

Exploring predictors for academic achievement using machine learning on LMS log data from the online course

Ji Won You

Professor, Department of Education, Graduate school of Education, Gachon University

This study aimed to identify online learning behavior patterns and explore key variables for predicting academic achievement by analyzing LMS log data. Data were collected from 48 students who enrolled in a 15-week online teacher education course at A Graduate School of Education. From 18,270 rows of log data, a total of 123 features—capturing both weekly learning activities and overall course trends—were used to build the prediction model. Academic achievement prediction was conducted using the machine learning technique elastic net, and the findings were twofold. First, an analysis of weekly learning patterns revealed that behaviors differed according to the type of learning content (e.g., video, link, file). Furthermore, contrary to concerns about cramming or procrastination, most learners did not engage in cramming or procrastination, with approximately 15% of learners exhibiting cramming behavior. Second, the predictive modeling yielded 19 variables that were reliably selected for the final model. Key predictors of higher academic achievement included: a higher frequency of content viewing and material downloads in specific weeks; an earlier start to weekly learning; a greater number of completed assignments; more frequent content access during the final-exam preparation period; and enrollment in an A-F grading system. Based on these results, the study discusses the significance of the selected variables within the course context, as well as the advantages and implications of utilizing machine learning techniques in learning analytics.

Key words : online learning, LMS log data, machine learning, learning analytics, elastic net

<요 약>

온라인 강좌의 LMS 로그 데이터를 활용한 기계학습 기반 학업성취도 예측 변수 탐색*

유지원†

가천대학교 교육대학원 교육학과 교수

본 연구는 학습관리시스템(Learning Management System, LMS)을 기반으로 운영된 온라인 강좌에서 수강생들의 로그 데이터를 수집하여 온라인 학습행동 양상을 파악하고, 학업성취도 예측에 중요 변수를 탐색하고자 수행되었다. 연구를 위한 자료 수집은 수도권 소재 A교육대학원의 15주간 운영된 교과과목을 수강한 48명 학생들의 LMS 로그 데이터를 수집하였다. 18,270행의 LMS 로그 데이터로부터 주차별 학습행동을 포착할 수 있는 변수와 전체 교과목 전체에서의 경향을 파악할 수 있는 지표를 추가로 생성하여, 총 123개의 설명변수를 활용하여 학업성취도 예측 모델링을 수행하였다. 학업성취도 예측은 기계학습 기법인 elastic net을 활용하였으며 주요 연구결과는 다음과 같다. 첫째, 학생들의 주차별 학습 양상을 살펴본 결과, 학습콘텐츠 유형(동영상, 링크, 파일 등)에 따라 학습행동 횟수에 차이를 보였다. 또한 온라인 수업에서 우려하는 마감일에 임박한 학습양상을 살펴본 결과, 대부분의 학습자는 학습지연 또는 벼락치기 학습을 하지 않는 것으로 파악되었고 약 15%의 학습자가 벼락치기 학습행동을 보였다. 둘째, 예측 모델링 결과, 최종 모형에 안정적으로 선택된 19개의 변수를 도출하였다. 학업성취도 예측에 중요 변수는 특정 주차에서의 학습콘텐츠 열람 횟수가 높을수록, 자료를 다운로드 할수록, 주차의 학습을 일찍 시작할수록, 과제 완료 횟수가 높을수록, 기말시험 준비기간에 학습콘텐츠 열람 횟수가 많을수록, 학점유형을 A-F로 신청한 학습자 경우에 학업성취도가 높게 예측되었다. 이러한 결과를 통해 해당 교과 맥락에서 선정 변수의 의미를 해석하고 논의하였으며, 학습분석에서 기계학습 기법 활용의 장점과 시사점을 논의하였다.

주제어 : 온라인 학습, LMS 로그 데이터, 기계학습, 학습분석, elastic net

* 본 논문은 2025년도 가천대학교 교내연구비 지원에 의한 결과임(GCU-202504220001).

