

Discovering Youth Job Search Bundles Using Association Rule Mining

Hyoungrae Kim*, Jongdeok Lim*, Jeongrae Kim**

*Researcher, Korea Employment Information Service, Eumseong-gun, Korea

**Professor, Dept. of Artificial Intelligence Software Convergence, Korea Polytechnics Chuncheon Campus, Gangwon, Korea

[Abstract]

Given the persistent challenges of youth unemployment and skills mismatch, a granular understanding of how young job seekers actually prepare for the labor market is critical for policy intervention. While most research on this topic examines the effectiveness of individual activities, understanding the combinations of these activities as strategic 'patterns' is crucial for developing effective employment policies. For this purpose, this study utilizes youth panel data, employing association rule mining (ARM) and logistic regression. The analysis reveals that job search patterns are distinctly differentiated by sociodemographic characteristics such as gender, educational attainment, and region. For instance, a strong association was found between {female residents in the metropolitan area} and {preparing for official English tests} (Lift=1.282). Furthermore, residents in the metropolitan area tended to focus on preparing for practical, job-related certifications. This suggests that youth job search activities are not uniform but are the result of strategic choices based on an individual's socio-structural position. These findings highlight the need for customized support policies tailored to the specific characteristics of different groups.

▶ **Key words:** Youth job seekers, Job preparation, Association rule mining, Panel data, Competency preparation activities, Job search effort activities, Customized employment support

[요 약]

지속적인 청년 실업과 기술 불일치라는 사회적 과제를 고려할 때, 청년 구직자들이 어떻게 취업 준비하는 지에 대한 이해가 정책 마련을 위해 중요하다. 기존 청년 구직 연구가 개별 활동의 효과성에 집중한 반면, 본 연구는 활동 간 조합과 연관성에 주목해 구직 준비 '패턴'을 실증 분석했다. 이를 위해 청년패널자료를 연관규칙분석 및 로지스틱 회귀분석으로 분석했다. 분석 결과, 이러한 구직 패턴은 성별, 학력, 지역 등 인구사회학적 특성에 따라 뚜렷하게 분화되었다. 예를 들면, {여성, 수도권} 거주자의 경우 {공인영어성적} 준비가 높게(Lift=1.282) 나타났다. 또한 수도권 거주자는 실용적인 직무 관련 자격증 준비에 집중하는 경향을 보였다. 이는 청년 구직 활동이 획일적이지 않으며, 개인의 사회구조적 위치에 따른 전략적 선택의 결과임을 시사한다. 이러한 발견은 각 집단의 특성을 고려한 맞춤형 지원 정책의 필요성을 제기한다.

▶ **주제어:** 청년 구직자, 취업 준비, 연관규칙분석, 패널데이터, 역량 준비 활동, 취업 노력 활동, 맞춤형 취업지원

- First Author: Hyoungrae Kim, Corresponding Author: Jeongrae Kim
- *Hyoungrae Kim (goddoes8@gmail.com), Korea Employment Information Service
- *Jongdeok Lim (jdpro@naver.com), Korea Employment Information Service
- **Jeongrae Kim (ceright@gmail.com), Dept. of Artificial Intelligence Software Convergence, Korea Polytechnics Chuncheon Campus
- Received: 2025. 09. 08, Revised: 2025. 09. 29, Accepted: 2025. 10. 22.

I. Introduction

청년층의 고용 불안정성은 전 세계적인 정책 과제이며, 특히 한국에서는 '고스펙 청년 실업'이라는 역설로 심화되고 있다[1]. 높은 수준의 인적자본을 축적했음에도 다수의 청년들이 장기 구직을 경험하는 현상은, 거시 경제나 산업 구조만으로는 온전히 설명되지 않는다. 이러한 괴리의 중심에는 구직자가 수행하는 취업준비 활동의 이질성(heterogeneity)이 존재한다.

기존 연구들은 주로 인적자본이론[2]이나 신호이론[3]에 근거하여 학력, 자격증 등 개별 자산(asset)이 취업 성과에 미치는 효과를 검증하는 데 집중해왔다. 그러나 이러한 분석적 접근은 구직 과정이 상호 보완적인 활동들로 구성되는 전략적 '묶음(bundle)'이라는 본질을 간과하는 한계를 지닌다.

본 연구는 이러한 공백을 넘어, 취업준비 활동들의 '구성'에 주목하는 통합적 관점을 제안한다. 즉, "청년 구직자들은 어떤 활동들을 함께 묶어서 수행하는가?" 그리고 "이러한 활동의 조합 패턴은 인구사회학적 특성에 따라 어떻게 분화되는가?"라는 질문에 답하고자 한다. 이를 위해 본 연구는 행동 간의 비선형적 연관성과 잠재적 규칙을 탐색하는 데 강점을 지닌 연관규칙분석(Association Rule Mining, ARM)을 핵심 연구 방법으로 채택한다[4, 5]. 이를 통해 구직 전략의 잠재적 유형을 식별하고, 로지스틱 회귀분석으로 배경 변수에 따른 이질성을 검증하여, 데이터 기반의 개인화된 정책 설계를 위한 실증적 토대를 제공하고자 한다.

II. Theoretical Background

1. Job Search Behavior

구직자 행동을 설명하는 이론적 논의는 인적자본이론[2], 신호이론[3], 그리고 탐색이론을 중심으로 전개되어 왔다. 이러한 이론적 틀을 바탕으로, 선행 연구들은 다양한 방법론을 통해 청년 구직 전략을 유형화하려는 시도를 해왔다. 예를 들어, 군집분석(Cluster Analysis)은 유사한 활동 패턴을 보이는 구직자 집단을 분류하는 데 사용되었으나[6], 연구자가 사전에 정의한 거리 측정 방식에 결과가 민감하다는 한계가 있다. 또한, 회귀모형(Regression Models)은 특정 활동이 취업 성과에 미치는 영향을 분석하는 데 유용하지만, 다수의 활동이 복합적으로 조합되어 나타나는 '묶음(bundle)'의 효과를 모델링하는 데는 제약

이 따른다. 이러한 연구들은 대부분 개별 활동의 효과를 분석하거나 구직자 그룹을 유형화하는 '원자론적 접근(an atomistic approach)'에 머물렀다. 이는 실제 구직자들이 다수의 활동을 전략적으로 '조합'하여 포트폴리오를 구성하는 현실을 간과하는 한계를 낳는다. 따라서 본 연구는 활동들을 하나의 '구성체'로 바라보는 통합적 관점을 통해, 기존 방법론들이 포착하지 못한 청년 구직 전략의 내재적 구조를 규명하고자 한다.

