

# TLBoost: 해석 가능성을 높인 학습자 성취도 예측 모델

## TLBoost: Student achievement prediction model with increased interpretability

박선우<sup>†</sup> · 김현철<sup>††</sup>

Sunwoo Park<sup>†</sup> · Hyeoncheol Kim<sup>††</sup>

### 요 약

본 연구는 학습자의 성취도를 예측하기 위한 TLBoost 모델을 제안한다. TLBoost 모델은 범주형 데이터와 연속형 데이터를 효과적으로 처리하며, 예측 결과의 해석 가능성을 제공하는 트리 기반의 딥러닝 모델이다. TLBoost는 TabNet과 LightGBM 모델을 앙상블하여 각 알고리즘의 장점을 결합함으로써 성능을 개선하고 모델의 안정성과 일반화 능력을 강화한다. TLBoost는 TabNet이 추출한 특징을 LightGBM의 입력으로 사용하여, LightGBM이 최종 예측 성능을 향상시킨다. 연구 결과, TLBoost 모델은 수학과 언어 성취도 예측에서 기존 모델보다 뛰어난 성능을 보였다. 해석 가능성을 높이기 위해 두 가지 주요 방법을 사용하였다. 첫째, TabNet의 어텐션 메커니즘을 활용하여 데이터의 중요한 특징을 시각적으로 강조하였다. 둘째, LightGBM 모델의 특징 중요도 분석을 통해 각 특징이 모델 예측에 미치는 영향을 시각적으로 표현하였다. 이를 통해 교수자는 학습자의 성적에 영향을 미치는 주요 요소들을 이해하고 맞춤형 학습 지원을 제공하여 기술의 발전과 교육 현장 적용 사이의 간극을 좁히고, 교육 정책의 실효성을 높일 수 있다.

**주제어:** 학습자 성취도 예측, 범주형 데이터, 앙상블 모델, 특징 분석, 설명 가능성, 현장 적용

### ABSTRACT

This study proposes the TLBoost model to predict student achievement. The TLBoost model effectively handles both categorical and continuous data while providing interpretability of prediction results through a tree-based deep learning approach. By ensembling TabNet and LightGBM models, TLBoost combines the strengths of each algorithm to enhance performance and improve model stability and generalization capabilities. TLBoost uses features extracted by TabNet as inputs for LightGBM, thereby improving the final predictive performance of LightGBM. The results show that the TLBoost model outperforms existing models in predicting math and language achievement. To increase interpretability, two main methods were employed. First, the attention mechanism of TabNet was utilized to visually highlight important features in the data. Second, feature importance analysis of the LightGBM model was conducted to visually represent the impact of each feature on the model's predictions. Through this approach, educators can understand the key factors influencing student performance and provide tailored learning support, thus bridging the gap between technological advancements and practical application in educational settings, and enhancing the effectiveness of educational policies.

**Keywords:** Prediction of Student Achievement, Categorical Data, Ensemble Model, Feature Analysis, Interpretability, Practical Application

## 1. 서론

최근 교육부는 교육 현장에 컴퓨터 기반 학업 성취도 평가의 전면 도입을 계획하고 있다. 컴퓨터 기반 학업

성취도 평가는 2022년 초등학교 6학년, 중학교 3학년, 고등학교 2학년을 대상으로 시작하여, 2024년에는 초등학교 3학년부터 고등학교 2학년까지 대상 학습자를 확대하여 적용될 예정이다[1]. 평가의 목적은 컴퓨터

<sup>†</sup>정 회 원: 고려대학교 일반대학원 컴퓨터학과 박사수료

<sup>††</sup>중 심 회 원: 고려대학교 컴퓨터학과 교수(교신저자)

시스템을 이용하여 측정한 객관적인 학습자 진단 데이터를 활용하여 개인별 맞춤형 학습 지원을 제공하는 것이다[2]. 이를 위해 적용 대상 학교에 평가도구를 제공하고 진단 데이터를 빅데이터 형태로 구축하여 디지털 교수 및 학습 통합 플랫폼과 연계할 계획이다[3].

그러나, 학교 현장에서 사용할 플랫폼의 실효성은 미지수이며, 기술을 교육에 유의미하게 적용하기 위해 극복해야 할 장벽들이 있다. 첫째, 학교 현장에서 주로 사용하는 범주형 데이터를 기반으로 하는 모델이 부족하다. 성취도 예측을 위해 고도화된 인공지능 기법들이 적용되었지만 문항에 따른 정 오답 기반의 단순한 이진 데이터 기반 모델이 많다[4]. 또한 현장에서는 오프라인 교재 사용이 일반적이며 성취도 점수 외에 개별 문항에 대한 정 오답 데이터를 효과적으로 수집하기 어려운 상황이다[5]. 학습자의 성취도는 정 오답 뿐만 아니라 범주형 데이터에 해당하는 학습자의 개인적, 환경적 요인에 의해 달라질 수 있다[6]. 예를 들어 학습자 부모의 사회경제적 지위, 개인의 동기부여, 낙제 경험 등은 학습자 성취도에 영향을 미칠 수 있다[7][8]. 현장에서 수집되어 사용되는 학습자 데이터셋에는 성취도 점수인 숫자형 변수와 요인을 다루는 범주형 변수가 혼합되어 있다. 범주형 데이터를 사용하기 위해서 one-hot 인코딩이 필요하지만 one-hot 인코딩은 Tree-based 모델의 성능을 저하시킨다[9]. 이에 범주형 데이터의 중요도에도 불구하고 학습자의 성취도를 예측할 때 범주형 변수를 제외하고 연속형 데이터를 사용한다[10][11].

둘째, 기존 성취도 예측 모델의 해석성 부족으로 현장 적용 시 실효성이 부족하다. 교수자가 학습자들에게 지도를 제공하기 위해서는 모델의 예측 근거를 명확히 이해할 필요가 있다. 그러나 성취도 예측 모델이 딥러닝을 기반으로 하고 있어, 블랙박스 현상이 발생하고 이에 따라 예측의 근거를 해석하는 데 어려움이 있다[12].

