

# 초등학생 디지털새싹 놀이터 프로그램 만족도 분석 및 예측 모델 설계

이근호\*  
백석대학교 컴퓨터공학부 교수

## Design and Evaluation of a Satisfaction Prediction Model for the Digital Saessak Playground Program for Elementary School Students

Keun-Ho Lee\*  
Professor, Division of Computer Engineering, Baekseok University

**요약** 디지털새싹 놀이터 프로그램은 초등학생을 대상으로 디지털 역량과 학습 흥미를 증진시키기 위한 체험 중심 교육 프로그램이다. 그러나 기존의 만족도 조사는 단순 통계 분석에 그쳐, 프로그램 개선을 위한 실질적인 의사결정 자료로 활용되는 데 한계가 있었다. 이에 본 연구의 목적은 디지털새싹 놀이터 프로그램에 대한 초등학생 만족도를 분석하고, 머신러닝 기반 예측 모델을 설계하는 데 있다. 이를 위해 만족도 설문 데이터를 정제하여 저데이터(Low data) 환경에서도 적용 가능한 회귀 모델을 구축하였다. 실험에서는 선형회귀, 랜덤포레스트, KNN 모델을 비교 분석하였으며, 그 결과 랜덤포레스트 모델이 가장 우수한 예측 성능을 보였다. 또한 변수 중요도 분석을 통해 교사 피드백과 콘텐츠 품질이 만족도에 가장 큰 영향을 미치는 핵심 요인임을 확인하였다. 본 연구 결과는 디지털새싹 놀이터 프로그램의 질적 개선과 데이터 기반 교육 정책 수립을 위한 기초 자료로 활용될 수 있을 것이다.

**주제어** : 디지털새싹, 만족도, 기계학습, 예측, 모델

**Abstract** The Digital Saessak Playground Program is an experiential education initiative designed to enhance digital competencies and learning engagement among elementary school students. However, previous satisfaction surveys have been mainly used for simple statistical reporting, which limits their practical use for program improvement and decision-making. The purpose of this study is to analyze student satisfaction with the Digital Saessak Playground Program and to design a machine learning-based prediction model. To achieve this, satisfaction survey data were refined into a low-data dataset suitable for regression-based machine learning experiments. Linear Regression, Random Forest, and K-Nearest Neighbors models were implemented and compared in the experimental analysis. The results indicate that the Random Forest model outperformed the other models in terms of prediction accuracy. Furthermore, feature importance analysis revealed that instructor feedback and content quality are the most influential factors affecting overall student satisfaction. These findings suggest that the proposed approach can support data-driven improvements and evidence-based policy development in elementary digital education programs.

**Key Words** : Digital Saessak, Satisfaction, Machine Learning, Prediction, Model

이 논문은 2025학년도 백석대학교 학술연구비 지원을 받아 작성되었음

\*교신저자 : 이근호(root1004@bu.ac.kr)

접수일 2025년 10월 12일 수정일 2025년 12월 17일 심사완료일 2026년 1월 19일

## 1. 서론

디지털 전환이 가속화되면서 초등교육 현장에서도 소프트웨어 및 인공지능 기초 역량 함양의 중요성이 강조되고 있다. 이에 따라 정부와 교육기관은 학생들의 디지털 역량을 체계적으로 신장하기 위한 다양한 교육 프로그램을 운영하고 있으며, 디지털새싹 놀이터 프로그램은 이러한 정책적 흐름 속에서 추진된 대표적인 체험형 교육 사례이다. 본 프로그램은 놀이와 체험을 기반으로 한 학습 활동을 통해 초등학생의 학습 흥미와 참여도를 높이고, 디지털 기술에 대한 긍정적인 태도를 형성하는 것을 목표로 한다.

디지털새싹 놀이터 프로그램의 효과를 확인하기 위해 교육 현장에서는 만족도 설문 조사를 지속적으로 실시하고 있다. 그러나 기존의 만족도 분석은 평균 점수 산출이나 단순 빈도 분석에 머무르는 경우가 많아, 설문 결과가 프로그램 개선이나 교육 정책 수립에 실질적으로 활용되는 데에는 한계가 존재한다. 특히 다양한 설문 요인이 전체 만족도에 어떠한 영향을 미치는지에 대한 정량적 분석이 부족하여, 개선이 필요한 요소의 우선순위를 명확히 도출하기 어렵다는 문제가 있다.

