

# 인공지능 기반 이직 가능성 예측을 위한 데이터 학습 시스템에 관한 연구

양지원, 전은경, 오승훈, 이세형, 유홍연, 손동훈, 임권섭

한국전자통신연구원

jiwonyang@etri.re.kr, ekjeon@etri.re.kr, osh93@etri.re.kr, seihyoung@etri.re.kr, keister@etri.re.kr, dhson78@etri.re.kr, kaide.lim@etri.re.kr

## A Study on Development of An AI-based Employee Turnover Prediction System

Yang Ji Won, Jeon Eun-Kyoung, Oh, Seung-Hun, , Lee Seihyoung, Yu Hong Yeon, Son

Dong Hoon, Lim Kwon-Seob

Electronics and Telecommunications Research Institute

### 요 약

본 논문은 인공지능 기술을 도입하여 IT 산업 종사자 개개인의 성격 및 업무 유형을 분석하고 이직 가능성을 도출하는 시스템을 제안한다. 제안된 시스템은 웹 인터페이스를 통해 재직자들의 접근을 유도하여 설문/응답 형태로 휴먼리소스 정보를 수집하고, 이를 가공 및 인공지능 기반 학습을 수행한다. 본 연구에서 설계된 시스템의 타당성을 검증하기 위해 상용 데이터셋을 기반으로 학습 모델을 설계하고 모델 성능을 검증하였다.

### I. 서 론

인력 관리 및 유지는 기업의 성장과 경쟁력 확보에 있어 아주 중요한 요소이다. 코로나19 팬데믹으로 인해 사회 전반적인 디지털화가 가속화됨에 따라 IT 전문인력의 수요가 급격하게 증가하였음은 물론, 이에 따른 IT 전문인력 부족 현상이 사회문제로 부각되고 있다. 이러한 인력확보의 과잉 경쟁 구도에서 기업의 인건비 부담은 자연스레 늘어나고 있으며, 이로 인해 구인난이 발생하는 악순환이 조성되고 있다. 특히 구직자들의 중소기업 기피 현상과 더불어 중소기업을 빅테크 기업으로 가기 위한 징검다리 역할로 활용함에 따라 이직 활동이 빈번해지는 양상도 두드러지고 있다 [1].

이러한 대규모 전문인력 이직 현상을 분석하는 데 있어서 사회적 요소뿐만 아니라 개인의 관점에서의 이직 의도에 높은 비중을 두고 있으며, 여기에는 조직문화, 보상 정도, 직무 스트레스, 일과 생활의 균형 정도 등이 주로 고려되고 있다 [2]. 개개인의 업무 성격 특징을 분석한 결과를 토대로 기업은 개인에게 적합한 업무 환경과 조직문화 형성이 가능해지며, 나아가 사회 전반적으로 양질의 일자리 수급이 확대되어 구직난과 구인난 해소 및 의미 채용률 향상에 크게 기여할 수 있다.

따라서, 본 논문에서는 IT직군 종사자를 대상으로 개인별 성격 및 업무 성향 데이터를 수집하고 이직 가능성 지표를 추론해낼 수 있는 AI 학습 시스템 설계에 관해 연구하였다.

### II. 이직 가능성 예측을 위한 데이터 처리 시스템 설계

#### 가. 이직 가능성 평가 요인

<표 1>에서 볼 수 있듯이 IT 전문인력의 이직 의사를 결정하는 요인은 여러 측면에서 분석될 수 있다. 이 중, 대부분의 연구에서 직무 만족도와 급여 수준을 비롯하여 조직 내 대인관계 지표, 일과 생활의 균형 등이 주요 평가 지표로 고려되었다. 이해정 외[2]의 연구에서는 그 외에도 직무에서 겪는 스트레스와 고용안정 등을 이직요인으로 제시하였고, IBM[3]

의 연구는 나이, 성별, 결혼 여부 등 인구통계학적 지표도 이직에 영향을 끼친다고 보았다. 또한, 최진욱 외[4]는 개인의 발전 기회와 성취에 대한 보상이 적절히 제공되고 있는지를 이직 결정요인으로 꼽기도 하였다.

