

Developing a Machine-learning Algorithm for Classifying Parents' Responses to Web-based Early Language Screening

Ji Young Na^a, Gyung-Hun Hong^a, Jung-Yeon Kim^b

^aDepartment of Communication Disorders, Korea Nazarene University, Cheonan, Korea

^bICT Convergence Research Center, Soonchunhyang University, Asan, Korea

Correspondence: Jung-Yeon Kim, PhD
ICT Convergence Research Center, Soonchunhyang University, 22 Soonchunhyang-ro, Asan 31538, Korea
Tel: +82-41-530-4776
Fax: +82-41-530-1596
E-mail: betterwithme@sch.ac.kr

Received: October 15, 2024
Revised: November 16, 2024
Accepted: November 18, 2024

This research was supported by the National Research Foundation of Korea (NRF), funded by the government (Ministry of Science and ICT) in 2023 (RS-2023-00218176).

Objectives: This study aims to address the issue of decreased reliability caused by inaccurate responses in web-based early language screening tests for infants and toddlers by proposing a method for detecting the sincerity of parental responses using a machine-learning algorithm. **Methods:** The language development screening test used in this study was developed so that parents could respond directly online, and a total of 453 parents participated in the screening test. Additionally, 226 randomly generated responses were created using a random number generator. According to the criteria set by the researchers, the response types were categorized into sufficient effort responses, insufficient effort responses, and randomly generated responses. Various machine learning algorithms, including XGBoost, LightGBM, and MLP, were used to train the response data, and the accuracy of the model was assessed using K-fold cross-validation. **Results:** Based on the established criteria, the responses were classified as 220 sufficient effort responses, 233 insufficient effort responses, and 226 randomly generated responses. The ensemble voting demonstrated a high accuracy of approximately 89.7% in detecting insufficient effort responses and randomly generated responses, effectively classifying parents' insufficient effort responses. **Conclusion:** The machine-learning-based approach proposed in this study can be effectively utilized not only in web-based language screening for infants and toddlers but also in detecting insufficient effort responses in self-reported surveys. This approach is expected to enhance the reliability of research data and facilitate more accurate assessments.

Keywords: Machine-learning, Web-based language screening, Early language screening, Classifying responses

발달지연이란 발달의 주된 영역인 운동, 언어, 인지, 정서 및 사회성 영역에서 한 가지 이상이 평균 연령의 또래보다 현저하게 늦는 경우를 말한다. 이 중 언어장애는 의사소통 수단(언어, 구어, 몸의 움직임 등)을 이용해 메시지를 서로 주고받는 과정에서 메시지에 담겨 있는 내용을 수용, 처리 및 전달하는 능력에 현저한 결함을 가지고 있는 상태를 일컫는다. 일반적으로 언어장애를 가진 아동들은 발달 초기에 발달지연을 나타내는 경우가 많은 것으로 보고되고 있다(Hong & Kim, 2005; Lee, 2011; Mondal et al., 2016). 국내 5세 미만의 발달지연 출현율은 약 1-3% 수준이다(Jung, Kim, &

Jung, 2001; Kim & Ha, 2014). 이러한 발달상의 지연을 전문가가 확인하는 과정 중, 초기 단계에서 이루어지는 것이 선별검사이다. 선별검사는 전문적인 추후 정밀 진단이 필요한지를 결정하는 과정으로 정상적인 발달의 범주 내에 들어가지 않는 아동을 확인하고자 하는 과정이다.

특히, 영유아기(0-3세)는 언어발달의 예후를 좌우하는 결정적인 시기(critical period)로, 이 시기에 언어발달의 적절한 평가와 조기 선별을 하는 것은 아동의 추후 장기적인 발달에 있어서 매우 중요하다(Owens, 2019; Rescorla & Dale, 2013). 조기에 발견된 언어발

달 지연은 조기 중재를 가능하게 할 수 있으므로, 영유아를 대상으로 한 언어발달 선별검사는 아동의 성장과 발달을 지원하는 중요한 도구로 자리 잡고 있다.

현재 영유아의 언어발달을 평가하기 위해 다양한 선별검사들이 사용되고 있다. 이러한 검사들은 대부분 부모 보고형 체크리스트나 전문가의 아동 관찰, 혹은 부모 면담 등을 통해 이루어진다. 예를 들어, Korean developmental screening test for infants & children (K-DST, 한국 영유아 발달선별검사)은 발달지연을 조기에 선별하기 위해 부모가 체크리스트 문항에 응답하는 형태이며, 영유아 언어발달검사(Sequenced language scale for infants, SELSI)의 경우에는 언어발달을 중심으로 전문가의 평가와 부모 면담을 통해 평가하고 있다. 부모 보고형 선별검사는 제한적이고 구조화된 치료실 내에서 전문가가 아동을 관찰하고 평가하는 직접 검사보다 해당 아동의 언어 사용 형태를 잘 알고 있는 부모가 평가한다는 점에서 아동의 언어발달에 대한 대표성이 있는 정보를 얻을 수 있다는 장점이 있다. 또한 부모가 직접 평가를 실시하므로 전문가의 서비스 개입이 최소화된다는 점에서 경제적이라고 제안하고 있는 선행 연구들도 있다(Gartstein & Marmino, 2008; Han & Lim, 2018).

최근에는 영유아 언어발달에 대한 선별평가를 간편하게 할 수 있는 온라인 선별검사 도구들이 개발되고 있다. 이는 최근에 부모가 자녀의 발달 문제와 관련된 정보를 얻는 주요 경로로 온라인(43.8%)을 가장 많이 활용한다고(Kim, 2024) 밝힌 연구결과와 시대적 맥락을 함께 한다고 볼 수 있다. Seo와 Chang (2013)은 기존의 영아발달 선별검사 도구(Developmental assessment for the Early intervention Program planning, DEP)를 온라인 형태의 선별검사 도구로 변환하여 e-DEP를 개발하였다. 그리고 새롭게 개발된 온라인 형태의 e-DEP와 기존의 DEP의 효용성을 비교 분석한 결과, 접근성, 타당성 및 유용성 등을 포함한 전반적인 선별검사 도구에 대한 만족도가 e-DEP가 더 높은 것으로 나타났다고 보고하였다.

그러나 온라인에서 수행되는 선별검사에는 응답 결과 해석을 왜곡시킬 수 있는 참여자들의 ‘응답 성실도’라는 새로운 도전 과제도 존재한다. 온라인 검사의 특성상 양육자의 응답이 부정확하거나 성실하지 않을 수 있으며, 응답자가 양육자인가에 대한 사실 여부의 확인이 어렵다는 점에서도 검사 결과의 신뢰도에 영향을 미칠 수 있다. 특히 온라인 언어 선별검사에서는 부모가 아동 발달과 관련된 용어나 질문의 의미를 완벽히 이해하지 못해 부적절한 답변을 하거나, 답변의 일관성이 부족해 보이는 경우가 많으며, 이러한 오타나 혼동이 불성실한 답변으로 간주될 가능성이 높다. 즉, 고의로 불성실하게 온라인상에서 응답하는 경우만이 아니라, 어떤 이유이든 적절하지 못한 응답을 하는 경우는 온라인 형태의 설문 결과의

분석과 해석을 왜곡할 수 있는 것이다. 온라인 형태의 설문 결과를 분석하는 최근 연구들에서는 설문에 답한 응답자들이 주의와 노력을 기울이지 않고 설문에 답한 형태의 응답을 ‘불성실 응답’(careless responding 또는 insufficient effort responding)으로 개념을 정립하고 있다(DeSimone, DeSimone, Harms, & Wood, 2018; Huang & Wang, 2021; Ward & Meade, 2018). 이러한 불성실 응답의 형태는 주로 1) 설문을 통해 파악하고자 하는 현상과 연관이 없거나 거리가 먼 응답과, 2) 충분한 노력을 기울이지 않은 응답 또는 부주의한 응답으로 구분해 볼 수 있다.