연구의 목적은 범주형 데이터를 활용하여 성취도 예측의 정확도를 높이고 기본 연구의 블랙박스 문제를 해결하는 것이다. 이를 위해 학습자의 성취도를 예측하기 위한 TLBoost 모델을 제안한다. TLBoost 모델은 TabNet과 LightGBM 모델을 앙상블하여 성능을 개선하고, 예측 결과의 해석 가능성을 제공함으로써 교육 현장에서 실질적으로 활용될 수 있는 모델을 개발하는 것을 목표로 한다. TabNet의 어텐션 메커니즘을 통해 중요한 특징을 추출하고, 이를 LightGBM의 입력으로 사용하여 예측 성능을 높였다. 실험 결과 같은 계열의 모델에서 가장 뛰어난 성능을 보였다. 예측에 사용된 중요도 특징 추출과 시각화를 통해 예측의 해석성을 제

공하며 블랙박스 문제를 해결한다. 본 연구는 학습자의 성취도를 예측하는 새로운 모델을 제안함으로써, 교수자들이 학습자의 성적에 영향을 미치는 주요 요소들을 이해하고 맞춤형 학습 지원을 제공하는 데 기여할 수 있다. 또한, 기술의 발전과 교육 현장 적용 사이의 간극을 좁히고, 교육 정책의 실효성을 높이는 데 도움을 줄 수 있다. 본 연구에서 사용된 데이터는 포르투갈 학습자의 실제 데이터로 국내 연구에 직접 적용하기 한계가 있을 수 있다. 실험은 수학 및 언어 과목의 성취도를 예측하여 다른 과목이나 다양한 교육 환경에 대한 적용 가능성은 추가 연구가 필요할 수 있다.

## 2. 관련 연구

### 2.1 학습자 성취도 분석

#### 2.1.1 학습자 성취도 예측

학습자의 학습 활동에 대한 데이터 분석이 정교화됨에 따라, 데이터를 통합 및 분석하여 학습자에게 맞춤형 학습을 제공하려는 시도가 확산되고 있다[13]. 학습자들의 성적을 예측하는 것은 학습 분석의 주요 과제이다. 성취도 예측은 교수자들이 학습자들을 조기에 식별하고 교육적인 조치를 취할 수 있게 하기 때문이다. 조기 식별과 조치는 학습자들의 성적 향상에 도움을 줄 수 있다[14]. 기존에는 학습자의 정오답 데이터만 사용한 모델이 많았기에 학습자의 현실을 반영하여 효과성 있는 맞춤형 학습을 하기에 한계가 있었다. 최근 연구에서는 학습자의 정보를 다양하게 수집하여 학습자의 성취도를 예측하는 연구들이 이루어지고 있다[15]. 성취도 예측 모델은 학습자의 정오답 데이터만을 가지고 성취도를 예측했을 때보다 학습자의 다양한 요인과 함께 성취도를 예측했을 때 정확도가 높다는 연구 결과가 유의미하게 존재한다[16]. 학습자의 요인과 성취도 분석은 단순한 통계 및 회귀 분석으로도 가능하지만, 더 높은 성능과 다양한 활용을 위하여 딥러닝과 머신러닝을 활용하는 사례가 증가하고 있다[10].

#### 2.1.2 환경적 및 개인적 요인 분석

학습자의 성취도는 여러 요인에 의해 영향을 받을 수 있으며 요인들은 환경적 요인과 개인적 요인으로 구분할 수 있다. 환경적 요인은 교육 환경 및 과정과 관련된

요인이다. 부모님의 학력, 직업, 주 양육자, 부모와 거주 여부, 가족의 크기, 시골 및 도시 거주 여부, 학교에서 집까지 걸리는 시간, 학교의 교육적 지원, 가정에서의 교육적 지원, 사교육 여부, 집에서 인터넷 연결 여부, 가족과의 관계 등이 있을 수 있다. 그리고 개인적 요인은 학습자의 개인적 상황과 성향과 관련된 요인이다. 학습자의 성별, 나이, 학교 선택 이유, 공부 시간, 결석 횟수, 낙제 경험, 고등 교육으로 진학 동기 여부, 연애 여부, 방과 후 자유 시간의 정도, 친구와 만나는 시간 정도, 알코올 소비량, 건강 상태 등이 있다. 위의 요인은 본 연구에서 가지고 있는 데이터를 선행연구[8]를 바탕으로 분류한 것이며, 표 2에서 확인할 수 있다.

그러나 환경적 요인과 개인적 요인들은 일반적인 성취도 예측 모델에서 활용되지 않고 있다. 요인은 범주형 데이터로 이루어져 있지만, 성취도 예측 딥러닝 모델에서 사용하는 대규모 데이터셋이 문항 정보의 정오답 여부에 초점이 맞추어져 있기 때문이다[17]. 종합적인 진단과 정확한 성취도 예측을 위하여 환경적 및 개인적 요인에 따라 학습자의 성취도를 예측할 수 있는 범주형 데이터를 다루는 모델의 연구가 필요하다.

## 2.2 앙상블 예측 모델

앙상블 학습 방법은 여러 개의 학습 알고리즘을 결합하여 정확한 예측 모델을 구축하는 기법이다[19]. 기법들은 모델의 약점을 상호 보완함으로써 전반적인 성능 향상을 목표로 하며 주로 분류 및 회귀 문제에서 사용된다[18].

GBM(Gradient Boosting Machine)은 앙상블 기법 중 한 종류로 여러 개의 약한 결정 트리를 순차적으로 학습시켜 각 단계에서 이전 모델의 오차를 줄이는 방식으로 작동한다. 각각의 약한 학습기가 서로를 보완하며 전체 모델의 예측 성능을 향상시킨다[19]. GBM은 LightGBM, CatBoost 알고리즘 등과 같은 발전된 부스팅 방법론의 기초가 된다. 교육용 데이터의 주류인 테이블 형태의 정형 데이터에서 딥러닝 보다 높은 설명 가능성을 제공할 수 있으며 데이터의 어떠한 특성이 모델의 최종 예측에 중요했는지 근거를 제공할 수 있다[20]. GBM 모델은 이진, 이산, 범주형 데이터에서 사용될 수 있지만 범주형 변수를 직접 처리할 수 없어 데이터 전처리 단계에서 수치형으로 변환하기 위해서 시간과 비용이 든다. 범주형 데이터 인코딩 문제를 해결하기 위하여 범주형 데이터를 직접 처리할 수 있는 LightGBM, CatBoost 알고리즘을 주로 사용한다[21].

