조 영 일 교수 지도

박사학위 청구논문

인적자원관리(HRM)분야에서 빅데이터 분석의 편향 식별 및 처리

: 학습자료와 이론적 근거의 중요성을 중심으로

2021

성신여자대학교 대학원

심리학과

구 소 희

인적자원관리(HRM)분야에서
빅데이터 분석의 편향 식별 및 처리

: 학습자료와 이론적 근거의 중요성을 중심으로

조 영 일 교수 지도

이 논문을 박사학위 논문으로 제출함

2021년 4월

성신여자대학교 대학원
심리학과
구 소 희

인 준 서

구소희의 박사학위 논문으로 인준함.

2021년 4월

심사위원장 김 명 선 (인)

심 사 위 원 신 경 수 (인)

심 사 위 원 차 옥 균 (Okkyoon Cha)

심 사 위 원 김 혜 영 (인)

심 사 위 원 조 영 일 (인)

성신여자대학교 대학원

논문개요

본 논문은 사회과학 분야, 특히 산업 및 조직심리 장면에서 인적자원관리를 할 때 인공지능 알고리즘을 사용하여 데이터를 분석하는 방법과 기존의 통계분석 방법의 장단점에 대해 설명하고 분석 결과를 비교하였다.

본 논문은 세 가지 연구로 구성되었으며 첫 번째 연구는 빅데이터에 대해 소개하고 산업 및 조직 심리 장면에서 고용 및 직무배치에 사용되는 AI면접의 장점과 단점에 대해 논의하였다. 이어서 두 번째 연구는 빅데이터 기반의 기계학습을 실시할 경우, 편향된 학습자료를 활용하게 되면 발생하는 예측 및 분류 오류를 확인하였다. 이에 따라 발생하는 편향성과 차별성으로 인해 인재 선발에 부적절한 영향을 미치지 않도록 조심해야 한다는 점을 논의하였다. 마지막으로 세 번째 연구에서는 고용노동부와 한국고용정보원이 제공하는 경험적 자료인 '대졸자 직업이동 경로조사'를 활용하여 기존의 설명중심 통계모형과 예측중심의 기계학습의 분석 결과를 비교하였다. 결과적으로 예측중심의 기계학습인 랜덤포레스트가 설명중심 통계모형인 로지스틱 회귀분석보다 나은 예측력과 분류정확도를 보였으나 결과에 대한 해석이 모호하거나 고용정책법에 위반되는 변수들이 예측에 유의미한 역할을 수행하는 것으로 확인하였다.

빅데이터를 사용하는 인공지능의 역할이 점차 중요해지는 4차 산업혁명 시대에 무엇보다 필요한 것은 방대한 양의 데이터를 통해 얻는 통찰력과 예측력뿐만이 아니라 인과성에 기반을 둔 결과에 대한 설명가능성과 투명성이다. 이를 위해서는 분석 전에 학습자료의 편향성을 검토해야 하며, 이론과 경험적 자료의 상호보완적인 사용을 통해 안정성이 보장된 공정하고 편향되지 않은 의사결정을 내려야 한다. 인공지능의 정확성과 투명성은 상충관계에 있기 때문에 설명가능 인공지능(XAI) 개발을 통해 높은 정확성을 유지하되 알고리즘의 편향을 방지하고 보다 설명 가능하고 투명한

기계학습 기술을 연구해야 한다.

주요어: 빅데이터, 산업 및 조직 심리, 인적자원관리, 인공지능, 기계학습, 로지스틱 회귀분석, 랜덤 포레스트, 설명가능 인공지능(XAI)

목 차

표 목 차

그 림 목 차

I. 서론

1. 연구의 필요성 및 목적

전 세계에 현재 존재하는 모든 데이터의 90%는 불과 지난 2년 동안 생산되었다(Devakunchari, 2014; Skilton & Hovsepian, 2017). 방대한 양의 데이터를 분석하면 이전에 알지 못했던 또는 기대하지도 못했던 사실을 발견하고 요인 간의 관계를 확인할 뿐만 아니라 탐색적으로 접근함으로 인해 많은 통찰력을 얻을 수 있다(Mazzocchi, 2015; Mazzocchi, 2020). 과학기술의 발달과 함께 데이터의 규모가 커짐에 따라 인공지능을 학습시켜 의료진단, 법적 의사결정과 같이 중요한 의사결정, 즉 인간의 인지능력이 필요한 작업을 컴퓨터가 대신 수행하기 시작했다(Tambe et al.,, 2019). 이로 인해 20년 전에는 공상 과학으로만 여겨졌던 것들이(예를 들어, 사물인터넷 IoT, 가상현실 VR 또는 말하는 컴퓨터 AI 등) 이제는 세계를 제 4차 산업혁명으로 이끌어 가고 있는 주역이 되었다. 4차 산업 혁명은 빅데이터의 활용으로 기계의 지능화와 자율화를 통해 새로운 가치 창출에 핵심을 두는 것이다(석현덕 등, 2017). 패러다임의 변화를 통하여 새로운 가치를 창출하는 과정에서, 4차 산업혁명은 인간이 환경 및 서로와 상호작용하는 방식뿐만 아니라 정치, 경제 사회 등 다양한 분야에서 변화를 주도하고 있다(Ghislieri et al., 2018). 초기 수준의 기계학습을 넘어서 신경망을 사용하는 딥러닝 또한 데이터가 풍부해짐에 따라 더 보편화 되었고 인간의 의사결정을 모방하는 능력을 가진 진정한 인공지능에 한껏 더 가까워지고 있다(석현덕 등, 2017; Tambe et al., 2019).

Mayer-Schönberger & Cukier (2013)는 빅데이터 분석 접근방식이 전례

없는 풍부한 데이터는 분석에 대한 더 폭넓은 시야를 제공할 수 있다고 밝혔다. 동일한 문제의 여러 측면을 조사하여 임의의 부분에 초점을 맞추지 않고 포괄적이며 전체적인 그림을 제공할 수 있다. 즉 규모가 큰 표본으로 인하여 표집으로 인한 표본 간의 변산성이 줄어든다. 이는 결과적으로 표집오차가 줄어들어 표본에 기초한 해석에 대한 우려가 줄어들 수 있다. 또한 빅데이터를 통해 정확성에 대한 갈망을 줄일 수 있다. 내적타당도는 측정하고자 하는 것을 정확히 측정했는가를 보여준다. 따라서 통제되고 단순화 된 조건에서는 보다 높은 내적타당도를 기대할 수 있다. 하지만 분석결과를 일반화할 수 있는 가능성은 줄어들 수밖에 없다. 이러한 문제는 외적타당도에 해당하는데 빅데이터로 인해 연구자들은 현실의 복잡성을 반영한 데이터의 복잡성을 볼 수 있게 되어 외적타당도를 높일 수 있다.

빅데이터를 기반으로 하는 인공지능은 의료산업, 자동차 산업, 소셜 미디어, 광고 및 마케팅 분야 등 인간 활동의 모든 영역에서 빠른 속도로 발전하고 있다. 뿐만 아니라 심리학 영역에도 빅데이터 분석이 소개되면서 연구 주제와 연구 방법에서 다양한 변화가 예상되고 있다(김청택, 2019). 특히 산업 및 조직 심리학자는 인사 선발과 인적자원관리에서 인공지능을 활용하여 의사결정의 직무 관련성, 타당성 및 공정성 문제를 해결하고자 한다(Oswald, 2020). 이 뿐만 아니라 다양한 이유 때문에 인공지능이 사용될 수 있다. 예를 들어, 구직 지원자 참여 증가, 새로운 측정 방식을 통한 직무 관련 특성 이해(VR, 비디오 게임 등), 실제 자아를 측정하기 위해 자연스럽고 눈에 띄지 않는 방법으로 구직자를 측정(비디오 녹화를 통한 행동, 얼굴 및 신체적 표현과 구두대화 분석)하기 위함 등에 인공지능이 활용되고 있다. 이를 위한 빅데이터 및 인공지능 평가 도구로는 인지측정(작업기억 측정 및 처리속도), 비디오 게임(가상현실 게임), 비디오와 오디오 기반 상호 작용평가(AI 인터뷰) 또는 온라인 자료

통합(이력서, 소셜 미디어 프로필 및 게시물) 등이 있다. 이처럼 다양한 데이터에 기계학습이 적용되어 인재선발, 인사평가, 직무배치 등의 인사결정이 내려지기도 한다(Oswald, 2020).

하지만 인간에게 중요한 의사결정에 데이터 분석의 결과가 적용되어 개개인의 삶에 영향을 미쳤을 때, 이러한 의사결정이 내려지는 방식과 기준이 사회가 일반적으로 중요하다고 생각하는 기준과 갈등을 일으키면 다양한 문제가 발생할 수 있다. 특히 기계학습 알고리즘은 인간에 의해 생성된 과거 데이터를 기반으로 학습하기 때문에 인간이 가진 편견을 학습할 위험이 있다. 그리고 빅데이터에서 학습하는 알고리즘은 특정 사회적 범주를 통계적 규칙성, 고정관념 및 과거 차별과 연관시킬 수 있는 위험성이 있다(Williams et al., 2018). 따라서 기계학습 알고리즘은 인종적, 성별적 편향 등을 나타내는 결과를 보여줄 수 있으며 이는 특히나 인적자원관리 분야에서 윤리적 문제를 야기하여 고용정책기본법 등을 위반하는 결과를 초래할 위험이 있다(Lum & Isaac, 2016). 인적자원관리 분야에서 빅데이터를 사용한 기계학습 알고리즘의 결과를 그대로 받아들이게 되면 현재 그대로의 상황을 유지할 뿐 향후 성장 가능성에 대한 기회와 고려는 배제될 것이며 나아가 사회에 현존하는 편향과 편견을 고착시킬 위험성이 존재한다. 이를 방지하기 위해 빅데이터 분석 시 먼저 학습자료의 편향성을 검토해보아야 한다. 이를 위해서 인적자원 관련 데이터의 주기적인 업데이트가 필요하고 다양한 학습자료의 활용으로 결과비교를 통해 편향성을 확인해보는 것이 필요하다.

이에 따라 연구 1에서는 먼저 개관연구로 산업 및 조직 심리 분야에서 빅데이터 사용에 대해 소개하고 연구 2에서는 모의실험을 통해 편향된 학습자료를 사용함으로써 발생하는 예측 및 분류 오류를 검증하였다. 또한 학습자료의 편향성 검토뿐만 아니라 기계학습 분석결과가 타당한지에 대한

뒷받침 되는 설명이 함께 제시되어야 함을 강조하였다. 편향성과 타당성에 대한 검증을 통해서 인적자원관리 분야에서 효율적이고 정확하며 공정한 의사결정을 내릴 수 있음을 보였다. 또한 공정한 의사결정을 위해서는 이론과 데이터의 순환 반복적인 과정을 거쳐야 함에 주목하였다. 이를 검증하기 위해서 연구 3에서는 자료에 기반을 둔 기계학습 분석결과의 해석이 타당한지 알아보았다. 경험자료를 사용하여 전통적인 통계분석의 결과와 기계학습 결과의 비교를 통해 각각의 장단점을 알아보고 이론과 데이터 사이에 어떠한 균형이 효율적이고 정확하며 공정한 의사결정을 도모할 수 있는지 확인하였다.

결론적으로 본 논문을 통해 인적자원관리 분야에서 빅데이터를 사용하는 경우, 인적자원을 공정하고 합리적이며 효율적으로 선발 및 배치하기 위해 기계학습 알고리즘이 사용되는 학습자료의 편향성을 검증하고 확인하였다. 또한 빅데이터를 사용한 기계학습의 결과를 인적자원관리 의사결정에 적용할 시에 평가의 다양성과 공정성을 위해 결과 해석의 투명성을 강조하였다. 탐색적인 방법을 통해 다양하고 창의적인 이론을 구축하고 확인적인 방법을 통해 이론을 기반으로 하는 해석용이성을 갖는 상호보완적인 방법을 정교화 시키는 것을 제안하였다. 또한 과거데이터를 학습함으로써 생기는 편향성을 방지하기 위해 학습데이터의 주기적인 업데이트가 필요하며 여러 다른 학습 데이터를 활용한 결과의 비교를 통한 편향성의 재확인이 중요하다. 또한 모형의 적합도, 타당도 검토가 선행되는 것이 바람직하며 선행연구나 이론을 통해 중요한 독립변수를 식별하고 데이터를 클리닝 하는 작업을 거쳐야 한다.

II. 이론적 배경

1. 인적자원관리 분야에서 빅데이터 활용의 실태

인공지능과 기계학습은 우수한 구직자를 식별하고 고용하려는 조직에서 널리 사용되고 있다. 하지만 산업 및 조직 심리학자들의 전문적 참여의 양, 다양성 및 속도는 인재 평가 및 선발을 위해 기계학습 프로그램을 개발하고 평가하기에는 아직 제한적이다(Gonzalez et al., 2019). 또한 조직의 맥락에서 기계학습 도구의 신뢰성, 타당성 및 공정성을 조사하는 경험적 연구가 부족한 실태이다(Angrave et al., 2016; Huselid, 2018; Scholz, 2017). 따라서 조직의 맥락에서 기계학습 및 인공지능 애플리케이션을 개발, 구현 및 평가하는데 있어 산업 및 조직 심리학자와 컴퓨터 과학자, 법률학자 및 기타 전문 분야 구성원 간의 협력 증가가 필요하다(Gonzalez et al., 2019).

인공지능과 기계학습으로 인해 인적자원관리 중 인재확보 분야에서 큰 추진력을 얻은 것은 사실이다. 많은 조직들이 더 빠른 속도와 효율성으로 구직자를 식별, 모집 및 선택하기 위해 빅데이터를 사용하여 인적자원관리를 도모하고 있다(Stephan et al., 2017). 예를 들어 인공지능을 이용하는 채용 플랫폼 TalVista는 더 다양한 지원자를 식별하기 위해 직무기술에서 편향을 제거하고자 노력하고 있으며 HireVue나 Montage처럼 비디오 면접을 통해 채용을 위한 인터뷰 점수를 매기는 방법도 생기고 있다(Gonzalez et al., 2019). 이 밖에도 Pymetrics와 Knack은 기계학습 알고리즘과 게임을 융합하여 인간의 인지, 감정, 사회적 능력을 평가하는 신경심리 검사를 개발해 개인의 성과를 측정한다(Gonzalez et al., 2019). 하지만 이러한 빅데이터를 사용한 기계학습의 결과에 대한 신뢰성 타당성 및 공정성에 대한 증거와 뒷받침되는 이론이 아직 부족한

실태이다(Gonzalez et al., 2019). 인적자원관리 분야에서 빅데이터를 통한 기계학습의 개발과 구현속도는 과학적 연구 및 법적 지침을 앞지르고 있다(Gonzalez et al., 2019; Scholz, 2017). 따라서 현재 조직은 빅데이터를 사용한 기계학습이나 인공지능에 대한 잠재적 한계와 윤리적 문제에도 불구하고 이에 대한 과도한 신뢰와 수용을 하고 있는 상태이다(Gonzalez et al., 2019).

인적자원관리 분야에서 빅데이터 분석을 통해 얻을 수 있는 이점은 분명하다. 조직이 더 짧은 시간 내에 더 많은 지원자 데이터를 식별하고 검토함으로써 효율성이 높아진다(Das et al., 2018). 기계학습 알고리즘으로 인해 다양하고 새롭고 더 나은 직원은 배출하지 못한다 하더라도 기존 평가방법보다 더 빠른 의사결정 속도와 효율성을 제공하여 시간과 비용을 절약하는 것은 사실이다(Gonzalez et al., 2019). 하지만 기계학습을 이용한 결과가 기존의 방법보다 더 정확한지, 또는 어느 부분에서 얼마큼 개선되었는지는 아직 밝혀지지 않았으며 이에 대한 증거 데이터와 후속 연구가 필요한 상황이다(Gonzalez et al., 2019).

Sajjadani et al. (2019)는 16,071명의 지원자에 대한 종단 빅데이터에 기계학습 기술을 적용하여 업무 경험 관련성, 임기 이력 및 비자발적 이직 이력, 더 나은 직업을 찾아본 이력 등에 대한 해석 가능한 측정치를 개발했다. 이에 따라 빅데이터 분석을 통해 조직 내의 새로운 데이터에 대한 접근을 제공하고 정교한 알고리즘을 적용하여 조직에 더 높은 효율성을 제공할 수 있다고 밝혔다. Bara et al. (2015)은 인적자원관리를 위한 지원자들의 선발을 돕기 위해 프로토타입 시스템을 개발했고 Indira and Kumar(2016)는 기존의 주요 검색어를 사용하는 응용 프로그램과 달리 문장의 동작을 이해하는 텍스트 분석을 사용하여 이력서를 걸러내는 응용프로그램을 개발하고 소개했다. 또한 Park et al. (2015)는 소셜 미디어

페이스북의 66,732개 사용자의 자가 보고된 인터뷰를 통해 얻은 5가지 성격특성과 예측모형을 통해 얻은 성격 특성을 비교, 예측하는 연구를 진행하였다. 이를 통해 기계학습이 인간보다 훈련하는데 더 빠르기 때문에 효율적이며 인간의 의사결정의 정확성은 사람 내에서 그리고 사람 간에 불안정하게 변동한다는 점을 감안할 때 기계학습이 더 신뢰성을 향상 시킬 수 있다고 주장했다. 따라서 조직은 이러한 빅데이터 분석을 통해 심리적 변수(예: 언어 유창성, 정직성, 정서성, 공격성 등)에 대한 측정치를 계산하여 중요한 직원 및 조직의 결과를 예측할 수 있다(Tausczik & Pennebaker, 2010).

산업 및 조직심리학자들이 HR전문가, 데이터 과학자, 조직 의사결정권자, 변호사 및 정책 입안자와 인적자원관리에서 빅데이터를 사용한 기계학습의 사용에 대해 협력할 때 이에 대한 실질적 및 잠재적 강점과 의미, 그리고 한계점 사이에서 균형을 맞춰야 한다(Gonzalez et al., 2019; Scholz, 2017). 첫 번째로는 시간, 노력, 비용 및 인적자원 측면에서 비용을 절감할 수 있다는 장점은 있지만 결과의 예측력은 학습 데이터의 품질과 사용된 알고리즘의 적합성에 의해 결정되기 때문에 데이터 품질 및 의사결정에 대한 검토가 필요하다. 두 번째는 과학 기술의 발달로 방대한 양의 데이터를 처리할 수 있는 성능은 향상되고 있지만 기계학습의 알고리즘은 분석과정을 설명할 수 없는 블랙박스가 존재함으로 결과 예측에 대한 이유가 명확하지 않다는 것을 염두 해두어야 한다. 세 번째로 대량의 데이터, 소셜 미디어 컨텐츠, 인터뷰 응답 등 예전에는 분석이 어려웠던 데이터 형식에 대한 접근성이 향상하고 있지만 개인정보보호와 관련된 향후 윤리적 및 법적 문제의 가능성에 대해 고민해야 한다. 마지막으로 중요한 개인 및 조직의 관심결과에 대한 예측 정확도가 향상하고 있지만 구직자, 언론, 일반 대중의 비판적인 반응에 대한 가능성도 고려해야 한다(Gonzalez

et al., 2019). 이처럼 인적자원관리 분야에서 빅데이터 분석을 적용할 때 실질적인 강점과 한계점 사이의 균형을 맞추기 위해서 빅데이터 분석방법에 대한 다양한 연구와 활용도 중요하지만 동시에 한계점과 주의점에 대한 경각심 또한 가져야 한다. 특히 기계학습 알고리즘의 블랙박스 문제로 결과에 대한 이유가 명확하지 않기 때문에 인적 자원관리 분야에서 발생할 수 있는 윤리적 및 법적 문제를 예방하기 위해 의사결정에 대한 검토와 데이터의 품질과 편향에 대한 검토가 필요하다.

2. 빅데이터 활용 시 발생할 수 있는 편향

빅데이터 분석 시 인간에 의해 생성된 데이터를 알고리즘이 학습하기 때문에 인간이 인종차별을 하는 것처럼 알고리즘 또한 인종차별을 할 수 있다(O'Donnell, 2019). 예를 들어 인종적으로 편향된 형사 사법 시스템의 맥락에서 사용되는 예측치안유지(predictive policing) 알고리즘도 예외가 아니다(Brantingham, 2017). 왜냐하면 예측치안유지 알고리즘은 인간에 의해 생성된 과거범죄 데이터로 학습하기 때문에 유색 인종에 대한 인종적 편향이 포함되어 인종차별에 심하게 영향을 받는다. 특히 장소기반 예측치안유지 알고리즘은 특정 영역에 대한 위험점수를 생성하기 위해 빅데이터의 패턴을 설명하고 어느 지역에 경찰을 배치해야 하는지 결정을 내린다(Brantingham, 2017; O'Donnell, 2019). 인종이 범죄와 상관이 있다는 인종적 왜곡이 존재하는 미국 형사 사법 제도의 현실로 인해 인종적으로 왜곡된 데이터에서 두드러지는 패턴이 발견된다. 따라서 과거 범죄 데이터에 의존하여 이전에 발생한 범죄를 기반으로 미래 범죄가 발생할 위치를 예측하는 경우, 주로 흑인 및 히스패닉 지역에 범죄가 더 많다는 결과를 보였다(O'Donnell, 2019). 게다가 기계학습의 블랙박스 특성 때문에

경찰관은 알고리즘의 결과가 실제로 인종 차별에 감염이 되었음에도 불구하고 이 결과를 중립적으로 생각하고 무작정 믿을 수 있다. 이러한 악순환으로 인해 기계학습 알고리즘은 인종 차별적인 정책을 영속화 하는 위험이 발생할 수 있다. 알고리즘은 인종과 범죄를 연관 시키는 것이 옳다고 인식하고 후속 반복에서 이 연관성에 더 의존하여 의사결정을 하게 되면 인종 차별적인 정책이 악화되는 것이다(O'Donnell, 2019).

또한 예측 정책 알고리즘에 적용되는 전제를 정확하게 조사하기 위해 Lum & Isaac(2016)은 HRDAG(Human Rights Data Analysis Group)에서 주로 쓰이는 예측치안유지 알고리즘인 PredPol이 백인 인구를 대상으로 하는 방식과 흑인 인구를 대상으로 하는 방식을 구체적으로 비교 조사하였다. PredPol에서 사용된 경찰 기록 데이터의 효과를 조사하기 위해 캘리포니아 오클랜드에서 발표한 마약 범죄 기록 알고리즘을 사용하였다. 그 결과 경찰 데이터의 명백한 편향을 수정하는 대신 알고리즘이 편향을 강화한다는 것을 발견했다. 또한 PredPol이 피드몬트와 같은 백인 지역보다 흑인이 많은 지역인 서부 오클랜드에 경찰을 더 많이 배치한 다는 것을 입증했다. 또한 흑인이 백인보다 두 배의 비율로 알고리즘의 표적이 될 수 있었다. 이는 실제로 오클랜드에서 마약 사용 패턴이 인종 간에 동일하다는 사실과 대조되는 결과였다. 범죄는 모든 곳에서 발생할 수 있지만 경찰은 제한된 곳에서만 범죄를 찾고 있다는 결과이다. 이에 따라 예측치안유지 알고리즘은 역사적으로 과도하게 치안된 지역사회만 치안하여 지역사회의 치안이 점점 불균형 하게 된다는 시사점을 밝혔다.

이 밖에도 현재는 사라졌지만 마이크로소프트사의 자동 챗봇 테이는 알고리즘이 인종을 고려하도록 코딩 되지 않은 경우에도 편향된 학습자료로 인해 인종 차별적 알고리즘을 생성한 대표적인 예이다(Zemčík, 2021). 테이는 트위터를 통해 인종차별적인 논평과 편견에 노출되었고 이로 인해

인간이 가지는 인종 차별 주의적 성향을 알고리즘에 내면화 시키게 되었다(Zemčík, 2021).

또한 미국의 비영리 인터넷 언론기관은 예측 알고리즘이 백인 피고인에 비해 거의 두 배에 달하는 비율로 흑인 피고인을 미래의 범죄자로 잘못 예측하는 결과를 표시할 가능성이 있다고 밝혔다(Shapiro, 2017). 또한 백인 피고인은 흑인 피고인에 비해 낮은 위험인자로 잘못 분류되는 경우가 더 잦았음을 발견했다. 이러한 인종 차별적인 결과가 흑인 피고인의 이전 범죄 또는 경찰이 흑인 피고인을 체포한 범죄 유형에 기인할 수 있다는 우려를 없애기 위해서 범죄 기록, 재범률, 나이 및 성별에서 인종의 영향을 통제하였다. 하지만 흑인 피고인은 미래에 폭력 범죄를 저지를 위험이 77% 더 높다는 결과가 나왔다. 또한 흑인 피고인은 미래 범죄의 가해자로 식별될 가능성이 45% 더 높았다. 이 연구로 인해 기계학습 알고리즘이 인종과 범죄를 연관시키는 것을 학습했기 때문에 흑인 피고인에 대한 편견을 드러내기 시작한 것이라는 결론을 지었다.

