

인성 데이터를 활용한 조기 퇴사자 예측

김 영 박 · 김 형 중*

고려대학교 정보보호대학원 빅데이터 응용 및 보안학과

Predicting Early Retirees Using Personality Data

Young Park Kim · Hyoung Joong Kim*

Department of Big Data Application And Security, Korea University, Seoul 02841, Korea

[요 약]

본 연구는 기업에서 채용 전형 시 진행되는 인성시험 결과 데이터를 기반으로, 입사 3년 미만의 조기 퇴사자를 분석하였다. 예측 모형은 적합성 및 향후 활용성을 고려하여 제조(manufacture)직군과 R&D직군 2개 그룹으로 구분하여 분석하였으며, 독립변수 선택은 전진(stepwise)선택법에 따라 직군별로 유의미한 독립변수를 선택하였다. 예측 모형은 지도학습(supervised learning) 방법 중 로지스틱 회귀분석 알고리즘을 선택하였으며, 과잉적합(overfitting) 또는 과소적합(underfitting)을 방지하고자 교차 검증(cross validation)을 통해 예측 모형을 훈련시켰다. 혼동행렬(confusion matrix)을 통해 2개 그룹의 정확도(accuracy)를 확인하였으며, 조기 퇴직에 가장 영향을 많이 미치는 요인으로 제조직군에서는 ‘몰입’, R&D직군에서는 ‘반사회성’ 항목으로 확인되었다. 기존 퇴직 관련 연구는 설문 방식으로 데이터를 수집하고, 퇴직과 관련성이 높은 요인을 확인하는데 집중하였다면, 본 연구는 채용 전형 시 진행되는 인성 결과 분석을 통해 향후에도 지속 가능한 조기 퇴직 예측 모형을 제시했다는 면에서 의의를 갖는다.

[Abstract]

This study analyzed the early retired employees who stayed in company no longer than 3 years based on a certain company's personality evaluation result data. The predicted model was analyzed by dividing into two categories; the manufacture group and the R&D group. Independent variables were selected according to the stepwise method. A logistic regression model was selected as a prediction model among various supervised learning methods, and trained through cross-validation to prevent over-fitting or under-fitting. The accuracy of the two groups were confirmed by the confusion matrix. The most influential factor for early retirement in the manufacture group was revealed as “immersion,” and for the R&D group appeared as “antisocial.” In the past, people concentrated on collecting data by questionnaire and identifying factors that are highly related to the retirement, but this study suggests a sustainable early retirement prediction model in the future by analyzing the tangible outcome of the recruitment process.

색인어 : 로지스틱 회귀분석, 조기 퇴직, 교차 검증

Key word : Logistic regression, Early retirement, Cross validation

<http://dx.doi.org/10.9728/dcs.2018.19.1.141>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 27 October 2017; Revised 09 November 2017

Accepted 29 January 2018

*Corresponding Author; Hyoung Joong Kim

Tel: +82-02-3290-4895

E-mail: khj-@korea.ac.kr

I. 서론

산업의 쌀'로 불리는 반도체는 우리나라를 대표하는 수출 효과 산업으로, '17년 반도체 수출액은 단일품목으로는 사상 최고치인 900억 달러를 돌파할 것으로 예상된다[1]. 특히, AI(인공지능), IoT(사물인터넷), VR(가상현실), AR(증강현실), 자율주행차 등 4차 산업혁명을 대표하는 신기술들이 빠르게 성장하고 있어, 향후 반도체 수요는 더욱 증가할 것으로 전망되고 있다[2].

