

## 환자이탈군 특성요인과 이탈환자 예측모형에 관한 연구 -데이터마이닝을 활용하여-

김광환<sup>1\*</sup>

<sup>1</sup>건양대학교 병원관리학과

### A Study on the Characteristics of Prematurely Discharged Patients and Establishing a Model for Predicting Prematurely Discharged Patients -Using Data Mining-

Kim, Kwang Hwan<sup>1\*</sup>

<sup>1</sup>Department of Hospital Management Konyang University

**요 약** 본 연구는 1999년 7월 1일부터 2000년 6월 30일까지 1년 동안 일개 대학병원의 퇴원환자 전체를 조사대상으로 의료기관내에 생성된 의무기록 정보데이터를 활용하여 환자 이탈군 특성요인을 확인하고 이탈환자 예측모형을 구축하고자 하였다. 연구결과 회귀분석모델이 이탈환자 방지모형 적용에 가장 우수 하였으며, 이를 이탈환자에 적용하였을 경우 입원경로가 외래인 경우는 응급실 입원보다 이탈할 가능성이 적게 나타났고, 내원사유는 질병에 비하여 외상환자가 이탈할 가능성이 높게 나타났다. 아울러 Threshold 0.7 기준에서 이탈환자를 예측하면 예상 이탈환자 920명 중에서 실제로 이탈한 환자는 136명이 되어 검출효율은 14.78%로 나타났다. 이것은 Lift value 측면에서 보면 무작위 추출에 비해서 2.9배(14.78/5.15)의 효과가 있음을 알 수 있다.

**Abstract** This research was based on the purpose of establishing predicted model of prematurely discharged patients using the mandatory information data, recorded in the medical institutes based on discharged patients of a University Hospital for the period of 1 year, from July.

The result showed that the regression analysis model was the most excellent method of application model for preventing discharged patients, and when this is applied to discharged patients who are outpatients, the possibility of discharge can be less than staying in the emergency room. In addition, based on threshold 0.7, when we expect the discharged patients, out of 920 discharged patients, the actual patients who are discharged can become 136, showing the extract effectiveness of 14.78%. Based on the perspective of lift value, compared to random extract, this is 2.9 times (14.78/5.15) more effective.

**Key Words** : Discharged Patient, Data mining, Prematurely Discharged Patients

### 1. 서론

최근 의료 기관에서는 퇴원요약정보[1-3], 즉 의료기관의 관심사와 필요사항을 중심으로 추출된 정보 항목들을 질병역학, 보건정책 및 병원마케팅 등에 다양하게 활용하고 있다. 이로 인한 대량의 데이터를 포함하고 있는 정

보시스템을 구축, 관리할 수 있는 데이터웨어하우스(data warehouse)가 각광 받고 있다[4]. 이와 함께 방대한 양의 데이터로부터 함축적이며 잠재적 유용성이 있는 정보가 더욱 증대되고 있다[5]. 특히 의료기관 방문 환자들을 대상으로 한 라이프스타일 분석은 효율적인 홍보 전략수립 등을 위한 마케팅 도구로서 의원보다는 병원이나 종합병

\*교신저자 : 김광환(kkh@konyang.ac.kr)

접수일 09년 09월 18일

수정일 09년 10월 17일

게재확정일 09년 11월 12일

원에서 사용할 수 있을 것이라고 밝혔다. 환자 이탈의 원인은 환자들이 진정으로 원하는 가치를 그들이 원하는 방식으로 제공하지 못했기 때문이다. 이로부터 병원의 문제점을 파악하기 위해서는 불만을 표시하도록 기회를 제공해야 하며, 이러한 불만을 수렴해서 해결해 주어야 환자 이탈을 막을 수 있다. 극단적으로 표현해 환자와의 관계 형성에 실패한 병원은 성공할 수 없다. 환자중심의 마케팅이란, 환자가치와 환자만족을 창조하고자 하는 것으로, 환자들의 욕구를 이해하고 이를 만족시키는 것을 마케팅의 최우선 목표로 삼고 있다[6]. 이러한 상황에서 새로운 홍보수단인 인터넷 마케팅의 특징이자 장점은 시간과 공간을 초월해 고객(의료소지자)과 만난다는 것이다. 병원은 DB마케팅을 위해 병원이 우선 해야 할 일은 정보의 확보와 체계적인 정리일 것이다. 특히 대형병원의 경우 DB를 활용하려면 산재해 있는 데이터를 체계화하는 작업이 필요하다. 그리고 병원이 추구하는 환자 관리에 맞는 솔루션을 구축해야 한다[7-9]. 이와 같은 맥락에서 환자의 불만을 초래하는 이탈환자의 이탈이유를 살펴볼 필요가 있다. 따라서 이 연구에서 의료기관내에 생성된 의무기록 정보데이터를 활용하여 환자이탈군 특성요인을 확인하고 이를 통해 이탈환자 예측모형을 구축하고자 하였다.