<표 1> IT 직군 종사자들의 주요 이직 결정요인

문헌	주요요인
이 외(Lee et al., 2014) [2]	일의 특성, 급여 수준, 조직문화, 스트레스, 대인관계, 직장 유무, 일과 생활의 균형, 고용안정 등
IBM HR(IBM., 2017) [3]	나이, 출장빈도, 통근 거리, 직원 규모, 성별, 급여 수준, 업무 참여도, 결혼 여부, 대인관계 만족도, 경력 기간, 일과 생활의 균형, 근속 연수 등
최 외(Choi et al., 2021) [4]	급여 수준, 공정한 보상과 인정, 성장을 위한 기회 제공, 복지 제도 등
세렌코(Serenko et al., 2022) [5]	직무 만족도, 동료와의 관계, 개인 성취, 일의 피로도, 직장-가정 갈등 등

본 연구에서는 <표 1>의 선행연구를 참고하여, 이직 가능성 추론 모델을 학습하기 위해 수집할 데이터 구조를 결정하였다. 여기에는 조직문화, 복지제도, 회사 규모 등과 같이 통계학적인 정보뿐만 아니라 설문 응답을 통해 개개인으로부터 드러나는 업무 성격, 그리고 직무 스트레스 수준 등을 포함하여 이직 의도 결정요인을 설정하였다. 따라서 본 연구에서는 검사지 기반으로 개인의 성격, 업무 유형 및 직무 스트레스 지표를 종합적으로 수집하여 종사자의 이직 가능성을 예측하기 위한 학습 시스템을 구성하였다.

#### 나. 데이터 수집 및 분석 시스템의 구성

그림 1은 설문지 기반으로 종사자들의 특성 및 심리 파악 및 이직 가능성 예측을 위한 전반적인 시스템의 구성을 나타낸다. 본 시스템의 사용자는 재직자 그룹과 심리 전문가 및 인사 담당자 그룹으로 구분하며, 각 그룹은 별개의 사용자 인터페이스를 통해 학습 데이터 확보를 위한 일련



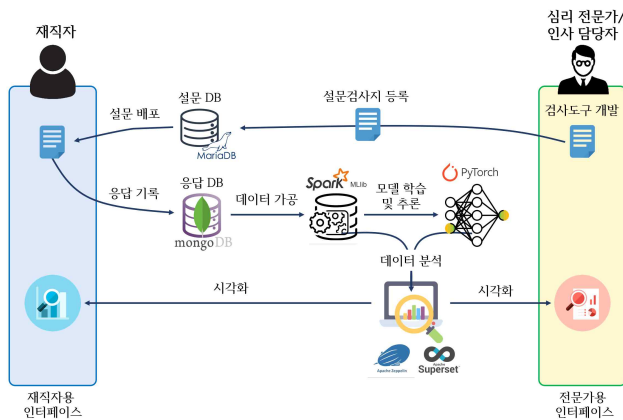


그림 1. 인공지능 기반 개인 특징 및 이직 가능성 예측 시스템의 구성도

의 과정을 수행한다.

먼저, 심리 및 인사 전문가는 IT 기업 재직자들의 개인 특성과 업무 환경을 진단할 수 있는 검사 도구를 개발하고 이를 플랫폼의 설문 DB에 등록한다. 이후 해당 설문지는 검사 대상에 해당하는 재직자들에게 배포되고, 재직자들이 사용자 인터페이스를 통해 응답한 내용은 가명 처리되어 응답 DB에 기록된다. 응답 DB에 저장된 데이터는 데이터 가공 모듈에서 발생시킨 SQL 쿼리를 통해 실시간으로 조회되어 데이터 분석 및 인공지능 모델 학습을 위한 표본 추출 과정을 거친다. 제안된 시스템에서는 비일관적인 답변을 한 불성실 응답자 색출 작업과 결측치 처리 과정이 데이터 가공 단계에 포함되어 있다. 또한, 인공지능 모델 학습을 위한 사전 단계로써 설문 응답 내용 중 숫자로 입력하는 항목에는 최소-최대 정규화를 수행하였고, 범주형 데이터에 대해서는 One-hot encoding 기법을 적용하여 전처리를 수행한다. 데이터 학습 모델의 경우, 새로운 데이터가 지속적으로 추가되는 것을 고려하여 온라인 학습 방식으로 설계하여 학습과 추론을 동시에 병행한다. 마지막으로 데이터 분석 단계는 가공된 데이터와 학습 모델의 결과물을 재직자 그룹과 전문가 그룹에게 종합적으로 시각화해주는 역할을 수행한다. 재직자들은 개인의 검사 결과를 열람하고, 전문가 그룹은 재직자 그룹에 대한 업무 성향 및 이직 예측 통계를 토대로 업무 환경이나 제도 개선 계획을 수립할 수 있다.