그러나 반도체 산업의 성장과는 다르게 업계에서는 인력수급이 원활하지 않다고 하는 우려의 목소리가 높아지고 있다. [2016년 산업기술인력수급통계조사결과]에 따르면, 반도체 분야 인력부족 현상이 여전히 개선되지 않는 상황에서(부족률 1.5%), 더욱 심각한 현상은 조기 퇴사율이(입사 1년 내) 38.2%나 되는 것으로 나타났다. 이런 조기 퇴직은 기업에서 비용과 조직문화 측면에서 부정적인 영향을 미친다. 신입사원의 재교육을 위해 소요되는 평균 기간은 19.5개월이며, 이를 비용으로 환산할 경우 직접비용과 간접비용을 포함하여 신입사원 1인당 총 6,088만 4천원을 쓰고 있다[3]. 또한 구성원의 이직은 해당조직 잔류 구성원에게 부정적인 영향을 주는 것으로 나타나는데 동료의 이직으로 인해 업무량 증가와 심리적 상실감에 따른 조직몰입의 감소로 인해 다른 조직구성원의 퇴사까지 유발할 수 있다[4]. 이렇게 구성원의 관리(retention)가 매우 중요함에도 불구하고 대부분의 기업들은 퇴직자에 대한 사전 감지 시스템이 미흡한 상황이며, 퇴직 결정 후 조치가 이뤄지고는 상황이다.

조기 퇴직에 대한 선행 연구들은 대부분 설문방식으로 데이터를 수집하였으며, 독립변수의 원인 규명을 목적으로 하여 향후 활용성에 대한 한계가 지적되었다. 본 연구는 인성결과 데이터를 기반으로 조기 퇴사자 예측모형 구축을 목적으로 한다. 인성시험은 입사 지원 시 반드시 거쳐야 하는 관문으로서 매년 많은 취업 준비생들이 지원한 기업의 인성시험에 응시하고 있으며[5], 본 연구는 A사의 채용전형 시 진행되는 인성 결과 데이터를 활용하였다. 수집된 데이터는 향후 활용성을 위해 제조(manufacture)직군과 R&D직군으로 나눠 로지스틱 회귀분석을 통해 직군별 분석을 진행하였으며, 추정 후 혼동행렬(confusion matrix)을 통해 정확도(accuracy)를 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 로지스틱 회귀 분석의 개요와 선행 연구를 검토하고, 3장과 4장에서는 선행연구들과의 차별성과 인성 데이터를 활용한 조기 퇴사자 예측 모형을 제안한다. 5장에서는 테스트를 통해 모형의 정확도를 확인하고, 마지막으로 6장에서는 연구결과와 향후 연구방향에 대해 논의하겠다.

II. 로지스틱 회귀분석 개요

로지스틱 회귀분석(logistic regression analysis)은 D. R. Cox가 1958년[6] 제안한 확률 모델로서 임의의 범위를 갖는 입력 값으로부터 0과 1사이의 값을 예측이나 분류할 때 적용하는 대표적인 통계기법이다. 로지스틱 회귀의 모델이 선형 모델과의 차이점은 종속 변수가 이항적(binary)이기 때문에 조건부 확률($P(y|x)$)의 분포가 이항 분포를 따른다는 점이며, 독립변수와 종속변수의 관계가 비선형, 즉 S곡선 형태를 이루게 된다.

모형 추정에 사용하는 방법은 최대 우도법(maximum-likelihood method)이며, 대표적인 검증 방법은 의사결정계수(R^2)로 모델 적합도 평가에 주로 쓰인다.

어떤 사건이 발생할 확률을 p 라고 하면, 독립변수와 상관없이 종속변수 범위가 항상 0과 1 사이에 있도록 식 (1)과 같이 오즈비(odds ratio)를 식 (2)와 같이 로짓(logit)으로

$$Odds\ ratio = \frac{p(y=1|x)}{1-p(y=1|x)} \quad (1)$$

$$Logit = \log \frac{p(y=1|x)}{1-p(y=1|x)} = a + bx \quad (2)$$

변환시킨다. 그 결과 독립변수 개수가 i 개인 로지스틱 회귀함수는 모형은 식 (3)와 같다.

$$P(y=1|x_1, x_2, \dots, x_i) = \frac{e^{a+b_1x_1+b_2x_2+\dots+b_ix_i}}{1+e^{a+b_1x_1+b_2x_2+\dots+b_ix_i}} \quad (3)$$

로지스틱 회귀분석을 통해 만들어진 예측 모형으로부터 각 관측 개체에 대한 적합 확률이 계산되며, 분류기준점(cutoff value)에 따라 0.5보다 크면 Y 값은 1, 0.5보다 작으면 Y 값은 0으로 분류한다. 다만 오분류의 중요성을 고려하여 분류기준점(cutoff value)은 상황에 따라 변경될 수 있다[7].