## 2. 연구대상 및 방법

### 2.1 연구대상

1999년 7월 1일부터 2000년 6월 30일까지 1년 동안 일개 대학병원의 퇴원환자 전체를 조사대상으로 하였다. 조사대상 22,873명 중 내과계 10,375명 45.4%, 외과계 12,498명 54.6%로 내과계보다 외과계가 많았으며, 과별로는 내과가 27.5%로 가장 많았으며, 소아과 10.5%, 산부인과 10.4%순이었으며, 마취과가 0.1%로 가장 적었다.

### 2.2 연구방법

수집정보 항목은 의무기록과의 퇴원요약정보[1-3] 중, 미국의 UHDDS[10]퇴원요약 정보와 공동으로 설정하고 있는 항목으로 하였다.

일반적 특성에 관련된 변수로는 성별, 연령, 보험유형, 과별, 지역별 5항목으로 구성하였다. 입원상태에 관련된 변수로는 입원경로, 질병발생요인, 내원형태 3항목으로 구성하였다. 진료형태에 관련된 변수로는 협의진료유무, 수술유무, 전과유무, 주진단명 21대분류, 주진단명 질병

15대 순위 5항목으로 구성하였다. 퇴원상태에 관련된 변수로는 퇴원요일, 퇴원계절, 재원일수 3개 항목 등 총 16개 항목을 전산화된 주전산기에서 텍스트 형태의 일정한 형식을 정하여 download를 받았다. 이탈환자 중 주전산기에 없는 입원이유와 이탈이유의 항목은 전화통화로 자료를 수집하였으며, 이들 자료를 통계처리에 알맞게 SAS, Office tool을 이용하여 재편집하였다.

### 2.3 자료처리 및 분석

조사대상의 일반적 특성에 따른 정상퇴원환자와 이탈환자에 대한 유의성 검증은  $\chi^2$ -test 를 실시하였고, 환자 이탈방지 모형을 개발하고자 데이터 마이닝을 이용하여 다음과 같은 분석기법을 사용하였다. 의사결정나무(tree) 분석은 Tree 옵션에 제공되는 3가지 분석방법 중 Splitting criterion으로 Gini reduction 분석을 하였다. 다층인식자 신경망(MLP, neural network)분석은 E-Minor에서 사용할 수 있는 신경망으로 RBF(radial basis function, 원형기준함수) 신경망과 이를 확장한 EBF(elliptical basis function, 타원형기준함수)를 사용할 수 있으나 본 이탈환자 예측 모형에서 MLP 신경망을 사용하였다. 로지스틱 회귀분석(logistic regression)은 목표 변수가 이탈환자 여부(1,0)인 binary 형태를 취하고 있어 일반적으로 많이 사용되는 로지스틱 모형을 선택해서 단일 모형간에 성능을 비교하였다. 연구에 사용된 모든 자료처리와 분석은 통계소프트웨어 SAS(ver. 6.12)와 SAS의 데이터 마이닝 툴인 Enterprise Miner(ver. 3.01)를 이용하였다.

## 3. 연구결과

### 3.1 조사대상자의 일반적 특성

조사대상자의 일반적 특성은 표 1과 같이 총 22,873명 중, 정상퇴원환자 21,695(94.8%), 이탈환자 1,178( 5.2%)로 나타났다.

성별로는 남자 53.6%, 여자 46.4%로 여자보다 남자가 많았으며, 이탈환자를 보면, 남자 6.3%, 여자 3.9%로 여자보다 남자가 많았다( $p<0.001$ ).