설문 연구와 같이 연구 참여자들의 응답을 분석하는 연구에서는 분석 대상 응답에서 이러한 불성실 응답을 구분하여 제외하지 않고 분석에 활용할 경우, 연구결과가 현실을 반영하지 못하고 왜곡된 결과가 도출될 위험이 있다고 알려져 있다(Huang, Liu, & Bowling, 2015). Huang, Liu 등(2015)은 불성실 응답이 설문 데이터에 미치는 영향을 검증하기 위하여, 성격 설문지, 시뮬레이션 및 직원 표본을 사용한 세 가지 실험을 통해 불성실 응답의 영향을 분석하였다. 실험에 사용한 척도의 평균이 중간점에서 벗어날 때 불성실 응답이 관찰된 상관관계를 증가시킬 수 있음을 보여줌으로써 불성실 응답의 존재가 어떻게 결과를 왜곡시키는지 보여주었다.

이러한 응답 성실도의 문제를 해결하기 위한 방안 중 하나로, 온라인 설문 응답의 성실도를 자동으로 탐지하고 분석할 수 있는 시스템을 개발하는 방법이 있다. 불성실한 응답을 탐지하기 위한 전통적인 방법은 크게 세 가지로 구분할 수 있다. 첫번째로는, 설문에서 불성실한 응답으로 구분해 내기 위한 문항을 포함하여 해당 문항에 응답을 기준으로 판단하는 ‘직접적 방법’(Huang, Bowling, Liu, & Li, 2015; Oppenheimer, Meyvis, & Davidenko, 2009)이며, 두번째로는, 설문 과정에서 수집된 응답시간(DeSimone & Harms, 2018) 또는 참여자의 응답패턴(Dunn, Heggstad, Shanock, & Theilgard, 2018; Meade & Craig, 2012)과 같은 정보와 기초적인 통계 기법을 활용하는 ‘비개입적 방법’이 있다. 마지막으로, 성실한 응답을 기준으로 좀 더 복잡한 통계 방법을 사용하여 응답 분포의 이상치를 분석하여 불성실한 응답을 탐지하는 ‘통계적 방법’(Curran, 2016; Meade & Craig, 2012)이 있다.

최근에는 이러한 전통적인 방법과 함께 기계학습과 같은 인공지능 기술을 활용하여 불성실한 응답을 탐지하려는 연구들이 발표되고 있다. 기계학습은 지도학습과 비지도학습으로 구분할 수 있는데, 지도학습은 입력 데이터와 그 입력 데이터의 정답(또는 클래스 레이블)이 같이 주어지며, 기계학습 모델이 입력 데이터와 정답 간의 관계를 학습하여 새로운 데이터에 대한 클래스 예측 또는 분류 등의 문제를 해결하는 방법이다. 반면에, 비지도학습 방법은 정답

에 대한 정보가 없는 입력 데이터에서 패턴이나 특정 구조를 찾아내는 방법으로, 데이터의 분포나 군집(또는 클러스터링)을 분석하는데 주로 사용하는 학습 방법이다(Bishop, 2009). Ozaki (2024)는 불성실 응답을 탐지하기 위한 효과적인 방법을 파악하기 위하여 설문문항에 대한 응답과 설문 응답에 소요된 시간을 기계학습 모델 훈련을 위한 입력 데이터로 활용하였으며, 기계학습 모델과 통계적 방법 간의 불성실 응답 탐지 성능을 검증하였다. 연구결과, 통계를 활용한 방법과 비교하여 기계학습 기반의 불성실 응답 탐지 방법의 성능이 부분적으로 더 좋은 것을 확인하였다. Kang, Choi와 Kwon (2020)은 기계학습 모델 가운데 인간의 뇌 신경망 구조를 모방한 인공신경망(Artificial Neural Network, ANN)과 데이터를 고차원 공간으로 변환하여 최적의 결정 경계를 찾는 방법인 서포트 벡터 머신(Support Vector Machine, SVM) 같은 지도학습 모델을 활용하여 설문 응답 중에서 불성실 응답을 탐지하고자 하였는데, Ozaki (2024)의 연구와 마찬가지로 설문에 대한 응답 외에 설문 응답에 소요된 시간을 변수로 사용하였다. 이외에도, 응답자의 성별, 연령, 응답시간, 설문에 응답할 때 사용한 장치 유형(PC 또는 Mobile) 및 부정응답 횟수와 같은 변수를 기계학습 모델 훈련을 위한 추가 데이터로 사용하였다. 또한, 연속된 문항에 동일한 응답을 일관되게 유지한 응답 유형 및 시간이 촉박한 상황으로 인해 질문에 대한 고민없이 성급하게 설문조사에 임하는 참여자의 상황을 가정하여 불성실 응답을 생성하고, 이를 설문 참여자가 성실하게 응답한 사례와 함께 분석하여 불성실 응답 판별을 위한 모형을 검증하였다. 이에 반해 Jebreel 등(2020)은 클러스터링 및 이상치 검출 성능, 그리고 실행시간과 같은 성능지표를 기준으로 우수한 성능을 가지는 것으로 알려진 비지도학습 모델, Density-based spatial clustering of application with noise (DBSCAN), Isolation forest (IForest)를 활용하여 불성실 응답 탐지 방법을 검증하였다. DBSCAN과 IForest는 모두 데이터의 분포나 밀도를 기반으로 이상치나 특이점을 찾는 비지도 학습 알고리즘이다. 해당 연구에서는 응답자로부터 총 15개의 문항(인적사항과 관련된 범주형 문항 9개 및 개인소득과 관련된 연속형 문항 6개)에 대한 응답을 수집하고, 이로부터 20%의 응답을 무작위로 선택하여, 15개 문항 중 6개 문항을 무작위 응답으로 대체한 불성실 응답 데이터(부분 불성실 응답 데이터)와 15개 문항 모두를 무작위 응답으로 대체한 불성실 응답 데이터(전체 불성실 응답 데이터)를 사용하여 성실 응답과 불성실 응답 판별 모델을 개발하였다.