2. Youth as a Period of Uncertain Transition

현대 사회에서 청년기는 교육에서 노동시장으로 이행(School-to-Work Transition)하는 과정이 길고 복잡해진다는 특징을 가진다[7]. 경력 자산이 부족한 청년들은 교육 인플레이션 속에서 과도한 '스펙' 경쟁에 직면하며, 장기화된 구직 과정은 자존감 하락 등 심리적 자원을 소진시킨다[8]. 이처럼 현대 청년층의 구직 과정은 단순히 일자리를 찾는 행위를 넘어, 불확실성 속에서 자신의 가치를 증명하려는 복합적인 전략적 행위로 이해해야 한다.

3. Association Rule Mining in Social Science

본 연구의 핵심 질문인 '구직 활동의 조합 패턴'을 탐색하기 위해 연관규칙분석[4]을 채택한 근거는 데이터의 특성과 분석의 목적에 있다. 청년 구직 데이터는 다수의 활동 참여 여부를 나타내는 이진형 변수들로 구성된다. 이때 핵심 과제는 독립 변수와 종속 변수 간의 선형적 관계를 찾는 것이 아니라, 수많은 활동들 사이에서 어떠한 조합이 의미 있게 함께 나타나는지를 발견하는 것이다. 이러한 목적에 있어 단순 교차분석은 두 활동 간의 관계는 보여줄 수 있지만, 셋 이상의 활동으로 구성된 복합적인 패턴을 탐색하기는 어렵다. 네트워크 분석은 활동을 노드로 설정할 경우 양자관계나 전체적인 구조를 보여주는 장점이 있다[9]. 반면 ARM은 셋 이상의 활동으로 구성된 다차원적인 조합 패턴(전략적 묶음)을 발견하고, 그 관계의 방향성과 강도(신뢰도, 향상도)를 정량적으로 제시하는데 더 특화되어 있다. 요인분석(Factor Analysis)은 변수들을 소수의 잠재 요인으로 축소하는 데 유용하지만, 구직 활동처럼 각각이 독립적인 전략적 선택일 수 있는 경우에 부적합할 수 있다[9]. ARM은 생활 습관 위험 행동의 패턴을 분석[10]하거나 인터넷 사용 패턴을 파악[11]하는 등의 다양한 사회과학 연구에 성공적으로 사용되었다. 따라서 ARM은 기존 연구가 간과했던 '구직 전략 묶음'의 실체를 탐색적으로 밝히는 데 최적화된 모델이다.

III. Research Methods

본 연구는 청년층의 취업준비 활동 패턴을 다각적으로 분석하기 위해, 탐색적 데이터 마이닝과 예측 모델링을 결합한 단계적 접근법을 채택했다. 1단계에서는 핵심 방법론인 연관규칙분석을 적용하여 활동들 간의 조합적 연관 구조, 즉 '구직 전략 묶음'을 탐색한다. 2단계에서는 로지스틱 회귀분석을 통해 특정 활동 참여에 영향을 미치는 인구사회학적 요인의 개별적 효과를 추정함으로써, 탐색적 결과를 보완하고 심화한다.

1. Discovering Job Search Strategies

청년 구직자들이 수행하는 활동들 간의 잠재적 조합 패턴을 발견하기 위해, 본 연구에서는 각 구직자를 '거래(transaction)', 각 취업준비 활동을 '항목(item)'으로 간주하여 분석을 수행한다. 생성된 규칙들 중에서 의미 있는 패턴만을 선별하기 위해, 세 핵심 지표를 활용한다.

- 향상도 (Lift): 두 활동이 통계적으로 독립일 때의 기대 확률 대비 실제 동시 발생 확률의 비율. Lift 값이 1보다 클 경우 두 활동 간에 우연 이상의 양(+)의 관계가 존재함을 의미하며, 규칙의 유용성을 판단하는 척도
- 신뢰도 (Confidence): 특정 활동(e.g., 인터넷 구직정보 습득)을 한 구직자가 다른 활동(e.g., 이력서 제출)도 수행했을 조건부 확률로, 규칙의 예측력을 의미
- 지지도 (Support): 전체 구직자 중 특정 활동 조합(e.g., {여성, 수도권})을 동시에 수행한 비율로, 규칙의 보편성을 나타냄

또한, 본 연구는 연관규칙 탐색 과정에 이론적 제약조건을 설정하여, 분석 결과의 해석 타당성을 확보하고 데이터의 내적 일관성을 검증하는 장치로 활용한다. 구직 과정의 논리적 선호관계를 고려할 때, 개인의 인구사회학적 특성은 취업준비 활동에 선행하며, 장기적인 역량 축적(역량 준비)은 단기적인 구직 실행(취업 노력)에 앞서는 것이 일반적이다. 이에 따라 다음 세 가지 제약조건을 설정했다.

- 제약조건 1: 인구사회학적 변수 → 역량 준비 활동
- 제약조건 2: 인구사회학적 변수 → 취업 노력 활동
- 제약조건 3: 역량 준비 활동 → 취업 노력 활동

먼저, 전체 활동 항목(item) 집합 I 를 다음과 같이 세 개의 상호 배타적인 부분집합으로 정의한다.

- I_D : 인구구조변수 항목 집합(예, {지역=수도권})
- I_C : 역량 준비 활동 항목 집합(예, {공인영어성적=참여})
- I_E : 취업 노력 활동 항목 집합(예, {이력서제출=참여})

구직과정의 논리적 선호관계를 반영하면, 제약을 위반하는 규칙들은 다음과 같이 정의된다:

$$R_{invalid} = \{ (X \rightarrow Y) \mid (Y \cap I_D \neq \emptyset) \vee (X \subseteq I_E \wedge Y \subseteq I_C) \} \quad (1)$$

일반적으로 제약조건은 방대한 규칙 후보군을 사전에 제거하여 연산 효율성을 높이는 데 사용된다[12]. 그러나 본 연구에서는 이러한 접근법을 달리 적용하여, 먼저 모든 유의미한 규칙을 도출한 뒤 이 규칙들이 상기된 제약조건을 위반하는지를 사후적으로 검증했다. 만약 제약조건에 위반되는 규칙이 발견되어 $R_{found} \cap R_{invalid} \neq \emptyset$ 이라면, 이는 설문 응답 과정에서의 오류나 데이터의 잠재적 문제를 시사할 수 있기 때문이다[13]. 따라서 이 과정은 설문 데이터의 신뢰도를 간접적으로 확인하는 중요한 절차라 할 수 있다.

본 분석에서는 지지도와 신뢰도 기준은 최소한으로하여 1차적으로 규칙을 생성한 후, 향상도(Lift)가 1.1 이상이고 해당 규칙의 카이제곱 검정 p-value가 0.05 미만인 통계적으로 유의미한 규칙들만을 최종 분석 대상으로 선정했다[14]. 이를 통해 해석 가능성이 높고 신뢰할 수 있는 구직 전략 묶음을 도출하고자 했다.