연령별로는 60세 이상군에서 23.8%로 가장 많았고, 30세-44군 21.3%, 45세-59세군 19.0%순이었으며, 15세-29세군에서 17.0%로 가장 적었다. 이탈환자를 보면, 60세 이상군에서 7.7%로 가장 많았고, 45세-59세군 5.7%, 30세-44세군 5.1% 순이었으며, 14세 이하군에서 2.8%로 가장 적었다( $p<0.001$ ).

[표 1] 조사대상자의 일반적 특성

단위 :인원수(%)

일반적 특성 /		정상퇴원환자	이탈환자	합 계
성 별***	남자	11,482(93.7)	768( 6.3)	12,250(53.6)
	여자	10,213(96.1)	410( 3.9)	10,623(46.4)
연 령***	≤14	4,240(98.0)	86( 2.0)	4,326(18.9)
	15-29	3,715(95.4)	179( 4.6)	3,894(17.9)
	30-44	4,617(94.9)	249( 5.1)	4,866(21.3)
	45-59	4,090(94.3)	247( 5.7)	4337(19.0)
	60≤	5,033(92.3)	417( 7.7)	5,450(23.8)
보험유형***	보 험	18,762(95.4)	904( 4.6)	19,666(86.0)
	산 재	269(94.1)	17( 5.9)	286( 1.3)
	보 호	1,313(93.5)	92( 6.5)	1,405( 6.1)
	자 보	1,091(90.9)	109( 9.1)	1,200( 5.2)
	일 반	260(82.3)	56(17.7)	316( 1.4)
합 계		21,695(94.8)	1,178( 5.2)	22,873(100.0)

\*\*\* p<0.001

[표 2] 데이터 분할(Data Partition)

구 별	분석용데이터 (Training Data)	평가용데이터 (Validation Data)	검증용데이터 (Test data)	합 계
이탈환자	472(40.0) (50.0)	353(30.0) (50.0)	353(30.0) ( 5.2)	1,178(100.0) ( 5.2)
정상퇴원환자	472(40.0) (50.0)	353(30.0) (50.0)	6,501(30.0) (94.8)	21,695(100.0) (94.8)
합 계	944(200.0)	708(100.0)	6,854(100.0)	22,873(100.0)

보험유형별로는 보험이 86.0%로 가장 많았고, 의료보호와 자동차 보험이 각각 6.1%, 5.2%로 나타났으며, 산재 보험이 1.3%로 가장 적었다. 이탈환자를 보면, 일반이 17.7%로 가장 많았고, 자동차보험 9.1%, 의료보호 6.5% 순이었으며, 보험이 4.6%로 가장 적었다(p<0.001).

대적으로 적기 때문에 모형의 학습을 보장하였다. 분석용(training)데이터와 평가용(validation)데이터는 정상 비율을 50대50으로 하여 변수선택과 모형의 학습에 이탈환자의 특성을 보다 명확하게 반영시켰다. 또한 검증용(test)데이터는 선택된 모형의 검증 측면에서 전체 모집단과 동일한 이탈환자 비율(5.2%)로 랜덤 샘플링하였다.

### 3.2 이탈환자 데이터 마이닝을 이용한 스코어링 모형

#### 1) 모형평가

모형의 타당성을 평가하고[표 2], 여러 모형을 비교 검토 하고자 데이터를 분석용(training)데이터, 평가용(validation)데이터, 검증용(test)데이터로 분할하였다.

분석용(training)데이터를 이용하여 모형을 구축하고 평가용(validation)데이터와 검증용(test)데이터를 이용하여 모형의 비교 및 최종적인 평가를 수행하였다.

