

CRT 방법을 활용한 한국인의 고위험 음주 예측모형

변해원¹⁾

Prediction Modeling of High Risk Drinking in Korea Using CRT Method

Haewon Byeon¹⁾

요 약

이 연구는 서울시 복지패널 데이터를 이용하여 한국인의 고위험 음주를 예측할 수 있는 통계학적 모형을 개발하고 고위험 음주를 예방할 수 있는 기초자료를 제공하였다. 2010년 서울시 복지패널에 참여한 20세 이상 성인 3,304명(남 2,169명, 여 1,135명)을 분석하였다. 결과변수는 1회 음주량을 기준으로 남성은 소주 7잔(알코올 60g), 여성은 소주 5잔(알코올 40g) 이상으로 정의 한 고위험 음주여부로 설정하였고, 설명변수는 연령, 성, 최종학력, 현재 취업 상태, 가구 월 평균 총 소득, 배우자 유무, 흡연 여부, 주관적 건강상태, 정기적인 운동여부, 가족관계, 교우관계, 지난 1달 간 우울증상 여부, 지난 1년 간 금주 프로그램 참여여부, 현재 질병 여부로 설정하였다. 분석방법은 데이터마이닝의 CRT(Classification And Regression Tree) 알고리즘을 이용하였다. 한국인의 고위험 음주 예측 모형을 구축한 결과, 유의미한 예측변수는 성, 연령, 최종학력, 흡연 여부, 배우자 유무이었다. 본 예측 모형을 기반으로 고위험 음주 예측 집단의 건전한 음주 문화 정착을 위한 절주 프로그램 개발과 체계적인 관리가 필요하다.

핵심어 : 데이터마이닝, 문제음주, 위험요인, 예측모형, 의사결정나무모형

Abstract

The purpose of this study was to analyze the factors that affects the high risk drinking in Korean adults. Data were from the A Study on the Seoul Welfare Panel Study 2010. Subjects were 3,304 persons (2,169 male, 1,135 female) aged 19 and older living in the community. A prediction model was developed by the use of a CRT (Classification And Regression Tree) algorithm of data-mining approach. In the CRT algorithm analysis, gender, age, education level, smoking, marital status were significantly associated with high risk drinking in Korean adults.

Keywords : Classification And Regression Tree, Decision tree, Risk factor, High Risk Drinking

1. 서론

Received (March 05, 2015), Review Request(March 06, 2015), Review Result(March 24, 2015)

Accepted(April 14, 2015), Published(June 30, 2015)

¹506-706 Dept. Speech Language Pathology & Audiology, Univ., Chumdan 23, Gwansangu, Gwangju, Korea

email: byeon@nambu.ac.kr

국민건강증진에 대한 정부의 지속적인 정책에도 불구하고, 우리나라의 음주 소비량은 큰 변화가 없으며, 오히려 고위험 음주는 소폭 증가하는 추세이다. 2014년 식품의약품안전처가 만 15세 이상 남녀 2000명을 대상으로 발표한 주류 소비·섭취 설문조사 결과, 조사대상자 중 95.0%가 음주 경험이 있으며, 세계보건기구에서 정의하는 남성을 기준으로 1회 음주량이 소주 7잔(알코올 60g), 여성은 소주 5잔(알코올 40g) 이상 섭취하는 고위험 음주자는 2012년 68.2%에서 2013년 82.5%로 증가하였다[1]. 또한, 음주를 처음 접한 연령도 2012 평균 20.6세에서 2013년 평균 19.7세로 낮아졌으며, 음주 경험자 중에서 약 절반은 폭탄주를 경험한 것으로 보고되었다[1].

고위험 음주는 국민 건강을 위협하는 주요 건강위험행위이다. 만성적인 고위험 음주는 간경변의 주요 원인일 뿐만 아니라, 간암, 구강암, 식도암, 후두질환 등의 위험요인이다[2-6]. 최근에는 과도한 음주가 중추신경계를 자극하여, 인지장애를 유발한다는 결과도 보고되었다[7]. 또한, 과도한 음주는 자동차 사고, 폭행, 소아 학대 등의 사회적 문제의 원인으로 작용한다[2][8].