III. 선행연구

3-1 개인성향 분석을 통한 퇴직연구

본 연구와 관련 선행연구로 이직(turnover)의도 및 행동에 대한 선행 연구도 함께 확인한다. 이직은 개인이 직장을 옮기거나 직업을 바꾼다는 뜻으로 국내에서는 1997년 외환위기 이후 평생직장의 개념이 사라지면서 이직 관련 연구가 많이 증가하였다. 1980년대부터 2008년까지의 인사 관련 학술지에 게재된 논문 주제를 분석해 보면 1980년 5.1%수준에서 2000년대 15.8%까지 증가하였다[8].

‘정보통신 업종 대졸 초기 경력자의 이직의도와 이직행동 결정요인 분석’[9]에서는 대졸자 패널 26,544명을 대상으로 이직행동에 가장 영향을 미치는 독립변수를 분석해본 결과 조직 만족도, 임금 만족도, 복리후생 만족도, 노동조합 가입여부 등이 영향을 미치는 것으로 나타났고, ‘2단계 로지스틱 회귀모형을 이용한 직무만족도와 이직행동에 관한 연구’[10]에서는 대졸자 패널 23,728명을 대상으로 이직행동에 영향을 미치는 독립변수로 분석해본 결과 이직의사, 직무만족도, 상용직여부, 전반적 만족도, 근무 개월수, 수입, 정규직 여부, 노조 유무 등이 영향을 미치는 것으로 나타났다. 그러나 앞선 연구의 경우 다양한 직업군을 대상으로 설문방식으로 데이터가 수집되었으며, 선정된 독립변수들도 원인 분석에만 초점을 맞춰 한계가 있었다.

또한 서비스 기업 종사자의 경우 감성노동 강도가 심화되는 상황에서 개인 성향과 이직의도간의 상관관계에 대한 선행연구들도 활발히 진행되고 있었다.

‘성격유형(big 5)이 감성노동 및 이직의도에 미치는 영향’[11]에서는 국내 항공사 객실 승무원 243명 대상으로 이직행동에 영향을 미치는 독립변수로 분석해본 결과 외향성은 표면행위에 (-)영향을 미치며, 신경증은 (+)영향을 미치는 것으로 나타났다. 그러나 독립변수들이 보편적인 개인 성향 항목들로 선정되어, 조직생활에 특화된 성향을 분석하기에는 한계점이 있었다.

3-2 선행 연구들과의 차별점

본 연구는 기존 연구와 다음의 차별성을 가지고 있다. 첫째, 기존 연구의 데이터 수집은 대부분 설문 방식으로 진행되었으며, 이는 수집 당시 개인별 상황에 따라 상이한 결과가 나올 수 있다. 본 연구에서는 공간 및 시간 등 모두 동일한 환경에서 데이터를 수집하여 신뢰성을 높였다. 둘째, 기존 연구는 직무만족도, 고용형태, 조직문화, 인구통계학적 데이터 등 대부분 외부환경에 해당하는 항목을 독립변수로 선정하여 조기 퇴직과 관련된 유의미한 독립변수를 확인에 집중하였으나 실제 활용성 측면에서는 제한적이었다. 또한 개인 성향을 독립변수로 선정한 경우에도 성격의 5가지 요인(불안정성, 외향성, 개방성, 친화성, 성실성)등 개인 성향에 대한 일반적인 항목을 독립변수로 선정하였다. 본 연구는 독립 변수가 기업의 조직 생활에 특화된 항목으로 구성되었으며, 채용 전형에 포함되어 있어 향후 조기 퇴직자 예측 모형에 대한 정확도를 향상시킬 수 있다.

IV. 조기 퇴직 예측 모형 구축

본 연구에서는 인성시험 결과를 기반으로 조기 퇴직자를 예측하는 모형을 제시하는 것을 목표로 한다.

4-1 활용 데이터 정의

본 연구에서는 기업에서 인재 선발의 첫 관문으로 시행하고 있는 A社の 인성시험 결과를 연구 데이터로 선정하였으며, 데이터의 세부내용은 다음과 같다.