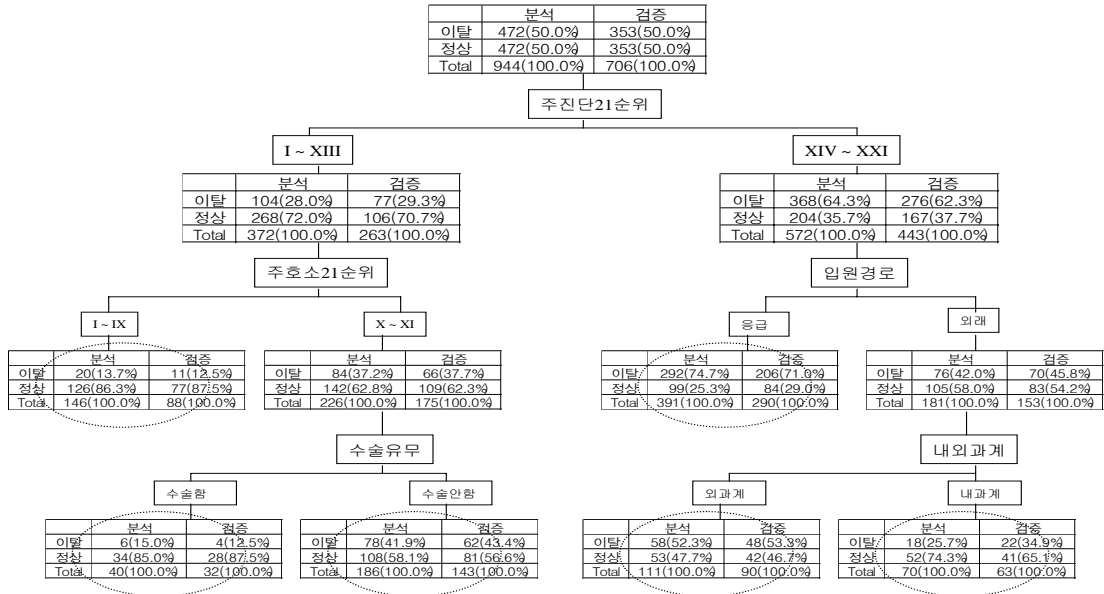
전체 데이터의 이탈환자 비율이 5.2%로 그 비율이 상

#### 2) 이탈환자의사결정나무분석(decision tree)

이탈환자 의사결정나무모형을 통해 검증한 결과[표 3] 한국표준사인 주 진단명 21대분류, 한국표준사인 주 호소증상 21대분류, 입원경로, 수술유무, 내·외과계에 따라 분류군이 결정되었다.

한국표준사인 주 진단명 21대분류군이 I(특정감염성 및 기생충성 질환)부터 VIII(귀 및 유양돌기의 질환) 일 때 이탈율이 29.3%, XVI(주산기에 기원한 특정병태)부터 XX I(건강상태 및 보건서비스) 일 때 이탈율이 64.3%였다 .

[표 3] 이탈환자의 의사결정나무 구조



- I: 특정감염성 및 기생충성 질환 II: 신생물 III: 혈액 및 조혈기관 질환 IV: 내분비, 영양 및 대사질환  
 V: 정신 및 행동장애 VI: 신경계의 질환 VII: 눈 및 눈부속기 질환 VIII: 귀 및 유양돌기 질환  
 IX: 순환기계 질환 X: 호흡기계 질환 XI: 소화기계 질환 XII: 피부 및 피하조직 질환  
 XIII: 근골격계 및 결합조직 질환 XIV: 비노생식기계 질환 XV: 임신, 출산 및 산욕  
 XVI: 주산기에 기원한 특정병태 XVII: 선천성기형, 변형 및 염색체이상 XVIII: 달리 분류되지 않은 증상  
 XIX: 손상 및 중독의인에 의한 특정 결과 XXI: 건강상태 및 보건서비스

3.3 이탈환자 데이터마이닝을 이용한 하이브리드(hybrid)모형

1) 최종 Lift값 모형에 대한 평가 및 선택

최종 Lift값 모형에 대한 평가 및 선택 표 4를 위해 전체 이탈환자를 모형화하고 그 결정요인을 분석하기 위해 의사결정나무분석(decision tree)+ 다층인식자 신경망분석(MLP, neural network) 하이브리드(hybrid)모형, 의사결정나무분석(decision tree) + 로지스틱 회귀분석(logistic regression) 하이브리드(hybrid)모형을 이용한 모형평가(assessment)를 한 결과 Lift 값 측면에서 상위 10%에서 이탈환자를 예측했을 때 로지스틱 회귀분석(logistic regression) 모형은 무작위 추출에 비해서 3.51배, 의사결정나무분석(decision tree) + 신경망분석(MLP, neural network)모형은 2.61 배, 의사결정나무분석(decision tree) + 로지스틱 회귀분석(logistic regression) 모형은 2.88배의 증가를 나타내 로지스틱 회귀분석(logistic regression) 모형이 이탈환자를 설명할 수 있는 가장 높게 나타났다.