2. Determinants of Activity Participation

마지막으로, 연관규칙분석이 활동 '간의' 관계에 집중하는 것과 상보적으로, 개별 활동 '참여 자체'에 영향을 미치는 요인을 식별하기 위해 이항 로지스틱 회귀분석(binary logistic regression)을 수행했다. 각 취업준비 활동 참여 여부(참여=1, 불참=0)를 종속변수로 설정하고, 성별, 연령대, 최종학력, 거주지역을 독립변수로 투입했다. 이 모델은 각 인구사회학적 요인이 특정 활동 참여 확률에 미치는 독립적인 영향력을 오즈비(Odds Ratio, OR)를 통해 설명해준다[15]. 이 분석은 ARM을 통해 발견된 특정 집단의 조합 선호 경향에 대한 통계적 추론의 깊이를 더해주는 역할을 한다.

이항 로지스틱 회귀분석과 오즈비(OR)의 관계는 다음 수식을 통해 설명된다.

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_j x_j \quad (2)$$

오즈에 대한 수식으로 만들기 위해 양변에 자연상수 e 를 지수화한다. 오즈비는 두 오즈의 비율이다. 따라서, 오즈비(OR)는 독립변수의 입력값이 1단위 변했을 때, 종속변수의 오즈(Odds)가 몇 배 변하는지를 나타내는 값이다. 예를 들어 오즈비가 1.5라면 독립변수가 1단위 변할 때 종

속변수가 발생할 오즈가 1.5배가 된다고 해석할 수 있다. 로지스틱 회귀모델의 특정 계수 β_j 를 오즈비로 변환하는 수식은 다음과 같다.

$$OR = \frac{Odds_1}{Odds_2} = \frac{e^{\beta_0 + \beta_j}}{e^{\beta_0}} = e^{\beta_j} \quad (3)$$

총 K 개의 역량준비 및 취업노력 활동 각각에 대해 별도의 모델을 구축하고, i 번째 활동의 참여여부에 대한 개인의 인구사회학적 특성을 c 라고 구성한다. 이때 각 활동(i)이 인구사회학적 특성에 통계적으로 유의미한 영향을 받는지를 판별할 수 있다. 전체 활동 집합 $Y = \{y_1, y_2, \dots, y_K\}$ 중에서 인구사회학적변수에 의해 유의미한 영향을 받은 인구구조 종속적 활동의 부분집합 Y_S 를 다음과 같이 정의한다.

$$Y_S = \{y_i \in Y \mid \exists j \in \{1, \dots, c\} s.t. p_{i,j} < \alpha\} \quad (4)$$

여기에서 $p_{i,j}$ 계수에 대한 p-value 이며 유의수준은 본 연구에서 0.05이다. 즉, 통적 활동에 대한 회귀모델에서 하나 이상의 인구사회학적변수 계수가 통계적으로 유의하게 나타날 경우 해당 활동을 인구구조종속적 활동으로 분류했다. 이 절차는 청년들의 구직 활동 중 어떤 것들이 보편적 행동이고, 어떤 것들이 사회구조적 위치에 따라 분화하는지를 식별하는 근거가 된다. 다만, 어떤 활동이 Y_S 집합에 포함되었다고 해서 모두 동일한 수준으로 사회구조적 위치에 따라 분화된다고 볼 수는 없다. 영향요인의 광범위성과 영향요인의 강도 등을 다차원적으로 기준을 함께 고려하여야 한다. 구직활동을 대부분의 청년이 수행하는 보편적 활동, 특정 상황에서만 분화되는 특화적 활동, 여러 사회구조적 요인이 복합적으로 작용하는 고도로 구조화된 활동 등 다양하게 유형화하여 해석할 필요가 있다. 이러한 분석 결과는 청년 구직 전략의 복잡성을 보다 정밀하게 포착하기 위한 장치이다.

데이터 처리 및 분석은 Python 3.10 환경에서 수행되었으며, pandas와 scikit-learn 라이브러리를 데이터 전처리와 연관규칙분석에, statsmodels 라이브러리를 회귀분석에 활용했다[16].

IV. Datasets and variables

1. Data Source and Sample

본 연구는 한국고용정보원(Korea Employment Information Service, KEIS)에서 제공하는 청년패널조사 (Youth Panel, YP) 2021년 1차 웨이브 자료를 활용했다

[17]. YP는 대한민국 청년층을 대표하도록 설계된 대규모 종단면 조사로서, 청년들의 교육, 직업훈련, 구직 활동, 노동시장 이행 과정 전반에 대한 포괄적이고 심층적인 정보를 수집하고 있다. 특히, 본 연구의 핵심 주제인 다양한 취업준비 활동에 대한 상세한 문항을 포함하고 있어, 청년 구직 전략의 이질성을 분석하는 데 최적의 자료라 할 수 있다. YP 2021의 전체 응답자 12,213명 중, 조사 시점에서 하나 이상의 취업준비 활동을 수행했다고 응답한 1,919 명의 구직자를 최종 분석 대상으로 한정했다.

2. Measures

분석에 사용된 변수는 크게 취업준비 활동 변수와 인구사회학적 특성 변수로 구성된다. 모든 변수는 연관규칙분석 및 로지스틱 회귀분석에 적합한 형태로 변환하여 사용했다.

2.1. Job Preparation Activities

원자료의 총 27개 취업준비 활동 항목을 분석 대상으로 삼았다. 각 활동은 응답자의 참여 여부에 따라 ‘참여(1)’ 또는 ‘불참(0)’의 이진 변수(binary variable)로 변환했다. 연관규칙분석 결과의 해석적 깊이를 더하기 위해, 활동의 기능적 특성에 따라 다음과 같이 역량 및 스펙 준비활동과 구직활동으로 구분할 수 있다.

Table 1 에서 역량준비는 노동시장에서의 경쟁력을 강화하기 위한 인적자본 투자 활동들이다. 이는 크게 어학능력, 자격 및 경력, 학업/경쟁력, 기타 활동으로 구분할 수 있다.

구직자가 자신의 역량 및 스펙을 갖춘 후에는 취업노력을 하게 된다. 구직활동은 직접적인 활동과 비공식적인 네트워크를 활용하는 방식으로 구분할 수 있다. 이런 비공식적인 활동은 사회적 자본을 활용하는 비공식적 경로의 활동들이다.