이 같은 음주로 인한 우리나라의 사회경제적 비용은 2004년을 기준으로 20조 990억원으로 GDP 대비 2.9%를 상회하는데[9], 이는 일본의 1.9%, 캐나다의 1.09%, 스코틀랜드의 1.19%와 비교했을 때, 최대 2배 이상 더 높다[10-12]. 따라서 고위험 음주에 영향을 미치는 관련 요인을 규명하고, 올바른 음주 문화를 형성하는 것은 국민 건강증진 측면에서 중요한 주제일 뿐만 아니라, 불필요한 사회적 지출을 줄이기 위해서도 중요하다.

현재까지, 한국인의 고위험 음주를 결정하는 요인으로 성, 직업, 연령, 경제수준, 학력, 음주에 대한 인식, 흡연, 우울증 등이 보고되었다[5][13][14]. 그럼에도 불구하고, 선행연구들은 주로 의료기관을 내원한 환자를 대상으로 수행되었거나, 사회경제학적 요인 등의 개별 요인만을 탐색한 결과이며, 다양한 요인을 심층적, 복합적으로 규명한 연구는 미흡한 실정이다. 특히, 한국인의 고위험 음주는 술을 권하는 것이 미덕으로 여겨지는 한국 사회의 음주문화와 무관하지 않기 때문에 인구사회학적 요인 외에도 사회적 환경, 경제적 환경, 건강수준, 건강행위 등 다양한 요인을 고려한 예측 모형의 구축이 필요하다.

이 연구는 한국인의 고위험 음주를 예측할 수 있는 데이터마이닝에 근거한 통계학적 모형을 개발하여, 고위험 음주를 예방할 수 있는 기초자료를 제공하였다.

2. 연구 방법

2.1 자료원 및 연구대상

이 연구의 자료원은 서울특별시에 거주하는 서울시민을 대상으로 서울시복지재단에서 조사한 2010년도 서울시복지패널조사(Seoul Welfare Panel Study)이다. 서울시복지패널조사는 서울시 거주 가구의 복지수준을 파악하고 복지취약계층의 실태파악 및 복지서비스 수요를 추정하기 위한 목적에서 2009년 통계청의 승인(제20113호)을 받아서 수행되었다[15]. 모집단은 2005년 인구주택총조사

대상가구 중 조사 시기를 기준으로 서울시 소재 가구이며, 표본추출방식은 서울시 25개 구를 대상으로 한 층화 2단 집락추출방법을 이용하였다. 조사항목은 소득, 경제수준, 건강, 생활여건, 복지서비스 수요 등으로 구성되었고, 조사방법은 면접원이 조사 대상 가구를 방문하여 휴대용 컴퓨터에 구조화된 설문에 따라 응답한 내용을 입력하는 컴퓨터를 이용한 대면면접조사(CAPI: Computer Assisted Personal Interviewing)방법을 이용하였다. 2010년도 서울시복지패널조사는 2010년 6월 1일부터 2010년 8월 31일까지 진행되었다.

이 연구에서는 서울시복지패널조사 완료자 7,761명 중에서 19세 이하 536명, 비음주자 2,906명, 지난 1년간 금주자 1,015명을 제외한 3,304명(남 2,169명, 여 1,135명)을 분석대상으로 하였다.

2.2 변수의 측정 및 정의

결과 변수는 지난 1년간 고위험 음주여부(예, 아니오)로 정의하였다. 고위험 음주는 세계보건기구의 기준을 참고하여 1회 음주량이 남성은 소주 7잔(알코올 60g), 여성은 소주 5잔(알코올 40g) 이상 섭취로 정의하였다.