챗봇 테이에 관한 연구와 프로퍼블리카의 연구에 따르면 과거 정책 데이터 및 편견으로 가득 찬 인터넷 검색과 같은 편향된 데이터를 기계학습 알고리즘 학습에 사용하면 알고리즘이 인종과 범죄 사이의 상관성을 잘못 도출할 수 있다(O′Donnell, 2019; Shapiro, 2017; Zemčík, 2021). 알고리즘이 인종을 변수로 사용하지 않도록 명시된 경우에도 이런 결과는 발생할 수 있다. 최근 사이언스 간행물에 따르면 기계 학습 알고리즘을 훈련하는데 사용되는 학습데이터는 종종 역사적인 편견을 반영하여 알고리즘이 인종과 범죄 같은 용어를 연결하도록 유도한다고 밝혔다(O′Donnell, 2019).

이처럼 기계학습은 인간이 생성한 데이터를 학습하기 때문에 인간의 편향까지 학습할 위험이 있다. 더불어 인적자원관리 분야에서 편향된 학습자료를 사용하여 결과를 도출하고 의사결정을 내릴 때 이론적 근거가

충분하지 않다면 이런 결과는 기업에게 큰 도움이 될 수 없다. 왜냐하면 빅데이터 분석의 목표는 분명 가치 창출이지만 단지 주어진 데이터의 분석을 통해 통찰력을 도출하여 의사결정을 내리는 것은 아무런 의미가 없기 때문이다. 어떠한 가치를 위해 어떠한 통찰력이 필요한지부터 설계하고 분석을 시작해야한다. 즉 원하는 가치를 먼저 정한 후, 그 가치를 창출하기 위해 어떤 인사이트를 추출할 수 있을지 판단하고 이를 위해 어떠한 이론, 데이터, 그리고 분석방법을 사용해야 하는지 기획해야 한다. 즉 신중한 연구 설계는 빅데이터 분석 전에 선행되어야 한다(조성준, 2019). 이와 같이 경험적 자료가 없어서 검증할 수 없는 이론은 가치가 없고 마찬가지로 엄청난 양의 데이터는 의미 있는 이론을 구성할 때까지는 쓸모가 없다(Bhattacherjee, 2012). 따라서 빅데이터로부터 새로운 지식을 창출하고 의미 있는 의사결정을 내리기 위해서는 데이터부터 시작하는 탐색적 단계와 이론부터 시작하는 확인적 단계가 순환적으로 이루어져야 한다.

3. 데이터와 이론

탐색적 분석은 귀납적 추론방식을 사용하여 상향식으로 데이터부터 분석하여 패턴이나 연관성을 찾아낸다(Bhattacherjee, 2012). 이론이 아닌 데이터로부터 분석이 시작되기 때문에 데이터로부터 패턴이 탄생하고 패턴에 대한 설명을 함으로써 가설이 생성된다. 이러한 의미에서 기존의 과학적 방법의 가설 검증의 특성과 달리 가설 생성의 과정으로 볼 수 있다(Mazzocchi, 2015; Mazzocchi, 2020). 기계학습의 알고리즘은 빅데이터 분석에서는 방대한 양의 데이터를 기반으로 자료주도적 접근방식을 사용하기 때문에(McAbee et al., 2017) 탐색적 분석의 형태를

띤다(Mazzocchi, 2015; Mazzocchi, 2020). 많은 예측변수가 포함된 다차원적인 데이터에서 규칙성과 상관관계를 발견하여 과학적으로 흥미로운 통찰력(예를 들어 패턴이나 규칙성)을 추출하고자 한다(Maass et al., 2018). 이렇게 새롭게 발견된 가설을 기반으로 향후 의사결정을 하게 된다. 따라서 많은 변수들을 분석에 모두 포함함으로서 예측력을 높이는데 초점을 둔다(Yarkoni & Westfall, 2017).

하지만 사회과학 분야에서는 빅데이터 분석처럼 방대한 자료로부터 이론을 구축하는 일은 특히 더 어려울 수 있다. 왜냐하면 이론적 개념(조작적 개념)이 관찰가능하지 않아서 애매모호한 특성을 지니고 있거나, 사용하는 도구가 측정에 부적절하거나, 관심이 있는 현상에 영향을 미칠 수 있는 설명되지 않은 오염변수들이 존재할 수 있기 때문이다(Maass et al., 2018). 이 때문에 사회과학 분야에서는 이론이 풍부(theory-rich)하고 데이터가 부족(data-poor)한 환경에서 연구가 진행되어 왔고 연역적 추론을 기반으로 하는 이론검증의 연구인 모형주도적 접근방식을 주로 사용해왔다(Hales, 2013). 모형주도적 접근 방식은 가설부터 시작하는 하향식으로 연구설계에 따라 데이터를 수집 및 분석한 후 가설을 검증하고 그 결과를 기반으로 이론적 결론을 도출하는 과학적인 탐구 방법이다(Maass et al., 2018). 모형주도적 접근 방식을 사용하는 경우 사전연구나 이론에 근거하여 가설을 구축하기 때문에 가설 지향적이며 이론검증 (theory-testing) 연구라고도 하며 여기서 이론검증의 목표는 이론을 검증하는 것뿐만 아니라 이론을 개선 및 확장하는 것을 포함한다. 또한 연역적적 추론방식으로 전제된 가설을 검증하는 과정을 거치기 때문에 확인적인 접근방식이다(Bhattacherjee, 2012). 또한 탐색적 분석과 달리 예측의 정확성 보다는 주로 현상에 대한 이해와 절차적 설명력을 높이는데 초점을 둔다(Yarkoni & Westfall, 2017). 탐색적 분석과 확인적 분석의

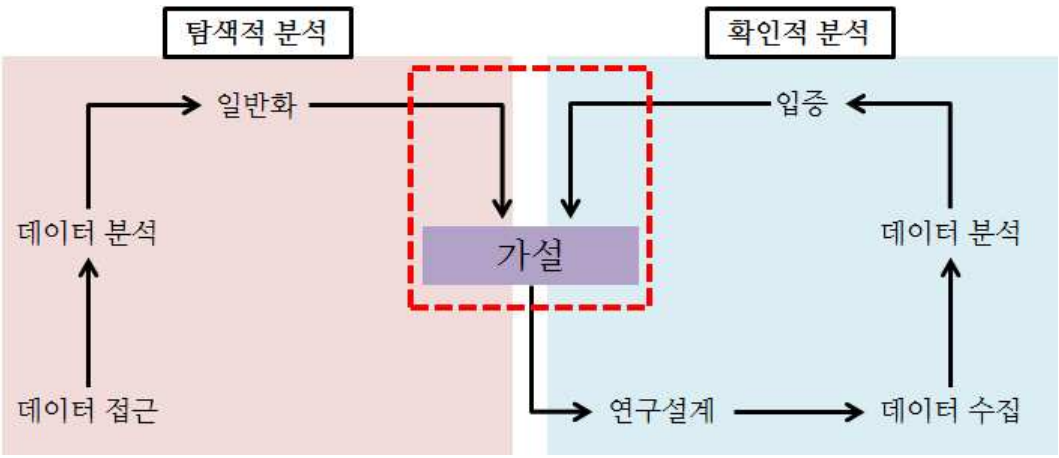
차이를 <표 1>에 정리하였다.

<표 1> 탐색적 분석과 확인적 분석의 비교

	탐색적 분석	**확인적 분석**
사전 지식	사전지식 없음	사전지식 있음
추론 방식	귀납적(inductive) 추론	연역적(deductive) 추론
접근 방식	자료주도적(data-driven) 접근	모형주도적(model-driven) 접근
설계 방법	상향식(bottom-up)	하향식(top-down)
주요 목표	예측(expectation) 중심	설명(explanation) 중심
분석 방법	기계학습 (machine learning)	과학적 방법 (scientific method)

예측력과 더불어 결과에 대한 이론적 설명이 필수인 사회과학 분야에서는 패턴 또는 규칙성을 찾기 전에 패턴이 무엇인지에 대한 정의를 전제로 하고 가설 혹은 모형 즉 이론에 기반을 두어야 한다. 따라서 사회과학 분야에서만큼은 빅데이터 분석을 시작하기 전에 이론을 필요로 해야 한다(Hales, 2013). 따라서 빅데이터는 이론에서 멀어지는 분석이 아니라 분석을 시작하기 전에 오히려 이론을 필요로 하는 분석이라고 할 수 있다(Hales, 2013). 탐색적 연구와 확인적 연구는 연구자의 목적에 따라 선택되어 사용될 수 있지만 서로 정확히 구분해서 사용하기 보다는 상호보완적으로 융합하는 방법도 건설적인 연구 설계가 될 수 있다(Orcan, 2018; Suhr, 2006). 두 가지 분석 방법이 상호보완적으로 사용되는 절차를 <그림 1>에 나타냈다. 탐색적 분석과 확인적 분석이 각각 실시되는 순서를 나타내고 있으며 빨간 점선상자로 표시한 부분에서 서로 교차하며 다시

구분되는 과정을 나타낸다.



<그림 1> 탐색적 분석과 확인적 분석의 상호보완적 역할
(Caliebe, Leverkus, Antes & Krawczak, 2019)

빅데이터 분석이 많은 분야에서 각광을 받기 시작하면서 Anderson(2008)은 페타바이트 정보 및 슈퍼 컴퓨팅 시대에 기존의 가설 지향적인 과학적 분석 방법은 더 이상 쓸모없어질 것이라고 주장했다. 뿐만 아니라 과학적 방법의 아버지라 불리는 Francis Bacon은 Novum Organum(1620)에서 과학적 지식은 선입견이 아니라 실험 데이터에 기반을 두어야 하고 연역적 추론은 결국 제한적이라 주장했다. 왜냐하면 실험 전에 가설을 설정하면 그 가설과 일치하도록 추론이 제한되기 때문이라고 했다. 이에 따라 Bacon(1620)은 하향식 접근을 하는 연역적 추론 대신 상향식 접근 방법을 옹호했다.

분명 자료 기반 접근 방식이 과학적 연구를 위한 새로운 도구가 될 것은 확실하다. 하지만 이것이 수세기에 걸친 인지적, 방법론적 절차를 대체할 것이라는 의미는 아니다. “이론의 종말”Anderson(2008)이 아니라 새로운

기회가 생긴 것이다(Hales, 2013). 이론적 가정과 가설의 역할을 무시하고 오랫동안 확립되어온 기존의 과학적 연구 방식이 모두 컴퓨터로 대체되는 것은 다시 생각해 보아야 할 문제이다. 자료주도적 연구를 통한 분석결과가 진정한 지식 생산 방식인지 아니면 무엇보다도 잠재적으로 유용한 정보를 식별하는 도구인지 생각해보아야 한다(Mazzocchi, 2015; Mazzocchi, 2020). 또한 자료주도적 접근을 사용할 시에 장기적으로 발생할 수 있는 결과뿐만 아니라 그에 따른 가치와 편향에 대해 좀 더 신중하게 살펴보아야 한다고 밝혔다(Mazzocchi, 2015)

알고리즘은 데이터 안에서 설명하지 못하고 있는 패턴이나 현상이 있을 수 있으며 이로 인해 편향된 결과를 나타낼 수 있다. 또는 패턴이 특정 데이터만을 설명하는 과적합으로 인해 주어진 데이터 보다 넓은 현상에 대해 일반화 시킬 수 없는 경우가 생길 수도 있다. 따라서 적절한 검증방법을 선택하는 것은 연구자가 어떤 연구문제를 가지고 있느냐에 따라 달라진다. 그래서 사회과학을 추구하려면 최소한 가져야 할 질문은 연구자가 무엇을 찾고 있는지, 연구문제가 무엇이고 가설이 무엇인지이며 이에 대한 답을 가지고 있어야 한다. 이는 의미 있는 과학이 되기 위해서 뿐만 아니라 연구문제에 따라 달라지는 알고리즘, 즉 분석결과가 윤리적, 정치적으로도 영향을 미칠 수 있기 때문이다. 다시 말해 기계학습 알고리즘으로 도출된 결과를 통해 사회적 현상에 대해 견해를 가지게 되고 이것은 다시 사회적 현실로 굳어지게 된다. 만약 분석결과 이런 질문에 대한 일관된 답이 나오지 않는다면 그 연구는 방향이 없고 최악의 경우 목표 자체를 인식하지 못한 것이다. 물론 데이터에서 패턴이나 규칙성을 찾는 것 자체가 연구 목표일 수 있다. 하지만 찾고자 하는 규칙성이 무엇인지에 대한 조작적 정의가 필요하며 그것은 가설이나 모형 즉 이론을 기반으로 해야 한다(Caliebe et al., 2019; Hales, 2013).

빅데이터 분석은 탐색적 방식을 통해 과거 경험을 기반으로 하여 현재의 지식, 실제 관찰 및 주관적 추론을 결합하여 가설생성을 촉진시킬 수 있다. 그리고 모든 가설은 평가가 필요(Caliebe et al., 2019)하기 때문에 생성된 가설을 확인적 방식을 통해 검증가능하다. 과학적 연구는 귀납적 추론과 연역적 추론의 반복을 통해 발생하기 때문에 의사결정을 내리기 위한 결과는 결코 탐색적 방법 또는 확인적 방법 하나로만 검증될 수 없다. 따라서 두 가지 방법이 가지는 각각의 단점을 최대한 피하고 두 가지 모두의 장점을 가질 수 있도록 연구문제에 알맞은 분석방법을 택하는 것이 중요하다. 특히 인적자원관리의 경우 효율성 추구와 공정성은 언제나 중요시 되는 부분이기 때문에 효율성과 적절성 측면에서 인공지능의 이점을 모두 발전시키는 방법을 찾아야한다(Tambe et al., 2019).

따라서 본 연구에서는 인적자원관리 분야에서 빅데이터 분석을 실시하는 경우에 확인적 방법과 탐색적 방법의 상호보완적 방법을 통해 이론적 근거를 가지고 학습자료의 편향성에 대한 위험을 최소화해야 한다고 제안했다. 연구 1에서는 인적자원관리 분야에서 빅데이터 활용에 대한 개관을 하고 연구 2에서는 편향된 학습자료를 사용함으로써 발생하는 분류와 오류의 수준을 모의실험을 통해 검증하였다. 마지막으로 연구 3에서는 경험적 자료를 사용하여 랜덤포레스트와 로지스틱 회귀분석 결과를 비교하여 두 방법의 장점을 모두 가질 수 있도록 데이터와 이론의 균형이 필요하다고 제언했다.

추가적으로 기계학습이란 규칙 기반 프로그래밍에 의존하지 않고 데이터로부터 직접 학습하는 알고리즘이며 통계 모델링은 수학적 방정식의 형태로 데이터에 존재하는 변수 간의 관계를 형식화 한 것이다(Shah, 2016). 하지만 이 같은 정의는 각 개념의 극단적 입장에 초점을 맞추는 경향이 있다(Mayo, 2017). 통계 모델링의 현대적인 정의는 불확실성을 측정하고

통제하며 의사소통하기 위해 데이터로부터 얻는 과학적 사실이다(Kuonen, 2004). 마찬가지로 기계학습의 단순화된 정의에는 데이터의 유무, 모형 또는 추정치, 그리고 손실이나 비용을 최소화 하는 세 가지로 구성된다(Mayo, 2017). 즉 기계학습은 인간이 손으로 해결할 수 있는 통계적 문제를 통해 데이터의 양과 질을 증가시켜가며 손실함수를 최적화 하는 과정이다(Shah, 2016). 따라서 회귀분석은 이러한 요구사항을 충족시키기 때문에 기계학습의 일종이라고 말할 수 있다(Mayo, 2017). 특히 연구 2에서는 학습자료의 생성에 사용되는 회귀식과 실제자료의 생성에 사용되는 회귀식이 다르게 사용되며 두 개의 회귀식에 기초하여 자료를 생성한 후 학습자료를 활용하여 알고리즘(중다회귀 방정식과 로지스틱 회귀 방정식)을 추정하기 때문에 기계학습의 일종으로 볼 수 있다.

빅데이터 분석 방법 중에 데이터에서 변수들 간의 연관성과 패턴을 찾아 결과를 예측하는 방법을 데이터 마이닝이라고 한다(Chen & Wojcik, 2016). 데이터 마이닝은 크게 지도학습과 비지도학습으로 나눌 수 있는데 지도학습은 입력값 X가 주어지면 입력값에 대한 Y를 주어 이들의 관계를 학습시키는 것을 말한다, 즉 정답이 있는 데이터를 사용하여 학습시키고 기계가 답을 잘 맞혔는지를 알 수 있다. 지도학습 알고리즘의 종류로는 k-최근접 이웃(k-nearest neighbors), 선형회귀, 로지스틱 회귀, 서포트 벡터 머신(SVM; Support Vector Machine), 결정트리(Decision Tree)와 랜덤포레스트 그리고 신경망(Neural Network)등이 있다(김청택, 2019). 서포트 벡터 머신은 서로 다른 결과를 잘 분류해주는 특징(feature)들의 선형 결합을 찾는 방법이며 랜덤포레스트는 데이터 마이닝 기법 중 하나로 기계학습에서 분류와 회귀분석에 사용되는 앙상블 학습 방법이다. 즉 많은 나무(tree)들을 통해 각각 분류 결과가 나타나면 다수결로 최종결과를 결정하는 방법이다. 하지만 랜덤포레스트는 다중 클래스를 분류하지만

서포트 벡터 머신은 다중 클래스 분류를 위한 데이터 처리가 필요하다. 가령 레이블 중 하나와 나머지 모두를 구별하거나 모든 클래스를 일대일로 이진 분류하는 방법이 있다. 또한 행이 10,000개 이상으로 많거나 데이터를 사전 처리할 시간이 충분하지 않거나 범주형 및 수치형 자료가 혼합된 경우는 랜덤 포레스트를 사용하는 것을 추천한다(Thanh Noi & Kappas, 2018; Liu et al., 2013; Ahmad et al., 2018). 따라서 연구 3에서 사용하는 데이터는 자료의 형식이 혼합되어 있고 다중 클래스 분류를 위한 데이터 처리를 추가적으로 하지 않았으므로 랜덤 포레스트를 사용하였다. 랜덤포레스트는 앙상블 기법 중 배깅(bagging)을 기반으로 한 대표적인 기계학습의 알고리즘(오세웅, 2017; 유진은, 2015)으로 비정상적인 값과 노이즈에 대한 내성이 좋고 예측 정확도가 높다고 확인되어 인적자원관리 분야에서도 활발하게 사용된다(Gao et al., 2019).

III. 연구 1
“인적자원관리 연구에서 빅데이터의 활용”

1. 서론

“제 4차 산업혁명” 이라는 용어는 비영리재단인 세계경제포럼(World Economic Forum;WEF)의 창시자 Klaus Schwab이 처음 사용하였다. 여기서 산업혁명을 기존 시스템의 운영방식을 발전시키고 변화시킨다는 측면에서 볼 때, 1차 산업혁명은 증기기관 발달과 같은 새로운 동력원으로 육체노동을 기계로 대체하는 기계화를 촉진시켰다. 2차 산업혁명 시기에는 전기 에너지 발달로 인한 대량 생산화 혁명을 이루어서 시장경제가 활성화되었고 효율성이 증가하였다. 3차 산업혁명은 IT기술인 인터넷과 컴퓨터 발달로 지식노동을 컴퓨터로 대체하는 지식정보화 및 자동화 시스템을 구축시켰다. 이번 4차 산업혁명은 3차 산업혁명처럼 하드웨어 중심이 아닌, 소프트웨어와 기계를 융합하여 데이터를 통한 가치 창출에 핵심을 두고 있다. 4차 산업혁명은 데이터의 확대로 빅데이터를 만들고 이를 기반으로 하는 기계학습(Machine learning; ML)과 인공지능(Artificial Intelligence; AI)을 통해 지능화를 촉진시키고 있다. 지능화를 통해 연결된 모든 사물, 기계, 시스템 또는 서비스의 자율화를 추진하여 생산과정의 최적화를 통해 지능정보사회를 구현하고자 한다.

사물들 간의 단순한 연결 과정뿐만 아니라 기업의 경영과 운영 과정에서도 인간의 행동 또는 태도에 대한 빅데이터가 다양한 국면에서 수집되고 저장되고 있다(권영진, 정우진, 2019). 이에 따라 산업장면에서 인간의 행동을 예측하기 위해 빅데이터를 분석하고 활용하기 위한 심리학을

사용하려는 움직임이 보인다. 예를 들어 효율적인 인적자원 관리, 확보 그리고 유지를 위해 빅데이터 분석을 활용한 인적자원분석 시스템이 도입되기 시작했다(Jia et al., 2018).

본 연구에서는 빅데이터에 대하여 설명하고 이를 심리학 연구에 활용할 수 있는 가능성과 주의해야 할 점에 대해 논의할 것이다. 특히 심리학 분야 중, 산업 및 조직 심리 장면에서 인적자원관리(Human resource management; HRM)를 하는 방법 중 하나인 인적자원분석(Human resource analytics; HRA)을 중심으로 논의할 것이다. 또한 최근에 국내에서 인적자원을 채용하고 직무에 배치할 때 사용하는 AI 면접을 설명하고, AI 면접의 장점과 단점에 대해 논의할 것이다.

2. 빅데이터란?

컴퓨터 공학, 비즈니스 그리고 심리학 등에서 학문의 경계가 없이 빅데이터를 분석하는 다양한 방법을 다루는 학문분야를 자료과학(data science)이라 일컫는다(김청택, 2019). 이 밖에도 인공지능, 기계학습, 딥러닝 등 많은 개념들이 일반적으로 구분없이 언급되고 있지만 이들은 서로 구분되어야 한다. 자료과학이 데이터를 이용해 분석하는 모든 분석방법에 대한 학문이라면, 인공지능은 그 중에 인간이 가진 지적 능력으로 할 수 있는 사고, 학습 등을 컴퓨터가 인공적으로 대신할 수 있도록 하는 것이다(Wang & Siau, 2019). 기계학습은 인간이 학습을 하는 것과 같이 컴퓨터에게 데이터를 입력시키고 이를 활용하여 스스로 학습하도록 하는 방법에 주안점을 둔다(Wang & Siau, 2019). 최근 빅데이터 분석에서 주목받는 딥러닝(deep learning)은 머신러닝보다 발전된 형태이기 때문에 컴퓨터에 학습할 데이터를 입력시키지 않아도 스스로 학습하여 데이터에

나타나는 패턴을 예측하거나 의사 결정을 한다. 이 모든 과정에서 기초가 되는 빅데이터는 어디에나 존재하기 때문에 빅데이터를 이용하여 연구하고 분석하는 방법들은 심리학에서도 발전되고 있다(Harlow & Oswald, 2016).

빅데이터는 용량(volume), 속도(velocity), 그리고 다양성(variety)의 세 가지 측면으로 정의될 수 있다(Laney, 2001). 용량은 데이터의 규모로서 자료에 포함되는 대상 및 변수의 수를 일컫는다. 어느 정도의 양이어야 빅데이터라는 기준은 없지만 빅데이터는 기존의 데이터보다 더 많은 사람에 대한 자료를 포함하거나 한 시점이 아닌 여러 시점 즉 여러 시간과 공간에서 측정한 자료를 포함하게 된다(김청택, 2019). 속도는 데이터 생성 및 이를 처리하고 활용하는 프로세스에서 요구되는 정도를 말한다. 빅데이터의 마지막 구성요소인 다양성은 정형화된 숫자 데이터 및 텍스트 문서, 오디오, 비디오 및 소셜 미디어 등의 비정형을 포함하여 빅데이터가 취할 수 있는 다양한 형식을 말한다. 이에 따라 심리학에서도 빅데이터 활용을 높이기 위해 사회관계망 서비스를 통해 수집된 비정형 데이터를 이용하여 사용자의 행동과 마음의 패턴을 분석하기도 한다. 그리고 시간의 흐름에 따른 변화를 연구하고 다양한 연구문제를 구체화시키기 위해 종단 자료 분석을 위한 패널 데이터도 늘어나고 있는 추세이다.