Table 1. Competency Preparation and Job Search Effort

구분	역량준비	구분	취업노력
어학 능력	공인영어성적	직접 적활 동	사설학원수강
	영어회화능력		직업적성/흥미검사
	제2외국어능력		면접준비/훈련
	한자능력		공공기관 구직등록
자격 및 경력	컴퓨터 자격증	동	이력서 제출
	업무/직무 관련 자격증		인터넷 구직정보 습득 및 등록
	고시준비		신문/정보지 구인란 구독
	인턴경험		취업전문회사 자문
학업 /경 쟁력	직무경험	네트 워크 활용	그룹 스터디
	수상경력		친인척/선배/친구에게 취업부탁
기타 활동	선박사학위	활용	선생님 (교수)에게 취업부탁
	해외경험		
	봉사활동경험		
	동아리활동		
	기타활동		

2.2. Sociodemographic Variables

구직 활동 패턴에 영향을 미칠 것으로 예상되는 주요 인구사회학적 특성을 독립변수 및 통제변수로 활용했다. 각 변수의 조작적 정의는 다음과 같다.

- 성별: '남성'과 '여성'으로 구성된 이분형 변수
- 연령: '19-21세', '22-24세', '25-27세', '28-30세'의 구간
- 최종학력: '고졸 이하', '전문대학', '대학', '대학원' 범주
- 거주지역: 거주지를 기준으로 '수도권(서울/경기/인천)', '경상권', '충청권', '호남권', '강원권', '제주권'의 6개 권역

V. Results

1. Analysis of Job Search Activity Patterns

청년 구직자들의 취업준비 활동 간 조합 패턴을 탐색하기 위해 연관규칙분석을 수행한 결과, 설정된 최소 기준($p < .05$)을 충족하는 총 44개의 통계적으로 유의미한 연관규칙이 도출되었다. 본격적인 내용 해석에 앞서, 연구방법론에서 제시한 논리적 제약조건을 기준으로 데이터의 타당성을 검증한 결과, 도출된 44개 규칙 중 제약조건을 위배하는 비논리적인 규칙은 발견되지 않았다. 이는 본 연구에 사용된 데이터의 논리적 일관성과 신뢰도가 높음을 시사하며, 이어지는 해석이 견고한 데이터에 기반하고 있음을 밝힌다. Table 2는 이 44개 규칙 중 향상도(Lift) 값을 기준으로 정렬한 주요 연관규칙의 일부를 보여준다.

연관규칙분석을 수행한 결과, 청년들의 구직 활동은 무작위적인 행동의 나열이 아니라, 몇 가지 뚜렷하고 의미 있는 '전략적 묶음'으로 구성되어 있음을 확인했다.

첫째, 구직 과정의 가장 기본적인 순차적 행동 패턴이 명확하게 드러났다. 가장 강력한 규칙은 {인터넷 구직정보 습득 및 등록} → {이력서 제출} (Lift=3.083)로 나타났다. 이는 정보 탐색 활동이 실제 지원 행위로 이어지는, 구직 과정의 가장 보편적이고 논리적인 경로를 실증적으로 보여준다.

둘째, 구직자들은 자신의 조건에 맞춰 서로 다른 '전략적 전문화' 묶음을 구성하는 것으로 나타났다. 한 가지 경로는 {거주지: 수도권} → {업무/직무 관련 자격증} (Lift=1.144) 규칙과 같이 실용적 직무 역량에 집중하는 패턴이다. 이와 대조적으로, '공인 영어성적'과 '학점 관리' 등을 연계하는 '학업 연계형 고스펙'을 구축하는 패턴도 뚜렷했다. 이러한 두 갈래의 전략은 구직자들이 노동시장에 자신의 생산성을 증명하기 위해 각기 다른 종류의 신호(signal)를 선택적으로 구축하고 있음을 시사한다[3].

셋째, 특정 인구사회학적 특성이 특정 활동 묶음을 선택할 확률을 높이는 경향이 뚜렷했다. 예를 들어, {성별: 여

성, 거주지: 수도권} → {공인 영어성적} (Lift=1.282) 규칙은 여성 및 수도권 거주라는 조건이 결합될 때, 공인 영어성적을 준비할 확률이 유의미하게 높아짐을 의미한다. 이는 성별과 지역이라는 사회적 위치가 구직 전략의 선택에 중요한 영향을 미치고 있음을 시사하는 결과이다.

요약하면, 연관규칙분석을 통해 청년 구직 활동은 단일 경로가 아닌, '기본적 구직 경로', '전략적 전문화 경로(실용적/학문적)' 등 몇 가지 이질적인 전략적 묶음으로 구조화되어 있음을 확인했다. 또한, 이러한 전략의 선택은 성별이나 거주지역과 같은 개인의 배경 특성과 밀접하게 연관되어 있음 또한 확인되었다. 이러한 탐색적 결과는 맞춤형 취업 지원 정책의 필요성에 대한 중요한 경험적 근거를 제공한다.

2. Determinants of Participation in Job Search Activities

앞선 연관규칙분석이 활동 '간의' 조합 패턴을 탐색한 것과 상보적으로, 본 절에서는 개별 활동 '참여 자체'에 영향을 미치는 인구사회학적 요인의 독립적인 효과를 분석했다. 이를 위해 총 27개의 주요 역량강화 및 취업노력 활동 각각의 참여 여부를 개별 종속변수로 하는 일련의 이항 로지스틱 회귀분석을 수행했다. Table 3 와 Table 4 은 각 모델의 분석 결과를 오즈비로 요약한 것이다. 분석 결과, 청년 구직자의 활동 선택은 개인의 배경 특성에 따라 체계적으로 분화되는 뚜렷한 경향성을 보였다. 인구사회학적 영향을 받는 항목을 살펴보면, 역량준비 활동으로는 공인영어성적, 업무/직무 관련 자격증, 직무경험 등이 나타났다. 취업노력 활동으로는 직업적성/흥미검사, 공공기관 구직등록, 선생님(교수)에게 취업부탁, 인터넷 구직정보 습득 및 등록, 이력서 제출 등이 나타났다.

2.1. Gender Differences in Job Search Strategies

구직 전략 선택에서 뚜렷한 성별 차이가 발견되었다. 특히 여성 구직자들은 남성에 비해 광범위한 역량준비 활동과 일부 취업노력 활동에서 더 높은 참여 확률을 보였다. 구체적으로, 여성은 남성보다 공인영어성적(OR=1.27, $p=.027$), 영어회화능력(OR=1.48, $p=.002$), 해외경험(OR=2.34, $p=.003$) 등 다양한 역량준비 활동에 유의하게 더 많이 참여했다. 동시에, 공공기관 구직등록(OR=1.53, $p=.037$)과 같은 공식적인 취업노력에서도 여성의 참여가 더 활발했다. 이는 여성 구직자들이 취업 활동에 보다 적극적으로 대응하기 위해, 객관적으로 검증 가능한 자격 축적과 공식적 구직 채널 활용 모두에서 남성보다 더 폭넓은 노력을 기울이는 전략적 경향성을 가짐을 시사한다[18].

