

청년세대의 이직 요인 분석 및 분류 예측 모형 개발: 직장 및 개인의 특성과 환경을 중심으로

박세연*, 황성아**, 허수정***

<요 약>

최근 신입 직원의 조기 퇴사율 증가 현상은 중요한 문제로 대두되고 있다. 본 연구는 심층적 분석을 통하여 청년세대의 주된 이직 요인을 다각도로 파악하고 개인의 이직 여부를 예측하는 모형을 개발하는 것을 목적으로 한다. 한국고용정보원의 대졸자 직업 이동 경로 조사 데이터를 가설에 따라 분류하고 구조방정식을 세워 이직에 중요한 변수 23개를 추출한다. 이어서 중요 변수를 기반으로 머신러닝 알고리즘과 베이지안 최적화 기법을 통해 이직 여부 예측 모형을 개발한다. 분석 결과, 직장의 특성, 직장에 대한 개인의 특성, 직장에 대한 개인의 만족도, 개인의 환경 요인은 이직과 인과관계를 가지며, 만족도의 경우 만족도가 높을수록 이직률이 줄어들 것이라는 본 연구의 가설이 성립하는 것을 확인했다. 본 연구를 통해 청년세대의 근속률 향상을 위한 정부 및 기업 차원의 제도적 장치 마련에 기여할 수 있을 것으로 기대한다.

주제어: 분류 예측 모형, 구조방정식, 청년세대 이직, 요인 분석

* 연세대학교 정보대학원 비즈니스 AI 빅데이터 분석 트랙, 석사과정

** 연세대학교 정보대학원 비즈니스 AI 빅데이터 분석 트랙, 석사과정

*** 연세대학교 정보대학원 비즈니스 AI 빅데이터 분석 트랙, 석사과정

I. 서론

최근 한국의 청년세대(만 19~39세) 사이에서 이직률이 증가하고 있다. 통계청의 「일자리이동통계(2017~2021년)」에 따르면, 근속기간 1년 미만의 일자리 이동 비율은 2017년 약 0.43%에서 2021년 약 0.80%로 매년 유의미하게 증가하였다. 이러한 추세는 청년층의 직장 만족도, 삶의 지향 태도, 일 가치관, 기업 규모, 워라벨(Work-Life Balance, 일과 삶의 균형) 등 다양한 요인들과 관련이 있다는 연구들이 진행되고 있다. 특히, 이직의 주요 요인들을 분석하고, 이를 바탕으로 개인의 이직 여부를 예측하는 연구는 개인과 기업 및 사회 전체에 매우 중요한 시사점을 제공한다.

<그림 1> 청년 고용동향(고용률 및 실업률)



자료: 고용노동부 「청년 고용동향」

고용노동부의 2010~2023년 「청년 고용동향」 자료에 따르면, 코로나19 팬데믹 이후 청년 고용률은 2020년에 42.2%에서 2023년에는 46.5%로 증가하여 역대 최고치를 기록하였다. 이를 통해 고용시장이 빠르게 회복되고 있음을 알 수 있다. 그러나 고용노동부 및 한국고용정보원의 「2023년 하반기 기업 채용동향 조사」에 따르면, 연간 신규입사자 중 평균 16.1%가 1년 내 퇴사하며, 그중 신입직원이 57.2%에 달하는 것으로 나타났다. 즉, 고용률은 비교적 안정적인 데에 반해 조기 퇴사자가 많은 상황이다.

코로나19로 인해 청년들이 취직에 어려움을 겪다가 고용시장이 회복된 상황에서, 신입 직원의 1년 내 퇴사율이 증가한 것은 중요한 문제로 대두되고 있다.¹⁾ 이는 고용률이 역대 최고치를 기록한 것과 대비되는 현상으로, 청년들이 직장에서 안정적으로 근무하지 못하는 상황을 반영한다. 조기 퇴사로 인한 기업의 손실비용은 채용, 교육 업무 인수인계 등에 대해 1인당 2천만 원 이상으로 추산되기 때문에 기업 입장에서도 심각한 문제이다.²⁾ 이에 이직 원인을 분석하는 연구는 불필요한 이직 및

1) 동아일보의 2024년 4월 2일자 '대기업 신규 입사 16%가 1년 내 퇴사... 기업들 "인당 2000만원 손해"' 기사에 따르면 기업들은 조기 퇴사자가 증가하여 발생하는 손해 및 부담을 느끼고 있으며, 이를 방지하는 것이 근로자와 회사 모두에게 중요하다고 보고 있다.

2) 고용노동부의 2024년 3월 24일자 "선도기업의 채용 결정요소 1위 '직무관련 일경험'"이라는 제목의 보도자료에 따르면 기업의 75.6%는 조기 퇴사로 인한 기업의 손실비용이 2천만 원 이상이라고 답

조기퇴사를 방지하고 개인 및 기업에 긍정적인 영향을 미칠 것이다.

따라서 본 연구에서는 청년세대의 이직 요인을 심층적으로 분석하고자 한다. 이를 통해 청년세대의 주된 이직 요인을 파악하고, 정부 및 기업이 청년층의 근속률 향상을 위해 마련할 수 있는 제도적 장치를 제안하는 것을 목표로 한다. 나아가, 본 연구는 다양한 요인들을 고려하여 이직에 중요한 변수들을 추출하고 개인의 이직 여부를 예측할 수 있는 모형을 구축하고자 한다. 나아가, 구축한 모형의 해석과 함께 개인이 처음부터 자신에게 적합한 직업을 선택하는 데에 도움을 주고, 기업의 입장에서 효과적인 채용 방안을 제공하고자 한다.

이를 통하여 청년층의 노동 시장 안정화에 기여할 뿐만 아니라, 기업들이 인재 유출을 방지하고 효율적인 인사 관리를 하도록 돕는 데 중요한 역할을 할 것이다. 또한, 청년층의 이직 요인들을 다각적으로 분석함으로써, 그들이 원하는 직장 환경과 근로 조건을 이해하고, 이를 토대로 보다 나은 일자리 정책을 설계하는 데 기여할 수 있을 것이다.

II. 선행연구

1. 이직 요인 분석

기존 연구의 동향을 살펴보면, 이직과 중요하다고 판단한 소수 요인에 초점을 맞추어 이직의 원인을 찾아왔다. 이와 관련된 선행 연구는 아래와 같다.

박진아(2016)는 청년층의 직장 만족도가 낮을수록 이직률이 높아진다는 연구를 제공한다. 해당 연구는 직장 및 직무 만족도가 이직률에 직접적인 영향을 미치는 중요한 요인 중 하나라고 주장한다. 직장 만족도란 직장에서 제공하는 근로 조건, 복지 혜택, 근무 환경 등이 직원 개인의 기대와 얼마나 부합하는지를 의미한다. 이는 청년들이 직장에서 기대하는 바와 실제 직장 환경 간의 괴리에서 비롯된다고 볼 수 있다.

