

## 머신러닝을 활용한 청년 구직자의 강소기업 선호 예측모형 개발 및 요인별 상대적 중요도 분석\*

조운주\*\* · 김진수\*\*\* · 배환석\*\*\*\* · 양성병\*\*\*\*\* · 윤상혁\*\*\*\*\*

### 〈목 차〉

I. 서론	IV. 연구결과
II. 이론적 배경	V. 결과 토의 및 시사점
2.1 강소기업	5.1 연구결과 토의 및 활용방안
2.2 구직선호도 관련 선행연구	5.2 연구의 시사점
2.3 분류모델 기반 머신러닝	5.3 연구의 한계점 및 향후 연구방향
III. 연구방법	참고문헌
3.1 데이터 수집	<Abstract>
3.2 연구설계 및 분석절차	

### I. 서론

경제활동과 노동시장에서 중요한 개념인 구인(employment)과 구직(job seeking)은 단순한 경제적 상호작용을 넘어, 서로 유기적으로 얽혀 있으면서 상호 영향을 주고받는 관계를 가지고 있다. 이 둘은 퍼즐 조각과 같이 서로 맞물려야만 완성되는 요소로, 기업의 성장과 노동자의 직업만족도, 그리고 국가경제의 안정에 큰 역할

을 담당하고 있다. 과거에는 이러한 구인과 구직이 이뤄지는 과정이 비교적 단순했으나, 이제는 기업과 노동자의 욕구와 기대치가 서로 조화롭게 일치해야만 성공적인 채용이 이뤄지는 시대가 되었다(류옥현, 이충석, 2010). 기업은 단순한 노동력 제공자로서의 역할을 넘어, 창의력과 유연성을 지닌 인재를 찾고 있는 반면, 구직자들은 단순히 급여만을 중시하는 것이 아니라, 자신의 역량을 펼칠 기회와 미래에 대한 보

\* 본 논문은 2023년 한국정보시스템학회 추계학술대회 우수 논문으로 선정되었음.  
\*\* 경희대학교 빅데이터응용학과 석사과정, jyj0510@khu.ac.kr(주저자)  
\*\*\* 경희대학교 빅데이터응용학과 석사과정, wlstm010203@khu.ac.kr  
\*\*\*\* 경희대학교 빅데이터응용학과 석사과정, srbae@khu.ac.kr  
\*\*\*\*\* 경희대학교 경영학과/빅데이터응용학과 교수, sbyang@khu.ac.kr  
\*\*\*\*\* 한국기술교육대 산업경영학부 조교수, yoonsh@koreatech.ac.kr(교신저자)

장을 요구하고 있기 때문이다.

이러한 상황은 우리나라 구인 및 구직 동향에도 뚜렷하게 나타나고 있는데, 특히 청년 실업 문제 악화와 함께 중소기업의 가중되는 구인난이 동시에 나타나고 있다는 점에서 문제의 심각성을 확인할 수 있다. 통계청의 자료에 따르면, 2022년 하반기 평균 고용률은 76.6%로 하락 추세를 보이는 반면, 2022년 3분기의 채용 계획 대비 미충원 인력 수는 185,000명으로, 통계조사가 시작된 2008년 이후 가장 높은 수치를 기록한 것으로 나타났다(김혜나, 2023; 통계청, 2023). 이 중 중소기업의 미충원 비율은 93.7%로, 대기업, 중견기업 및 공공기관의 미충원 비율에 비해 압도적으로 높은 것으로 조사되어, 국내 구인·구직 간 불균형 문제가 얼마나 심각한지를 보여주고 있다(고차원, 2023). 이러한 상황을 해결하기 위해 정부는 다양한 중소기업 지원 프로그램을 추진하고 있으며, 경쟁력을 갖춘 강소기업에는 특별한 지원을 더욱 강화하고 있다. 여기서 ‘강소기업’은 기업규모는 작지만 강한 경쟁력을 갖춘 중소기업을 의미하는데, 이들의 성장과 발전은 국가경제에 긍정적인 영향을 미치고, 중장기적으로 안정된 일자리를 보장하는 데 큰 기여를 할 수 있을 것으로 평가받고 있다(나중호, 2022). 즉, 대기업에 대해서는 구직난이 일어나고, 중소기업에서는 구인난이 일어나는 극단적인 구인·구직 간 불균형 상황 하에서, 이러한 강소기업이 문제해결의 대안으로 떠오르고 있는 것이다.

한편, 강소기업의 지속적인 성장과 발전을 위해서는 구직자들이 이러한 기업을 선택하는데 있어 어떤 요인을 중요하게 생각하는지를 파악하는 것이 필수적이다. 여러 선행연구에서

국내 구직자들의 직장선택 선호도가 성별, 나이, 지역 등 구직자의 다양한 특성에 따라 다르게 나타난다는 사실이 밝혀진 바 있다(이정록, 2019; 이충석, 류옥현, 2011). 최근에는 대학을 갓 졸업한 청년 구직자들의 기업 선택 선호도와 관련된 연구 또한 활발히 이뤄지고 있다(김영직, 조민호, 2015; 최재우, 강운선, 2015). 그러나, 대부분의 선행연구가 구인을 원하는 특정한 기업군(강소기업, 대기업, IT기업 등)에 초점을 맞추기보다는, 대학 졸업자, 청년 구직자, 일반 구직자 등 구직자의 입장에서만 이뤄진 한계로 인해, 실질적인 구인·구직 간 불균형 해소를 위한 통찰을 제공하기에는 미흡한 측면이 있다. 특히, 구인 측면에서 문제해결의 실마리를 가지고 있는 강소기업을 대상으로, 구직층 가운데 가장 큰 비중을 차지하고 있는 청년 구직자의 입장에서(공민경, 2023), 어떠한 요인이 선호도에 유의하게 작용하는지, 이들 선호요인 간 상대적 중요도는 어떻게 되는지를 살펴보는 것은 정부의 강소기업 지원정책의 효과적인 방향성 제시를 위해서라도 큰 의미가 있음에도, 이러한 시도는 아직 이뤄지고 있지 못한 상황이다.

이에, 본 연구에서는 강소기업 관련 구인 및 구직 정보를 통해 (1) 강소기업에 대한 청년 구직자의 선호에 영향을 미치는 요인들을 도출하고, (2) 머신러닝(machine learning) 기법을 활용하여 예측모형을 개발한 후, (3) 이들 선호요인 간 상대적 중요도를 분석하고, (4) 그 결과를 바탕으로 정부의 실제 정책수립에 활용될 수 있는 유의미한 방안을 제안해 보고자 한다. 이를 위해 고용노동부의 ‘2023 강소기업 선정 현황’ 공공데이터 및 청년위크넷 홈페이지(2023)

의 ‘청년친화 강소기업’ 정보를 크롤링(crawling)하여 데이터를 수집한 후, 다양한 분류모델 기반 머신러닝(Logistic Regression, Random Forest Classifier, Gradient Boosting Classifier, LGBM Classifier, XGB Classifier) 기법을 활용하여 청년 구직자의 강소기업 선호 예측모형을 개발하고자 한다. 본 연구는 국내 구인·구직 간 불균형 문제해결을 위해 청년 구직자들의 강소기업 선호 예측모형을 개발하고, 선호요인 간 상대적 중요도를 분석한 최초의 연구라는 점에서 학술적 의의를 지닌다. 본 연구를 통해 청년 구직자의 강소기업 선호요인을 파악함으로써, 정부, 기업 및 구직자 모두에게 유의미한 대응방안 마련을 위한 통찰을 제공할 수 있을 것으로 기대한다.