이 밖에도 변산성(variability), 시각화(visualization), 가치(value) 그리고 정확성(veracity)등을 빅데이터의 특성에 추가시키기도 한다(김청택, 2019). 그 중 정확성은 빅데이터를 구성하는 요소들의 모호하고 다양한 품질을 처리해야 하는 문제를 의미한다. 즉 측정하고자 하는 대상을 얼마나 잘 측정하고 있는지에 대한 데이터 자체의 타당성이다. 분석에서 활용되는 빅데이터의 정확성이 충분히 보장되지 않는다면 의사결정 과정에서 빅데이터를 사용한 인공지능을 사용하는 것은 오히려 차별적이고 편향된 결과를 만든다(Houser, 2019). 머신러닝에 기초한 인공지능은 인간의 지적

능력을 컴퓨터가 인공적으로 대신하게끔 자료에 기반한 학습을 시키는 것이기 때문에(Wang & Siau, 2019) 정확성이 결여된 자료를 활용하게 되면 인간이 가진 편향(bias) 또는 고정관념(stereotype)이 그대로 학습된다(Houser, 2019). 일부 심리학자들은 빅데이터를 사용하여 도출한 결과에서 인간의 가진 편향을 재현할 위험이 있다는 것(garbage in, garbage out)과 알고리즘 결과에 대한 이유를 이해하지 못해서 편향을 감지할 수조차 없는 블랙박스(black box) 문제를 주요한 문제점으로 언급했다(Crawford, 2013; Houser, 2019). 특히나 산업심리 장면에서 인적관리분석 시 이와 같은 문제가 발생할 경우, 빅데이터가 사회에 현존하는 불평등을 더욱 강화시키고 편향을 확대하는 상황이 발생할 수 있다.

3. 빅데이터 분석

심리학과 같은 사회과학 분야에서 아직 빅데이터 분석이 크게 자리 잡지 못했던 이유는 빅데이터 분석이 모형주도적 접근(model-driven approach)보다 자료주도적 접근(data-driven approach)으로 이루어지기 때문이다(김청택 2019). 즉 빅데이터 분석은 보통 심리학과 같은 사회과학 분야에서 사용하는 가설-지향적 접근과 반대의 방식으로 진행된다. 가설-지향적 연구는 하향식(top-down) 구조화된 접근방식으로 결정을 내리기 위해 가설부터 시작한다. 하지만 빅데이터 분석에서는 일반적으로 탐색적 분석을 하게 되며 상향식(bottom-up) 추론 접근방식을 사용한다(Chen & Wojcik, 2016). 가설-지향적 접근은 주로 모형주도적 분석을 사용하는데 통계학의 기법으로 주로 회귀분석, 분산분석, 요인분석, 구조방정식 등이 있다. 심리학을 포함하는 경험적 사회과학연구에서 많이

쓰이는 방법으로 결과를 예측하기보다는 어떠한 과정에 대해 이해하고 분석하는 것에 더 무게를 둔다. 따라서 사전 연구나 이론에 근거하여 결과나 과정에 대한 가설, 또는 이론적 모형을 먼저 구축하게 된다. 그 후 이론적 모형을 검증할 수 있는 방법을 찾아 실시하고 구성개념에 대한 측정치를 구하게 된다. 그리고 구성개념 간의 관계를 기초로 변수간의 관계를 설명하고 예측하는 방식이다(김청택, 2019). 이런 경우 대부분 연역적 추론을 하게 되는데 이를 지식기반 인공지능이라고도 한다(조성준, 2019).

하지만 인간은 살아가면서 연역적 추론보다는 주로 귀납적 추론을 하게 된다. 귀납적 추론이란 여러 데이터 즉 경험에 의해 새로운 명제 또는 사실을 생성해내는 것이다. 그리고 귀납적 추론 방법을 컴퓨터에게 가르쳐 인공지능을 구현하는 것이 기계학습이다(조성준, 2019). 귀납적 추론은 위에서 언급한 상향식 추론 접근 방식과 유사하며 자료주도적 분석방법을 사용한다. 기계학습 기법의 예로는 신경망, 지지벡터기계, 딥러닝, 그리고 데이터 마이닝 기법 등이 있다. 자료주도적 분석방법은 어떠한 과정에 대한 설명보다는 결과를 예측하고 그에 대한 정확성에 더 중점을 두는 방법이다. 따라서 이론이나 가설 또는 이론적 모형을 세우기 전에 자료를 먼저 수집한 이후에 주어진 자료의 패턴을 분석하여 자료의 특성에 기초한 가설을 설정하고 분석한다(김청택, 2019).

빅데이터 분석 방법에는 크게 텍스트 분석, 지도학습 방법 그리고 비지도 학습방법이 있다. 텍스트 분석은 컴퓨터 공학, 전산 언어학 및 인문학을 포함한 많은 분야에 널리 사용되고 있다(Weiss et al., 2010). 데이터 마이닝 방법은 크게 지도학습과 비지도학습으로 나뉘게 된다. 데이터 마이닝은 유용한 결과를 예측하기 위해 데이터에서 구성요소들 간의 연관성과 패턴을 찾는 것이다. 시청자가 좋아할 만한 새로운 영화를 추천하는 것과 같이

수집된 데이터 범위를 넘어 유용한 정보를 추출하여 예측을 수행하는 모형을 구축한다(Chen & Wojcik, 2016).

지도학습 알고리즘은 데이터에서 일련의 예측변수(feature variables)와 목표변수(target outcome variable)의 관계를 학습한 후, 이들의 연관성을 특징짓는 것을 목표로 한다. 지도학습은 목표변수의 결과를 알고 있을 때, 훈련 데이터 세트(training data set)에 알고리즘을 훈련(모형을 구축)시키는 것이다. 따라서 답을 이미 알고 있기 때문에, 알고리즘은 데이터를 통해서 답에 가장 적합한 모형을 추론할 수 있다. 일반적으로 지도학습의 목표는 아직 알려지지 않은 결과를 예측하기 위해 미래의 데이터에 사용될 모형을 구축하는 것이다. 지도학습의 예로 분류(classification)와 회귀(regression)가 있다. 반면에, 비지도 학습방법은 데이터에서 흥미로운 구조나 패턴을 나타내는 기술모형을 구축하는 것이다. 지도학습과는 달리 데이터 샘플에 결과변수가 표시되어 있지 않고 결과에 대한 정보도 없으며 주어진 목표변수도 존재하지 않는다. 따라서 이는 탐색적 방법이며 결과는 분석을 기반으로 데이터에서 발생하게 된다. 비지도 학습의 예로는 군집화가 있고 연관규칙학습이 있다. 이 밖에도 데이터 분석 방법에는 오디오, 이미지 및 비디오를 포함한 다양한 소스의 광범위한 데이터 형식을 다루는 멀티미디어 분석, 네트워크를 계산한 후 복잡한 관계 패턴을 요약하는 네트워크 분석 등이 있다(Chen & Wojcik, 2016).

4. 심리학 연구에서 빅데이터 적용의 가능성과 주의할 점

빅데이터는 주로 자료주도적 접근이 수행된 컴퓨터 공학 및 비즈니스와 같은 분야에서 전통적으로 활용되어 왔다. 하지만 행동이론과 경험적 연구를 하는 심리학 연구 및 응용 실무 또한 빅데이터 시대에 크게

기여하고 진입하고 있다(Cheung & Jak, 2016). 김청택(2019)은 빅데이터를 심리학 연구에 적용할 수 있는 가능성에 대해 논의하였다. 첫 번째로 심리학 또는 사회과학에서 참여관찰과 같은 질적 연구를 통해 인간의 인지과정과 행동을 빅데이터 분석법을 이용하여 탐색적으로 파악하는 것을 제안했다. 즉 탐색적 연구를 하되 자료주도적 접근이 아닌 기존 사회과학에서 주로 쓰이는 모형 주도적 분석방법으로 이를 검증하자는 것이다. 많은 양의 데이터에 접근하고 분석할 수 있는 컴퓨터 공학의 최첨단 기술과 이해와 예측과 개입이 가장 필요한 영역에 대한 통찰력을 제공하는 사회과학의 이론 및 행동 과학의 통합과 협업을 통해 이점을 얻을 수 있다(Harlow & Oswald, 2016).

두 번째로, 인간의 심리구조나 과정보다 인간 행동의 결과 예측이 더 중요한 심리학 또는 사회과학 영역에 빅데이터 분석기법을 적용시키자는 것이다. 빅데이터 분석 즉 자료주도적 분석방법의 가장 큰 장점이 예측과 정확성이기 때문이다. 심리학 연구자들은 추론의 무의식적 오류가 인간의 판단을 왜곡하는 무의식적 편향(unconscious bias)과 현재 기분, 날씨, 배고픔의 정도와 같이 주제와는 상관없는 요인 때문에 의사결정이 무작위적으로 영향을 받을 수 있는 노이즈(noise)가 사람들이 생각하는 것보다 더 자주 발생한다는 사실을 발견했다(Kahneman & Tversky, 2013). 무의식적 편향은 사람들이 정보를 처리하는데 사용하는 정신적 지름길을 알아차리지 못했을 때 발생한다. 그리고 노이즈는 우연이나 상관없는 요인으로 인한 인간의 의사결정에서의 변동성을 나타낸다(Houser, 2019). 사람들은 자신이 객관적이라고 생각하지만 편향과 노이즈 때문에 부정확하고 일관성 없는 주관적인 결정을 하게 되는 경우가 자주 발생한다. 따라서 알고리즘 자체에 편향과 노이즈가 없는 빅데이터를 분석 할 때, 인간의 의사결정에 영향을 미치는 오류들을 완화시킬 수 있을

것이다(Houser, 2019). 또한 페이스북, 트위터 및 기타 소셜 미디어를 활용함으로써 광범위한 모집단의 태도와 행동에 대한 심리학적 창을 제공하는 대규모 데이터를 이용할 수도 있다(Harlow & Oswald, 2016). 세 번째로, 빅데이터는 실험단계에서의 자극이나 설문지 등의 문항을 개발하는 데에도 도움을 줄 수 있을 것이라고 예상했다.

다만 김청택(2019)은 빅데이터를 사용한 자료주도적 분석이 회의적인 이유 또한 언급하였다. 첫째로, 데이터의 선택, 해당 데이터에서 예측변수의 선별, 분석 방법의 결정에서 주체는 연구자가 되어야 한다고 주장한다. 이에 따라 겉보기에는 이론이 없이 진행되는 것으로 보일지언정 연구 설계 과정에서 연구자 자신만의 이론이 담겨 있다는 것이다. 두 번째로, 주어진 데이터에 기반하여 어떤 현상을 설명하거나 결과를 예측하는 과정에서도 연구자의 이론이 개입된다는 것이다. 빅데이터를 사용하여 데이터가 먼저인 자료주도적인 분석을 한다 해도 빅데이터 자체도 인간이 이루어낸 결과물이고 이를 해석하고 설명하는 것 또한 연구자의 몫이다. 따라서 결과의 해석과정에서 연구자가 이미 가지고 있는 편향과 고정관념을 재현할 위험이 있다(Houser, 2019). 세 번째로, 자료기반 분석을 할 때에는 체계적으로 반복검증을 할 수 없으며 “반증할 수 없는 사실은 과학적 결과로 받아들일 수는 없는 일이다(김청택, 2019; Popper, 1959)” 라고 주장했다. 그럼에도 불구하고, 과학 기술의 변화는 인간의 사고방식과 행동을 변화시키기 때문에 심리학에서도 빅데이터에 대한 이해와 이를 분석하는 방법들이 논의되어야 한다(김청택, 2019). 네 번째로, 공공영역 및 사기업영역에서 얻은 대규모 데이터를 분석할 때 윤리적 고려사항에 대해 세심하게 식별하고 해결해야 한다(Harlow & Oswald, 2016). 가령, 데이터 수집 및 보안, 사용자의 신원 보호, 정보가 어떻게 사용 및 해석될 것인지에 대한 결정 등 빅데이터 프로젝트에 관한 윤리적인 문제를 신중하게 다루는

것이 매우 중요하다는 것이다. 마지막으로, 데이터에서 개발된 모형을 별도의 데이터 세트 또는 hold-out 표본에 적용하여 빅데이터에서 예측 모형을 검증해야 한다(Harlow & Oswald, 2016). 가령, 큰 데이터 세트가 존재하는 경우 이론이나 가설이 아직 형성되지 않았기 때문에, 빅데이터의 초기분석은 탐색적 분석이나 데이터 마이닝 수준에 있는 경우가 많다. 이런 초기 데이터를 일반화하기 위해서는 별도의 데이터에 대한 하나 이상의 후속 분석이 필요하다. 특히 예측과 관련된 많은 변수는 있을 수 있지만 추가 예측 및 계획을 통해 얻을 수 있는 최상의 척도는 아닌 경우가 포함된다.

5. 산업 및 조직 심리학 장면에서 빅데이터의 활용

심리학 분야 중에 산업 및 조직(industrial-organizational) 심리학자는 조직심리, 인사심리, 공학 심리, 직업심리 등의 분야에서 직무스트레스, 작업 수행, 리더십, 직무태도, 조직 문화 및 풍토, 팀, 인력 관리와 같은 다양한 주제를 다루는 연구들을 수행한다(유태용 등, 2018). 이에 따라서 산업 및 조직 심리학자는 연구 및 실습 과정에서 데이터를 생성하기도 하고 다른 데이터 세트와 결합하여 더 거대한 데이터를 생성하기도 한다. 실제로 박정아와 임혜빈(2018)은 기업들이 빅데이터를 통하여 소비자의 심리를 이해하고 파악함으로써 효율적인 마케팅에 활용하는 방안에 대하여 관심을 가지고 있으며, 빅데이터를 활용하여 광고 또는 마케팅에 활용되는 데이터 분석 방법과 사례에 대해 소개하였다. 정혜정과 오경화(2016)은 소셜 빅데이터 분석을 사용하여 6년간 SNS 상에서 언급된 윤리소비 연관어들을 바탕으로 윤리소비유형, 동기, 감정의 변화양상을 시계열적으로 살펴보았다.

산업 및 조직심리에서 인적자원관리(Human resource management;

HRM)란 인적자원 정책 및 기업의 해당 관리 활동을 의미한다(Jia et al., 2018). 관리 활동에는 주로 기업 인사 전략 수립, 직원 채용 및 선발, 교육 및 개발, 성과 관리, 보상 관리, 직원 이동성 관리, 직원 관계 관리, 안전 및 건강 관리가 포함된다(Noe et al., 2006). Jia et al. (2018)은 HRM이 여섯개의 차원으로 구성되며 이들은 서로 상호작용하여 효과적인 인적자원을 형성한다고 말한다.

HRM 과정에서 인공지능 기술을 사용하면 더 큰 경제적 이익을 얻을 수 있고 HRM의 효율을 높이는 것이 HRM 발전에 있어 중요한 트렌드가 되고 있다(Jia et al.. 2018). 인공지능 기술은 직원을 관리하고 회사가 인적자원을 합리적으로 할당하는 것을 도와주며, 모든 과정에서의 고정관념을 해결하고 관리작업을 수행하는 보조자 역할을 할 수 있다. 이에 따라 HRM 활동 중에 산업장면에서 조직 구성원과 관련된 의사결정에 빅데이터를 활용하는 인적자원분석(Human resource analytics; HRA)이 주목을 받기 시작했다. HRA는 통계적인 기법을 사용하여 데이터를 분석한 후 시사점을 도출하고 인적자원 관행을 조직의 성과와 연결하여 인재경영 관련 의사결정에 기여하고자 하는 활동이다(김성준, 2013; 원지현, 2014).

HRM 과정 중 채용과 직무배치의 의사결정 과정에서 HRA를 통하여 인공지능을 사용하는 방법은 크게 AI에 의한 자소서 평가와 AI 면접으로 나눌 수 있다. 즉, 서류전형 전 단계에서 활용되기도 하고 서류전형에서 면접으로 가는 단계에서 활용되기도 한다(이원갑, 2019)[1]. 인공지능이 자기소개서를 평가하게 되면 인사담당자보다 빠르게 평가하게 되어 효율성과 비용 절감을 누릴 수 있게 된다(최현주, 2018)[2]. 기업에서 원하는

1) 서류전형의 전 단계, 혹은 서류전형에서 면접으로 넘어가는 단계에서 활용되기 시작했다. (이원갑, 2020.09.20.검색).

2) 인공지능이 자기소개서를 평가하는데 걸리는 시간은 평균 3초다. 1만 명의 자기소개서를 평가하는 데 8시간이 걸린다.(최현주, 2020.09.20.검색).

면접 질문과 여태까지 쌓여온 인적자원 데이터를 인공지능에 학습시키고 이로 인해 기업이 선호하는 인재상을 채용할 수 있도록 하는 것이다. AI를 활용한 서류전형은 2000년대 초부터 해외에서 사용되고 있으며, 최근에는 미국 경제지 포춘이 선정한 50대 기업의 대부분도 AI를 채용 과정에서 사용하고 있다(마이다스아이티, 2018)[3]. 국내에서도 2010년대에 들어서 신입사원의 서류 전형에 AI를 도입하여 활용하고 있다(최현주, 2018). AI 면접이란 지원자를 면접관이 직접 만나는 대면면접 대신 마이크와 웹캠을 사용하여 영상을 AI가 직접 분석하여 면접 지원자를 평가하는 시스템이다. 지원자의 표정, 목소리, 제스처 등을 포착하여 감정과 상태를 분석하고 성격에 맞는 부서를 파악하거나 돌발상황에 대한 대처능력 등을 파악한다. AI 면접이 진행되는 동안 파악한 지원자의 목소리, 얼굴의 근육 움직임, 뇌파 등을 분석하여 적합한 업무에 효율적으로 지원자를 배치할 수 있도록 한다(최현주, 2018). AI 면접을 통해 서류검사 또는 인적성 검사를 시행하면 기존의 인적성 검사보다 문제 출제의 의도를 파악하기 힘들고 정답이 있지 않으므로 면접관에게 잘 보이기 위한 선택이나 사전학습이 필요하지 않아 사회적 바람직성이 나타나는 것을 방지할 수 있다. 따라서 이를 효율적으로 잘 활용한다면 기업은 인사관리와 채용과정에서 편견을 줄이고 효율성과 공정성을 얻을 수 있다.

6. AI 면접의 시사점

하지만 채용과 직무배치 과정에서 국내의 많은 대기업들이 인사관리와 채용과정에서 AI를 도입하고 있는 것과 반대로 일반인의 절반 이상은 AI 면접에 대해 부정적인 것으로 나타났다(이광석, 2019). 또한 전문가들은

3) 50대 기업의 대부분도 AI를 채용 과정에서 사용(마이다스아이티, 2020.09.20.검색).

인공지능이 사회에 현존하는 편향도 그대로 학습하게 된다는 것에 부정적이다. 실제로 2014년에 아마존은 자체 개발한 인공지능 채용 시스템이 자기소개서 또는 이력서 평가에서 여성 지원자에게 불이익을 준다는 것이 알려졌다. 10년간 학습한 데이터에서 여자보다 남자 지원자가 많은 것을 통해 판단한 결과이다. 시스템의 알고리즘 자체에 편향이 생겨 잘못된 결과를 도출하는 것도 큰 문제지만 인공지능에 의해 도출된 결과는 통계적 검증과 합리적인 계산에 의한 결과라는 사실로 인하여 기존의 차별이나 편향보다 더 정당하게 받아들여질 가능성이 있다.

산업 장면에서도 기업이 수집한 데이터를 AI에게 학습시키는 경우, AI는 주어진 데이터에만 기반하여 지원자를 판단하고 채용여부를 결정한다. 이는 기업 입장에서는 현재 그대로의 상황을 유지할 뿐 향후 성장 가능성에 대한 기회와 고려가 없다는 부작용이 있을 수 있다(최현주, 2018). 즉, 기업 또는 인간이 가지고 있는 편향과 고정관념을 AI가 그대로 학습하여 이를 기반으로 인적자원을 평가하고 관리하게 된다(Houser, 2019). 빅데이터 분석은 귀납적 추론을 하기 때문에 결과를 예측하고 그 결과에 대한 정확성에 초점을 둔다. 따라서 인적자원 즉 사람에 대해 결과적으로 피상적인 것만을 도출해 낼 위험이 있다. 그러면 채용과 직무배치 의사결정에서 어떠한 과정과 실제적인 원인, 또는 변수들의 관계는 보지 못하게 되는 것이다. 이렇듯 빅데이터 세상은 효율성과 편의성을 제공하지만, 수많은 도전과 잠재적인 위험을 제시하기도 한다. 따라서 빅데이터의 출현과 함께 파생된 문제나 현상에 대한 인식을 높이고 이에 대한 해결 방안을 제시하는 것이 중요하다. 또한 빅데이터 세계가 계속해서 발전하고 조직 및 산업심리학자의 경험과 영향력이 커짐에 따라 제시되는 권장사항도 함께 발전해야 한다(Guzzo et al., 2015).

7. 결론 및 논의

빅데이터를 활용하는데 발생할 수 있는 위험은 데이터의 종류보다는 해당 데이터에 대한 알고리즘 또는 머신러닝의 적용에서 비롯될 가능성이 높다. 전통적인 통계적 방법(가령 t검정)은 이러한 위험이 없지만 알고리즘 및 머신러닝 시스템의 구현은 특히 데이터의 복잡도가 높을 때 예상치 못한 방식으로 피해를 받을 수 있다. 알고리즘이 복잡할수록 예측하기 어렵고 바람직하지 않은 결과가 도출될 수 있다. 가령, 산업 및 조직심리 장면에서도 알고리즘은 편향이 시스템에 기록되어 평가 다양성이 부족해질 수 있다. 회사의 인력이 과도하게 동질적이거나 일부 지원자 군집에는 고용 기회가 전혀 제공되지 않을 수 있기 때문이다.

두 번째로는 빅데이터는 표집과 관련한 한계를 가질 수 있다. 표본의 크기가 크다고 해서 대표하는 표본과 완벽하게 똑같은 것은 아니다. 데이터 세트가 커도 비무작위 표본과 관련된 문제들은 발생할 수 있으며 이에 따른 설명 및 추론의 잠재적 편견이 발생할 수 있다. 통계적 추론을 할 때는 모수, 즉 모집단을 알 수 없기 때문에 표본을 이용하여 모집단을 추론한다. 이때 확률 표집을 통해 모든 관측치가 동일한 확률로 추출되게 하는 무작위 표본을 사용해야 추론의 정확성과 일반화 가능성이 높이지게 된다. 하지만 비무작위 표본을 무작위 표본으로 잘못 인식하고 사용하게 되면 선택 편향이라는 오류가 발생하게 된다. 선택편향은 통계 분석을 왜곡시키는 오류 중 하나로 통계적 유의성을 추정할 때 편향이 발생될 수 있다(성태제, 2014). 선택 편향이 일어나게 되면 대표성을 확보하기 힘들기 때문에 데이터 분석 결과를 세상에 일반화하여 적용하기 어려워진다(성태제, 2014). 따라서 빅데이터를 적절하게 사용하려면 표집 방법을 정확하게 제시해야 하며 가지고 있는 표본의 한계에 대한 논의를 제공해야 한다.

세 번째는 빅데이터에서 간과되고 있지만 여전히 중요한 측면은 빅데이터를 구성하고 있는 측정치들의 품질에 관한 것이다. 기존의 설문조사와 달리 첨단 장치를 이용해 얻은 데이터라도 측정치의 신뢰도와 타당도는 입증되어야 한다. 연구자가 측정치의 신뢰도를 파악하여야 데이터에서 잘못된 추론을 할 가능성을 낮출 수 있기 때문이다. 데이터가 여러 출처에서 얻어지고 정교한 도구로 측정되어도 반드시 “좋은” 데이터는 아니다. 따라서 기존처럼 측정치의 품질에 대한 우려는 계속되어야 하며 이를 뒷받침하는 증거도 제공되어야 한다.

네 번째로 측정치의 품질과 관련된 문제인 데이터 자체의 정확성이다. 정확성은 통계적 개념에서 타당도와 유사한 개념으로써 연구결과가 얼마나 타당한지, 즉 측정하고자 하는 것을 정확하게 측정했는지를 나타내는 지표이다(성태제, 2014). 만약 조작적 정의가 제대로 되지 못함으로써 측정하고자 하는 것이 아닌 다른 것을 측정하고 있다면 연구자가 기대하는 결과와 전혀 다른 결과가 도출될 수 있다. 가령 기업에서 직원만족도를 측정하고자 하는데 연봉에 대한 만족도만을 직원만족도에 포함하고 직원복지에 대한 만족도나 직무에 대한 만족도 등을 누락하는 경우 기업이 애초에 원한 직원만족도를 정확하게 측정할 수 없게 되는 것이다. 데이터의 용량, 속도, 그리고 다양성에 따라 데이터의 정확성을 확인하기 어려울 수 있기 때문에, 빅데이터를 적절하게 사용하기 위해서는 데이터의 정확성을 평가하고 의심스러운 데이터는 식별 및 수정(또는 폐기)해야 한다.