[표 4] 최종모형 평가

모형종류 / 상위누적	10 %	20%
Logistic_Lift Value	3.51	2.71
Tree+MLP 신경망_Lift Value	2.61	2.65
Tree+Logistic_Lift Value	2.88	2.25

2) 최종 정분류율 모형에 대한 평가 및 선택

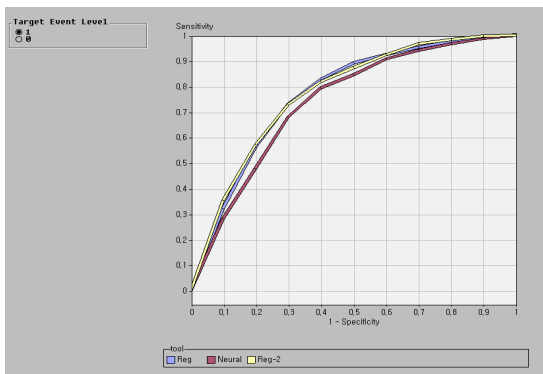
모형의 정분류율을 파악하기 위해[표 5] 의사결정나무 분석(decision tree) + 다층인식자 신경망분석(MLP, neural network) 하이브리드(hybrid)모형, 의사결정나무분석(decision tree) + 로지스틱 회귀분석(logistic regression) 하이브리드(hybrid)모형을 이용한 모형평가(assessment)를 한 결과 로지스틱 회귀분석(logistic regression) 모형이 77.5%로 설명력이 가장 높았다.

[표 5] 최종 이탈환자 정분류율표

모형종류 /	분석용 결과	검증용 결과
Logistic_Lift Value	77.5%	68.8%
Tree+MLP 신경망_Lift Value	76.3%	62.6%
Tree+Logistic_Lift Value	74.5%	67.1%

3) 최종ROC(receiver operation characteristic) 모형에 대한 평가 및 선택

최종 ROC(receiver operation characteristic) 모형에 대한 평가 및 선택 그림 1은 의사결정나무(decision tree), 다층인식자신경망(MLP, neural network) 및 로지스틱 회귀분석(logistic regression)모형간 성능비교 시 유용한 ROC Chart는 좌 상향일수록 우수한 모형이다. 따라서 위의 ROC Chart 결과는 로지스틱 회귀분석(logistic regression)모형이 전체적으로 가장 좌 상향에 위치하고 있어 가장 우수한 모형임을 나타내고 있다.



[그림 1] 최종모형 ROC Chart

3.4 최종 선택된 로지스틱 회귀분석(logistic regression) 이탈환자 예측모형 평가

선택된 최종 모형의 결과를 보면 표 6과 같다.

[표 6] 최종 선택된 로지스틱 회귀분석(logistic regression) 이탈환자 예측모형

Parameter	DF	Estimate	Standard Error	Ward Chi-square	Pr> Chi-square	Standardized Estimate	Exp(Est)
절편(Intercept)	1	-2.3477	0.4949	22.51	<.0001		0.096
입원경로 1(외래)	1	-0.4841	0.0826	34.34	<.0001		0.616
내원사유 1(교통사고)	1	-0.3567	0.1790	3.97	0.0463		0.700
내원사유 2(외상환자)	1	0.3768	0.1768	4.54	0.0330		1.458
보험유무 0(무)	1	0.5482	0.3124	3.08	0.0793		1.730
수술유무 0(무)	1	0.3873	0.0971	15.91	<.0001		1.473
주진21대순위	1	0.0829	0.0188	19.52	<.0001	0.274164	1.086
연령대	1	0.1676	0.0561	8.93	0.0028	0.131069	1.182