### 2.2.2 LightGBM 알고리즘

LightGBM 알고리즘은 트리 기반의 모델로 범주형 데이터를 직접 다루며 손실 함수를 최소화하는 방향으로 모델을 반복적으로 업데이트한다[22]. 각 단계에서 그라디언트(Gradient)와 헤시안(Hessian)을 계산하여 가장 최적의 분할을 찾고 데이터를 분할하여 최대의 이득을 얻는 방향을 모색하며 정보 이득은 수식 (1)과 같이 계산한다[22]. EFB는 고차원 특징 공간 문제를 해결하기 위해 상호 배타적인 특징들을 번들링한다. 이는 동시에 0이 아닌 값을 가지는 경우가 드문 특징들을 하나로 묶어 계산 복잡도를 줄이고 성능을 향상시키는 방법이다. LightGBM은 범주형 데이터 인코딩을 위하여 내장 범주형 처리, one-hot 인코딩, 순위 인코딩, 빈도 인코딩, 타겟 인코딩을 사용한다. 또한, 각 카테고리 특성에 따른 중요도를 출력할 수 있어, 결과물의 근거를 시각화하고 해석성을 높인다.

$$G = \frac{1}{2} \left( \frac{(\sum_{i \in L} g_i)^2}{\sum_{i \in L} h_i + \lambda} + \frac{(\sum_{i \in R} g_i)^2}{\sum_{i \in R} h_i + \lambda} - \frac{(\sum_i g_i)^2}{\sum_i h_i + \lambda} \right) - \gamma \quad (1)$$

### 2.2.3 CatBoost 알고리즘

CatBoost 알고리즘은 트리 기반 모델로 Cardinality가 높은 범주형 변수를 기존의 one-hot 인코딩 방식 대신 데이터의 순서를 고려하는 Ordered Target 인코딩 방식으로 해결한다. 기존의 GBM 모델이 잔차에 집중하여 과적합되었던 문제점을 각 단계에서 샘플링을 변경하여 과적합을 줄이는 기법을 수식 (2)와 같이 사용하여 변수의 Cardinality가 높아도 성능 저하 없이 학습을 진행할 수 있게 한다[23].

$$h^t = \arg \min_{(h \in H)} \frac{1}{n} \sum_{(k=1)}^n [-g^t(x_k, y_k) - h(x_k)]^2$$

$$\hat{x}_k^i = (\sum_{x_j \in \Omega} [1_{x_j = x_k^*} y_j] + a^* P) / (\sum_{x_j \in \Omega} [1_{x_j = x_k^*} y_j] + a) \quad (2)$$

Ordered TBS(target-based statistics)는 카테고리 별로 타겟 값의 평균을 이용하여 범주형 변수를 변환하는 기존 인코딩 방법을 개선한 것으로, 각 데이터 포인트를 처리할 때 해당 데이터 포인트 이후의 타겟 값을 사용하지 않고 이전의 값들만을 사용한다. Ordered Boosting은 각 부스팅 단계마다 데이터셋을 독립적으로 샘플링하고, 학습 중인 모델을 적용하여 잔차를 계산하

는 방식으로 GBM 모델의 데이터 누수 문제를 해결한다.

## 2.3 범주형 데이터 딥러닝 모델

### 2.3.1 TabNet

TabNet은 범주형 및 수치형 데이터를 처리하는 딥러닝 알고리즘으로, 테이블 형식의 데이터를 다루는 데 최적화되어 있다. TabNet은 가공하지 않은 기본 데이터에 그래디언트를 기반으로 최적화를 함으로써 end to end 사용이 가능하다. TabNet은 의사결정 단계에서 변수 선택이 가능하도록 어텐션 구조(attention mechanism)를 활용하였다[24]. 이 때 순차적인 어텐션 메커니즘을 사용함으로써 모델의 성능과 해석성을 향상시킬 수 있다.

$$h_i^{(t+1)} = ReLU(W_i h_i^{(t)} + b_i) \quad (3)$$

$$a_i^{(t)} = Sparsemax(M_i^{(t)} h_i^{(t-1)}) \quad (4)$$

$$x_i^{(t+1)} = x_i^{(t)} \odot a_i^{(t)} \quad (5)$$

주요 구성 요소는 Feature Transformer 수식(3)과 Attentive Transformer 수식 (4), (5)이다[24]. Feature transformer는 전처리 과정 없이 입력 값을 받아 변수를 처리하며 중요한 정보를 보존하고, 다음 의사결정 단계에 전달한다. Attentive Transformer는 이전 의사결정 단계에서 선택된 변수를 반영하여, 현재 단계에서 마스킹(masking)할 변수를 결정한다. 수식 (6)로 최종 예측을 통해 TabNet은 중요한 특성을 선택하고 학습하여 모델의 성능을 높인다[24]. 모델의 각 단계에서 선택된 특성과 중요도를 확인할 수 있어 해석이 용이하며, 대규모 데이터셋에서도 효율적인 학습이 가능하다.

$$\hat{y}_i = \sigma(W_{out} h_i^{(L)} + b_{out}) \quad (6)$$

### 2.3.2 TabNet과 LightGBM의 결합

TabNet 모델과 LightGBM의 결합한 연구들을 살펴본다. Tianyu Xu, Yongchuan Yu, Jianzhuo Yan, Hongxia Xu은 TabNet 모델과 LightGBM을 결합하여 강우 예측을 하였다. 해당 논문에서 제안된 모델 결합 방법은 각 모델을 개별적으로 학습 시킨 다음, TabNet

과 LightGBM 모델을 각각 최적의 성능 상태로 만드는 것이다. 이 후 각 학습된 모델의 예측 결과를 결합하고, 이 때 단순 평균 방식을 사용하였다[25]. Sanket Vilas Salunke, Abdelkader Ouda은 LightGBM, XGBoost, Neural Network, TabNet, 1D-CNN 모델을 앙상블 하였고, 하드 투표(Hard Voting), 소프트 투표(Soft Voting), 스택킹(Stacking) 방법을 사용하였다. 앙상블 한 모델은 총 4가지로 LightGBM, XGBoost, Neural Network, TabNet를 앙상블, LightGBM, XGBoost, Neural Network를 앙상블, LightGBM, XGBoost를 앙상블, Neural Network, TabNet, 1D-CNN를 앙상블하며 어떤 방법이 가장 좋은 앙상블 방법과 조합인지 확인하는 연구이다[26].

본 연구는 기존연구와 같이 단순히 예측 결과를 결합하는 것이 아니라, TabNet이 추출한 중요한 특징 정보를 LightGBM이 학습하도록 하여 두 모델의 장점을 극대화하는 방식으로 차별성을 가진다.