배성숙(2023)의 연구에서는 기업의 복지 수준이 청년층의 직장 만족도에 큰 영향을 미치며, 이는 이직률과 밀접한 관련성을 파악한다. 특히, 근로소득, 고용 안정성, 직무 내용, 근무환경(시설, 안전, 위생상태 등), 근로시간, 개인의 발전 가능성, 인간관계, 복리후생 제도, 승진 제도, 일자리 및 하고 있는 일에 대한 사회적 평판, 일의 자율성과 권한, 자신의 적성 및 흥미의 일치, 직무교육 또는 훈련 등이 직장 만족도를 구성하는 주요 요소들이다.

문영만·홍장표(2017)의 연구에 따르면, 기업 규모는 청년층의 이직률에 중요한 영향을 미치며 기업의 규모가 클수록 청년들의 이직률이 낮아지는 경향을 보인다. 대기업에 근무하는 청년들은 중소기업에 비해 상대적으로 더 나은 복지와 안정성을 제공받기 때문에 이직률이 낮다고 주장한다.

송민선·박정운(2023)에 따르면 청년층의 삶 지향 태도와 일 가치관은 이직률에 중

했으며 이에 따른 경제적 비용이 상당한 것으로 판단한다.

요한 영향을 미친다. 삶 지향 태도란 개인이 삶에서 추구하는 목표와 가치관을 의미하며, 이는 직장 선택과 근무 환경에 대한 선호도에 영향을 미친다.

안세진·엄기수·서철승(2023)은 워라밸이 청년층의 이직률에 미치는 영향을 보여준다. 이들은 워라밸이 잘 유지되는 직장에서 근무하는 청년들이 그렇지 않은 경우보다 이직률이 낮다는 결과를 도출한다. 워라밸이란 일과 개인 생활의 균형을 의미하며, 이는 개인의 전반적인 삶의 질에 큰 영향을 미친다.

2. 이직 여부 분류 예측 모형 개발의 해외 사례

해외에서는 다양한 요인을 기반으로 직원의 이직률을 예측하는 연구들이 활발히 진행되고 있다. 이와 관련된 선행 연구는 다음과 같다.

Liao·Chenyu(2023)는 머신러닝 기반 모델을 사용하여 직원의 이직 예측 중요성을 강조한다. 해당 연구는 직무 만족도, 근로 시간, 급여 수준, 근무 환경 등의 요인을 활용하여 직원의 이직 확률을 예측하고, 이를 통해 기업이 인사 관리 전략을 수립할 수 있도록 지원한다. 이는 머신러닝을 통한 직원 이직 예측이 기업의 인적 자원 관리에 중요한 역할을 할 수 있음을 시사한다.

Sardar·Devadarshan K·Sidhant Chourasiya·V. Vijyalakshmi(2023)은 직원의 개인 특성과 업무 환경을 고려하여 이직 확률을 예측한다. 직장 만족도, 직무 스트레스, 경력 발전 기회, 상사와의 관계 등 다양한 요인을 활용해 직원 이직 예측의 정확성을 향상시키는 것에 집중한다.

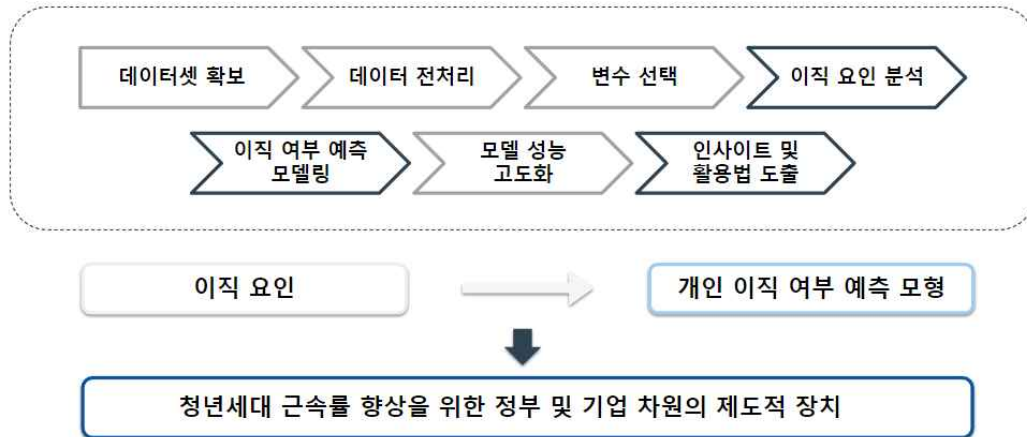
Li·Silin(2022)는 데이터 마이닝 알고리즘을 기반으로 직원 이직 예측 모델을 구축한다.. 해당 연구는 근로 조건, 복지 혜택, 직무 적합성, 조직 문화 등을 요인을 기반으로 직원 이직 가능성을 분석 및 예측하고, 이를 통해 기업의 인적 자원 관리 효율성에 기여한다.

앞선 선행 연구에서 활용한 요인들과 해외 이직 여부 분류 예측 모형 사례를 바탕으로 다각도에서 살펴본 100개 이상의 요인을 기반으로 구조방정식을 이용해 신뢰성 높은 중요변수를 추출하여 이직 분류 예측 모형을 개발한다. 이는 다양한 방면에서 이직의 요인을 찾아볼 수 있기에 중요한 연구 과제이며, 기존 연구들이 현재 재직 중인 직장에 대한 요소에만 집중한 것과는 차별화된다. 또한 해외의 이직 여부 분류 예측 모형에 대한 기존 연구는 해외 사례를 토대로 구축되었기에 우리나라의 경우와 그 의미와 이유에 있어 큰 차이가 있다. 따라서 우리나라의 직장 및 개인의 환경과 특성을 반영한 이직 원인을 찾고 이직 여부 분류 예측 모형을 개발하고자 한다.

본 연구는 청년세대의 이직 요인을 폭넓게 파악하여 근속률 향상을 위한 시사점을 제공하고자 한다. 이를 통해 정부 및 기업의 인사 관리에 유용한 정보를 제공함으로써, 청년들이 보다 안정적이고 만족스러운 직장 생활을 할 수 있도록 돕고, 나아가 한국 사회의 노동 시장 안정화에 기여할 것이다.