$$Y = - 2.3477 - 0.4841*(입원경로(외래)) - 0.3567*(내원사유(교통환자)) + 0.3768* (내원사유(외상환자)) + 0.5482*(보험유무(무)) + 0.3873*(수술유무(무)) + 0.0567*(주호21순위) + 0.0829*(주진21순위) + 0.1676*(연령대)$$

$$\text{예상 이탈환자 확률} = \text{EXP}(Y) / (1 + \text{exp}(Y))$$

즉, 입원경로가 외래인 경우는 응급실 입원보다 이탈할 가능성이 적게 나타났고, 내원사유는 질병에 비하여 외상환자가 이탈할 가능성이 높게 나타났으며, 교통 환자의 경우는 이탈할 가능성이 적은 것으로 나타났다.

보험유형별로는 비보험 환자가 보험환자 보다 이탈할 가능성이 높게 나타났으며, 외과계 환자의 수술유무에 따라서는 수술이 없는 경우가 수술이 있는 경우보다 이탈할 가능성이 높게 나타났다.

한국표준사인 주 주진단 21대 분포로는 XV(임신,출산 및 산욕), XVI(주산기에 기원한 특정병태), XVII(선천성기형, 변형 및 염색체이상), XVIII(달리분류되지 않은 증상), XX(손상 및 중독외인에 의한 특정 결과), XXI( 건강상태 및 보건서비스)에서 이탈할 가능성이 높은 것으로 나타났으며, 연령대별로는 연령이 높아질수록 이탈할 가능성이 높은 것으로 나타났다.

3.5 로지스틱 회귀분석(logistic regression) 모형에 Threshold별 적용한 결과

로지스틱 회귀분석(logistic regression) 모형에 Threshold별 적용한 결과 표 7과 같다. Threshold 0.7 기준에서 이탈환자를 예측하면 예상 이탈환자 920명 중에서 실제로 이탈한 환자는 136명이 되어 검출효율은 14.78%로 나타났다. 이것은 Lift value 측면에서 보면 무작위 추출에 비해서 2.9배(14.78/5.15)의 효과가 있음을 알 수 있다. 실제 이탈환자이면서 예상 이탈환자로 예측된 136명은 전체 이탈환자 353명 중에서 38.5%(검출력)으로 나타났다. 전체 환자에서 이탈환자의 비율 5.15%(이탈환자 353명 / 전체환자 6,854명)

[표 7] 로지스틱 회귀분석(logistic regression)모형에 Threshold별 적용한 결과

Threshold	검출력 (실제이탈환자/전체이탈환자)	검출효율 (실제이탈환자/예상이탈환자)
0.95	0.57%( 2/353)	66.67%( 2/3)
0.90	3.68%(13/353)	39.39%(13/33)
0.85	7.93%(28/353)	25.00%(28/112)
0.80	13.31%(47/353)	22.38%(47/210)
0.75	27.76%(98/353)	18.99%(98/516)
<b>0.70</b>	<b>38.53% (136/353)</b>	<b>14.78% (136/920)</b>
0.65	49.01%(173/353)	14.22%(173/1217)
0.60	56.94%(201/353)	13.09%(201/1535)
0.55	60.91%(215/353)	11.87%(215/1812)
0.50	72.52%(256/353)	11.20%(256/2286)

#### 4. 결론 및 토의

총 조사대상 22,873명 중, 정상퇴원환자 21,695명 이탈환자 1,178명으로 이탈율이 5.2%였다. 홍준현 등[11]은 재입원환자 중 이탈율이 3.1%로 연구되어, 본 조사보다 낮게 나타났고, 신순문[12]은 신생아 호흡곤란증후군에서는 이탈환자가 11.2%라고 하였으며, 이 등(1998)은 최근 1년간 21,652명의 응급환자 분석중 이탈환자 6.6%로 나타나 본 조사와 비슷한 경향을 보였다. 박미숙[13]은 이탈환자를 법적 성격으로 진료계약의 해지라고 하였으며, 해지란 계속적인 채권관계에 있어서 계약의 효력을 장래에 향하여 소멸하게 하는 일방적 행위라고 하였다. Threshold를 이용하여 모형평가(assessment) 한 결과 Lift 값 측면에서 상위 10%에서 이탈환자를 예측했을 때 의사결정나무(decision tree)모형은 2.33배 이탈환자를 추출 가능한 것으로 나타났다. 그러나 분석용 자료(training data)에만 의존하는 의사결정나무는 새로운 자료의 예측에서는 불안정(unstable)할 가능성이 높다. 이와 같은 현상은 분석용 자료의 크기가 너무 작은 경우와 너무 많은 가치를 가지는 의사결정나무를 얻는 경우에 빈번히 발생한다.