마지막으로 데이터에 고품질의 과학적 접근법을 사용하는 것은 중요하다. 과학적 지식이 객관적이라고 할 수 있는 이유는 관찰 또는 실험을 통해 연구자가 세운 가설이 검증되거나 기각되었기 때문이다. 과학철학자 Popper(1959)가 말했다시피 반증가능성(falsifiability)이 있어야만 주어진 이론 또는 가설이 옳다고 말할 수 있는 것이다. 따라서 데이터를 기반으로

하는 주장은 과학적인 방법에 의해 뒷받침 되어야 한다. 표본 크기가 커지게 되면 기존의 통계적 유의성 검정으로 분석하는 것은 적합하지 않을 수 있기 때문에 특정 결과의 의미를 모호하게 만들 수 있다. 또한 변수의 수가 많아지면 통계적 검정의 순열 또는 모형 생성이 끝없이 많아질 수 있다. 이렇게 끝없는 가능성이 주어지게 되면 결과를 해석할 때 신중해야 하며 허위 상관의 가능성을 인정해야 한다. 또한 통계적 발견이 연구 및 실무에 유용하다는 결론을 내리기 전에 반복 및 교차검증을 수행해 보아야 한다. 연구자가 세운 가설이 지지되는 연구결과를 얻었다 하더라도 다른 연구자들에 의해 다른 상황과 연구방법으로도 같은 결과를 얻는 반복검증이 이루어져야만 그 결과가 일반화될 수 있기 때문이다. 수많은 대체변수, 모형 또는 분석이 동일하게 잘 수행되어야 한다는 점을 고려해야 한다. 또한 동일하게 유효한 변수, 모형 또는 분석에서 연구자들은 자신이 선호하는 특정 결과만을 선택하는 경향을 피하고 조사된 합리적인 대안의 전체범위를 제시해야 한다.

2014년에 최초의 다양성 보고서가 기술 산업에서 여성의 부족현상을 언급한 이후부터 기업들은 직원들에게 무의식적 편견 교육을 제공하기 위해 컨설턴트를 고용하기 시작했다. 하지만 안타깝게도 최근의 보고서에서는 이러한 현상이 크게 개선되지 않았다. Human Capital Institute의 설문조사에 따르면 거의 80%의 리더가 인재 관리에 영향을 미치는 결정을 내리기 위해 여전히 직감과 개인적인 의견을 사용하고 있다고 밝혔다(Houser, 2019). 이에 대한 해결방안으로 AI를 채용 결정과정에 통합함으로써 인간의 의사결정에서 발생하는 무의식적 편견과 변산성(또는 노이즈)을 완화시킬 수 있다. 기존의 연구들은 AI를 사용해서 얻은 이득에 관한 연구가 대부분이었으나 본 연구에서는 AI 알고리즘 결과에서 인간이 가진 무의식적 편견을 재현할 위험이 있다는 것(garbage in, garbage out)과

알고리즘 결과에 대한 이유를 이해하지 못해서 재현된 편견을 찾아낼 수도 없는 블랙박스 문제에 대한 우려가 존재한다는 이론적 시사점을 갖는다. 실제로 2020년 8월 13일 코로나 19사태로 인해 시험을 치르지 못한 영국 고등학교 학생들이 인공지능이 부여한 학점에 대해 항의하는 사태가 일어나기도 했다(Reuters, 2020). 인공지능이 기존 데이터가 가지고 있는 편향을 확대하여 나타난 현상으로 인공지능이 우리 사회에 존재하는 불평등을 강화할 것이라는 학자들의 우려가 현실로 드러나는 사례였다. 물론 빅데이터를 이용한 인공지능은 효율성, 편의성, 공정성, 경제적 이익성 등 많은 이점을 가져다 줄 수 있다. 따라서 편향이 섞인 결과에 대한 두려움 때문에 AI를 의사결정 과정에 통합시키는 것을 지연시키는 것은 합리적인 해결책이 아니다. AI를 사용하여 도출된 결과에서 발견된 잠재적인 편향은 AI 사용 자체가 문제가 아니라 데이터에 포함된 인간의 편향에서 기인하는 것이기 때문이다. 그러므로 4차 산업혁명 시대에서 빅데이터를 기반으로 하는 인공지능 또는 기계학습의 혜택을 받음과 동시에 데이터의 정확성을 평가하고 결과물에 대한 당위성에 대하여 고민해보아야 한다.

IV. 연구 2

“기계학습의 활용에서 발생하는 편향과 차별: 학습자료의 활용”

1. 서론

우리나라는 4차 산업혁명이라는 새로운 시대에 진입하고 있다. 4차 산업혁명은 인공지능이 활용됨으로써 산업 환경에서 자동화와 연결성이 극대화되고, 이를 활용하여 산업 및 의사결정의 효율성을 높이고자 한다. 예를 들어, 사물인터넷과 자율주행 자동차처럼 4차 산업혁명에서 주목받고 있는 기술의 기초에는 빅데이터 처리 모형의 확대와 인공지능의 진보가 뒷받침하고 있다. 특히, 인공지능은 2016년 이세돌과 알파고와의 바둑 대국이 이루어진 이후에 일반 대중들에게 4차 혁명의 대명사로 취급받고 있다(김지연, 2017).

4차 산업혁명의 출현과 함께 인공지능은 사회의 다양한 분야로 확대되고 있다. 이러한 추세는 인문사회과학 중에서도 경영학을 비롯한 산업심리학 분야에서 주요하게 묵도되고 있다(구소희 등, 2020). 특히, 빅데이터 및 인공지능에 기초한 의사결정은 HR의 다양한 분야에서 활용될 수 있다. 예를 들어, 기계 학습에 기초한 인공지능이 선발 장면(예, 서류검사, 인적성 검사 및 면접 등)에서 주목을 받기 시작한 이후에 2021년 현재 다양한 기업체에서 활발하게 활용되고 있다(최현주, 2021). 보다 구체적으로, 기계학습에 기초한 인공지능은 과거 지원자들에게서 수집된 자기 소개서의 정보에 기초하여 새롭게 기업에 지원한 사람들의 자기소개서에 기술된 지원자의 잠재성을 평가한다(주영재, 2021). 또한, 기계학습에 기초한 인공지능은 다양한 안면인식

기술을 활용하여 지원자의 특성에 맞는 면접 질문을 생성하고 답변을 평가하여 직무적합성에 대한 정보를 수집하고 입사결정에 대한 정보로 활용된다(형인우, 2021).

기계학습에 기초한 인공지능의 발전은 사회 및 경제에서 국가 생산성과 개인 삶의 질의 향상이라는 긍정적인 효과를 불러왔다. 하지만 이와 함께, 인공지능의 출현이 노동 시장에 대변혁을 가져옴으로써 고용 시장 구조의 변화를 초래하였고, 일부 채용 분야에서는 고용의 대폭적인 축소라는 부정적인 결과를 가져왔다(나준호, 2006). 예를 들어, 복합적인 업무를 수행하는 전문직 직무의 경우, 많은 양의 정보와 복잡한 직무, 정량적 분석의 중시, 업무 속도 증가와 같은 이유로 사람들은 한계에 봉착하고 있다. 따라서 이러한 직무들은 인공지능으로의 대체 및 인공지능과의 협업의 가능성이 증대하고 있다. 인공지능은 복잡한 논리적, 단계적 규칙에 따라 계속적으로 객관적 데이터를 분석, 판단, 실행한다(신현석, 정용주, 2017). 단계적 분석을 통하여 직무를 수행하는 과정에서 만나게 되는 다양한 문제들은 과거에 유사한 문제의 해결에 사용되며 축적된 대용량 데이터 분석에 기초한 의사결정을 통해 해결된다. 따라서 직무 수행 과정에서 겪게 되는 다양한 문제에 대한 해결책의 질은 축적된 데이터의 함수로 귀결된다. 즉, 축적된 데이터에 편향된(biased) 자료가 포함되면 문제에 대한 타당한 해결책을 제시하지 못하고 업무 수행과 관련된 의사결정의 수준이 떨어지게 된다.

인공지능 로봇이나 알고리즘의 등장으로 인하여 인간은 실제 인간보다는 기계와 빈번히 접촉하고 교감하게 될 것이며, 이에 따라 외부와 단절한 상태에서 오로지 기계와 소통함으로써 발생하는 인간관계의 퇴화나 인간성 상실이 우려된다. 이러한 문제보다 더 심각하고 현재화된 문제는 편향된 자료의 활용으로 인하여 발생한 의사결정의 편향성, 차별성의 폐해다. 그동안 실전에 배치된 기계학습에 기초한 인공지능의 알고리즘이 편향된 데이터를

학습한 결과 편향성을 보이고, 결과가 인종이나 성별에서 차별성을 보인다는 여러 사례가 나오면서 기계학습의 편향성이나 차별성의 문제는 사회적으로 큰 반향을 일으켰다.

2. 이론적 배경

현재 인공지능 활용의 기초에는 기계학습 및 딥러닝이 주를 이룬다. 지도학습에 속하는 기계학습에 기초한 인공지능의 핵심은 주어진 데이터를 입력값과 출력값 사이의 관계에 대하여 상관 분석 실시하고, 상관 관계에 기초한 패턴을 추정하는 학습 기능이다. 학습된 모형을 활용하여 새로운 자료에 대한 분석과 예측 기능을 수행한다. 기계학습의 분석과 예측 기능은 자율학습(self-learning) 개념에 기초한다. 자율학습은 기계학습을 활용하는 인공지능이 학습에 사용될 데이터와 경험정보를 활용하여 통계 모델링을 개발하고, 이를 적용하여 패턴을 발견하며 새롭게 입력된 데이터를 활용하여 모델링의 성능을 스스로 개선하는 것을 일컫는다. 그림 2와 같이 자율학습에 기초한 기계학습에서는 명령의 입력보다는 데이터의 입력에 기초하여 모형과 결과(예, 예측 및 분류 등)가 추정된다.

<그림 2> 데이터 입력에 기초한 자율학습 모형

기계학습은 크게 지도학습(supervised learning), 비지도학습(unsupervised learning), 강화학습(reinforcement learning)으로 구분된다. 지도학습에서는

입력변수와 출력변수로 기분류된 자료를 제공하여 이들 간의 상관관계에 기초하여 모형을 추정하고 학습한다. 지도학습에 적합한 알고리즘의 예로는 회귀분석, 의사결정나무, 신경망, SVM(Support Vector Machine) 등이 있다. 지도학습과 달리, 비지도학습에서는 입력변수들이 사전 분류 없이 모형에 투입된 입력변수들 간의 상관을 추정하고 분석함으로써 모형을 학습한다. 비지도학습에 적합한 알고리즘의 예로는 군집분석, 연관성분석, 네트워크분석, 차원 축소 등이 있다. 강화학습은 알고리즘을 활용하여 예측 모형을 추정하기 위하여 임의의 시행에서 발생하는 오차에 대한 피드백이 사용된다. 이전 시행에서 발생한 피드백에 기초하여 추정된 모형을 수정하는 과정을 반복적으로 수행하고, 오차가 일정 값으로 수렴하는 단계에서 최종모형이 선택되고 학습된다.

기계학습이 활용되는 인공지능에서는 인공지능이 학습하게 될 입력자료의 중요성이 간과될 수 없다. 기계학습을 위하여 입력된 자료에서 모형을 추정하기 위한 정보가 충분히 포함되지 않거나, 편향된 정보로 추정된 모형은 부적합하거나 편향된 결과를 만들어 낸다. 특히, 기계학습에 사용되는 자료에 개인화된 검색의 결과로 인하여 확증편향이 일어나는 필터 버블(filter bubble)과 자신과 유사한 생각을 가진 사람들과만 소통하는 에코 챔버(Echo Chamber) 효과가 반영될 수 있다(Bozdag, 2013; Pariser, 2011). 학습자료에 특정한 성향을 가진 구성원들만의 반응이 포함될 경우에는 불완전한 모형이 추정될 수 있다. 덧붙여서, 기계학습에 기반한 인공지능이 활용하는 모형의 불완전성(artificial stupidity)은 기계학습을 위해서 인공지능에 입력되는 자료가 앞으로 예측이나 분류를 위해 활용될 수 있는 모형에 적용될 수 있는 모든 방면에 대한 정보를 포함하여 불완전한 정보에 기초하여 추정된 모형을 의미한다. 즉, 입력된 학습자료가 포함하지 않은 현상을 예측하는 경우에는 타당하지 못한 결과를 만들게 된다.

Leslie(2019)는 인공지능이 유발할 수 있는 위험성에 대한 여섯 개의 시나리오를 제시하였다. 본 연구에서는 인공지능으로 인한 발생하는 위험들 중 편향과 차별에 해당하는 네 가지 시나리오에 대하여 구체적으로 살펴보고자 한다. 기계학습에 기초한 인공지능이 편향적인 자료에 기초한 학습으로 인하여 역기능적으로 활용될 수 있다. 예를 들어, Google의 온라인 광고시스템은 고소득 직업에 대한 구인광고를 여성보다 남성에게 훨씬 자주 노출시켰다(Datta et al., 2015). 또한, 흑인을 연상하게 하는 이름을 검색할 때 형사 사건과 관련된 종류의 광고를 게재할 가능성이 25% 더 높았다(Sweeney, 2013).

기계학습에 기초한 인공지능이 편향된(biased) 결과를 만들어 내는 근거는 다음과 같다(한국지능정보사회진흥원, 2018). 첫째, 편향된 데이터셋을 학습함으로써 편향이 발생한다. 기계학습 기반 인공지능 알고리즘이 현실 세계의 편향성이 반영된 데이터셋을 학습하게 됨으로써 편향된 결과가 발생하는 것이다. 특히, 현실이 반영된 학습데이터에 의한 편향성은 인식하기도 쉽지 않고 학습데이터의 특정 계층에 불리한 결과를 제시할 수 있다. 예를 들어, 2016년 3월 MS사가 선보인 딥러닝 기반 챗봇 '테이'에게 일부 이용자들이 인종차별, 성차별 발언과 욕설을 학습하도록 유도하여, 챗봇이 더 이상 활용되지 않았다.

둘째, 학습자료를 디지털로 가공하는 과정에서 편향이 발생할 수 있다. 원자료(raw data)를 알고리즘이 학습할 수 있는 형태인 디지털로 가공하는 과정에서 편향된 결과가 발생한다. 또한, 원자료에 이전에는 문제가 되지 않았지만, 이미 존재하는 편향성을 학습하고 결과를 도출하게 되면 과거 편향성을 그대로 노출될 수 있다. 예를 들어, 영국의 Saint George 병원은 의사 모집 과정을 관리하기 위해 자동 지원자 평가 시스템을 구현하였다. 새로운 시스템은 더 빠르고 사람이 개입할 여지가 없었기 때문에 더 객관적일

것이라고 생각되었으나 이 시스템은 여성과 비유럽식 이름을 가진 사람들을 차별하는 것으로 밝혀졌다(Lowry & Macpherson, 1988). 새로운 시스템의 개발자는 과거의 모집 담당자가 의사 결정을 내린 데이터를 사용하였으나, 데이터에 이미 편견이 포함되어 있었던 것이다. 따라서 과거에 업적 자격과 무관하게 여성은 전반적으로 고려 대상이 되지 않았던 점이 알고리즘에 학습되어 편향을 유발하였다.

셋째, 모델을 통해 편향성이 확대될 수 있다. 즉, 입력자료와 무관하게 기계학습에 기초한 인공지능의 블랙박스인 알고리즘 모델은 편향성을 일으키는 결과를 발생시킬 수 있다. 이를 해결하기 위해 모델에서 편향성을 일으키는 정보를 제거하는 방법을 고려할 수 있으나 알고리즘이 무용지물이 되는 결과를 초래할 수 있다. 예를 들어, 버클리 대학 연구팀은 기계학습 기술을 사용하여 늑대들이 찍힌 사진들과 허스키들이 찍힌 사진들을 구분하도록 시스템을 훈련시켰다. 인공지능이 사진 속의 동물이 늑대라고 판단할 때, 인공지능은 실제로 늑대들 사진에 있는 주변 환경인 눈(snow)에 주의를 기울이고 있음을 발견하였다(Ribeiro et al., 2016). 학습에 사용된 늑대의 모든 사진에는 눈이 있었기 때문에 허스키 사진에 눈이 있으면 그것이 늑대라고 생각하는 잘못된 상관관계를 학습하였다. 즉, 인공지능이 실제로 무엇에 주목하고 있는가를 알아야 어디에 실수가 일어나고 있는지를 분명히 알 수 있다.

넷째, 기계학습에 기초한 인공지능이 유발한 편향성이 사람들에게 영향을 미친다. 예를 들어, 트위터, 페이스북 등에 내재된 알고리즘의 편향성 양극화가 사람들의 생각과 행동에 영향을 주어 현실에서 증폭되는 현상이 발생한다. 따라서 무분별한 알고리즘의 사용은 사회에서 지속되는 차별의 기존 패턴을 더욱 악화시킬 수 있다. 미국과 인도에서 미결정 유권자에 대한 일련의 연구 결과에 따르면 검색 엔진 결과는 투표 결과를 20% 정도 변경할 수 있었던

것으로 밝혀졌다(Epstein et al., 2017)

위에서 기술된 네 가지 시나리오는 기본적으로 학습자료에 존재하는 편향이 기계학습을 통해서 추정된 알고리즘에 반영됨으로써 발생하는 현상에 기초한다. 즉, 기계학습 알고리즘의 추정에 활용되는 학습자료와 실제 연구문제를 해결하기 위해서 사용되는 실제자료가 대표하는 모집단 간의 동질성이 확보되지 않음으로써 기계학습 알고리즘을 활용함으로써 얻을 수 있을 것으로 예상되는 효과를 보이지 못할 수 있다. 이에 덧붙여 보다 심각한 문제는 학습자료에서 결과변수를 예측함에 있어서 효과적이기는 하지만 윤리적 혹은 법률적인 문제로 인하여 포함되어서는 안 되는 변수가 포함되고 이에 기초하여 기계학습 알고리즘을 추정하고 실제자료에서 활용함으로써 기존에 존재하는 편향이 확대될 수 있는 가능성이 높아짐을 보여준다.

따라서 본 연구에서는 편향된 학습자료를 사용함으로써 발생할 수 있는 예측 및 분류의 오류의 수준을 검증하였다. 특히, 편향된 자료가 발생하는 상황을 구체화하고 이러한 상황에서 발생하는 편향된 알고리즘의 편향의 정도를 실증적으로 확인하였다. 편향이 발생할 수 있는 상황을 모집단의 동질성이 확인되지 못하는 경우와 편향된 구성원의 속성이 포함되는 경우로 구분하였다.

구체적으로, 모집단을 대표하지 못하는 집단에서 수집된 학습자료에서 추정됨으로써 편향된 알고리즘(예, 중다회귀분석)을 실제 자료에 적용하였다. 이를 통하여 인공지능을 활용하여 얻을 수 있는 것으로 기대되는 예측이 어느 정도의 오류를 포함하는지를 검증하였다. 이에 덧붙여서, 학습자료에서 준거변수와 부적인 상관을 가지고 부적인 영향을 미치는 편향된 변수가 포함된 경우에 추정되는 편향된 알고리즘(예, 로지스틱 회귀분석)에서 발생할 수 있는 편향의 정도 및 효과를 검증하였다. 이러한 편향의 정도를 검증하기 위하여 학습자료에서 추정된 알고리즘에 포함된 편향된 변수의 영향으로 실제

자료에서 발생하는 분류 오류를 살펴보았다.

3. 모의실험 방법 및 결과

3.1. 모의실험 1

3.1.1 연구 방법

첫 번째 모의실험에서는 알고리즘의 추정을 위해서 활용된 학습자료와 추정된 알고리즘을 활용하여 예측에 사용되는 실제 자료의 동질성이 보장되지 못하는 상황을 가정하였다. 즉, 학습자료에서 추정된 알고리즘을 실제 자료에 적용하는 과정에서 두 자료 사이의 동질성이 확보되지 못할 때 실제 자료에 대한 추정된 알고리즘의 설명력과 예측 오차의 정도를 검증하였다. 첫 번째 모의실험에서는 지도 학습에 속하며 결과변수의 예측에 활용될 수 있는 회귀분석이 사용되었다. 회귀모형에서는 학습자료에서 포함된 예측변수를 활용하여 결과변수를 예측하는 알고리즘이 추정된다.

이러한 상황 및 연구 문제를 검증하기 위하여, 2개의 모의 자료가 생성되었다. 다만, 2개의 자료에서 가정하는 준거변수와 예측변수 간의 관계가 서로 다르기 때문에, 모의 자료 간의 동질성이 가정되지 않았다. 즉, 학습자료와 실제 자료에서 사례 수가 1,000명인 자료가 5개의 예측변수와 1개의 준거변수의 측정을 가짐을 가정하였다[4]. 다만, 5개의 예측변수와 1개의 준거변수 간의 관계가 2개의 모의 자료에서 동일하지 않았다. 보다

4) 편향의 효과를 검증하는 모형의 간명성을 위하여 모의자료를 생성하는 모형을 간소화하였다. 이를 위하여 예측변수의 수를 5개로 제한하였다. 특히, 본 연구에서는 학습자료와 실제자료가 대표하는 모집단의 차이로 인하여 발생하는 기계학습 추정 알고리즘의 편향 효과를 검증하는 것이기 때문에, 학습 모형의 복잡성은 고려의 대상으로 삼지 않고 간명한 모형을 설정하였다.

구체적으로, 학습자료의 생성에는 식 (1)의 회귀식이 활용되는 반면에 실제 자료의 생성에는 식 (2)의 회귀식이 사용되었다. 2개의 회귀식에서 오차항은 N(0, 1)인 표준정규분포를 따른다고 가정하였다. 2개의 회귀식에 기초하여 자료를 생성한 이후에 학습자료를 활용하여 알고리즘(즉, 중다회귀방정식)을 추정하였다. 마지막 단계에서는, 학습자료에서 추정된 알고리즘을 실제 자료에 적용하여 설명력 과 추정의 오차를 검증하였다.[5)]

$$z_y = 0.5z_{x1} - 0.3z_{x2} + 0.2z_{x3} + 0.1z_{x4} - 0.1z_{x5} \qquad (1)$$

$$z_y = 0.5z_{x1} - 0.3z_{x2} + 0.2z_{x3} + 0.5z_{x4} - 0.5z_{x5} \qquad (2)$$

3.1.2 결과

학습자료와 실제 자료의 동질성이 유지되지 못함으로 인하여 발생하는 편향을 검증하기 위한 첫 번째 모의실험의 결과는 타당하지 않은 학습자료의 사용으로 인하여 인공지능 알고리즘이 기대되는 예측력을 충분히 보여주지 못함을 확인하였다. 구체적으로, 모의 자료로 생성된 학습자료에서 추정된 인공지능 알고리즘은 식 (3)과 같은 중다회귀식을 추정하였다.

$$\hat{z}_y = 0.51z_{x1} - 0.29z_{x2} + 0.21z_{x3} + 0.08z_{x4} - 0.11z_{x5} \qquad (3)$$

식 (3)의 추정된 중다회귀식을 활용하여 실제 자료에서의 준거변수의 예측값을 구하고, 실제 자료의 관찰값과 예측값의 차이에 기초하여 R^2을

5) 학습자료와 실제자료 간의 모집단의 동질성이 유지되지 않는 상황을 가정하기 위해서 모의자료 생성에 사용된 회귀분석의 식 (1)과 (2)에서 z4와 z5의 회귀계수가 다르게 설정되었다. 즉, 서로 다른 회귀계수를 설정함으로써 학습자료와 실제자료 간의 모집단의 동질성이 유지되지 않고 있음을 보였다.

구하였다. 학습자료와 실제 자료간의 동질성이 확보된 경우에는 식 (1)에 기초하여 기대되는 설명량은 .40이다. 이에 반하여, 모의실험 1에서 추정된 설명량은 .12이었다. 즉, 학습자료와 실제자료 간의 동질성이 확보되지 않음으로써 설명력이 70% 감소하였다.

3.2. 모의실험 2

3.2.1 연구 방법

두 번째 모의실험에서는 알고리즘의 추정을 위해서 활용된 학습자료에 포함된 변수에 편향이 있는 경우를 가정하였다. 학습자료에는 준거변수와 관련이 있는 예측변수가 활용되어 있으나, 실제 자료에서는 해당 변수가 준거변수와의 관련성을 가지지 못하여 분류의 정확성을 낮추는 상황을 가정하였다. 즉, 학습자료에는 준거변수에 편향된 영향을 미치는 예측변수가 포함되어, 학습자료에 기초하여 추정된 알고리즘이 편향된 변수의 영향이 없는 실제 자료의 분류에 미치는 영향을 검증하였다.