## II. 이론적 배경

### 2.1 강소기업

최근 중소기업의 구인난과 청년 취업층의 구직난이 동시에 발생하며, 취업시장에서의 구인·구직 간 불균형이 심화하고 있다. 이러한 상황의 원인 중 하나로 청년 구직자들의 기업 선택기준 변화가 지적되고 있다. 대한상공회의소의 조사에 따르면, 낮은 고용안정성에 대한 우려와 불투명한 미래 성장의 이유로 많은 청년 구직자가 중소기업을 기피하고 있는 것으로 나타났다(대한상공회의소, 2023). 한편, 이러한 문제해결의 대안으로 떠오르고 있는 강소기업의 경우, 대기업만큼의 복지와 기술경쟁력을 바탕으로 높은 고용안정성과 낮은 산업사고율을

지닌 기업이 대부분임에도 불구하고, 다수의 청년 구직자가 이를 인지하지 못해 올바른 기업 선택의 기회를 놓침으로써 구인·구직 간 불균형을 심화시키는 상황을 초래하고 있다. 독일의 연구자 헤르만 지몬(Hermann Simon)에 의해 처음 소개된 바 있는 개념인 강소기업은 (1) 특정 분야(산업)에서 대중에게 잘 알려지지 않은, (2) 작은 규모의 기업임에도 불구하고, (3) 시장경쟁력과 발전가능성을 인정받고 있는 기업을 의미한다(이병환, 허문구, 2014).

강소기업에 대한 올바른 인식 확산 및 구직자들의 선호도 제고가 절실한 국정과제로 떠오르며, 우리나라 정부도 ‘글로벌 강소기업 육성 정책’, ‘월드클래스300 프로젝트’, ‘글로벌 전문기업 육성’ 등의 다양한 지원 정책을 시행하고 있다(길재욱 등, 2017). 하지만, 주무부처가 다른 기관에서 개별적으로 시행되는 지원 정책들로 인해 그 효과와 성공요인에 대한 심도있는 분석과 검토가 필요하다는 의견 또한 지속적으로 제기되고 있다(김수욱, 2010). 즉, 강소기업 관련 정부의 지원 정책들이 한국 경제의 성장과 고용창출에 크게 기여하고 있다는 사실은 분명하지만, 앞서 언급한 심화된 구인·구직 간 불균형 문제해결을 위해서는 좀 더 종합적이고 적극적인 정책 마련이 요구되고 있는 것이다.

한편, 국내 강소기업 관련 선행연구들은 대부분 분야별 글로벌 및 산업분류에 초점을 맞춰 진행된 바 있다(김수욱, 2010; 이영주, 윤동진, 2015; 장호준, 고영희, 2014; 정인교 등, 2014). 이러한 연구들은 주로 사례연구와 실증 분석 등을 활용하여 강소기업 육성 및 발전 관련 다양한 시사점을 도출했다는 점에서 의의를

찾을 수 있으나, 구인·구직 간 불균형 문제해결을 위한 목적으로 강소기업 대상 청년 구직층의 선호를 연구한 시도는 부족한 실정이다. 이에, 본 연구에서는 청년워크넷에서 제안한 선정 기준을 참고하여, 임금, 일과 생활의 균형, 고용 안정성, 청년고용 실적 등이 우수한 중소기업으로 강소기업을 정의하고, 이에 대한 청년 구직층의 선호 예측모형을 개발해 보고자 한다. 강소기업 관련 구체적인 선정기준은 <부록>에 제시하였다.

## 2.2 구직선호도 관련 선행연구

최근 극도의 취업난에 허덕이는 청년층이 구직활동이나 취업준비 없이 일상을 보내는 경우가 점차 증가하고 있다(공민경, 2023; 통계청, 2023). 이러한 현상은 국가의 경제발전에 부정

적인 영향을 주기 때문에, 저조한 고용시장의 상황을 개선하는 것이 중요한 국정과제로 떠오르고 있다. 이를 위해, 고용시장의 핵심층인 청년 구직자들이 취업을 희망하는 기업에 대해 어떠한 요인을 좀 더 중요하게 인식하고 있는지를 살펴보는 일은 매우 중요하다. 이에, 선행 연구에서는 이러한 청년층을 포함한 구직자 입장에서의 기업선택에 대한 선호요인을 분석하려는 시도가 다수 진행된 바 있다(<표 1> 참고). 구직선호도 관련 대표적인 선행연구를 살펴보면, 우선, 김영직과 조민효(2015)는 청년 고용정책의 효과성을 높이기 위해 로지스틱 회귀분석을 활용해 대학 졸업 청년층의 가치와 구직선호 요인을 분석하고, 이를 통해 대학 졸업자의 가치선호도를 기반으로 한 구직선호 유형을 도출하였으며, 최재우와 강운선(2015)은 4년제 대학 졸업자들의 구직선호도를 분석한 결과, 성

<표 1> 구직선호도 관련 선행연구

구분	저자(연도)	연구내용	연구방법	
연구대상 및 범위	대학 졸업자	김영직, 조민효(2015)	청년 고용정책 수립을 위해 대학 졸업을 한 청년들을 기준으로 가치와 선호도를 파악하고자 함	로지스틱 회귀분석
		최재우, 강운선(2015)	4년제 대학 졸업자들을 중심으로 구직선호도의 잠재적 유형 도출	설문조사
	청년 구직자	김영생(2011)	청년층을 구직자를 중심으로 구인 구직의 불일치 상태와 문제점을 분석	설문조사
		김형석, 이정환(2021)	청년 구직자를 대상으로 중소기업 일자리에 대한 인지적, 심리적 인식 확인	설문조사
		이정환, 김동욱(2022)	브랜드의 인식이 청년 구직자의 취업 의향에 미치는 영향	회귀분석
	일반 구직자	이정록(2019)	잠재적 구직자를 대상으로 지역 일자리에 대해 구직자의 선호도 분석	설문조사
		이충석, 류옥현(2011)	성별, 나이별, 지역별 등 다양한 요인으로 구직자 유형을 나누고 직장선택 요인에 관한 탐색적 연구	설문조사
		허성호(2020)	대기업, 중소기업 그리고 구직자의 특성을 파악해 채용 경향성 비교	사회연결망 분석

별과 졸업한 대학의 위치가 구직선호도에 큰 영향을 미친다는 사실을 확인한 바 있다. 한편, 대학 졸업자를 포함한 청년 구직자를 대상으로 한 연구로, 김영생(2011)은 중소기업과 청년층 간 구인·구직 불일치 문제를, 김형석과 이정환(2021)은 청년 구직자들의 중소기업에 대한 인식을 각각 설문문을 통해 조사한 바 있다. 나아가, 이정환과 김동욱(2022)은 브랜드 인식이 청년 구직자의 취업 의향에 어떠한 영향을 미치는지를 회귀분석을 통해 실증하기도 하였다. 마지막으로, 나이대를 구분하지 않은 채 일반 구직자를 대상으로 진행된 연구도 다수 존재하는데, 이정록(2019)은 잠재적 구직자를 대상으로 지역 일자리 선호도 분석을 진행하였으며, 이충석과 류옥현(2011)은 설문조사를 이용하여 일반 구직자의 성별, 나이, 지역별 특성을 파악한 후, 이를 기반으로 직장선택 요인을 도출하는 탐색적 연구를 수행한 바 있다. 또한, 허성호(2020)는 기업의 특성과 구직자의 특성 간 채용 경향에 대한 차이 분석을 시도하였다.