## 2.4 실험 평가 지표

본 연구는 학습자의 성취도를 예측하는 회귀 문제이기 때문에 RMSE,  $R^2$  Score를 사용한다[25]. RMSE(Root Mean Squared Error)는 식 (7) 과 같이 예측 값과 실제 값 차이를 제곱한 후 평균을 계산한 값의 제곱근이다[27]. MSE에 루트는 씌워서 에러를 제곱해서 생기는 값의 왜곡을 줄인다. 오차의 크기를 원점수의 단위로 해석할 수 있게 한다. 값이 낮을수록 높은 성능을 말한다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (7)$$

$R^2$  Score는 Adjusted R-squared, 결정계수로 독립 변수의 수가 모델에 미치는 영향을 조정한다[26]. 독립 변수의 수가 많은 학습자 범주형 데이터이기 때문에 조정된  $R^2$  Score를 사용한다. 위의 RMSE의 값은 모델마다 다를 수 있어 절대 값만 보고 성능을 판단하기 오류가 있을 수 있다. 이에 식(8)에서  $R^2$  Score로 실제 값의 분산 대비 예측 값의 분산 비율을 확인한다[28]. 값이 1에 가까울수록 성능이 높다. 본 연구의 실험에서 사용된 GBM 계열인 LightGBM, CatBoost 모델 및 딥러닝 모델 TabNet, TCBoost, TLBoost 모두 RMSE와  $R^2$  Score를 통해 모델 성능을 평가한다.

$$R^2 = 1 - (SSE/SST)$$

$$R^2 = 1 - (\sum_{(i=1)}^n (y_i - \hat{y}_i)^2) / (\sum_{(i=1)}^n (y_i - \bar{y})^2) \quad (8)$$

### 3. TLBoost: 학습 성취도를 위한 설명가능한 앙상블 예측 모델

본 연구는 학습자의 성적을 예측하는 데 중점을 두고 있으며, 범주형 데이터를 포함한 정형 데이터에 대한 예측 성능과 해석성을 탐구한다. 학습자의 성취도에 영향을 미치는 다양한 요소들을 개인적 특성, 가정 환경, 학습 습관 범주형 데이터로 분류하고, 예측 근거인 특징 중요도를 시각화하여 예측 결과의 해석가능성을 높였다.

TLBoost 모델은 TabNet과 LightGBM 모델을 앙상블하여 각 알고리즘의 장점을 취합한다. 앙상블을 통해 TabNet과 LightGBM이 개별적으로 학습한 정보를 모두 사용할 수 있으므로, 단일 모델보다 성능이 개선되고 모델의 안정성과 일반화 능력을 강화시킬 수 있다. TabNet은 비선형적인 복잡한 패턴을 잘 학습하고, LightGBM은 Boosting을 통해 높은 예측 성능을 제공하기 때문에 상호 보완적이다.

이 때 TabNet을 먼저 사용하고 LightGBM를 결합한다. TabNet은 신경망 기반 모델로 복잡한 선형 및 비선형 관계를 학습할 수 있으며, 범주형 및 연속형 데이터를 모두 처리할 수 있어 특징을 학습하는 데 효과적이다. 또한 데이터의 중요한 특징을 선택하고 강조하는 어텐션 메커니즘을 사용하여 초기 단계에서 중요한 패턴을 추출하는 데 효과적이다. 이에 TabNet이 추출한 데이터의 특징을 LightGBM의 입력으로 사용함으로써 두 모델의 강점을 결합한다. 즉, TabNet의 예측값을 새로운 특징으로 추가함으로써, LightGBM이 최종 예측 성능을 향상시키며 차별성을 가진다.

또한, TLBoost 모델은 해석 가능성을 크게 향상시킨 점에서 차별화된다. TabNet의 어텐션 메커니즘을 활용하여 데이터의 중요한 특징을 시각화하고, LightGBM의 특징 중요도 분석을 통해 예측에 기여한 각 특징의 영향을 시각적으로 표현한다. 이는 교수가 학습자의 성적에 영향을 미치는 주요 요소들을 명확히 이해할 수 있도록 돕는다. 기존 연구에서는 이러한 해석 가능성에 중점을 두지 않은 반면, TLBoost 모델은 두 가지 독립적이며 상호보완적인 시각화 기법을 통해 해석 가능성을 극대화한다. 이를 통해 교수는 학습자의 성적에

영향을 미치는 주요 요소들을 식별할 수 있다. 모델의 구성은 그림 1과 같다. TLBoost 모델의 구체적인 차별성은 3.1과 3.2에서 설명한다.

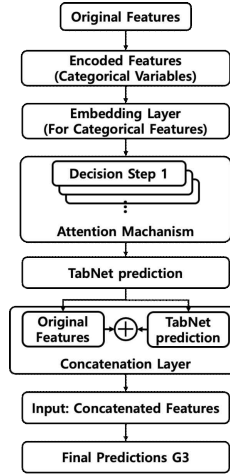


Figure 1. TLBoost Architecture

#### 3.1 TabNet 예측값을 새로운 특징으로 추가

TabNet 모델은 수식 (9)과 같이 입력 데이터에 따라 대상 값을 사용하여 학습하고 학습된 모델은 예측 값을 생성한다.

$$\begin{aligned} \hat{y}_{(TabNet,train)} &= f_{TabNet}(X_{train}) \\ \hat{y}_{(TabNet,test)} &= f_{TabNet}(X_{test}) \end{aligned} \quad (9)$$

TabNet 모델의 예측값  $\hat{y}_{TabNet}$ 을 원본 입력 데이터  $X_{train}$ 과  $X_{test}$ 에 새로운 특징으로 추가하여 LightGBM 모델의 입력 데이터  $X_{TLBoost}$ 를 수식 (10)과 같이 생성한다.  $X_{train TLBoost}$ 와  $X_{test TLBoost}$ 은 TabNet 예측값을 추가한 새로운 학습 및 테스트 입력 데이터이다.

$$\begin{aligned} X_{train TLBoost} &= [X_{train}, \hat{y}_{(TabNet,train)}] \\ X_{test TLBoost} &= [X_{train}, \hat{y}_{(TabNet,test)}] \end{aligned} \quad (10)$$

LightGBM 모델은 새로운 입력  $X_{train TLBoost}$ 와 대상 값  $y_{train}$ 을 사용하여 학습하고 학습된 모델은 수식 (11)과 같이 최종 예측값을 생성한다.

$$\begin{aligned} \hat{y}_{(LightGBM,train)} &= f_{LightGBM}(X_{train} TLBoost) \\ \hat{y}_{(LightGBM,test)} &= f_{LightGBM}(X_{test} TLBoost) \end{aligned} \quad (11)$$

### 3.2 해석 가능성 향상

TLBoost 모델의 해석 가능성은 독립적인 두 가지 주요 방법으로 향상된다. 첫째, TabNet의 어텐션 메커니즘을 활용하여 데이터의 중요한 특징 부분을 강조한다. 어텐션 메커니즘은 각 특성의 어텐션 가중치를 계산하여 중요도를 부여하며 소프트맥스 함수를 사용하여 수식 (12)와 같이 계산된다. 둘째, LightGBM 모델은 각 특성의 중요도를 계산하여 모델 예측에 대한 기여도를 평가한다. 특정 중요도는 각 트리에서 특성의 기여도를 합산하여 계산된다. 특정 중요도 계산은 수식 (13)과 같다. 이를 토대로 시각화하여 해석성을 향상시킨다.

$$\text{Attention Weights}_i = \frac{e^{a_i}}{\sum_{j=1}^n e^{a_j}} \quad (12)$$

$$\text{Feature Importance}_i = \frac{1}{T} \sum_{t=1}^T \Delta L_{t(i)} \quad (13)$$

## 4. 실험 방법 및 결과

### 4.1 데이터 수집

본 연구에서 분석에 활용된 데이터는 포르투갈의 실제 두 학교—Gabriel Pereira와 Mousinho da Silveira—의 학습자들로부터 수집된 것으로, Kaggle을 통해 획득했다. 데이터는 수학과 언어 과목에 대한 학습자들의 성적을 포함하며, 범주형 및 숫자형 변수로 이루어져 있다. 데이터셋에는 학습자들의 인구통계학적 특성, 사회적 배경, 학교 환경과 관련된 특성이 포함되어 있으며, 학교 보고서 및 설문지를 통해 수집되었다. 데이터에 대한 구체적인 설명은 표 1에 제시되어 있다.

