같은 구조적 요인들과 부모자녀관계와 같은 과정적·기능적 요인들이 있다.

경제 수준에 따라서는 다양한 결과들이 보고된다. 빈곤가정 청소년들의 가출율이 더 높다는 보고가 있는 반면(안창규 등, 1995), 이종화 등(2006)의 연구와 한정훈(1999)의 연구에서는 경제 수준이 잘 사는 편인 청소년들의 가출 경험 비율이 오히려 어려운 편이나 보통인 경우보다 더 높았다. 또 다른 결과로, 김경희와 김지수(2007)는 잘사는 편과 어려운 편인 청소년들이 보통인 가정의 청소년들보다 유의하게 가출경험이 더 높은 것으로 보고하였다. 이처럼 각기 다른 연구결과는 경제수준이 단일변인으로서 청소년들의 가출을 예측하지는 못하며, 다른 조건들을 함께 고려해야 함을 시사한다. 가족 형태면에서는 대부분 한부모 혹은 기타 다른 형태 가정의 청소년들의 가출 경험이 양친부모 가정에서보다 유의하게 더 높았다(김경희·김지수, 2007; 정혜경·권혜진, 2001; 정혜경·안옥희, 2001; De Man et al., 1993; Sanchez et al., 2006).

부모자녀관계와 관련하여서는, 일관되게 역기능적인 가족이 청소년의 가출에 유의한 영향을 미치는 것으로 나타난다. 어머니와 아버지가 거부적일수록 청소년들은 가출충동을 더 심하게 느꼈으며(배문조·전귀연, 2002), 가출 청소년들을 대상으로 인터뷰한 국외의 질적 연구들에서도 그들은 대부분 부모와 갈등을 겪고 있었고 부모의 사랑을 받지 못해 늘 외로움을 느꼈다고 보고하였다(Khong, 2009; Martinez, 2006). 반면, 가족간의 애착이 높을수록(김경희·김지수, 2007), 어머니와 아버지가 수용적이거나 반응적·지지적일수록(배문조·전귀연, 2002; Englandera, 1984), 부모와의 관계의 질이 높다고 느낄수록(정혜경·안옥희, 2001; De Man et al., 1993) 청소년의 가출 충동 혹은 경험이 적어지는 것으로 나타나 가정생활 만족도가 청소년 가출을 예방할 수 있을 것으로 시사된다.

개인요인으로는 흡연과 음주, 이성친구, 자살사고, 자아평가 등이 주 예측변인으로 언급된다. 특히, 흡연과 음주는 개인요인들 중에서 청소년의 가출에 영향을 미치는 매우 중요한 변인으로 나타난다(김경희·김지수, 2007; 이종화 등, 2006; 정혜경·안옥희, 2001). 이종화 등(2006)의 연구에서 흡연경험과 음주경험은 성격형, 친구 중 성격형자, 친구 중 임신경험자, 성격특성 등 다른 어떤 개인 요인보다도 가출에 더 큰 영향을 미쳤다. 정혜경과 안옥희(2001)의 연구에서도 흡연 청소년과 음주 청소년의 각각 절반은 가출 경험이 있었으며 이는 가출 경험이 없는 청소년들과 매우 유의한 차이로 나타났다. 이성친구도 가출에 중요한 요인으로 작용하였는데(김경희 등, 2007; 정혜경·권혜진, 2001), 이성친구가 있는 청소년들일수록 가출 경험이 매우 유의하게 높았다. 이는 자극을 추구하는 사춘기 청소년들이 이성친구와 보다 많은 시간을 보내기 위해 행하는 현상인 것으로 보인다. 그러나, 우울하고 침체되어 있어 자살에 대한 생각을 가지고 있는 청소년들도 가출을 할 가능성이 높은 것으

로 나타났다(De Man et al., 1993). 자신에 대한 부정적인 평가도 감수성이 예민한 청소년들을 가출로 이끄는 요인으로 작용한다. 김경희 등(1997)은 가출 경험이 있는 청소년들과 없는 청소년들의 자아개념을 비교하였는데, 가출청소년들일수록 전반적인 자아개념이 낮았으며, 학급 및 성취 자아개념, 가족 자아개념, 정서 자아개념 등이 모두 낮았다. 정혜경과 안옥희(2001)의 연구와 Englandera(1984)의 연구에서도 자아존중감이 낮을수록 가출 충동을 더 많이 느끼거나 실제 가출 경험이 더 많았다.

한편, 개인적 특성 중 성별에 대해서는 다양한 결과들이 보고되고 있다. 남학생이 여학생보다 더 많은 가출을 경험한다는 결과(김경희·김지수, 2007; 정혜경·안옥희, 2001)가 있는 반면, 여학생이 남학생보다 더 많은 가출을 경험한다는 결과(Sanchez et al., 2006)가 있고, 성별은 가출과 무관한 영향을 가진다는 결과(김경희 등, 2007) 등이 혼재되어 나타난다. 이는 성별이 다른 어떠한 조건들과 결합하느냐에 따라 결과가 다르게 나타날 수 있음이 시사되는 바이다.