III. 연구 방법

<그림 2> 연구 흐름도



1. 연구 자료

본 연구에서는 한국고용정보원의 대졸자 직업 이동 경로 조사 데이터(Graduate Occupational Mobility Survey, GOMS)(통계청 승인번호: 327004)를 사용한다. 2019년 기준 59페이지, 1,333개의 문항으로 이루어진 설문조사 데이터로 매년 신규로 노동시장에 진입하는 전문대학교 이상의 대학졸업자 18,000명을 대상으로 조사하였으며 교육과정, 재학 중 경력개발과 취업 경험 등의 정보를 포함한다. 본 연구에서는 해당 데이터 중 2010년부터 2019년도까지의 데이터를 사용한다. 이때, 설문 문항 중 각 문항별로 이직에 영향을 끼칠 중요도를 연구 팀 내부에서 자체적으로 판단하여 205개의 문항을 선택하고, 아르바이트 형태가 아닌 상용근로자 데이터만을 가져오기 위해 관련 설문조사 응답을 필터링하여 사용한다.

2. 데이터 이해 및 전처리

1) 가설 설정에 따른 변수 분류

<표 1> 사용할 데이터 변수 및 개수

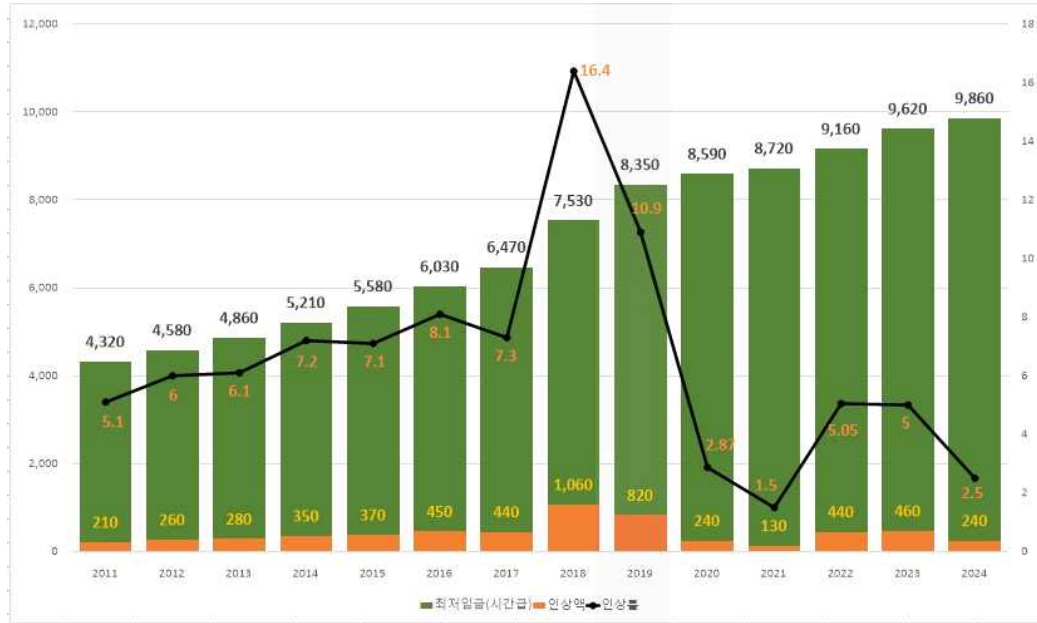
구분		변수 설명	개수
종속변수		이직 준비 여부	1
독립변수	직장의 특성	직장 규모, 직장 위치, 직장 유형, 직장 내 복지	42
	직장에 대한 개인의 특성	직장 내 지위, 근로시간, 근로소득, 초임, 출퇴근 거리, 노동조합 가입 여부, 기술 수준 일치 정도	72
	직장에 대한 개인의 만족도	근로소득, 고용 안정성, 직무 내용, 근무 환경, 개인의 발전 가능성, 인간관계, 복리후생 제도, 승진 제도, 일자리 및 하고있는 일에 대한 사회적 평판, 일의 자율성과 권한, 자신의 적성 및 흥미의 일치, 직무교육 또는 훈련, 일 수준 일치 정도, 기술 수준 일치 정도의 만족도 결과	30
	개인의 환경	전공 만족도, 졸업 평점, 학자금 대출 및 상환 여부, 휴학 및 졸업유예 경험, 재학 중 경험한 일자리 횟수, 졸업 전 취업 목표와 달성 여부, 자격증 개수, 현재 건강상태, 삶에 대한 만족도, 혼인 여부, 배우자와 부모님의 월평균 소득, 부양 자녀 수, 가족의 경제적 지원, 금융 및 부동산 소득, 월평균 지출액	61

이직의 요인을 분석하기 위하여 종속변수를 이직 준비 여부로 두고 연구의 가설을 크게 4가지로 설정하여 독립변수를 구분한다. <표 1>에 따르면, 설정한 가설과 해당되는 변수는 다음과 같다. 첫째, 이직은 직장의 특성과 인과관계를 가진다. 직장 규모, 직장 위치, 직장 유형, 직장 내 복지 등이 이에 해당된다. 둘째, 이직은 직장에 대한 개인의 특성과 인과관계를 가진다. 직장 내 지위, 근로시간, 근로소득, 초임 등이 해당된다. 셋째, 이직은 직장에 대한 개인의 만족도가 높을수록 이직과 부의 관계를 가진다. 개인의 발전 가능성, 인간관계, 복리후생 제도, 고용 안정성 등의 변수가 이에 해당된다. 넷째, 이직은 개인의 환경과 인과관계를 가진다. 전공 만족도, 졸업 평점, 학자금 대출 및 상환 여부, 휴학 및 졸업유예 경험 등이 해당된다.

2) 파생변수 생성

(1) 최저임금을 반영한 소득 관련 파생변수 생성

<그림 3> 연도별 최저임금 결정현황



자료: 최저임금위원회 「최저임금 현황」

$$\text{변환계수} = \begin{cases} 209 \times 12 & (\text{연봉일 경우}) \\ 209 & (\text{월평균일 경우}) \\ 209 \div 4.345 & (\text{주당소득일 경우}) \\ 8 & (\text{일당일 경우}) \\ 1 & (\text{시간당소득일 경우}) \end{cases} \quad (1)$$

$$\text{급여 배수} = \frac{(\text{급여액}) \times 10000}{(\text{변환계수}) \times (\text{해당 연도의 최저임금})} \quad (2)$$

소득과 관련된 월근로소득, 희망연봉, 급여 형태, 급여액은 각 연도별로 최저임금이 상이하므로, 이를 고려한 소득 계산이 필요하다. 최저임금위원회의 최저임금 현황 자료를 토대로 각 소득 데이터를 해당 연도별 최저임금 대비 비율로 바꿔 환산한다. 또한 연봉, 주당 소득, 일당 등 설문조사 기간 별로 다양한 형태로 값이 존재하므로 기간을 통일하는 변환 계수를 구하여 연도별 시간당 최저임금 대비 비율로 변환한 파생변수를 생성한다.