이와 같이, 지도학습 또는 비지도학습 모델을 활용하여 연구 참여자들의 불성실 응답을 판별하기 위하여 다양한 방법을 적용한 연구들이 수행되고 있다. 기계학습 기술을 활용한 응답 성실도 탐

지 시스템은 검사 결과의 정확성을 높이고, 영유아의 언어발달 상태를 보다 신뢰성 있게 평가하는 데 기여할 수 있을 것으로 기대된다. 하지만, 설문을 위한 문항의 수, 응답자의 성향 및 불성실 응답 패턴 등 불성실 응답 판별 방법 개발에 영향을 미칠 수 있는 요인의 다양성에 비해 불성실 응답 판별방법 개발 관련 연구의 수는 제한적이므로 추가적인 연구가 필요한 실정이다. 또한, 불성실한 응답은 설문조사의 내용과 관련 없는 무작위 형태의 응답을 임의로 제공하는 ‘무작위 불성실 응답’뿐만 아니라, 설문조사의 내용과는 관련 없이 똑같은 응답을 반복하거나 특정 패턴의 응답을 반복하는 ‘비무작위 불성실 응답’과 같은 다양한 유형으로 나타날 수 있지만(Bowling & Huang, 2018; Kam & Meyer, 2015), 기존의 관련 연구들은 ‘무작위 불성실 응답’을 파악하기 위한 방법들을 주로 연구하였다. 그러나, 자기보고식 설문조사 응답 표본에 불성실 응답이 차지하는 비율이 아주 낮더라도 연구결과에 미치는 영향이 적지 않으므로(Huang, Liu et al., 2015), ‘비무작위 불성실 응답’도 불성실 응답의 유형으로 포함하여 ‘무작위’와 ‘비무작위’로 판별할 수 있는 방법 개발의 중요성이 부각되고 있다.

따라서 본 연구의 목적은 온라인을 활용한 비대면 언어발달 선별검사에 적용하기 용이한 기계학습 기반 불성실 응답 탐지 모델을 제안하는 데 있다. 이를 위하여, 전문가 기준에 따른 응답 성실도와 기계학습 모델을 활용한 응답 성실도 결과를 비교하는 방법을 활용하였다. 구체적으로는, 온라인 선별검사에서 불성실한 응답을 탐지하기 위하여 불성실 응답을 두 가지 유형(‘무작위 응답’과 ‘비무작위 응답’)으로 구분하여 판별하고, 기계학습 모델 중 지도 학습 방법을 활용하여 정확도가 높은 불성실 응답 탐지 방법을 개발하는 것을 목표로 하였다. 이러한 연구목적을 위한 본 연구의 구체적인 연구문제는 다음과 같다. 1) 온라인 설문지(Google 설문지)를 활용한 영유아 대상 비대면 온라인 언어발달 선별검사의 응답에서 전문가 기준에 따른 성실 응답과 불성실 응답(‘무작위 응답’과 ‘비무작위 응답’)의 비율은 어떠한가? 2) 기계학습을 기반으로 한 응답 성실도 탐지 방법은 성실 응답과 불성실 응답(‘무작위 응답’과 ‘비무작위 응답’)을 적절하게 탐지하는가?

연구방법

연구대상

본 연구의 대상자는 생후 0-36개월 영유아 자녀를 둔 부모 453명이었다. 설문 응답의 대상이 된 영유아는 연령별로 0-5개월 65명, 6-11개월 92명, 12-17개월 49명, 18-23개월 37명, 24-29개월 79명, 30-36개월 131명이었다. 전체 453명 아동 중 남자는 211명, 여자는

242명이었다.

본 연구에 참여한 부모의 학력은 고등학교 졸업이 24명, 대학 졸업이 387명, 대학원 이상 졸업이 42명이었다. 부모의 성별은 아동의 아버지가 76명, 어머니가 377명이고, 부모의 연령대는 20대가 26명, 30대가 382명, 40대가 45명이었다. 연구 참여자의 주거지는 서울이 191명, 수도권이 85명, 그 외 지역이 177명이었다. 이와 같은 연구 참여자 정보는 Table 1과 같다.

연구도구

본 연구에서는 온라인 선별검사 문항의 개발을 위하여, 발달지연

을 조기에 선별하기 위한 목적으로 사용되는 Korean child development inventory (K-CDI, 아동발달검사), Korean children development review-revised (KCDDR-R, 영유아 발달선별검사) 등과 언어발달을 중점으로 평가하는 Korean adaptation of the LENATM development snapshot (K-SNAP, 영유아 언어 및 의사소통 발달 선별검사), SELSI를 참고하였다. 또한 어휘발달, 언어발달 그리고 표준화 예비연구와 관련한 선행연구를 고려해 문항을 설계하였다. 본 연구에 활용된 언어 선별검사 문항들은 0-36개월 영유아의 언어발달을 평가할 수 있도록 구성되었다. 각 문항은 부모가 직접 관찰한 영유아의 행동을 기반으로 응답할 수 있도록 구성하였다. 최종 선별검사의 문항 수는 총 173개였으며, Table 2와 같이 연령 집단을 3-5개월 간격으로 구분하여 연령 집단별로 언어인지, 언어이해, 언어표현 세 가지 영역의 문항 11-25개로 구성하였다. 예를 들어, 9-11개월의 언어인지 영역에 ‘성인이 보거나 가리키는 물건을 향해 아이가 함께 바라본다.’, 12-14개월의 언어이해 영역에 ‘동사를 포함한 2 낱말의 문장을 이해한다(예: 00이가 빠, 아빠한테 앉아 등)’, 33-36개월의 언어표현 영역에 ‘다음 예시 중 5개 이상의 동사를 산출한다(예: 고쳐, 두드려, 떼, 뛰어, 가, 그려, 꺼내)’와 같은 문항이 포함되어 있다.

개발된 모든 문항에 대해 0-36개월 영유아의 언어발달을 선별하기에 타당한지 언어치료학과의 교수 5인을 대상으로 타당도를 검증한 결과, 5점 만점에 전체 평균 4.4점으로 영유아 언어발달을 선별하기에 적절한 것으로 나타났다.

온라인 선별검사의 구성은 아동과 응답자인 부모의 기본정보와 개인정보 수집 및 활용에 동의서, 최종 선정된 선별검사 문항, 그리고 원격 선별검사에 대한 만족도 문항의 순으로 Google 설문지를 활용하여 구현하였다. Google 설문지의 자동화 기능을 활용하여, 응답자가 입력한 아동의 월령에 따라 아동의 연령대에 맞는 각기 다른 문항 세트가 자동으로 응답자에게 제공되었으며, 모든 설문 응답은 실시간으로 수집되도록 설정하였다. 또한, 설문지에는 응답자의 편의를 고려하여 설문을 완료하기까지 남은 문항의 비율을 ‘진행표시줄(progress-bar)’ 형태로 표시하여, 응답자가 설문의 총 길이와 남은 문항의 양을 시각적으로 쉽게 파악할 수 있도록 했다.

Table 1. Characteristics of study participants

Demographic information	n	%
Child's age (month)		
0-5	65	14
6-11	92	20
12-17	49	11
18-23	37	8
24-29	79	18
30-36	131	29
Child's sex		
Male	211	47
Female	242	53
Caregiver's highest education level		
High school graduate or below	24	5
College graduate	387	85
Above college graduate	42	10
Relationship with the child		
Father	76	17
Mother	377	83
Caregiver's age range		
20 s	26	6
30 s	382	84
40 s	45	10
Residence		
Seoul	191	42
Metropolitan city	85	19
Other regions	177	39
Total	453	100

Table 2. Number of items by domain

Language domain	Age group of the item											
	0-5	6-8	9-11	12-14	15-17	18-20	21-23	24-26	27-29	30-32	33-36	Total
Cognitive linguistics	6	9	8	6	5	5	4	4	3	3	3	56
Language comprehension	3	3	4	5	5	5	5	5	5	6	6	52
Language expression	2	3	4	3	6	6	8	6	6	5	16	65
Total	11	15	16	14	16	16	17	15	14	14	25	173

이를 통해 응답자의 참여 부담감을 줄이고, 중도 포기를 방지해 높은 응답률을 유도하였다.