가정요인과 개인요인 만큼 가출의 중요한 변인으로 언급되지는 않지만, 몇몇 연구에서 학교생활과 관련된 문제도 청소년의 가출에 영향을 미치는 것으로 나타난다. 학교생활이 불만족스러울수록 가출 경험이 증가하게 된다(김경희·김지수, 2007). 배문조와 전귀연(2002)은 학업성적문제, 진로문제 및 교사와의 갈등을 경험하고 있는 청소년들일수록 가출충동을 유의하게 더 많이 느끼는 것으로 보고하였다. 특히 학업성적이 가출을 예측하는 변인으로 제시되고 있는데(이종화 등, 2006; 정혜경·권혜진, 2001), 국외의 연구에서는 성적이 가출의 중요한 예측변인으로 거의 언급되지 않았으며, 이는 학업능력이 매우 중요한 우리나라의 사회적 분위기에 기인하는 것으로 보인다.

이와 같이, 가출은 단일변인이 아닌 다양한 요인들의 역동적인 과정을 고려해야 하는 것으로 시사된다. 그러나, 지금까지의 연구들은 가출에 대하여 가정, 학교 및 개인 변인 등 그 원인을 다양한 영역에서 통합적으로 살펴보았다 하더라도 대부분 변인들을 한번에 하나씩만 고려할 수 있는 ANOVA를 사용하였거나 여러 변인들 간의 상대적 영향력만을 살펴볼 수 있는 중회귀분석으로 검증하였다. 이는 설정된 여러 예측변인들 간의 수많은 상호작용의 관계를 모두 다 살펴보기 못한다는 한계를 가진다. 선행연구들에서 경제수준과 성별이 가출에 미치는 결과가 연구마다 각기 다르게 나타나는 것은 바로 이러한 점이 간과되었기 때문이다. 동일한 경제수준과 동일한 성별이라 하더라도 개개인 가지고 있는 수많은 다양한 변인들이 존재한다. 그렇기 때문에 개인이 가지고 있는 변인들 간의 조합관계에 따라 동일한 경제수준 내에서도 어떤 청소년들은 가출을 하는 반면 어떤 청소년들은 가출을 하지 않게 된다. 선행연구에서는 이성친구가 가출에 영향을 미치는 것으로 나타났다지만, 사실 이성친구가 있는 청소년들이 모두 다 가출을

행하지는 않는다. 그렇다면 어떠한 조건들에 있는 청소년들이 이성친구가 있을 때 가출을 하게 될 위험이 높아지는지 파악할 수 있다면 보다 실질적인 가출 예방 대책 수립에 도움이 될 것이다. 그러나, 기존의 연구 분석방법으로는 이러한 다양한 변인들 간의 수많은 상호작용을 모두 고려할 수 없으며, 이는 데이터 마이닝의 의사결정나무분석 방법을 통해 효율적으로 수행될 수 있다. 이런 측면에서 데이터 마이닝은 개인별 '맞춤형 개입방안' 으로 언급되기도 한다.

따라서, 본 연구에서는 데이터 마이닝을 사용하여 실업계 고등학교 학생의 가출을 설명하는 예측모형을 구축하고자 한다. 이처럼 개인이 가지고 있는 다양한 변인들 간의 조합관계를 통해 청소년 개인의 상황을 고려한 예측모형이 제시된다면 가출 청소년에 대한 이해를 보다 명확히 할 수 있게 되어 가출 예방에 대한 방안을 효율적으로 마련할 수 있을 것이다.

## II. 연구 방법

### 1. 연구 대상

본 연구에서는 한국직업능력개발원의 한국교육고용패널(Korean Education and Employment Panel: KEEP) 1차년도(2004년) 실업계 고등학교 3학년 데이터를 활용하였다. 표집지역은 제주도를 제외한 전국이며, PDA를 이용한 1:1 면접조사로 이루어졌다. PDA 조사는 coding 검증 등에 따른 비용 지출이 절약되고, 데이터 생성까지 소요되는 시간이 단축되는 장점이 있다. 단, PDA 조사가 곤란할 경우 설문지를 사용한 조사가 병행되었다(한국직업능력 개발원, 2006a).

연구대상자는 총 2000명이며 남학생 1105명(55.3%), 여학생 895명(44.8%)였다. 아버지 교육수준은 초등학교 졸업 이하 294명(17.7%), 중학교 졸업 413명(24.9%), 고등학교 졸업 795명(47.9%), 2~3년제 대학교 졸업 40명(2.4%), 4년제 대학교 졸업 102명(6.1%), 대학원 이상 15명(0.9%)였으며, 어머니 교육수준은 초등학교 졸업 이하 349명(20.0%), 중학교 졸업 553명(31.8%), 고등학교 졸업 773명(44.4%), 2~3년제 대학교 졸업 19명(1.1%), 4년제 대학교 졸업 43명(2.5%), 대학원 이상 4명(0.2%)였다. 아버지 직업은 기능원 및 관련기능 종사자 292명(19.0%), 농업·임업 및 어업 숙련 종사자 270명(17.5%), 장치·기계조작 및 조립 종사자 260명(16.9%), 단순노무 종사자 218명(14.2%) 등이었고, 어머니 직업은 서비스 종사자 353명(29.0%), 농업·임업 및 어업 숙련 종사자 195명(16.0%), 단순노무 종사자 157명(12.9%), 기능원 및 관련기능 종사자 137명(11.3%) 등이었다. 월평균 가구 소득은 100만원 이하 256명(13.3%), 100만원~200만원 650명(33.7%), 200만원~300만원 532명(27.5%), 300만원~500만원 389명(20.2%), 500만원 이상 104명(5.4%) 이었다(한국직업능력개발원, 2006b).