(2) 대학별 학점 만점을 고려한 학점 비율 파생변수 생성

$$\text{만점 대비 비율} = \frac{(\text{실제 학점})}{\begin{cases} 4.5 \text{ if 학점만점이 } 4.5 \text{ 점일 때} \\ 4.3 \text{ if 학점만점이 } 4.3 \text{ 점일 때} \\ 4.0 \text{ if 학점만점이 } 4.0 \text{ 점일 때} \end{cases}} \quad (3)$$

대학별로 학점의 만점 기준이 상이하기 때문에, 만점 대비 비율로 변환한 파생변수를 생성할 필요가 있다. 따라서 실제 학점을 해당 학점 만점으로 나누어 비율로

환산하는 방법을 사용한다. 예를 들어, 학점이 3.5이고 학점 만점이 4.5일 경우, $3.5 \div 4.5 = 0.778$ 값으로 변환한다.

3) 전처리 과정

설문조사 데이터의 특성상 결측값이 존재한다. 따라서 결측값이 전체의 20% 이상인 변수는 삭제하고, 나머지 결측값은 각 변수의 특성에 따라 최빈값 또는 평균값으로 대체한다. 수치형이 아닌 변수에 대해 범주형은 One-Hot Encoding을 적용하고, 데이터 타입을 카테고리형으로 변환한다. 여부에 대한 변수인 경우 Boolean 처리한다. 또한 데이터의 분포의 왜곡도가 |2|이상일 경우 왼쪽분포면 x^2 , 오른쪽분포면 $x^{0.5}$ 로 변환하고, 이상치는 박스플롯을 통해 제거하고, Min-Max 정규화를 적용하여 [0, 1] 범위 내에서 변수의 범위를 재조정함으로써 모든 피처에 걸쳐 균일한 스케일링을 보장한다. 또한 변수의 개수가 많아 다중공선성이 불가피하므로 변수 간 상관계수가 절대값 0.8 이상인 경우 둘 중 하나의 변수만 남겨 다중공선성 문제를 해결한다. 추가로, 종속변수인 이직 준비 여부에서 이직 생각이 없다고 대답한 수가 이직 생각이 있다고 대답한 수에 비해 약 5.7배 많아 불균형을 띄는 문제를 해결하기 위해 K-Nearest Neighbor 중 랜덤으로 하나의 샘플을 선택하는 Over-sampling 기법 중 하나인 SMOTE를 적용한다. 이를 통해 이직 생각이 없다고 대답한 수와 이직 생각이 있다고 대답한 수를 동일하게 23,597로 맞춰 불균형 문제를 완화한다.

3. 구조방정식

구조방정식은 경로 분석, 회귀 분석, 요인 분석이 합성되어 발전된 통계 방법으로, 변수들 간의 복잡한 관계를 모델링하고 검증하는 데 사용된다. 이를 이용하여 잠재변수와 관측변수간의 관계를 분석하고, 동시에 측정오차를 포함한 여러 변수들 간의 인과관계를 추정할 수 있다. 구조방정식은 측정 모형과 구조적 모형로 구성된다. 먼저 측정 모형 파트에서 잠재변수와 관측변수의 관계 정의를 통해 관측변수에 대한 각 잠재변수의 대표성을 평가한다. 그리고 해당 측정 모형을 바탕으로 구조적 모형을 설계하고, 이 구조적 모형을 통하여 잠재변수 간 인과관계를 분석할 수 있다.

4. 머신러닝 기반 이진 분류 모델링

1) 이진 분류 예측 모형

머신러닝 기반의 이진 분류 예측이 가능한 모형은 Logistic Regression, K-Nearest Neighbor, Naive Bayes, Decision Tree가 있다. Logistic Regression은 선형 결합을 사용하여 입력 변수와 출력 사이의 관계를 모델링하며, 로지스틱 함수를 통해 확률로 변환하여 이진 분류가 가능하다. K-Nearest Neighbor는 주어진 데이터 포인트와 가장 가까운 K개의 이웃 데이터 포인트들의 다수 클래스를 예측값으로 사용하여 분류한다. Naive Bayes는 베이즈 정리를 기반으로, 각 입력 변수가

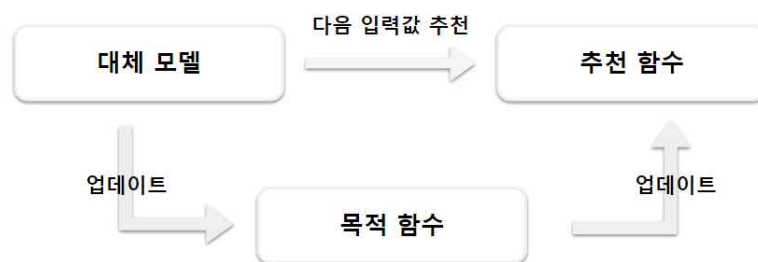
독립적이라고 가정하여 각 클래스에 대한 사후 확률을 계산하고, 이를 기반으로 분류를 수행한다. 그리고 Decision Tree는 데이터 분할을 통해 나무 구조를 형성하며, 각 노드에서 최적의 분할 기준을 찾아 가치를 분할한다. 이때, 정보 이득이나 지니 불순도와 같은 지표를 사용하여 분할을 결정한다. 이들은 주로 정확도, 재현율, 정밀도, F1-스코어를 성능지표로 설정하여 그 성능을 평가한다.

2) Decision Tree 기반 앙상블 모형

앙상블 모형은 여러 개의 개별 모형을 결합하여 예측 성능을 향상시키는 기법이다. 앙상블 모형은 단일 모형의 한계를 극복하고, 예측 정확도와 일반화 능력을 높이는 데 사용한다. Decision Tree 기반 앙상블 모형로는 Random Forest, LightGBM, XGBoost가 있다. Random Forest는 여러 개의 Decision Tree를 독립적으로 학습하고, 각 트리의 예측을 결합하여 최종 예측을 수행하는 앙상블 학습 모형이다. 이는 배깅 기법을 사용하여 트리들의 예측을 다수결 투표 방식으로 결합함으로써, 과적합을 방지하고 예측의 안정성을 높인다. LightGBM은 여러 개의 Decision Tree를 순차적으로 학습하는 부스팅 방식 중 하나로 Gradient Boosting 알고리즘을 사용하여 트리의 성장을 리프 중심으로 확장하는 특징이 있다. 이는 학습 속도를 높이고, 메모리 사용량을 줄여 대규모 데이터셋에서도 효율적으로 동작한다. 마찬가지로, XGBoost도 부스팅 방식의 앙상블 모형로, Gradient Boosting 알고리즘을 확장하여 다양한 최적화 기법과 정규화 기법을 포함하고 있다. XGBoost는 L1, L2 정규화를 통해 과적합을 방지하며, 병렬 학습을 지원하여 학습 속도를 대폭 향상시킨다. 또한, 희소 데이터를 효율적으로 처리할 수 있는 기능을 갖추고 있어 자주 사용되는 모형 중 하나이다.