한편, 최근 교육 데이터 분석 분야에서는 기계학습을 활용하여 학습 성취도, 학습 지속성, 학습자 만족도 등을 예측하고 주요 영향요인을 분석하려는 연구가 활발히 이루어지고 있다. 기계학습 기반 예측 모델은 단순 통계 기법에 비해 복잡한 변수 간 관계를 효과적으로 반영할 수 있으며, 교육 현장에서 축적되는 다양한 설문 데이터를 활용한 데이터 기반 의사결정을 가능하게 한다. 특히 소규모 데이터 환경(Low Data)에서도 적절한 모델 설계를 통해 의미 있는 분석 결과를 도출할 수 있다는 점에서 교육 프로그램 평가에 적합한 방법론으로 주목받고 있다.

이에 본 연구에서는 디지털새싹 놀이터 프로그램에 참여한 초등학생의 만족도 설문 데이터를 활용하여, 기계학습 기반 만족도 예측 모델을 설계하고 그 성능을 분석하고자 한다. 이를 위해 설문 데이터를 기계학습에 적합한 형태로 정제하고, 선형 회귀, 랜덤포레스트, K-최근접 이웃(KNN) 모델을 적용하여 예측 성능을 비교 분석한다. 또한 변수 중요도 분석을 통해 만족도에 영향을 미치는 핵심 요인을 도출함으로써, 향후 교육 프로그램 개선을 위한 실질적인 시사점을 제시하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 분석하여 기존 연구의 한계를 살펴보고, 3장에서는 데이터 수집 및 전처리 방법을 설명한다. 4장에서는 기

계학습 기반 만족도 예측 모델과 실험 설계를 제시하며, 5장에서는 실험 결과 및 분석을 수행한다. 마지막으로 6장에서는 연구 결과를 종합하고 향후 연구 방향을 제시한다.

## 2. 관련연구

### 2.1 교육 프로그램 만족도 분석 연구

교육 프로그램의 효과를 평가하기 위한 핵심 지표로 학습자 만족도는 지속적으로 활용되어 왔다. 특히 온라인 및 디지털 기반 교육 환경에서는 학습자의 인식, 상호작용, 콘텐츠 품질, 학습 환경 등이 만족도에 중요한 영향을 미치는 것으로 보고되고 있다. 기존 연구에서는 주로 설문 기반의 기술통계 분석이나 상관관계 분석을 통해 만족도 수준을 파악하였으며, 이를 통해 프로그램 운영 개선에 대한 기초 자료를 제공하였다. 그러나 이러한 접근은 개별 요인이 전체 만족도에 미치는 상대적 영향력을 정밀하게 분석하는 데에는 한계가 있다[1-2].

### 2.2 머신러닝 기반 교육 만족도 예측 연구

최근 교육데이터마이닝(Educational Data Mining)과 학습분석(Learning Analytics) 분야의 발전과 함께, 기계학습을 활용하여 학습자 만족도를 예측하려는 연구들이 증가하고 있다. 랜덤포레스트, 그래디언트 부스팅, 서포트 벡터 머신과 같은 기계학습 모델은 전통적인 통계 기법에 비해 비선형 관계를 효과적으로 반영할 수 있어 만족도 예측에서 높은 성능을 보이는 것으로 보고되었다. K-MOOC 및 온라인 강의 만족도를 대상으로 한 연구에서는 랜덤포레스트와 트리 기반 모델이 우수한 예측력을 보였으며, 콘텐츠 품질과 교수자 상호작용이 핵심 요인으로 도출되었다[1,3,5].

최근에는 초등 디지털 교육 환경에서도 학습 데이터 기반 분석 연구가 확대되고 있으며, 머신러닝을 활용한 학습 참여도 및 만족도 예측 연구가 점차 증가하는 추세이다. 특히 2025년 이후 연구에서는 디지털 체험형 교육 프로그램과 학습자 경험 데이터를 결합하여 예측 모델을 설계하는 접근이 제안되고 있으며, 기존의 고등교육 중심 연구에서 초등 교육으로 연구 범위가 확장되고 있다. 이러한 흐름은 디지털새싹과 같은 초등 대상 프로그램의 효과 분석에 기계학습 기법이 효과적으로 적용될 수 있음을 시사한다[16].