## 2. 조사 도구

### 1) 목표변인

KEEP의 설문 문항 중 가출 경험 여부(귀하는 가출 경험이 있습니까? - 예/아니오)를 묻는 1문항을 사용하였다.

### 2) 예측변인들

본 연구에서는 가정요인으로 보호자, 보호자와의 동거 여부, 가정 경제 형편에 대한 고민, 가정생활 만족도, 어머니와의 관계, 아버지와의 관계, 가족과의 활동, 부모와의 불화에 대한 고민 등을, 학교 요인으로 학교생활만족도, 학업성취, 진로에 대한 고민, 교사 애착 등을, 개인요인으로 성별, 휴연, 음주, 이성친구 유무, 이성친구에 대한 고민, 자살 사고, 자아평가 등을 예측변인으로 설정하였다. 사용된 예측변인들에 대한 구체적인 정보는 <표 1>을 통해 제시하였다.

## 3. 분석방법

본 연구에서는 실업계 고등학교 학생의 가출을 가장 효율적으로 예측하는 설명모형을 구축하기 위해 Answer tree 2.1을 사용하여 데이터마이닝의 의사결정나무분석을 하였다. 이는 선행 연구를 바탕으로 이론적 모형을 먼저 구축한 후, 그것을 검증하는 기존의 연구 분석방법과 완전히 상반되는 방식이다. 데이터 마이닝에서는 수많은 자료들을 투입하면 그 속에서 가장 적합한 모형이 자동적으로 산출되어 목표변인을 설명하는 가장 효율적인 예측모형이 구축된다. 또한, 데이터 마이닝에서는 특정한 통계적 가정을 필요로 하지 않기 때문에 본 연구에서와 같이 척도가 혼재된 자료에 적합하다. 척도의 급간 뿐 아니라, 명목 척도나 등간 척도 등 각기 다른 척도 수준의 사용에 있어서도 제약이 없으므로 변수에 대한 특별한 사전 조치를 필요로 하지 않기 때문이다(임희진·유제민, 2007; 최중후 등, 2003).

의사결정 나무분석은 본 연구에서와 같이 다양한 변인들 간에 이루어지는 수많은 상호작용을 모두 계산하여 그 중 종속 변인을 가장 잘 설명하는 조합관계들로 예측모형을 산출해낸다(최중후 등, 2003). 변인들 간의 수많은 상호작용을 모두 고려하는 것은 회귀분석과 같은 기존의 통계 분석방법으로는 불가능하다. 그러나, 이처럼 변인들 간의 상호작용을 고려하지 않는다면 가출과 같은 인간의 행동을 보다 명확히 설명할 수 없다. 인간의 행동은 복합적인 특성을 지니기 때문이다. 더구나, 청소년 각각이 처해 있는 상황은 개별적이기 때문에 다양한 변인들 간의 상호작용은 반드시 고려되어야 한다. 따라서, 가출 예측모형을 구축하기 위한 본 연구에서는 의사결정 나무분석이 매우 적절하다고 볼 수 있다.

본 연구에서는 CART(Classification and Regression Trees) 방식을 적용하여 분석하였으며, 분리기준으로는 지니 지수가

&lt;표 1&gt; 예측변인들에 대한 문항들

예측변인	문항의 예	척도급간 (Cronbach $\alpha$ )
보호자	'귀하의 보호자는 누구입니까?' 1문항	(어머니, 아버지/아버지/어머니/ 조부모/손위형제/친척 등)
보호자와의 동거 여부	'귀하는 보호자와 함께 살고 있습니까?' 1문항	예/아니오
가정 경제 형편에 대한 고민	가정의 경제적 형편에 대한 고민의 심각성을 묻는 1문항	4점 Likert 척도
가정생활 만족도	'귀하는 현재 가정생활에 만족하고 있습니까?' 1문항	5점 Likert 척도
어머니와의 관계	'나의 의견이나 감정을 믿고 존중해준다', '나와 대화를 자주 한다', '나는 그 분을 존경한다' 등 7문항	5점 Likert 척도 (.977)
아버지와의 관계	'어머니와의 관계' 문항과 상동	5점 Likert 척도 (.984)
가족과의 활동	'귀하는 고등학교 재학 중 가족과 함께 다음의 활동을 얼마나 자주 하였습니까?' 5문항 (외식하기/영화, 연극, 뮤지컬/ 박물관, 미술관, 음악회/ 여행, 등산, 낚시/ 운동)	5점 Likert 척도 (.718)
부모와의 불화에 대한 고민	부모와의 불화에 대한 고민의 심각성을 묻는 1문항	4점 Likert 척도
학교생활 만족도	'귀하는 재학하고 있는 학교 생활 전반에 대해 만족하고 있습니까?' 1문항	5점 Likert 척도
학업성취	국어, 영어, 수학, 과학영역, 사회영역 등 5과목에 대하여 각각 성취 수준을 응답함	5점 Likert 척도 (.715)
진로에 대한 고민	진학, 진로 문제에 대한 고민의 심각성을 묻는 1문항	4점 Likert 척도
교사 애착	'존경하는 선생님이 있다', '나를 이해해 주는 선생님이 있다' 등 4문항	예/아니오 (.694)
성 별	'귀하의 성별은 무엇입니까?'	남자/여자
흡 연	'담배를 피우니까?' 1문항	피운다/안피운다
음 주	'술을 마십니까?' 1문항	마신다/안 마신다
이성친구 유무	'귀하는 현재 사귀고 있는 이성친구가 있습니까?' 1문항	있다/없다
이성친구에 대한 고민	이성친구 문제에 대한 고민의 심각성 1문항	4점 Likert 척도
자살 사고	'귀하는 자살을 심각하게 고민해 본 경험이 있습니까?' 1문항	예/아니오
자아 평가	자신에 대해 '춤, 노래, 운동, 유머 등 재주가 있다', '리더십이 있다', '친구관계가 좋다' 등 7문항에 대해 동의하는 정도	5점 Likert 척도 (.644)