이처럼 구현된 Google 설문지를 통하여, 연구 참여 부모의 자녀 월령에 따라 제공된 문항은 Table 3과 같다. 연구 참여자는 1) 참여 아동의 월령대에 대부분 습득한다고 알려진 문항, 2) 현재 월령보다 6-12개월 더 어린 월령대에서 주로 습득하는 문항, 그리고 3) 아동의 월령보다 6-9개월 높은 월령대에서 주로 습득하는 문항에 대하여 설문 응답을 제공하였다. 예를 들어, 연구 참여자의 자녀가 18-23개월 월령 집단에 속한다면, 해당 월령의 문항들뿐만(18-23개월) 아니라 9-17개월에 대부분의 아동이 습득하는 문항들과 24-29개월에 주로 습득하는 문항들도 함께 응답하도록 하였다. 이러한 예시는 Table 3에서 굵은 글씨체로(bold) 표기되었다. 예외적으로, 월령이 5개월 이하인 경우에는 아동의 월령보다 6-12개월 더 어린 월령대에서 주로 습득하는 문항이 제공되지 않았으며, 반대로 월령이 30개월보다 높은 경우에는 아동의 월령보다 6-9개월 높은 월령대에서 주로 습득하는 문항이 제공되지 않았다.

연구절차

설문 데이터 수집

온라인 선별검사는 2024년 4월부터 7월까지 실시하였다. 천안, 아산, 청주 지역 내 어린이집과 문화센터의 이메일로 본 연구의 목적 및 내용 설명과 Google 설문지로 연결되는 QR 코드가 포함된 안내문을 배포하였다. 또한 경기도 소재의 어린이집을 통해 어린이 집 협의회와 동의를 얻어 SNS로 배포하는 방법을 통하여 연구 참여자를 모집하였다. 누락 데이터(missing data)를 방지하기 위하여 Google 설문지 시스템상에서 중간에 중단하거나 문항을 모두 응답하지 않은 경우에는 최종 데이터에 포함되지 않도록 하였다. 참여자의 응답 결과 분석 시, 응답자의 자녀인 아동의 언어발달 지연이 의심되는 경우에는 언어재활사인 연구자가 별도로 응답자에게 표준화된 선별검사인 국민건강보험 사이트에서 제공하는 K-DST를 추가로 실시해보도록 권유하는 안내문을 전달하였다. 또한, 연

구에 참여한 부모에 대한 보상 및 응답률 향상을 위하여 선별검사를 마지막 문항까지 완료한 참여자에게는 소정의 기프티콘을 제공하였다.

데이터 처리

불성실 응답 탐지를 위한 기계학습 분류기 훈련을 위해서 응답의 유형을 성실 응답, 비무작위 불성실 응답, 무작위 불성실 응답으로 구분하였으며(Bowling & Huang, 2018; Kam & Meyer, 2015), 훈련 데이터 셋(설문응답 679건) 구축을 위한 절차는 응답 데이터 레이블링 부분에 자세하게 기술하였다.

응답 데이터 레이블링

응답 유형 구분에 앞서, 온라인 선별검사 문항에 대한 응답을 ‘예’라고 답한 경우 1점, ‘아니오’라고 답한 경우 0점을 부여해 점수화하였다. 본 연구에서는 선별검사 문항에 대한 응답을 1) 성실 응답, 2) 비무작위 불성실 응답, 3) 무작위 불성실 응답, 총 세 가지 유형으로 구분하였다. 불성실 응답은 월령에 따른 언어발달과 상관없이 같은 응답을 반복하거나, 일정한 패턴을 보이는 불성실 응답인 비무작위 불성실 응답과 특정 패턴을 찾아볼 수 없는 무작위 형태의 응답으로 구분된다.

먼저, 선별검사 응답 데이터 중, 성실 응답과 불성실 응답을 구분하기 위하여, 언어치료 분야의 박사학위 소지자이자 최소 10년 이상의 임상 및 연구 경력을 가진 2명의 연구자들이 논의를 통해 구분 기준을 도출하였다. 불성실 응답의 기준은 응답자의 자녀 월령 집단을 기준으로, 더 어린 월령대 문항에서 ‘아니오’가 50% 이상인 경우와 더 높은 월령대 문항에서 ‘예’가 50% 이상인 경우로 간주하였다. 이 불성실 응답의 기준은 발달 단계와 기대되는 반응 패턴을 고려한 것이다. 보통 아동이 연령에 맞는 질문에 답할 때는 그 연령대에서 경험하거나 이해할 수 있는 수준의 항목에 “예”라고 응답하는 경향이 있으며, 반대로 나이가 어리거나 발달적으로 적절하지 않은 질문에 “아니오”라고 답할 확률이 높다. 따라서, 해당 연령대

Table 3. Screening items presented by age group

Age group	Age group of the item											Number of items
	0-5	6-8	9-11	12-14	15-17	18-20	21-23	24-26	27-29	30-32	33-36	
0-5	○	○	○									42
6-11	○	○	○	○	○							72
12-17		○	○	○	○	○	○	○				109
18-23			○	○	○	○	○	○	○			108
24-29						○	○	○	○	○	○	101
30-36						○	○	○	○	○	○	101

Parents who have children in 18-23 month range answered 108 question items (Bold in letters).

보다 어린 연령대에서 “아니오”가 50% 이상 나온 경우나 더 높은 연령대 문항에서 “예”가 50% 이상인 경우는 발달 수준과 불일치하는 반응으로 보았다. 이는 응답자가 무작위로 답변했거나 문항의 의미를 이해하지 못했을 가능성을 시사하므로, 불성실 응답의 근거로 제시될 수 있다. 즉, 이는 응답자가 일반적인 언어발달 형태를 고려할 때 자녀의 실제 월령에 부합하지 않는 응답 패턴을 보일 경우 불성실 응답으로 탐지하기 위함이다. 해당 규칙을 활용해 응답을 구분하기 위한 예를 살펴보면 다음과 같이 응답 유형을 구분할 수 있다. 20개월 된 아동을 예로 들면, 해당 아동의 월령은 18-23개월 집단에 해당되어 9-29개월 문항들이 제공되는데, 더 어린 월령대 문항인 0-17개월 문항(72문항)에서 ‘아니오’가 50%인 36개 이상이거나 더 높은 월령대 문항인 24-36개월 문항(68문항)에서 ‘예’가 50%인 34개 이상인 경우, ‘비무작위 불성실 응답’으로 분류된다.

