

# Development of the Drop-outs Prediction Model for Intelligent Drop-outs Prevention System

Mi-Young Song\*

## Abstract

The student dropout prediction is an indispensable for many intelligent systems to measure the educational system and success rate of all university. Therefore, in this paper, we propose an intelligent dropout prediction system that minimizes the situation by adopting the proactive process through an effective model that predicts the students who are at risk of dropout.

In this paper, the main data sets for students dropout predictions was used as questionnaires and university information. The questionnaire was constructed based on theoretical and empirical grounds about factor affecting student's performance and causes of dropout. University Information included student grade, interviews, attendance in university life. Through these data sets, the proposed dropout prediction model techniques was classified into the risk group and the normal group using statistical methods and Naive Bays algorithm. And the intelligence dropout prediction system was constructed by applying the proposed dropout prediction model.

We expect the proposed study would be used effectively to reduce the students dropout in university.

▶ Keyword: Dropout Prediction Model, Naive Bays Algorithm

## I. Introduction

대학 학령 인구의 감소로 2018년부터 대학 입학정원이 고교 졸업자수를 초과하게 될 것으로 예상되며, 이는 대학의 충원율이 감소될 것으로 전망하고 있는 가운데, 일부 대학은 입학자원의 부족을 직접적으로 경험하고 있다. 이에 대학은 향후 지속적인 대학 학령 인구의 감소를 현실로 인식하고 확보된 자원인 입학생의 중도 탈락을 사전에 예방하면서 학생들을 지속적으로 충원하는 문제에 집중해야 한다[1].

이 뿐만 아니라 대학생의 중도탈락률은 해마다 증가하는 추세를 보이고 있어 문제가 심각한 상황이다. 교육통계연구센터의 2015 교육통계분석자료집(전문대학)을 통해 대학생의 중도탈락률을 살펴보면, 전문대학의 경우 2014학년도 전문대학의 중도탈락률은 전 학년도 대비 약 0.1%p 감소한 것으로 나타났지만, 중도탈락률은 1990학년도부터 2007학년도까지 증가세

를 보이다가, 2008학년도 이후로는 큰 변동없이 7%대를 유지해오고 있다.

연도별로 전문대학의 설립별 중도탈락률 변화를 [표1]에서 살펴보면, 2014학년도에는 국·공립 7.3%, 사립은 7.5%로 나타났다. 국·공립의 경우 1990학년도에 3.1%로 중도탈락률이 가장 낮게 나타났고, 2012학년도에는 15.1%로 가장 높게 나타났다. 사립 역시 국·공립과 마찬가지로 1990학년도에 2.2%로 중도탈락률이 가장 낮게 나타났으며, 2007학년도에는 8.3%로 가장 높게 나타났[2].

---

• First Author: Mi-Young Song, Corresponding Author: Mi-Young Song  
\*Mi-Young Song (songsnail@naver.com), Dept. of Mobile Media, Suwon Women's University.  
• Received: 2017. 09. 22, Revised: 2017. 10. 11, Accepted: 2017. 10. 19.  
• This study was conducted by the support of Paper Research Projects of Suwon Women's University in 2017

Table 1. Number of dropouts and dropout rates by year(as of 1990 ~ 2014year)[2]

(unit:person, %)

	all		national · public		private	
	number of dropouts	dropout rates	number of dropouts	dropout rates	number of dropouts	dropout rates
1990	7,224	2.2	837	3.1	6,387	2.2
1995	19,066	3.3	1,544	7.2	17,522	3.2
2000	50,496	5.5	1,725	4.6	48,771	5.6
2005	63,853	7.5	2,902	8.0	60,951	7.5
2006	67,323	8.2	3,462	10.3	63,861	8.1
2007	65,697	8.3	2,339	7.7	63,358	8.3
2008	59,958	7.8	2,276	8.5	57,682	7.7
2009	55,237	7.3	1,947	7.7	53,290	7.2
2010	56,658	7.4	1,316	6.1	55,342	7.4
2011	55,163	7.1	1,122	6.0	54,041	7.1
2012	58,923	7.7	2,533	15.1	56,390	7.5
2013	57,651	7.6	1,576	9.8	56,075	7.6
2014	55,239	7.5	1,124	7.3	54,115	7.5

그리고 2014학년도 지역별 중도탈락률 [표2]에서 살펴보면, 수도권 전문대학은 7.5%이고 비수도권 전문대학은 7.4%로 수도권 전문대학의 중도탈락률이 비수도권 전문대학보다 약간 높았다[2].

Table 2. Number of dropouts and dropout rates by region(as of 2014 year)[2]

(unit:person, %)

	all		national · public		private		
	number of dropouts	dropout rates	number of dropouts	dropout rates	number of dropouts	dropout rates	
total	55,239	7.5	1,124	7.3	54,115	7.5	
capital region	total	23,757	7.5	110	3.9	23,647	7.6
noncapital region	total	31,482	7.4	1,014	8.1	30,468	7.4

대학생 중도탈락률의 증가에 따른 재학생 수의 감소는 대학 재정 운영의 안정성을 위협하는 요인이 될 수 있으며, 대학 교육의 질 저하 및 1인당 교육비의 증가를 초래할 수 있다. 특히 우리나라 전체 대학의 약 80% 이상을 차지하고 있는 사립대학의 경우 재정의 대부분을 학생의 등록금에 의존하는 구조이기 때문에 학생의 높은 등록률은 곧 안정적 재정 구조를 의미한다[3]. 이는 교육부가 대학을 평가하는 지표인 재학생 충원율과 직결되는 것으로 대학 경쟁력을 좌우하는 중요한 비중을 차지한다.