2.2. Differences in Job Search Strategies by Educational Attainment

최종학력은 구직 활동의 유형을 결정하는 가장 강력한 예측 변수 중 하나였다. 특히 학력 수준이 높을수록 자신의 전문성을 입증하기 위한 활동과 학문적 네트워크에 의존하는 경향이 극명하게 나타났다. 대졸 이상 집단은 공인 영어성적(OR=4.85, p=.012)과 같은 객관적 스펙을 준비할 확률이 유의하게 높았다. 더 나아가, 석사학위 이상 소지자는 '선생님(교수)에게 취업 부탁'과 같은 비공식적 인적 자원을 활용할 확률이 다른 집단에 비해 압도적으로 높았

다(OR=13.65, p=.001). 이는 고학력 구직자들에게 학문적 네트워크가 중요한 사회적 자본으로 기능하며[19], 이를 활용하는 것이 일반적인 구직 전략의 일부임을 보여준다.

2.3. Regional Disparities in Job Search Activities

거주지역은 구직 활동의 선택에 있어 또 다른 중요한 분화 요인이었다. 수도권(기준집단)과 비수도권 거주자 간에는 뚜렷한 활동 참여의 차이가 존재했다. 수도권 거주자는 비수도권에 비해 공인영어성적, 업무/직무 관련 자격증, 직무경험 등 주요 역량준비 활동에서 전반적으로 높은 참여

Table 2. Results of Association Rule Analysis

Antecedents	Consequences	Lift	Confidence	Support	빈도	chi2	p
(취업노력) 인터넷 구직정보 습득 및 등록	(취업노력) 이력서 제출	3.083	0.688	0.132	253	563.5	0
(역량준비) 학점, (역량준비) 공인영어성적	(역량준비) 영어회화능력	3.066	0.594	0.11	211	444	0
(역량준비) 영어회화능력, (역량준비) 학점	(역량준비) 공인영어성적	3.04	0.938	0.11	211	469.8	0
(역량준비) 영어회화능력	(역량준비) 공인영어성적	2.693	0.831	0.161	309	586.7	0
(역량준비)영어회화능력, (역량준비)공인영어성적	(역량준비) 학점	2.663	0.683	0.11	211	348.7	0
(역량준비)컴퓨터 자격증, (역량준비)공인영어성적	(역량준비) 학점	2.643	0.678	0.107	206	333.6	0
(역량준비) 컴퓨터 자격증, (역량준비) 학점	(역량준비) 공인영어성적	2.598	0.802	0.107	206	335.5	0
최종학력_고졸, (역량준비) 공인영어성적	(역량준비) 학점	2.548	0.653	0.102	196	291.4	0
거주지역_수도권, (역량준비) 공인영어성적	(역량준비) 학점	2.415	0.619	0.111	213	287.1	0
(역량준비) 학점, 거주지역_수도권	(역량준비) 공인영어성적	2.414	0.745	0.111	213	297.5	0
(역량준비) 학점, 성별_여성	(역량준비) 공인영어성적	2.368	0.731	0.103	198	261.3	0
(역량준비) 영어회화능력	(역량준비) 학점	2.359	0.605	0.117	225	291.6	0
성별_여성, (역량준비) 공인영어성적	(역량준비) 학점	2.355	0.604	0.103	198	248.1	0
(역량준비) 공인영어성적	(역량준비) 학점	2.339	0.6	0.185	355	526.6	0
(역량준비) 학점, 최종학력_고졸	(역량준비) 공인영어성적	2.229	0.688	0.102	196	223.6	0
(역량준비) 학점, (역량준비) 공인영어성적	(역량준비) 컴퓨터 자격증	2.07	0.58	0.107	206	192.4	0
(역량준비) 영어회화능력	(역량준비) 컴퓨터 자격증	2.014	0.565	0.109	210	182.9	0
(역량준비) 학점	(역량준비) 컴퓨터 자격증	1.863	0.522	0.134	257	190.4	0
(역량준비) 컴퓨터 자격증	(역량준비) 공인영어성적	1.832	0.565	0.158	304	229	0
(역량준비) 컴퓨터 자격증	(역량준비) 업무/직무 관련 자격증	1.63	0.524	0.147	282	139.4	0
(역량준비) 업무/직무 관련 자격증	(역량준비) 학점	1.599	0.41	0.132	253	111.4	0
거주지역_수도권	(역량준비) 직무경험	1.536	0.209	0.103	197	82.6	0
최종학력_대졸	(역량준비) 고시준비	1.444	0.386	0.13	249	69	0
(역량준비) 공인영어성적	(역량준비) 업무/직무 관련 자격증	1.334	0.429	0.132	254	44.7	0
성별_여성, 거주지역_수도권	(역량준비) 공인영어성적	1.282	0.395	0.106	204	24.4	0
거주지역_수도권	(역량준비) 영어회화능력	1.247	0.242	0.119	228	26.7	0
최종학력_대졸	(역량준비) 공인영어성적	1.241	0.383	0.129	247	24.7	0
거주지역_수도권	(역량준비) 학점, (역량준비) 공인영어성적	1.221	0.226	0.111	213	20	0
성별_여성, 거주지역_수도권	(역량준비) 업무/직무 관련 자격증	1.218	0.391	0.105	202	15.4	0
연령대_25-27	(역량준비) 고시준비	1.213	0.324	0.124	238	19.2	0
거주지역_수도권	(취업노력) 이력서 제출	1.189	0.265	0.13	250	18.5	0
거주지역_수도권	(역량준비) 컴퓨터 자격증	1.184	0.332	0.163	313	23.9	0
거주지역_수도권	(취업노력) 인터넷 구직정보 습득 및 등록	1.183	0.227	0.112	214	14.4	0
거주지역_수도권	(역량준비) 학점	1.183	0.303	0.149	286	20.9	0
거주지역_수도권	(역량준비) 공인영어성적	1.182	0.365	0.179	344	27	0
연령대_22-24	(역량준비) 학점	1.178	0.302	0.115	221	12.5	0
성별_여성	(역량준비) 영어회화능력	1.17	0.227	0.115	220	13.2	0
거주지역_수도권	(역량준비) 업무/직무 관련 자격증	1.144	0.368	0.181	347	17.9	0
최종학력_고졸	(역량준비) 학점	1.115	0.286	0.149	285	9.1	0.003
성별_여성	(역량준비) 학점, (역량준비) 공인영어성적	1.103	0.204	0.103	198	4.5	0.034
성별_여성	(역량준비) 공인영어성적	1.096	0.338	0.171	328	7.8	0.005
연령대_22-24	(역량준비) 공인영어성적	1.094	0.337	0.129	247	4.4	0.035
성별_여성	(역량준비) 컴퓨터 자격증	1.092	0.306	0.155	297	6.2	0.013
성별_여성	(역량준비) 학점	1.09	0.279	0.141	271	5.2	0.023