Threshold에 의한 단일모형을 이용한 모형평가(assessment) 한 결과 Lift 값 측면에서 상위 10%에서 이탈환자를 예측했을 때 로지스틱 회귀분석(logistic regression)모형은 무작위 추출에 비해서 3.51배, 다층인식자신경망(MLP, neural network)모형은 2.84 배, 의사결정나무(decision tree)모형은 2.33배의 증가를 나타나 로지스틱 회귀분석(logistic regression)모형이 이탈환자를 설명할 수 있는 가장 높은 기법으로 나타났다[14]. 또한 ROC(receiver operation characteristic)곡선은 우리가 구축한 모형의 성능을 민감도와 특이도에 의해 판단하고자

하는 곡선으로 의사결정나무(decision tree), 다층인식자신경망(MLP, neural network) 및 로지스틱 회귀분석(logistic regression)모형간 성능비교 시 유용한 ROC Chart는 좌 상향일수록 우수한 모형이다[15]. 따라서 위의 ROC Chart 결과는 로지스틱 회귀분석(logistic regression)모형이 전체적으로 가장 좌 상향에 위치하고 있어 가장 우수한 모형임을 알 수 있었다.

단순모형과 복합모형인 하이브리드 모형을 비교하여 최종모형을 선택하고자 하였다.

이탈환자 의사결정나무(decision tree) + 다층인식자신경망(MLP, neural network) 하이브리드(hybrid)모형의 모델링을 수행 했을 때, Threshold를 이용하여 모형평가(assessment)하였을 때 Lift 값 측면에서 상위 10%에서 이탈환자 예측결과 의사결정나무분석(decision tree + 신경망 분석(MLP, neural network)모형은 2.61 배의 이탈환자를 추출 가능한 것으로 나타났다.

이탈환자 의사결정나무(decision tree) + 로지스틱 회귀분석(logistic regression) 하이브리드(hybrid)모형의 모델링을 수행해 본 결과 최종 모형을 선택하고자 단일모형 중 가장 우수한 로지스틱 회귀분석(logistic regression)모형과 하이브리드 모형의 Threshold, ROC(receiver operation character)곡선으로 비교하여 최종 모형을 선택하였다.

오분류표를 통하여 단일모형에서 선택된 로지스틱 회귀분석(logistic regression)모형과 하이브리드(hybrid)모형을 모형평가(assessment) 한 결과, Threshold에 의한 단일모형에 얻은 로지스틱 회귀분석(logistic regression)모형과 하이브리드(hybrid)모형을 이용한 모형평가(assessment) 한 결과 Lift 값 측면에서 상위 10%에서 이탈환자를 예측했을 때 로지스틱 회귀분석(logistic

regression) 모형은 무작위 추출에 비해서 3.51배, 의사결정나무분석(decision tree + 신경망분석(MLP, neural network)모형은 2.61 배, 의사결정나무분석(decision tree) + 로지스틱 회귀분석(logistic regression) 모형은 2.88배의 증가를 나타내 로지스틱 회귀분석(logistic regression) 모형이 이탈환자를 설명할 수 있는 가장 높은 기법으로 나타났다. ROC Chart 결과 또한 로지스틱 회귀분석(logistic regression)모형이 전체적으로 가장 좌 상향에 위치하고 있어 가장 우수한 모형임으로 나타났다. 전체 환자 중 무작위로 이탈환자를 추출할 때 이탈환자를 뽑을 확률은 5.2%였으나, 최종 선택된 로지스틱 회귀분석(logistic regression)모형에 의해서 이탈환자를 예측 할 경우 threshold 0.7기준에서 이탈환자를 예측하면 예상 이탈환자 136명으로 추출에 비해서 14.78배의 이탈방지 효과가 있는 것으로 나타났다. 최종 선택된 로지스틱 회귀분석(logistic regression)에 의한 이탈요인으로는 입원경로가, 외래인 경우는 응급실 입원보다 이탈할 가능성이 적게 나타났다. 외래를 통한 입원환자가 많았는데 이는 외래 환자의 경우 전문의를 통한 입원이 예정된 환자들의 입원을 의미한다. 응급실에서는 급하게 입원하게 되며 이때 레지던트를 만나서 입원을 하게 되어[13] 응급상황이 종료되었거나 의료진에 대한 신뢰가 영향을 있을 것으로 생각된다.