중도탈락은 학생 개인에게 학업 지속의 교육기회를 상실하게 되는 부정적인 결과를 초래하지만, 대학 입장에서는 교육의 질적인 저하와 재정난을 가중시켜 대학의 존립까지 위협하게 된다. 따라서 대학은 중도탈락 문제를 해소시켜 학생 자원을 안정적으로 확보하고 유지하는 일이 무엇보다 급선무라 할 수 있다[2]

이와 같이 학생중도탈락 예측은 어떤 대학의 교육 질적 수준과 경쟁력을 측정하는 지능형 시스템 개발의 불가결한 요소로 중도탈락 상황을 최소화하기 위해서 가능성이 높은 중도탈락의

위험 학생을 사전에 예측할 수 있는 효과적인 방법을 개발하는 것이 필수적이다.

따라서 이 연구는 진단 속성에 대한 통계적 방법과 가우시안 나이브 베이즈 알고리즘을 이용하여 중도탈락생을 예측하는 것이다. 이 개발된 예측 모델을 이용하여 웹 기반 중도탈락 예방 시스템을 구축하였다. 구축된 시스템을 통해서 대학의 중도탈락 문제에 대하여 교육현장에서 학생별 중도 탈락 성향이 높은 주요 원인을 다양한 기준을 가지고 분석하여 예방차원의 적극적인 지도를 할 수 있는 기반을 마련할 것이다. 그리고 잠재적 중도탈락생을 대상으로 애로 사항을 해결하기 위한 교육현장 중심의 실질적인 상담과 교내 프로그램을 집중 진행하여 중도탈락률을 줄이기 위해 효과적으로 이용될 것으로 기대한다.

이 연구의 구성은 2장에서는 중도탈락의 원인 분석 및 요인, 예측 시스템에 대한 관련 연구를 알아보고, 3장에서는 중도탈락 예방 시스템 구축을 위한 중도탈락예측의 처리과정을 살펴본다. 4장에서는 구축된 웹 기반 중도탈락 예방시스템의 구성에 대해 살펴보고, 5장에서는 향후 진행에 대해 제시하며 결론을 맺는다.

## II. Research Background

중도탈락에 대한 연구는 대부분 중도탈락 예측 시스템 구축보다는 중도탈락의 원인 분석 및 요인 등에 관한 것이다.

### 1. Causal Analysis and Factors of Drop-out

[4]에서는 중도탈락의도를 결정하는 요인을 개인 관련 변인(인구사회학적, 학업관련 특성), 교육 기관 관련 변인(학문적, 환경적 요인), 사회적 변인(사회적 환경, 사회적 인식) 등으로 구분하여 분석하였고, 이는 중도탈락을 예방하거나 중도탈락률 예방을 위한 대처방안을 모색하기 위한 기초자료로 활용될 수 있는 연구 결과를 제시하였다.

[5]에서는 중도탈락 사례분석을 위해 중부 소재의 S대학의 재학생과 제적생을 대상으로 중도탈락에 영향을 미치는 변수를 분석하였다. 중도탈락과 관련된 개인적 영역, 학문적 영역, 사회적 영역, 환경적 영역의 요인을 독립 변인으로, 이를 중도탈락여부를 나타내는 이분 범주로 나누어 중속 변인으로 사용하였다. 여기서는 인적자원개발의 관점을 적용하여 고등교육의 중도탈락 현상을 분석하고 대안을 모색하였다.

[3]에서는 대학알리미에서 ‘대학’으로 분류된 학교 가운데 186개 대학을 대상으로 학생 개인적 변인과 대학 경험 변인을 제외한 대학 수준의 특성이 중도탈락에 미치는 영향을 분석하였다. 즉, 대학 평판도 및 교육·연구 활동, 대학 재정·교육비, 교육 여건 변인을 독립변수로 투입하여 위계적 회귀분석을 통해 분석하였다.

[6]에서는 중도탈락 위험이 높은 학생들의 개별 면담을 통해서 입학 전·후 중도탈락을 결정할 때까지 경험한 심리적·환경적 요인들을 심층적으로 분석하여 학습자 요인, 구조적 요인, 대학 경험 요인으로 구분하였다. 이러한 분석과정을 통해서 중도탈락의 모형은 지역여건 요인, 자아실현 욕구, 일과 학업을 병행하는 것 등으로 분류하였다. 그리고 중도탈락의 결정 과정의 모형 및 패턴을 ‘잠정적 이론’의 형태로 구체적인 사례와 함께 제시하였다.

[7]에서는 중도탈락의 요인을 학교, 경제, 가족, 개인적 요인 4가지로 나누었다. 학교 요인은 교사의 학생에 대한 관심부족, 비효율적인 처벌과 처벌의 불공평성, 무단결석, 학교처벌, 부정적인 학교경험 등, 경제 요인은 사회경제적 낮은 계층과 가난 등, 가족요인은 부모나 형제자매가 중도탈락의 유경험자이거나 가족의 낮은 교육목표 등, 개인 요인은 낮은 자아개념과 내외통제성 등이 중도탈락의 중요한 요인이라고 하였다.

[8]에서는 중도탈락의 주된 요인이 가정/가족, 학업흥미, 정서/정신건강, 대인관계와 관련된 문제들을 요인으로 제시하였다.

[9]에서는 대학 1학년을 대상으로 3년 후에 중도탈락여부를 조사하여 대학 학업지속 집단과 학업중단 집단 간의 개인, 가정, 대학환경 변인의 차이를 분석함으로써 대학 중도탈락생의 이동 경로 구조 분석의 기초를 제공하고자 하였다. 대학 중도탈락생의 학업중단 경로 구조 분석 결과, 대학기관이 대학 1학년의 학문적·사회적 통합성을 향상시키기 위해 적극적으로 개입한다면 학업중단 비율을 낮출 수 있는 효과를 가져 올 수 있음을 제안하였다.