† 교신저자: 유지원, 가천대학교, 교수(E-mail: uimagine@gachon.ac.kr)

I. 서론

디지털 대전환 시대를 맞이하여 교육 분야도 빠르게 디지털화가 진행되고 있다. 교육현장에서는 수업 설계, 학습자원 개발, 수업 활동과 평가, 학습지원 및 학습관리에 이르기까지 디지털 교육 플랫폼과 다양한 에듀테크 도구가 보편적으로 활용되는 추세이다. 대학의 경우, 대부분 학습 관리시스템(learning management system, 이하 LMS)을 운영하고 있으며(박지우, 임규연, 2023; Arizmendi et al., 2023), LMS를 활용하는 수업 유형도 블렌디드 수업, 플립러닝, MOOC, 비실시간 온라인 교과 등 다양하게 나타난다. 학습자는 LMS를 통해 수업 자료 접근, 학습활동 참여, 과제 제출, 구성원과의 상호작용 등 다양한 학습행동을 수행하고, 그 과정에서 시간 정보(time-stamp)가 포함된 로그 데이터가 LMS에 자동으로 기록된다. 이러한 로그 데이터는 학습자의 학습행동을 객관적으로 파악하고 분석할 수 있는 토대를 제공한다는 점에서 잠재적인 활용 가치가 높다.

LMS 로그 데이터를 포함한 학습데이터에 대한 관심이 높아지면서, 데이터를 교수학습 개선의 핵심 자원으로 활용하는 학습분석(learning analytics) 연구도 활성화되고 있다(박지우, 임규연, 2023; Chen & Cui, 2020). 최신 학습분석의 정의는 “학습자와 그들의 학습에 대한 데이터에 대해 수집, 분석, 해석, 소통하여 학습과 교육을 향상시킬 수 있는 이론적으로 타당하고 실행가능한 통찰을 제공하는 것”(SoLAR, 2025, p. 2)이다. Long과 Siemens(2011)는 특히 대학이 다양한 데이터가 축적되어 있음에도 불구하고 비효율적으로 데이터를 관리하거나 분석 지연으로 인해 효과적인 개입 기회를 놓치고 있다고 지적하고, LMS에 기록되는 학습행동 데이터로부터 학습자의 학습과정에 대한 이해와 지원 방안을 모색할 수 있다고 역설하였다(Long & Siemens, 2011). LMS 로그 데이터를 활용한 선행연구에서는 학생들의 학습양상 분석(홍유정 외, 2025), 학습성취 예측(유진은, 2020; Arizmendi et al., 2023; Conijn et al., 2017; You, 2016), 중도탈락 예측 모델 구축(Brooks & Thompson, 2017), 학업성취에 대한 조기 예측(early prediction) 모델 개발(Arnold & Pistilli, 2012; Howard et al., 2018), 대시보드 설계 및 개발(박연정, 조일현, 2014; 신중호 외, 2018) 등이 주를 이룬다. 이와 같이 LMS는 학습분석학의 중요한 도구로 인식되며, 비간섭적으로 수집할 수 있는 로그 데이터를 통해 학습자의 온라인 학습행동을 관찰하고 의미 있는 패턴을 탐색하여 통찰을 얻을 수 있는 정보원의 역할을 한다(서진선, 이재경, 2020).

한편 온라인 학습환경은 오프라인 수업에 비해 학습자의 자율성이 높은 만큼 자기조절학습이 필수 요인으로 강조되며, 학습분석 연구에서 LMS 데이터와 자기조절학습 이론을 연계하는 것이 중요한 과제이다(박지우, 임규연, 2023; 이성혜 외, 2021; Jovanović et al., 2021; Winne, 2022; You, 2016). 자기조절학습이란 학습자가 자신의 학습목표를 성취하기 위해 인지, 동기, 행동, 환경을 능동적으로 조절하는 과정을 뜻한다(Pintrich, 2004). 특히 온라인 학습환경에서 자기조절학습 실패는 전통적인 교실 학습상황보다 피해가 더 크다고 논의된다(Dabbagh & Kitsantas, 2004; Warnock et al., 2012). 자기조절학습에 실패한 학습자는 마감일에 압박해서 벼락치기식 학습행동

을 하거나(Nieberding & Heckler, 2021; Theobald et al., 2021), 학습 시작을 미루는 학습지연 행동을 보이며, 학습지연 행동은 궁극적으로 학습성취에 부정적인 영향을 미치게 된다(You, 2016). 이러한 맥락에서 연구자들은 자기조절학습 이론을 토대로 LMS 로그 데이터로부터 자기조절학습 행동을 대리하는 변수를 활용하여 연구를 수행하고 있다(박지우, 임규연, 2023; 조일현, 김윤미, 2013; Sonnleitner et al., 2025; You, 2016).