이상의 구직선호도 관련 선행연구를 종합하면, 대부분의 선행연구가 대학 졸업자, 청년 구직자, 일반 구직자 등 구직자의 입장에서 중소기업에만 초점을 맞춰 진행되었음을 알 수 있다. 이러한 한계로 인해, 실질적인 구인·구직 간 불균형 해소를 위한 통찰을 제공하기에는 다소 부족한 측면이 있다. 이에, 본 연구에서는 문제 해결의 실마리를 가지고 있는 강소기업을 대상으로, 청년 구직자가 취업을 고려할 때 어떠한 요인을 더욱 중점적으로 살펴보는지, 이들 선호 요인 간 상대적 중요도는 어떻게 되는지를 살펴보고자 한다. 또한, 대부분의 선행연구가 설문조사 및 회귀분석을 바탕으로 이루어진 것을

고려할 때, 본 연구에서는 분류모델 기반 머신러닝 기법을 활용하여 강소기업에 대한 구직자들의 선호 예측모형을 개발하고, 나아가 이들 선호요인 간 상대적 중요도를 도출해 보고자 한다.

### 2.3 분류모델 기반 머신러닝

분류모델은 복잡한 데이터 패턴을 학습하는 핵심 데이터 분석방법론 중 하나로, 새로운 데이터의 범주나 클래스를 예측할 때 활용된다(Novaković et al., 2017). 이 방법론은 수학적 도구와 기법을 활용해 데이터에 잠재된 중요한 정보와 관계를 찾아내는데 효과적인 것으로 알려져 있는데, 최근 다양한 머신러닝 알고리즘이 이 분야에 활용되기 시작하면서 분류의 정확성과 효과성이 크게 증가하고 있다(Sarker, 2021). 이에, 본 연구는 청년 구직자가 강소기업을 선택하는 데 영향을 주는 핵심 선호요인들을 밝혀내고, 이들 간의 상대적 중요도를 도출하기 위해 분류모델 기반 머신러닝을 활용하고자 한다. 대표적인 분류모델 기반 머신러닝 기법으로는 (1) Logistic Regression, (2) Random Forest Classifier, (3) Gradient Boosting Classifier, (4) LGBM Classifier, (5) XGB Classifier 등을 꼽을 수 있는데, 여기서, Logistic Regression은 이진 또는 다중 클래스 분류에 적합한 통계적 방법론으로 널리 사용되고 있으며(Hosmer and Lemeshow, 2013), Random Forest Classifier는 여러 의사결정 트리를 학습하여 종합적인 결과를 도출하는 앙상블(ensemble) 기법으로(이동훈, 김태형, 2020), 개별적 의사결정 트리 모델보다 우수하며 보다 일반적인 결과를 제공해

주는 것으로 알려져 있다(Breiman and Cutler, 2001). 또한, Gradient Boosting Classifier는 약한 학습기들을 순차적으로 학습시켜 오차를 감소시키는 부스팅 방식을 채택하고 있으며(Friedman, 2001), LGBM Classifier는 Microsoft에서 개발한 Gradient Boosting 알고리즘으로 대용량 데이터 세트(large data set) 처리에 특히 효율적인 것으로 알려져 있다(Ke et al., 2017). 마지막으로, XGB Classifier는 GBM(gradient boosting model)을 최적화하여 빠르고 정확한 성능을 제공하는 알고리즘으로 알려져 있다(Chen and Guestrin, 2016).

### III. 연구방법

#### 3.1 데이터 수집

본 연구에서 데이터는 고용노동부의 ‘2023 강소기업 선정 현황’ 공공데이터와 청년위크넷 홈페이지에서 얻은 ‘청년친화 강소기업’ 정보를 통합해 활용하였다. 고용노동부의 데이터는 2023년 기준 최신 정보를 담고 있으며, 강소기

업으로 선정된 기업의 사업장명, 업종, 상세주소 등의 정보를 포함하고 있다. 한편, 청년위크넷의 데이터는 파이썬(Python)의 Selenium 라이브러리를 활용한 동적 웹크롤링(dynamic web crawling) 기법을 활용하여 수집되었으며, 해당 데이터에는 기업명, 선정분야, 구직자들의 관심을 나타내는 관심기업 건수 등과 같은 변수가 포함되어 있다(<표 2> 참조).

동적 웹크롤링 결과, 고용노동부에서 27,737개, 청년위크넷에서 27,780개의 강소기업 관련 데이터를 수집하였으며, 두 기관으로부터 수집한 데이터를 비교하여 중복된 기업 정보를 제거한 후, 최종적으로 27,737개의 강소기업 데이터를 최종 분석에 활용하였다. 고용노동부와 청년위크넷 데이터의 차이점은 포함된 변수들에서 비롯된다. 청년위크넷이 제공하는 27,780개의 데이터는 지역을 기준으로 하여, 다른 지역에 위치하면 데이터를 나누었고, 이에 따라 중복된 기업이 존재한다. 반면, 고용노동부의 27,737개 데이터는 기업을 중심으로 정리되어, 하나의 기업이 여러 지역에 위치하더라도 하나의 기업으로만 표기하였으므로 기업의 총개수에 차이가 생긴다. 청년위크넷의 데이터를 추가

<표 2> 수집 데이터의 변수 정의

변수명	상세정보
기업명	기업(강소기업)의 이름
관심기업	기업이 받은 총 관심기업 득표수
직원 수	기업의 총 직원 수
기업규모	기업의 규모(대기업, 중견기업, 중소기업 등)
소재지(대)	기업의 위치(광역시)
소재지(중)	기업의 위치(시, 군, 구)
브랜드명	기업이 인증받은 강소기업유형
대표자명	기업 대표자의 이름
업종(중분류)	기업의 주요 업종 분류