사용되었다. 지니 지수는 CART 방식에서 본 연구에서와 같이 목표변인이 이산형일 경우 사용된다. 부모마디(parent node)와 자식마디(child node)는 각각 25와 1로 설정하였으며, 정지규칙으로서 나무깊이는 3수준으로 설정하였다.

### III. 연구결과

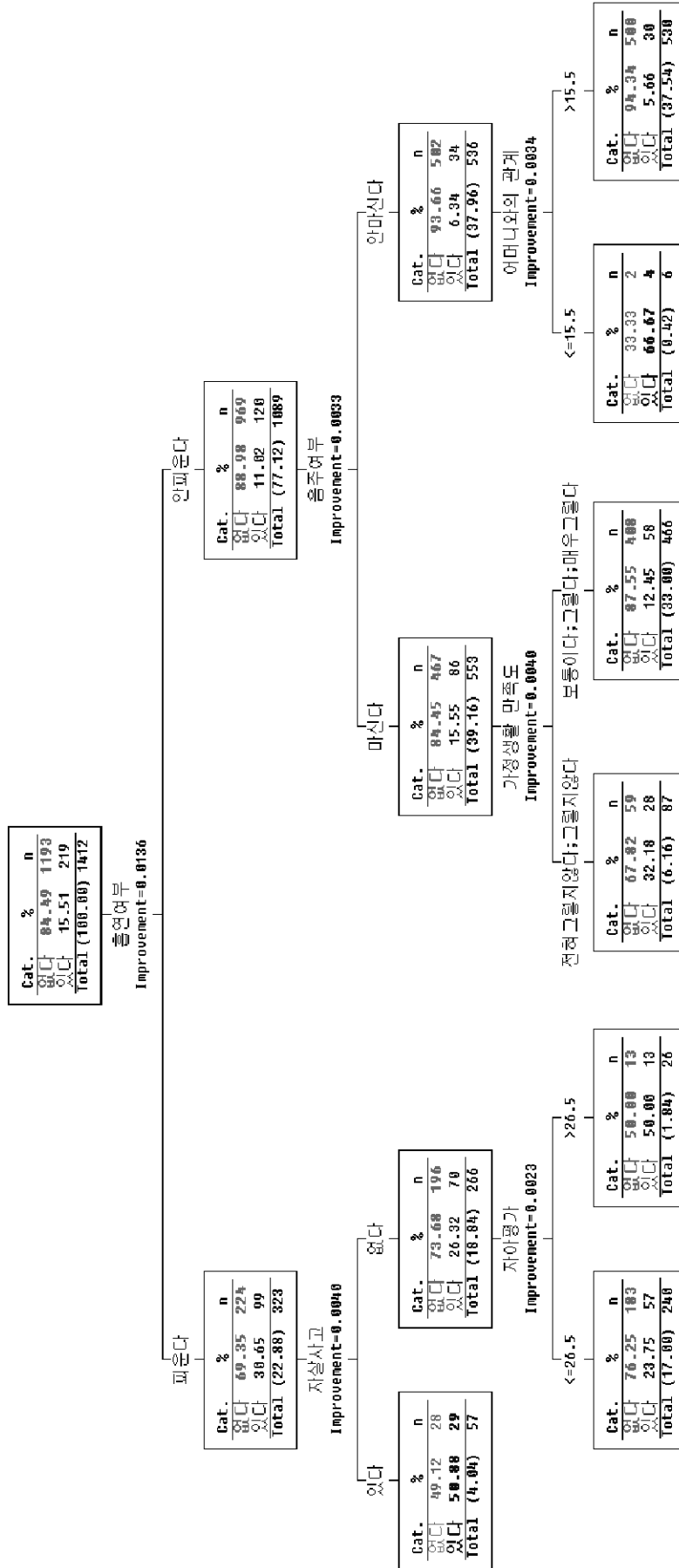
#### 1. 실업계 고등학교의 가출 예측모형

실업계 고등학교의 가출 예측모형에 대한 의사결정 나무분석 결과는 <그림 1>과 같다. 예측변인이 아무것도 투입되지 않은 뿌리마디에서 15.5%가 가출을 경험한 것으로 나타나,

실업계 고등학교의 약 15.5%가 최소 1번 이상의 가출을 경험한 것으로 볼 수 있다. 나무모형의 결과도표를 통해 실업계 고등학교의 가출에 영향을 미치는 예측변인들이 나타나 있다. 여기서 상위마디일수록 중요한 예측변인이 되며, 따라서 가장 상위의 마디에 있는 예측변인이 목표변인을 설명하는 가장 중요한 변인이 된다. 본 연구에서 실업계 고등학교의 가출을 예측하는 가장 중요한 변인은 <그림 1>에서 보는 바와 같이 '흡연' 여부였다. 흡연을 하게 되면 가출 경험이 약 30.7%로 증가하였으며, 반면 흡연을 하지 않는 경우 가출 경험은 11%로 감소하였다.