이러한 상황 및 연구 문제를 검증하기 위하여, 첫 번째 모의실험과 동일하게 2개의 모의자료를 생성하였다. 다만, 2개의 자료에서 가정하는 준거변수와 유의한 관계를 보여주는 예측변수의 수에서 차이가 있음을 가정하였다. 또한 편향된 변수의 포함이 분류에 미치는 효과를 평가하기 위하여 이분형 범주변수를 준거변수로 설정하였다. 즉, 학습자료와 실제 자료에서 자료는 사례 수 1,000명이 5개의 예측변수와 1개의 준거변수를 가지고 있는 것으로 가정하였다. 학습자료와 실제 자료에서 4개의 예측변수와 1개의 준거변수 간의 관계가 2개의 모의자료에서 동일하였다. 다만, 학습자료에서는 5번째 예측변수와 준거변수 간의 상관이 통계적으로

유의하지만, 실제 자료에서는 해당 예측변수와 준거변수 간의 상관이 없음을 가정하였다.[6] 보다 구체적으로, 학습자료의 생성에는 식 (4)의 회귀식이 활용되는 반면에 실제 자료의 생성에는 식 (5)의 회귀식이 사용되었다. 2개의 회귀식에서 오차항은 N(0, 1)인 표준정규분포를 따른다고 가정하였다. 2개의 회귀식에 기초하여 자료를 생성한 이후에, 학습자료를 활용하여 알고리즘(즉, 로지스틱회귀방정식)을 추정하였다. 마지막 단계에서는, 학습자료에서 추정된 알고리즘을 실제 자료에 적용하여 분류의 오차를 추정하였다.

$$\text{logit}(z_y) = 0.5z_{x1} - 0.3z_{x2} + 0.2z_{x3} + 0.1z_{x4} + 0.5z_{x5} \quad (4)$$

$$\text{logit}(z_y) = 0.5z_{x1} - 0.3z_{x2} + 0.2z_{x3} + 0.1z_{x4} - 0.0z_{x5} \quad (5)$$

3.2.2 결과

편향된 변수가 포함된 학습자료의 활용이 분류 예측에 미치는 효과를 검증하기 위한 실시된 두 번째 모의실험의 결과는 편향된 변수를 분류의 예측에 사용되는 경우에는 인공지능 알고리즘이 기대되는 분류의 정확도를 충분히 보여주지 못함을 확인하였다. 구체적으로, 모의자료로 생성된 학습자료에서 추정된 인공지능 알고리즘은 식 (6)과 같은 로지스틱회귀식을 추정하였다.

---

6) 학습자료와 실제자료간의 결과변수를 예측하는 통계적으로 유의미한 변수의 차이로 인하여 발생하는 분류 정확도의 차이를 검증하고자 하였다. 이를 위하여, 학습자료에서는 결과변수를 통계적으로 유의미하게 예측하여 분류에 도움이 되지만, 실제자료에서는 결과변수를 통계적으로 유의미하게 예측하지 모하여 분류에 도움이 되지 않는 상황을 가정하였다. 특히, 실제자료에서는 분류의 정확도를 높이기 위해서 전혀 도움이 되지 않기 때문에 분류에 사용이 되어서는 안되는데 편향의 효과로 인하여 결과변수의 집단 분류에 사용되는 경우를 가정함으로써 실제 결과변수의 분류에 사용되어서는 안되는 변수가 사용되어 오히려 분류의 정확도가 떨어지는 편향의 효과를 검증하고자 하였다.

$$\mathrm{logit}(\hat{z}_y) = 0.50z_{x1} - 0.31z_{x2} + 0.19z_{x3} + 0.09z_{x4} + 0.49z_{x5} \quad (6)$$

식 (6)의 추정된 로지스틱회귀식을 활용하여 실제자료에서의 준거변수의 예측값을 구하고, 실제자료의 관찰값과 예측값의 차이에 기초하여 분류표를 추정하고 분류의 정확도를 추정하였다. 편향된 변수가 포함되지 않은 경우에 분류의 정확도를 추정하기 위하여 실제자료의 자료를 생성하기 위해서 사용된 식 (4)를 활용하였다. 식 (4)를 활용한 경우에 분류의 정확도는 93%이었다. 이에 반해서 식 (6)의 추정된 로지스틱회귀식을 활용하여 준거변수를 분류한 경우에는 분류의 정확도가 47%이었다. 즉, 학습자료의 편향된 변수가 실제자료에 사용되는 경우에는 49% 포인트((93%-47%)/93%) 감소하였다.

4. 경험자료 연구

4.1. 연구 방법

두 번의 모의실험 결과가 실제 경험자료에서 재현될 수 있는지 검증하기 위하여 두 개의 경험연구를 수행하였다. 두 개의 경험연구에서 사용된 자료는 첫 번째 모의 실험에서 가정한 상황을 대표할 수 있는 경우를 활용하였다. 즉, 첫 번째 경험연구에서는 종단자료를 사용함으로써 학습자료와 실제자료가 대표하는 모집단이 동일한 경우를 대표하였다. 즉, 종단자료를 사용함으로써 학습자료 및 실제자료를 생성하는 모집단이 동일함을 가정할 수 있다. 따라서 학습자료에서 추정된 알고리즘을 실제자료에 적용하여 결과를 예측하는 과정에서 편향이 발생하지 않음을 가정하였다. 이에 반하여 두 번째 경험연구에는 횡단자료를 사용함으로써 학습자료와 실제자료가 대표하는

모집단이 동일하지 못한 경우를 대표하였다. 즉, 횡단자료를 사용함으로써 학습자료 및 실제자료를 생성하는 모집단이 동일하지 않을 수 있음을 가정할 수 있다. 따라서 학습자료에서 추정된 알고리즘을 실제자료에 적용하여 결과를 예측하는 과정에서 편향이 발생할 수 있는 것을 가정하였다

첫 번째 자료는 종단자료가 사용되었다. 첫 번째 경험연구에서 활용된 자료는 한국청소년정책연구원에서 실시하여 공개한 종단자료인 '한국아동·청소년패널조사(korean children & youth panel survey; 이하 KCYPS)'에서 초1, 초4, 중1 패널 자료를 사용하는 것에 대한 동의를 받았다. 이 종단자료는 2010년도 기준 전국의 초등학교 1학년, 4학년, 중학교 1학년 재학생을 모집단으로 하여 다단층화집락표집(stratified multi-stage cluster sampling)에 근거하여 연구대상자들을 추출하여 표본으로 선정하였다. 본 연구에서는 초등학교 4학년 패널이 사용되었다. 초등학교 4학년 패널이 초등학교 6학년이 되는 3차 년도 응답자료는 학습을 위하여 사용되었고, 중학교 3학년이 되는 6차 년도 응답자료는 추정된 모형의 유용성을 검증하기 위한 실제자료로 활용되었다. 2개의 응답 자료를 활용하여 정신건강을 예측하는 회귀모형을 추정하였다(<표 2> 참조).

이에 반하여, 두 번째 자료는 횡단자료가 활용되었다. 두 번째 경험연구에서 활용된 자료는 한국청소년정책연구원에서 실시하여 공개한 횡단자료인 '청소년활동 참여 실태조사'이었다. 본 횡단자료는 전국 초등학교 4~6학년, 중학교 1~3학년, 고등학교 1~3학년에 재학 중인 9000명의 학생을 대상으로 수집되었다. 특히, 학습자료는 2014년도에 조사된 자료가 활용되었고, 학습된 회귀모형의 유용성은 2018년도 응답자료가 활용되었다. 경험자료 연구에서 사용된 회귀모형은 청소년들의 정의적 특성인 행복도를 측정하기 위한 회귀모형을 추정하였다(<표 3> 참조).

4.2. 연구 결과

종단 자료를 활용하여 추정된 회귀 모형은 결과는 <표 2>에 보고되었다. 3차 년도 자료를 사용하여 추정된 회귀모형에서 5개의 예측변수가 통계적으로 유의하였다. 특히, 본 회귀모형의 설명력은 .335이었다. 이에 반하여, 3차 년도를 활용하여 추정된 회귀모형의 회귀계수를 6차 년도 응답자료에 적용한 경우에서의 예측력을 추정하였다. 6차 년도 자료의 설명량을 추정하기 위하여 SST(sum of squared, 총제곱합) = SSR (sum of regressed, 설명된 제곱합) + SSE(sum of error, 오차 제곱합)이 사용되었다. 분석 결과, 1755.15=565.83+1189.34로 설명량은 .322이었다. 즉, 동일한 응답자들이 반복적으로 측정되는 종단자료에서는 시점에 따라서 동일한 회귀계수를 사용하는 모형은 비록 시점이 다르더라도 유사한 수준에서의 설명력을 보여주었다.

<표 2> 청소년의 정신건강을 예측하는 중다회귀모형

	B	se	β	t	p	R^2	ΔR^2
(상수)	0.633	.194		3.258	.001		
성별	0.250	.024	.206	10.256	.000		
학교지역(시도)	0.001	.006	.019	0.250	.803		
자택지역(시도)	0.001	.006	.008	0.106	.916		
주거 형태	-0.015	.011	-.028	-1.361	.174		
아버지 근로 여부	0.020	.027	.018	0.758	.449	.335	.112
어머니 근로 여부	-0.010	.013	-.017	-0.789	.430		
보호자 근로 여부	-0.089	.039	-.045	-2.249	.025		

보호자 건강 상태	0.000	.024	.000	-0.006	.995
보호자 삶의 만족도	0.038	.020	.045	1.885	.060
키	0.001	.001	.019	0.829	.407
몸무게	-0.001	.001	-.020	-0.896	.370
건강상태 평가	0.074	.015	.102	5.031	.000
전체성적 만족도	0.066	.014	.098	4.815	.000
학교적응도	0.210	.036	.141	5.887	.000
공동체의식	0.021	.022	.022	0.939	.348
학업관련시간	0.000	.000	-.027	-1.319	.187
여가시간	0.000	.000	-.018	-0.904	.366

덧붙여서, 횡단 자료를 활용하여 추정된 회귀 모형은 결과는 <표 3>에 보고되었다. 2014년도 자료를 활용 추정된 회귀모형에서 7개의 예측변수가 통계적으로 유의하였다. 특히, 본 회귀모형의 설명력은 .557이었다. 이에 반하여, 2014년도 자료를 활용하여 추정된 회귀모형의 회귀계수를 2018년도 조사자료에 적용한 경우의 예측력도 추정하였다. 2018년도 자료의 설명량은 2533.20(SST) = 2133.87(SSE) + 399.33(SSR)에 기초하여 .158이었다.

즉, 동일한 응답자들이 반복적으로 활용되는 종단자료와는 다르게 매 조사년도(2014년도와 2018년도)에 다른 연구 참여자들이 응답하는 횡단자료에서는 회귀모형의 동질성이 유지되지 않았다. 이러한 결과로 인하여 2014년도 회귀모형의 설명량은 .557에서 2018년도 회귀 모형의 설명량은 .158로 감소하였다.

<표 3> 행복도를 예측하기 위한 중다회귀모형 결과

	B	se	β	t	p	R^2	ΔR^2
(상수)	1.193	0.063		18.86	<.001		
자기존중감	0.363	0.014	.342	26.11	<.001		
진로성숙도 (계획성)	-0.036	0.014	-.039	-2.61	.009		
진로성숙도 (독립성)	0.08	0.012	.008	0.65	.517		
진로성숙도 (태도)	0.108	0.014	.100	7.64	<.001	.557	.310
진로성숙도 (자신지식)	0.131	0.013	.154	10.09	<.001		
진로성숙도 (진로행동)	0.122	0.013	.126	9.31	<.001		
외재적 동기	-0.020	0.011	-.022	-1.76	.078		
내재적 동기	0.064	0.12	.068	5.43	<.001		
무동기	-0.113	0.011	-.132	-10.44	<.001		

5. 결론

본 연구에서는 기계학습에 기초한 인공지능 알고리즘을 활용함에 있어 자료 혹은 변수의 편향성으로 인하여 실제 자료 분석에서 발생하는 오류를 검증하였다. 예를 들어, 학습자료와 실제자료가 대표하는 모집단의 속성이 서로 다른 경우에는 학습자료에서 추정된 알고리즘이 실제자료를 제공하고 있는 구성원들의 속성을 제대로 반영하지 못한다. 이로 인하여, 적절치 못한 알고리즘을 활용함으로써 실제자료의 예측에서 발생하는 설명량의 감소를 확인하였다. 둘째로, 학습자료에 준거변수에 편향된 영향을 미치는 예측변수가 포함된 경우에 발생하는 오류를 검증하였다. 즉, 편향된 예측변수를 실제자료의 예측 및 분류에 포함시킴으로써 예측 및 분류의 오류가 증가함을

확인하였다.

본 연구는 서론에서 살펴본 편향들이 인공지능 알고리즘을 실제로 활용함에 있어서 발생할 가능성이 높음을 확인하였다. 특히, 본 연구에서 살펴본 선행 연구들과 연구 결과는 학습자료의 중요성을 다시 한번 강조하였다. 즉, 선행연구에서는 인공지능 알고리즘을 사용하는 경우에 편향된 결과가 발생하는 원인을 정리하였다. 기술된 원인들을 요약할 수 있는 단어는 '학습자료의 중요성'으로 정리될 수 있다. 인공지능을 학습시켜 실제 문제 해결에 활용될 수 있는 알고리즘의 추정에 있어서 편향되지 않고 타당한 변수들로 구성된 학습자료가 가장 중요하면서도 기본적인 기초임을 보여주고 있다. 특히, 본 연구에서 수행된 간단한 시뮬레이션 연구와 실증자료 검증을 통해서 편향된 자료 및 변수의 사용으로 인하여 예측 및 분류에 있어서 오류가 무시할 수 없는 수준으로 증가함을 확인하였다.

본 연구의 결과를 통해서 인공지능 알고리즘을 활용함에 있어서 발생할 수 있는 제한점은 다음과 같다. 첫째, 편향된 표본에서 추출된 학습자료는 연구문제를 적용하고자 하는 모집단의 특성을 적절하게 반영하지 못하기 때문에, 편향된 표본에서 추출된 인공지능 알고리즘을 활용하여 실제 자료에서 연구문제를 해결하는 경우에 편향된 결과를 만든다. 편향된 연구 결과보다 더욱 문제가 되는 것은 빅데이터 및 인공지능을 활용함에 있어서 학습자료를 구축하기 위해서 사용되는 표본이 전집을 적절하게 대표하고 있는지에 대한 검증이 충분히 이루어지지 않는다는 것이다. 예를 들어, 최근까지 인공지능을 활용한 알고리즘이 주요하게 활용되는 분야는 경영 분야에서 마케팅이나 생산관리 등이었다. 이러한 분야에서는 학습자료를 구성하기 위해서 빅데이터의 활용이 가능하다. 하지만 경영 분야에서 인적자원관리는 다른 분야와 달리 학습자료를 구성하는 측면에서 빅데이터의 활용에 제한을 가진다. 특히, 마케팅 분야에서 학습자료와 실제 자료 간에는 시차적으로 거의

차이가 없이 동시적으로 구성이 될 수 있는 반면에, 인적자원관리는 학습자료에서의 알고리즘의 추정과 실제 자료로의 적용 간의 시간의 차이가 존재할 가능성이 높다. 이러한 시간적 차이로 인하여 학습자료와 실제 자료가 대표하는 모집단의 차이로 인한 편향성이 발생할 수 있다. 학습자료는 과거 데이터에 기초하기 때문에 과거의 관습과 편향이 새로운 시대에 창의적인 역량을 가진 인재의 선발을 방해할 수 있다.

보다 집중되어야 하는 것은 현대 사회에서 자료의 수집이 인터넷이나 SNS에 기초하여 수집되는 경우가 증가이다. 검색엔진이나 SNS와 같은 도구의 활용을 통한 자료수집은 필터버블이나 에코챔버와 같은 현상으로 편향된 사고를 가진 구성원들을 표집하여 모집단을 타당하게 대표하지 못할 가능성이 높다. 다만, 이러한 현상으로 학습자료의 크기를 확대함으로써 확률적으로 편향된 정보를 사후적으로 수정할 수 있다. 이에 반하여, 가치-중립적인 측면에서 편향된 정보는 기계학습에서 활용된 알고리즘에 따라서는 유사한 태도나 가치를 가진 조직 구성원들로 인하여 더욱 강화될 가능성이 높다는 문제점이 발생한다. 특히, 인적자원관리에서 기계학습에 사용되는 자료는 해당 조직 구성원들의 가치-중립적인 측면에서의 편향된 정보가 반영될 수 있다. 이러한 편향성으로 인하여 조직 내 구성원들이 공유하는 태도나 가치에서의 편향성이 확대될 가능성이 상존한다. 이러한 문제점을 해결하기 위해서 학습자료에 대한 태도나 가치의 편향성에 대한 외부 전문가의 감사 및 확인이 필요하다.

둘째, 학습자료에 기초하여 추정된 알고리즘의 형태에 대한 확인이 다소 제한적이다. 비록 본 연구에서는 학습자료에 기초하여 인공지능 알고리즘의 추정이 확인되어 예측 변수들의 상대적인 영향력을 비교하고 알고리즘의 유의성을 검증할 수 있다. 하지만, 딥러닝 인공지능에 대한 정보는 확인이 불가능한 경우가 대다수이다. 즉, 인공지능에 의해서 추정된 알고리즘에 대한

경험적 근거가 다소 미흡한 경우가 존재한다. 이러한 제한점으로 인하여 인공지능 알고리즘에 대한 설명이 불가능하기 때문에, 경영학의 인적자원개발 같은 경우는 인공지능 알고리즘의 활용이 다소 제한적일 수 있다. 특히, 인적자원에 대한 의사결정 과정에서 이의가 제기되는 경우에는 해당 과정에 대하여 설명할 수 없기 때문에 딥러닝에 기초한 인공지능의 활용이 제한될 수 있다.

셋째, 학습자료와 실제자료가 대표하는 모집단의 차이로 인하여 기계학습 알고리즘의 활용에 편향이 발생할 수 있다. 즉, 학습자료의 수집이 되는 과정인 표집에 대한 검증과 이를 통하여 학습자료가 모집단을 대표하는 정도에 대한 확인이 필요하다. 이에 덧붙여, 기계학습 알고리즘을 활용하여 실제 사회에서 발생하는 문제에 대한 실증적으로 연구하는 경우에는 학습자료와 실제자료간의 모집단의 동질성이 확보되었는지에 대한 사전적인 확인이 필요하다. 특히, 빅데이터를 활용한 기계학습 알고리즘이 탐색적인 방법으로 주요하게 활용되는 현실을 감안한다면 이론적 모형이 부재한 경우에 탐색적 방법이 사용되는 경우에 발생할 수 있는 오류에 대한 사전적인 검토가 필요하다. 심리학을 포함한 사회과학 분야의 연구들에서 탐색적 연구가 가지는 한계인 오류의 모형화(capitalization by chance)로 인하여 기계학습 알고리즘의 일반화의 한계에 대한 검토가 필요하다(MacCallum et al., 1992). 보다 구체적으로, 학습자료에서 추정된 기계학습 알고리즘을 실제자료에 적용하기에 앞서 실제자료의 일부분만을 표집하여 효과성을 검증하는 사전단계의 도입이 필요하다.

이러한 제한점들에도 불구하고, 인공지능을 활용하여 다양한 알고리즘들은 다양한 분야에서 생성되고 활용될 수 있다. 다만, 인공지능을 활용한 알고리즘/을 타당하게 적용하기 위해서는 학습자료의 중요성이 강조되어야 한다. 인공지능의 활용에 가장 기초가 되는 것은 인공지능이나 알고리즘의

종류를 선택하기에 앞서, 사회과학 분야에서 도출되는 다양한 연구문제를 해결하기 위해서 연구문제 및 해결책을 적용할 수 있는 모집단의 특성을 타당하게 반영하는 표본에 기초한 학습자료의 수집이다.

V. 연구 3

“설명중심의 통계모형과 예측중심의 기계학습의 비교”

1. 서론

사회과학에 속하는 심리학은 역사적으로 인간의 행동을 이해하고 설명하기 위해 변수들 간의 인과관계를 탐색하고 검증하는 것에 관심을 가져왔다(Yarkoni & Westfall, 2017). 인간 행동을 이해하기 위해 무선 표집과 엄격하게 통제된 실험방법이 심리학 연구의 중심에 자리 잡았고, 이를 통해서 어떠한 변수들이 어떻게 또는 왜 인간의 행동을 예측 혹은 설명하는지에 대한 연구가 진행되어왔다. 그러나 인간 행동의 원인을 설명하는 것에 초점을 두기 때문에 심리 메커니즘의 복잡한 이론을 검증하고 논의하는 연구들이 다수인 실정이다(Yarkoni & Westfall, 2017).

설명 중심 전략(explanation-focused approach)은 상대적으로 적은 수의 변수들을 포함하는 제한된 단순한 모형을 구축하기 때문에 미래의 행동 예측에 대한 정확성이 떨어질 수 있다. 반면에, 예측 중심 전략(prediction-focused approach)은 예측력을 높이기 위해서 많은 변수들이 분석에 포함될 수 있다. 이를 바탕으로 미래를 예측하는 기계학습의 원리와 기술이 결합됨으로써 보다 예측력이 높은 모형을 추정할 수 있다(Yarkoni & Westfall, 2017). 그러나 예측중심 전략을 사용하는 기계학습의 경우 변인들 간의 논리적 관계를 설명하는 이론에 기초하지 않고 활용 가능한 모든 데이터를 분석에 활용하여 통계 모형을 추정하기 때문에 어떤 변수가 어떻게 또는 왜 사용되는지 알 수 없다. 결과적으로,

분석 결과를 해석함에 있어 이론적 근거 부족으로 인하여 추정된 모형의 타당화와 일반화 가능성이 상대적으로 떨어질 수 있다. 반면에 설명 중심 전략을 사용하는 전통적인 통계분석 방법은 이론을 기반으로 하여 중요 변수를 선택하고 이를 통계 모형의 추정에 사용하기 때문에 분석 결과에 대한 설명이 용이해지고, 모형의 일반화 가능성이 상대적으로 높다(Yarkoni & Westfall, 2017; Cawley & Talbot, 2010).

분석 결과의 예측력을 상대적으로 더 중요하게 여겨지는 장면에서 빅데이터의 활용은 큰 이익을 가져다줄 수 있다. 예를 들어 공장 관리 장면에서 빅데이터를 활용함으로써 생산 시간을 단축하거나 생산부품의 결함 또는 기계적 결함을 선제적으로 예측하여 보다 빠르게 대처할 수 있다. 생산관리 같은 장면에서는 과정 및 부품의 오류 및 결함을 예측함에 있어서 인간이라는 구성요소가 포함되지 않는다. 특히, 예측의 과정에서 인간의 인지 편향이 포함되지 않기 때문에 통계 모형을 활용함에 있어서 사람에게 불평등하거나 공정하지 않은 결과를 초래할 가능성이 현저하게 낮다. 따라서 생산관리와 같은 장면에서는 상황을 설명하는 것보다 결과를 빠르고 정확하게 예측하는 것이 상대적으로 중요하게 여겨진다. 하지만 경영 장면에서의 인사선발 및 인사평가처럼 인간에 대한 깊은 이해와 이들에 대한 처분에서의 공정성이 요구되는 장면에서는 분석 결과의 예측력뿐만 아니라 의사결정에 대한 설명 가능하고 합당한 근거가 필요하다. 이론적 근거 혹은 법률적(혹은 도덕적) 제한 등을 고려하지 않고 예측력에만 초점을 둔다면 불평등을 초래하는 변수를 포함하게 되어 분석 모형의 해석 및 적용에 편향된 결론을 유발할 위험성이 상존한다(Fan, Han & Liu, 2014). 기계학습에서 기계를 학습하기 위해 사용되는 데이터는 사회 전반에 만연하고 있는 편견을 그대로 반영하고 있기 때문에 이러한 데이터에 기반하여 학습된 모형은 사회에 이미 존재하는 배타적인 면과

불평등을 포함하지만 예측에는 유용한 관계성을 포함할 수 있다(Brocas & Selbst, 2016).