LMS 로그 데이터 연구에 활용되는 지표를 살펴보면, 학습시스템 로그인 횟수, 학습콘텐츠 클릭 횟수, 동영상 학습시간, 토론글 열람 또는 작성 횟수, 자료 다운로드 횟수, 퀴즈 시도 횟수 등 학습콘텐츠와의 상호작용 지표가 빈번히 활용된다(Bernacki et al., 2020; Jovanović et al., 2021; Romero et al., 2013; Sonnleitner et al., 2025). 이 지표들은 로그 데이터에 기록된 원자료로부터 쉽게 추출할 수 있으며, 학습자의 학습을 위한 노력 행동을 양적으로 가시화한 지표이다. 학습자의 시간관리 차원의 자기조절학습을 가늠할 수 있는 지표를 활용하는 연구들도 등장하였는데, 예로 조일현과 김윤미(2013)는 학습 접속시점 간격의 규칙성 지표를 만들고, 이 지표가 학습시간과 학습 횟수보다 학습성취 예측에 효과적이라고 논의하였다. 김향미와 전광호(2025)는 다음 학습 접속까지 걸리는 시간 간격이 길어질수록 학습성취를 낮게 예측한다고 밝혔다. 또한 박지우와 임규연(2023)은 학습 횟수, 학습 시점 간격, 학습시점 규칙성 변인을 활용하여 자기조절학습 예측을 분석한 결과, 인지조절 영역과 맥락조절 영역에는 학습 시점의 규칙성과 학습 시점 간격 변수가 중요 변수로 나타났고, 행동조절 영역에는 규칙적인 학습 시점 변수가 선택되었다. 저자들은 학습 횟수 변수도 전반적으로 중요하게 나타났으나 단순히 양적인 클릭 횟수보다는 적절한 시간 관리를 가늠할 수 있는 변수가 상대적으로 중요하게 나타났다고 논의하였다. 이상의 논의를 종합하면, LMS 로그 데이터로부터 도출한 다양한 행동 변수는 학습자의 온라인 학습행동을 이해하는데 유용한 정보를 제공한다. 또한 온라인 학습에서는 학습일정을 관리하면서 학습콘텐츠에 접속하여 학습을 완료하고 과제를 제출하는 등의 노력 행동과 학기 전반에 걸쳐 꾸준히 규칙적인 학습을 지속하는 시간관리 행동이 무엇보다 중요하다. 이는 학습목표 성취를 위한 노력 및 시간을 계획, 모니터링, 조절하는 행동이 온라인 학습에 핵심 요소임을 시사한다. 따라서 LMS 로그 데이터로부터 자기조절학습에 관련된 다양한 지표를 활용하여 학습성취 예측에 유용한 변수를 탐색하는 것은 학습분석을 기반의 온라인 수업 개선과 학습촉진 방안 모색에 기여할 수 있을 것이다.

한편, 교육에서 빅데이터 생성과 학습분석의 활성화는 데이터 분석 방법에도 변화를 가져왔다. 전통적인 통계분석은 예측보다 설명에 초점을 두고 있으며(Bzdok et al., 2018), 수백 개 이상의 변수와 다양한 정형 또는 비정형 데이터 유형을 분석하는데 한계가 있다. 예로 전통적인 회귀분석의 경우 하나의 모형에 수많은 변수를 투입할 수 없어서 변수 개수를 대폭 줄이기 위해 집계 변수(aggregate variable)를 만들어 투입한다. 또한 도출한 모형이 현재 자료에는 높은 설명력을 보이나 새로운 데이터에 대해 예측력이 낮게 나타나는 과적합(overfitting) 문제가 단점으로 지

적된다(유진은, 2020). 그러나 기계학습은 사례 수보다 입력 변수가 더 많은 ‘wide data’를 다룰 때에도 유용하며(Bzdok et al., 2018), 원자료(raw data)의 특징을 그대로 유지하면서 분석을 수행할 수 있다. 기계학습 분석 방법에는 랜덤 포레스트(random forest), 의사결정나무(decision tree), 서포트 벡터머신(Support Vector Machine), 그래디언트 부스팅(Gradient Boosting), k-평균 군집분석(k-means clustering), 신경망 분석 등 다양하며 분석의 목적과 데이터 종류와 특징, 예측 변수의 특성, 모형 성능, 예측 결과의 해석 가능성 등에 따라 적합한 알고리즘을 선택한다.