[10]에서는 중도탈락을 미연에 방지할 수 있는 새로운 방안으로 입학전형별 중도탈락 실태를 파악한 후, 그 결과를 입학전형 방법과 절차에 반영함으로써 중도탈락을 개선하는 방법을 제시하였다. 또한, 입학전형별로 대학생할 만족도와 중도탈락률 간의 비교를 통해 중도탈락률을 개선하는 방법도 제시하였다.

## 2. Drop-Outs Prediction Model

다음은 중도탈락에 대한 예측 모형을 컴퓨터 알고리즘으로 접근한 연구이다.

[11]에서는 고등학교 중도탈락자를 대상으로 중도탈락의 원인이 학교, 가정, 사회, 개인 등의 변수를 고려하여 중도탈락에 예측모형을 구축하였다. 여기서 중도탈락예측모형은 의사결정트리와 연관규칙을 결합한 모형을 제안하였다. 제안된 모형을 통해서 교육현장에서 중도탈락예방을 위해 지도를 할 수 있는 기반을 제시하였다.

[12]에서는 중도탈락예측변수로 학교, 가족, 환경 등의 측정된 값을 사용하여 NAIVE BAYES와 J48 알고리즘의 정확도 메트릭을 사용하여 두 가지 분류 기법을 평가한 후 마지막으로 NAIVE BAYES 알고리즘으로 최상의 분류 알고리즘을 얻는다. 이 기법을 사용하여 사용 가능한 모든 속성을 활용하고 데이터 분류 및 예측에 널리 사용되는 데이터 마이닝 도구인 WEKA 도구를 통해 가장 좋은 속성을 선택하므로 높은 정확도로 어떤 학생이 중도탈락자인지를 예측하였다.

[13]에서는 의사결정트리 알고리즘을 이용하여 대학 신입생

의 사회-인구 통계학적, 학문적 및 제도적 데이터에서 관련 속성을 확인하고 학생이 학업을 지속하거나 중도탈락의 위험이 있는지를 자동으로 인식할 수 있는 프로토타입의 학습 도구를 설계하였다. 강력한 의사결정 도구를 위해 사회 인구 통계학적 데이터, 부모의 태도와 제도적 요인 등의 다양한 매개 변수를 이용하였다. 생성된 지식은 대학 등록 비율을 증가시키고 예방 및 자문 조치를 취하도록 하여 학생의 중도탈락을 줄이고 이에 관련 정책 및 전략을 개발하기 위해 교사와 대학의 관리에 매우 유용하게 활용하고 중도탈락 학생에 영향을 미치는 이유와 관련 요소를 찾는 데 활용하였다.

[14]에서는 지속적인 학습의 가능성이 있는 중도탈락생의 위험 그룹을 예측할 수 있는 인공지능 기술 중 하나인 FUZZY-ART MAP 신경망을 이용한 강력하고 혁신적인 잠재력을 가진 시스템 구축을 제안하였다. 제안된 시스템은 중도탈락할 가능성이 있는 학생들의 그룹에 대해 76%의 정확성과 신뢰성으로 조기 발견할 수 있음을 제시하였다.

[15]에서는 학생의 성과와 행동 사이의 관계를 분석하고 학생들의 학업 성취도 예측 모델을 개발하기 위해서 SVM(Support Vector Machine) 분류와 커널 위한 K-MAP 클러스터링 기법을 사용하는 데이터 마이닝을 위한 K-MAP 방법을 적용하였다. 이를 통해 학생의 성과를 예측함으로써 실패의 위험이 있는 학생들을 파악할 수 있으므로 경영기관은 적시에 도움을 제공하고 학생을 지도하여 성과를 향상시키는 데 도움을 주었다.

## III. Drop-Outs Prediction Process

이 연구는 진단설문 및 대학생활의 자료를 활용하여 통해서 학생의 중도탈락 가능성을 예측하여 분류한다.

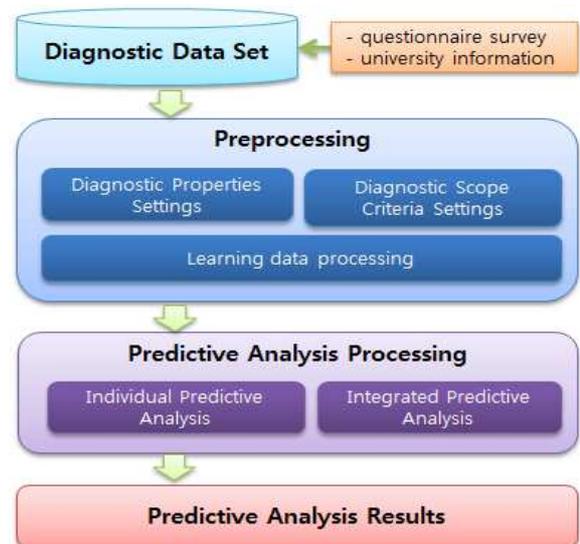


Fig. 1. Diagram of Dropout Prediction System

구축된 중도탈락 예방시스템에서 중도탈락생을 예측 분석하는 과정은 [그림1]과 같다. 중도탈락생의 예측분석을 위해 설문, 면담, 성적, 출석 등을 진단 속성으로 사용한다. 이러한 진단 속성별로부터 중도탈락의 가능성이 높은 위험군과 가능성이 낮은 정상군으로 분류하기 위한 중도탈락 가능성의 진단 기준 범위를 설정하고, 설정된 진단 기준 범위에 따라 진단 속성별 분석을 통해서 중도탈락의 가능성을 예측한다. 또한 해당 진단 기준 범위에 속하는 평균, 분산을 산출하고 이를 가우시안 나이브 베이즈를 적용하여 모든 진단 속성에 대한 종합적 분석을 통해서 중도탈락의 가능성을 예측한다.