<표 3> 기술통계표

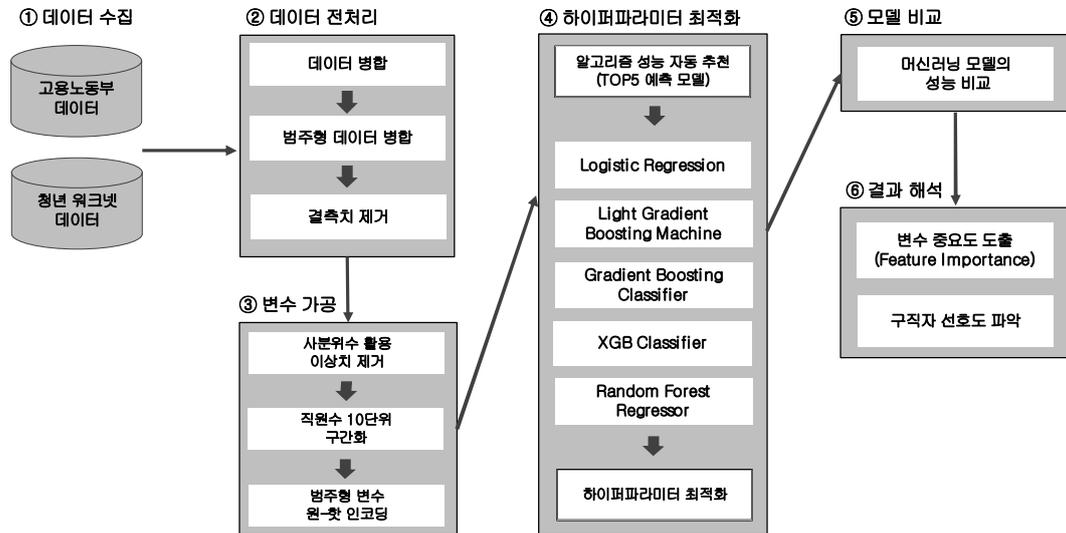
변수명	최솟값	최댓값	평균	표준편차	분산
직원 수	0	5,007	50.14	100.1	10,026.2
관심 기업	0	966	11.35	22.5	506.356

한 이유는 고용노동부 데이터에 없는, 구직자들이 관심을 가질 수 있는 '관심 기업' 변수를 활용하기 위함이다. 여기서, 기업명은 강소기업 선정 시 부여받은 고유한 이름을 사용하였으며, 선정분야는 이노비즈와 메인비즈 및 기타로 구성되어 있다. 업종은 중분류 기준으로 기업이 속한 산업에 따라 분류하였고, 기업규모는 대기업(분사), 중견기업(분사), 중소기업 등으로 구분하였다. 소재지 변수의 경우, 강소기업의 위치를 광역과 시, 군, 구 등의 단위에 따라 각각 개별적인 변수(소재지\_대, 소재지\_중)로 구분하였으며, 직원수의 경우, 총 10개의 구간으로 구분하였다. 마지막으로, 본 연구에서 결과변수로 활용된 관심기업\_투표수 변수의 경우, 청년워크넷에서 구직자들의 관심을 반영하여 기록된 총 투표의 수를 나타낸다. 그 결과, 청년워크넷에서만 제공하는 관심 기업에 대한 투표수는 평균 11회, 최대 966회에 달하며, 기업의 직원수는 평균 50.14명, 최대로는 5,007명의 직원을 둔 기업이 존재함을 알 수 있다(<표 3> 참조).

### 3.2 연구설계 및 분석절차

본 연구의 분석절차는 강소기업 관련 (1) 데이터 수집, (2) 데이터 전처리, (3) 변수 가공, (4) 데이터 분석(하이퍼파라미터 최적화), (5) 모델 비교, (6) 결과해석의 순서로 진행하였으며, 이를 도식화하여 <그림 1>에 나타내었다.

우선, 데이터 수집 단계에서는 고용노동부 공공 데이터와 청년워크넷 청년친화 강소기업 관련 데이터를 병합하여 사용하였다. 사용한 변수들은 기업명, 관심기업\_투표수, 직원수, 기업규모, 소재지\_대, 소재지\_중, 선정분야, 대표자명, 업종(중분류) 등이 있다. 다음으로, 데이터 전처리 단계에서는 먼저, 업종 이름 중 제조업이 포함된 값들을 복잡성을 줄이기 위해 모두 제조업으로 통일시켰다. 이후, 다른 표현으로 작성된 동일 선정분야를 모두 일치시키는 작업을 진행하였으며, 병합으로 인해 생성된 결측치 304건을 모두 삭제하였다. 또한, 연번, 주소 등과 같이 본 연구의 목적에 부합하지 않는 변수들은 모두 제거하였다. 이후, 데이터 분석의 결과를 왜곡시키는 이상치(outlier)를 제거하기 위해 각 수치형 변수들의 사분위 수를 분석하였고, 사분위 수의 범위가 IQR(Inter Quartile Range)\*1.5 값을 넘어선 경우, 이상치로 판단하고 이를 모두 제거하였다(Rousseeuw and Hubert, 2011). 다음으로, 직원 수가 기업별로 어떻게 분포하는지를 파악하고 이들 간의 패턴을 이해하기 위해 구간화(binning)를 진행, 총 10개로 구성된 구간별 '직원수\_해당구간' 변수를 추가하였다(Freedman and Diaconis, 1981). 또한, 업종(중분류) 변수에 제조업과 서비스업이 대다수를 차지하고 있으므로, 이 두 업종을 구분하는 변수를 새롭게 추가하였다. 마지막으로, 소재지\_대, 소재지\_중, 기업규모 등과 같은



<그림 1> 분석절차

범주형 변수들들의 경우, 분석의 용이성을 위해 원-핫 인코딩(one-hot encoding)으로 변환하였다(Harris et al., 2020).

결과변수인 관심기업\_투표수의 경우, 중위수 4를 기준으로 관심도 높음(1), 관심도 낮음(0)을 구분함으로써, 대표적인 분류모델 기반 머신러닝 적용이 가능하도록 하였다. 머신러닝 모델로는 Logistic Regression, Random Forest Classifier, Gradient Boosting Classifier, LGBM Classifier, XGB Classifier 등 총 다섯 가지 분류모델을 활용하였으며 이는 AUTO ML을 통해 도출한 모델이다. 이후, 적용한 머신러닝 모델들의 하이퍼파라미터 튜닝(hyperparameter tuning)을 위해 Optuna 최적화 알고리즘을 적용하였는데, Optuna 최적화 알고리즘의 경우, 기존에 TPE (tree-structured Parzen estimator)를 활용한 효율적 하이퍼파라미터 탐색으로 불필요한 시행을 조기에 중단하여 계산시간을 줄일 수 있는 장점이 있다(Akiba

et al., 2019). 이후, 다섯 가지 머신러닝 모델의 성능을 비교하여 최적의 모델을 선정하고, 선정된 모델을 활용하여 선호요인 변수 간 상대적 중요도를 도출하고, Logistic Regressor의 회귀 계수를 추가로 활용하여 그 의미를 해석하였다(Lundberg and Lee, 2017).