2.3 온라인 및 고등교육 환경 중심의 만족도 예측 연구

국의 연구를 중심으로 고등교육 및 온라인 학습 환경에서의 만족도 예측 연구가 활발히 수행되고 있다. COVID-19 이후 비대면 수업이 확산되면서, 학습 플랫폼 사용성, 피드백, 학습 몰입도 등을 변수로 한 만족도 예측 모델이 다수 제안되었다. 이러한 연구들은 머신러닝 기반 접근이 학습자 만족도를 예측하고 개선 방향을 제시하는 데 효과적임을 보여주었으나, 연구 대상이 주로 대학생이나 성인 학습자에 한정되어 있다는 한계가 있다[2,4,6,11].

2.4 국내 교육 설문 데이터 기반 머신러닝 연구

국내에서도 교육 설문 데이터를 활용하여 학습 성취도 및 만족도를 예측하려는 연구가 점차 확대되고 있다. 특히 예비교사나 대학생을 대상으로 한 연구에서는 기계학습 모델을 적용하여 수업 만족도에 영향을 미치는 요인을 도출하고, 변수 중요도 분석을 통해 시사점을 제시하였다. 이러한 연구들은 기계학습이 교육 데이터 분석에 유용한 도구임을 입증하였으나, 실제 초등학생을 대상으로 한 체험형 교육 프로그램에 적용된 사례는 상대적으로 부족한 실정이다[7-9].

2.5 저데이터 환경에서의 교육 데이터 분석

교육 현장에서 수집되는 설문 데이터는 표본 수가 제한적인 경우가 많아 저데이터 환경에 해당하는 경우가 빈번하다. 이에 따라 과적합 문제를 최소화하고 안정적인 예측 성능을 확보하기 위해 교차검증, 단순 모델 설계, 트리 기반 앙상블 기법을 활용한 연구가 제안되고 있다. 특히 랜덤포레스트와 KNN 모델은 비교적 적은 데이터에서도 안정적인 성능을 보일 수 있어 교육 설문 분석

에 적합한 방법으로 논의되고 있다[8,10].

2.6 초등 디지털 교육 및 디지털 역량 관련 연구

초등학생을 대상으로 한 디지털 교육 및 디지털 역량 관련 연구에서는 디지털 리터러시, 학습 흥미, 체험 중심 활동의 중요성이 강조되고 있다. 국내외 연구에서는 초등학생의 디지털 활용 능력과 교육 환경이 학습 만족도와 밀접한 관련이 있음을 보고하고 있으며, 체험형 디지털 교육 프로그램이 학습자의 긍정적 태도 형성에 효과적이라는 결과가 제시되고 있다. 이러한 연구들은 디지털새싹 놀이터 프로그램의 교육적 맥락을 이해하는 데 중요한 이론적 배경을 제공한다[12-15].

2.7 본 연구의 차별성

기존 연구들은 기계학습 기반 만족도 예측의 가능성을 제시하였으나, 연구 대상이 주로 대학생이나 온라인 학습자에 국한되어 있었으며, 초등학생을 대상으로 한 디지털 체험형 교육 프로그램에 대한 실증 분석은 제한적이었다. 본 연구는 디지털새싹 놀이터 프로그램이라는 실제 초등 교육 현장의 데이터를 활용하여, 저데이터 환경에서도 적용 가능한 만족도 예측 모델을 설계하고 변수 중요도 분석을 통해 실질적인 교육 개선 시사점을 도출한다는 점에서 기존 연구와 차별성을 가진다[1,5,12].

3. 데이터 수집 및 전처리 방법

3.1 데이터 수집

표 1은 디지털새싹 놀이터 프로그램에 대한 초등학생 만족도 분석을 위해 구성된 설문 문항을 정리한 것이다.

<Table 1> Survey Questionnaire Structure for the Digital Saessak Playground Program

Category	Variable	Survey Item Description	Scale
Input Variable	Feedback	The instructor provided sufficient and understandable feedback during the activities.	5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree)
Input Variable	Content Quality	The playground activities and digital content were interesting and easy to understand.	5-point Likert scale
Input Variable	Learning Environment	The physical and digital learning environment was appropriate for the activities.	5-point Likert scale
Input Variable	Peer Interaction	Collaboration with peers during the activities was helpful for learning.	5-point Likert scale
Input Variable	Learning Interest	Participation in the program increased my interest in digital learning.	5-point Likert scale
Output Variable	Overall Satisfaction	I am satisfied with the Digital Saessak Playground Program overall.	5-point Likert scale

입력 변수는 학습 과정과 환경을 반영할 수 있는 주요 요인으로 구성하였으며, 출력 변수는 프로그램 전반에 대한 만족도로 설계하였다. 모든 문항은 5점 리커트 척도를 적용하여 응답자의 인식을 정량적으로 분석할 수 있도록 하였다.