### 1. Diagnostic Data Set

중도탈락생의 예측을 위한 진단 데이터는 [그림2]와 같이 진단설문자료와 대학생활자료를 활용한다.

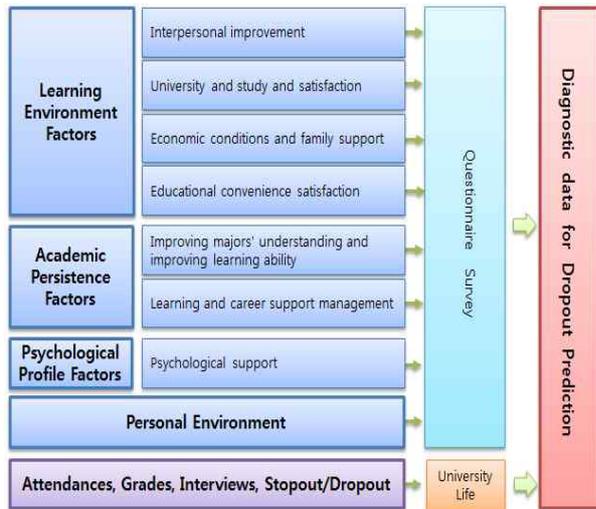


Fig. 2. Diagnostic Data Set of Dropout Prediction

#### 1.1 Questionnaires

진단설문자료는 학생들을 대상으로 45여개의 설문문항을 조사한다. 진단설문지는 [표3]과 같이 크게 학습 환경, 학업, 사회심리 등의 요인별 설문 내용과 개인 환경별 설문 내용으로 각각 45여개의 설문 문항으로 구성하여 5점 척도로 평가한다 [16]. 진단설문지에 대한 조사는 학기별로 이루어진다.

Table 3. Questionnaire by factor

Classification1	Classification2	Questionnaire Items
Learning environment factors	Interpersonal improvement	I stay with others smoothly. I hang out with my colleagues. I am satisfied with the professors' lectures. I think that the professors have an interest in the students. If I have a problem, I have the right to resolve it by consulting with the professor.
	University	I am satisfied with my admission to

	and study and satisfaction	my university. I am satisfied with our department. I think my aptitude and majors match I decided on the school and the department according to my hope I am currently worried about changing my career. I decided to pursue my college graduation career goals.
	Economic conditions and family support	I have financial difficulties in continuing my studies. I am considering taking a leave to prepare tuition and living expenses. My family supports my college life positively. I have a good relationship with my family.
	Educational convenience satisfaction	I have difficulty in commuting to school due to school distance and transportation. I am satisfied with my college education facility (classroom, laboratory, equipment, etc.) I am satisfied with my college life support service (scholarship, club activities, festivals, event support, etc.)
academic persistence factors	Improving majors' understanding and improving learning ability	I understand the contents of the professors' lectures. I have the basic knowledge I need to study my major. I work hard to prepare and prepare for assignments related to the lesson.
	Learning and career support management	I am in agreement with my major and my career plan. I do not feel late or absent in class. I am working hard to improve my academic performance. I am preparing for my career after graduation (qualification and language). I plan to take a separate language course (leave of absence) during the semester.
psychological profile factors	Psychological support	I try to set goals and execute them as planned. I do not give up easily even if I have difficulties. I have a sense of self esteem.

Table 4. Personal environment questionnaire

Classification	Contents
A high school region	City
A graduated high school	High school / general high school, industrial high school, commercial high school, agricultural high school.
Average grade in high school 3rd grade	90 points or more, 80 to 89 points, 70 to 79 points, 60 to 69 points or less, 59 points or less
Final academic goal	After graduating from a professional college, transferred to a four-year program, graduated from a domestic graduate school, Graduated from domestic graduate school, graduated from overseas studying master's and Ph.D.
Difficulty of study	Illness, economic circumstances, maladjustment, poor academic

	performance, aptitude inconsistency, language training, home conditions, Change of course, distance to school, etc.
Usual counselor	Friends (department friends), advisor (professor), parents, family members other than parents, etc.
housemate	Living with family, living with friends, living with friends, leaving alone, other
The cost of tuition and living expenses	Family full support (No.40), family and I together (No.39), Some relatives support(No.39), national support(No.39), others
What you need most now	Student loan, part-time job, scholarship, leave of absence or leave after work, etc.
Time required to attend school	Less than 30 minutes, less than 30 minutes to less than 1 hour, less than 1 hour to less than 1 hour 30 minutes, Less than 1 hour 30 minutes ~ 2 hours, more than 2 hours
Career planning	Employment in the field related to the major, employment in the major and other fields, Transfer to a 4-year college, college other departments, re-entry, etc.
A language training plan	(2 months), 6 months, 1 year, 1 year or more
Academic satisfaction	Very regretful, regretful, normal, satisfactory, very satisfied
Continued school life	After going on, I am going to graduate and leave school (No.45) I'm going to drop out (No.45)
Reason to leave school or drop out (2 priorities)	Dissatisfaction with majors and aptitude, lectures, dissatisfaction with educational level, dissatisfaction with school facilities, dissatisfaction with school facilities, economic difficulties, planning to advance to other universities, family problems, interpersonal problems,

1.2 University Information

대학생활자료는 학생의 출석, 성적, 면담, 휴학/자퇴 여부 등을 활용한다. 출석은 매주 주차별로 출결을 기록하고, 성적은 학기별 성적을 활용한다. 면담은 교수와의 면담을 통해 면담내용을 기록하고 진단 설문지의 요인별 분류와 같이 학생의 중도탈락 가능성여부를 면담자가 5점척도로 평가한 결과를 활용한다. 휴학/자퇴 여부도 면담과 같이 동일하게 활용한다.