#### IV. 연구결과

본 연구에서는 분류모델 기반 머신러닝 기법을 활용한 예측모형 성능을 평가하기 위해, 대표적 분류성능 평가지표들인 Precision(정밀도), Specificity(특이도), Sensitivity(민감도) 및 F1-Score를 사용하였다(Lee et al., 2022). 각 분류모델의 성능평가 결과는 <표 4>에 정리하였으며, 각 모델은 동일한 하이퍼파라미터 최적화 방식에 기반하여, 성능을 비교하였다.

<표 4> 분류모델 기반 머신러닝 예측모형 성능비교

Model	Precision	Specificity	Sensitivity (Recall)	F1-Score
Logistic Regression	0.653	0.695	0.684	0.668
Random Forest Classifier	0.664	0.715	0.671	0.667
Gradient Boosting Classifier	0.671	0.727	0.662	0.666
LGBM Classifier	0.674	0.730	0.665	0.670
XGB Classifier	0.667	0.715	0.678	0.672

<표 5> 선호요인 변수의 상대적 중요도

Feature	Importance (LGBM Classifier)	Coefficient (Logistic Regression)
업종(중분류)_제조업	849	0.776
업종(중분류)_서비스업	586	0.251
소재지_대_경기	551	-0.535
소재지_대_서울	515	-1.476
직원수_0-10	438	-1.755

다섯 가지 머신러닝 활용 선호요인 예측모형의 성능을 비교·평가한 결과, 정밀도와 특이도 면에서 LGBM Classifier가 각각 0.674와 0.730으로 다른 모델들에 비해 가장 뛰어난 성능을 보인 것으로 나타났다. 반면, 민감도 면에서는 Logistic Regression과 XGB Classifier가 각각 0.684, 0.678로 높은 성능을 보인 것으로 나타났다. 마지막으로, F1-Score의 경우, XGB Classifier와 LGBM Classifier가 각각 0.672 및 0.670으로 높은 성능을 보였다. 전반적으로, LGBM Classifier는 정밀도와 특이도에서 최고의 성능을 보이면서 다른 성능 지표들에서도 균형잡힌 결과를 보인 것으로 나타나, 본 연구에서는 LGBM Classifier를 최종 선호요인 예측모형으로 선정하였다.

본 연구는 청년 구직자의 관심도를 높이는 강소기업의 주요 선호요인을 파악하기 위한 목

적으로 수행되었다. 이를 위해, 높은 성능의 예측모형(LGBM Classifier)을 이루는 선호요인 변수 가운데, 어떤 변수가 성능을 결정하는데 큰 영향을 미치는지를 알아보기 위해 변수의 상대적 중요도를 추가적으로 계산하였다. 변수의 상대적 중요도는 머신러닝 분야에서 널리 활용되는 도구로, 이를 통해 각 변수가 결과변수에 미치는 영향력을 수치적으로 파악할 수 있는 장점이 있다(Lee et al., 2022). 그러나, 변수의 상대적 중요도만으로는 해당 변수가 결과변수에 어떤 방향(정(+))의 방향 혹은 부(-)의 방향)으로 영향을 미치는지 정확히 알 수 없으므로, 본 연구에서는 Logistic Regressor의 회귀계수를 추가로 활용하여 각 변수의 영향 방향을 파악하였다(정동균 등, 2021).

상대적 중요도 분석결과, 업종(중분류)\_제조업, 업종(중분류)\_서비스업, 소재지\_대\_경기,

<표 6> 제조업종 강소기업의 지역 비율

지역	비율
서울	4.61%
경기	36.98%
비수도권	58.41%

소재지\_대\_서울, 직원수\_0-10 등의 순으로 선호요인 변수의 상대적 중요도가 높은 것으로 나타났다(<표 5> 참조). 우선, 가장 높은 중요도를 보인 두 변수는 업종(중분류)\_제조업과 업종(중분류)\_서비스업으로, Logistic Regression의 계수(coefficient) 값이 각각 0.776, 0.251으로 정(+)의 방향을 나타내고 있으므로, 청년 구직자들이 강소기업 중에서도 제조업종과 서비스업종에 속한 기업을 특별히 더 선호한다는 것을 알 수 있다. 다음으로, 소재지\_대\_경기와 소재지\_대\_서울 변수가 상대적으로 높은 중요도를 가진 것으로 나타났는데, Logistic Regression의 계수가 각각 -0.535와 -1.476으로 부(-)의 방향을 나타내고 있음을 알 수 있다. 이러한 결과는 청년 구직자들이 경기와 서울지역에 위치한 강소기업에 대한 선호도가 상대적으로 낮음을 의미하는데, 이는 일반적인 예상과 배치되는 결과로, 향후 추가적인 연구를 통해 밝혀내어야 할 부분으로 판단된다. 다만, 이러한 결과가 도출된 원인으로 다음의 두 가지를 꼽을 수 있는데, 첫째, 서울과 경기지역은 다른 지역에 비해 대기업 및 중견기업의 비중이 상대적으로 높은 지역으로, 청년 구직층의 관심이 강소기업보다는 대기업과 중견기업에 더 몰렸을 가능성이 있다. 둘째, <표 6>의 제조업종 강소기업의 지역별 비중을 살펴보면, 서울(4.61%) 및 경기(36.98%) 지역을 제외한 비수

도권 지역에 58.41%가 위치해 있음을 알 수 있는데, 청년 구직층의 제조업종 기업에 대한 높은 선호(상대적 중요도 1위)가 반영된 결과로 해석할 수 있다.

마지막으로, 직원수\_0-10 변수는 강소기업의 직원 수, 즉, 기업의 규모가 청년 구직자의 관심도와 밀접한 연관성을 가질 수 있음을 나타낸다. Logistic Regression의 계수가 -1.755로 부(-)의 방향을 나타내고 있음을 볼 때, 청년 구직자들은 강소기업의 직원 수가 10명 이하일 경우 그 선호도가 급격히 낮아질 수 있음을 알 수 있다.

## V. 결과 토의 및 시사점

### 5.1 연구결과 토의 및 활용방안

본 연구를 통해 선정된 예측모형 및 분석결과는 정부, 강소기업, 청년 구직자 입장에서 다양한 방면으로 활용될 수 있다. 특히, 청년 구직자의 관심도나 선호도를 반영한 정책 수립이나 기업활동 계획 수립에 큰 도움을 줄 것으로 기대한다. <표 7>은 도출된 변수의 상대적 중요도 분석결과를 바탕으로 제안된 실무적 활용방안이다.