#### 다. AI 학습 기반의 이직 가능성 추론 모델 설계

설계된 시스템의 상용화에 앞서, 공개된 설문기반 이직 가능성 예측 데이터셋을 활용하여 인공지능 모델을 구현하고 예측 성능을 평가하였다. 실험에 도입된 데이터셋은 앞서 소개한 IBM HR [3] 분석 자료에 수록된 IBM에 종사하고 있거나 종사했었던 1,470명의 직원을 대상으로 실시한 설문결과이다. 사용한 데이터셋은 이직 여부를 포함한 총 35가지의 응답 항목을 포함하고 있으며, 전체 1,470개의 표본 중 237개의 표본이 이직자들에 해당하는 응답 결과이다. 이직자와 일반 재직자 간의 표본 크기 차이가 상당하므로, 본 연구에서는 오버샘플링 (Oversampling) 기법을 통해 클래스 간 표본 수의 균형을 맞춰주었으며, 결과적으로 2,466개의 데이터를 얻었다. 이중 전체의 80%에 해당하는 양의 데이터를 무작위로 추출하여 학습 단계에, 나머지를 모델 검증 단계에 사용했으며 학습 모델로는 5개의 은닉층으로 구성된 심층 신경망을 활용하였다.

그림 2는 모델 학습 및 검증단계에서 산출된 여러 분류 지표를 보여준다. 그림 2a는 학습 과정이 진행됨에 따라 훈련 데이터셋과 검증 셋 각각에 대한 분류 정확도 (Accuracy) 결과를 나타낸다. 최종 학습 단계에서는 두 경우 각각 0.980, 0.919로 준수한 수치를 기록하였다. 마찬가지로 그림 2b에서 볼 수 있듯이, 학습된 모델은 검증셋에 대하여 정밀도 (Precision),

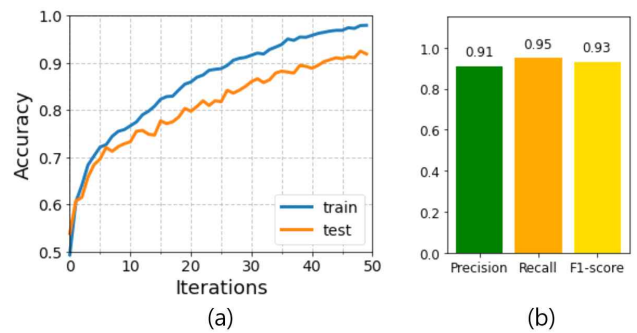


그림 2. IBM 이직 가능성 예측 모델의 성능 지표

재현율 (Recall) 및 F1-score 지표에서 모두 0.9 이상을 달성하였다. 3가지 지표 중, 재현율 측면에서는 꽤 우수한 성능을 얻었고, 정밀도 수치는 이에 다소 못 미치는 수준으로 나타났다. 이는 실제로 이직하지 않았음에도 모델이 이직을 하였다고 오판한 경우가 다른 경우들에 비해 상대적으로 잦은 편임을 의미한다. 앞서 학습 단계가 보여준 훈련 셋과 검증 셋의 성능 차이가 초반부터 벌어진 점도 같이 고려해보면, 사용된 데이터 수와 활용 가능한 특성의 가짓수가 다소 적은 편이었던 점을 주요한 원인으로 지적할 수 있다.

### III. 결론 및 향후 연구방향

본 논문에서는 개인 특징 평가를 기반으로 IT 전문인력의 이직 가능성을 예측하기 위한 데이터 수집 및 학습 시스템 구조를 제안하였다. 제안한 시스템은 심리 및 인사 전문가 그룹이 개발한 검사 도구를 기반으로 재직자들의 개인 특성 데이터를 추출한 뒤, 이를 가공하고 인공지능 모델 기반 추론 과정을 거쳐 개개인의 이직 가능성 유무를 판단한다. 설계된 시스템의 타당성을 검증하기 위해 공개된 데이터셋을 기반으로 학습 및 모델 성능 평가를 수행하였다. 향후 시스템의 성능 고도화를 위해 좀 더 다양한 지문으로 구성된 검사 도구와 방대한 양의 HR 데이터를 확보하여 추가적인 연구를 수행하고자 한다.

### ACKNOWLEDGMENT

본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음(22ZK1100, 호남권 지역산업 기반 ICT 융합기술 고도화 지원사업).

### 참 고 문 헌

- [1] 장지윤, 이종주, "SW개발자의 일자리 환경", SPRI Issue report, IS-121, July 2021.
- [2] 이해정, 김한별, 이정우, "IT 전문 인력의 커리어 가치관: 이직(移職)의 도의 Q 방법 분석", J. of KSSSSS, Vol. 28, 93-114, June 2014.
- [3] P. Subhash, "IBM HR analytics employee attrition & performance," (<https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset>).
- [4] 안병대, 최진욱, 서용무, "기업개성이 직원의 직무만족과 기업 이직률의 관계에 미치는 영향: 잡플래닛 기업 리뷰를 중심으로", J. of Information Technology Services, pp. 35-56, May 2020.
- [5] Alexander Serenko, Hiroshi Sasaki, Prashant Palvia, Osam Sato, "Turnover in Japanese IT Professionals: Antecedents and Nuances", Australasian Journal of Information Systems, Vol. 26, May 2022.