설명변수는 연령(20대, 30대, 40대, 50대, 60대, 70대 이상), 성, 최종학력(초등학교 이하, 중학교, 고등학교, 대학졸업 이상), 현재 취업 상태(취업, 미취업), 가구 월 평균 총 소득(100만원 미만, 100-200만원, 200-300만원, 300-400만원, 400-500만원, 500만원 이상), 배우자 유무(배우자가 있고 함께 살고 있음, 배우자가 있으나 함께 살고 있지 않음, 배우자 없음), 흡연 여부(비흡연, 과거흡연, 현재 흡연), 주관적 건강상태(좋음, 보통, 나쁨), 정기적인 운동 여부(없음, 있음) 주관적 가족관계(좋음, 보통, 나쁨), 교우관계(좋음, 보통, 나쁨), 지난 1달 간 우울증상 여부(없음, 있음), 지난 1년간 금주 프로그램 참여여부(없음, 있음), 현재 순환기, 내분비계, 근골격계, 호흡기계, 이비인후 질환, 간질환, 비노기계질환 등 질병 여부(없음, 있음)를 포함하였다.

2.3 자료의 분석 방법

2.3.1 음주의 잠재적 관련 요인 탐색

일반적 특성은 기술 분석을 이용하여 평균 및 백분율을 제시하였고, 고위험 음주 여부에 따른 집단 간의 차이는 카이제곱검정(Chi-square test)으로 분석하였다. 이 때, 유의미수준 0.5 이하인 설명변수는 고위험 음주의 잠재적 관련 요인으로 규정하고 데이터마이닝 분석에 포함하였다[16].

2.3.2 CRT 알고리즘

CRT(Classification And Regression Tree)는 Breiman이 1984년에 제안한 데이터마이닝 알고리즘으로 지니 계수(Gini Index)를 이용하여 불순도(impurity)를 측정하며, 부모마디로부터 자식마디가 2개만 형성되는 이진분류(binary split)를 수행하는 알고리즘이다[17]. CRT는 생성되는 규칙을 해석하기 쉽고, 연속형 변수와 범주형 변수를 모두 이용할 수 있다는 장점이 있다. CRT 알고리즘을 이

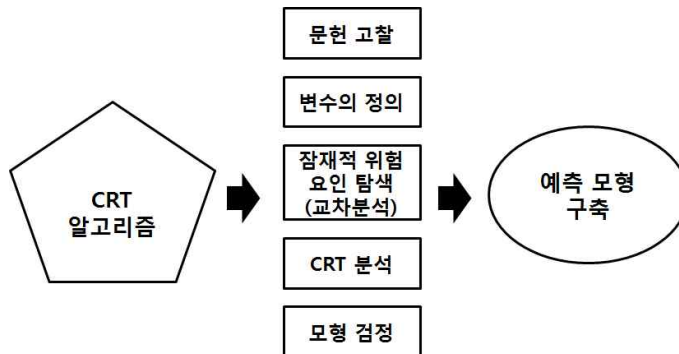
용한 연구 모형의 Flow Chart는 [그림 1]에 제시하였다.

지니 계수는 n 개의 원소 중에서 임의로 2개를 추출하였을 때, 추출된 2개가 서로 다른 그룹에 속할 수 있는 확률을 의미한다. 도출과정은 먼저, 각 마디에서 도수가 가장 많은 목표변수의 오분류 확률을 계산하며 통계식은 식(1)과 같다.

$$Gini\ Index(t) = 1 - \sum_j [P(j/t)]^2 \quad (1)$$

여기서 $P(j/t)$ 는 t 번째 자식마디에서의 j 변수의 상대도수를 의미하는데, 이진분리를 실시한 하위 마디에서 지니 계수를 각각 계산한 후 지니 계수가 최소화 되는 경우 최상의 분리가 수행되었다고 판단한다.

지니 계수의 감소량이 계산되면, 알고리즘의 마지막 과정으로 지니 계수를 가장 감소시켜 주는 예측 변수와 최적 분리를 자식 마디로 선택한다.



[그림 1] CRT 알고리즘을 이용한 연구 모형의 Flow Chart

[Fig. 1] Flow Chart for Prediction model

이 연구의 모형에서 CRT 알고리즘에 대한 의사결정규칙(decision rule)의 분리 및 병합 기준값은 0.05로 설정하였고, 부모마디의 수는 200명, 자식마디 수는 100명, 분지가 개수는 5개로 제한하였다.

생성된 모형의 타당성 평가는 K 배 교차 검증법(K -fold cross-validation method)을 이용하여 평가하였다[18]. 이 때, 모형의 신뢰도를 높일 수 있는 K 의 값은 10일 경우가 적절하다고 알려져 있기 때문에 이 연구에서도, K 의 값은 10으로 설정한 교차 검증법을 적용하였다[19].