2. Preprocessing

2.1 Diagnostic Properties and Scope Criteria Settings

진단 속성은 설문지, 면담, 성적, 출석을 선택하고 진단 속성별로 위험군, 정상군의 범위를 구분하는 진단 범위 기준으로  $R_Q, R_F, R_G, R_A$ 을 설정한다. 여기서 진단 속성과 진단 기준 범위는 학과/학부의 특성에 따라 중도탈락 예측 시스템에서 조정할 수 있다.

$$\{\text{설문지, 면담, 성적, 출석}\} = \{R_Q, R_F, R_G, R_A\} \quad (1)$$

식(1)은 진단 속성인 설문지, 면담, 성적, 출석에 대한 백분율의 진단범위 기준을 나타낸다.

2.2 Data processing by diagnostic properties

전체 학생에 대한 진단 속성별 결과를 산출하여 진단 속성별 범위기준에 따라 위험군과 정상군으로 분류한다. 그리고 각 분류에 대한 진단 속성별 평균( $\mu$ )과 분산( $\sigma^2$ )을 산출한다.

Table 5. Calculation formula for each diagnostic property

Property	Calculation Formula
Questionnaire	$Q_n = \sum_{i=1}^l qst_i$ (2) n : student number(1 to S), l : total number of questionnaires
Interview	$I_n = (\sum_{i=1}^m inv_i) / \text{횟수}$ (3) n : student number(1 to S), m : total number of interviews
Grade	$G_n = (\sum_{i=1}^p grd_i) / p$ (4) n : student number(1 to S), p : total number of semesters
Attendance	$A_n = \sum_{i=1}^q atd_i$ (5) n : student number(1 to S), q : total number of absences

위의 표(5)는 모든 학생에 대해 진단 속성별 결과치를 산출하는 식이다. 즉, 설문지 결과치( $Q_n$ )은 n번째 학생이 총 설문항목수에 대한 응답 점수의 합산이다. 면담 결과치( $I_n$ )은 n번째 학생이 총 면담횟수에 대한 면담 점수의 합산이다. 성적 결과치( $G_n$ )은 n번째 학생이 총 시험횟수에 대한 성적 점수의 합산이다. 출석 결과치( $A_n$ )은 n번째 학생이 총 출석횟수에 대한 결석횟수의 합산이다.

이와 같이 식(2~5)을 통해 산출된 전체 학생수에 대한 각 진단 속성별 결과치가 진단 속성별 진단 범위 기준인  $R_Q, R_i, R_g, R_A$  이내이면 위험군, 그 외는 정상군으로 분류하여 각 진단 속성별 위험군과 정상군의 평균과 분산을 각각 ( $\mu_{risk}, \sigma_{risk}$ ) ( $\mu_{normal}, \sigma_{normal}$ )을 산출한다.

3 Predictive Analysis Processing

이 연구에서는 중도탈락의 예측분석은 개별적 예측분석 및 통합적 예측분석을 통해 분석한다.

개별적 예측분석은 진단 속성별로 위험군과 정상군으로 분류한다. 또한 통합적 예측분석은 진단 속성에 대해 종합적인 예측분석을 위해서 가우시안 나이브 베이즈를 적용하여 위험군과 정상군으로 분류한다.

### 3.1 Individual Predictive Analysis

개별적 예측분석에서는 설문응답, 면담 성적, 출석 등에 대해 각각 중도탈락의 위험군과 정상군을 예측 분석한다.

설문응답에 대한 예측분석은 식(2)에서의 설문응답에 대해 문항별 5점 만점 척도를 합산한 결과에 대해 식(1)에서 설정한 진단 범위 기준  $R_Q$  이하인 경우 중도탈락의 위험군으로 판단한다.

$$\begin{aligned} C_{risk}(Q_n) &= PERCENTRANK(Q_1, \dots, Q_s, Q_n) \leq R_Q \\ C_{normal}(Q_n) &= PERCENTRANK(Q_1, \dots, Q_s, Q_n) > R_Q \end{aligned} \quad (6)$$

식(6)에서  $C_{risk}(Q_n)$ 과  $C_{normal}(Q_n)$ 은 설문응답에 대해 중도탈락여부를 판단하기 위한 진단대상자  $Q_n$ 에 대해 각각 중도탈락의 위험군과 정상군으로 예측한 결과이다.

여기서  $Q_1 \dots Q_s$ 은 설문응답에 대한 전체 학생의 집합이다.

$PERCENTRANK(Q_1, \dots, Q_s, Q_n)$ 은 진단대상자  $Q_n$ 이 전체 집합  $Q_1 \dots Q_s$ 에서의 백분율 순위를 나타낸다.

따라서 설문응답을 통한 진단대상자  $Q_n$ 이 전체 학생에 대해 백분율 순위가 진단 범위 기준  $R_Q$ 에 포함되는 경우 중도탈락 위험군으로 예측하고 그 외는 정상군으로 예측 판단한다.