첫째, A社에서 ‘13년부터 15년까지 3년 동안 진행된 인성시험 응시자 총 1,179명을 대상으로 데이터를 수집하였다. 둘째, 제조직군을 그룹 A, R&D직군을 그룹 B로 표기하며, 세부 현황은 표 1과 같다. 셋째, 인성시험은 총 500문항으로 응시자는 동일한 문항과 제한시간에서 평가가 진행된다. 넷째, 평가항목은 총 24가지 항목으로 그 내용은 표 2와 같다.

표 1. 그룹별 인력현황

Table 2. workers's status by Group

Variables	Total	Group A	Group B
sum	1179	641	538
currently employee	1096	591	505
retiree	83	50	33

표 2. 인성평가 항목

Table 2. The list of Personality Evaluation

No	List	No	List
1	honesty	13	flexibility
2	universality	14	positive bias
3	irresponsibility	15	analytical skills
4	antisocial	16	Independence
5	impulsiveness	17	ethicality
6	Achievement	18	sociality
7	spontaneity	19	leadership
8	responsibility	20	teamwork skill
9	particularity	21	interpersonal skills
10	endurance	22	self discipline
11	immersion	23	stress tolerance
12	innovation	24	elasticity

4-2 연구방법

조기 퇴직 예측 연구방법은 그림 1과 같은 순서로 진행토록 한다. 인성시험 데이터 결과와 구성원의 재직 여부를 매칭하여 그룹 A와 B로 나누었다.

회귀분석에서 대표적으로 독립변수를 선정하는 방법으로는 크게 전진선택(forward selection; FS)방법, 후진제거(backward elimination; BE)방법, 단계적 선택(stepwise) 방법 등 총 3가지가 있다. 전진선택 방법은 아무런 설명 변수가 없는 상황에서 중요한 독립변수를 추가해 나가는 방식으로(F 검정의 p -value 0.05이하) R^2 값이 유의하게 향상되

는 독립변수가 없을 때까지 추가된다. 반대로, 후방제거 방법은 모든 독립변수를 모형에 포함시킨 뒤에 가장 덜 중요한 독립변수를 하나씩 제거해 나가는 방식으로(F 검정의 p -value 0.1 이상), R^2 값에 유의한 영향을 미치지 않으면서 제거될 수 있는 독립변수가 없을 때까지 진행한다. 단계적 선택은 전진선택과 후진제거의 두 방법을 적절히 조합한 단계적 방법이 가장 많이 사용되고 있다. 본 연구에서도 단계적 방법을 이용하여 로지스틱 회귀분석에서 사용할 독립변수를 선정하였다.