모든 분석은 MINITAB version 13(Minitab Inc., State College, Pennsylvania, USA)과 Decision Tree version 20.0(IBM Inc., Chicago, Illinois, USA)을 이용하였다. 유의수준은 양측검정에서 0.05로 설정하였다.

3. 결과

3.1 연구대상의 일반적 특성과 고위험 음주의 잠재적 요인

고위험 음주 경험 여부에 따른 대상자의 일반적 특성 및 잠재적 요인은 [표 1]에 제시하였다. 전체 대상자 3,304명 중에서 고위험 음주 경험이 있는 대상자는 1,655명(50.1%) 이었다. 카이제곱검정 결과, 고위험 음주 경험이 있는 대상자와 고위험 음주 경험이 없는 대상자는 성, 연령, 최종학력, 현재 취업 상태, 현재 질병 여부에서 통계적으로 유의미한 차이가 있었다($p < 0.05$). 남성(63.9%), 20세 이상 40세 미만(52.8%), 대학 졸업 이상(52.1%), 현재 취업자(59.0%), 현재 순환기, 내분비계, 근골격계, 호흡기계, 이비인후 질환, 간질환, 비뇨기계질환 등의 질병이 있는 집단(51.7%)에서 고위험 음주 경험이 높았다.

[표 1] 고위험 음주 여부에 따른 대상자의 일반적 특성

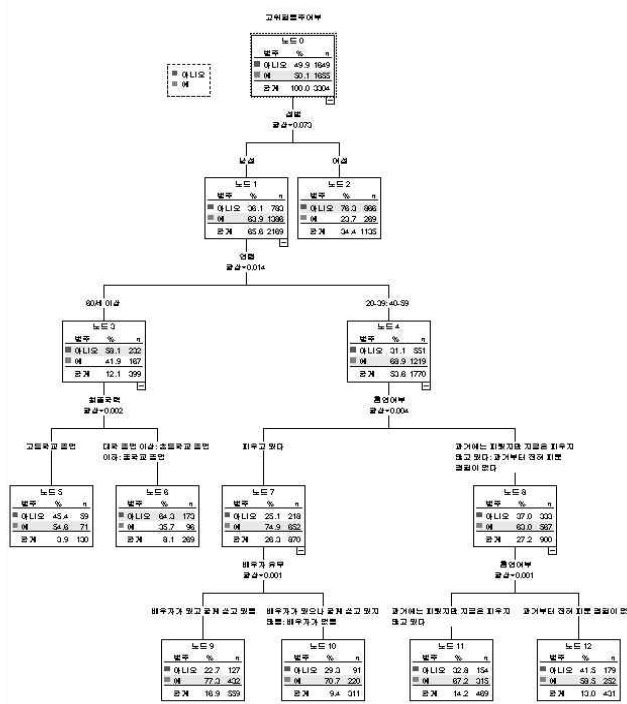
[Table 1] General characteristics of the subjects by high risk drinking, n (%)

Variables	High risk drinking		P
	No (n=1,649)	Yes (n=1,655)	
성별			<0.001
남	783 (36.1)	1,386 (63.9)	
여	866 (76.3)	269 (23.7)	
연령			<0.001
20-39세	720 (47.2)	806 (52.8)	
40-59세	611 (47.8)	668 (52.2)	
60세 이상	318 (63.7)	181 (36.3)	
최종학력			<0.001
초등학교 졸업이하	137 (65.6)	72 (34.4)	
중학교 졸업	134 (57.5)	99 (42.5)	
고등학교 졸업	481 (48.5)	510 (51.5)	
대학 졸업 이상	897 (47.9)	974 (52.1)	
현재 취업 상태			<0.001
취업	826 (41.0)	1190 (59.0)	
미취업	823 (63.9)	465 (36.1)	
가구 월 평균 총 소득			0.094
100만원 미만	160 (56.1)	125 (43.9)	
100만원-200만원	313 (51.5)	295 (48.5)	
200만원-300만원	305 (46.6)	349 (53.4)	
300만원-400만원	270 (47.9)	294 (52.1)	
400만원-500만원	219 (49.0)	228 (51.0)	
500만원 이상	382 (51.2)	364 (48.8)	
배우자 유무			0.562
동거	1140 (50.4)	1123 (49.6)	