또한, 본 연구에서는 부모가 제출한 선별 문항 응답 외에 특정 월령의 언어발달과 관련이 없으며, 특정 응답 패턴 또한 확인할 수 없는 ‘무작위 불성실 응답’ 데이터 셋을 구축하기 위하여 예측하기 어려운 일련의 숫자를 생성하는 난수 발생기(random number generator)를 활용하였다. 무작위 불성실 응답 데이터 생성은 월령을 난수 발생기로 먼저 생성한 후, 해당 월령에 해당되는 응답에 대해서 ‘예’ 또는 ‘아니오’ 응답을 무작위로 발생시켜 어떤 특정 패턴을 가지지 않는 무작위의 응답을 생성하였다. 따라서 본 연구에서 기계 학습 기술에 기반을 둔 불성실 응답 탐지 모델 개발 및 검증은 위하여 실제 부모의 응답을 기준으로 한 성실 응답 및 비무작위 불성실 응답과 난수 발생기를 활용한 무작위 불성실 응답, 총 세 가지 유형의 응답 데이터 셋을 구축하였다. 온라인 선별검사에 참여한 부모 중 응답을 완료한 453건의 응답 자료와 특정 월령에 해당하는 무작위 불성실 응답 226건을 합하여 총 679건의 응답 데이터 셋을 확보하고, 통계분석에 활용하여 연구결과를 도출하였다.

응답 데이터 결측치 전처리

본 연구에서 사용된 선별검사는 월령에 따라 응답해야 하는 문항의 종류와 수가 다르기 때문에 다양한 월령에 속한 아동의 부모가 응답한 모든 데이터를 분석에 활용하기 위해서는 결측치를 처리하는 전처리 과정이 필요하였다. 아동 월령에 따른 제시 문항 외 나머지 문항에 대한 결측치를 더 어린 월령대 문항은 모두 ‘예(1)’, 더 높은 월령대 문항은 모두 ‘아니오(0)’로 대체하였다. 예를 들어, 12-17개월 사이에 해당하는 아동을 자녀로 둔 부모의 응답에서 제시되지 않은 어린 월령대 문항인 0-5개월 문항은 모두 ‘예(1)’, 높은 월령대 문항인 27-29개월, 30-32개월 및 33-36개월 문항은 모두 ‘아니오(0)’로 대체하여 데이터 전처리를 수행하였다.

기계학습 기반 불성실 응답 탐지 모델 학습 및 검증

본 연구에서는 정형적 데이터 분석을 위해 지도학습 기계학습 모델 사용하여 정상적 응답 탐지기를 개발하고자 하였다. 이를 위하여 다양한 연구 분야에서 활용되고 있는 대표적인 지도학습 알고리즘인 K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Multilayer Perceptron (MLP), Gradient Boosting Classifier (GBC), Adaptive Boosting (AdaBoost), eXtreme Gradient Boosting (XGBoost) 및 Light Gradient-Boosting Machine (LightGBM)을 사용하였다. 불성실 응답 탐지 방법 개발을 위해 학습 데이터 셋과 검증 데이터 셋을 7:3의 비율(학습 데이터 셋: 475건, 검증 데이터 셋: 204건)로 분리한 후 학습 데이터 셋에 대하여 K-fold 교차 검증(K=10)을 실시하였으며, 데이터 셋을 분리할 때 응답 유형의 비율을 고려하여 분리하였다. 또한, K-fold 교차검증 시 정확도 기준으로 불성실 응답 탐지 성능이 가장 좋은 상위 5개 모델을 선정 후 앙상블 학습에서 각 모델의 예측을 결합하는 두 가지 방법인 Soft Voting 및 Hard Voting 방식을 적용하였고, 204건의 검증 데이터 셋을 활용하여 Soft Voting과 Hard Voting 방법에 따른 설문 응답 유형 분류 정확도를 검증하였다. 앙상블 학습은 여러 개의 기계학습 모델이 가진 개별적인 예측력을 결합하여 향상된 정확도와 안정성을 도출하는 방법이다(Zhou, 2012). Soft Voting은 각 모델이 출력하는 클래스의 확률 값에 대한 평균을 구한 후 가장 높은 확률을 가지는 클래스를 출력하지만, Hard Voting은 각 모델이 예측한 결과 중 가장 많은 모델이 예측한 클래스를 결과로 출력하는 방법이다(Zhou, 2012). 짝수 개의 모델을 사용하여 예측 결과를 결합하면, Hard Voting 방식의 경우, 두 개의 클래스를 예측한 모델의 수가 동일한 경우가 발생할 수 있고, 이는 모델의 성능 저하 및 예측 결과 신뢰성 저하의 원인이 될 수 있다(Dietterich, 2000). 따라서 예측 결과를 앙상블링하기 위한 모델의 수를 홀수로 선정하였다.

특정 클래스의 데이터가 현저히 많을 경우, 그 특정 클래스로 결과가 예측되었을 때 정확도가 높아질 수 있는 모델 과적합 문제가 발생할 수 있기 때문에, 기계학습 기반 불성실 응답 분류 모델을 학습하기 위해 사용된 데이터 셋의 데이터 불균형 여부를 확인하고자 응답 유형 별 데이터 분포를 먼저 확인하였다. 이후, 선별검사의 언어인지, 언어이해, 언어표현의 각 영역별로 문항을 나눠서 응답 유형 탐지 성능평가를 수행한 후 모든 영역의 문항을 통합하여 응답 유형 탐지 성능에 대한 모델 검증을 수행하였다. 모델 성능평가는 분류 정확도를 이용하였으며 K-fold 교차검증 시 얻어진 정확도의 평균과 표준편차로 결과를 보고하였다.

연구결과

전문가 기준 기반 선별검사 응답 유형 분포

Table 4는 월령에 따라 기대되는 언어발달 수준과 연계되어 도출된 기준에 따라 분류된 성실 응답과 비무작위 불성실 응답으로 판단된 응답 그리고 난수 발생기를 활용하여 무작위로 생성된 무작위 불성실 응답의 분포를 제시하고 있다. 본 연구에서 활용된 기준에 따라 확인된 성실 응답은 220건이며 월령 집단별로 6-11개월 사이의 자녀를 둔 부모의 응답이 가장 많은 것으로 나타났다. 전문가 기준에 따라 비무작위 불성실 응답으로 분류된 233건의 응답 중에는 30-36개월 사이의 자녀를 둔 부모의 비율이 가장 높은 것으로 나타났다. 무작위 불성실 응답은 226건으로 확인되었다. 이는 전체 응답 중 33.2%의 비율을 차지하는 것으로 파악되며, 전체 응답의 32.4%를 차지하는 성실 응답과 34.3%를 차지하는 비무작위 불성실 응답과 비교하였을 때, 무작위 불성실 응답이 유사한 비율로 생성된 것을 확인할 수 있다.