**Table 1.** Environmental & Personal data description

Variable	Description
school	student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
sex	student's sex (binary: 'F' - female or 'M' - male)

age	student's age (numeric: from 15 to 22)
address	student's home address type (binary: 'U' - urban or 'R' - rural)
famsize	family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
Pstatus	parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
Medu	mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
Fedu	father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
Mjob	mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
Fjob	father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
reason	reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
guardian	student's guardian (nominal: 'mother', 'father' or 'other')
traveltime	home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
studytime	weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
failures	number of past class failures (numeric: n if 1<=n<3, else 4)
schoolsup	extra educational support (binary: yes or no)
famsup	family educational support (binary: yes or no)
paid	extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
activities	extra-curricular activities (binary: yes or no)
nursery	attended nursery school (binary: yes or no)
higher	wants to take higher education (binary: yes or no)
internet	Internet access at home (binary: yes or no)
romantic	with a romantic relationship (binary: yes or no)
famrel	quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
freetime	free time after school (numeric: from 1 - very low to 5 - very high)
gout	going out with friends (numeric: from 1 - very low to 5 - very high)
Dalc	workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
Walc	weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
health	current health status (numeric: from 1 - very bad to 5 - very good)
absences	number of school absences (numeric: from 0 to 93)
G1	first period grade (numeric: from 0 to 20)
G2	second period grade (numeric: from 0 to 20)
G3	final grade (numeric: from 0 to 20, output)

	target)
--	---------

실험에 사용된 데이터셋은 학습자들의 수학과 언어 과목에 대한 데이터를 포함한다. 데이터는 28개의 범주형 변수와 3개의 숫자형 변수로 구성되어 있다.

**Table 2.** Data Description

Dataset	Math	Language
Student	395	649
School (GP)	349	423
School (MS)	46	226
Features (Numerical)	5	5
Features (categorical)	28	28

수학 과목의 데이터는 395명, 언어 과목의 데이터는 649명을 포함하여 총 1,044명의 학습자 데이터로 이루어져 있으며 구체적인 설명은 표 1과 표 2에 제시되어 있다.

## 4.2 실험방법

### 4.2.1 비교모델 실험방법

본 연구에서는 학습자 성취도 예측을 위해 LightGBM, CatBoost, TabNet 모델을 활용하였으며, 제안하는 모델인 TLBoost와의 성능 비교를 위해 TCBoost 모델을 개발하였다. TCBoost는 TabNet과 CatBoost를 앙상블한 모델로, TLBoost와 유사하게 TabNet 모델의 예측값을 새로운 특징으로 추가하여 CatBoost로 학습하고 최종 예측을 수행한다. LightGBM과 CatBoost는 모두 특징 추출 및 해석이 가능한 트리 기반 알고리즘으로, 비교 대상으로 선정하였다.

실험에 사용된 데이터는 학습자들의 배경 정보와 성취도를 포함하고 있으며 각 변수는 범주형 데이터와 연속형 데이터를 포함한다. 타겟 변수는 학생 성취도 점수(G3)이며, 특징은 나머지 모든 변수들로 구성된다. 전처리 과정에서 LightGBM는 범주형 변수를 카테고리 타입으로 변환하였다. CatBoost는 범주형 변수를 식별한 뒤 데이터 프레임의 열 이름을 사용하여 범주형 변수의 인덱스를 추출하여 모델에 입력하였다. TabNet은 범주형 변수를 직접 처리할 수 있도록 카테고리 인코딩이 필요하여, 범주형 변수를 숫자로 인코딩하였다. TCBoost와 TLBoost는 범주형 특징 인코딩 기법을 활용하여, 선택된 범주형 특징을 하나씩 순회하며 각 특

징을 숫자형 코드로 변환하였다.

실험에 사용한 알고리즘 모두 전처리가 완료된 데이터를 학습: 평가 = 8: 2 비율로 분할하였다. 교차검증을 위하여 5개의 폴드로 나누고, 데이터를 무작위로 섞기 위해 shuffle=True와 random\_state=42를 설정하였다. 모델의 하이퍼파라미터를 최적화하기 위해 Optuna 프레임워크를 사용하고 50회의 시도를 수행하였다. 최적화 목적 함수를 정의하고, 함수 내에서 하이퍼파라미터를 샘플링하고 각 모델을 학습 및 평가하였다. 과적합을 방지하기 위하여 데이터를 분할하고 조기 종료를 사용할 수 있게 하였다. 학습된 모델을 분할한 데이터셋을 사용하여 학습하고, 타겟 변수 G3를 예측하여 RMSE와 R<sup>2</sup> Score를 통해 모델 성능을 평가하였다. 학습된 모델의 중요도를 추출하였고 중요도 값을 바탕으로 시각화하였다. LightGBM, CatBoost, TabNet, TCBoost, TLBoost 모델 모두 CPU(x86\_64)를 사용하였다.

### 4.2.2 TLBoost 실험방법

데이터를 특징(X)과 타겟(y)로 분리하였으며, 타겟은 학습자 성취도를 나타내는 G3이다. 범주형 특징을 식별하여 범주형 타입으로 변환 후, 숫자로 인코딩하였다. 데이터를 학습 데이터와 테스트 데이터로 분할하였으며 비율은 8:2로 설정하고 random\_state를 42로 고정하였다. Optuna를 사용하여 총 50번의 시도로 TabNet 및 LightGBM 모델의 하이퍼파라미터를 최적화하였다. 목적 함수는 TabNet 모델의 하이퍼파라미터를 샘플링, TabNet 모델 학습한 예측값, LightGBM 모델의 하이퍼파라미터 샘플링, LightGBM 모델을 TabNet의 예측값을 포함한 학습 데이터로의 학습, 최종 예측값을 계산하여 반환하는 것을 포함한다. 최적화된 하이퍼파라미터를 사용하여 TabNet 모델을 다시 학습하였으며, TabNet 모델의 예측값을 사용하여 학습 데이터를 확장하였다. 확장된 학습 데이터를 사용하여 LightGBM 모델을 학습하였다. 최종적으로 LightGBM 모델의 예측값을 얻어 RMSE와 R<sup>2</sup> Score를 계산한다. 특징 중요도의 해석성을 위하여 TabNet 모델의 Attention score를 나타내는 distribution를 각 결정 단계에서 시각화한다. 각 결정 단계마다 별도의 distribution를 생성하여 특징 중요도를 비교하였다. 또한 LightGBM 모델의 특징 중요도를 추출하여 이를 그래프로 시각화하였으며 중요도가 높은 순으로 정렬하였다.