첫째, 정부는 청년 구직자의 선호도를 활용

<표 7> 정부, 기업, 구직자 차원의 실무적 활용방안 제안

구분	분석결과 활용방안
정부	구직자 선호도를 기반으로 강소기업 선정 정책 기준에 활용 가능
	취업지원 교육프로그램 선정 시 새로운 업종 도입 가능
	선호도를 활용한 산업규모 및 성장도에 따른 투자전략 수립 가능
강소기업	추후 강소기업의 채용 및 구인활동의 참고자료로 활용 가능
	기존 강소기업들의 사업운영 지표로 활용 가능
	업종(제조업, 서비스업) 특성을 기반으로 예비 창업인들의 창업계획 수립에 활용 가능
청년 구직자	기업 선택 시 참고지표로 활용 가능
	구직자 개인에게 적합한 기업을 추천받을 수 있음
	강소기업에 대한 새로운 인식 제고

하여 강소기업 관련 정책을 개선할 수 있다. 본 연구결과를 바탕으로, 기관별로 다르게 적용되던 강소기업 선정기준을 통일하는 것이 가능해질 것이며, 이 기준은 강소기업 선정에 직접 활용될 수 있다. 더불어, 취업지원 교육프로그램을 선정하는 과정에서도 새로운 업종의 도입이 고려될 수 있다. 예컨대, 국민 내일배움카드나 국비지원 프로그램에서는 이런 연구결과를 통해 구직희망자들의 취업영역을 확대할 수 있을 것으로 기대한다. 또한, 구직자의 선호도를 기반으로 특정 중소기업에 투자하거나 지원하는 전략 수립도 가능할 것으로 기대한다.

둘째, 강소기업은 구직자의 선호도를 채용 및 구인 시의 참고자료로 활용할 수 있다. 특히, 강소기업은 자원과 정보가 한정적이기 때문에 (윤병운, 이성주, 2010), 구직자 선호도를 이해하면 그 한계 내에서도 효과적인 복지제공 전략을 구축할 수 있다. 현재 강소기업에 해당하는 기업들은 구직자의 선호도를 참고해 기업의 구조나 방향성을 재정립할 수 있을 것이다. 예를 들면, 강소기업의 직원규모나 사업장 위치 등의 변수를 재검토하여 기업의 경쟁력을 높일 수 있을 것으로 기대한다. 더불어, 예비창업자

의 경우에도 본 연구를 통해 구직자들의 특정 선호도를 이해하고 이를 바탕으로 신뢰를 얻을 수 있을 것으로 기대한다. 연구결과에 따르면, 서울과 경기의 강소기업, 10명 이하의 규모, 특정 업종 등에 대한 선호도가 상대적으로 낮게 나타났다. 이러한 정보는 예비창업자가 기업위치, 초기 직원규모, 사업영역 등을 결정할 때 훌륭한 지침이 될 수 있을 것으로 기대한다.

마지막으로, 청년 구직자들은 연구결과를 참고지표로 활용하여 원하는 직장을 선택하는데 도움을 받을 수 있다. 특히, 직장경험이 부족하거나 취업준비가 미숙한 청년 구직자들은 다른 구직자들의 선호도를 참고하여 본인에게 더욱 적합한 직장을 선택하는 데 훌륭한 참고자료로 활용할 수 있을 것이다. 예를 들어, 강소기업에 대한 불안정한 인식이 있음에도 불구하고, 경영성과, 직원 수, 업종 및 소재지 등과 같은 실제 지표를 바탕으로 안정적인 직장을 찾아내는 것도 가능할 것으로 기대한다.

## 5.2 연구의 시사점

본 연구는 2023년 고용노동부에서 발표한 강

소기업 선정현황 결과와 청년워크넷의 청년친화 관심기업 데이터를 바탕으로 청년 구직자의 강소기업 선호요인 예측모형을 개발하고, 선호요인 간 상대적 중요도를 분석하였으며, 그 결과를 바탕으로 활용방안을 제안하였다. 이로부터 다음과 같은 이론적 시사점을 도출할 수 있다. 첫째, 본 연구는 강소기업에 대한 전통적인 분석을 넘어 새로운 차원의 접근법을 도입하였다. 기존의 강소기업 관련 연구는 주로 강소기업 성공사례를 소개하거나 분야별 산업분류에 초점을 맞춰 진행되어 왔다(김수욱, 2010; 이영주, 윤동진, 2015; 장호준, 고영희, 2014; 정인교 등, 2014). 그러나, 이러한 선행연구들은 최근 대두되고 있는 구인·구직 간 불균형 문제를 해결하기에는 한계가 존재한다. 이에, 본 연구에서는 청년 구직자 입장에서 강소기업에 대한 선호요인을 도출하고, 이들 간 상대적 중요도를 실증함으로써, 강소기업 관련 연구에 새로운 패러다임을 제시하였다는 점에서 학술적 의의가 있다.

둘째, 본 연구는 분석도구로써 머신러닝을 적극 활용하였다. 즉, 기존 선행연구에서 주로 활용한 방법론인 사례연구 및 실증연구에 그치지 않고, 최근 경영, 관광 및 환대산업, 레저 및 스포츠 산업 등 많은 영역에서 활발히 적용되고 있는 머신러닝을 기법을 활용하여 청년 구직자의 강소기업 선호요인 예측모형을 개발하고, 선호요인 간 상대적 중요도를 분석함으로써, 선호요인과 강소기업 관심도 간 비선형적인 영향관계를 실증할 수 있었다. 이러한 시도는, 향후 강소기업 및 구직선호도 관련 연구를 방법론적인 측면에서 확장할 수 있는 계기를 마련함으로써, 향후 머신러닝, 나아가 딥러닝

(deep learning) 방법론 활용을 위한 디딤돌을 마련했다는 점에서 학술적 의의가 있다.

마지막으로, 본 연구는 구인난을 겪고 있는 강소기업의 입장, 구직난을 겪고 있는 청년 구직자의 입장, 나아가 구인·구직 간 불균형 문제 해결을 통한 안정된 일자리 보장 및 국가경제 발전을 도모하고자 하는 정부의 입장을 모두 반영하여 분석결과 활용방안을 제안하고 있다는 점에서 학술적 의의를 찾을 수 있다. 기존 구직선호도 관련 선행연구들이 중소기업과 같이 구인자의 입장 혹은 청년 구직자를 중심으로 한 구직자의 입장 등 어느 한쪽의 입장에서 살펴본 연구가 대부분인 반면, 본 연구에서는 기업(강소기업)의 입장과 구직자(청년 취업층)의 입장을 일치시켜 이를 동시에 살펴봄으로써, 한쪽에 치우치지 않은 균형잡힌 결과를 도출하고, 이를 통해 실질적인 청년층 구인·구직 불균형 문제에 대한 해결책을 제시했다는 점에서 의의를 찾을 수 있다.