따라서 편향된 자료와 모형에 기초한 기계학습 결과에 무조건 의존하고 이를 활용하게 된다면 역사적으로 불우하고 취약한 집단이 일관되게 불리한 방식으로 사회에 참여할 수 있는 기회가 제한되는 효과가 발생할 수 있다. 나아가 이렇게 발생한 차별은 인간의 의식적인 선택이 아닌 알고리즘 사용으로 발생하는 자동적이며 의도치 않게 나타나는 현상이기 때문에 문제의 원인을 파악하는 것이 더욱 어려워질 수 있다(Brocas & Selbst, 2016; Fan et al., 2014). 따라서 빅데이터를 활용한 기계학습은 통계 모형의 활용이라는 측면에서 유용할 수 있지만, 해당 모형을 타당하게 현실 문제에 적용하기 위해서는 처음부터 올바른 자료를 학습할 수 있도록 자료에 포함되는 변수에 대한 고찰이 필요하다. 또한 변수 및 자료의 특성에 맞는 모형 혹은 알고리즘을 고려하여 이론적 근거를 토대로 모형 추정의 결과에 대한 설명이 용이하도록 해야 한다.

이처럼 빅데이터에 기초한 기계학습을 현실문제의 해결에 사용할 경우에 주의해야할 점과 문제점에 대한 다양한 연구가 진행되어 왔다. 가령, Fan et al. (2014)는 빅데이터의 방대한 표본 크기, 변수의 수가 많아짐에 따라 나타나는 차원의 확장은 노이즈에 대한 민감성[7], 허위상관[8] (Cohen et al., 2013)과 같은 통계적 문제를 야기한다고 밝혔다. 또한 이러한 결과로 인해 검증하기 힘들거나 잘못된 통계적 추론을 이끌어 결과적으로 잘못된

7) 빅데이터 분석 시 많은 변수를 동시에 투입하여 모형을 추정하기 때문에 발생하는 추정오차에 해당되며 분석에 사용될 때 제거되거나 무시되어야 하는 데이터를 의미한다. 학습데이터가 이런 노이즈에 너무 치중되어 설명하게 되면 과적합이 발생해 실제 변수들 간의 관계를 오해석 할 수 있다(Fan, Han & Liu, 2014).

8) 허위상관(spurious correlation)은 오염변수나 우연의 일치로 인해 두 개 이상의 변수가 통계적으로 유의한 상관을 나타내지만 인과적으로 상관이 없는 관계를 말한다. 상관성만 고려한다면 예측력은 높아질 수 있지만 상관관계는 곧 인과관계를 나타내는 것은 아니기 때문에 정확한 문제해결이 도움이 되지 않을 수 있다(Fan, Han & Liu, 2014; Cohen, Cohen, West & Aiken, 2013).

결론으로 이어질 수 있다는 문제점을 제시하였다.

또한, Sivarajah et al. (2017)은 빅데이터가 다양한 사회문제에서 실적인 대안을 제시할 수 있지만 잠재력을 완벽하게 실현하기 위해서는 해결할 과제가 많다고 밝혔다. 가령 빅데이터의 특성상 발생하는 문제, 기존의 분석방법으로 빅데이터를 분석할 시 생기는 문제, 그리고 데이터 처리 시스템의 한계에 대해 소개하였다. Crawford(2013)는 빅데이터에 기초하여 추정된 결과가 믿을 만한지에 대해 회의적으로 접근하였다. 예를 들어, 인간이 데이터를 생성하는 것에 관여하기 때문에 데이터 자체가 객관적이지 못하며, 더 나아가서 인간에 의해 생성 및 수집되는 데이터로 추론과 의사결정을 하는 것은 숨겨진 편견으로 인한 위험성을 내포될 수 있다고 밝혔다.

이에 따라 본 연구에서는 고용노동부와 한국고용정보원이 제공하는 '대졸자 직업 이동 경로조사' 데이터를 사용하여 설명중심 전략을 이용하는 전통적인 통계분석 결과와 예측 중심 전략을 사용하는 기계학습을 통한 분석 결과를 비교해보고자 한다. 전통적인 통계분석의 경우, 선행 연구에 기반하여 대졸자 취업 여부에 영향을 미치는 중요한 요인들을 선택하여 분석에 포함하고 로지스틱 회귀분석을 실시하였다. 이에 반하여, 기계학습의 경우, 주어진 모든 데이터를 학습시켜 랜덤 포레스트(random forest) 알고리즘으로 대졸자 취업여부에 어떤 변수들이 영향을 미치는지 알아보았다. 이후 설명 중심 전략을 사용하는 기존의 로지스틱 회귀분석 결과와 예측 중심 전략을 사용하는 랜덤 포레스트 분석 결과를 비교하였다. 각 분석 방법에서 중요하게 여기는 변수를 비교하였고 주어진 변수에 따라 취업 여부에 대한 예측력 또한 비교하였다. 두 모형의 추정 결과를 비교함으로써 빅데이터 활용 시 설명 중심 전략과 예측 중심 전략의 장단점을 알아보고 어떤 방법이 가장 효율적이며 정확하고 공정한

의사결정을 도모할 수 있는지 확인하였다.

2. 연구방법

2.1 로지스틱 회귀분석과 랜덤 포레스트 분석방법

본 연구에서는 이분형 결과변수를 예측에 활용될 수 있는 두 가지 접근법에서의 효율성 및 편향이 비교되었다. 로지스틱 회귀분석과 랜덤 포레스트는 데이터마이닝 기법중 하나이며, 의사결정을 위해 빅데이터에서 변수간의 관계나 패턴 등을 탐색하여 필요한 지식을 도출하는 방법이다(오세웅, 2017). 구체적으로, 동일한 자료를 활용하여 연구 문제를 검증하기 위하여 로지스틱 회귀분석과 랜덤 포레스트 모형이 활용되었다.

설명 중심 전략을 사용하는 전통적인 로지스틱 회귀분석은 결과변수가 범주형일 경우 예측변수를 사용하여 관측값을 분류하거나 예측하는데 사용되는 분석기법이다. 결과변수가 연속변수가 아니기 때문에 확률의 값이 0과 1사이의 값을 취하게 되어 예측변수와 결과변수가 선형 관계를 이룬다는 전제가 적절하지 않다. 따라서 확률을 로짓(logit)으로 변환하여 예측변수와 결과변수의 관계를 선형함수로 표현하는 일반화 선형모형(generlalized linear model)을 적용한다(홍세희, 2005). 하지만 로지스틱 회귀분석도 회귀분석의 일종이기 때문에 회귀계수를 사용하여 변수에 대한 해석이 가능하다. 또한 설명 중심 전략이기 때문에 연구자가 이론을 바탕으로 결과변수를 가장 효율적으로 설명할 수 있는 예측변수를 선택하게 된다. 따라서 투입한 변수들이 결과변수에 미치는 영향력을 비교할 수 있지만 사례수가 커지는 경우 검증력이 높아져 전체적인 예측력이 낮아질 수 있는 단점을 가지고 있다(홍세희, 2005).

예측 중심 전략을 사용하는 랜덤 포레스트는 로지스틱 회귀분석과는 달리 결과변수가 범주형 또는 연속형인 경우 모두 사용 가능한 분류분석의 일종이다. 랜덤 포레스트는 의사결정나무(decision tree) 모형을 여러 개 만들어서 더 정확한 예측 결과를 나타낸다(오세웅, 2017; 유진은, 2015). 랜덤 포레스트의 알고리즘 과정은 모집단으로부터 표집한 학습 데이터에서 환원표집을 이용해 부트스트랩 표본을 생성한다. 이런 과정을 반복하여 n개의 부트스트랩 데이터를 생성하여 의사결정나무 알고리즘을 적용해 랜덤하게 m개의 예측변수를 선택하게 된다(오세웅, 2017). 이렇게 선택한 예측변수를 이용하여 훈련목적 함수를 최대로 만드는 분할을 찾아 노드 분할 함수의 최적값을 구한다. 빅데이터에서 수천 개의 예측변수를 활용할 수 있다는 장점이 있으며 예측 중심 전략이기 때문에 중요한 변수를 찾아내는데 효율적인 분석 방법이다. 하지만 랜덤 포레스트 분석 결과에 대한 이론적인 설명이나 해석이 어렵다는 단점을 가지고 있다(유진은, 2015).

로지스틱 회귀분석과 랜덤 포레스트는 모두 데이터 마이닝 중 분류분석의 일종으로 동일하지만 알고리즘에서 차이를 보인다. 따라서 본 논문에서는 빅데이터를 두 가지 분석 방법으로 각각 분석하여 결과를 비교하였다. 로지스틱 회귀분석 결과 예측변수들의 회귀계수를 제공하고 통계적으로 유의한지 살펴보았다. 그리고 랜덤 포레스트를 위하여 500개의 부트스트랩 표본을 생성하였으며, 변수의 중요도 지수인 지니 지수를 제공하였다.

2.2 자료

본 연구는 고용노동부와 한국고용정보원이 제공하는 '대졸자 직업이동 경로조사' 를 활용하여 분석을 실시하였다. 이는 대졸자의 교육과정,

구직활동, 일자리 경험, 직업훈련, 자격증 등에 대한 조사이며, 2006년 조사 시작 당시에는 종단 패널 조사로 설계되었지만 2012년 이후 횡단 조사로 실시되고 있다. 구체적으로 대졸자의 노동시장 진입 및 이동에 대한 데이터를 분석하기 위해 2018년에 2~3년제 대학 및 4년제 대학 졸업자들을 대상으로 실시한 자료를 사용하였다. 전통적인 방법인 로지스틱 회귀 분석 시 전체 사례 수 18,081명 중 296명의 결측치를 제외한 17,785명의 데이터가 사용되었다. 그리고 기계학습을 이용한 랜덤 포레스트 분석 시 전체 예측변수 1,182개 중에 결측값이 없는 변수만 사용하여 총 178개의 예측변수가 분석에 사용되었다. 추가로 id 변수도 제외되었으며 53개 이상의 범주를 가진 변수는 분석에 사용불가 하여 최종적으로 총 164개의 예측변수가 사용되었다. 로지스틱 회귀분석의 컴퓨터 분석 시스템은 SPSS 22.0을 사용하였으며, 랜덤 포레스트는 R 3.1.1을 사용하여 분석하였다.

대졸자 취업 여부에는 많은 요인이 복잡하게 영향을 미치며 주로 취업 준비생의 인구/사회/경제적 요인과 취업역량 요인 그리고 교육배경 요인 등으로 나누어서 연구가 진행된다(염동기 등, 2017). 최기성과 조민수(2016), 이종찬과 박지현(2015)는 인구/사회/경제적 요인으로 성별, 연령, 부모의 사회/경제적 배경을 포함하였다. 그리고 대졸자의 취업에 영향을 미치는 취업역량 요인으로는 대학학점, 어학연수, 자격증 등을 포함하였으며 교육배경 요인으로는 전공계열, 학교소재지역, 학교특성 등을 포함하였다. 따라서 본 연구에서는 전공계열, 학교유형, 학교소재지역, 고등학교 계열, 졸업대학 입학 구분, 대학 입학 모집방법, 복수전공/부전공/연계전공 여부, 어학연수 경험, 자격증 소지 여부를 대졸자의 취업을 예측하는 변수로 포함하였다.

<표 4> 취업성과에 미치는 영향요인 분류

연구자	분류	요인
이종찬·박지현 (2015)	인구/사회/ 경제적 요인	성별, 연령, 부모의 사회경제적 배경
	취업역량 요인	학점, 어학연수, 재학 중 근로경험
	교육배경 요인	학력(교육수준), 대학특성, 대학지명도, 대학소재지, 전공계열, 취업지원서비스
최기성·조민수 (2016)	인구/사회/ 경제적 요인	성별, 연령, 부모의 사회·경제적 배경(학력과 가구소득), 출신대학소재지
	교육배경 요인	최종학교, 대학유형, 전공, 명성
	취업역량 요인	대학성적, 어학연수, 자격증, 교육훈련, 일, 취업지원 프로그램 경험
최일수·신은종 (2016)	인구/사회/경제적 요인	성별, 연령, 학력, 부모의 사회적 배경(부모의 학력과 소득), 전공계열, 출신학교의 소재지, 대학입학성적(고교내신 성적, 수능성적)
	취업역량 요인	학점, 외국어능력, 현장실습
	취업역량 요인	어학연수 경험, 자격증 취득, 취업클리닉 참여

고용노동부, 한국산업인력공단, 대한상공회의소가 제공하는 블라인드채용 가이드북(2018)은 채용과정에서 편견이 개입될 수 있는 차별적인 요인을 제외하여 직무능력 중심 평가를 통한 채용을 적용할 수 있도록 정보를 제공한다. 블라인드 채용은 채용의 공정성과 우수 인재 채용을 통한 기업의 경쟁력 강화, 그리고 공정한 채용을 통한 사회적 비용감소를 목적으로 한다. 블라인드 채용을 실시하기 위해서 먼저 「고용정책기본법」 제7조 “성별, 신앙, 연령, 신체조건, 사회적 신분, 출신지역, 학력, 출신학교, 혼인·임신, 병력 등”에 따르는 차별적인 항목과 직무상 필요한 요인인지를 결정해야 한다. 이에 따라 차별적 항목이지만 직무 필요조건에 해당된다면 포함가능하며 직무 필요조건에 해당되지 않는 차별적 항목은 제외하도록 지시된다.

따라서 본 논문에서는 이론을 바탕으로 변수 선택을 실시하여 분석을 진행하는 전통적인 회귀분석에서는 「고용정책기본법」 제7조에 따른 차별적인 항목에 포함되는 성별, 연령, 신체조건, 출신지역을 제외하였다. 그리고 기계학습에서는 이론이 포함되지 않기 때문에 랜덤 포레스트 분석 시 사용 가능한 모든 변수를 투입하여 분석하였다.

3. 연구결과

전통적인 로지스틱 회귀 분석 결과 포함된 변수 중 전공계열, 학교유형, 학교소재지역, 고등학교계열, 대학 입학 모집방법, 자격증 소지여부가 통계적으로 유의하게 나타났고 졸업 대학 입학 구분과 복수전공/부전공/연계전공 여부 그리고 어학연수 경험은 유의하지 않게 나타났다(<표 5>). 전공계열의 경우, 인문은 사회, 공학, 의약과 차이가 없었지만 인문보다 교육, 자연, 예체능은 취업률이 높았다. 학교유형의 경우, 2~3년제와 4년제는 차이가 없었지만 2~3년제보다 교육대는 취업률이 높았다. 학교소재지역의 경우, 서울을 기준으로 부산, 광주, 강원, 전북, 전남, 경북, 경남의 취업률이 낮게 나타났으며 나머지 지역은 서울과 유의미한 차이가 없었다. 고등학교 계열의 경우, 일반계고(문과)보다 자율형 사립고/자율형 공립고와 방송통신고등학교/대안학교 등은 취업률이 낮았다. 반면에 일반계고(문과)보다 특성화고(공업계)의 취업률이 유의하고 높았다. 대학 입학 모집 방법의 경우, 정시보다 수시가 취업률이 높았으며 자격증 소지 여부의 경우, 자격증을 소지한 경우가 아닌 경우보다 취업률이 유의하게 높았다.

<표 5> 취업 여부에 관한 로지스틱 회귀분석 계수 및 오즈비

(취업=0, 비취업=1)

변수설명				*B*	*S.E.*	Wald	*df*	*p*	Exp(B)
g171majorcat	전공계열	전공계열				202.127***	6	.000	
		사회(1)	인문(0)	0.105	0.072	2.173	1	.140	1.111
		교육(1)	인문(0)	−0.157	0.064	6.030 *	1	.014	0.855
		공학(1)	인문(0)	0.072	0.086	0.701	1	.403	1.074
		자연(1)	인문(0)	−0.353	0.068	26.898 ***	1	.000	0.703
		의약(1)	인문(0)	0.085	0.071	1.421	1	.233	1.088
		예체능(1)	인문(0)	−0.934	0.095	95.826 ***	1	.000	0.393
g171school	학교유형	학교유형				58.704 ***	2	.000	
		4년제(1)	2,3년(0)	−0.055	0.047	1.376	1	.241	0.946
		교육대(1)	2,3년(0)	−1.821	0.238	58.687 ***	1	.000	0.162
g171area	학교소재지역	학교소재지역				79.894 ***	16	.000	
		부산(1)	서울(0)	0.183	0.065	8.063 **	1	.005	1.201
		대구(1)	서울(0)	0.105	0.088	1.414	1	.234	1.111
		대전(1)	서울(0)	0.140	0.080	3.042	1	.081	1.150
		인천(1)	서울(0)	−0.165	0.115	2.080	1	.149	0.848
		광주(1)	서울(0)	0.252	0.090	7.956 **	1	.005	1.287
		울산(1)	서울(0)	0.003	0.164	0.000	1	.984	1.003
		경기(1)	서울(0)	−0.074	0.059	1.592	1	.207	0.928
		강원(1)	서울(0)	0.199	0.084	5.669 *	1	.017	1.220
		충북(1)	서울(0)	0.120	0.089	1.814	1	.178	1.127
		충남(1)	서울(0)	−0.091	0.080	1.308	1	.253	0.913
		전북(1)	서울(0)	0.374	0.081	21.200 ***	1	.000	1.453
		전남(1)	서울(0)	0.365	0.108	11.436 **	1	.001	1.440
		경북(1)	서울(0)	0.274	0.069	15.676 ***	1	.000	1.316
		경남(1)	서울(0)	0.285	0.078	13.307 ***	1	.000	1.330
		제주(1)	서울(0)	−0.140	0.207	0.455	1	.500	0.870
		세종(1)	서울(0)	0.097	0.155	0.397	1	.529	1.102
g171f009	고등학교계열	고등학교 계열				24.926 **	9	.003	
		일반이과(1)	일반문과(0)	0.064	0.046	1.970	1	.160	1.067
		외국어고, 과학고, 국제고(1)	일반문과(0)	0.053	0.098	0.294	1	.588	1.054
		예술고, 체육고(1)	일반문과(0)	0.171	0.153	1.251	1	.263	1.187
		마이스터 고등학교(1)	일반문과(0)	−0.490	0.344	2.029	1	.154	0.613
		상업·정보계(1)	일반문과(0)	−0.022	0.085	0.068	1	.794	0.978

		공업계(1)	일반문과(0)	−0.254	0.102	6.186 *	1	.013	0.776
		농생명, 수산, 해양고 등(1)	일반문과(0)	0.245	0.213	1.324	1	.250	1.277
		자율고(1)	일반문과(0)	0.351	0.140	6.315 *	1	.012	1.420
		기타(1)	일반문과(0)	0.507	0.255	3.937 *	1	.047	1.660
g171f010	졸업 대학 입학 구분	편입(1)	입학(0)	−0.035	0.078	0.208	1	.648	0.965
g171f013	대학 입학 모집방법	모집방법				8.551 *	2	.014	
		수시(1)	정시(0)	−0.100	0.034	8.389 **	1	.004	0.905
		기타(1)	정시(0)	0.055	0.231	0.056	1	.813	1.056
g171f023	복수전공, 부전공 여부	안함(1)	함(0)	0.078	0.048	2.650	1	.104	1.081
g171i001	어학연수 경험	아니오(1)	예(0)	0.040	0.053	0.565	1	.452	1.041
g171m001	자격증 소지 여부	아니오(1)	예(0)	0.150	0.035	18.957 ***	1	.000	1.162
상수항				−0.864	0.105	67.208	1	.000	0.421

기계학습에 기초한 랜덤 포레스트 분석 결과, 분석에 사용된 총 164개의 예측변수 중 취업 여부에 영향을 미치는 요인들 중 상위 10% 요인인 16개를 순서대로 나열하였다(<표 6>). 가장 영향력이 높았던 변수 두 가지인 'SQ2. 지난 1주간 주로 한일' 과 'SQ1. 지난 4주간 주로 한일' 은 결과변수인 취업 여부를 직접적으로 응답하는 형태임으로 제거하였다. 취업 여부에 큰 영향을 미치는 변수 16개 중 「고용정책기본법」 제7조에 따라 4개(연령, 현재 거주지, 몸무게, 키)의 차별적 항목이 포함됨이 확인되었다. 이 밖에도 이론적으로 설명이 어려운 변수 5가지가 함께 포함되었다(함께 거주하는 가구원 수, 하루 평균 수면시간, 음주 빈도, 출생 월, 일주일 평균 운동 시간). 즉 총 16가지 변수 중 9개인 50% 이상의 변수가 고용정책기본법에 위반되는 변수를 포함하거나 이론적으로 설명 불가한 변수를 포함하였다.

<표 6> 취업 여부에 관한 랜덤 포레스트 분석 결과 (취업=0, 비취업=1)

변수설명			Gini(불순도) 점수
g171p041	P15.	가족에게 경제적 지원을 받는지 여부	663.527
g171k109	K5.	향후 1년간 계획	552.237
g171p039	P14.	지난 한 해 소득_만원	431.373
g171p017	P6.	함께 거주하는 가구원 수	131.696
g171p043	P16.	가족에게 경제적 지원을 하는지 여부	130.359
g171age		연령(2018년 9월 기준)	102.292
g171p025	P8.	현재 거주지 시군구(구/시/군)	100.798
g171q014	Q11.	몸무게_kg	93.415
g171g001	G1.	대학 졸업 후 학교(대학/대학원)에 다닌 경험 유무	90.215
g171k027	K4.	일자리 정보를 얻은 방법_1순위	86.893
g171q013	Q10.	키_cm	85.627
g171j021	J7.	유보임금 (희망하는 최소 임금)_연봉_만원	78.943
g171q003	Q3.	하루 평균 수면시간	71.762
g171q006	Q5.	음주 빈도	68.940
g171birthm		출생 월	65.068
g171q002	Q2.	일주일 평균 운동 시간	64.802

결과적으로 로지스틱 회귀분석과 랜덤 포레스트 분석 결과 공통적으로 유의하거나 중요한 변수로 나타난 변수는 없었다. 다만 로지스틱 회귀분석에서는 분석결과 유의미한 변수들에 대한 해석이 가능했지만 랜덤 포레스트 분석시에는 해석 보다는 어떤 변수가 더 설명력이 높은지를 차례대로 알아볼 수 있었다.

추가로 각 분석 방법의 예측정확도를 알아보기 위해 오차행렬(confusion matrix)을 살펴본 결과(<표 7>, <표 8>) 랜덤 포레스트는 정답률(hit

ratio)이 92.03%로 전통적인 분석 방법을 사용한 로지스틱 회귀분석의 정답률(hit ratio)인 68.61% 보다 정확한 예측 성능을 보였다. 로지스틱 회귀분석의 경우 취업을 한 경우 취업을 했다고 올바르게 예측하는 예측력은 99.82%로 93.17%인 랜덤 포레스트 분석 결과보다 높았으나 비취업인 경우 비취업이라고 올바르게 예측한 예측력은 0.23%로 89.55%인 랜덤 포레스트 분석 결과보다 낮았다. 또한 로지스틱 회귀분석 시 2종 오류가 0.18%로 사례수가 많아짐에 따라 검증력 또한 함께 과도하게 높아져 비취업 상태임에도 불구하고 취업상태라고 예측하는 1종 오류가 99.77%로 나타났다.

결과적으로 랜덤 포레스트가 로지스틱 회귀분석보다 전체적으로 나은 예측력을 보였지만 중요한 변수 16가지 중에 50% 이상이 고용정책기본법에 위반되는 변수 (연령, 현재 거주지, 몸무게, 키) 또는 변수 선택을 뒷받침하는 이론적 배경이 없는 변수(함께 거주하는 가구원 수, 하루 평균 수면시간, 음주 빈도, 출생 월, 일주일 평균 운동 시간)가 나타났다. 고용정책기본법에 위반되는 변수들이 포함됨으로써 인사선발 및 인사평가에서 불평등을 초래할 수 있는 위험성을 수반했다. 또한 변수추출에서 이론적 근거가 부족하여 선택된 예측변수가 결과변수에 중요한 변수로 작용하는 이유에 대한 직관적인 해석이 어려웠다.

<표 7> 로지스틱 회귀분석 결과 오차행렬

Logistic Regression (n=17,785)		Observed		Total
		비취업	취업	
Prediction	비취업	13 (0.23%)	21 (0.18%)	
	취업	5,561 (99.77%)	12,190 (99.82%)	
Total		5,574	12,211	17,785 (68.61%)

<표 8> 머신러닝 랜덤 포레스트 분석 결과 오차행렬

Test Data Set (n=3,641)		Reference		Total
		비취업	취업	
Prediction	비취업	1,020 (89.55%)	171 (6.83%)	
	취업	119 (10.45%)	2,331 (93.17%)	
Total		1,139	2,502	3,641 (92.03%)

4. 논의

본 논문에서는 설명 중심 전략을 사용하는 기존의 로지스틱 회귀분석과 예측 중심 전략을 사용하는 랜덤 포레스트 분석 결과를 비교해보았다. 두 모형의 결과를 비교함으로써 빅데이터를 사용할 때 설명중심전략과 예측중심전략의 장점 모두를 가지는 효율적이고 정확한, 그리고 공정한 의사결정을 내릴 수 있는 방법을 알아보았다.