LMS 데이터를 포함한 학습자 데이터를 기계학습으로 분석한 연구를 살펴보면, 학습자의 성취 예측모형을 구축하고 성능을 평가하는 예측 모델링이 주를 이룬다(Howard et al., 2018; Sonnleitner et al., 2025; Tiukhova et al., 2024; Tomasevic et al., 2020). Howard와 동료들(2018), Chen과 Cui(2020) 등은 LMS 데이터로 여러 기계학습 기법을 적용하여 학업성취를 조기 예측하는 모형을 구축하고 적용한 알고리즘 간 성능을 비교하고 논의하였다. 하지만 LMS 데이터는 교과목, 수강환경(온라인, 블렌디드, 오프라인 등), 수업에서 정책(출석 규칙, 참여도 규칙 등), 평가 방식, 수강학생의 특성 변화 등 연구 맥락마다 달라지는 특성이 있다(Sonnleitner et al., 2025; Tiukhova et al., 2024). 이에 학자들은 비선형 모델(XGBoost, 딥러닝 등)은 대체로 높은 예측력을 보이지만 맥락 변화에 민감하여 새로운 교과목이나 학기에 적용할 때 기존 모형의 성능이 저하될 수 있고 블랙박스 기반의 알고리즘으로 결과 해석이 어렵다고 지적한다(Sonnleitner et al., 2025). 이러한 배경에서 연구자들은 간명한 선형모형 사용을 추천하며, 분석 결과에 대한 해석이 용이해야 교수자와 학생 간의 상호작용 개선에 기여할 수 있고, 결과에 대한 신뢰도를 향상시킬 수 있다고 주장한다(Bernacki et al., 2020; Khosravi et al., 2022).

해외에서 학습분석 연구에 기계학습 기법이 활발히 적용되는 것에 비해(Arizmendi et al., 2023; Chen & Cui, 2020; Howard et al., 2018; Sonnleitner et al., 2025; Tiukhova et al., 2024; Tomasevic et al., 2020), 국내에서는 LMS 데이터를 기계학습으로 분석한 연구가 적다. 최근 10년간 국내 학습 분석 연구를 분석한 김혜빈과 동료들(2023)에서는 전통적인 통계분석을 활용한 연구가 70%, 질적 분석이 약 11%에 해당하였고, 지도 및 비지도 학습의 기계학습 기법을 활용한 연구는 16% 정도에 그쳤다고 보고하였다. 이는 학습데이터의 분석 방법 측면에서 개선할 필요가 있음을 시사하는 것이다.

이에 본 연구는 LMS에서 수집한 데이터를 기계학습 방법으로 분석하고자 한다. 분석방법으로 별점회귀모형인 elastic net을 선정하였다. 이 기법은 많은 설명변수가 투입될 때 불필요한 변수를 걸러내는 lasso regression의 장점과 유사한 변수로 인한 다중공선성 문제를 완화시키는 ridge regression의 장점을 모두 지닌 분석방법이다(Zou & Hastie, 2005). Elastic net은 일반화 성능이 안정적이고, 선택된 변수에 대한 해석이 직관적이어서 학업성취 예측에 추천된다(Arizmendi et al., 2023). 유진은(2020)은 플립러닝 강좌의 LMS에서 동영상 시청 빈도 데이터와 학생 배경변수를 합쳐서 총 100개의 설명변수로 elastic net을 적용하여 기말고사 점수를 예측하였다. 저자는 수업에