Table 4. Response classes based on the criterion developed by professionals in this study

Class	Sufficient effort responses	Non-random insufficient effort responses	Randomly generated insufficient effort responses	Total
Number of responses	220	233	226	679
Ratio (%)	32.4	34.3	33.2	100

Table 5. Classification performance obtained by training individual machine learning models

Tested models	Accuracy									
	Cognitive linguistics		Language comprehension		Language expression		Entire language domain		Performance by model	
	M	SD	M	SD	M	SD	M	SD	M	SD
KNN	.838	.045	.835	.055	.837	.045	.843	.043	.838	.003
DT	.848	.038	.845	.046	.845	.052	.855	.370	.848	.004
GBC	.860	.042	.855	.048	.860	.036	.852	.046	.857	.003
RF	.832	.047	.805	.076	.787	.049	.865	.032	.822	.029
SVM	.815	.035	.828	.049	.837	.060	.873	.032	.838	.022
MLP	.852	.032	.843	.047	.863	.048	.862	.041	.855	.008
AdaBoost	.835	.051	.823	.070	.780	.076	.800	.070	.810	.021
XGBoost	.857	.054	.852	.046	.850	.026	.852	.036	.853	.003
LightGBM	.845	.037	.858	.043	.865	.038	.870	.033	.860	.009
Performance by domain	.842	.013	.838	.016	.836	.030	.852	.021	-	-

Table 6. Classification accuracy achieved by voting methods (soft and hard voting)

Voting methods	Accuracy				
	Cognitive linguistics	Language comprehension	Language expression	Entire language domain	Mean
Soft voting	.890	.878	.866	.878	.878
Hard voting	.878	.866	.866	.915	.881
Mean	.884	.872	.866	.897	-

기계학습 기반 설문 응답 유형 탐지 성능 평가 결과

선별검사에 사용된 영역별 문항을 언어인지, 언어이해 및 언어표현으로 나누어 기계학습 모델을 학습시킨 후 응답 유형을 분류한 결과와 모든 영역의 문항을 학습에 사용하여 도출한 분류 정확도를 Table 5에 정리하였다. 언어발달 영역별로 불성실 응답 탐지 정확도를 비교하면, 전체 영역의 문항 응답을 모두 사용하였을 때 평균적으로 가장 높은 85.2% ($\pm 2.1\%$)의 정확도를 보였고, 언어인지 영역 문항에 대한 응답이 두 번째로 높은 84.2% ($\pm 1.3\%$)의 정확도를 보인 반면에 언어표현 영역의 응답을 사용하였을 때 가장 낮은 정확도(83.6% $\pm 3.0\%$)를 보였다. 전문가들의 불성실 응답 기준에 근거하여 응답 분류 정확도를 모델별로 비교할 경우, LightGBM과 GBC를 활용한 응답 유형 분류 정확도가 각각 86.0% ($\pm 0.9\%$), 85.7% ($\pm 0.3\%$)로 가장 높은 수준임을 확인할 수 있었다. 이 두 모델에 이어 높은 성능을 보이는 기계학습 알고리즘은 MLP로 85.5% ($\pm 0.8\%$)의 정확도를 보였으며, 테스트에 사용된 기계학습 모델 중 가장 낮은 정확도를 보이는 알고리즘은 AdaBoost로 81.0% (\pm

2.1%)의 불성실 응답 탐지 정확도를 보였다. 본 연구에 사용된 기계 학습 모델들의 성능을 비교하면, LightGBM 및 MLP가 모든 영역에서 불성실 응답을 탐지하는데 높은 정확도를 보이는 상위 5개 모델에 속하는 것으로 나타나 불성실 응답 탐지에 뛰어난 성능을 보이는 모델로 확인되었다.

불성실 응답 탐지 성능을 기준으로 선정된 상위 5개 모델의 예측 결과를 결합하여 각 언어 영역별로 앙상블 모델의 성능을 평가하였으며, 그 결과를 Table 6에 정리하였다. 앙상블 모델의 불성실 응답 탐지 성능은 선별검사 전 영역의 항목 모두 사용하였을 때 가장 높은 정확도(89.7%)를 보였으며, 언어인지 영역 문항의 응답을 사용하였을 때의 분류 정확도가 두 번째로 높은 88.4%였다. 언어표현 영역의 응답만을 사용하였을 때는 86%의 불성실 응답 탐지 정확도를 보여 가장 낮은 정확도를 보였다. Soft Voting과 Hard Voting 앙상블 모델 간의 평균 정확도는 각각 87.8%와 88.1%로 비슷한 불성실 응답 탐지 성능이 비슷한 것으로 나타났다.

논의 및 결론

본 연구는 온라인 언어발달 선별검사서에서 응답자의 신뢰도를 확인하기 위한 방법을 탐지하기 위하여 응답 데이터를 기계학습 알고리즘을 활용하여 1) 성실 응답, 2) 무작위 불성실 응답, 3) 비무작위 불성실 응답으로 분류하는 방법을 검증하고자 하였다. 본 연구에서는 언어발달 단계와 관련된 정보가 내포되어 있는 설문응답 데이터를 활용하여 기계학습 모델 훈련을 통해 불성실 응답을 탐지하는 방법을 제안하였다. 본 연구에서 활용한 설문조사 문항은 참여 아동의 현재 월령보다 6-12개월 어린 월령대에서 주로 습득하는 문항과 6-9개월 높은 월령대에서 주로 습득하는 문항에 대한 설문으로 설계되어 아동의 월령에 따른 언어발달 단계와 관련된 설문 응답 데이터를 수집하였고, 불성실 응답 탐지 성능 검증을 시행하였다.

연구결과, 응답자의 아동 월령대를 기준으로, 전문가들의 논의에 따라 실제 월령에 부합하지 않는 응답 패턴을 불성실 응답으로 간주해 구분한 결과, 453개의 응답 중 성실 응답은 220개, 불성실 응답(비무작위 불성실 응답)은 233개로 나타났다. 이에 난수 발생기로 생성한 무작위 불성실 응답 226개를 포함하여 총 679개의 응답 데이터를 활용해 불성실 응답 탐지를 실시하였으며, 본 연구에서 사용한 데이터 셋의 응답 유형 분포의 균형 검증 결과, 성실 응답, 비무작위 불성실 응답 및 무작위 불성실 응답의 비율이 각각 32.4%, 34.3% 및 33.3%로 나타났으며, 어떤 특정 클래스의 데이터가 다수를 차지하는 응답 데이터의 불균형은 찾아볼 수 없었다. 기

계학습 모델을 활용하여 선별검사 응답 데이터 중 불성실 응답을 구분하기 위한 방법을 탐구한 결과, 본 연구를 위해 수집된 데이터 셋에 대해서 약 85% 수준의 정확도로 성실, 무작위 불성실, 비무작위 불성실 응답 유형 분류가 가능함을 확인하였다.

본 연구에서 사용한 지도학습 모델의 분류 정확도는 기존 관련 연구들의 불성실 응답 탐지 정확도와 비교하였을 때 정확도가 상대적으로 낮은 것으로 나타났다. 지도학습 모델을 활용하여 불성실 응답을 탐지하고자 한 Kang 등(2020)의 연구에서는 SVM과 ANN 두 가지 모델을 사용하였는데, 약 98% 이상의 정확도로 불성실 응답 분류가 가능하다고 보고하였다. 이 연구의 결과는 본 연구에서 확인한 불성실 응답 탐지 정확도와 비교하여 다소 높은 수치이다. 그러나 이 연구는 본 연구와 마찬가지로 불성실 응답의 유형을 분류하고자 하였지만, 사용한 설문조사 변수의 수가 6개로 적은 편이고 모델에서 결정을 해야 할 클래스도 두 개로 적었기 때문에 해당 연구의 불성실 응답 탐지 정확도와 본 연구결과를 직접적으로 비교하기는 어렵다고 할 수 있다. 또한, 본 연구에서 검증한 불성실 응답 탐지 방법은 Kang 등(2020)의 연구에서 제안한 방법과 비교하여 분류하고자 하는 응답 유형의 개수가 많을 뿐만 아니라 기계학습 모델을 훈련하기 위해 사용된 데이터 수도 적기 때문에 분류 난이도가 더 높다고 할 수 있으며, 본 연구에서 제안한 방법의 응답 유형 분류 정확도가 낮다고 단정 짓기 어렵다.