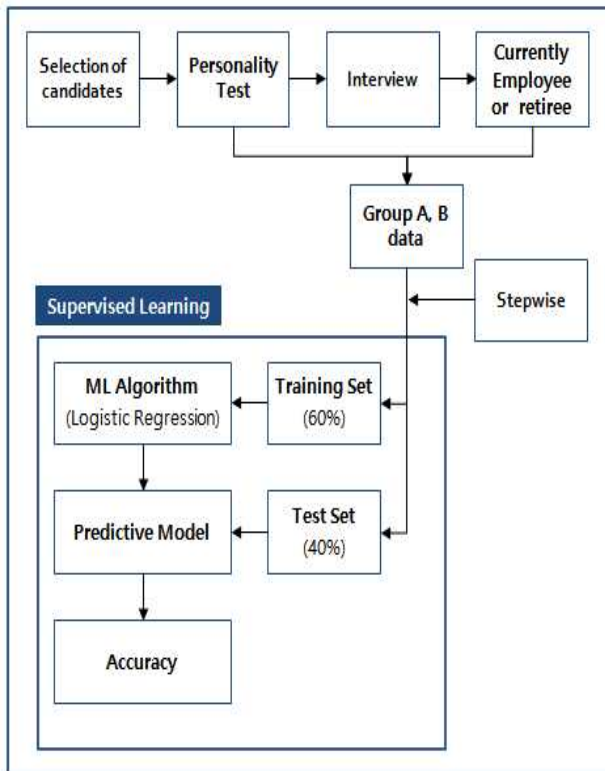


그림 1. 연구 모형
Fig. 1. Research model

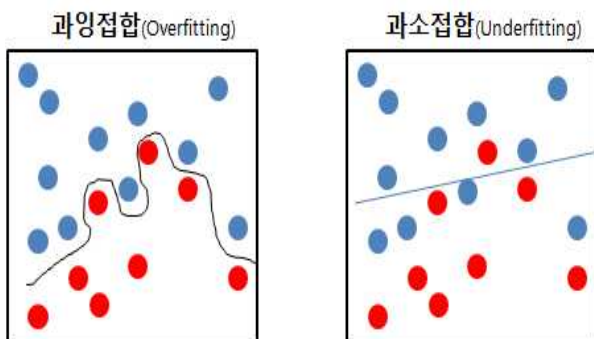


그림 2. 과잉적합 vs. 과소적합
Fig. 2. over-fitting vs. under-fitting

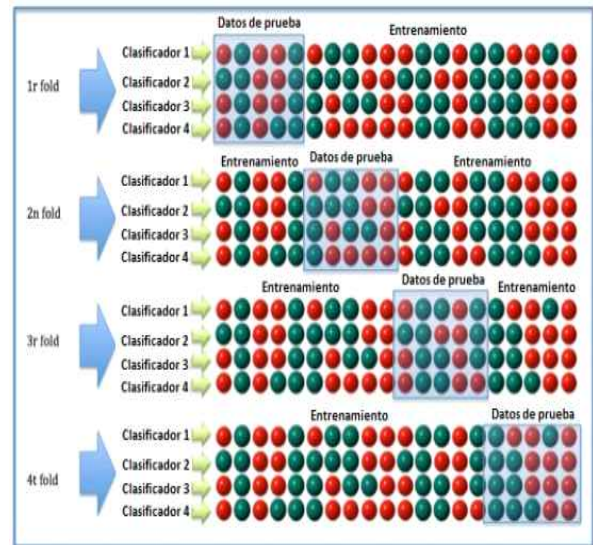


그림 3. K-fold 교차 검증
Fig. 3. K-fold cross validation
출처 : <https://github.com/rmaestre/K-fold-cross-validation>

수집된 데이터에 대한 레이블(label)이 주어졌기 때문에 지도학습(supervised learning) 방식으로 분석하며, 종속변수가 이분 범주형 데이터로 주어져 분류(classification)의 대표적인 머신러닝 알고리즘인 로지스틱 회귀분석을 통해 예측 모형을 제안하도록 한다. 다만 모든 데이터로 예측모형을 만들었을 경우 그림 2와 같이 과잉적합(overfitting) 또는 과소적합(underfitting)이 발생할 수 있기 때문에 예측 모형의 성능 최적화를 위해서 그림 3과 같이 K -fold 교차검증(cross validation) 방식으로 훈련시키도록 한다. 분석 시 K 값은 5로 하며, 데이터는 트레이닝 세트와 테스트 세트로 각각 6대 4대로 분리하여 진행한다. 향후 모형의 성능 평가는 표 3과 같이 혼동행렬에서 식 (4)와 같이 정확도를 산출하여 모형의 평가결과를 확인하였다. 분석 방법은 SPSS statistics 23과 R 3.4.1를 통해 분석하였다.

표 3. 혼동행렬
Table 3. Confusion Matrix

Variables	Result P	Result N
Prediction P	True Positive (TP)	False Negative (FN)
Prediction N	False Positive (FP)	True Negative (TN)

$$\text{Accuracy} = \frac{TP+TN}{(TP+FP+FN+TN)} \quad (4)$$

V. 조기 퇴직자 예측 모형 구축 결과

5-1 A그룹 분석결과

전진 선택법을 통해 변수선택을 하였으며, 총 24개의 독

립변수 중에 솔직성, 무책임성, 반사회성, 책임감, 몰입 등 총 5개의 독립변수가 유의미한 것으로 나타났다. 인성시험 결과와 조기 퇴직과의 관계를 로지스틱 회귀분석으로 분석한 결과는 표 4과 같다.