본 연구는 강소기업의 채용 전략, 청년 구직자의 강소기업 선호요인 도출, 그리고 정부의 구인·구직 불균형 해소를 위한 정책 수립에 깊은 통찰을 제공했다는 점에서 실무적 시사점을 갖는다. 첫째, 본 연구는 2023년의 응답자의 기억 및 평소의 인식에 기반한 설문데이터 대신, 강소기업 현황 관련 최신의 빅데이터(big data)를 통합하여 분석을 위한 자료로 사용함으로써, 청년 구직자의 강소기업에 대한 변화하는 선호도를 정확히 반영하는 결과를 도출하였다. 따라서, 강소기업 혹은 강소기업으로 인증받기를 원하는 중소기업의 경우, 본 연구결과를 바탕으로 변화하는 구직자의 욕구와 기대사항 및 선호도를 파악하고, 이를 바탕으로 한 구직자

맞춤 채용전략 마련이 가능할 것으로 기대한다. 둘째, 본 연구는 강소기업과 청년 구직자 간의 상호작용에 초점을 맞추었다. 강소기업들은 청년 구직자의 선호와 기대를 충족시키는 전략을 마련하고, 그 전략을 통해 우수 인재를 유치하고 장기간 유지하는 방법을 탐색할 수 있을 것이다. 반면, 구직자들은 자신의 선호도와 기대치를 명확히 인식하고, 그에 부합하는 기업을 선택하는 데 본 연구의 결과를 활용할 수 있을 것으로 기대한다. 마지막으로, 정부, 대학 등 청년층의 안정적인 일자리 창출을 도모하는 기관 또한, 본 연구결과를 기반으로 강소기업 지원 및 발전전략 정책 마련의 방향성을 제공받을 수 있을 것으로 기대한다. 분석결과와 관련한 정부, 강소기업, 청년 구직자 차원의 실무적인 활용방안은 <표 7>에 정리하여 나타내었다.

### 5.3 연구의 한계점 및 향후 연구방향

본 연구의 한계점 및 향후 연구방향은 다음과 같다. 첫째, 본 연구에서 활용한 데이터 세트에서는 강소기업에 대한 충분한 선호요인 변수가 존재하지 않아, 청년 구직자들의 다양한 선호요인을 전반적으로 파악하지 못한 한계가 존재한다. 본 연구에서는 고용노동부와 청년워크넷서 제공하는 데이터를 활용하여 강소기업의 특성을 분석하였지만, 분석대상으로 선정된 기업들은 이미 어느 정도의 기준을 충족한 기업이 대부분이었다. 즉, 구직자들이 선호하는 요인으로 도출된 결과는 기업의 업종, 소재지, 직원 수 등과 같이 단기간에 수정 및 보완할 수 없는 요인들이 많아 현재 강소기업의 위치에 있는 기업이 이들 변수를 홍보하는 데에는 활

용이 가능한 반면, 강소기업의 위치를 희망하는 중소기업으로서는 즉각적인 반영이 어려운 한계가 존재한다. 이에, 향후 연구에서는 사람인, 잡코리아, 인크루트 등의 다양한 취업정보 사이트를 통한 기업정보 비교, 그리고 키워드네트워크 분석을 활용한 익명 게시판에서의 구직자의견조사 등의 방법을 통해, 더욱 다양한 구직자 선호요인을 도출할 수 있기를 기대한다.

둘째, 본 연구결과를 해석하는 데 있어 일반화의 한계가 존재한다. 본 연구는 청년워크넷 홈페이지에서 제공하는 ‘관심기업’이라는 지표를 활용하여 결과변수를 측정하였는데, 이 홈페이지는 청년 대상의 다양한 혜택을 제공하기 위해 설계되어 있다. 따라서, 본 연구의 결과를 청년층이 아닌 일반적인 구직자로 확대하여 해석하기에는 무리가 따른다. 최근의 고령화 및 정년 나이 하락 추세에 힘입어 청년층 외에도 60대 이상의 고령 구직자 비율 또한 증가하고 있다. 이에, 향후 연구에서는 청년층을 넘어 다양한 연령대의 구직자에 대한 강소기업 관련 선호요인을 조사할 필요가 있다.

마지막으로, 본 연구에서는 최신 머신러닝 또는 딥러닝 기법을 고려하지 못한 방법론적 한계가 존재한다. 예컨대, 딥러닝, 특히 트랜스포머(transformer)와 같은 최신의 고급 예측모형들은 정형 및 비정형 빅데이터에 잠재되어 있는 복잡한 패턴과 상호작용을 포착하는 데 탁월한 성능을 보여주고 있다. 이러한 딥러닝 기반 예측모형은 대용량 데이터 세트에서 비선형 관계나 상호작용을 파악하는 데 특화되어 있기 때문에, 본 연구에서 최신 딥러닝 기법을 적용하지 않은 점은 본 연구결과 타당성을 일정 부분 제한할 수 있다. 이에, 향후 연구에서

는 최신 딥러닝 기법을 활용한 데이터 분석을 고려하여 분류성능을 제고함으로써, 더욱 세밀하고 엄격한 연구가 수행될 수 있기를 기대한다.

## 참고문헌

- 길재욱, 김송희, 이은정, “강소기업을 위한 히든 챔피언 선정 효과에 관한 연구,” 한국증권학회지, 제46권, 제3호, 2017, pp. 687-722.
- 김수욱, “글로벌 강소기업의 성공 요인 분석,” 기업가정신과벤처연구, 제13권, 제1호, 2010, pp. 87-114.
- 김영생, “중소기업의 취업 정보와 청년층 구직 활동의 미스매치,” 취업진로연구, 제1권, 제1호, 2011, pp. 127-144.
- 김영직, 조민호, “4년제 대졸 청년들의 구직선호 유형별 진로경향 및 근로여건 분석: 잠재적 집단분석과 다항 로지스틱 회귀 분석을 통하여,” 한국정책학회 동계학술발표논문집, 2015, pp. 1095-1127.
- 김형석, 이정환, “청년구직자의 중소기업 취업과 선호일자리 정보가치에 대한 인식 차이 비교 연구,” 경영과정보연구, 제40권, 제4호, 2021, pp. 49-66.
- 대한상공회의소, “청년 구직자 일자리 인식조사”, 대한상공회의소 고용노동정책팀 기고, 2023.05.25.
- 류옥현, 이충석, “중소기업 구인 해결을 위한 구직자 중심의 기업평가모형 개발”, 중소기업연구, 제32권, 제1호, 2010, pp. 89-105.
- 매일일보, “중소기업 인력난 심화에도...‘그냥 쉬는’ 청년 50만명 육박”, 김혜나 기고, 2023.03.27.
- 이동훈, 김태형. “머신러닝 기법을 활용한 대졸 구직자 취업 예측모델에 관한 연구.” 정보시스템연구, 제29권, 제2호, 2020, pp. 287-306.
- 윤병운, 이성주, “중소기업의 오픈 이노베이션 모형,” 기술혁신학회지, 제13권, 제1호, 2010, pp. 160-183.
- 이병환, 허문구, “강소기업의 전략, 환경 및 조직구조가 기업활동과 성과에 미치는 영향: 구성 형태적 접근,” 전략경영연구, 제17권, 제3호, 2014, pp. 139-161.
- 이영주, 윤동진, “글로벌 강소기업으로 성장하기 위한 역량,” 국제경영연구, 제26권, 제2호, 2015, pp. 33-61.
- 이정록, “지역 일자리에 대한 잠재적 구직자의 선호도 분석: 광양제철소 협력사를 사례로,” 한국경제지리학회지, 제22권, 제3호, 2019, pp. 337-350.
- 이정환, 김동욱, “청년구직자의 중소기업 고용주 브랜드 인식이 취업의향에 미치는 영향 분석,” 한국콘텐츠학회논문지, 제22권, 제4호, 2022, pp. 292-300.
- 이충석, 류옥현, “구직자 유형별 직장선택 요인에 대한 탐색적 연구,” 중소기업연구, 제33권, 제4호, 2011, pp. 5-18.
- 중소기업신문, “독일 강소기업의 성장비결은”, 나중호 기고, 2022.10.31.
- 정동균, 이종화, 이현규, “머신러닝을 이용한 국내 수입 자동차 구매 계약 예측 모델 연