### 참고문헌

[1] 김광환. 의무기록정보를 활용한 내과 질병통계에 관한 연구. *Journal of the Korean Data Analysis Society*. 9권 5호. pp 2233-2243. 2007.

[2] 김광환, 김용하, 한상태, 강현철. 의무기록정보를 활용한 48시간이내 사망환자 특성에 관한연구. *Journal of the Korean Data Analysis Society*. 9권4호. pp 1653-1659, 2007.

[3] 전은주, 김광환. 의료정보를 활용한 14세 이하 외상환자 응급실 총 경유시간에 미치는 요인. *병원경영학회지*. 13권4호. pp 27-45, 2008.

[4] 김광환, 한상태, 강현철. 보건정보를 활용한 산부인과 응급실 재원시간 실태에 관한 연구. *Journal of the Korean Data Analysis Society*. 8권3호. pp 989-996, 2006.

[5] 김광환. 환자이탈군 특성요인과 이탈환자 예측모형에 관한 연구. *계명대학교 공중보건학과 박사학위논문*, 2001.

[6] 김광환, 한상태, 강현철, 보건정보를 활용한 요일별 응급실 내원환자 특성에 관한 연구. *Journal of the Korean*

*Data Analysis Society*., 6권5호. pp 1400-1414, 2004.

[7] 강수원, 김미숙, 김옥남, 김재영, 민병옥, 서순원, 신종연, 이현실, 조은희, 의무기록정보관리학. *대한의무기록협회*. pp 258-259, 2007.

[8] Jeffrey A. Mello, *Strategic Human Resource Management*, 2ed., Thomson. pp. 348-349, 2006.

[9] 김광환, 한상태, 강현철, 손방용, 보건정보를 활용한 신환 환자 예약부도 분석. *Journal of the Korean Data Analysis Society*. 6권6호, 1817-1827, 2004.

[10] UHDDS. Uniform Hospital Discharge Data Set. Department of Health and Human Services. Health Information Policy Council, Washington, DC. 1984.

[11] 홍준현, 최귀숙, 이정화, 이은미, 진료정보 DB를 이용한 한 대학병원 환자들의 병원이용 패턴에 관한 연구 (병원이용 패턴에 관한 연구). *대한의료정보학회지*. 6권 4호, 23-33, 2000.

[12] 신손문, 신생아학 전문의 입장에서 자의퇴원. *제3회 신생아학 연속강좌 심포지엄*. 12-23, 1998.

[13] 박미숙, 일개대학병원의 자의퇴원환자 특성분석. *인제대학교 보건대학원석사논문*, 2000.

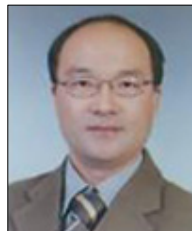
[14] 강현철, 한상태, 최종후, 김은석, 김미경. SAS Enterprise Miner를 이용한 데이터마닝방법론 및 활용. *자유아카데미*. 13-18, 1998.

[15] 최종후, 한상태, 강현철. SAS Enterprise Miner를 이용한 데이터마닝 기능과 사용법. *자유아카데미*. 33-45, 1999.

“상기 논문은 저자의 박사학위논문 데이터 재정리하여 논문 작성하였음”

### 김 광 환(Kwang-Hwan Kim)

[정회원]



- 2001년 2월 : 계명대학교 대학원 (보건학박사)
- 2006년 3월 ~ 현재 : 건양대학교병원 의무기록실장
- 2006년 3월 ~ 현재 : 건양대학교 병원관리과 조교수

<관심분야>  
의무기록정보, 보건관리, 병원관리