### 4.3 실험 결과

실험 결과는 표 3과 같이 TLBoost 모델이 수학과 언어 과목 성취도 예측 모두에서 최고의 성능을 보였다. LightGBM과 CatBoost, 그리고 TabNet을 단일 모델로 사용했을 때의 성능보다 각 모델을 앙상블 하였을 때 성능이 높게 나왔다. TabNet과 LightGBM을 결합하였을 때 가장 높은 성능을 보였으며 이는 앙상블이 모델의 예측 정확도와 안정성을 크게 향상시켰음을 의미한다. 또한 TLBoost 모델의 높은 R<sup>2</sup> Score는 모델이 실제 성취도 변동을 설명할 수 있음을 나타낸다.

**Table 3.** Prediction with different models

Subject	Math		Language	
	RMSE	R <sup>2</sup> Score	RMSE	R <sup>2</sup> Score
LightGBM	2.0295	0.7991	1.2716	0.8341
CatBoost	2.0077	0.8034	1.3509	0.8128
TabNet	2.6799	0.6497	1.4624	0.7806
TCBoost	1.7519	0.8503	1.2057	0.8509
<b>TLBoost</b>	<b>1.7040</b>	<b>0.8583</b>	<b>1.0694</b>	<b>0.8827</b>

본 연구에서는 수학 성취도와 언어 성취도를 예측하기 위해 나누어 실험을 수행하였다. 이에 각 과목별로 모델의 성능을 평가하고, 예측 중요도를 설명한다. 다음 표와 그림은 은 각 과목에 따른 모델의 예측 결과 및 시각화한 내용이다.

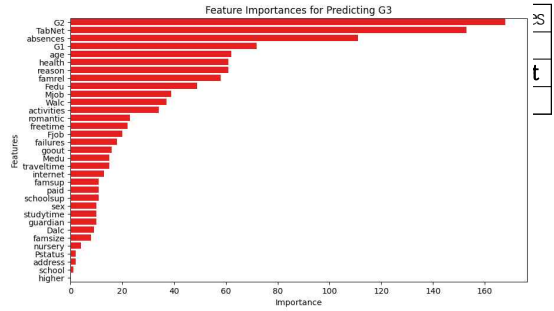
#### 4.3.1 수학 성취도 예측

표 3에서 수학 성취도 예측에서 TLBoost 모델은 가장 낮은 RMSE(1.7040)와 가장 높은 R<sup>2</sup> Score(0.8583)를 기록하여 높은 예측 성능을 보였다. 비슷한 계열의 앙상블인 TCBoost 모델도 높은 성능을 보였으며, RMSE는 1.7519 R<sup>2</sup> Score는 0.8509로 나타났다.

그림 2은 TLBoost 모델이 최종 예측에서 수학 성취도 G3를 예측할 때 중요하게 사용된 특징을 중요도 순으로 시각화한 것이다. 실험에 사용된 모델 모두 특징 중요도를 추출할 수 있어 표 4와 같이 정리하였다.

**Table 4.** Feature Importances for math G3

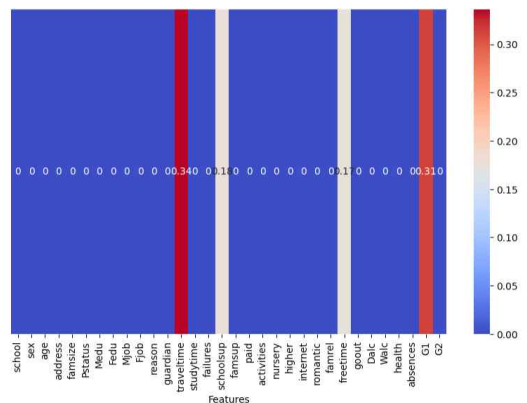
Subject	Math				
	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>
LightGBM	G2	G1	activities	romantic	guardian



**Figure 2.** Feature Importance for math G3

표 4에서 LightGBM 모델은 G2, G1, activities, romantic, guardian 순으로 중요도가 높은 특징을 사용하여 예측했다. CatBoost 모델은 G2, 결석(absences), G1, 나이(age), activities 순으로 중요도가 높은 특징을 사용하여 예측했다. TabNet 모델은 G2, G1, 아버지 교육 수준(Fedu), 자유 시간(freetime), 나이(age) 순으로 중요도가 높은 특징을 사용하여 예측했다. TCBoost 모델은 G2, 결석(absences), G1, 나이(age), activities 순으로 중요도가 높은 특징을 사용하여 예측했다. TLBoost 모델은 G2, TabNet 모델의 예측 결과, 결석(absences), G1, 나이(age) 순으로 중요도가 높은 특징을 사용하여 예측했다.

표 4에서 “TabNet“이라는 항목은 TabNet 모델에서 생성된 예측 결과를 LightGBM 모델에 입력으로 사용한 것을 나타낸다. 즉, TabNet의 예측 값이 하나의 특징으로서 LightGBM 모델에 제공되었으며, 최종 예측에서 데이터의 33개의 특징 중에서 두 번째로 중요하게 사용되었다. 이는 TabNet 모델의 예측이 다른 모델에서도 중요한 예측 변수로 활용될 수 있음을 보여준다. 표 3과 그림 2의 결과로 TabNet의 예측이 최종 예측에 사용될수록 예측 성능이 높은 것을 확인할 수 있다.



**Figure 3.** Attention distribution of features for math 그림 3은 TabNet 모델의 Attention distribution을 시



각화 한 것이다. Attention score를 나타내며, 무작위로 추출한 결정 단계 3에서 특징의 중요도를 점수로 표현한다. 붉은색일수록 높은 값을 나타내며 중요한 특징으로 표시된다. X축은 특징들의 이름을 나타내며, Y축은 중요도를 나타낸다.

#### 4.3.2 언어 성취도 예측

표 3에서 언어 과목의 성취도 예측 또한 TLBoost 모델이 가장 우수한 성능을 보였다. RMSE는 1.0694, R<sup>2</sup> Score는 0.8827로, 다른 모델들에 비해 높은 예측 성능을 보였다. TCBoost 모델이 두 번째로 높은 성능을 기록하였다.