서 제공한 47개에 대한 동영상 시청 빈도 변수만으로 학업성취도 예측이 가능하다고 밝히고, 특히 2-3주차 동영상 시청 행동이 중요 변수로 나타남에 따라 학습에 어려움을 겪는 학습자를 조기 선별하여 교수자의 효과적인 개입 가능성을 논의하였다. Arizmendi와 동료들(2023)은 플립러닝 교과에서 다양한 학습자 관련 데이터로 학기말 학점을 elastic net을 활용한 로지스틱 회귀로 예측 모형을 구축하였다. 연구자들은 예측 모델에 58개의 특징변수가 선택되었고 주차별로 생성한 행동지표가 예측력에 기여했다고 밝히고, 교과목 전체에 걸친 집계 변수보다 주차별 세분화된 데이터로 분석했을 때 학습자가 언제, 어떤 행동을 얼마나 하는지 시점을 파악하여 교수자 개입과 학습촉진에 기여할 수 있다고 논의하였다.

본 연구의 필요성과 의의는 다음과 같다. 첫째, 국내에서 LMS 로그 데이터를 기계학습으로 분석한 사례가 적고, 특히 표본이 작은 수업에 적용한 사례는 드물다. 본 연구는 수강생 규모가 작은 온라인 교과목에서 기계학습을 적용한 학습분석 사례를 제공함으로써, 향후 다수 온라인 교과목에서 학습자 행동 이해와 교수학습 개선을 위한 학습분석적 접근 시도가 확대되길 기대한다. 둘째, 기존 연구에서는 LMS 로그 데이터로부터 학습빈도, 학습시간, 학습시점 규칙성 지표가 주로 사용되고 있으나, 학습성취 예측에 새로운 변수의 발굴과 검증이 추천된다. 기계학습 분석을 활용하면, 투입변수 개수의 제약 없이 동시에 많은 변수를 활용할 수 있고, 주차별로 세분화된 변수를 사용할 수 있다. 이는 기존 연구에서 사용하지 않았던 새로운 행동지표에 대한 검증과 교과목 운영 시점에 따른 학습자 행동을 구체적으로 파악하고 개선 방안을 마련하는데 기여할 수 있다. 요약하면, 본 연구의 목적은 대학에서 LMS 기반으로 운영된 온라인 교과목의 로그 데이터로부터 다양한 온라인 학습행동 지표를 활용하여 기계학습 방법으로 학업성취 예측에 중요 변수를 탐색하는 것이다. 주요 연구문제는 다음과 같다.

첫째, 대학에서 온라인 수업으로 운영된 교과목의 LMS 로그 데이터로부터 파악한 학습자의 온라인 학습 양상은 어떠한가? 둘째, 대학에서 운영된 온라인 교과목의 LMS 로그 데이터로부터 학업성취 예측에 중요한 변수는 무엇인가?

II. 연구 방법

1. 연구 대상 및 맥락

가. 연구 대상

본 연구는 수도권 소재 A대학교의 교육대학원에서 2025학년도 1학기에 온라인으로 운영된 교직교과목 ‘디지털 교육’ 수강생의 자료를 수집하였다. ‘디지털 교육’은 필수 교직소양 교과목으로 전공에 상관없이 예비교사가 이수하는 교과목이다. 총 48명이 수강하였으며, 학생들의 전공은

음악교육 10명, 체육교육 14명, 영양교육 7명, 유아교육 12명, 교육심리 및 상담 전공이 5명으로 나타났다(<표 1> 참조).

<표 1> 연구참여자 배경정보

	구분	빈도	%
입학년도	2024학년도 입학생	23	47.9
	2025학년도 입학생	25	52.1
전공	음악교육	10	20.8
	체육교육	14	29.2
	영양교육	7	14.6
	유아교육	12	25.0
	교육심리 및 상담	5	10.4
	총 계	48	100

나. 교과목 구성 및 운영

해당 교과목의 구성은 다음과 같다. 교과목은 15주로 구성되었으며, 1주차부터 14주차까지의 모든 수업은 LMS상에서 강의, 활동, 과제로 운영되었고, 15주차에는 오프라인에서 기말고사를 실시하였다. 1주차는 교수자의 오리엔테이션 내용과 함께 교과목에 대한 개요를 강의한 영상, 본 교과목에서 수행해야 하는 과제를 설명하는 자료, Padlet(링크 제공)에서 학습자 소개하는 공유 활동으로 이루어졌다. 2주차부터는 연구자가 만든 강의동영상, 강의영상에 활용한 강의노트, 학습자원 파일, 외부 학습 링크, 의견 공유 활동, 과제 제출 등으로 구성되었다. 각 주차별 학습에 대한 세부 구성은 <표 2>와 같다.