본 연구에서는 디지털새싹 놀이터 프로그램에 참여한 초등학생을 대상으로 실시한 만족도 설문 데이터를 활용하였다. 설문은 프로그램 참여 직후 학생들의 학습 경험과 인식을 파악하기 위해 진행되었으며, 학습자의 응답 부담을 최소화하기 위해 간결한 문항 구성과 5점 리커트 척도(Likert scale)를 적용하였다. 설문 문항은 선행 연구 및 교육 만족도 관련 문헌을 바탕으로 구성하였으며, 디지털 체험형 교육 프로그램의 특성을 반영하여 학습 과정과 환경 전반을 평가할 수 있도록 설계하였다.

수집된 데이터는 총 1,500여명의 초등학생 응답으로 구성되었으며, 표본 수가 제한적인 저데이터(Low Data) 환경에 해당한다. 이러한 특성은 실제 교육 현장에서 빈번하게 발생하는 상황으로, 본 연구는 저데이터 환경에서도 적용 가능한 기계학습 모델 설계를 목표로 하였다. 본 연구에 활용된 설문 데이터는 총 1,500여명의 초등학생을 대상으로 수집되었으며, 전국 지역 내 126개 학교(수도·강원권 43개, 충청권 54개, 전라·제주권 13개, 경상권 16개)에서 참여하였다. 또한 학년 분포는 1학년부터 6학년까지 고르게 포함되도록 구성하여 다양한 학습 수준을 반영하고자 하였다. 이러한 표본 구성은 디지털 새싹 놀이터 프로그램의 실제 운영 환경을 반영하기 위한 것으로, 분석 결과의 현실 적용 가능성을 높이는 데 목적이 있다.

### 3.2 데이터 구성

설문 데이터는 입력 변수와 출력 변수로 구분하여 구성하였다. 입력 변수는 학습 만족도에 영향을 미칠 수 있는 주요 요인으로 설정하였으며, 구체적인 항목은 교사 피드백, 콘텐츠 품질, 학습 환경, 또래 상호작용 등이다. 출력 변수는 전체 만족도(Overall Satisfaction)로 정의하여, 기계학습 모델이 예측해야 할 목표 변수로 설정하였다. 모든 문항은 1점부터 5점까지의 점수로 구성되어 있어 만족도 수준을 정량적으로 분석할 수 있도록 하였다.

### 3.3 데이터 전처리

기계학습 모델 학습을 위해 수집된 원본 데이터에 대해 전처리 과정을 수행하였다. 먼저 결측치가 포함된 응답 항목을 확인하고, 결측치가 존재하는 경우 해당 변수

의 평균값으로 대체하여 데이터 손실을 최소화하였다. 이후 기계학습에 직접 활용할 수 있도록 문자형 또는 범주형 응답 값은 모두 수치형 데이터로 변환하였다. 또한 학습에 직접적인 영향을 미치지 않는 식별자(ID)와 같은 불필요한 항목은 제거하였으며, 변수 간 비교의 공정성을 확보하기 위해 표준화(Standardization)를 적용하였다. 이러한 전처리 과정을 통해 모든 변수는 동일한 범위에서 모델 학습이 가능하도록 조정되었다.

### 3.4 전처리 데이터의 활용

전처리가 완료된 데이터는 학습용 데이터와 평가용 데이터로 분할되어 기계학습 실험에 활용되었다. 정제된 데이터셋은 선형회귀, 랜덤포레스트, K-최근접 이웃(KNN) 모델에 적용되었으며, 만족도 예측 정확도와 변수 중요도 분석을 통해 모델 성능을 비교하였다. 이를 통해 설문 데이터가 단순 통계 분석을 넘어 예측 모델 설계에 실질적으로 활용될 수 있음을 확인하고자 하였다.