별거	29 (53.7)	25 (46.3)	
없음	480 (48.6)	507 (51.4)	
흡연			<0.001
비흡연	1060 (67.8)	503 (32.2)	
과거흡연	295 (42.6)	398 (57.4)	
현재흡연	294 (28.1)	754 (71.9)	
주관적 건강상태			0.229
좋음	855 (49.7)	865 (50.3)	
보통	545 (48.7)	573 (51.3)	
나쁨	249 (53.4)	217 (46.6)	
정기적인 운동 여부			0.427
안한다	936 (49.3)	962 (50.7)	
한다	713 (50.7)	693 (49.3)	
지난 1년간 금주 프로그램 참여			0.689
없음	1638 (49.9)	1642 (50.1)	
있음	11 (45.8)	13 (54.2)	
주관적 가족관계			0.182
좋음	1166 (50.6)	1137 (49.4)	
보통	394 (48.9)	411 (51.1)	
나쁨	71 (43.6)	92 (56.4)	
주관적 교우관계			0.503
좋음	758 (50.9)	730 (49.1)	
보통	749 (49.3)	770 (50.7)	
나쁨	142 (47.8)	155 (52.2)	
지난 한 달 간 우울증상			0.260
없음	1457 (50.3)	1442 (49.7)	
있음	192 (47.3)	214 (52.7)	
현재 질병 여부			0.001
없음	422 (55.1)	344 (44.9)	
있음	1227 (48.3)	1311 (51.7)	

3.2 CRT 알고리즘을 이용한 고위험 음주 예측 모형

CRT 알고리즘을 이용한 고위험 음주 예측 모형은 [그림 2]에 제시하였다. 카이제곱 검정에서 고위험 음주의 잠재적 요인으로 설정된 변수들을 예측 모형에 포함한 후 CRT 알고리즘 분류모형을 구축한 결과, 유의미한 영향을 미치는 분류 변수는 성, 연령, 최종학력, 흡연 여부, 배우자 유무이었다. 가장 우선적으로 관여하는 예측 요인은 성별 이었다. 다음으로는 연령이 관여되는 분류 변수였다. 세 번째로는 20세 이상 60세 미만에서는 흡연여부가 분류 변수였고, 60세 이상에서는 최종학력이 관여되는 분류 변수이었다. 마지막으로, 현재 흡연자는 배우자 유무가 분류 변수이었다.



[그림 2] 한국인의 고위험 음주 예측 모형

[Fig. 2] Prediction model for high risk drinking in Korean adults

[표 2] CRT 알고리즘에 의한 이득 도표

[Table 2] Gains chart of predictor variable by CRT algorithm

Node no	Node n (%) ¹	Gain n (%) ²	Response % ³	Gain Index % ⁴	Characteristics
9	559 (16.9)	432 (26.1)	77.3	154.3	배우자와 함께 살고 있고, 현재 흡연을 하는 20세 이상 60세 미만의 남성
10	311 (9.4)	220 (13.3)	70.7	141.2	배우자가 없거나 별거중이고, 현재 흡연을 하는 20세 이상 60세 미만의 남성
11	469 (14.2)	315 (19.0)	67.2	134.1	과거 흡연 경험이 있는 20세 이상 60세 미만의 남성
12	431 (13.0)	252 (15.2)	58.5	116.7	평생 흡연 경험이 없는 20세 이상 60세 미만의 남성
5	130 (3.9)	71 (4.3)	54.6	109.2	최종 학력이 고등학교 졸업인 60세 이상 남성