Table 3. Logistic Regression Results for Competency Preparation Activities (N=1,919): Odds Ratios (p-values)

구분	학점	공인영 어성적*	영어회 화능력	제2외 국어능 력	한자 능력	컴퓨터 자격증	업무직 무 관련 자격증*	고시 준비	해외 경험	직무 경험*	수상 경력	석박사 학위	봉사활동 동경험	동아리 활동	기타
const	0.45 (.228)	0.51 (.29)	0.18 (.022)	0.03 (.002)	0.00 (.0)	0.27 (.038)	0.40 (.13)	0.02 (.0)	0.01 (.004)	0.04 (.0)	0.21 (.352)	0.00 (.003)	0.31 (.317)	0.58 (.637)	0.01 (.242)
성별_여성	1.30 (.022)	1.27 (.027)	1.48 (.002)	1.61 (.016)	1.86 (.023)	1.31 (.014)	1.14 (.22)	1.01 (.964)	2.34 (.003)	1.46 (.01)	1.75 (.059)	1.66 (.182)	1.90 (.002)	1.59 (.024)	0.84 (.794)
연령	1.00 (.954)	1.00 (.96)	1.01 (.713)	1.05 (.301)	1.11 (.119)	1.02 (.394)	1.01 (.627)	1.11 (.0)	1.05 (.5)	1.08 (.036)	0.91 (.215)	1.12 (.293)	0.93 (.164)	0.91 (.066)	0.99 (.946)
최종학력_전문대졸	0.45 (.0)	0.36 (.0)	0.53 (.004)	0.28 (.004)	0.49 (.16)	0.66 (.018)	1.13 (.444)	0.77 (.162)	0.60 (.314)	1.11 (.623)	0.95 (.908)	0.59 (.634)	1.09 (.77)	0.81 (.515)	1.45 (.674)
최종학력_대졸	0.84 (.231)	1.49 (.003)	1.43 (.02)	1.32 (.24)	1.31 (.391)	0.94 (.666)	0.97 (.844)	1.77 (.0)	1.87 (.062)	0.84 (.345)	2.02 (.052)	6.17 (.001)	1.22 (.444)	1.16 (.577)	1.03 (.975)
최종학력_석사학위이상	3.08 (.059)	4.85 (.012)	2.55 (.128)	0.75 (.792)	1.33 (.794)	1.25 (.707)	1.67 (.377)	1.29 (.689)	4.27 (.101)	0.29 (.247)	0.00 (.942)	22958 (.0)	3.37 (.149)	1.67 (.638)	0.00 (.995)
거주지역_경상권	0.45 (.0)	0.55 (.0)	0.55 (.0)	0.38 (.0)	0.29 (.002)	0.46 (.0)	0.53 (.0)	1.79 (.0)	0.17 (.0)	0.22 (.0)	0.33 (.008)	0.36 (.042)	0.32 (.0)	0.48 (.006)	0.00 (1.0)
거주지역_충청권	1.28 (.092)	0.70 (.017)	0.73 (.071)	0.35 (.001)	0.48 (.078)	0.96 (.767)	1.09 (.56)	2.42 (.0)	0.18 (.005)	0.33 (.0)	0.72 (.38)	0.41 (.156)	0.72 (.227)	0.77 (.337)	0.00 (1.0)
거주지역_호남권	0.30 (.0)	0.48 (.0)	0.28 (.0)	0.11 (.002)	0.22 (.039)	0.52 (.002)	0.53 (.001)	2.79 (.0)	0.00 (.929)	0.18 (.0)	0.13 (.047)	0.10 (.038)	0.06 (.006)	0.07 (.008)	1.45 (.643)
거주지역_강원권	0.88 (.814)	0.11 (.031)	0.00 (.791)	0.00 (.946)	0.00 (.79)	0.99 (.989)	0.82 (.693)	4.24 (.002)	0.00 (.964)	0.00 (.916)	0.00 (.976)	0.00 (.933)	0.00 (.674)	0.76 (.788)	0.00 (.978)
거주지역_제주권	0.48 (.082)	0.57 (.13)	0.35 (.047)	0.00 (.95)	0.45 (.432)	0.58 (.165)	0.52 (.095)	2.69 (.004)	2.49 (.074)	1.29 (.497)	1.25 (.765)	0.00 (.91)	1.96 (.148)	1.69 (.293)	0.00 (.97)
Log-Likelihood	-1043	-1130	-902	-443	-277	-1111	-1184	-1043	-254	-704	-250	-144	-432	-434	-57
AIC	2107	2283	1826	908	576	2244	2390	2108	530	1430	522	309	886	890	135
BIC	2169	2344	1888	969	637	2306	2451	2169	592	1491	583	371	947	951	196

* 일부 항목은 해석을 위해 강조 표시됨. 이 변수들은 인구사회학적 특성과 유의미한 연관성을 보일 뿐만 아니라, 본문의 논의장에서 이론적·정책적 시사점과 함께 심층적으로 해석되는 주요 활동들임.

Table 4. Logistic Regression Results for Job Search Effort Activities (N=850): Odds Ratios (p-values)