반면에, 본 연구에서 확인한 기계학습 기반 불성실 응답 탐지 방법은 응답의 유형을 성실 응답, 무작위 불성실 응답 및 비무작위 불성실 응답으로 세분화된 분류 결과 도출이 가능하기 때문에 그 의미가 있다고 할 수 있다. 기계학습 모델을 활용하여 불성실 응답 탐지를 위한 방법을 제안한 또 다른 연구(Jebreel et al., 2020)에 따르면 DBSCAN 및 IForest와 같은 비지도학습 모델들의 예측 결과를 결합하였을 때 불성실 응답을 약 91-96% 정확도로 탐지할 수 있음을 확인하였다. 하지만, 해당 연구의 목적이 불성실 응답의 형태를 분류하는 방법이 아닌 불성실 응답의 비중을 기준으로 불성실 응답을 탐지하는 방법인 연구이기 때문에 본 연구와 직접적인 비교가 어렵다고 할 수 있다. 또한, 이 연구는 15개의 설문 문항에 대한 약 4만 5천 건의 응답 데이터를 두 개의 클래스로 분류하였는데, 173개의 문항에 대한 679건의 응답을 활용한 본 연구와 비교할 때, 비지도 기계학습 모델 훈련을 위해 비교적 충분한 데이터를 활용하였을 뿐만 아니라, 분류 항목의 개수가 적어 분류 난이도가 상대적으로 낮았기 때문에 다소 높은 정확도를 얻을 수 있었다고 유추해 볼 수 있다. 그러나 이러한 불성실 응답 탐지 정확도 차이에도 불구하고, 이 연구에서도 개별 모델의 성능보다는 각 모델을 앙상블링하여 불성실 응답을 탐지하는 방법의 정확도가 더 높다는 결

과를 제시하였으며, 이는 본 연구에서도 동일하게 확인되었기 때문에 다수의 모델을 통합하여 불성실 응답 탐지의 성능을 향상시킬 수 있다는 사실은 중요한 시사점이라 할 수 있다.

인터넷을 활용한 자기보고식 온라인 설문은 대면 또는 전화를 통한 조사 방법에 비해 비용절감과 응답자의 설문 참여 용이성, 그리고 익명성 보장 등의 장점이 있으며(Fricker & Schonlau, 2002), 이러한 이유로 인해 연구자 또는 설문조사 참여 대상자들이 선호하는 조사방식이라 할 수 있다. 하지만, 자기보고식 온라인 설문조사는 동일한 참여자가 여러 번 설문조사에 참여하거나 의도적으로 응답을 조작할 수 있다는 점이 단점이 있으며(Teitcher et al., 2015), 전통적인 대면 방식의 인터뷰나 전화 조사에 비해 조사원의 통제에서 자유롭기 때문에 설문조사에 불성실하게 임할 위험성이 높은 것으로 알려져 있다(Brühlmann, Petralito, Aeschbach, & Opwis, 2020). 이러한 불성실 응답은 자기보고식 설문 응답의 신뢰성에 부정적 영향을 미칠 수 있는 요인으로 확인되었다(Niessen, Meijer, & Tendeiro, 2016). 이러한 문제를 해결하기 위한 방안 중 하나로 본 연구에서는 영아의 언어발달과 관련된 선별검사 설문 문항을 활용하여 기계학습 모델에 기반을 둔 불성실 응답 탐지에 활용 가능한 방법을 제시하였다. 본 연구에서 제안한 방법이 온라인에서 영유아의 언어발달 검사를 위한 부모보고식 선별검사뿐만 아니라 온라인 상에서 수행되는 자기보고식 설문조사에서 불성실 응답을 효과적으로 탐지할 수 있는 도구로서 온라인 선별검사 연구데이터의 신뢰성 확보에 도움이 되기를 기대한다. 구체적으로는 이 모델을 언어 선별검사 과정에 통합함으로써 불성실한 응답을 실시간으로 식별하고 필터링하여, 아동의 실제 언어능력을 더욱 정확하게 반영하는 결과를 얻을 수 있는 데 도움을 줄 수 있을 것으로 사료된다. 이는 임상 현장에서 언어발달장애의 조기 진단을 위한 정확한 설문 응답 수집 및 개별 아동에게 최적화된 중재 계획 수립을 위한 효과적 인 도구로서의 역할도 기대할 수 있을 것으로 판단된다.

REFERENCES

- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Bowling, N. A., & Huang, J. L. (2018). Your attention please! Toward a better understanding of research participant carelessness. *Applied Psychology, 67*(2), 227-230.
- Brühlmann, F., Petralito, S., Aeschbach, L. F., & Opwis, K. (2020). The quality of data collected online: an investigation of careless responding in a crowd-sourced sample. *Methods in Psychology, 2*, 100022.
- Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66*, 4-19.
- DeSimone, J. A., DeSimone, A. J., Harms, P. D., & Wood, D. (2018). The differential impacts of two forms of insufficient effort responding. *Applied Psychology, 67*(2), 309-338.
- DeSimone, J. A., & Harms, P. D. (2018). Dirty data: the effects of screening respondents who provide low-quality data in survey research. *Journal of Business & Psychology, 33*, 559-577.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In J. Kittler & F. Roli (Eds.), *Multiple classifier systems. MCS 2000. Lecture notes in computer science* (Vol. 1857, pp. 1-15). Springer.
- Dunn, A. M., Heggstad, E. D., Shanock, L. R., & Theilgard, N. (2018). Intra-individual response variability as an indicator of insufficient effort responding: comparison to other indicators and relationships with individual differences. *Journal of Business & Psychology, 33*, 105-121.
- Fricker, R. D., & Schonlau, M. (2002). Advantages and disadvantages of internet research surveys: evidence from the literature. *Field Methods, 14*(4), 347-367.
- Gartstein, M. A., & Marmion, J. (2008). Fear and positive affectivity in infancy: convergence/discrepancy between parent-report and laboratory-based indicators. *Infant Behavior & Development, 31*(2), 227-238.
- Hong, K. H., & Kim, Y. T. (2005). A longitudinal study of predictors for expressive vocabulary development of late-talkers. *Korean Journal of Communication & Disorders, 10*(1), 1-24.
- Han, J. Y., & Yim, D. S. (2018). Korean brief parent report measures of language development in children with vocabulary delay. *Journal of Speech-Language & Hearing Disorders, 27*(2), 69-84.
- Huang, J. L., Bowling, N. A., Liu, M., & Li, Y. (2015). Detecting insufficient effort responding with an infrequency scale: evaluating validity and participant reactions. *Journal of Business & Psychology, 30*, 299-311.
- Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: examining an insidious confound in survey data. *Journal of Applied Psychology, 100*(3), 828-845.
- Huang, J., & Wang, Z. (2021). Careless responding and insufficient effort responding. *Oxford Research Encyclopedia of Business and Management*. Retrieved from <https://oxfordre.com/business/view/10.1093/acrefore/9780190224851.001.0001/acrefore-9780190224851-e-303>.
- Jebreel, N. M., Haffar, R., Singh, A. K., Sánchez, D., Domingo-Ferrer, J., & Blanco-Justicia, A. (2020). Detecting bad answers in survey data through unsupervised machine learning. *Proceedings of the Privacy in Statistical Databases: UNESCO Chair in Data Privacy, International Conference*, 309-