중요한 특징들을 순위별로 정리한 것이다. LightGBM 모델은 G2, G1, nursery, 성별(sex), activities 순으로 특징을 사용하여 예측했다. CatBoost 모델은 G2, G1, 결석(absences), 나이(age), 평일 음주(Dalc) 순으로 특징을 사용하여 예측했다. TabNet 모델은 G2, 어머니 교육 수준(Medu), 통학 시간(traveltime), 실패 횟수(failures), 나이(age) 순으로 특징을 사용하여 예측했다. TCBoost 모델에서는 TabNet의 예측 값이 두 번째로 중요한 특징으로 사용되었고, TLBoost 모델에서는 가장 중요한 특징으로 사용되었다. TLBoost뿐만 아니라 좋은 성능을 내는 TCBoost에서도 예측을 위한 중요한 특징으로 사용되었다. TabNet이 높은 특징 중요도로 사용될 때 모델의 성능을 향상시키는데 기여할 수 있다는 것을 알 수 있다.

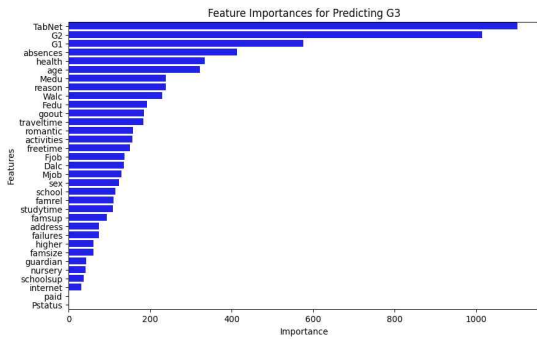


Figure 4. Feature Importance for Language G3

그림 4는 TLBoost가 언어 성취도를 최종 예측할 때 중요하게 사용된 특징을 중요도 순서로 나열하여 시각화한 것이다. X축은 특징의 중요도를 나타내며 값이 클수록 해당 특징이 예측 모델에 중요함을 의미한다. Y축은 모델에 사용된 특징의 이름이며, TabNet도 포함되어 있다. 그래프에 따르면 TabNet 모델에서 생성된 예측값이 TLBoost 모델에서 가장 중요한 특징으로 사용되어 있다.

Table 5. Feature Importances for Language G3

Subject	Language				
	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>	5 <sup>th</sup>
LightGBM	G2	G1	nursery	sex	activities
CatBoost	G2	G1	absences	age	Dalc
TabNet	G2	Medu	traveltime	failures	age
TCBoost	G2	TabNet	G1	absences	age
TLBoost	TabNet	G2	G1	absences	failures

표 5는 다양한 모델이 언어 성취도(G3)를 예측할 때

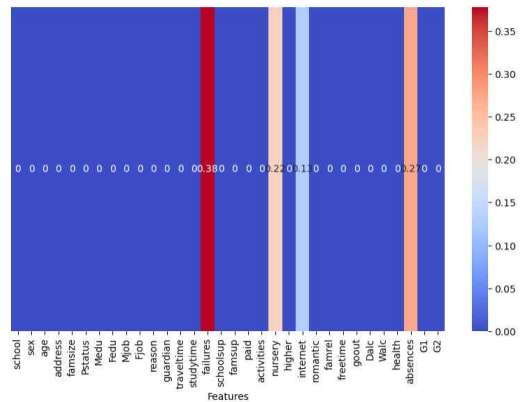


Figure 5. Attention distribution of features for Language G3

그림 5는 TabNet 모델의 Attention distribution을 시각화한 것으로, 무작위 추출로 결정 단계 35에서 각 특징의 중요도를 나타낸다. 그래프를 통하여 TLBoost가 TabNet 모델을 사용할 때 각 결정 단계에서 어떤 특징을 얼마나 중요하게 사용하는지를 시각적으로 보여준다. 이는 모델의 해석 가능성을 높여주며, 특정 특징이 예측에 어떻게 기여하는지 이해하는 데 도움을 주며 해석성을 향상시킨다. 또한 중요도가 높은 특징을 파악하고 추가적인 데이터 수집을 통하여 모델의 예측 성능을 향상시킬 수 있다. 또한 교수자들이 성취도 예측 모델을 사용할 때 어떤 특징이 중요한지 이해함으로써 효과적인 맞춤형 학습 방법을 찾을 수 있다.

## 5. 결론 및 제언

본 연구는 TLBoost 모델을 사용하여 학습자의 성취도를 예측하는 방법을 제안하고 성능을 검증하였다. TLBoost는 TabNet과 LightGBM을 결합한 앙상블 모델로 두 알고리즘의 장점을 결합하여 예측 정확도와 해석 가능성을 동시에 향상시켰다. 구체적으로 TLBoost 모델은 수학 성취도 예측에서 RMSE 1.7040,  $R^2$  Score 0.8583, 언어 성취도 예측에서는 RMSE 1.0694,  $R^2$  Score 0.8827로 나타났다. 단일 모델인 LightGBM, CatBoost, TabNet 모델 그리고 앙상블 모델인 TCBoost 모델보다 높은 예측 성능을 기록하며, TLBoost 모델의 효용성을 확인하였다. 예측 성능 향상에 특징 중요도가 사용되었다. TabNet 예측 값이 LightGBM 모델에서 높은 중요도를 차지하여 예측 성능 향상에 기여하였음을 확인하였으며, 이는 TabNet이 학습한 복잡한 패턴이 LightGBM 모델의 예측 정확도를 높이는 데 중요한 역할을 함을 시사한다. TabNet 예측 값 외에 수학 과목에서는 G2, G1, absences, age 등의 특징이 중요한 역할을 하였고, 언어 과목에서는 G2, G1, absences, Dalc 등이 중요하게 사용되었다. 해석 가능성 향상을 위하여 TLBoost 모델은 TabNet의 어텐션 메커니즘을 활용하여 예측 과정에서 중요한 특징들을 시각적으로 강조하였고, LightGBM의 특징 중요도 분석을 통해 예측에 기여한 요소들을 확인하였다. 이를 통해 교수자들이 학습자의 성취도에 영향을 미치는 주요 요소들을 이해하고, 맞춤형 학습 지원을 제공할 수 있다.

본 연구에서 활용된 포르투갈 학습자 데이터는 국내 교육 현장에 직접적으로 적용되기에는 한계가 있다. 그러나 국내에 충분한 교육 데이터가 부재한 상황에서 해외의 실제 데이터를 기반으로 한 실험은 필요한 데이터 수집 방법론을 개발하는데 기초 자료로 활용될 수 있으며, 도입을 압둔 AI 기반 성취도 시스템의 현장 적용을 위한 방안을 모색할 수 있다. 결론적으로, 본 연구는 성취도 예측 모델의 해석성을 높여 기술의 발전과 교육 현장 적용 사이의 간극을 좁히고, 교육 현장에서 실질적으로 활용 가능한 예측 모델을 제시함으로써 교육 정책의 실효성을 높이고, 개인별 맞춤형 학습 지원을 제공하는 데 기여할 수 있다.