결과적으로 전통적인 로지스틱 회귀분석은 각각의 예측변수들이 결과변수에 미치는 영향력을 비교할 수 있었고 예측변수의 범주 간의 차이를 비교할 수 있었다. 하지만 빅데이터에서 사례수가 너무 커짐에 따라 검증력이 높아져서 예측력에서 문제가 나타나는 것으로 확인되었다. 이 같은 현상을 오경보 또는 거짓양성(false alarm 또는 false positive)이라고 하는데 영가설이 사실임에도 불구하고 영가설을 기각하고 귀무가설을 택하는 오류이다. 거짓양성이 나타나서 통계적 추론에 오류가 생기면 정확한 원인 진단을 하기 힘들어진다. 이로 인해 잘못된 의사결정, 또는 공정하지 않은 판단을 초래할 위험이 높아진다(홍세희, 2005). 랜덤 포레스트는 빅데이터를 분석하는 경우 기존 통계 분석 방법보다는

기계학습을 이용한 분석 방법이 예측력에서 더 효과적이었다. 하지만 뒷받침되는 이론 없이 모든 변수가 분석에 포함되기 때문에 해석이 모호하거나 고용정책기본법에 위반되는 변수, 즉 공정성에 어긋나는 변수가 사용됨을 보였다. 따라서 빅데이터 분석 시 모든 데이터를 분석에 사용하기보다는 상황에 맞게 변수를 정리하고 데이터를 클리닝 하는 과정을 거쳐야 한다. 이에 따라 기계학습을 통한 정확한 예측력뿐만 아니라 이론에 기반하여 결과를 해석할 수 있고 형평성에 어긋나지 않는 의사결정을 내릴 수 있다.

또한 로지스틱 회귀분석과 랜덤 포레스트 분석 결과가 서로 다르게 나타난 이유는 두 모형 자체에 차이가 존재하기 때문으로 보인다. 로지스틱 회귀분석은 단변량으로 진행되지만, 랜덤 포레스트는 다변량으로 진행되어 변수 선택을 자체적으로 진행한 후 생성된 여러 개의 의사결정 트리의 투표를 통해 1/0 데이터를 분류한다. 따라서 다변량을 사용하는 랜덤 포레스트의 경우 다른 변수에 의해 변수의 영향도가 간섭받는 다중공선성의 문제가 발생할 수 있다. 즉 기계학습을 기반으로 하는 빅데이터 분석 시에는 많은 변수가 투입되기 때문에 예측변수 간의 상관관계가 존재하여 추정치에 대한 해석이 어려워지는 다중공선성 문제가 발생하지 않도록 주의해야 한다.

본 논문에서는 학습자료와 평가자료를 나누어 분석하기는 했으나 애초에 학습자료와 평가자료가 동일한 자료에서 나누어졌으므로 과적합이 발생하지 않은 것으로 보인다. 하지만 빅데이터 분석 시 과적합의 문제는 늘 염두해 두고 조심해야 할 사항 중 하나가 돼야 한다. 왜냐하면 모형의 복잡도는 분석에서 사용되는 변수의 개수로 결정되는데 정교하고 복잡한 모형(complex model)의 경우 편향(bias)은 작아질 수 있지만 변산도(variance)가 커지게 된다. 반대로 간결한 모형 (robust model)은

편향은 커질 수 있지만 추정치의 변산도는 작아지게 된다. 여기서 편향은 추정값(즉, 기댓값)와 모수 (즉, 실제값)의 차이로 가설과 실제 현상의 차이를 말하며 편향이 클수록 학습 데이터에서 예측변수가 결과변수를 잘 설명하고 있는 것이다. 변산도는 예측값과 평균의 차이로 데이터가 평균과 떨어진 정도를 나타낸다. 이런 현상을 편향과 분산의 트레이드오프(bias-variance tradeoff)라고 일컫는데 빅데이터를 사용한 기계학습처럼 고려하는 변수가 많은 복잡한 모형을 구축할 시, 노이즈(noise)에 민감하게 반응하여 과적합(overfitting) 문제가 발생할 가능성이 높아지는 것을 의미한다(Yarkoni & Westfall, 2017). 변산도가 크다는 것은 학습자료에 너무 의존하여 주어진 데이터에 대한 설명력은 높지만 새로운 데이터가 입력되는 경우 일반화 가능성이 떨어져서 예측력이 낮아질 수 있는 것이다(Ying, 2019).

과적합이 발생하는 이유는 첫째로 학습자료와 실제 자료가 대표하는 모집단이 달라서 실제 자료를 설명하기에 학습자료가 더이상 대표성을 가지지 못하는 것이다. 두 번째로는 고려하는 변수가 너무 많아지면 차원(dimension)이 커지고 데이터 공간이 함께 커져서 분석에 필요한 데이터 건수가 증가하여 주어진 데이터로는 공간을 설명할 수 없을 때 과적합이 발생하는 것이다. 따라서 과적합이 발생했는지 알아보고 기계학습의 효과를 검증하기 위해 원본 데이터는 주로 두 가지 데이터로 나누어서 진행한다. 알고리즘이 학습하기 위한 데이터인 학습자료(training set)와 학습자료를 학습 시키고 난 후 모형의 예측 정확도를 가늠하는 실제데이터인 평가자료(test set)이다. 평가자료는 모형의 성능이 검증된 후 모형을 구축할 때 포함되지 않았던 데이터로 학습시킨 모형의 최종 성능을 평가하고 알고리즘이 실제 상황에서 얼마나 잘 수행되며 사회현상을 얼마나 잘 설명하는지 알려 준다(Brownlee, 2017)[9]. 만약 학습자료 분석결과

정확도가 높게 나오는데 평가자료에서는 정확도가 낮게 나온다면 이 모형은 과적합 되었다고 볼 수 있다. 즉 학습자료는 잘 학습하여 과거의 패턴은 잘 설명하고 있지만 너무 학습자료에만 의존하고 집중한 결과, 새로운 자료가 입력되는 경우 예측력이 떨어지는 것이다.

인공지능은 데이터를 그대로 학습하기 때문에 사회에 현존하는 편향도 온전히 학습하게 된다는 것에 대해 몇몇 전문가들은 부정적인 의견을 보이기도 한다(구소희 등, 2020). 따라서 학습 자료에만 너무 의존하면 과적합이 발생하여 기존 학습자료와 상이한 새로운 자료가 입력될 시에 새로운 현상을 예측하고 분류하기 어려워질 수 있다. 학습자료와 실제 자료의 시간적 차이로 인해 인적자원관리에서 적합한 인재의 선발을 방해할 수 있다(국가정보화백서, 2018). 기계학습과 같이 복잡한 모형을 구축하는 경우 편향이 알고리즘에 함께 학습됨에 따라 인적자원관리 장면에서 평가 다양성이 부족해질 수 있다. 이로 인해 조직 내의 인력이 과도하게 동질성을 띠거나 일부 지원자 집단에 고용 기회가 제공되지 않는 현상이 발생할 수 있다(구소희 등, 2020).

일부 학자들은 인공지능을 사용한 알고리즘이 인간의 직관보다 더 나은 모형을 추정하고 인간의 편견을 제거한다고 주장한다(Brocas & Selbst, 2016). 하지만 알고리즘은 주어진 데이터, 즉 학습자료가 우수한 경우에만 그 결과도 우수할 수 있다. 보다 많은 데이터가 예측력이 높은 결과를 가져다 줄 수 있지만 이론적으로 뒷받침된변수를 선택하고 이로 인해 적절한 복잡도를 갖는 모형을 구축하는 것이 중요하다. 기계학습의 알고리즘은 의사결정자의 편견을 그대로 물려받아 학습하기 때문에 데이터가 불안정할 수 있다(Brocas & Selbst, 2016). 따라서 빅데이터

9) Datasets,『What is the Difference Between Test and Validation Datasets ?』,<https://machinelearningmastery.com/difference-testvalidtiondatasets/, > (Accessed 2021.2.4.)

자체를 온전히 받아들이기보다는 이론을 기반으로 하여 주성분 분석 등과 같은 변수 추출(feature extraction)을 통해 차원을 축소시키는 방법을 사용할 수 있다. 이론을 바탕으로 변수에 대한 조작적 정의를 사용하여 구성개념을 생성하는 측정모형을 고민해보는 방법이다. 또는 분석 예측력에 중요하게 영향을 미치는 변수인데 고용정책기본법에 위반이 되는 변수라면 해당 변수와 상관이 높은 선행변수 또는 유사변수(proxy variable)를 찾아서 대체하는 방법을 강구해볼 수 있다. 앞으로 이에 대한 연구를 진행하여 인적자원관리 장면에서 예측력을 높임과 동시에 불평등을 해소할 수 있는 분석 방법을 탐색할 수 있을 것이다.

VI. 논의

본 논문은 산업 및 조직심리 분야에서 빅데이터 분석을 사용하여 인적자원관리를 시행할 시에 얻을 수 있는 장점과 단점에 대해 알아보았다. 본 논문은 총 세 가지 연구로 이루어졌으며 첫 번째 연구에서는 빅데이터에 대해 소개하고 심리학 연구에서 빅데이터를 활용할 수 있는 가능성과 빅3데이터를 활용할 때의 주의점에 대해 논의하였다. 특히 인적자원을 고용할 때 사용되는 AI면접의 시사점에 대해 알아보았다. 그 결과 빅데이터 분석 시 데이터의 복잡도가 증가함에 따라 예측하기 어렵고 타당하지 않은 결과를 얻을 수 있는 위험성이 있었고 그로 인해 편향이 발생하여 다양성이 부족한 고용 결과를 도출할 가능성이 존재했다. 또한 표집방법을 정확하게 밝혀 빅데이터의 대표성을 확보하고 일반화 가능성을 높여야 한다고 제시했다. 그리고 빅데이터를 이루고 있는 측정치들의 품질에 대한 우려를 언급하며 측청치에 대한 신뢰도와 타당도를 입증하고 이를 뒷받침 할 수 있는 증거도 함께 제공해야 한다고 제언했다. 또한 데이터의 정확성을 판단해 의심스러운 데이터는 식별 및 수정 또는 폐기해야 한다. 마지막으로 데이터를 기반으로 하는 분석이기 때문에 반증가능성과 같은 과학적 방법에 의해 뒷받침되고 결과에 대한 해석이 신중해야 함을 강조했다.

두 번째 연구에서는 기계학습을 기반으로 하는 인공지능 알고리즘을 사용함에 있어서 변수나 데이터의 편향으로 인해 실제 데이터 분석에서 발생하는 오류를 검증하였다. 첫 번째 연구에서 언급했던 표집의 문제를 자세하게 알아보기 위해 학습자료와 실제자료가 서로 다른 모집단에서 표집된 경우 학습자료를 학습한 알고리즘으로 실제자료를 설명할 때 설명량이 감소함을 확인했다. 또한 편향된 영향을 미치는 예측변수가 학습자료에 포함되는 경우에 예측 및 분류의 오류가 증가함을 확인했다. 따라서

인적자원관리에서 사용하는 학습자료는 실제자료와의 시간 차이가 존재할 가능성이 높아 각각이 대표하는 모집단의 차이로 편향성이 발생할 위험성을 설명했다. 또한 인적자원관리의 의사결정 과정에 대해 설명할 수 없는 경우 산업 및 조직심리 분야에서 인공지능의 활용은 제한 될 수 있음을 밝혔다. 마지막으로 사회과학 분야의 연구들에서 탐색적 연구의 한계인 오류의 모형화로 인해 기계학습 결과의 일반화가 한계를 가질 수 있다는 점을 제안하였다. 인적자원관리 분야에서 수집되는 빅데이터의 업데이트는 일반적인 마케팅이나 산업공학처럼 순식간에 이루어지는 것이 아니라 연단위로 이루어지고 있다. 그렇기 때문에 학습자료와 실제자료 간의 시간지연(time-lag)이 존재할 가능성이 높다. 따라서 횡단자료와 종단자료 모두에서 편향이 발생할 수 있다. 또한 일반적으로 산업 및 조직심리 분야에서 학습데이터는 재직자들을 중심으로 자료가 구성되고 수집되는 반면에 실제 예측자료는 지원자를 중심으로 구성된다. 이는 두 번째 연구에서 모집단의 동질성이 확보 되지 못하는 경우에 해당함으로써 특히 인적자원관리 분야에서 학습자료의 중요성에 대해 강조하였다.

세 번째 연구에서는 실제 경험적 데이터를 사용하여 설명중심의 통계모형인 로지스틱 회귀분석과 예측중심의 기계학습인 랜덤 포레스트를 실시하여 결과를 비교하였다. 그 결과 랜덤 포레스트가 예측력에서는 더 효과적이었으나 결과에 대한 이론적 설명이 불가하여 해석이 모호하거나 고용정책기본법에 위반되는 변수들이 분석에 포함되기도 했다. 따라서 인적자원관리 시 평가 다양성과 공정성에 부정적 영향을 끼치지 않기 위해 이론적으로 뒷받침 될 수 있는 변수를 선택하여 분석에 포함시키고 이로 인해 적절한 복잡성을 갖는 모형을 구축하여 높은 예측력이라는 장점 또한 함께 가질 수 있는 분석방법을 제안하였다. 특히 인적자원관리 분야에서 효율적이며 정확하고 공정성에 위반되지 않는 의사결정을 내리기 위해서

기계학습 분석결과를 그대로 받아들이는 것은 위험하다고 강조하였다. 따라서 분석결과의 타당성을 뒷받침하기 위한 이론과 예측력을 높이기 위한 방대한 양의 데이터 사이에 적절한 균형이 필요하다고 제안하였다.

결론적으로 본 논문은 빅데이터 분석 시 학습자료의 중요성과 분석결과에 대한 해석 투명성에 대해 강조하였다. 즉 경험적 자료가 있어야 가설을 검증할 수 있고 이론을 바탕으로 해야 자료에 대한 해석이 용이해지는 것이다. 그러므로 다양하고 창의적인 전략을 위해 이론을 구축하는 과정에서 탐색적으로 분석하고 투명하고 공정한 의사결정을 위해 이론을 검증하는 과정에서 확인적 분석을 실시하여 이론과 데이터 사이의 순환 반복적인 방법을 정교화 시켜야 한다(Harrison & Rouse, 2015). 이론이 반드시 데이터에 선행하는 것이 아니며 그 반대의 경우도 마찬가지다(McAbee et al., 2017). 오히려 데이터 기반의 탐색적 방법을 통해 정확하고 흥미로운 이론 개발에 기여할 수 있다. 또한 확인적 방법을 통해 이론과 가설을 기반으로 의사결정에 대한 해석을 용이하게 하고 연구자의 계획된 설계를 통해서 예기치 못한 편향을 방지할 수 있다(McAbee et al., 2017). 따라서 산업 및 조직 심리 장면에서 빅데이터 분석 시 탐색적 방법과 확인적 방법을 대조적으로 생각하지 않고 서로 보완하는 것으로 고려해야 한다. 즉 탐색적 방법을 통해 빅데이터를 분석하여 구축된 이론을 확인적 방법으로 검증하는 방법을 강구해야 한다. 어느 방법이 본질적으로 결함이 있거나 부적절 한 것이 아니기 때문에 어느 하나의 방법을 희생하면서 다른 방법에 과도하게 의존하게 된다면 산업 및 조직 심리 장면에서 제한된다. 따라서 산업 및 조직 심리 분야의 미래 성장에 필요한 두 가지 접근방식의 통합이 필요하며 빅데이터 분석의 사용으로 더 큰 기회를 제공할 것이다.

인간이 생성한 데이터를 학습할 수밖에 없는 인공지능은 인간이 지닌

편견과 편향을 그대로 학습할 수밖에 없다(O'Donnell, 2019). 비록 인적자원관리 분야에서 데이터의 수집이 주로 연단위로 이루어지고 있지만, 학습자료의 편향을 줄이기 위해서는 인적자원 데이터의 보다 주기적인 업데이트가 필요할 것이다. 그리고 다양한 학습데이터를 여러 가지 활용하여 결과 비교를 통한 편향성을 확인해볼 수도 있다. 또한 알고리즘을 적용하기 전에 모형의 적합도와 타당도 검증을 먼저 실시하고 이론과 선행연구를 통해 중요한 예측변수 또는 제거해야 할 예측변수를 식별하고 데이터를 클리닝 하는 작업이 필요하다.

또한 인적자원관리 데이터의 규모는 일반적인 데이터 과학에 사용되는 규모보다 상대적으로 작은 경향이 있다(Tambe et al., 2019). 직원이 많지 않거나 충분한 성과 평가가 수행되지 않았거나 기업에서 원하는 사건들(해고했거나 애초에 불합격해서 입사하지 못한 직원에 대한 데이터)에 대한 데이터가 충분하지 않기 때문이다. 상대적으로 작은 데이터 세트는 예측력을 낮추고 과적합이 생길 위험이 존재한다(Junqué de Fortuny et al., 2013; Tambe et al., 2019). 인적자원관리 분야에서 보유한 데이터의 양이 적어서 데이터 분석에서 얻을 수 있는 정보가 적어지게 되기 때문에 관심 있는 결과변수에 영향을 미치는 예측변수를 식별하기 위해 이론과 선행연구들에 더욱 의지해야 한다(Tambe et al., 2019). 이론에 더욱 의지 한다는 것은 데이터에 기반을 둔 분석결과 뿐만 아니라 이론적 모형과 데이터가 얼마나 잘 부합하는지 즉 예측변수와 결과 변수 간에 관계성을 이론을 바탕으로 검토해보는 절차가 필요하다. 만약 해석이 불투명한 관련성이 발견된다면 설명이 뒷받침 되는 새로운 변수를 추가하거나 변수 간의 경로를 재설정 하는 방안으로 요인 또는 현상을 잘 설명할 수 있도록 모형화 해야 한다.

인공지능은 인간의 지능이 필요한 작업을 대신 수행할 수 있는 기계를

만드는 것을 목표로 한다. 인공지능은 검색, 상징적 추론 및 논리적 추론, 통계적 기술 및 행동 기반 접근법과 같은 모든 기계학습 기술로 구성된다. 기술의 발전뿐만 아니라 인간을 둘러싼 자연과 인간의 마음에 대한 이해가 발전함에 따라 인공지능을 구성하는 것에 대한 개념이 바뀌어 가고 있다. 또한 생성되고 저장되며 분석에 사용될 수 있는 디지털 정보의 양이 크게 증가함에 따라 인공지능은 점차 더 중요한 역할을 하게 될 것이다. 인공지능 시스템을 구축하는 한 가지 주요한 이유는 인간과 유사한 성과를 내기 위함이 아니라 심지어 이를 초과하기 위함이기도 하다. 어떠한 의사결정에 대해 수백 개 수 만개의 입력이 주어진 경우, 인간은 수많은 입력값(또는 변수)과 그들 간의 상호작용 사이의 복잡한 관계를 한 번에 파악하기 어렵기 때문에 보다 작은 입력값과 작은 상호작용들에 초점을 맞추게 된다. 하지만 인공지능을 통해서는 이런 복잡한 관계를 파악하는 것이 가능하기 때문에 이로 인해 비용절감, 위험관리, 의사결정향상, 생산성 향상 등 많은 이점을 얻을 수 있다. 그래서 앞으로 인공지능은 주요 혁신자로서 모바일 애플리케이션, 보안시스템, 음성인식 시스템, 금융 관련 사업, 사물 인터넷, 스마트 시티, 자동차 산업, 생물과학 등 다양한 산업을 변화시킬 것으로 예상된다.

인공지능은 규제기관과 참여자들이 선택한 어느 소수뿐만이 아니라 최대한 모든 사람에게 포용적이고 혜택을 주기위한 기술 혁명이다. 그러나 딥러닝, 랜덤포레스트, SVM(support vector machine)등과 같은 복잡한 인공지능 알고리즘을 사용하면 모형에 '블랙박스' 가 생겨 투명성이 부족해 질 수 있다(Nott, 2017). 투명성의 문제는 딥러닝이나 복잡한 모형에만 국한된 것이 아니다. 커널머신(kernel machines), 선형모형, 로지스틱 회귀모형 또는 의사결정나무와 같은 다른 모형도 차원이 많아지는 경우 해석이 어려워 질수 있다(Lipton, 2018). 이런 블랙박스 문제 때문에 알고리즘이

어떠한 결정을 내린 이유를 알 수 없으며 인공지능은 사용자가 받아들이거나 또는 무시해야 하는 답변만을 제공한다(Griffin, 2017). 금융안전위원회에 따르면 금융 부문에서 불투명한 모형을 사용하는 경우 거시적 수준의 위험을 줄 수 있는 해석의 부족 (lack of interpretability) 또는 감사가능성(auditability) 부족을 야기할 수 있다고 밝혔다. 금융안전위원회가 강조한 바와 같이 인공지능은 알고리즘의 산출물에 대한 해석 가능성과 함께 발전하는 것이 중요하다(Board, 2017).

이처럼 결과에 대한 설명가능성(explainability)은 인공지능으로부터 인간이 요구하는 중요한 사항이다(Bloomberg, 2018). 즉 인공지능의 알고리즘이 어떻게 만들어지고 어떻게 작동하는지 이해해야만 인공지능으로부터 얻은 결과를 설명할 수 있다. 만약 법정장면처럼 인간의 생사를 결정하는 의사결정의 경우 인공지능이 제공한 결과에 대한 설명가능성이 절대적으로 중요한 역할을 하게 된다. 산업 및 조직 심리 장면도 예외는 아니다. 인간의 생사는 아닐 수 있지만 공정성에 대한 민감도가 높고 결과적으로 개인과 사회에 심각한 결과를 초래할 수 있기 때문이다.

많은 연구자들은 의사결정을 위해 빅데이터를 사용하면 더 근거 있고 편견이 적은 결과를 얻을 것이라고 생각한다. 이것은 마치 만약 과학이 세상에서 행동(action)하는 것에만 관심을 가지고 스키너의 행동심리학이론처럼 입력되는 자극(stimulus)과 출력되는 반응(response)만을 연구하게 되는 것과 같다. 즉 인간의 마음, 정신, 또는 뇌 안에서 어떤 일이 일어나는지는 알 필요가 없는 것이다(Bowker, 2014). 하지만 중간에 존재하는 블랙박스에 대한 고려가 없이 과거 데이터를 사용하여 훈련된 기계학습의 예측 알고리즘은 과거 편향을 코드화 하고 편견을 증폭시킬 수 있다(Cowgill, 2019). 이런 알고리즘 편향에 대해

우려하는 학자들은 사법적인 의사결정(Angwin et al., 2016), 고용 관련 의사결정(Datta et al., 2015; Lambrecht & Tucker, 2016), 그리고 타겟광고(Sweeney, 2013) 와 관련하여 여러 가지 문제를 지적했다.

이에 따라 영국 AI 위원회(UK Parliament AI committee)는 인공지능이 우리 사회에서 통합되고 신뢰할 수 있는 도구가 되기 위해서는 지능적인 AI 시스템 개발이 필수적이라고 밝혔다(Cowgill, 2019). 기술적 투명성, 설명가능성 또는 둘 다 필요한지의 여부는 주어진 상황이나 이해관계에 따라서 다를 수 있다. 하지만 설명가능성은 시민과 소비자들에게 분명 더 유용한 접근 방식이 될 것이다. 인공지능이 내리는 결정에 대해 완전하고 만족스러운 설명을 제시할 수 없다면 인간의 삶에 상당한 영향을 미칠 수 있는 결정에 인공지능 시스템을 사용하는 것은 용납될 수 없다. 더불어 심층신경망(deep neural network)처럼 내린 결정에 대한 철저한 설명을 아직 생성할 수 없는 경우, 대체 솔루션을 찾을 때까지 특정 용도에 대한 사용을 지연 시킬 수 있다(Lords, 2018). 따라서 광범위하고 책임감 있고 신뢰할 수 있는 인공지능이 인간의 삶과 산업에 긍정적인 영향을 미칠 수 있도록 설명 가능한 인공지능(Explainable AI; XAI)의 개발이 필요하다.