독립변수간 상관관계 높을 경우 발생하는 다중공선성(multicollinearity) 문제를 진단하기 위해 공차한계와 VIF(variance inflation factor) 값을 확인하였다. 다중공선성 진단은 공차한계가 0.1 이하거나 VIF가 10 이상이면 다중공선성의 문제가 있다고 보는데[12][13], 분석 결과 독립변수들의 공차한계 값이 0.365~0.666로 나타났고, VIF 값은 1.50~2.74로 나타나 변수간 다중공선성은 없는 것으로 확인하였다. 독립변수 중에서는 ‘몰입’이 음의 방향으로 영향력이 큰 것으로 확인되었으며, 예측 모형은 (5)와 같다.

표 4. 인성데이터 분석 결과

Table 4. The result of personality data analytics

Variables	B	S.E	Wald	p-value	Tolerance limit	VIF
honesty	.017	.008	4.54	.033	.571	1.75
irresponsibility	-.024	.010	5.70	.017	.365	2.74
antisociality	-.015	.008	3.44	.064	.567	1.76
responsibility	-.020	.009	5.50	.019	.424	2.36
immersion	-.020	.007	7.17	.007	.666	1.50
constant	.400	1.04	.148	.701		

$$\log \frac{Px}{1-Px} = 0.4 + 0.17x_1 - 0.24x_2 - 0.15x_3 - 0.2x_4 - 0.2x_5 \quad (5)$$

표 5. A그룹 혼동행렬

Table 5. The confusion matrix of Group A

Variables	Result P	Result N	Accuracy
Prediction P	8	16	0.8906
Prediction N	12	220	

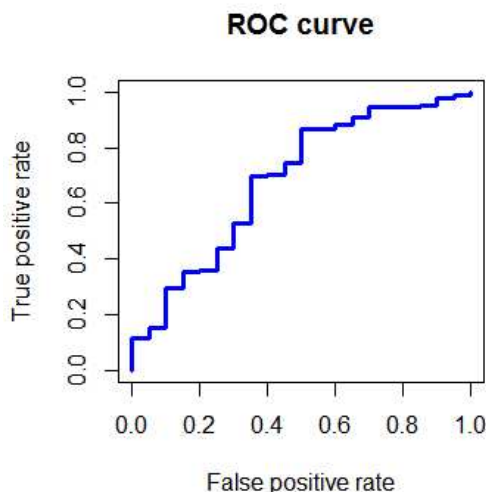


그림 4. 그룹 A ROC 커브

Fig. 4. The ROC curve of Group A

예측 모형에 트레이닝 데이터 60% (385개), 테스트 데이터 40%(256개)를 교차검증 적용하여 혼동행렬로 분석한 결과 표 7과 같다. 또한, ROC 커브는 그림 4와 같으며, AUC 값은 0.6837로 확인되었다.

5-2 B그룹 분석결과

그룹 A와 마찬가지로 전진선택법을 통해 변수선택을 하였으며, 총 24개의 독립변수 중에 반사회성, 성취지향성, 자발성, 회복탄력성 등 총 4개의 독립변수가 유의미한 것으로 나타났다. 인성시험 결과와 조기 퇴직과의 관계를 로지스틱 회귀분석으로 분석한 결과는 표 6과 같다. 다중공선성은 분석 결과 독립변수들의 공차한계 값이 0.466~0.689로 나타났고, VIF 값은 1.45~2.15로 나타나 변수간 다중공선성은 없는 것으로 확인하였다. 독립변수 중에서는 반사회성이 양의 방향으로 영향력이 큰 것으로 확인되었으며, 예측 모형은 (6)과 같다.

표 6. 인성데이터 분석 결과

Table 6. The result of personality data analytics

Variables	B	S.E	Wald	p-value	Tolerance limit	VIF
antisocial	.035	.009	14.9	.000	.689	1.45
Achievement	.018	.009	3.61	.057	.503	1.99
spontaneity	-.028	.011	6.93	.008	.466	2.15
elasticity	.032	.01	9.98	.002	.630	1.59
constant	-6.01	.976	37.9	.000		

$$\log \frac{Px}{1-Px} = -6.01 + 0.35x_1 + 0.18x_2 - 0.28x_3 + 0.32x_4 \quad (6)$$

예측 모형에 트레이닝 데이터 60% (323개), 테스트 데이터 40%(215개)를 교차검증 적용하여 혼동행렬로 분석한 결과 표 7과 같다. 또한, ROC 커브는 그림 5와 같으며, AUC 값은 0.7098로 확인되었다.