3) 베이지안 최적화 기법

<그림 4> 베이지안 최적화 알고리즘



베이지안 최적화 기법은 사전 정보를 최적값 탐색에 반영하여 불필요한 하이퍼 파라미터 반복 탐색을 줄이고, 보다 빠르게 최적 하이퍼 파라미터를 찾을 수 있는 최적화 기법이다. 해당 기법은 미지의 목적함수를 최대화 혹은 최소화하는 최적해를 찾기 위해 사용되며, 현재까지 조사된 입력값과 함수값 쌍들에 대해 목적 함수에 대한 확률적 추정을 수행하는 모형인 대체모형과 그 대체모형을 활용해 다음 입력

값 후보를 추천해주는 함수인 추천모형로 구성되어 있다. <그림 4>에서 볼 수 있듯이 입력값과 함수값의 조합을 기반으로 대체 모형을 구축하고 순차적으로 업데이트해 나가면서 예측값과 실제 값이 거의 일치하는 지점에서 최적값을 선택한다.

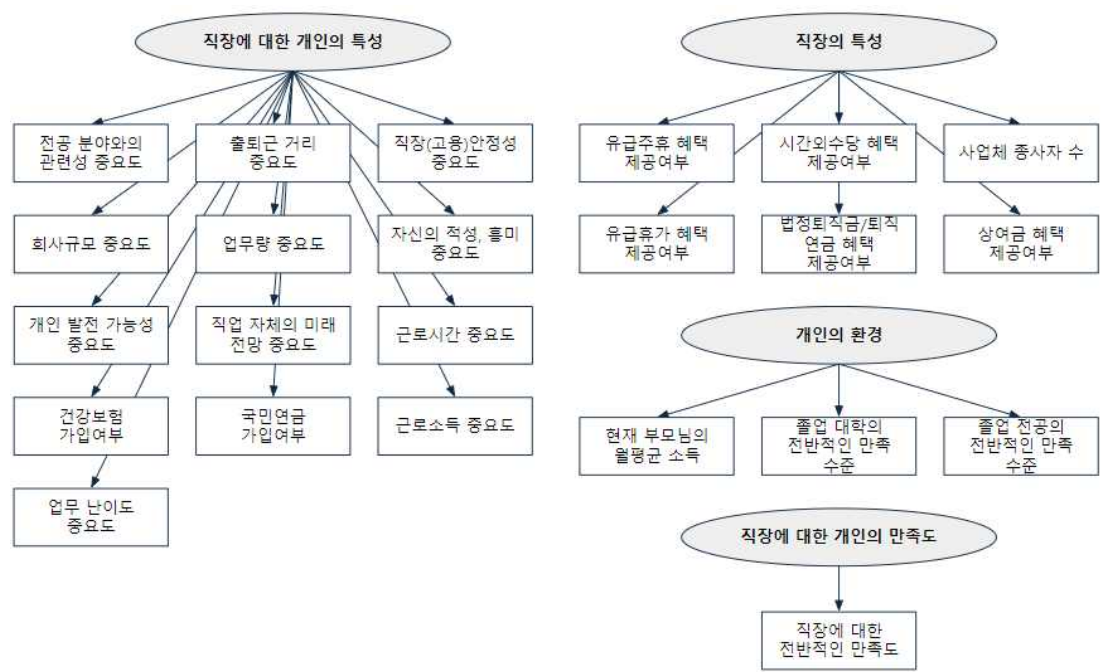
5. SHAP 모델 해석 기법

SHAP(Shapley Additive exPlanations)은 Lloyd Stowell Shapley의 이론을 기반으로 Shapley Value와 특징 사이의 독립성을 활용하여 설명력을 제공한다. SHAP를 사용함으로써 결과가 항상 일정하도록 보장하고 상호 영향과 부정적 영향의 가능성을 설명할 수 있다. SHAP의 Shapley Value는 각 특징이 전체 성능에 기여하는 정도를 수치화한 것이다. 구체적으로, Shapley Value는 모형이 표현할 수 있는 모든 특징의 조합에서 값을 구하고자 하는 전체 모형에 기여하는 특정 특징을 빼고 평균 내어 계산하며, SHAP은 모형의 출력을 각 특징의 기여도로 분해하여 설명한다.

IV. 연구 결과

1. 요인 분석 결과

<그림 5> 요인 분석 결과 시각화



전처리한 데이터셋에 대하여 구조방정식을 통한 요인 분석을 진행한다. p-value 값이 0.05 이하인 관찰변수 중에서 최종 분류 예측 모형을 생성하였을 때, 정확도가 0.9 이상이 되도록 하는 Estimates 값의 기준을 정하고, 그 기준을 기반으로 최종 중요변수의 개수를 결정한다. 그 결과, 총 205개의 관찰변수 중 중요한 변수 23개를

추출하였으며, 해당 변수는 Estimates 값이 1.5이상인 것으로 확인되었다. 추출된 변수는 구체적으로 직장에 대한 개인 특성 관련 변수 13가지, 직장의 특성 관련 변수 6가지, 직장에 대한 개인 만족도 변수 1가지, 개인의 환경 변수 3가지로 선정되었다. 이렇게 추출된 관찰변수와 잠재변수 간 Estimates 값의 관계를 시각화한 그림은 <그림 5>와 같다.

<표 2> 추출된 23개의 중요변수

잠재변수	관찰변수명	p-value ³⁾	Estimates
직장에 대한 개인의 특성	전공 분야와의 관련성 중요도	**	23.22
	출퇴근 거리 중요도	**	22.10
	직장(고용)안정성 중요도	**	25.04
	회사규모 중요도	**	22.84
	업무량 중요도	**	29.43
	자신의 적성, 흥미 중요도	**	21.21
	개인 발전 가능성 중요도	**	23.78
	직업 자체의 미래 전망 중요도	**	25.88
	근로시간 중요도	**	24.70
	건강보험 가입여부	**	1.59
	국민연금 가입여부	**	2.33
	근로소득 중요도	**	21.06
	업무 난이도 중요도	**	27.74
직장의 특성	유급주휴 혜택 제공여부	*	1.69
	시간외수당 혜택 제공여부	**	1.78
	사업체 종사자 수	*	11.49
	유급휴가 혜택 제공여부	*	1.83
	법정퇴직금/퇴직연금 혜택 제공여부	**	1.63
	상여금 혜택 제공여부	**	1.78
직장에 대한 개인의 만족도	직장에 대한 전반적인 만족도	**	1.54
개인의 환경	현재 부모님의 월평균 소득	*	6.11
	졸업 대학의 전반적인 만족 수준	**	3.54
	졸업 전공의 전반적인 만족 수준	**	4.79

<표 2>는 선정된 관찰변수에 대한 설명이다. 실제 직장을 다녀야만 정확한 정보를 알 수 있는 관찰변수가 주를 이루었으며, 일자리 지원 시 무엇을 우선적으로 고려하는지를 판별하는 중요도나 졸업대학 관련 만족도 등 개인에 따라 다른 결과값

3) *: $0.01 \leq p < 0.05$, **: $p < 0.01$

을 갖는 개인 특성 관련 변수가 중요변수의 가장 큰 비율을 차지하였다.