흡연 집단의 경우를 먼저 살펴보면, 흡연 청소년이 자살에 대하여 심각하게 고민해본 경험이 있다면 가출의 위험성은 이전의 30.7%에서 약 50.9%까지 증가하였다. 반면, 흡연 청소년이 자살에 대해 심각하게 고민해 본 경험이 없다면 가출

가계 정형 (Training Sample)



<그림 1> 실업계 고등학교의 가계 예측 모형

&lt;표 2&gt; 실업계 고등학생의 가출 예측모형에 대한 이익도표

node	이익지수				누적지수			
	node:n	node:%	gain(%)	index(%)	node:n	node:%	gain(%)	index(%)
13	6	0.42	66.67	429.83	6	0.42	66.67	429.83
3	57	4.04	50.88	328.03	63	4.46	52.38	337.73
10	26	1.84	50.00	322.37	89	6.30	51.69	333.24
11	87	6.16	32.18	207.51	176	12.46	42.05	271.09
9	240	17.00	23.75	153.13	416	29.46	31.49	203.03
12	466	33.00	12.45	80.25	882	62.46	21.43	138.16
14	530	37.54	5.66	36.50	1412	100.00	15.51	100.00

node: 마디번호, node:n: 해당마디번호에서의 사례수, node:%: 해당마디번호에서의 백분율, gain(%): 해당 마디에서 올바르게 분류된 개체 수/해당마디의 전체 개체수, index(%): 해당 마디에서의 gain 값과 전체에서 gain 값의 비(최종후 등, 2003).

경험은 30.7%에서 26.3%로 다소 감소하였으나, 이 조건에서 자아평가가 높으면(26.5점 이상) 오히려 가출 경험이 50%까지 증가하였다.

비흡연 집단의 경우, 가출을 예측하는 두 번째 주요변인이 '음주' 여부로 나타났다. 흡연은 하지 않지만, 음주를 할 경우 가출 경험은 11%에서 약 15.6%로 증가하였으며, 흡연도 하지 않고 음주도 하지 않을 경우 가출 경험은 6.3%까지 감소하여 약 93.7%가 가출을 하지 않았다. 음주를 하는 청소년들이 가정생활 만족도가 좋지 않을 경우 가출 경험은 32.2%까지 증가하였으며, 반면 음주를 하더라도 가정생활 만족도가 보통이상으로 좋으면 가출 경험은 12.5%로 감소하였다. 흡연도 하지 않고 음주도 하지 않는 청소년들이라 하더라도 어머니와의 관계가 좋지 않으면(15.5점 이하) 가출 경험은 약 66.7%까지 증가하였으나, 어머니와의 관계가 좋으면(15.5점 이상) 가출 경험은 약 5.7%까지 감소하였다. 즉, 흡연과 음주를 모두 하지 않으면서 어머니와의 관계도 좋으면 94.3%는 가출을 하지 않는 것이다.

## 2. 실업계 고등학생의 가출 예측모형에 대한 이익도표

실업계 고등학생의 가출 예측모형에 대한 이익도표는 <표 2>를 통해 제시되었다. 목표 범주는 가출 경험이 '있다' 였다. node는 마디번호를 뜻하며, 나무도표에서 가장 상위의 왼쪽부터 1번이 된다(임희진·유계민, 2007). <표 2>에서 보는 바와 같이 13번 마디가 index 429.83%로 가장 상위를 차지하였다. 이는 아무런 예측변수도 투입되지 않았을 때와 비교했을 때, 13번 집단의 특성을 가진 청소년들의 가출 경험율이 429.83%, 즉 약 4배임을 뜻한다. 따라서, 13번 마디의 특성을 가진 실업계 청소년들의 가출의 위험이 가장 높은 것이다. 13번 마디의 특성은 흡연과 음주는 모두 하지 않으나, 어머니와의 관계가 좋지 않은 집단으로 이 집단의 66.7%는 가출을 경험하였다. 반면, 가출의 위험이 가장 적은 집단은 흡연과 음주를 모두 하지 않으면서 어머니와의 관계가 좋은 집단으로 이들의 94.3%는 가출을 경험하지 않았다(이 집단은 약 5.7%만이 가출을 경험함).

<표 2>에서 누적지수의 index(%)는 100%가 될 때까지 산출되는데, 이렇게 산출되어 나온 마디의 수로 의사결정나무를 최종 구성하는 것이 가장 효율적인 모델이 된다(임희진·유계민, 2007). 본 연구에서는 100%가 될 때까지 총 7개의 마디가 산출되어 7개의 마디로 의사결정나무가 마무리되었다.