수업 운영 규칙은 매주 학습할 영상과 자료들을 월요일 0시를 기준으로 공개하고, 학습자는 매주 월요일 0시부터 일요일 11:59:59pm까지 강의영상 시청을 완료해야 출석으로 인정된다. 학습자는 동영상 시청률에 따라 출석 상황을 학습이력 확인 메뉴에서 상시 확인할 수 있다. 해당 주차 이후에도 강의영상 시청은 자유롭게 할 수 있지만, 지정된 학습기간 이후 시청은 출석에 반영되지는 않는다.

2. 분석 자료

가. 로그 데이터 수집

학생들의 LMS에 기록된 로그 데이터 수집은 web scrapping 도구인 Web Scraper for chrome extension(<https://webscraper.io/>)을 활용하여 교과목 내 학생별로 제공되는 로그 데이터를 스크래핑하

<표 2> 주차별 학습 내용 및 구성

주차	학습주제	강의 동영상	강의 노트	파일	외부 콘텐츠 링크	공유 활동	과제함
1주차	교과목 오리엔테이션	1	1	1	-	1	-
2주차	디지털교육 플랫폼과 콘텐츠	1	1	1	-	1	-
3주차	에듀테크와 적응적 학습원리	3	1		-	-	-
4주차	AI literacy: 컴퓨팅사고력과 기계학습	3	3	-	2	-	-
5주차	AI literacy: 기계학습 과정 및 알고리즘	2	1	-	-	-	-
6주차	AI literacy: 딥러닝 이해	2	1	-	1	-	-
7주차	AI literacy: 개념 정리 및 확장 학습활동	-	-	1	-	-	1
8주차	AI literacy: 생성형AI 작동 원리와 모델	2	1	-	-	-	-
9주차	AI literacy: 생성형 AI 활용사례와 이슈	1	1	1	-	-	-
10주차	과제 1 수행 기간	-	-	-	-	-	1
11주차	AI literacy: 인공지능 윤리와 이슈	3	2	-	-	-	-
12주차	AI literacy: 최신 생성형 AI도구와 실습	1	1		-	-	1
13주차	AI · 디지털 교육 이슈 논의	-	-	-	1	-	-
14주차	과제 2 수행 기간	-	-	-	-	-	1
	합계	19	13	4	4	2	4

였다. 수집 기간은 2025년 3월 2일부터 기말고사 실시일인 2025년 6월 13일까지로, 총 18,270개의 로그가 수집되었다. 로그 데이터에 포함된 항목은 학생이름, 접속일시, 이벤트문맥, 구성요소, 발생이벤트이며, 이벤트문맥은 특정 영상, 파일, 링크, 과제함 등에 대한 정보를 저장하고, 발생 이벤트에는 학습자료/활동 클릭, 이수현황보고서 페이지 열람(클릭), 파일 다운로드, 게시판 공지글 클릭, 과제물 등 이벤트 정보를 제공한다. 개인정보 보호를 위하여 학생 이름은 삭제하고 일련번호를 부여하여 비식별화(de-identification)하였다.

나. 데이터 전처리 및 설명변수 선정

데이터 분석을 위한 정리는 크게 두 단계로 이루어졌다. 먼저 로그 데이터로부터 분석에 포함할 변수를 선별하여 코딩하였고, 두 번째 단계에서는 수집한 데이터의 변환, 결합, 추가 등의 전처리 과정을 거쳐 새로운 변수를 생성하였다. 로그 데이터에서 직접 수집한 변수는 <표 3>에, 새로 생성한 변수는 <표 4>에 제시하였으며, 변수에 대한 상세한 설명은 다음과 같다.