면담에 대한 예측분석은 식(3)에서의 개별 면담에 대해 면담 횟수에 따른 문항별 5점 만점 척도를 합산한 평균으로 식(1)에서 설정한 진단 범위 기준  $R_I$  이하인 경우 중도탈락의 위험군으로 판단한다.

$$\begin{aligned} C_{risk}(I_n) &= PERCENTRANK(I_1, \dots, I_s, I_n) \leq R_I \\ C_{normal}(I_n) &= PERCENTRANK(I_1, \dots, I_s, I_n) > R_I \end{aligned} \quad (7)$$

식(7)에서  $C_{risk}(I_n)$ 과  $C_{normal}(I_n)$ 은 면담에 대해 중도탈락 여부를 판단하기 위한 진단대상자  $I_n$ 에 대해 각각 중도탈락의 위험군과 정상군으로 예측한 결과이다.

여기서  $I_1 \dots I_s$ 은 면담에 대한 전체 학생의 집합이다.

$PERCENTRANK(I_1, \dots, I_s, I_n)$ 은 진단대상자  $I_n$ 이 전체 집합  $I_1 \dots I_s$ 에서의 백분율 순위를 나타낸다.

따라서 면담을 통한 진단대상자  $I_n$ 이 전체 학생에 대해 백분율 순위가 진단 범위 기준  $R_I$ 에 포함되는 경우 중도탈락 위험군으로 예측하고 그 외는 정상군으로 예측 판단한다.

성적에 대한 예측분석은 식(4)에서의 개별 성적에 대해 모든 학기별 평균을 산출한 후 식(1)에서 설정한 진단 범위 기준  $R_G$

이하인 경우 중도탈락의 위험군으로 판단한다.

$$\begin{aligned} C_{risk}(G_n) &= PERCENTRANK(G_1, \dots, G_s, G_n) \leq R_G \\ C_{normal}(G_n) &= PERCENTRANK(G_1, \dots, G_s, G_n) > R_G \end{aligned} \quad (8)$$

식(8)에서  $C_{risk}(G_n)$ 과  $C_{normal}(G_n)$ 은 성적에 대해 중도탈락여부를 판단하기 위한 진단대상자  $G_n$ 에 대해 각각 중도탈락의 위험군과 정상군으로 예측한 결과이다.

여기서  $G_1 \dots G_s$ 은 성적에 대한 전체 학생의 집합이다.

$PERCENTRANK(G_1, \dots, G_s, G_n)$ 은 진단대상자  $G_n$ 이 전체 집합  $G_1 \dots G_s$ 에서의 백분율 순위를 나타낸다.

따라서 성적을 통한 진단대상자  $G_n$ 이 전체 학생에 대해 백분율 순위가 진단 범위 기준  $R_G$ 에 포함되는 경우 중도탈락 위험군으로 예측하고 그 외는 정상군으로 예측 판단한다.

출석에 대한 예측분석은 식(4)에서의 총 결석 횟수를 산출한 후 진단 범위 기준  $R_A$  이상인 경우는 중도탈락의 위험군으로 판단한다.

$$\begin{aligned} C_{risk}(A_n) &= A_n \geq R_A \\ C_{normal}(A_n) &= A_n < R_A \end{aligned} \quad (9)$$

식(9)에서  $C_{risk}(A_n)$ 과  $C_{normal}(A_n)$ 은 출석에 대해 중도탈락여부를 판단하기 위한 진단대상자  $A_n$ 에 대해 각각 중도탈락의 위험군과 정상군으로 예측한 결과이다.

따라서 출석을 통한 진단대상자  $A_n$ 이 진단 범위 기준  $R_A$ 에 포함되는 경우 중도탈락 위험군으로 예측하고 그 외는 정상군으로 예측 판단한다.

### 3.2 Integrated Predictive Analysis

통합적 예측분석에서는 1차적으로 식(9)와 같이 결석횟수가  $R_A$  이상이면 무조건 위험군으로 판단한다. 이는 기존 연구 논문[8]에서 중도탈락 여부를 분류하는 주요 변수는 결석일수로서 중도탈락에 가장 큰 영향을 미치는 것을 알 수 있음을 밝히고 있다.

그리고 2차적으로 설문응답, 면담, 성적의 3가지 진단 속성에 대해 가우시안 나이브 베이스(Naive Bays Algorithm)로 식(10)을 적용하여 위험군과 정상군을 예측 분석한다.

$$p(x = \nu|c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(\nu - \mu_c)^2}{2\sigma_c^2}} \quad (10)$$

식(10)에서  $\mu$ 와  $\sigma$ 는 각각 진단 속성별에 대한 위험군과 정

상군의 평균과 분산을 나타낸다.

중도탈락에 대한 예측 분석은 진단대상자에 대해 식(10)을 활용하여 사후 확률을 산출한다.

$$\begin{aligned}
 & \text{posterior}(C_{risk}) \\
 &= \frac{P(C_{risk})p(Q_n|C_{risk})p(I_n|C_{risk})p(G_n|C_{risk})}{evidence} \\
 & \text{posterior}(C_{normal}) \\
 &= \frac{P(C_{normal})p(Q_n|C_{normal})p(I_n|C_{normal})p(G_n|C_{normal})}{evidence}
 \end{aligned}
 \tag{11}$$

식(11)에서  $\text{posterior}(C_{risk})$ 와  $\text{posterior}(C_{normal})$ 은 각각 위험군과 정상군의 사후 확률을 나타낸다. 여기서  $P(C_{risk})$ 와  $P(C_{normal})$ 은 위험군과 정상군의 분포에 따라 조정 가능한 가중치 값이고  $evidence$ 는 정규화 상수로 식(12)와 같다.

$$\begin{aligned}
 evidence &= P(C_{normal})p(Q_n|C_{normal})p(I_n|C_{normal})p(G_n|C_{normal}) \\
 &+ P(C_{risk})p(Q_n|C_{risk})p(I_n|C_{risk})p(G_n|C_{risk})
 \end{aligned}
 \tag{12}$$

따라서 중도탈락에 대한 통합적 예측 분석은 진단대상자에 대해 식(11)의 사후 확률을 산출하여 이 중 더 높은 결과치에 따라 위험군과 정상군으로 예측 판단한다.