- 구: H 수입차 딜러사 대상으로,” 정보 시스템연구, 제30권, 제2호, 2021, pp. 105-126.
- 정인교, 전재완, 조정란, 한하늘, 홍진영, “강소 기업의 성장 영향 요인 분석,” 인하대학교 정석물류통상연구원 연구총서, 2014, pp. 1-35.
- 전주 MBC, “1년 내내 채용 중... 중소기업 구인난 심각”, 고차원 기고, 2023.06.06.
- 장호준, 고영희, “한국의 글로벌 중소기업의 성공요인에 대한 사례 연구: IT 산업 분야를 중심으로,” 국제경영리뷰, 제18권, 제2호, 2014, pp. 25-56.
- 최재우, 강운선, “4년제 대학 졸업 구직자의 구직선호 유형과 성별과 졸업대학 소재지역에따른 차이,” 동아인문학, 제33권, 2015, pp. 633-652.
- 청년워크넷, (2023), 중소기업선정기업 홈페이지, <https://www.work.go.kr/jobyoung/main.do>
- 통계청, “연령/활동상태별(쉬었음) 비경제활동인구,” 고용통계과 기고, 2023.10.13.
- 허성호, “구인기업과 구직자 간의 채용경향성 비교 연구: 채용프로세스를중심으로,” 디지털융복합연구, 제18권, 제7호, 2020, pp. 85-92.
- KBS, “졸업 후 미취업 ‘청년 백수’ 126만명... ‘절반은 대졸 이상’, 공민경 기고, 2023.08.27.
- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M., “Optuna: A next-generation hyperparameter optimization framework,” *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019, pp. 2623-2631.
- Breiman, L., and Cutler, R. A., “Random forests,” *Machine learning*, Vol. 45, 2001, pp. 5-32.
- Chen, T., and Guestrin, C., “XGBoost: A scalable tree boosting system,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 2016, pp. 785-794.
- Freedman, D., and Diaconis, P., “On the histogram as a density estimator: L2 theory,” *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, Vol. 57, No. 4, 1981, pp. 453-476.
- Friedman, J. H., “Greedy function approximation: A gradient boosting machine,” *Annals of Statistics*, Vol. 29, No. 5, 2001, pp. 1189-1232.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., and Oliphant, T. E., “Array programming with NumPy,” *Nature*, Vol. 585, No. 7825, 2020, pp. 357-362.
- Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X., “Applied logistic regression,” *John Wiley & Sons*, Vol. 398, 2013.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen,

W., Ma, W., Ye, Q., and Liu, T.-Y., "LightGBM: A highly efficient gradient boosting decision tree," *Proceedings of the 31st Conference on Neural Information Processing Systems*, Vol. 30, 2017.

Lee, M., Song, Y. H., Li, L., Lee, K. Y., and Yang, S.-B., "Detecting fake reviews with supervised machine learning algorithms," *The Service Industries Journal*, Vol. 42, No. 13-14, 2022, pp. 1101-1121.

Lundberg, S. M., and Lee, S.-I., "A unified approach to interpreting model predictions," *Proceedings of the 31st Conference on Neural Information Processing Systems*, Vol. 30, 2017

Novaković, J. D., Veljović, A., Ilić, S. S., Papić, Ž., and Tomović, M., "Evaluation of classification models in machine learning," *Theory and Applications of Mathematics & Computer Science*, Vol. 7, No. 1, 2017, pp. 39-46.

Rousseeuw, P. J., and Hubert, M., "Robust statistics for outlier detection," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 1, No. 1, 2011, pp. 73-79.

Sarker, I. H., "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, Vol. 2, No. 3, 2021, pp. 160.

**조윤주 (Cho, Yoon Ju)**



경희대학교 일반대학원에서 빅데이터응용학과 경영전공으로 석사학위를 진행 중이다. 주요관심 분야는 데이터분석, 비즈니스 애널리틱스, 데이터 사이언스 등이다.

**김진수 (Kim, Jin Soo)**



경희대학교 일반대학원에서 빅데이터응용학과 공학전공으로 석사학위를 진행 중이다. 주요관심 분야는 시계열 예측, 딥러닝 아키텍처 등이다.

**배환석 (Bae, Hwan seok)**



경희대학교 일반대학원에서 빅데이터응용학과 공학전공으로 석사학위를 진행 중이다. 주요관심 분야는 금융 기술, 비즈니스 프로세스 자동화, 데이터 과학 등이다.

**양성병 (Yang, Sung-Byung)**



KAIST에서 경영공학 박사 학위를 취득하고, 한성대학교 경영학부, 아주대학교 e-비즈니스학과를 거쳐 현재 경희대학교 경영학과/빅데이터응용학과 교수로 재직 중이다. 주요 관심분야는 소셜미디어, 온라인 커뮤니티, 온라인 리뷰, 비즈니스 애널리틱스, 스마트 관광 등이다.

**윤상혁 (Yoon, Sang-Hyeak)**



연세대학교 정보대학원에서 박사 학위를 취득한 후, 현재 한국기술교육대학교 산업경영학부 조교수로 재직 중이다. 주요 연구분야는 생성형 인공지능, 디지털 마케팅, 비즈니스애널리틱스 등이다.

<Abstract>

## **Developing a Predictive Model of Young Job Seekers' Preference for Hidden Champions Using Machine Learning and Analyzing the Relative Importance of Preference Factors**

Cho, Yoon Ju · Kim, Jin Soo · Bae, Hwan seok · Yang, Sung-Byung · Yoon, Sang-Hyeak

### **Purpose**

This study aims to understand the inclinations of young job seekers towards "hidden champions" - small but competitive companies that are emerging as potential solutions to the growing disparity between youth-targeted job vacancies and job seekers. We utilize machine learning techniques to discern the appeal of these hidden champions.

### **Design/methodology/approach**

We examined the characteristics of small and medium-sized enterprises using data sourced from the Ministry of Employment and Labor and Youth Worknet. By comparing the efficacy of five machine learning classification models (i.e., Logistic Regression, Random Forest Classifier, Gradient Boosting Classifier, LGBM Classifier, and XGB Classifier), we discovered that the predictive model utilizing the LGBM Classifier yielded the most consistent performance.

### **Findings**

Our analysis of the relative significance of preference determinants revealed that industry type, geographical location, and employee count are pivotal factors influencing preference. Drawing from these insights, we propose targeted strategic interventions for policymakers, hidden champions, and young job seekers.

**Keyword:** Hidden Champions, Job-seeking preference, Machine Learning

\* 이 논문은 2023년 11월 7일 접수, 2023년 12월 11일 1차 심사, 2023년 12월 20일 게재 확정되었습니다.