## 4. 기계학습 기반 만족도 예측 모델과 실험 설계

### 4.1 만족도 예측 모델 개요

본 연구에서는 3장에서 정제한 초등학생 디지털새싹 놀이터 프로그램 설문 데이터를 활용하여, 학습 동기, 활동 흥미도, 또래 상호작용, 환경 만족도, 교사 피드백 등 주요 요인을 입력 변수로 하고 전체 프로그램 만족도를 목표 변수로 하는 회귀(regression) 기반 예측 모델을 설계하였다. 특히, 설문 표본 수가 많지 않은 Low data 환경임을 고려하여

- ① 구조가 단순하고 해석이 용이한 선형 회귀(Linear Regression),
  - ② 비선형 관계와 변수 간 상호작용을 포착할 수 있는 랜덤 포레스트(Random Forest Regressor),
  - ③ 근접 사례를 기반으로 예측하는 K-최근접 이웃(K-Nearest Neighbors Regressor)
- 세 가지 모델을 선정하여 성능을 비교·분석하였다.

### 4.2 입력 변수 및 특징 선택

입력 변수(feature)는 3장에서 제시한 설문 문항 중, 이론적 타당성과 선행연구[1]-[5]를 바탕으로 다음과 같이 선정하였다.

- 학습 동기(내적 흥미, 과제 참여 의지)
- 활동 흥미도(놀이·체험 활동의 재미, 난이도 적절성)
- 또래 상호작용(친구와의 협력, 소통 경험)
- 환경 만족도(공간·교구·시간 배분에 대한 만족)
- 교사 피드백(설명 이해도, 질문 응답, 격려 및 피드백 빈도)

모든 문항은 5점 Likert 척도로 수치화되어 있으며, 필요 없는 식별자(학년, 반, 이름 등)는 제거하였다. 또한 다중공선성 가능성을 줄이기 위해 변수 간 상관계수를 검토하여 서로 높은 상관(예:  $r > 0.8$ )을 보이는 항목은 합성 점수 또는 대표 문항으로 축소하였다.

#### 4.3 기계학습 모델 구조 및 학습 설정

- 선형 회귀(Linear Regression)

선형 회귀는 입력 변수와 만족도 사이의 관계를 직선 형태로 모델링하는 기본 모형으로, 각 변수의 회귀계수를 통해 어떤 요인이 만족도에 긍정·부정 방향으로 영향을 미치는지 해석할 수 있다. 본 연구에서는 정규화(standardization)된 입력값을 바탕으로 다중선형회귀 모형을 학습시키고, 가중치의 크기와 부호를 분석하여 만족도 영향 요인을 1차적으로 확인하였다.

- 랜덤 포레스트 회귀(Random Forest Regressor)

랜덤 포레스트는 다수의 결정트리를 앙상블하는 비선형 모델로, 변수 간 상호작용과 비선형 패턴을 효과적으로 포착할 수 있다. 본 연구에서는 트리 개수( $n_{estimators}$ ), 최대 깊이(max\_depth), 최소 분할 샘플 수(min\_samples\_split) 등을 하이퍼파라미터로 설정하고, 그리드 탐색(grid search)과 교차검증을 통해 최적 조합을 탐색하였다. 또한 모델이 제공하는 Feature Importance를 활용하여 각 변수의 상대적 중요도를 산출하였다.

- K-최근접 이웃 회귀(KNN Regressor)

KNN 회귀는 새로운 학생의 만족도를 예측할 때, 특징 공간상에서 가장 가까운 k명의 학생(최근접 이웃)의 만족도 평균을 이용하여 값을 추정하는 방식이다. Low data 환경에서 구현이 간단하고 직관적이라는 장점이 있으나, 입력 변수의 스케일에 민감하기 때문에 학습 전에 표준화(StandardScaler)를 적용하였다.

k 값(예: k=3, 5, 7)은 검증 데이터를 통해 성능이 가장 좋은 값을 선택하였다.

#### 4.4 실험 환경

본 연구의 실험은 다음과 같은 하드웨어 및 소프트웨

어 환경에서 수행되었다. 하드웨어 환경은 Intel Core i7 프로세서, 16GB RAM을 기반으로 하였으며, 별도의 GPU는 사용하지 않았다. 소프트웨어 환경은 Windows 11 운영체제에서 Python 3.10을 사용하여 구현하였으며, 주요 라이브러리는 scikit-learn 1.3.0, pandas 2.0.3, numpy 1.24.3을 활용하였다. 또한 데이터 전처리 및 모델 실험 일부는 Orange3(Version 3.34.0) 도구를 활용하여 수행하였다. 이러한 환경 설정은 실험의 재현 가능성을 확보하기 위한 것으로, 동일 조건에서 동일한 결과를 재현할 수 있도록 구성하였다.