추출된 각 중요변수와 종속변수간의 관계를 살펴보았다. 개인의 경우, 졸업 전 취업 목표가 없는 사람일수록 이직률이 높았다. 게다가, 출퇴근 거리, 근로시간, 업무량을 중요하게 생각할수록 이직률이 높다는 사실을 토대로 개인의 삶의 질 관련 요소가 이직에 영향을 미친다는 것을 확인 가능하다. 또한 퇴직연금, 시간외수당, 상여금, 국민연금, 건강보험 등의 혜택을 제공하는 기업일수록 이직률이 낮음을 알 수 있다. 마지막으로 업무 난이도, 개인 발전 가능성과 직업 자체의 전망, 고용안정성을 중요하게 생각할수록 이직률이 낮았으며 이는 이직을 할 때 직업에 대한 미래 지향적인 고려를 함께함을 알 수 있다.

2. 분류 예측 모델링 결과

구조방정식을 통해 선택된 중요변수 23개를 토대로 분류 예측 모형을 구축한다. 이때 예측 결과에 대한 성능은 정확도, 재현율, 정밀도, F1-스코어로 산출하여 평가한다. 본 연구에서는 실제로 이직할 자를 이직할 자로 예측하는 것이 중요하므로, 정확도 뿐만 아니라 실제 양성인 것 중 예측 모형이 양성으로 판정한 비율인 재현율을 중심으로 성능을 비교한다.

<표 3> 중요변수 23개 기반 머신러닝 모형 성능

모델	정확도	재현율	정밀도	F1-스코어
Logistic Regression	0.60	0.60	0.60	0.60
K-Nearest Neighbor	0.75	0.75	0.79	0.74
Naive Bayes	0.59	0.59	0.59	0.59
Decision Tree	0.83	0.83	0.83	0.83

<표 4> 전체 82개 변수 기반 머신러닝 모형 성능

모델	정확도	재현율	정밀도	F1-스코어
Logistic Regression	0.73	0.73	0.73	0.73
K-Nearest Neighbor	0.75	0.75	0.82	0.73
Naive Bayes	0.71	0.71	0.71	0.71
Decision Tree	0.83	0.83	0.83	0.83

우선 Logistic Regression, K-Nearest Neighbor, Naive Bayes, Decision Tree를 포함한 머신러닝 기반 이진 분류 모형 4가지를 활용한다. 전처리 결과로 추출된 총 82개의 변수로 구축한 예측 모형에 더불어, 앞서 구조방정식으로 추출한 중요변수 23개를 독립변수로 구축한 예측 모형의 성능을 비교하여 구조방정식을 통해 선택된 최종 변수들의 효과성을 검증한다. 그 결과는 위의 <표 3>, <표 4>와 같다. 중요변수 23개로 구성된 Logistic Regression과 Naive Bayes 기반의 모형은 모두 재현율

과 F1-스코어는 약 0.60 정도로 비슷한 성능을 보이며 4가지 모형 중 낮은 성능을 보인다. 반면 Decision Tree를 기반으로 한 예측 모형의 경우 정확도, 재현율, 정밀도, F1-스코어 면에서 모두 0.83을 기록하며 4가지 모형 중 가장 높은 성능을 보인다. 중요변수 23개로 구축한 모형의 경우 전체 82개의 모든 변수를 사용했을 때와 비교해 변수의 개수는 1/4로 줄임과 동시에 비슷한 정확도, 재현율, 정밀도 및 F1-스코어를 보인다. 따라서 중요변수 23개로 구축한 모형이 전체 변수를 사용했을 때보다 더 계산 비용이 적은 동시에 성능은 비슷하기 때문에 보다 효과적인 모형이라고 판단한다.

<표 5> 중요변수 23개 기반 앙상블 모형 성능

모 델	정 확도	재 현율	정 밀도	F1-스코어
Decision Tree	0.83	0.83	0.83	0.83
Random Forest	0.90	0.90	0.91	0.90
LightGBM	0.89	0.89	0.91	0.89
XGBoost	0.89	0.89	0.90	0.89

<표 6> 전체 82개 변수 기반 앙상블 모형 성능

모 델	정 확도	재 현율	정 밀도	F1-스코어
Logistic Regression	0.83	0.83	0.83	0.83
Random Forest	0.92	0.92	0.93	0.92
LightGBM	0.92	0.92	0.93	0.92
XGBoost	0.91	0.91	0.92	0.91

다음으로는 앞서 구축한 4가지 모형 중 Decision Tree 모형이 가장 높은 성능을 보여주었기 때문에, Decision Tree 기반의 3가지 앙상블 모형인 Random Forest, LightGBM, XGBoost를 이용해 예측 모형을 구축하고 이를 기본 Decision Tree 기반의 예측 모형과 그 성능을 비교해본다. 그 결과는 <표 5>, <표 6>과 같이 Random Forest 모형이 정확도 0.92, 재현율 0.92, 정밀도 0.93, F1-스코어 0.92로 가장 성능이 높았다. 마찬가지로 중요변수 23개로 구축한 모형의 경우 전체 82개의 모든 변수를 사용했을 때와 거의 비슷한 성능을 보여 구조방정식을 통해 선택된 최종 변수들의 효과성을 검증하였다.