표 7. B그룹 혼동행렬

Table 7. The Confusion Matrix of Group B

Variables	Result P	Result N	Accuracy
Prediction P	6	19	0.8790
Prediction N	7	183	

VI. 결론 및 향후 연구 방향

기존의 선행 연구는 설문 방식으로 데이터를 수집하고,

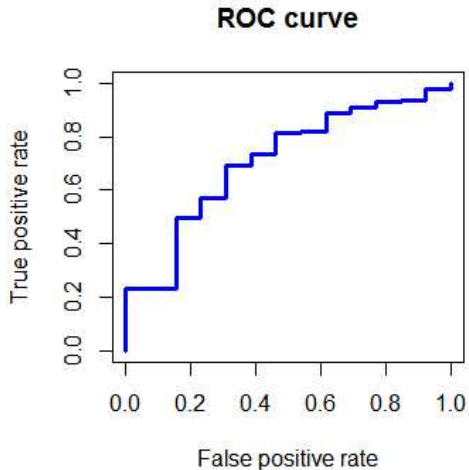


그림 5. B그룹 ROC 커브
 Fig. 5. The ROC curve of Group B

퇴직과 관련이 높은 독립 변수를 확인 하는데 집중하였다면, 본 연구는 채용 전형 시 진행되는 인성 결과 분석을 통해 향후에도 지속 가능한 조기 퇴직 예측 모형을 제시했다는 면에서 의의를 갖는다.

특히, 향후 예측모형의 적합성 및 활용성을 고려하여, 제조(manufacture)직군과 R&D직군으로 구분하여 분석하였으며, 각각 조기 퇴직에 영향을 많이 미치는 요인을 확인할 수 있었다. 제조직군에서는 ‘솔직성’, ‘무책임성’, ‘반사회성’, ‘책임감’, ‘몰입’ 등 총 5가지 항목으로 확인되었으며, 이 중에서 ‘몰입’ 항목이 가장 영향력이 미치는 것으로 확인되었다. R&D직군에서는 ‘반사회성’, ‘성취 지향성’, ‘자발성’, ‘회복탄력성’ 등 총 4가지 항목이 유의미한 항목으로 확인되었으며, 이 중에서 ‘반사회성’ 항목이 가장 영향력을 많이 미치는 것으로 확인되었다. 직군별 유의미한 독립변수들로 예측 모형을 제시하였으며, 데이터를 트레이닝과 테스트를 6대 4로 나눠서 훈련시켰다. 혼동행렬 및 ROC 커브로 모형을 검증한 결과 제조직군의 정확도는 89.0%, AUC 값은 0.6837로 확인되었으며, R&D직군의 정확도는 87.9%, AUC 값은 0.7098로 확인되었다. 본 연구를 통해 제시된 조기 퇴직 예측 모형이 향후 활용성 높은 연구 모형임을 확인할 수 있었다.

다만 본 연구에서도 여러 가지 한계점이 존재하였다. 수집된 데이터는 모두 인성시험에 합격한 사람이기 때문에 대부분의 평가항목이 유사한 점수대의 사람들로 구성되어 있다. 이에 특정 독립변수가 영향력이 매우 높기에는 한계가 있으며, 이는 직무별로 선정된 독립변수의 전체적인 p-value가 높지 않은 것과는 동일한 이유라고 판단된다. 또한 임의(random)로 트레이닝과 테스트 데이터를 설정하여 모형을 제시할 때 마다 정확도에 대한 오차가 발생하였다. 이는 누적된 데이터 많지 않아서 아직은 강인(robust)하지 못한 모형이라고 판단되며, 향후 데이터가 누적될수록

강인한 모형으로 향상 될 수 있을 것이라 예상된다.