### 참고문헌

- [1] Sin, J. (2022, September 6). Customized academic achievement assessment to be implemented from the 13th... Expansion from 2024 to grades 3 to 11 (맞춤형 학업성취도 평가 13일부터 시행...2024년부터 초3~고2 확대). Seoul Economic Daily. <https://www.sedaily.com/NewsView/26AYVJNKJA>
- [2] Ministry of Education. (2022). Announcement of 2021 national level academic achievement assessment results and response strategy (2021 국가수준 학업성취도 평가 결과 및 대응 전략 발표). Ministry of Education. <https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=294&lev=0&statusYN=W&s=moe&m=020402&opType=N&boardSeq=91788>
- [3] Park, Y. (2022, July 13). Introduction plan for computer-based academic achievement assessment. Yonhap News Agency. <https://www.yna.co.kr/view/GYH20220613001300044>
- [4] Shen, S., Liu, Q., Huang, Z., Zheng, Y., Yin, M., Wang, M., & Chen, E. (2024). A survey of knowledge tracing: Models, variants, and applications. IEEE Transactions on Learning Technologies. <https://doi.org/10.1109/TLT.2024.3383325>
- [5] Korea Educational Development Institute (KEDI). (2019). Study on personalized education support measures based on online learning analytics (온라인 학습분석 기반 맞춤형 교육지원 방안 연구). Available online at <https://www.kedi.re.kr/khome/main/research/selectPubForm.do?plNum0=12779>. Accessed 25 Apr 2024.
- [6] Park, S. H., & Sang, K. A. (2011). Characteristics of and factors affecting on students' attitude toward mathematics. School Mathematics, 13(4), 697-716.
- [7] Suhaini, M., Ahmad, A., & Harith, S. H. (2020). Factors influencing student achievement: a systematic review. International Journal of Psychosocial Rehabilitation, 24(5), 550-560.
- [8] Lee, J., & Shute, V. J. (2010). Personal and social-contextual factors in K-12 academic performance: An integrative perspective on student learning. Educational psychologist, 45(3), 185-202. DOI: 10.1080/00461520.2010.493471
- [9] Rodriguez, P., Bautista, M. A., Gonzalez, J., & Escalera, S. (2018). Beyond one-hot encoding: Lower dimensional target embedding. Image and Vision Computing, 75, 21-31. DOI: 10.1016/j.imavis.2018.04.004
- [10] Abdelrahman, G., Wang, Q., & Nunes, B. (2023). Knowledge tracing: A survey. ACM Computing Surveys, 55(11), 1-37. DOI: 10.1145/3569576
- [11] Yang, J., Moon, G., & Park, S. (2022). A study on extracting and predicting characteristics of learning difficulties using artificial intelligence (인공지능을 이용한 학습부진 특성 추출 및 예측 모델 연구). Journal of the Korea Institute of Information &

- Communication Engineering, 26(4), 510-518.
- [ 12 ] Villagr -Armedo, C. J., Gallego-Dur n, F. J., Llorens-Largo, F., Compa n-Rosique, P., Satorre-Cuerda, R., & Molina-Carmona, R. (2017). Improving the expressiveness of black-box models for predicting student performance. *Computers in Human Behavior*, 72, 621-631.
- [ 13 ] Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019, March). Student performance prediction and classification using machine learning algorithms. In *Proceedings of the 2019 8th international conference on educational and information technology* (pp. 7-11).
- [ 14 ] Shin, J., & Choi, J. (2019). Developing the prediction model of at-risk freshmen students and analyzing characteristics of cluster based on learning analytics. *Journal of Educational Technology*, 35(Special), 425-454. <https://doi.org/10.17232/KSET.35.2.425>
- [ 15 ] Wang, X. S., Perry, L. B., Malpique, A., & Ide, T. (2023). Factors predicting mathematics achievement in PISA: a systematic review. *Large-scale Assessments in Education*, 11(1), 24. DOI: 10.1186/s40536-023-00174-8
- [ 16 ] Zhang, L., Xiong, X., Zhao, S., Botelho, A., & Heffernan, N. T. (2017, April). Incorporating rich features into deep knowledge tracing. In *Proceedings of the fourth (2017) ACM conference on learning@ scale* (pp. 169-172). DOI: 10.1145/3051457.3053976
- [ 17 ] Liu, Q., Shen, S., Huang, Z., Chen, E., & Zheng, Y. (2021). A survey of knowledge tracing. *arXiv preprint arXiv:2105.15106*.
- [ 18 ] Sagi, O., & Rokach, L. (2018). Ensemble learning: A survey. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 8(4), e1249.
- [ 19 ] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- [ 20 ] Sagi, O., & Rokach, L. (2021). Approximating XGBoost with an interpretable decision tree. *Information sciences*, 572, 522-542.
- [ 21 ] Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M., & Suganthan, P. N. (2022). Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115, 105151. DOI: 10.1016/j.engappai.2022.105151
- [ 22 ] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- [ 23 ] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems*, 31.
- [ 24 ] Arik, S.  ., & Pfister, T. (2021, May). Tabnet: Attentive interpretable tabular learning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 35, No. 8, pp. 6679-6687).
- [ 25 ] Xu, T., Yu, Y., Yan, J., & Xu, H. (2020). Long-Term Rainfall Forecast Model Based on The TabNet and LightGbm Algorithm.
- [ 26 ] Salunke, S. V., & Ouda, A. (2023, July). Ensemble Learning to Enhance Continuous User Authentication For Real World Environments. In *2023 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)* (pp. 102-108). IEEE.
- [ 27 ] Botchkarev, A. (2018). Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. *arXiv preprint arXiv:1809.03006*.
- [ 28 ] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ computer science*, 7, e623.



박 선 우

2018년 서울시립대학교  
경영학부 (경영학학사)  
2021년 연세대학교  
교육행정학과 (교육학석사)

2024년 고려대학교 일반대학원 컴퓨터학과 박사 수료  
관심분야: AIED, CS/AI 교육, XAI  
E-Mail: sunwoosan@korea.ac.kr



김 현 철

1988년 고려대학교 전산과학과 (학사)  
1990년 Univ of Missouri-Rolla  
(전산학석사)  
1998년 Univ of Florida (전산정보학 박사)

1999년~현재 고려대학교 컴퓨터학과 교수  
관심분야: SW/AI 교육, 기계학습  
E-Mail: harrykim@korea.ac.kr