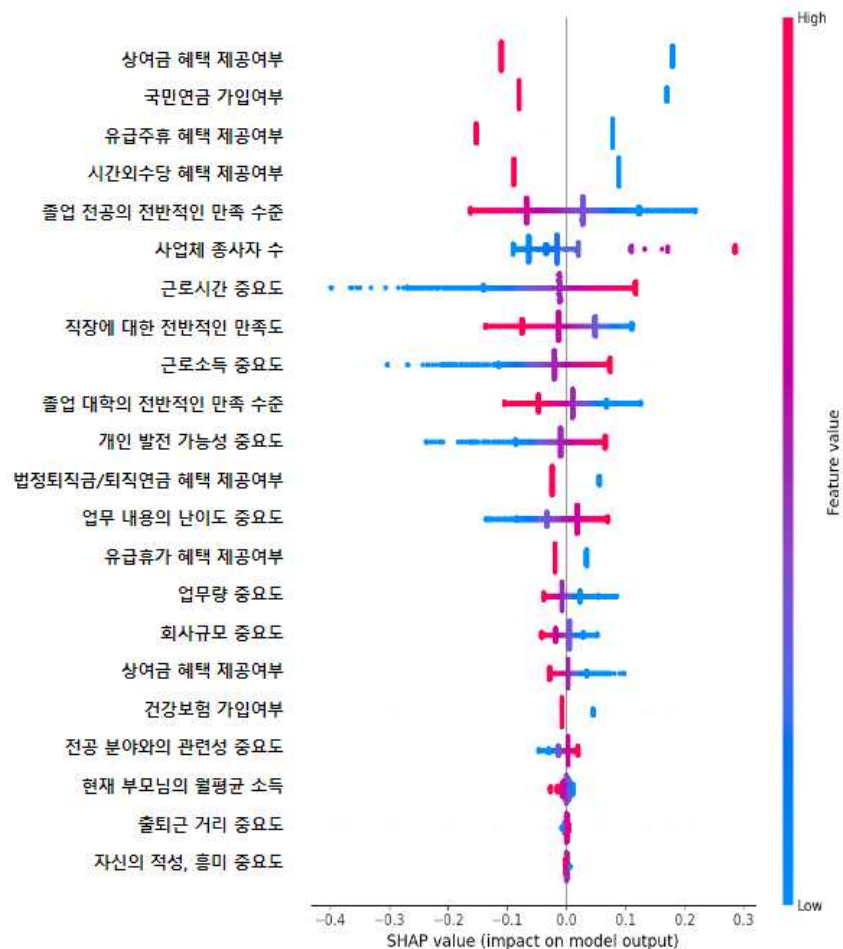
<표 7> 중요변수 23개 기반 베이지안 최적화 성능

모델	정확도	재현율	정밀도	F1-스코어
Decision Tree	0.84	0.84	0.84	0.84
Random Forest	0.91	0.91	0.91	0.91
LightGBM	0.90	0.90	0.91	0.90
XGBoost	0.90	0.90	0.91	0.90

위의 실험 결과를 토대로 하이퍼파라미터 튜닝 방법인 베이지안 최적화 기법을 통하여 모형의 성능을 고도화한다. 그 결과, 각 모형에 대한 최적의 max_depth, min_samples_split 및 n_estimators나 num_leaves와 같은 하이퍼파라미터 값을 찾을 수 있었으며, 이렇게 생성된 모형의 성능 중 가장 높은 랜덤 포레스트 기반 모형이 정확도, 재현율, 정밀도, F1-스코어 0.91을 기록하며 가장 안정적인 성능을 보였다. 따라서 본 연구에서는 max_depth가 399이며, min_samples_split이 4이고, n_estimators가 1788인 Random Forest 모형을 최종 분류 예측 모형으로 선정한다.

3. 모형 해석

<그림 6> SHAP을 통한 변수 해석 결과 시각화



앞서 선택한 최종 모형에 대하여 SHAP 알고리즘을 통하여 예측 결과를 해석해보므로써 4가지 연구 가설을 검증한다. <그림 6>의 해석을 통해 이직과 변수 간 인과관계를 찾을 수 있다. 몇 가지 변수와 이직과의 관계를 살펴보자면 우선, 직장의 특성인 대기업무일수록(종사자 수가 많을수록) 이직률이 높으며, 직장에서의 유급주휴 혜택을 제공할수록 이직률이 높다. 또한 직장에 대한 개인의 특성인 직장에서의 퇴직 연금이 실제 높거나 근로소득을 중요하게 생각할수록 이직률이 낮은 결과를 보여준다. 직장에 대한 개인의 만족도 관련된 일자리에 대한 전반적인 만족도 변수는 만족도가 높을수록 이직 가능성을 줄임을 알 수 있다. 마지막으로 개인의 환경 특성과 관련된 졸업 대학의 전공에 대한 전반적인 만족 수준이 높을수록 이직률이 낮다는 사실도 확인 가능하다. 따라서 직장의 특성, 직장에 대한 개인의 특성, 직장에 대한 개인의 만족도, 개인의 환경 요인은 이직과 인과관계를 가지며, 만족도의 경우 만족도가 높을수록 이직률이 줄어든 것이라는 본 연구의 가설이 성립한다.

V. 논의 및 결론

1. 연구 결과의 요약

본 연구는 청년세대의 이직 요인을 심층적으로 분석하고, 이를 바탕으로 개인의 이직 여부를 예측할 수 있는 모형을 개발하는 것을 목표로 하였다. 이를 위해 2010년부터 2019년까지의 한국고용정보원의 대졸자 직업 이동 경로 조사 데이터(Graduate Occupational Mobility Survey, GOMS)(통계청 승인번호: 327004) 데이터를 활용하였다. 다양한 전처리와 구조방정식을 통해 직장과의 특성, 직장 만족도, 개인의 환경 요인 등 총 23개의 주요 요인을 도출하였다.

이후, 도출한 주요 요인을 토대로 머신러닝 기반 분류 예측 알고리즘과 베이지안 최적화 기법을 적용하여 이직 여부를 예측하는 모형을 개발하였다. 이진 분류 모형인 Logistic Regression, K-Nearest Neighbor, Naive bayes, Decision Tree를 비교한 결과, Decision Tree 모형이 가장 높은 예측 성능을 보였다. 나아가 가장 높은 성능을 보인 Decision Tree 기반의 앙상블 모형인 Random Forest, LightGBM, XGBoost 모형의 성능을 비교한 결과, Random Forest 모형이 가장 높은 예측 성능을 보였다. 연구 결과, 직장의 특성, 개인의 특성, 직장 만족도, 개인의 환경 요인이 이직과 밀접한 관계가 있음을 확인하였다. 대기업이나 유급주휴 혜택이 있는 직장일수록 이직률이 높았고, 퇴직 연금이 높거나 근로소득을 중요시하는 경우와 전반적인 직무 만족도가 높은 경우가 이직률이 낮았다. 또한, 전공에 대한 만족도가 높을수록 이직률이 낮은 것으로 나타났다.

2. 연구의 의의

본 연구는 청년세대의 이직을 다각적으로 분석하여 이직 요인에 대한 종합적인 이

해를 제공한다는 점에서 의의가 있다. 기존 연구들이 소수의 특정 요인에만 집중한 것과 달리, 본 연구는 다양한 독립 변수들을 고려하여 종합적인 이직 예측 모델을 개발하였다. 특히, 이직 요인을 구조방정식과 머신러닝을 통해 분석한 것은 기존 연구와 차별화되며, 이는 다양한 요인들이 이직에 미치는 영향에 대하여 분석하고 이의 신뢰성을 높이는 데에 기여한다.