¹ Node n(%); node number, % to 3,304

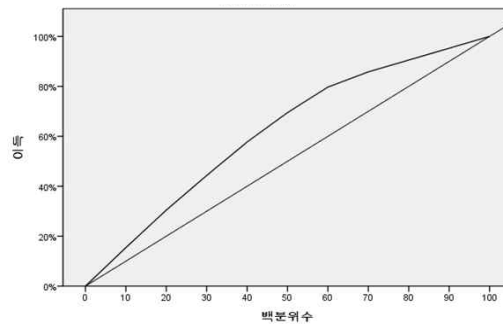
² Gain n(%); gain number, % to 1,655

³ Response (%): The fraction of the high risk drinking in Korean adults

⁴ Gain index (%):=154.3 in total 5 node

[표 2]는 고위험 음주 예측에 있어서 유의미한 경로를 이득율이 높은 순서 데로 제시한 CRT 알고리즘 기반 예측모형의 이익 도표이다. 마디번호는 최종마디의 번호이고, Gain Index (%)는 최종 노드에 대한 이익지표이다. 고위험 음주 예측 경로를 탐색하기 위해서 각 마디에 대한 이익지표를 도출했을 때, 유의미한 경로로 5개의 마디가 확인되었다. 먼저, 고위험 음주의 예측에 있어서 이익 지표 값이 가장 큰 제1경로는 배우자와 함께 살고 있고, 현재 흡연을 하는 20세 이상 60세 미만의 남성으로 이익지표는 154.3%이었다. 이익지표가 두 번째로 큰 제2경로는 배우자가 없거나 별거중이고, 현재 흡연을 하는 20세 이상 60세 미만의 남성으로 9.4%가 고위험 음주로 분류되었고, 이익 지표는 141.2%이었다. 제3경로는 과거 흡연 경험이 있는 20세 이상 60세 미만의 남성으로 14.2%가 고위험 음주로 분류되었고, 이익지표는 134.1%이었다. 제4경로는 평생 흡연 경험이 없는 20세 이상 60세 미만의 남성으로 13.0%가 고위험 음주로 분류되었고, 이익지표는 116.7%이었다. 마지막으로 제5경로는 최종 학력이 고등학교 졸업인 60세 이상 남성으로, 3.9%가 고위험 음주로 분류되었고, 이익지표는 109.2%이었다.

예측 모형의 분석이 완료되면, 개발된 예측 모형을 평가하기 위해서 10-fold 교차타당성 검정을 이용하였다. 도출된 모형의 안정성을 비교하기 위해서 10-fold 교차타당성 검정을 수행한 결과, 위험지수는 크로스 분류모형의 위험지수는 0.297, 오분류율은 30%로 도출되어, 예측모형의 위험지수 0.295 및 오분류율 30%와 동일하였다.



[그림 2] 최종 모형의 이득율
[Fig. 2] Gains percentile of final model

4. 결론

본 연구는 데이터마이닝의 CRT 알고리즘을 기반으로 한 한국인의 문제 음주 예측모형을 개발하였다. 통계학적 예측모형을 구축한 결과, 첫째, 배우자와 함께 살고 있고, 현재 흡연을 하는 20세 이상 60세 미만의 남성, 둘째, 배우자가 없거나 별거중이고, 현재 흡연을 하는 20세 이상 60세 미만의 남성, 셋째, 과거 흡연 경험이 있는 20세 이상 60세 미만의 남성, 넷째, 평생 흡연 경험이 없

는 20세 이상 60세 미만의 남성, 다섯째, 최종 학력이 고등학교 졸업인 60세 이상 남성은 고위험 음주의 가능성이 높은 집단이었다.

National Institute on Alcohol Abuse & Alcoholism[20]에 따르면, 과도한 음주는 알코올의 독성효과와 알코올 대사물질인 아세트알데하이드(acetaldehyde)로 인해 간경화 등의 원인이 될 수 있으며, 만성적인 음주 습관은 음주를 조절 할 수 없는 알코올의존을 초래할 수 있다.

따라서 본 연구에서 개발된 문제 음주 예측 모형을 근거로 건전한 음주 문화 정착을 위해서 고위험 음주 예측 집단을 대상으로 한 절주 프로그램 개발과 체계적인 보건교육이 요구된다.