한편, 본 연구에서 산출된 실업계 고등학생의 가출 예측모형이 얼마나 적절히 일반화 될 수 있는지 살펴보기 위해 데이터의 분할에 의한 타당성 평가를 실시하였다. 이는 전체 data를 일정 비율의 training data와 testing data로 나누어 비교하는 방법으로 보통 70:30의 비율로 설정하여 검증하는 것이 일반적이다. training data는 모형구축 자료로서 의사결정나무를 구축하는데 사용되며, testing data는 모형검증 자료로서 모형구축 자료를 통해 얻어진 결과를 적용해볼게 된다(최종후 등, 2003). 본 연구에서도 training data와 testing data의 비율을 70:30으로 나누어 검증하였다. 이 때, training data와 testing data의 각각의 위험 추정치의 차이가 크지 않아야 일반화의 가능성이 높아진다. 검증 결과, training data의 위험추정치 0.15(표준오차 0.01), testing data의 위험 추정치 0.20(표준오차 0.02)으로 나타났다. 따라서, 본 연구에서 구축된 실업계 고등학생의 가출 예측모형은 일반화에 큰 무리가 없을 것으로 판단된다.

## IV. 논의 및 결론

본 연구에서는 실업계 고등학생의 가출에 대하여 데이터 마이닝의 의사결정 나무분석을 통해 가정, 학교 및 개인 요인 등 다양한 변인들을 동시에 투입하여 여러 변인들 간의 수많은 상호작용을 검증하였다.

우선, 예측변인들을 아무것도 설정하지 않았을 때 실업계 고등학생의 가출 경험 비율은 약 15.5%로 나타났다. 이는 실업계 고등학생의 가출 경험이 18.7%로 나타난 김경희 등(2007)의 연구와 비슷한 수치이다. 평균 약 6명 중 1명 꼴로 가출을 경험하게 된다는 것이다. 인문계 고등학교에서는 약 10명 중 1명 꼴로

가출을 한다는 연구결과(김경희 등, 2007; 이종화 등 2006)와 비교해 볼 때 이는 상당히 높은 수준으로서, 실업계 고등학교의 가출에 대한 예방 대책이 시급히 필요함을 시사한다.

본 연구에서 의사결정 나무분석을 통해 어떠한 특성을 가진 실업계 고등학교 학생들이 가출을 하게 될 위험이 높은지 살펴본 결과, 그들의 가출을 예측하는 가장 중요한 변인은 '흡연' 여부였다. 이는 흡연이 다른 개인 요인(성경험, 친구 중 성경험자·임신경험자, 성격특성)에 비해 가출에 상대적으로 더 큰 영향을 미쳤다는 이종화 등(2006)의 연구와 일치하는 결과이다. 본 연구에서 흡연을 하게 되면 가출 경험이 약 30.7%로 증가하였으며, 반면 흡연을 하지 않는 경우 가출 경험은 11%로 감소하였다. 이렇게 흡연이 가출을 예측하는 가장 중요한 변인으로 나타난 것은 비행과 관련되는 것으로 보인다. 청소년들의 흡연은 분명한 일탈행동으로 간주되며, 또래집단에 민감한 영향을 받는 발달적 특성을 고려해 볼 때 흡연 청소년들은 비행친구들과 어울리며 가출을 경험했을 가능성이 있다.

흡연 청소년들의 가출에 대한 위험을 더욱 증가시키는 변인은 '자살사고'였다. 흡연 청소년들이 자살에 대해 심각하게 고민해 본 적이 있을 경우 가출 경험은 50.9%까지 증가하였다. 이러한 특성을 가진 청소년들은 2명 중 1명이 가출하는 꼴인 것이다. 자살생각을 하면서 가출을 하게 되면 주위사람들의 직접적인 보호도 받을 수 없게 되어 매우 위험한 상황에 처해질 수 있다. 따라서 이러한 조건의 청소년들에게는 특별한 관심과 보살핌이 요구된다.

한편, 동일한 흡연 청소년이라 하더라도 자살에 대해 심각하게 고민해 본 경험이 없다면 가출 경험은 26.3%로 감소하였다. 그러나, 이 조건에서 자아평가가 높은 청소년들은 오히려 가출 경험이 50%로 2배 가량 증가하였다. 이는 자신에 대해 부정적인 생각을 가지고 있는 청소년들일수록 가출의 위험이 더욱 높아진다는 선행연구들(김경희 등, 1997; 정혜경·안옥희, 2001; Englandera, 1984)과 반대되는 결과이다. 일반적으로 가출을 예방하는 보호요인으로 간주되었던 긍정적인 자아평가가 흡연을 하고 자살에 대해 심각하게 고민해 본 경험이 없는 청소년들에게는 오히려 가출을 촉발시키는 위험요인으로 작용한 것이다. 이에 대해서는 다음과 같이 생각해 볼 수 있다. 자살에 대해 생각해 본 적이 없다는 것은 적어도 심각한 심리적 부적응을 겪고 있지 않다는 것이며 이런 상태에서 자신에 대한 높은 자아평가가 우월감을 형성하여 어떠한 자극을 찾아 가출을 경험했을 수 있다. 이들이 흡연을 하고 있는 비행청소년이란 조건이 이를 더욱 뒷받침한다. 비행또래 집단에서의 비행행동은 자신들의 위치를 확보하는 것에 도움이 되어 보상으로 작용하기 때문이다. 이러한 본 연구결과를 지지하는 최근의 선행연구가 있다. Menon et al.(2007)은 일반적으로 알려져 있는 바와 달리 높은 자아존중감이 오히려 반사회적 행동을 이끌어 낸다고 보고하였다. 높은 자아존중감을 가진 반사회적 청소년들은 자신의

행동으로 인해 받게 되는 보상에 높은 가치를 두며 자신의 행동을 합리화시킨다는 것이다. 이처럼 일반적으로 자신에 대한 긍정적인 평가가 청소년들의 가출과 같은 비행을 예방하는 보호요인이 되는 것으로 알려져 있지만, 사실 그것이 모든 청소년들에게 동일하게 적용되는 것은 아니며 개별적인 특성에 따라 오히려 위험요인으로 작용하게 될 수도 있다는 것이 본 연구결과를 통해 나타난 것은 매우 의미있는 발견이다.