구분	사설학 원수강	직업적 성/흥미 검사*	면접 훈련	공공기 관 구직 등록*	친인척/ 선배/친 구에게 취업부탁	선생님 (교수)에 게 취업 부탁*	일간지 및 정보지 구인란 구독	인터넷 구직정보 습득 및 등록*	취업전 문회사 자문	이력서 제출*	그룹 스터디	인턴 경험
const	0.04 (.006)	0.24 (.251)	0.25 (.309)	0.01 (.0)	0.44 (.318)	0.49 (.53)	0.18 (.069)	0.05 (.0)	0.03 (.025)	0.04 (.0)	0.05 (.035)	0.00 (.002)
성별_여성	1.15 (.489)	1.38 (.143)	1.44 (.133)	1.54 (.037)	1.12 (.454)	1.24 (.268)	1.07 (.696)	1.32 (.06)	1.40 (.233)	1.15 (.337)	1.08 (.755)	0.89 (.69)
연령	1.06 (.22)	0.97 (.607)	0.96 (.465)	1.11 (.046)	1.00 (.903)	0.94 (.197)	1.02 (.588)	1.10 (.004)	1.03 (.658)	1.13 (.001)	1.02 (.694)	1.11 (.142)
최종학력_전문대졸	1.36 (.293)	1.09 (.808)	0.67 (.35)	1.62 (.107)	1.16 (.483)	2.29 (.003)	0.72 (.196)	1.23 (.329)	1.32 (.489)	1.60 (.027)	0.74 (.455)	1.22 (.666)
최종학력_대졸	1.60 (.068)	2.12 (.008)	1.99 (.026)	1.90 (.015)	0.95 (.776)	2.75 (.0)	1.07 (.756)	1.25 (.241)	1.38 (.371)	1.41 (.064)	1.97 (.023)	1.84 (.1)
최종학력_석사학위이상	0.87 (.898)	3.58 (.162)	1.97 (.55)	1.21 (.862)	0.75 (.738)	13.65 (.001)	0.00 (.961)	0.19 (.127)	0.00 (.977)	2.16 (.371)	1.67 (.649)	2.49 (.433)
거주지역_경상권	0.48 (.004)	0.15 (.0)	0.50 (.034)	1.31 (.258)	1.15 (.416)	0.77 (.28)	1.13 (.515)	0.96 (.806)	1.05 (.877)	0.93 (.678)	0.84 (.533)	0.66 (.24)
거주지역_충청권	0.62 (.236)	1.46 (.257)	1.10 (.81)	2.18 (.02)	1.15 (.626)	2.23 (.009)	0.83 (.575)	1.43 (.201)	0.87 (.805)	1.91 (.029)	1.39 (.401)	0.63 (.464)
거주지역_호남권	0.09 (.017)	0.10 (.022)	0.30 (.106)	0.77 (.596)	1.15 (.658)	0.67 (.368)	2.06 (.024)	0.75 (.377)	0.26 (.194)	0.93 (.81)	0.74 (.586)	0.00 (.974)
거주지역_강원권	0.00 (.674)	0.00 (.846)	0.00 (.933)	12.17 (.002)	0.79 (.783)	5.57 (.031)	2.32 (.278)	1.88 (.418)	5.73 (.046)	0.82 (.802)	0.00 (.874)	0.00 (.999)
거주지역_제주권	0.79 (.67)	1.90 (.169)	1.33 (.612)	2.12 (.131)	1.68 (.202)	0.42 (.251)	1.40 (.445)	1.36 (.45)	0.51 (.512)	0.65 (.307)	0.32 (.273)	2.45 (.122)
Log-Likelihood	-349	-301	-268	-343	-551	-376	-472	-567	-214	-566	-276	-192
AIC	719	623	559	708	1124	773	965	1156	450	1154	574	406
BIC	772	675	611	760	1176	826	1017	1208	502	1207	626	459

* 일부 항목은 해석의 강조를 위해 표시함. 이 변수들은 인구사회학적 특성과 유의미한 연관성을 보일 뿐만 아니라, 본문의 논의장에서 이론적·정책적 시사점과 함께 심층적으로 해석되는 주요 활동들임.

경향을 보였다. 반면, 특정 취업노력 활동에서는 비수도권의 참여가 더 낮은 현상이 관찰되었다. 예를 들어, 경상권(OR=0.15, $p < .001$)과 호남권(OR=0.10, $p = .022$) 거주자는 수도권 거주자에 비해 '직업적성/흥미검사'를 받을 확률이 유의하게 낮았다. 이러한 결과는 지역별로 접근 가능한 노동시장의 기회구조가 상이하며, 이것이 청년들의 구직 전략 분화에 직접적인 영향을 미치고 있음을 시사한다.

2.4. The Effect of Age on Job Search Activities

연령 또한 구직 활동의 성격에 유의미한 영향을 미치는 것으로 나타났다. 연령이 증가할수록, 청년들은 탐색적 활동보다는 실제 채용 과정과 직접적으로 연관된 '취업노력' 활동에 더 집중하는 경향을 보였다. 구체적으로, 연령이 증가할수록 공공기관 구직등록(OR=1.11, $p = .046$), 인터넷 구직정보 습득(OR=1.10, $p = .004$), 이력서 제출(OR=1.13, $p = .001$)과 같은 활동에 참여할 확률이 유의하게 증가했다. 이는 구직 기간이 길어짐에 따라, 보다 구체적이고 실행 중심적인 활동으로 전략을 전환하는 과정을 반영하는 것으로 해석할 수 있다. 종합하면, 로지스틱 회귀분석 결과는 청년층의 구직 활동이 개인의 합리적 선택일 뿐만 아니라, 성별, 학력, 지역, 연령과 같은 사회구조적 위치에 깊이 연동되어 있음을 보여준다.

VI. Discussion and conclusion

본 연구는 청년층의 취업준비 활동이 개별 행위의 총합이 아니라, 상호 연관된 '전략적 묶음'으로 구조화되어 있음을 실증적으로 규명했다. 연관규칙분석을 통해 드러난 '기본적 구직 경로'와 '전략적 전문화 경로'는, 청년들이 노동시장에 자신의 역량을 증명하기 위해 각기 다른 신호 포트폴리오를 구축하고 있음을 보여준다. 특히 '학업 연계형 고스펙' 묶음의 발견은, 구직자들이 단순히 스펙을 나열하는 것을 넘어 특정 노동시장(e.g., 대기업, 공공부문)이 요구하는 신호를 집중적으로 축적하고 있음을 시사한다. 로지스틱 회귀분석 결과는 이러한 전략 선택이 결코 개인의 자유로운 선택만은 아니며, 성별, 학력, 지역과 같은 사회구조적 위치와 그에 따른 기회구조의 제약 속에서 이루어지는 복합적인 과정임을 명확히 했다. 예를 들어, 여성이 남성보다 더 광범위한 스펙 준비에 참여하는 경향은 노동시장에서 여성이 직면하는 추가적인 장벽을 극복하기 위한 합리적 대응 전략일 수 있다. 또한, 수도권과 비수도권 간의 뚜렷한 활동 패턴 차이는, 구직 전략이 개인의 의지

만큼이나 거주 지역의 교육 및 산업 인프라에 의해 크게 좌우된다는 점을 보여준다. 결론적으로, 본 연구의 결과는 청년 구직 과정을 전통적인 인적자본 축적의 관점을 넘어, 자신의 사회적 위치와 가용한 기회구조에 맞춰 최적의 신호 조합을 탐색하는 복잡한 사회적 과정으로 이해해야 할 필요성을 제기한다.

본 연구의 발견은 다음과 같은 학술적, 방법론적, 그리고 정책적 시사점을 가진다. 학술적 시사점으로 본 연구는 구직 행동 연구의 분석 단위를 개별 활동에서 활동의 조합으로 확장함으로써, 인적자본 및 신호이론의 실증적 적용 범위를 넓혔다. 특히 '구성적 관점'을 통해 구직 전략의 이질성을 밝힘으로써, 청년층 내부의 분화된 노동시장 이행 경로를 보다 정밀하게 이해할 수 있는 이론적 토대를 마련했다. 방법론적 시사점은 사회과학 연구에서 아직 활용이 제한적인 연관규칙분석을 구직 행동 데이터에 성공적으로 적용하여, 전통적인 회귀분석이 포착하기 어려운 변수 간의 비선형적·조합적 관계를 규명하는 탐색적 분석의 유용성을 입증했다.