- 320.
- Jung, D. Y., Kim, H. I., & Jung, D. I. (2001). *A study on the prevalence of children with special education needs*. Ansan: Korea National Institute of Special Education.
- Kam, C. C. S., & Meyer, J. P. (2015). How careless responding and acquiescence response bias can influence construct dimensionality: the case of job satisfaction. *Organizational Research Methods*, 18(3), 512-541.
- Kang, C. W., Choi, S. B., & Kwon, M. S. (2020). False responses detection of survey system using machine learning. *Journal of The Korean Data Analysis Society*, 22(3), 1013-1021.
- Kim, E. S. (2024). Parenting of infants and toddlers at risk of disability and family-centered support measures. *Korean Institute of Child Care & Education*, 79, 17-30.
- Kim, J. Y., & Ha, E. H. (2014). Cluster analysis of CBCL 1.5-5 for infants and young children with developmental delays. *Proceedings of the Korean Psychological Association Conference*, 265.
- Lee, Y. (2011). The relationships among language, communicative abilities and motor, cognitive, and socio-emotional development in toddlers with language delays. *Korean Journal of Communication & Disorders*, 16(1), 1-12.
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437-455.
- Mondal, N., Bhat, B. V., Plakkal, N., Thulasingham, M., Ajayan, P., & Poorna, D. R. (2016). Prevalence and risk factors of speech and language delay in children less than three years of age. *Journal of Comprehensive Pediatrics*, 7(2), e33173.
- Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Detecting careless respondents in web-based questionnaires: which method to use? *Journal of Research in Personality*, 63, 1-11.
- Owens, R. E. (2019). *Language development: an introduction*. Pearson.
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4), 867-872.
- Ozaki, K. (2024). Detecting inattentive respondents by machine learning: a generic technique that substitutes for the directed questions scale and compensates for its shortcomings. *Behavior Research Methods*, 56, 7059-7078.
- Rescorla, L. A., & Dale, P. S. (Eds.). (2013). *Late talkers: language development, interventions, and outcomes*. Paul H. Brookes Publishing Co.
- Seo, S. J., & Chang, H. S. (2013). The reliability, validity, and usability of the on-line version for early development screening test (e-DEP) for young children. *Special Education Research*, 12(3), 23-42.
- Teitcher, J. E., Bockting, W. O., Bauermeister, J. A., Hoefler, C. J., Miner, M. H., & Klitzman, R. L. (2015). Detecting, preventing, and responding to "fraudsters" in internet research: ethics and tradeoffs. *The Journal of Law, Medicine & Ethics*, 43(1), 116-133.
- Ward, M. K., & Meade, A. W. (2018). Applying social psychology to prevent careless responding during online surveys. *Applied Psychology*, 67(2), 231-263.
- Zhou, Z. H. (2012). *Ensemble methods: foundations and algorithms*. Springer.

국문초록

기계학습을 활용한 영유아 온라인 언어 선별검사의 응답 성실도 탐지 연구

나지영¹ · 홍경훈¹ · 김중연²

¹나사렛대학교 언어치료학과, ²순천향대학교 ICT융합연구센터

배경 및 목적: 본 연구는 영유아 온라인 언어발달 선별검사에서 발생하는 부정확한 응답으로 인한 신뢰도 저하 문제를 해결하기 위해, 기계학습 모델을 활용한 부모의 응답 성실도 탐지 방법을 제안하고자 하였다. **방법:** 연구에 사용된 언어발달 선별검사는 부모가 직접 온라인으로 응답하는 형태로 개발되었으며, 총 453명의 부모에게 선별검사를 실시하였다. 추가적으로 난수 발생기를 통해 무작위 불성실 응답 226건을 생성하였다. 연구자가 설정한 기준에 따라 응답 유형을 성실 응답, 비무작위 불성실 응답, 무작위 불성실 응답으로 나누어 세분화했다. 그 후, XGBoost, LightGBM, MLP 등 대표적으로 많이 사용되는 기계학습 알고리즘을 활용하여 응답 데이터를 학습 시키고, K-fold 교차 검증을 통해 모델의 정확도를 평가하였다. **결과:** 설정한 기준에 따라 응답은 성실 응답 220건, 비무작위 불성실 응답 233건, 무작위 불성실 응답 226건으로 분류되었다. 기계학습 기반의 응답 유형 분류 결과, 앙상블 모델을 이용한 불성실 응답 탐지에서 약 89.7%의 높은 정확도를 기록하며, 부모의 불성실한 응답을 효과적으로 분류할 수 있음을 입증하였다. **논의 및 결론:** 본 연구에서 제안한 기계학습 기반 접근법은 영유아 온라인 언어발달 선별검사뿐만 아니라 자기보고식 설문 of 불성실 응답 탐지에 효과적으로 활용될 수 있다. 이러한 방법을 통해 연구 데이터의 신뢰성을 높이고 보다 정확한 평가가 이루어질 것으로 기대된다.

핵심어: 기계학습, 온라인 언어 선별검사, 영유아 언어 선별검사, 응답 성실도

이 연구는 2023년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2023-00218176).

참고문헌

- 강창원, 최승배, 권민수 (2020). 머신러닝을 활용한 설문조사 불성실응답 판별. *한국자료분석학회*, 22(3), 1013-1021.
- 김영훈 (2024). 장애위험 영유아 조기발견과 개입을 위한 서비스 체계 구축 필요성과 과제. *육아정책포럼*, 79, 4-5.
- 김은설 (2024). 장애위험 영유아육아 실태 및 가족중심 지원 방안. *육아정책포럼*, 79, 17-30.
- 김지예, 하은혜 (2014). 발달지체 영유아의 CBCL 1.5-5 유아행동평가척도 군집분석. *한국심리학회 학술대회 자료집*, 265.
- 서소정, 장혜성 (2013). 온라인 영아발달선별 검사도구(e-DEP)의 타당성 및 만족도에 관한 연구. *특수교육*, 12(3), 23-42.
- 이윤경 (2011). 언어발달지체 영유아의 언어 및 의사소통 능력과 인지, 운동 및 사회성 발달과의 관계. *언어청각장애연구*, 16(1), 1-12.
- 정동영, 김형일, 정동일 (2001). *특수교육 요구아동 출현율 조사 연구*. 안산: 국립특수교육원.
- 한지윤, 임동선 (2018). 부모 보고형 아동 언어 능력 평가도구(KBPR)의 표준화를 위한 예비연구. *언어치료연구*, 27(2), 69-84.
- 홍경훈, 김영태 (2005). 중단연구를 통한 ‘말늦은아동(late-talker)’의 표현어휘발달 예측요인 분석. *언어청각장애연구*, 10(1), 1-24.

ORCID

나지영(제1저자, 조교수 <https://orcid.org/0000-0002-1688-7887>); 홍경훈(공동저자, 교수 <https://orcid.org/0000-0002-6567-3885>); 김중연(교신저자, 연구교수 <https://orcid.org/0000-0001-7229-1466>)