비흡연 집단의 경우, '음주' 여부가 가출을 예측하는 중요한 변인으로 나타났다. 이 역시 음주가 흡연과 마찬가지로 청소년의 가출에 매우 중요한 변인으로 작용한다는 선행연구들(김경희·김지수, 2007; 이종화 등, 2006; 정혜경·안옥희, 2001)과 일치하는 결과이다. 흡연을 하지 않으면 가출 경험이 11%였지만, 흡연을 하지 않더라도 음주를 하게 될 경우 가출 경험이 15.6%로 증가하였다. 반면, 흡연도 음주도 하지 않을 경우 가출 경험은 6.3%로 낮아졌다.

비흡연 청소년들이 음주를 할 경우 가정생활에 대해 불만을 느끼게 되면 가출 경험은 15.6%에서 32.2%로 2배 가량 증가하였다. 선행연구들은 가족과 갈등을 겪고 있을 때 가출의 위험이 있다고 보고하였으나(Khong, 2009; Martinez, 2006) 그것은 모두에게 동일한 정도로 적용되기보다 이처럼 비흡연 이면서 음주를 하는 청소년들에게 특히 더 위험한 것임을 알 수 있다. 그러나, 가정생활에 보통 이상으로 만족을 느끼고 있다면 가출 경험은 12%로 감소하여 흡연을 하지 않는 음주 청소년들에게 가정생활만족도가 가출에 대한 보호요인으로 작용하였다. 가족간의 애착이 잘 형성되어 있어 기능적인 가족의 청소년들이 가출의 위험이 적다는 연구결과들(김경희·김지수, 2007; 배문조·전귀연, 2002; 정혜경·안옥희, 2001; De Man et al., 1993; Englandera, 1984)은 이처럼 특히 비흡연이지만 음주는 하는 청소년들의 가출에 대한 보호요인임을 지지한다.

비흡연이면서 비음주 집단의 청소년들은 어머니와의 관계가 좋지 않으면 가출 경험이 6.3%에서 66.7%로 약 10배 이상 증가하여 어머니와의 관계는 비흡연/비음주 집단의 청소년들에게 특히 중요한 변인으로 나타났다. 배문조와 전귀연(2002)에 따르면, 어머니가 거부적인 양육태도를 보일수록 청소년들은 더 심한 가출충동을 느낀다고 보고하였으나, 본 연구 결과에 따르면 흡연과 음주를 하는 비행청소년들에게는 어머니와의 관계가 가출을 예측하는 주요 변인으로 나타나지 않았으며 흡연과 음주를 하지 않는 청소년들 집단에서 가출을 예측하는 위험요인으로 나타난 것이다. 이처럼 어머니와의 관계가 중요한 변인으로 나타난 것은 현실적으로 아버지보다 어머니와 함께 하는 시간이 훨씬 많으므로 직접적인 갈등이 발생할 가능성도 높아지게 되는 것으로 유추할 수 있다.

한편, 흡연과 음주를 모두 하지 않으면서 어머니와의 관계도 긍정적인 경우 가출 경험은 약 5.7%까지 감소하여 이 집단의 94.3%는 가출을 하지 않았다. 이는 본 연구에서 가출 경험이

없는 비율이 가장 높은 집단으로, '비흡연'과 '비음주', 그리고 '어머니와의 긍정적인 관계'는 가출을 예방할 수 있는 최적의 조합관계인 것으로 나타났다. 특히, 흡연여부와 음주여부가 어머니와의 관계보다 상위의 변인으로 나타난 것은 의미 있는 결과이다. 항상 가출에 대한 예방책으로 가정문제를 먼저 떠올리고 그것에 집중하기 쉽지만, 이 결과에 따르면 사실 개인의 비행요소가 가출을 유발하는 더 큰 문제이므로 이에 대한 개입이 먼저 이루어져야 할 것이다.

지금까지 가출에 대한 예측변인을 살펴본 연구들은 대부분 개별적인 변인들 간 상대적 영향력 정도를 검증하는 것이 한계였다. 그러나, 본 연구에서는 데이터 마이닝의 의사결정 나무분석을 적용하여 이러한 한계점을 극복하고, 다양한 변인들 간의 수많은 상호작용의 조합관계를 모두 분석하였다. 이를 통해 실업계 고등학생의 가출에 대한 이해를 보다 명확하고 심도 깊게 할 수 있었으며, 다양한 상황과 특성을 가진 청소년들에게 각각 '맞춤형'의 개입방안을 제시하였다는 의의를 가진다.

본 연구의 제한점과 후속연구에 대한 제언을 하자면 다음과 같다.

첫째, 본 연구에서는 실업계 고등학생만을 대상으로 가출에 대한 예측모형을 구축하였다. 따라서, 후속연구에서는 인문계 고등학생을 대상으로 분석하여 그들의 가출에 대한 예측모형은 실업계 고등학생과 어떠한 차이가 나타나는지 비교해 볼 수 있을 것이다.