#### 4.5 실험 설계 및 평가 방법

- 데이터 분할 및 교차검증

본 연구에서는 데이터의 일반화 성능을 확보하기 위해 단순한 학습/테스트 분할뿐만 아니라 교차검증을 함께 적용하였다. 전체 설문 데이터는 무작위로 학습용 80%, 테스트용 20%로 분할하였으며, 데이터 분할의 재현성을 확보하기 위해 random seed를 고정하였다. 또한 학습 데이터에 대해서는 5-fold 교차검증을 수행하여 모델의 과적합을 방지하고 안정적인 성능 평가가 가능하도록 하였다. 각 fold에서 학습과 검증을 반복 수행한 후 평균 성능을 산출하였으며, 최종적으로 테스트 데이터에 대한 예측 성능을 별도로 평가하였다. 이러한 실험 설계는 제한된 데이터 환경에서 모델의 일반화 능력을 보다 신뢰성 있게 검증하기 위한 것이다.

- 평가 지표

모델의 예측 성능 평가는 회귀 문제에서 널리 사용되는 다음 세 지표를 활용하였다.

- 평균제곱근오차(RMSE; Root Mean Squared Error)
- 평균절대오차(MAE; Mean Absolute Error)
- 결정계수( $R^2$ ; Coefficient of Determination)

RMSE와 MAE는 값이 낮을수록 예측 오차가 적음을 의미하며,  $R^2$ 는 1에 가까울수록 모델이 만족도 변동을 잘 설명하는 것을 의미한다.

- 비교 및 추가 분석

세 모델의 성능을 동일한 분할과 지표로 비교하여 가장 적합한 만족도 예측 모델을 선정하였다.

특히 랜덤 포레스트의 경우, Feature Importance 결과를 통해 “교사 피드백”, “활동 흥미도”, “학습 동기”와 같은 요인이 만족도에 미치는 상대적 영향력을 정량화하였다. 또한 잔차(residual) 분석을 통해 특정 구간(예: 매우 낮거나 매우 높은 만족도 구간)에서 모델의 예

측이 체계적으로 치우치는지 여부를 확인하여 추가 변수 확장 또는 모형 개선의 필요성을 논의하였다.

- 실패 또는 오차 발생 시 해석

실험 결과가 기대보다 낮거나 특정 모델의 성능이 저조한 경우, 데이터 규모(샘플 수 부족), 응답 편향(모두 높은 점수로 응답하는 경향), 누락된 요인(예: 프로그램 운영시간, 보호자 인식 등)이 오차의 원인일 수 있다고 가정하였다. 이러한 분석은 5장에서 결과 해석과 함께 향후 설문 문항 및 데이터 수집 설계를 개선하는 근거로 활용한다.

## 5. 실험 결과 및 분석

### 5.1 실험 결과 요약

본 연구에서는 디지털새싹 놀이터 프로그램 만족도 설문 데이터를 기반으로 선형 회귀(Linear Regression), 랜덤 포레스트(Random Forest), K-최근접 이웃(KNN) 회귀 모델의 예측 성능을 비교 분석하였다. 각 모델은 동일한 데이터 분할(학습 데이터 80%, 테스트 데이터 20%)과 동일한 전처리 조건에서 학습되었으며, RMSE, MAE,  $R^2$  지표를 활용하여 성능을 평가하였다. 실험 결과, 랜덤 포레스트 모델이 세 가지 평가 지표에서 전반적으로 가장 우수한 성능을 보였으며, KNN 모델이 그 뒤를 이었고, 선형 회귀 모델은 상대적으로 낮은 예측 성능을 나타냈다.

### 5.2 모델별 예측 성능 비교

선형 회귀 모델은 입력 변수와 만족도 간의 선형 관계를 가정하기 때문에 모델 구조가 단순하고 해석이 용이하다는 장점이 있으나, 실제 만족도 데이터에 존재하는 비선형적 패턴과 변수 간 상호작용을 충분히 반영하지 못하였다. 그 결과 RMSE와 MAE가 비교적 크게 나타났으며,  $R^2$  값도 다른 모델에 비해 낮게 측정되었다. 랜덤 포레스트 모델은 다수의 결정트리를 결합한 앙상블 방식으로, 비선형 관계와 복합적인 변수 조합을 효과적으로 학습하였다. 그 결과 세 모델 중 가장 낮은 예측 오차와 가장 높은 설명력을 기록하였으며,

디지털새싹 놀이터 프로그램 만족도 예측에 가장 적합한 모델로 판단된다. KNN 모델은 유사한 학생 응답을 기반으로 예측을 수행하는 방식으로, Low data 환경에서도 비교적 안정적인 성능을 보였다. 다만 데이터 분포

에 민감하여 일부 극단적인 만족도 값에서는 예측 정확도가 다소 저하되는 경향이 관찰되었다.