#### 4. Predictive Analysis Results

개별적 예측 분석 및 통합적 예측분석 처리에서 분류된 위험군 중에서 하위 1/2에 해당하는 그룹을 고위험군의 진단범위로 설정하고 예측 분석 처리단계에서의 처리방법과 동일하게 하여 중도탈락의 고위험군과 위험군을 예측한다.

따라서 중도탈락의 예측 분석 결과는 개별적 예측분석 처리 방법과 통합적 예측분석 처리 방법에 따라 고위험군, 위험군 그리고 정상군으로 분류한다.

### IV. Implementation of Drop-Out Prevention System

진단설문자료, 대학생활자료 등의 데이터를 통해서 진단속성인 설문지, 성적, 면담, 출석 등을 통해 중도탈락을 예측 분석하여 사전에 잠재적 중도탈락생들을 대상으로 상담/연계 프로그램을 진행하여 대학의 중도탈락률을 줄일 수 있도록 활용할 수 있는 웹 기반 중도탈락 예방 시스템을 구축하였다.

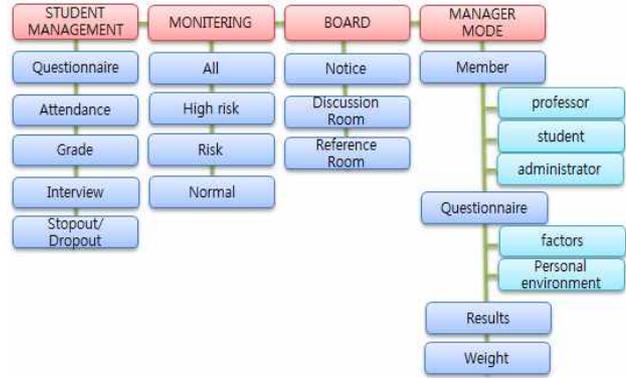


Fig. 3. Composition of Dropout Prevention System

웹 기반 중도탈락 예방시스템은 회원정보, 설문조사, 학생관리, 모니터링, 게시판, 관리자 메뉴로 구성하였다.

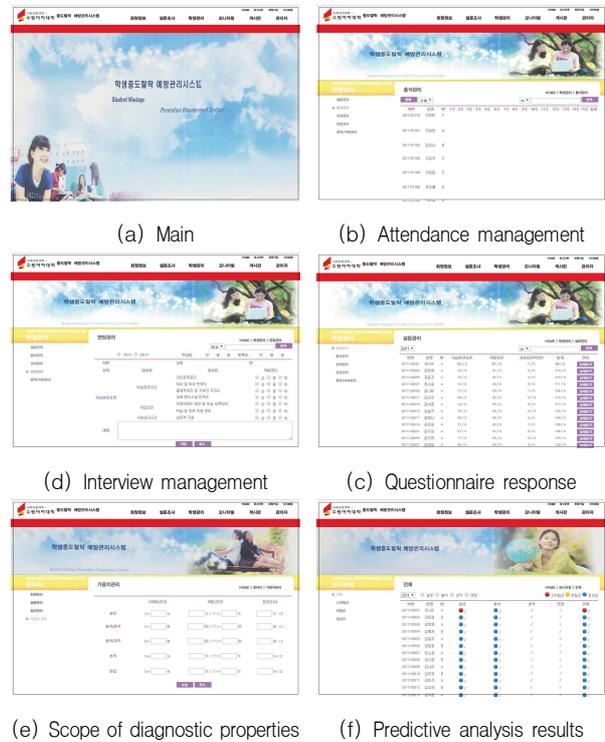


Fig. 4. Implementation screen of dropout prevention system

학생관리는 진단 속성인 설문지, 출석, 성적, 면담과 휴학/자퇴 등을 관리한다. 모니터링은 중도탈락의 예측 분석 처리과정을 통해 분류된 진단속성별 또는 전체적인 고위험군, 위험군, 정상군 등을 확인한다. 게시판은 학과/학부간의 공지사항, 토론펙, 자료실 등을 운영한다. 관리자는 회원관리(관리자, 교수, 학생)와 설문지 등록을 위한 설문관리(요인별설문, 개인환경설문) 및 진단속성별로 예측 분석된 결과관리, 진단속성별 진단범위 기준 설정을 위한 가중치 관리 등을 지원한다.

## V. Conclusions

학생 중도탈락 예측은 모든 대학의 교육 시스템과 성공률을 측정하는 데 없어서는 안 될 수많은 지능형 시스템의 필수 요소이다. 따라서 중도탈락의 사전 예방 처리를 채택하여 상황을 최소화 할 수 있도록 탈락 위험에 처한 학생들을 효과적으로 예측하는 방법을 개발하는 것이 필수적이다.

따라서 이 연구는 대학에서 중도탈락생을 사전에 예측하는 모형을 개발하였고 이를 통해 사용자가 쉽게 접근하여 학생의 중도탈락 가능성의 여부를 확인하고 분석할 수 있도록 웹 기반 중도탈락 예방 시스템을 구축하였다.

본 연구의 의의는 다음과 같이 요약할 수 있다.

첫째, 제안된 연구의 중도탈락 예측 모형을 활용하여 학생 대학 생활 및 학습 환경, 학업, 사회심리, 개인 환경 등의 학생별 중도탈락 요인을 다양하게 시각적으로 분석할 수 있도록 웹 기반 중도 탈락 예방 시스템을 구축하였다. 이를 통해서 중도탈락 원인을 분석하여 사전에 예방할 수 있는 방법을 제시하였다.