둘째, 본 연구에서는 가출에 대하여 '예' 또는 '아니오'로 응답하는 경험의 여부를 묻는 문항을 사용하였다. 이는 가출의 심각성 정도는 고려하지 못했다는 제한점을 가진다. 따라서, 후속연구에서는 첫가출-재가출, 가출의 빈도, 가출의 기간 등에 대해 보다 세밀히 분석해 볼 필요가 있겠다.

셋째, 본 연구에서는 2차 자료의 한계로 인해 가출한 친구와 관련된 변인은 추가하지 못하였다. 청소년들은 또래집단에 민감한 영향을 받는 특성을 가지고 있으므로, 가출 친구의 유무에 따라라도 가출 경험이 달라질 가능성이 있다. 따라서, 후속연구에서는 가출 친구 변인을 추가해 볼 것이 권장된다.

넷째, 본 연구에서는 KEEP의 2004년에 시행된 1차년도 자료에 근거하여 분석하였다. 이는 현 시점과 다소 차이가 있다는 한계를 가진다. 그럼에도 불구하고 본 연구에서 2004년도 데이터를 사용한 이유는, 연구대상자들이 2004년도에 고등학교 3학년에 재학중인 관계로 이후의 데이터들에서는 성인이 되었기 때문이다. 이는 청소년의 가출 예측모형을 살펴보고자 한 본 연구의 목적에 부합되지 않으므로 2004년도 자료를 사용하였다. 따라서, 후속 연구에서는 보다 최근의 데이터들을 대상으로 분석해 볼 필요가 있겠다.

## 【참 고 문 헌】

- 권윤아 · 김득성(1999). 실업계 여고생의 가출영향변인들 및 가출유형 분석-개인 · 가족 · 또래 및 학교 변인을 중심으로. **한국가정관리학회지**, 17(1), 1-13.
- 김경희 · 김지수(2007). 고등학생의 가출에 영향을 미치는 위험요인. **아동간호학회지**, 13(3), 338-348.
- 김경희 · 김지수 · 김미진 · 정승교(2007). 청소년의 자아개념, 자각추구측성 및 가출 경험에 관한 연구. **정신간호학회지**, 16(4), 348-356.
- 배문조 · 전귀연(2002). 청소년의 가출충동과 관련된 특성 연구: 개인 · 가족 · 학교환경 · 또래관계를 중심으로. **대한가정학회지**, 40(1), 23-35.
- 안창규 · 문선화 · 전윤식(1995). **가출청소년과 학교관리체계**. 서울: 집문당.
- 이종화 · 김경희 · 김희영 · 정혜경(2006). 여자 고등학생의 가출 경험에 영향을 미치는 예측요인.
- 임희진 · 유제민(2007). 청소년 진로상황의 불확실성에 대한 보호요인 및 위험요인의 탐색. **제 4회 한국청소년패널 학술대회 논문집**, 613-638.
- 정혜경 · 권해진(2001). 청소년 가출에 관한 예측모형. **정신간호학회지**, 10(1), 76-86.
- 정혜경 · 안옥희(2001). 청소년의 가출충동에 영향을 미치는 예측요인. **아동간호학회지**, 7(4), 483-493.
- 최종후 · 한상태 · 강현철 · 김은석 · 김미경 · 이성건(2003). **Answer Tree 3.0을 이용한 데이터마이닝 예측 및 활용**. 서울:한나래.
- 하순인(1997). **사회적 환경과 개인적 특성이 청소년 가출에 미치는 영향 연구**. 연세대학교 대학원 석사학위논문.
- 한정훈(1999). **실업계 고등학생의 가출원인 분석**. 전남대학교 교육대학원 석사학위논문.
- 한국직업능력개발원(2006a). 한국교육고용패널 1차(2004)~3차(2006)년도 조사 사용자지침서. 한국직업능력개발원.
- 한국직업능력개발원(2006b). 한국교육고용패널 기초분석보고서(2005)-제1차(2004)년도 자료 분석. 한국직업능력개발원.
- De Man, A., Dolan, D., Pelletier, R., & Reid, C.(1993). Adolescent runways: Familial and personal correlates. *Social behavior and personality*, 21(2), 163-168.
- Englandera, S. W.(1984). Some self-reported correlates of runaway behavior in adolescent females. *Journal of Consulting and Clinical Psychology*, 52(3), 484-485.
- Menon, M., Tobin, D. D., Corby, B. C., Menon, M., & Hodges, E. V. E.(2007). The developmental costs of high self-esteem for antisocial children. *Child Development*, 78(6), 1627-1639.



- Khong, L. Y. L.(2009). Runaway youths in Singapore: Exploring demographics, motivations, and environments. *Children and Youth Services Review*, 31, 125-139.
- Martinez, R. J.(2006). *Understanding runaway teens*. *Journal of Child and Adolescent Psychiatric Nursing*, 19(2), 77-88.
- Plass, P. S. & Hotaling, G. T.(1995). The intergenerational transmission of running away: Childhood experience of the parents of runaways. *Journal of Youth and Adolescence*, 24(3), 335-348.
- Sanchez, R. P., Waller, M. W. & Greene, J. M.(2006). Who runs? A demographic profile of runaway youth in the United States. *Journal of Adolescent Health*, 39, 778-781.

- 접수 일 : 2008년 01월 13일
- 심사 일 : 2008년 01월 30일
- 심사완료일 : 2009년 03월 19일