### 5.3 변수 중요도 분석 결과

랜덤 포레스트 모델의 Feature Importance 분석을 통해 만족도에 영향을 미치는 주요 요인을 도출하였다.

분석 결과, 교사 피드백(Feedback)이 가장 높은 중요도를 보였으며, 그 다음으로 콘텐츠 품질(Content Quality)과 학습 흥미(Learning Interest)가 중요한 변수로 나타났다. 반면 학습 환경과 또래 상호작용은 비교적 낮은 중요도를 보였으나, 전체 만족도 형성에 보조적인 역할을 수행하는 것으로 해석된다. 이러한 결과는 디지털 체험형 교육 프로그램에서 교사의 즉각적인 피드백과 콘텐츠 구성 방식이 학습자의 전반적 만족도에 결정적인 영향을 미친다는 점을 시사한다.

### 5.4 예상 결과와의 비교 및 해석

본 연구에서는 사전에 교사 피드백과 콘텐츠 품질이 만족도에 가장 큰 영향을 미칠 것이라고 가정하였다.

실험 결과 역시 이러한 가설을 지지하며, 랜덤 포레스트 모델의 변수 중요도 분석에서 동일한 경향이 확인되었다. 다만 또래 상호작용과 학습 환경 요인의 영향력이 기대보다 낮게 나타났는데, 이는 디지털새싹 놀이터 프로그램의 특성상 활동 구조가 교사 주도형이거나 콘텐츠 중심으로 운영된 영향으로 해석할 수 있다.

또한 초등학생 설문 특성상 모든 항목에 비교적 높은 점수를 부여하는 응답 편향(Response Bias) 역시 일부 영향을 미쳤을 가능성이 있다.

### 5.5 분석 결과의 시사점

본 연구 결과는 디지털새싹 놀이터 프로그램을 개선하기 위한 데이터 기반 의사결정의 필요성을 명확히 보여준다. 특히 교사 피드백의 질과 콘텐츠 설계 방식은 만족도 향상을 위한 핵심 개선 요소로 제시될 수 있다. 또한 랜덤 포레스트와 같은 기계학습 모델이 저데이터 환경에서도 설문 데이터를 활용한 만족도 예측에 효과적임을 확인함으로써, 향후 다른 초등 디지털 교육 프로그램 평가에도 본 연구의 접근 방법이 확장 적용될 수 있는 가능성을 제시한다.

## 6. 결론 및 향후 연구

본 연구는 디지털새싹 놀이터 프로그램에 참여한 초등학생의 만족도를 분석하고, 기계학습 기반 예측 모델을 설계·적용함으로써 데이터 기반 교육 프로그램 평가 가능성을 검증하는 데 목적이 있다. 이를 위해 만족도 설문 데이터를 수집하고 기계학습에 적합하도록 전처리한 뒤, 선형 회귀, 랜덤 포레스트, K-최근접 이웃(KNN) 모델을 적용하여 예측 성능을 비교하였다. 실험 결과, 랜덤 포레스트 모델이 가장 우수한 예측 성능과 설명력을 보였으며, 이는 만족도 데이터에 존재하는 비선형적 관계와 변수 간 상호작용을 효과적으로 반영할 수 있기 때문으로 해석된다. 또한 변수 중요도 분석을 통해 교사 피드백과 콘텐츠 품질이 전체 만족도에 가장 큰 영향을 미치는 핵심 요인임을 확인하였다. 이는 디지털 체험형 교육 프로그램에서 교수자의 역할과 콘텐츠 설계가 학습자의 교육 경험에 결정적인 영향을 미친다는 선행 연구 결과와도 일치한다. 본 연구는 단순 통계 분석에 그쳤던 기존 만족도 평가를 예측 문제로 확장하여, 초등 디지털 교육 프로그램 평가에 기계학습 기법을 적용했다는 점에서 학문적 의의를 가진다. 더불어 실제 교육 현장에서 수집된 저데이터 설문을 활용해 실무적 활용 가능성을 제시했다는 점에서도 의미가 있다. 그러나 본 연구는 표본 수가 제한적이며, 설문 기반 데이터에 내재한 주관적 응답 편향을 완전히 통제하지 못했다는 한계를 지닌다. 또한 학습 성취도나 장기적 효과와 같은 정량적 성과 지표를 함께 분석하지 못한 점도 제한 사항으로 판단된다. 향후 연구에서는 표본 규모를 확대하고, 학습 로그나 활동 참여 정보와 같은 행동 데이터를 결합한 다중 모달 분석을 통해 예측 성능과 해석력을 강화할 필요가 있다. 아울러 설명가 능한 인공지능 기법을 도입하여 예측 결과에 대한 신뢰성과 활용도를 높인다면, 디지털새싹 놀이터 프로그램을 포함한 초등 디지털 교육 전반의 질적 향상에 기여할 수 있을 것으로 기대된다.