기업에서 진행하는 인성시험은 연간 응시 규모나 관심도를 감안하면 국가고시와 견줄만하며, 이에 공정하고 업무에 적합한 신입사원 선발을 위해 많은 시간과 비용을 투자하여 인성 시험을 운영하고 있다. 그러나 이러한 인성관련 빅데이터를 일회성으로 선발여부를 판단하는 기준으로만 활용하는 아쉬움이 있었다. 본 연구는 인성 데이터를 기반으로 하여 인재관리가 필요로 하는 분야의 예측 시스템을 제시했다는 점에서 의의가 있으며, 향후 인성데이터를 활용하여 고성과자와 저성과자 예측, 제조직군과 R&D직군의 효율적인 인력 배치 등 인력관리를 위한 여러 가지 예측 시스템을 제안할 수 있을 것으로 기대 된다.

참고문헌

- [1] Yonhapnews, "Semiconductor exports exceed \$ 90 billion this year" [Internet]. Available: <http://www.yonhapnews.co.kr/bulletin/2017/09/13/0200000000AKR20170913037500003.HTML>
- [2] New Daily Economic, "4th industrial revolution semiconductor new golden age" [Internet]. Available: <http://biz.newdaily.co.kr/news/article.html?no=10123665>
- [3] Korea Employers Federation, "Survey on the recruitment and retraining status of new college graduates" [Internet]. Available: http://www.kefplaza.com/kef/kef_press_view.jsp?num=460
- [4] H. S. Jeon and E. J. Wang, "A study on an exit interview process, influencing the withdrawal of a turnover decision: Semiconductor manufacturing plant case," *Korean Journal of Industrial and Organizational Psychology*, vol. 27, no. 4, pp. 805-830, 2014.
- [5] MBC News, "The selection of new employee in half season", [Internet]. Available: http://imnews.imbc.com/replay/2017/nwdesk/article/4441410_21408.html.
- [6] D. R. Cox, "The regression analysis of binary sequences," *Journal of the Royal Statistical Society*, vol. 20, no. 2, pp. 215-242, 1958.
- [7] J. R. Baek, "Development of accident prediction model for military aircraft by using logistic regression," Master Dissertation, Yonsei University, Korea, 2012.
- [8] S. M. Lee, G. C. Yu, and W. S. Park, "Analysis of articles on HRM in the Korean Journal of Human Resource Management from 1980 to 2008," *Korean Academy of Organization and Management*, vol. 34, no. 1, pp. 177-218, 2010.
- [9] Y. M. Lee and K. J. Youn, "Analysis of influential factors that impact the turnover intention and turnover behavior of newcomers in information technology industries," *Korean*

Society for Learning and Performance, vol. 11, no. 1, pp. 59-77, 2009.

- [10] S. S. Chung and K. H. Lee, "A study on job satisfaction and turnover behavior with 2-stage logistic regression: In case of graduates occupational mobility survey," *Communications for Statistical Applications and Methods*, vol. 15, no. 6, pp. 859-873, 2008.
- [11] H. J. Jung, "The effects of big 5 on the emotional labor and turnover intention: focused on the flight attendant," *Journal of the Korean Data Analysis Society*, vol. 17, no. 3, pp. 1501-1511, 2008.
- [12] W. C. Seo, "A study on the internal reputation factors affecting the job satisfaction: Focusing on big data analysis in the social media for corporation reputation," *Journal of Digital Contents Society*, vol. 17, no. 4, pp. 295-305, 2016.
- [13] H. M. Park, C. S. Oh, and C. S. Yum, "An Empirical Study on the Factors Influencing Student Satisfaction of e-Learning," *Journal of Korean Institute of Information Technology*, vol. 9, no. 7, pp. 143-152, 2011.



김영박(Young Park Kim)

2006년 : 인하대학교 전자공학과 학사

2016년 : 고려대학교 빅데이터 응용 및 보안학과 (석사과정)

2006년~현 재: SK하이닉스

※관심분야 : 빅데이터, 로지스틱 회귀분석



김형중(Hyung Joong Kim)

1978년 : 서울대학교 전기공학과 학사

1986년 : 서울대학교 제어계측공학과(공학석사)

1989년 : 서울대학교 제어계측공학과(공학박사)

1989년~2006년: 강원대학교 교수

2006년~현 재: 고려대학교 정보보호대학원 교수

※관심분야 : 컴퓨터보안, 패턴인식, 가역정보은닉, 머신러닝, 빅데이터 분석 등