<표 3> LMS 로그 데이터로부터 1차 수집한 데이터 내용과 변수

구분	변수	변수 갯수	해석
전체	게시글 클릭 횟수(Notice_View), Q&A 글쓰기 횟수(QnA_Write), 학습이력현황 페이지 열람(클릭) 횟수(StatusCheck)	횟수: 3	A
주차별 학습	동영상 강의동영상 열람(클릭) 횟수(Wx_VODi_View), 강의동영상 첫 열람(클릭) 날짜(Wx_VODi_AccessDate)	클릭 횟수: 19 첫 클릭 날짜: 19	A, B
	노트 강의노트 열람(클릭) 횟수(Wx_Notei_View), 강의노트 다운로드 횟수(Wx_Notei_DownCount)	클릭 횟수: 13 다운로드: 13	A
	학습자료 파일 학습자료파일 열람(클릭) 횟수(Wx_Filei_View), 학습자료파일 첫 열람(클릭) 날짜(Wx_Filei_AccessDate), 학습자료파일 다운로드 횟수(Wx_Filei_DownCount)	클릭 횟수: 4 첫 클릭 날짜: 4 다운로드: 4	A, B, A
	링크 학습 링크 열람(클릭) 횟수(Wx_Linkj_View), 학습 링크 첫 열람(클릭) 날짜(Wx_Linkj_AccessDate)	클릭 횟수: 4 첫 열람날짜: 4	A, B
	학습활동 학습활동 링크 열람(클릭) 횟수(Wx_Activity_View), 학습활동 링크 첫 열람(클릭) 날짜(Wx_Activity_AccessDate), 학습활동 완료 여부(Wx_Activity_Complete)	클릭 횟수: 2 첫 클릭 날짜: 2 완료여부: 2	A, B, A
	과제 과제폴더 열람(클릭) 횟수(Wx_AssignmentFolder_View), 과제 첫 번째 제출 날짜(Wx_AssignSubmit_Date), 과제 제출 횟수(Wx_Assign_SubmitFreq)	클릭 횟수: 4 첫 제출날짜: 4 제출 횟수: 4	A, B, A

x: 1-14까지 주차; i: 순차 번호(예: W4_VOD2_View, 4주차 2번째 강의영상 열람횟수); A: 노력행동(성실도),
B: 시간관리

본 연구에서는 다양한 온라인 학습행동 지표를 활용하고자 하였으며, 변수 선정은 학습자의 성실한 학습행동을 보여주는 변수와 시간관리 행동을 보여주는 변수에 초점을 두었다. 성실한 학습 행동에는 학습자료에 빈번히 접속하고, 주어진 학습활동 및 과제를 완료하며, 중요 정보와 학습자료를 다운로드하여 학습을 준비하는 행동 등이 해당한다. 시간관리 행동 관련 변수는 주차별 업로드되는 학습콘텐츠를 미루지 않고 일찍 학습을 시작하는 행동, 벼락치기 학습행동, 규칙적인 학습패턴 등이 해당한다.

<표 3>을 살펴보면, 주차별 학습 유형에는 학습 콘텐츠 유형에 따라 강의동영상, 강의노트, 학습자료 파일, 외부 콘텐츠 링크로 세분화하였고, 이 외에 Padlet 링크에 접속하여 의견을 공유하는 학습활동 유형과 활동지 및 보고서를 과제함으로 제출하는 과제 유형으로 구분하였다. 강의동영상, 파일, 링크는 열람(클릭) 횟수와 첫 접속날짜를, 강의노트는 열람 횟수와 다운로드 횟수를 변수로 코딩하였다. 강의노트의 경우 동영상에 활용한 강의노트를 제공한 PDF 문서이므로 첫 접속날짜는 코딩에서 제외하였다. 학습활동은 특정 주제에 대한 의견을 Padlet에 게시하는 공

유 활동으로 총 3개의 변수(학습활동 링크 접속 횟수, 활동 링크 첫 접속날짜, 활동 참여 여부)로 코딩하였다. 과제는 학습 결과를 과제함으로 제출하는 경우로, 학습주제를 심화하는 활동지 제출 과제 2회, 보고서 작성 과제 2회 등 총 4회에 해당한다. 과제 관련 변수로는 과제함 열람 횟수, 과제 첫 제출 날짜, 과제 제출 횟수 등 총 3개의 변수로 코딩하였다. 주차별 학습 외에 수집한 데이터와 변수는 수업 공지게시판 게시글 열람 횟수(Notice_View), 질의응답게시판 글쓰기(QnA_Write)와 학습이력 현황 확인 횟수(StatusCheck)이다.