둘째, 대학의 전공별 학생 특성에 따라 중도 탈락의 주요 원인이 다를 수 있음에도 기존 연구는 학생 특성을 고려하지 않은 획일화된 기준에 따라 중도 탈락생을 분석하여 예측하고 있다. 그러나 제안된 본 연구는 중도탈락 진단 데이터의 속성별로 진단 범위를 조절할 수 있도록 지원하므로 전공별 학생 특성에 따라 중도 탈락생을 예측할 수 있다.

셋째, 이 제안된 연구의 분석된 결과를 통해서 학부/학과의 잠재적 중도탈락생들을 대상으로 애로 사항을 해결하기 위한 적극적인 상담과 교내 프로그램을 집중적으로 진행할 수 있도록 도움을 줄 수 있다.

향후에는 학생 중도탈락의 예측의 신뢰성을 높이기 위한 중도탈락의 다양한 요인별 데이터 수집에 대한 연구가 필요하다. 또한 제안된 예측 모형을 적용한 웹 기반 중도탈락 예방 시스템에 데이터 마이닝 기법을 적용하여 축적된 데이터를 통해서 학생별 중도탈락 성향이 높은 주요 원인의 가중치 부여 및 학생별 중도탈락 요인 간의 유사성 관계를 분석하는 연구가 이루어져 제안된 중도탈락 예측 모형의 정확성과 신뢰성을 높이고자 한다.

## REFERENCES

- [1] Soo-Yoen Kim, "A Study on Exploring the Academic Persistence and Dropout of Korean Universities and Colleges' Students", KEDI, The Journal of Korean Education, Vol.33, No.4, pp.33-62, 2006.
- [2] [2015 Educational Statistics Analysis data collection] College- II.student\_3. Student change situation 2)Drop out, Educational Statistics Research Center.
- [3] Jae-Young Chung, Mi-Sook Sun , Min-Ji Jeong , "An Analysis of Institutional Factors Affecting on College Dropout Rates", AJE, Vol.16, No.4, pp.57-76, 2015.
- [4] Gil-Soon Choi , Soon-Rye HAM "Factors Affecting College Students' Intention to Decide Dropout", KBR, Vol.3, No.1, pp.95-118, 2010.
- [5] Hye-lan Roh, Mi-na Choi, "A Study on Exploring the Dropouts in Korean University" KAHRD, Vol.11, No.1, pp.89-107, 2008.
- [6] Hyun Sim, "A grounded theory-based analysis on the factors that causing dropout of students in Korean national universities", Education Culture Research, Vol.23, No.2, pp.105-128, 2017.
- [7] Dupper, D. R., "Preventing school dropouts: Guidelines for school Social work in Practice", Social Work Education, 15(3), pp.141-149, 1993.
- [8] Brachman, R.J, and T. Anand, "Integrated Support for Data Archaeology", Proceeding of 1993 AAAI Workshop on Knowledge Discovery in Databases, 1993.
- [9] Soo-Yeon Kim , "A Study on College Freshmen's Adaptation Process Influencing on the Dropout - A Basic Research of the College Dropouts' Mobility Path and Structure", KSSCE, Career Education Research, Vol.25, No.3, pp.139-160, 2012.7.
- [10] Kil-Sung Choi, Yong-Chang Lee, "Measures to reduce Students' Withdrawal Rate : a case study on College D", Journal of KOCON, Vol.13, No.11, pp.979-987, 2013.
- [11] So-Young Chung, Soo-Tae Kwon, "Development of the Drop-outs Prediction Model Using Association Rule and Decision", Journal of KIIT, Vol.6, No.5, pp.202-210, 2008.
- [12] V.Vivekanandan1, A.Devipriya2, "Predition of student failure and Droptout Using Design Surppot System", KARPAGAM Journal of Engineering Research(KJER), Volume No.III, 2015.
- [13] ISweta Rai, IIPriyanka Saini, IIIAjit Kumar Jain, "Model for Prediction of Dropout Student Using ID3 Decision Tree Algorithm", IJARCSST 2014) Vol.2, pp.142-149, 2014.
- [14] Valquiria R. C. Martinho, Clodoaldo Nunes, Carlos Roberto Minussi, "Prediction of School Dropout Risk Group Using Neural Network", Proceedings of the 2013 Federated Conference on Computer Science and Information Systems, pp.111-114, 2013.
- [15] Mr. Bhushan S. Olokar, Prof. Ms. V.M.Deshmukh, "Application of Data Mining Technique for Prediction of Academic Performance of Student A Literature

- survey”, International Journal on Recent and Innovation Trends in Computing and Communication, ISSN:2321-8169, Vol.II, pp.3962 – 3965, 2014.
- [16] Seung-Kyu Lim, Jung Kim, Soo-Min Oh, “development of Prevention system and program for prevent students from dropping out”, Suwon Women's University Research Project, 2009.
- [17] Sung-Min Myoung, Hong-Ki Lee, “A Relationship of Learning flow and Dropout in Local University Students : The Mediating Effect of College Adaption”, Journal of The Korea Society of Computer and Information, Vol.20, No.6, pp.21-28, 2015.6.
- [18] Hyun-Sook Rhee, “An Intelligent Agent System using Multi-View Information Fusion”, Journal of the Korea Society of Computer and Information, Vol. 19(12), pp.11-19, 2014.12.

### Authors



Mi-Young Song received the M.S. and Ph.D. degrees in Computer Engineering from Dongguk University, Korea, in 1998 and 2004, respectively. Dr. Song is currently a Professor in the Department of Mobile Media, Suwon Women's University.

She is interested in Computer Games, Computer Graphics, Computer Vision, Image Processing.