정책적 시사점으로는 획일적 청년 고용정책의 한계를 명확히 확인한 바, 구직 전략 유형별로 세분화된 맞춤형 지원으로 전환되어야 한다. 예를 들어, '고스펙' 경로를 추구하는 구직자에게는 심화된 직무 연계형 프로그램을, 취업노력에 어려움을 겪는 비수도권 구직자에게는 온라인 기반의 직업적성/흥미검사 등에 대한 접근성 강화 프로그램을 제공하는 방식의 차별화된 접근이 요구된다.

본 연구의 한계로는 단면 자료를 활용하여 활동의 '효과성'까지 검증하지 못하였다는 점이다. 또한 구직자의 공급 측면에 집중하였기에, 특정 스펙 조합을 선호하는 기업들의 수요 측면 특성을 함께 고려하지는 못했다. 향후 패널 데이터를 활용한 종단 연구를 통해 각 전략 묶음이 실제 노동시장 성과로 어떻게 이어지는지를 분석하고, 머신러닝 기술[19] 연계를 확대하는 후속 연구가 필요하다.

REFERENCES

- [1] N. Lin, "Building a Network Theory of Social Capital," *Annual Review of Sociology*, Vol. 27, pp. ix-xxxiii, 2001. <https://doi.org/10.1146/annurev.soc.25.1.1>
- [2] D. H. Autor, "Skills, education, and the rise of earnings inequality among the 'other 99 percent'," *Science*, Vol. 344, No. 6186, pp. 843-851, 2014. <https://doi.org/10.1126/science.1251868>
- [3] M. Spence, "Job Market Signaling," *The Quarterly Journal of Economics*, Vol. 87, No. 3, pp. 355-374, 1973. <https://doi.org/10.1093/qje/87.3.355>

- 2307/1882010
- [4] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in Proceedings of the 1993 ACM SIGMOD international conference on Management of data, pp. 207-216, 1993. <https://doi.org/10.1145/170035.170072>
- [5] J. W. Kim, and H. K. Kang, "A visual exploration-based approach for extracting useful association rules," Journal of the Korea Computer Information Society, Vol. 18, No. 9, pp. 177-187, 2013. <https://doi.org/10.9708/jksci.2013.18.9.177>
- [6] J. Choi, and H. Chung. (2021). "K-means Clustering and Random Forest Analysis of Factors Influencing Group Classification by Job Preference." Korean Journal of Educational Research, 59(8), 113-144. <https://doi.org/10.30916/KERA.59.8.113>
- [7] W. R. Heinz, "Youth transitions in an age of uncertainty," in Youth and the new economy: A critical analysis, pp. 3-15, SensePublishers, 2009. https://doi.org/10.1163/9789087907297_002
- [8] K. I. Paul, and K. Moser, "Unemployment impairs mental health: Meta-analyses," Journal of Vocational Behavior, Vol. 74, No. 3, pp. 264-282, 2009. <https://doi.org/10.1016/j.jvb.2009.01.001>
- [9] Finch, W. H., French, B. F., & Hazelwood, A. (2023). A comparison of confirmatory factor analysis and network models for measurement invariance assessment when indicator residuals are correlated. Applied Psychological Measurement, 47(2), 106-122. <https://doi.org/10.1177/01466216231151700>
- [10] Park, S. H., Jang, S. Y., Kim, H., & Lee, S. W. (2014). An association rule mining-based framework for understanding lifestyle risk behaviors. PloS one, 9(2), e88859. <https://doi.org/10.1371/journal.pone.0088859>
- [11] Wibowo, W., Sari, N. P., Wilantari, R. N., & Abdul-Rahman, S. (2021, May). Association rule mining method for the identification of internet use. In Journal of Physics: Conference Series (Vol. 1874, No. 1, p. 012009). IOP Publishing. <https://doi.org/10.1088/1742-6596/1874/1/012009>
- [12] P. Fournier-Viger, J. C. W. Lin, B. Vo, T. T. Chi, J. Zhang, & H. B. Le, "A survey of itemset mining," WIREs Data Mining and Knowledge Discovery, Vol. 7, No. 4, 2017. e1207. <https://doi.org/10.1002/widm.1207>
- [13] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI magazine, Vol. 17, No. 3, pp. 37-54, 1996. <https://doi.org/10.1609/aimag.v17i3.1230>
- [14] M. Hahsler, B. Grün, and K. Hornik, "arules - A Computational Environment for Mining Association Rules and Frequent Item Sets," Journal of Statistical Software, Vol. 14, No. 15, pp. 1-25, 2005. <https://doi.org/10.18637/jss.v014.i15>
- [15] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, Applied logistic regression (3rd ed.), John Wiley & Sons, 2013. <https://doi.org/10.1002/9781118548387>
- [16] F. Pedregosa, et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, Vol. 12, pp. 2825-2830, 2011. <http://jmlr.org/papers/v12/pedregosa11a.html>
- [17] Korea Employment Information Service (KEIS), User's Guide for the 1st Wave of the Youth Panel 2021 (YP2021), KEIS, 2022. <https://survey.keis.or.kr/>
- [18] F. D. Blau, and L. M. Kahn, "The gender wage gap: Extent, trends, and explanations," Journal of Economic Literature, Vol. 55, No. 3, pp. 789-865, 2017. <https://doi.org/10.1257/jel.2016.0995>
- [19] Kim, J., Han-Joon, K., & Kim, H. (2019). Fraud detection for job placement using hierarchical clusters-based deep neural networks. Applied Intelligence, 49(8), 2842-2861. <https://doi.org/10.1007/s10489-019-01419-2>

Authors



Hyoungrae Kim earned his Ph.D. in Computer Science. He is currently an Associate Research Fellow at the Korea Employment Information Service (KEIS). His research focuses on enhancing

employment services through artificial intelligence, with specific interests in AI-powered job matching, artificial consciousness, and applied machine learning. He is a member of the Korea Society of Computer and Information.



Jongdeok Lim obtained a doctorate in industrial information system major at Seoul National University of Science and Technology, and is currently He is a researcher at the Digital Employment Service

Strategy Office of the Korea Employment Information Service. The main research fields are Generative AI, data analysis, technology management, information technology development, efficiency and economic evaluation, and are in charge of research and development in the field of digital employment services.



Jeongrae Kim received the M.S. degree in Defense Information Management from Korea National Defense University, Korea in 2013 and the Ph.D. degree in the School of Electrical and Computer Engineering from

University of Seoul, Korea in 2019. Dr. Kim is currently a Professor in the Department of Artificial Intelligence Software Convergence, Korea Polytechnics Chuncheon Campus. He is interested in fraud detection, time series data prediction, and data mining.