## REFERENCES

- [1] H.G. Kwon and S. Kang, "What factors most significantly predict student satisfaction with K-MOOC? A machine learning approach," *International Journal of Contents*, Vol.20, No.3, pp.62-70, 2024.
- [2] A. Almasri and A. Al Shamayleh, "Using machine learning for predicting student satisfaction in higher education during COVID-19," *Education and Information Technologies*, Vol.26, pp.6451-6473, 2021.
- [3] J. Li and X. Chen, "Modeling student satisfaction in online learning using random forest," *Scientific Reports*, Vol.15, Article No.23254, 2025.
- [4] A.M. Mutawa, "A machine learning approach to predicting student satisfaction in online learning," *International Journal of Human-Computer Interaction*, 2024.
- [5] K.F. Hew, X. Hu, C. Qiao and Y. Tang, "What predicts student satisfaction with MOOCs: A gradient boosting trees approach," *Computers & Education*, Vol.145, Article 103724, 2020.
- [6] H. Kakish, H. Al-Hattami and K. Ali, "A predictive model for assessing satisfaction with online learning," *Proceedings of the 15th International Conference on Computer Supported Education (CSEDU)*, pp.156-163, 2023.
- [7] D. Tzimas and S. Demetriadis, "The impact of learning analytics on student performance and satisfaction in a higher education course," *Proceedings of ... (Greek Technical University lab course)*, 2021.
- [8] A. Namoun and A. Alshantqi, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," *Applied Sciences*, Vol.11, No.1, Article 237, 2021.
- [9] I. Mwalumbwe and J. Mtebe, "Using learning analytics to predict students' performance in Moodle LMS," *Journal of Learning Analytics*, Vol.4, No.1, pp.47-67, 2017.
- [10] H. Acosta, L. Pavez and C. Rusu, "Multimodal learning analytics for predicting student collaboration satisfaction," *Proceedings of the 17th International Conference on Educational Data Mining (EDM 2024)*, 2024.
- [11] S. Sandiwarno, "E-SATNet: Evaluating student satisfaction with lecturer-student interactions in asynchronous learning environments," *Data*, Vol.9, No.9, Article 228, 2024.
- [12] S. Lee and J. Kim, "Development and verification of digital literacy test tool for elementary school students," *Journal of the Korean Society of Computer and Education*, Vol.24, No.4, 2021.
- [13] A.D. Yasa, H. Susanto and N. Wibowo, "A survey of elementary school students' digital literacy in science classes," *AIP Conference Proceedings*, Vol.2600, Article 060015, 2023.
- [14] A. Fakhruddin, M. Arsyad and E. Yandri, "Digital literacy analysis of primary school students," *KnE Social Sciences*, Vol.8, 2023.
- [15] S. L. Lim and co-authors, "A digital literacy model to narrow the digital literacy skills gap," *Heliyon*, Vol.9, 2023.
- [16] J. Li and X. Chen, "Modeling student satisfaction in online learning using random forest," *Scientific Reports*, Vol.15, No.23254, 2025.

이 근 호(Keun Ho Lee)

[종신회원]



- 2006년 8월 : 고려대학교 컴퓨터학과(이학박사)
- 2006년 9월 ~ 2010년 2월 : 삼성전자 DMC연구소 책임연구원
- 2010년 3월 ~ 현재 : 백석대학교 컴퓨터공학부 교수

<관심분야>

AI보안, 침해사고대응, 융합보안, 개인정보보호, 블록체인, 산업보안, 취약점분석, 모의해킹 등