References

- [1] Ministry of Food and Drug Safety, Study of Alcohol Consumption in Korea 2013, Ministry of Food and Drug Safety, O-song (2014).
- [2] J. Rehm and B. Fischer, Measuring harm: implications for alcohol epidemiology, In: M. Plant, E. Single and T. Stockwell (Eds) Alcohol: minimising the harm. Free Association Press, London (1997).
- [3] H. Byeon, The risk factors of laryngeal pathology in Korean adults using a decision tree model. Journal of Voice. (2015), Vol. 29, No. 1, pp.59-64.
- [4] H. Byeon, Exploring potential risk factors for benign vocal fold mucosal disorders using weighted logistic regression. International Journal of Bio-Science and Bio-Technology. (2014), Vol. 6, No. 4, pp.77-86.
- [5] H. Byeon, Y. Lee, Laryngeal pathologies in older Korean adults and their association with smoking and alcohol consumption. Laryngoscope. (2013), Vol. 123, No. 2, pp.429-433.
- [6] H. Byeon, Gender differences in risk factors of benign vocal fold disease in Korea: the fifth Korea National Health and Nutritional Examination Survey. Logopedics Phoniatrics Vocology. (2015), E-Pub, DOI: 10.3109/14015439.2015.1004365
- [7] H. Byeon, Y. Lee, S. Y. Lee, K. S. Lee, S. Y. Moon, H. Kim, C. H. Hong, S. J. Son and S. H. Choi, Association of alcohol drinking with verbal and visuospatial memory impairment in older adults: Clinical Research Center for Dementia of South Korea (CREDOS) study. Int Psychogeriatr. (2015), Vol. 27, No. 3, pp.455-461.
- [8] J. M. Shultz, D. P. Rice, D. L. Parker, R. A. Coodman, G. Stroh and N. Chalmers, Quantifying the disease impact of alcohol with ARDI software. Public Health Rep. (1991), Vol. 106, pp.443-450.
- [9] S. Lee, W. Chung, I. S. Kim, H. J. Kim, W. Cho, E. Shin, S. H. Ahn, K. H. Han and J. I. Myoung, Socioeconomic costs of alcohol drinking in Korea. J Korean Acad Fam Med. (2008), Vol.29, pp.201-212.
- [10] K. Nakamura, A. Tanaka and T. Takano, The social cost of alcohol abuse in Japan. J Stud Alcohol. (1993), Vol.54, No.5, pp.618-625.
- [11] E. Single, L. Robson, X. Xie and J. Rehm, The economic costs of alcohol, tobacco and illicit drugs in Canada, 1992. Addiction. (1998), Vol. 93, No. 7, pp.991-1006.

- [12] S. J. Varney and J. F. Guest, The annual societal cost of alcohol misuse in Scotland. *Pharmacoeconomics*. (2002), Vol. 20, No. 13, 891-907.
- [13] H. Ryu and M. J. Baek, Factors influencing the drinking behavior in female university students. *J Korean Acad Community Health Nurs* (2012), Vol. 23, No. 3, pp.307-315.
- [14] W. Jeong, The actual condition and the impact of psychosocial factors on problem drinking among the college students. *Korean Journal of Social Welfare* (2006), Vol. 58, No. 1, pp.347-372.
- [15] Seoul Welfare Foundation, Seoul Welfare Panel Study 2010, Seoul Welfare Foundation, Seoul (2010).
- [16] H. Byeon, The prediction model for self-reported voice problem using a decision tree model. *Journal of the Korea Academia-Industrial cooperation Society*. (2013), Vol. 14, No. 7, pp.3368-3373.
- [17] L. Brieman, J. Friedman, R. A. Olshen, C. J. Stone, *Classification and Regression Trees*, Chapman & Hall, New York (1984).
- [18] H. Byeon, The factors that affects the experience of discrimination in children in multi-cultural families using QUEST algorithm : focusing on Korean language education, *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*. (2014), Vol. 4, No. 2, pp.303-312.
- [19] L. Breiman, Bagging Predictors. *Machine Learning*. (1996), Vol. 24, No. 2, pp.123-140.
- [20] L. Gunzerath, V. Faden, S. Zakhari, and K. Warren, National Institute on Alcohol Abuse and Alcoholism report on moderate drinking. *Alcoholism: Clinical and Experimental Research*. (2004), Vol. 28, No. 6, pp.829-847.