더 나아가, 본 연구는 청년세대의 이직 요인을 보다 명확히 이해함으로써 정부와 기업이 청년층의 근속률을 향상시키기 위한 보다 효과적인 정책과 제도를 수립하는데 중요한 기초 연구로 작용할 수 있다. 예를 들어, 연구 결과를 토대로 정부는 청년 고용 정책을 개선하고, 기업은 인사 관리 전략을 수정하여 청년세대의 조기 이직률을 낮출 수 있을 것이다. 이는 청년세대의 노동시장 안정화에 기여하며, 기업의 인재 유출 방지도 중요한 역할을 할 것으로 기대된다.

마지막으로, 본 연구는 청년세대의 노동시장 내 행동 패턴을 이해하는 데에도 중요한 기여를 한다. 청년들이 왜 이직을 고려하는지, 어떤 요인들이 이직에 가장 큰 영향을 미치는지를 명확히 파악함으로써 정책 입안자와 기업 관리자는 청년들이 직장에서 장기간 근속할 수 있도록 지원 방향을 마련할 수 있다. 이는 장기적으로 청년세대의 직업 안정성을 높이고, 경제 전반의 생산성을 향상시키는 데 긍정적인 영향을 미칠 것이다.

3. 연구의 시사점

첫째, 대기업에서의 높은 이직률은 근로 환경 및 조직 문화 개선의 필요성을 시사한다. 대기업은 유급주휴 혜택 등의 복지 혜택을 제공하면서도 이직률이 높아질 수 있으므로, 복지 혜택의 질적 향상과 직원들의 의견을 반영한 맞춤형 복지 정책이 필요하다. 이를 통해 대기업 내 일과 삶의 균형을 개선하고 직원 만족도를 높이는 방향으로 정책을 설계하는 것이 중요하다.

둘째, 퇴직 연금이 높거나 근로소득을 중요시할수록 이직률이 낮다는 결과는 안정적인 연금제도와 공정한 급여 체계가 이직률 감소에 기여할 수 있음을 보여준다. 따라서 정부와 기업은 청년 근로자들에게 안정적인 연금 혜택을 제공하고 성과에 따른 공정한 급여 체계를 구축하는 것이 중요하다. 이를 통해 청년층의 경제적 안정을 도모하고, 장기근속을 유도할 수 있다.

셋째, 직무에 대한 전반적인 만족도가 높을수록 이직률이 낮아지는 결과는 지속적인 직무 교육 및 경력 개발 프로그램의 필요성을 강조한다. 기업은 직무 만족도를 높이기 위해 직무 교육, 경력 개발 프로그램, 상호 피드백 시스템 등을 도입하는 방법을 고려해볼 수 있다.

넷째, 졸업 대학에서 전공에 대한 만족도가 낮을수록 이직률이 낮다는 결과는 전공 불일치로 인한 직업 선택의 유연성을 높이고, 다양한 경로를 통한 경력 개발이 가능하도록 지원하는 정책이 필요함을 시사한다. 대학 교육과 직업 간 연결 고리를 강화하고, 청년들이 자신의 전공과 무관하게 다양한 직업 기회를 탐색할 수 있도록 지원하는 것이 중요하다.

4. 연구의 한계 및 후속연구 제언

본 연구에서는 「대졸자 직업 이동 경로 조사」의 2010년부터 2019년까지의 데이터를 사용하였다. 이중 2019년 데이터는 코로나 팬데믹의 영향이 반영된 시점 이후인 2020년 9월 이후에 수집되었다. 이에 따라 팬데믹의 영향력을 배제할 수 없으며, 이는 연구 결과 해석에 한계로 작용한다. 청년층의 인식과 행동 특성은 시계열적으로 특정 방향으로 변화하는 경향이 있으나, 팬데믹과 같은 사회적 위기 상황이 발생하면 일시적으로 다른 방향으로 변화할 수 있다. 따라서 후속 연구에서는 조사 대상 및 시점을 적절히 선정할 필요가 있다. 동 이유로, 「대졸자 직업 이동 경로 조사」는 팬데믹 전후를 비교할 수 있는 데이터가 없기 때문에 팬데믹의 영향력을 명확히 분리하여 분석하기 어려우므로, 후속 연구에서는 팬데믹 전후 시점을 비교할 수 있는 장기적인 데이터를 수집하여 분석하는 것이 필요하다. 이를 통해 시간의 흐름에 따른 이직 요인의 변화를 더 정확히 파악할 수 있을 것이다.

추가로, 다른 최신 알고리즘이나 기술을 도입하여 모형의 성능을 더욱 향상시킬 수 있을 것이다. 특히 딥러닝 활용한 예측 모형 개발이나 강화학습을 통한 지속적인 모형 개선이 고려될 수 있다. 또한, 청년층의 이직 결정에 영향을 미치는 사회적, 문화적 요인들에 대한 정성적 연구를 병행하여 보다 종합적인 이해를 도모할 필요가 있다.

참고문헌

- 문영만·홍장표(2017). "청년 취업자의 기업규모별 이직 결정요인 및 임금효과", 『산업노동연구』, 23 (2): 195-230.
- 박진아(2016), "대졸 청년층의 직장만족이 이직준비 여부에 미치는 영향", 『산업진흥연구』, 1.1: 35-40.
- 배성숙(2021), "대졸 청년층의 직무일치도와 직무만족도가 이직의도에 미치는 영향과 고용형태의 조절효과" 『기업경영리뷰』, 12 (3): 187-201.
- 송민선·박정윤(2023), "삶 지향 태도와 일 가치관이 미혼 청년 임금근로자의의직의도에 미치는 영향", 『인문사회』, 21 14 (3): 1809-1824.
- 안세진·엄기수·서철승(2023), "MZ세대 근로자의 일과 삶의 균형과 조직지원인식이 이직의도에 미치는 영향에 관한 연구", 『고객만족경영연구』, 25 (2): 119-139.
- Li, Silin(2022), "Construction and Evaluation of Employee Turnover Prediction Model based on Data Mining Algorithm" 2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI) : 531-535.
- Liao, Chenyu(2023), "Employee turnover prediction using machine learning models" International Conference on Mechatronics Engineering and Artificial Intelligence.
- Sardar, Devadarshan K, Sidhant Chourasiya and V. Vijyalakshmi(2023), "Employee Turnover Prediction by Machine Learning Techniques", 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT) : 265-272.
- <https://www.donga.com/news/Economy/article/all/20240401/124269295/1>(2024.04.02)
- https://www.moel.go.kr/news/enews/report/enewsView.do?news_seq=16352(2024.03.24)