영국 AI 위원회뿐만 아니라 독일총리부터 유럽 연합, 미국 평등 고용 기회 위원회(U.S. Equal Employment Opportunity Commission) 에 이르기까지 다양한 정책 담당자들은 알고리즘 편향의 방지를 목표로 하는 규정을 전면적으로 도입하기 시작했다(Cowgill, 2019). 또한 미국 국방부(US Department of Defense; DOD)도 XAI에 투자하고 있으며 새로운 기계학습 시스템은 결과에 대한 이유를 설명하고, 강점과 약점을 특성화 하고, 미래에 어떻게 행동할지에 대한 이해를 전달할 수 있는 능력을 갖추게 될 것이라 말했다(Bloomberg, 2018). 컴퓨터는 우리 삶에서 점점 더 중요한 부분이 되고 있으며 자동화는 시간이 지남에 따라 개선이 될 것이다. 따라서

복잡한 인공지능이나 기계학습 시스템이 내린 결정에 대한 이유를 아는 것은 점점 더 중요해질 것이다.

설명 가능한 인공지능(XAI)의 개념은 5가지 영역에서 주로 활발한 연구가 수행되고 있다(Wierzynski, 2018). 1) 투명성(Transparency): 인간은 인간이 이해할 수 있는 용어, 형식 및 언어로 인간에게 영향을 미치는 결정에 대한 설명을 가질 권리가 있다. 2) 인과성 (Causality): 데이터에서 모형을 학습할 수 있다면 이 모형이 올바른 추론뿐만 아니라 근본적인 현상에 대한 설명을 제공 할 수 있는가? 3) 편향성 (Bias): 인공지능 시스템이 학습자료 또는 목적함수의 단점을 기반으로 편향된 세계관을 학습했는지 또는 아닌지 어떻게 확인 할 수 있는가? 4) 공정성 (Fairness): 인공지능 시스템을 기반으로 결정이 내려진다면 그 결정이 공정하게 이루어 졌는지 확인 할 수 있는가? 5) 안정성 (Safety): 결론에 도달하는 방법에 대한 설명이 없이 인공지능 시스템의 신뢰성에 대한 확신을 얻을 수 있는가?

설명 가능한 인공지능 XAI는 첫째로 투명성 및 규정준수의 이점을 제공해야 한다. 주어진 예측과 관련된 모든 요인 및 연관성을 포함하는 감사 가능한 데이터를 제공해야 한다. 이를 통해 기업은 규정 준수 요구사항을 충족하고 조직이 숨기고 있거나 기계가 중요한 결정의 결과에 어떤 영향을 미치는지 알지 못한다는 우려를 없앨 수 있다. 두 번째로 알고리즘 결정을 공정하고 윤리적으로 방어 할 수 있는 감사 가능하고 입증 가능한 방법이 있는지 확인해야 한다.

투명성은 아무런 노력이 없이 얻을 수 있는 것이 아니다. 인공지능의 정확성과 투명성 사이에는 상충관계(trade-off)가 존재하며 이러한 상충관계는 인공지능 시스템의 내부 복잡성이 증가함에 따라 더 커질 것이다(Voosen, 2017). XAI는 높은 수준의 학습 성능 즉 정확성을 유지하면서 보다 설명 가능한 모형 즉 투명성을 생성하는 일련의 기계학습

기술을 만드는 것을 목표로 해야 한다(Holzinger et al., 2017). 또한 XAI는 결과에 대한 이론적 근거를 설명하고 인공지능의 강점과 약점을 특성화하고 미래에 어떻게 행동 할 것인지에 대한 이해를 전달할 능력이 있어야 한다. 이를 통해 인공지능이 제공한 결정을 신뢰하거나 불신할지를 결정할 수 있으며 이로서 투명성 및 인과관계를 충족하고 시스템 편향, 공정성 및 안정성 문제를 해결할 수 있을 것이다.

세계가 4차 산업 혁명으로 나아가면서 어디에나 존재하던 데이터는 이제 더욱 사용가능해지고 있다. 데이터 자체로는 구조화되지 않은 날 것의 상태이기 때문에 기업의 디지털 인프라를 어지럽히고 시스템을 저하시킬 수 있다(Skilton & Hovsepian, 2017). 하지만 이로부터 얻을 수 있는 통찰력은 계속해서 스마트 시대로 힘을 실어줄 것이다. 더 나은 제품과 서비스를 만들고 더 나은 의사결정을 내리기 위해서는 이론과 데이터 간의 상호작용을 통해 가치를 창출해야 한다. 탐색적 분석을 통해 새로운 패턴을 찾고 그 패턴이 어떤 가치를 나타내는지에 대한 조작적 정의가 필요하며 이것은 다시 이론을 기반으로 진행되어야 한다. 이로서 이론과 데이터의 순환과정에서 편향되지 않고 설명가능하며 투명하고 공정한 의사결정을 내릴 수 있을 것이다.

참 고 문 헌

고용노동부(2018). *블라인드채용 가이드북*. 세종: 고용노동부.

고용노동부(2020). *사업체 특성별 임금 분포 현황*. 세종: 고용노동부.

교육심리학용어사전(2000). *한국교육심리학회*

구소희, 조영일, & 박수진. (2020). 인적자원관리 연구에서 빅데이터의 활용. *인문사회 21, 11*(5), 1615-1630.

권영진, & 정우진. (2019). 기업의 빅데이터 투자가 기업가치에 미치는 영향 연구. *지능정보연구, 25*(2), 99-122.

김지연. (2017). 알파고 사례 연구: 인공지능의 사회적 성격. *과학기술학연구, 17*(1), 5-39.

김청택. (2019). 빅데이터를 이용한 심리학 연구 방법. *한국심리학회지: 일반, 38*(4), 519-548.

김현정. (2018). 정신건강분야에서 빅데이터 (의료이용자료) 의 활용. *한국심리학회 학술대회 자료집*, 120-120.

나준호. (2016). 인공지능의 발전과 고용의 미래. *Future Horizon,* (28), 14-17. Retrieved from 마이다스아이티 포털사이트, https://www.midasit.com/

박민정. (2014). 노인의 미충족 의료에 미치는 영향 요인: 2011 년도 한국의료패널자료를 이용하여. *Journal of The Korean Data Analysis Society, 16*(2), 1017-1030.

박정아, & 임혜빈. (2018). 빅데이터로 읽는 소비자 심리: 사례 및 분석. *한국심리학회 학술대회 자료집,* 200-200.

석현덕, 변승연, 정동열, & 김동훈. (2017). 4 차 산업혁명 기술의 임업분야

적용 및 발전방향. *한국농촌경제연구원 정책연구보고서*, 1-98.

성태제(2014). *현대 기초통계학: 이해와 적용*(제6판). 서울: 학지사.

송근배(2019.8.27). 건강한 조직문화, 공정함이 관건이다. *웰페어이슈*. Retrieved from http://www.welfareissue.com/news/articleView.html?idxno=1419

신현석, & 정용주. (2017). 제 4 차 산업혁명과 교육행정의 미래. *교육문제연구, 30*, 103-147.

양윤석, 오일석, & 강래형(2019). *R로 배우는 데이터 과학*. 서울: 한빛아카데미.

원지현(2014). *사람에 대한 데이터 분석 인재의 잠재력 살린다*. 서울: LG경제연구원.

염동기, 문상규, & 박성수. (2017). 대학졸업자의 취업성과 결정요인에 관한 실증분석. *취업진로연구, 7*(4), 45-68.

오세웅. (2017). *로지스틱 회귀모형과 랜덤포레스트를 혼합한 변수선택법에 기반하여 의사결정나무를 이용한 외래 관광객 만족도 분석* (Doctoral dissertation, 한양대학교).

유진은. (2015). *랜덤 포레스트*. 교육평가연구, 28, 427-448.

유태용, 이현준, 고윤진, 최효임, 김민경, 명민재, & 이의현. (2019). 창간 30 주년 '한국심리학회지: 산업 및 조직'내용분석 및 제언. *한국심리학회지: 산업 및 조직, 32*(3), 297-362.

이광석(2019). *4차 산업혁명 시대에서 정보인권 보호를 위한 실태조사*. 서울: 국가인권위원회.

이원갑(2019.10.7). KB 국민은행과 LH의 AI면접 '광탈' 피하기 위한 3가지 공략포인트. *뉴스투데이*. Retrieved from http://www.news2day.co.kr/138372

이종찬, & 박지현. (2015). 대학생 취업에 관한 이론적 접근과 NCS 기반

채용의 활용가능성 탐색. *취업진로연구, 5*(4), 139-160.

정혜정, & 오경화. (2016). 소셜 빅데이터를 통한 윤리소비유형, 동기와 감정 분석:“널리 인간을 이롭게 하라”. *한국심리학회지: 소비자· 광고, 17*(4), 875-893.

조성준(2019). *세상을 읽는 새로운 언어, 빅데이터*. 경기: 21세기북스.

주영재(2021). AI가 신입사원 채용 심사? “객관적” “회의적” 의견 분분. Retrieved from http://biz.khan.co.kr//khan_art_view.html?artid=201804162228005&code=920501

최기성, & 조민수. (2016). 대학 명성이 졸업생 취업 질에 미치는 효과와 시사점. *조사연구, 17*(2), 119-162.

최현주(2018.3.11). 인공지능(AI) 면접 치러보니…“표정, 목소리, 뇌파까지 분석”. *중앙일보*. Retrieved from https://news.joins.com/article/22430484

표준국어대사전(2021). (2021.4.1. 접속).

한국고용정보원(2018). *대졸자 직업이동 경로조사*. 충북: 한국고용정보원.

한국지능정보사회진흥원, 2018 국가정보화백서. 대구: 한국지능정보사회진흥원.

형인우(2021). AI가 자기소개서를 읽는다. 어떻게? 이렇게!. Retrieved from https://metnews.com/20191002000324

홍세희(2005). *이항 및 다항 로지스틱 회귀분석*. 서울: 교육과학사.

Ahmad, I., Basheri, M., Iqbal, M. J., & Rahim, A. (2018). Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. IEEE access, 6, 33789-33795.

Anderson, C. (2008). The end of theory: The data deluge makes the

scientific method obsolete. *Wired magazine, 16*(7), 16-07.

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: why HR is set to fail the big data challenge. *Human Resource Management Journal*, *26*(1), 1-11.

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 93-114). New York, NY: Routledge.

Bara, A., Simonca, I., Belciu, A., & Nedelcu, B. (2015). Exploring data in human resources big data. *Database Systems Journal*, *6*, 3-9.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. Calif. L. Rev., 104, 671.

Bassi, L. (2011). Raging debates in HR analytics. *People and Strategy, 34*(2), 14.

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods, 2*(2), 131.

Bhattacherjee, A. (2012). *Social science research: Principles, methods, and practices.*

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. the *Journal of machine Learning research, 3,* 993-1022.

Bloomberg, J. (2018). Don't Trust Artificial Intelligence? Time to Open the AI Black Box. Retrieved October, 23, 2018.

Board, F. S. (2017). Artificial intelligence and machine learning in financial services: Market developments and financial stability implications. *Financial Stability Board*, 45.

Bowker, G. C. (2014). Big data, big questions| the theory/data thing. *International Journal of Communication,* 8, 5.

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and information technology, 15*(3), 209-227.

Brantingham, P. J. (2017). The logic of data bias and its impact on place-based predictive policing. *Ohio St. J. Crim. L., 15*, 473.

Brownlee, J. (2017). What is the difference between test and validation datasets. *Machine Learning Mastery,* 14.

Caliebe, A., Leverkus, F., Antes, G., & Krawczak, M. (2019). Does big data require a methodological change in medical research?. *BMC medical research methodology, 19*(1), 1-5.

Cappelli, P., & Tavis, A. (2016). The performance management revolution. *Harvard Business Review, 94*(10), 58-67.

Cawley, G. C., & Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. The *Journal of Machine Learning Research, 11*, 2079-2107.

Chen, E. E., & Wojcik, S. P. (2016). A practical guide to big data research in psychology. *Psychological methods, 21*(4), 458.

Cheung, M. W. L., & Jak, S. (2016). Analyzing big data in psychology: A split/analyze/meta-analyze approach. *Frontiers in psychology, 7*, 738.

Cobb, A. N., Benjamin, A. J., Huang, E. S., & Kuo, P. C. (2018). Big data: More than big data sets. *Surgery, 164*(4), 640-642.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). Applied multiple regression/correlation analysis for the behavioral sciences.

Routledge.

Cowgill, B. (2019). Bias and productivity in humans and machines.

Crawford, K. (2013). The hidden biases in big data. H*arvard business review, 1*(4).

Das, N., Das, L., Rautaray, S. S., & Pandey, M. (2018). Big data analytics for medical applications. *International Journal of Modern Education and Computer Science*, *11*(2), 35.

Datta, A., Tschantz, M. C., & Datta, A. (2014). Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. arXiv preprint arXiv:1408.6491.

Devakunchari, R. (2014). Analysis on big data over the years. *International Journal of Scientific and Research Publications, 4*(1), 1-7.

Epstein, R., Robertson, R. E., Shepherd, S., & Zhang, S. (2017, June). A method for detecting bias in search rankings, with evidence of systematic bias related to the 2016 presidential election. In 97th annual meeting of the Western Psychological Association, Sacramento, CA. https://aibrt. org/downloads/EPSTEIN_et_al_2017-SUMMARY-WPA-A_Method_ for_Detecting_Bias_in_Search_Rankings. pdf.

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National science review, 1*(2), 293-314.

Fox, J. (2019). R*egression diagnostics: An introduction* (Vol. 79). Sage publications.

Gartner. (2013). Gartner IT glossary: big data. Gartner IT Glossary:

What is Big Data?.

Ghislieri, C., Molino, M., & Cortese, C. G. (2018). Work and organizational psychology looks at the fourth industrial revolution: how to support workers and organizations?. *Frontiers in psychology, 9,* 2365.

Gao, X., Wen, J., & Zhang, C. (2019). An improved random forest algorithm for predicting employee turnover. *Mathematical Problems in Engineering*, *2019*.

Gonzalez, M. F., Capman, J. F., Oswald, F. L., Theys, E. R., & Tomczak, D. L. (2019). "Where's the IO?" Artificial intelligence and machine learning in talent management systems. *Personnel Assessment and Decisions*, *5*(3), 5.

Griffin, A. (2017). Facebook's AI Creating Its Own Language Is More Normal than People Think Researchers Say. The Independent, 3.

Guzzo, R. A., Fink, A. A., King, E., Tonidandel, S., & Landis, R. S. (2015). Big data recommendations for industrial-organizational psychology. *Industrial and Organizational Psychology, 8*(4), 491.

Hagras, H. (2018). Toward human-understandable, explainable AI. *Computer, 51*(9), 28-36.

Hales, D. (2013). Lies, damned lies and big data. Aid on the Edge of Chaos, 1.

Harrison, S. H., & Rouse, E. D. (2015). An inductive study of feedback interactions over the course of creative projects. *Academy of Management Journal, 58*(2), 375-404.

Harlow, L. L., & Oswald, F. L. (2016). Big data in psychology:

Introduction to the special issue. *Psychological Methods, 21*(4), 447.

Henderson, H. V., & Velleman, P. F. (1981). Building multiple regression models interactively. *Biometrics,* 391-411.

Holzinger, A., Plass, M., Holzinger, K., Crisan, G. C., Pintea, C. M., & Palade, V. (2017). A glass-box interactive machine learning approach for solving NP-hard problems with the human-in-the-loop. arXiv preprint arXiv:1708.01104.

Hooker, J., & Kim, T. W. (2019). Ethical implications of the 4th Industrial Revolution for business and society. Business Ethics (Business and Society 360, Vol. 3. Bingley: Emerald Publishing Limited.

Huselid, M. A. (2018). The science and practice of workforce analytics: Introduction to the HRM special issue.

Houser, K. A. (2019). Can AI Solve the Diversity Problem in the Tech Industry: Mitigating Noise and Bias in Employment Decision-Making. Stan. Tech. L. Rev., 22, 290.

Huberty, C. J. (1989). Problems with stepwise methods-better alternatives. *Advances in social science methodology, 1*, 43-70.

Iliev, R., Dehghani, M., & Sagi, E. (2015). Automated text analysis in psychology: Methods, applications, and future developments. *Language and cognition, 7*(2), 265-290.

Imai, K., & Tingley, D. (2012). A statistical method for empirical testing of competing theories. *American Journal of Political Science, 56*(1), 218-236.

Indira, M. D., & Kumar, R. K. (2016). Profile screening and recommending using natural language processing (NLP) and leverage Hadoop framework for Bigdata. *International Journal of Computer Science and Information Security (IJCSIS)*, *14*(6).

Jia, Q., Guo, Y., Li, R., Li, Y., & Chen, Y. (2018, December). A conceptual artificial intelligence application framework in human resource management. In Proceedings of the International Conference on Electronic Business (pp. 106-114).

Junqué de Fortuny, E., Martens, D., & Provost, F. (2013). Predictive modeling with big data: is bigger really better?. *Big Data*, *1*(4), 215-226.

Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In Handbook of the fundamentals of financial decision making: Part I (pp. 99-127).

Kerlinger, F. N., Lee, H. B., & Bhanthumnavin, D. (2000). Foundations of Behavioral Research: The Most Sustainable Popular Textbook By Kerlinger & Lee (2000).

Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological methods, 21*(4), 493.

Kuonen, D. (2004). Data mining and Statistics: What is the connection?. *The Data Administration Newsletter*, *30*, 1-6.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes, 25*(2-3), 259-284.

Laney, D. (2001). 3D data management: Controlling data volume, velocity

and variety. *META group research note, 6*(70), 1.

Leslie, W. J. (2019). Zaire: Continuity and political change in an oppressive state. Routledge.

Lewis, M. (2007). Stepwise versus Hierarchical Regression: Pros and Cons. Online Submission.

Lipton, Z. C. (2018). The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue, 16*(3), 31-57.

Liu, M., Wang, M., Wang, J., & Li, D. (2013). Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. Sensors and Actuators B: *Chemical, 177*, 970-980.

Lords, H. O. (2018). AI in the UK: ready, willing and able?.

Lowry, S., & Macpherson, G. (1988). A blot on the profession. *British medical journal* (Clinical research ed.), 296(6623), 657.

Lum, K., & Isaac, W. (2016). To predict and serve?. *Significance*, *13*(5), 14-19.

Maass, W., Parsons, J., Purao, S., Storey, V. C., & Woo, C. (2018). Data-driven meets theory-driven research in the era of big data: opportunities and challenges for information systems research. *Journal of the Association for Information Systems, 19*(12), 1.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological bulletin, 111*(3), 490.

Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

Mayo, M(2017). Is Regression Analysis really Machine Learning? http://www.kdnuggets.com/2017/06/regression-analysis-really-machine-learning.html (2021년 5월 5일 검색)

Mazzocchi, F. (2015). Could Big Data be the end of theory in science? A few remarks on the epistemology of data driven science. *EMBO reports, 16*(10), 1250-1255.

Mazzocchi, F. On Big Data: How should we make sense of them?. *Mètode Science Studies Journal-Annual Review*, (11).

McAbee, S. T., Landis, R. S., & Burke, M. I. (2017). Inductive reasoning: The promise of big data. *Human Resource Management Review, 27*(2), 277-290.

Mehlberg, H. (1966). The theoretical and empirical aspects of science. In Studies in Logic and the Foundations of Mathematics (Vol. 44, pp. 275-284). *Elsevier.*

Noe, R. A., Hollenbeck, J. R., Gerhart, B., & Wright, P. M. (2006). Employee separation and retention. Human resource management: Gaining a competitive advantage, 5, 425-456.

Nott, G. (2017). Explainable Artificial Intelligence: Cracking Open the Black Box of AI. Computer world, 4.

O'Donnell, R. M. (2019). Challenging racist predictive policing algorithms under the equal protection clause. *NYUL Rev., 94*, 544.

Olsen, L., Aisner, D., & McGinnis, J. M. (2007). Institute of Medicine

(US). Roundtable on Evidence-Based Medicine. The learning healthcare system: workshop summary. *Washington National Academies Pr*.

Orcan, F. (2018). Exploratory and confirmatory factor analysis: Which one to use first. Journal of Measurement and Evaluation in *Education and Psychology, 9*(4), 414-421.

Oswald, F. L. (2020). Future research directions for big data in psychology.

Overton, W. F. (2007). Developmental psychology: Philosophy, concepts, methodology. Handbook of child psychology, 1.

Pariser, E., *(The)filter bubble*, 『생각 조종자들: 당신의 의사결정을 설계하는 위험한 집단』, 이현숙, 이정태 역, 시공사, 2011.

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... & Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of personality and social psychology*, *108*(6), 934.

Popper, K. (2005). The logic of scientific discovery. Routledge.

Popper, K. R., & Keuth, H. (2005). Logik der forschung (Vol. 11). Tübingen: Mohr Siebeck.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

Sajjadani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., & Mykerezi, E. (2019). Using machine learning to translate applicant work history

into predictors of performance and turnover. Journal of Applied Psychology, 104, 1207-1225. http://dx.doi.org/10.1037/apl0000405

Scholz, T. M. (2017). *Big data in organizations and the role of human resource management: A complex systems theory-based conceptualization*. Frankfurt a. M.: Peter Lang International Academic Publishers.

Schwab, K. (2017). The fourth industrial revolution. Currency.

Shah, A (2016) Wha's the Big Data? https://whatsthebigdata.com/2016/12/13/statistics-and-machine-learning/(2021년 5월 5일 검색)

Shapiro, A. (2017). Reform predictive policing. *Nature news*, *541*(7638), 458.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research, 70*, 263-286.

Skilton, M., & Hovsepian, F. (2017). The 4th industrial revolution: Responding to the impact of artificial intelligence on business. Springer.

Stephan, M., Brown, D., & Erickson, R. (2017). Talent acquisition: Enter the cognitive recruiter. Haettu, 5, 2018.

Suhr, D. D. (2006). Exploratory or confirmatory factor analysis?.

Sweeney, L. (2013). Discrimination in online ad delivery. *Communications of the ACM, 56*(5), 44-54.

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward.

California Management Review, 61(4), 15-42.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology, 29*(1), 24-54.

Thanh Noi, P., & Kappas, M. (2018). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors, 18*(1), 18.

Thomson Reuters (2020). COVID Creates Conflict Over English School Leavers' Results. https://www.usnews.com/news/world/articles/2020-08-13/covid-creates-conflict-over-english-school-leavers-results(2121.05.14접속)

Voosen, P. (2017). How AI detectives are cracking open the black box of deep learning. *Science.*

Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019, May). Designing theory-driven user-centric explainable AI. In Proceedings of the 2019 CHI conference on human factors in computing systems (pp. 1-15).

Wang, W., & Siau, K. (2019). Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: a review and research agenda. *Journal of Database Management (JDM), 30*(1), 61-79.

Weiss, S. M., Indurkhya, N., Zhang, T., & Damerau, F. (2010). Text mining: predictive methods for analyzing unstructured information. Springer Science & Business Media.

Wierzynski, C. The Challenges and Opportunities of Explainable AI, 2018.

Williams, B. A., Brooks, C. F., & Shmargad, Y. (2018). How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications. *Journal of Information Policy*, *8*, 78-115.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. Perspectives on *Psychological Science, 12*(6), 1100-1122.

Ying, X. (2019, February). An overview of overfitting and its solutions. *In Journal of Physics: Conference Series* (Vol. 1168, No. 2, p. 022022). IOP Publishing.

Zemčík, T. (2021). Failure of chatbot Tay was evil, ugliness and uselessness in its nature or do we judge it through cognitive shortcuts and biases?. *AI & SOCIETY*, *36*(1), 361-367.

ABSTRACT

Identifying and Processing biases in Big Data Analysis in the field of Human Resource Management(HRM) : Focusing on the importance of training set and theoretical background

Sohee Koo
Department of Psychology
The Graduate school of
Sungshin University

This paper presents the pros and cons of the traditional statistical analysis and machine learning algorithm using big data in the field of social science especially in the industrial and organizational psychology.

This paper is composed of three studies; the fist study introduces the concept of big data and discusses the advantages and disadvantages of AI interviews when hiring and for job placement. Subsequently the second study identified prediction and classification errors that occur during the use of machine learning based on big data analysis when biased trained data sets are used. Discussion points about being careful

not to adversely affect the employment decision due to bias and discrimination that arises were suggested. Finally in the third study, results of the explanation-centered statistical model and prediction-centered machine learning were compared using the empirical data. As a result, random forest, which uses predictive-centered machine learning algorithm, showd better predictive power and classification accuracy than logistic regression analysis, which is a explanation-centred statistical model. But the interpretation of the results was ambiguous or variables that violate the employment policy law were included in the analysis.

In the era of the Fourth Industrial Revolution, where the role of artificial intelligence using big data becomes increasingly important, what is needed the most is not only the insight and predictive power gained from vast amount of data but also the explainability and transparency of the results based on causality. This should enable fair and unbiased decision making with guaranteed stability.

Key Words: Big Data, Industrial and Organizational Psychology, Human Resource Management, Artificial Intelligence, Logistic Regression Analysis, Random Forest, Explainable AI(XAI)