다음은 1차로 정리한 데이터에서 값의 변환, 결함 등의 feature engineering으로 새로운 설명변수를 생성하였다(<표 4> 참조). 로그의 원자료 그대로 의미 있는 학습행동을 포착하기 어렵기 때문에, 원자료를 학습행동 간의 시간 간격, 규칙적 학습을 보여주는 지표로 변환함으로써 더욱 교육적 의미를 부여할 수 있다(Arizmendi et al., 2023). 본 연구에서는 세 종류의 시간관리 행동 변수를 생성하였다. 첫째, 매 주차 학습 또는 과제를 마감일 대비 얼마나 일찍 시작하는지를 보여주는 변수로, 첫 학습날짜 또는 첫 과제제출 날짜와 마감일 간의 차이를 -4부터 6에 이르는 값으로 변환하였다. 마감일 당일 접속 또는 제출은 0, 하루 전은 1, 이틀 전은 2, 같은 방식으로 6일전 이상은 6으로 변환하였고, 마감일 이후 일주일 이내 접속/제출은 -1, 4주 이내 -2, 8주 이내 -3, 미접속/미제출은 -4로 코딩하여 학습 또는 과제활동의 첫 학습시점을 파악할 수 있도록 하였다. 둘째, 학기 중 마감 당일 또는 마감일 이후 학습에 접속/제출한 빈도를 집계하여 벵락 치기 학습을 파악하는 변수(CramLateCount)를 생성하였다. 셋째, 학습시점의 규칙성 정도를 파악하기 위해 14주 동안 생성된 학습시점 데이터들의 최대값-최소값 차이, 평균, 표준편차를 계산하고 학습콘텐츠 유형별로 각각 생성하였다(예: Diff_Max_Min_VOD, Ave_Diff_VOD, Std_Diff_VOD 등).

<표 4> Feature engineering으로 생성한 변수에 대한 설명

변수 내용	정의 및 계산 방법	코딩	생성 변수	해석
다운로드 여부(0, 1)	강의노트 또는 파일의 다운로드 행동 구분	0=다운로드 하지 않음 1=다운로드 완료	Wx_Note _i _Down Wx_File _i _Down	학습노력 (성실도)
주차별 변수 변환	<ul style="list-style-type: none"> 학습콘텐츠 첫 접속일, 과제 첫 제출일 날짜 데이터로부터 계산 특정 학습자료에 첫 접속한 날짜와 마감일 간 차이를 수치화함. 마감일을 0으로 기준함 예: 동영상 시청 마감일이 일요일인데, 화요일 접속한 경우는 5, 토요일은 1, 마감일 접속은 0, 이후 1주 이내 접속은 -1, 4주 이내는 -2, 8주 이내 접속은 -3, 미접속은 -4로 코딩함 	6= 마감일 기준 6일 전(월요일), 5= 마감 5일 전, 4= 마감 4일 전, 3= 마감 3일 전, 2=마감 2일 전, 1=마감 1일 전, 0=마감일 접속, -1=마감 이후 1주 이내, -2=마감 이후 4주 이내, -3 마감 이후 8주 이내, -4= 그 이후 또는 미접속	Diff_Wx_VOD _i , Diff_Wx_File _i , Diff_Wx_Link _i , Diff_Wx_Activity, Diff_Wx_AssignSubmit, 총 33개 생성	시간 관리

<표 4> Feature engineering으로 생성한 변수에 대한 설명

(계속)

변수 내용	정의 및 계산 방법	코딩	생성 변수	해석
벼락치기 횟수 (0-33)	· 주차별 학습에서 마감일 대비 첫 학습시점(Diff_변수)을 구하는 변수 총 33개 중 0 또는 음수값 (늦은 제출 또는 미제출) 횟수	1주-14주 학습에서 생성한 첫 학습시점 변수의 값이 0 이하인 경우 횟수	CramLateCount	시간 관리
다운로드 비율(0-100)	· 총 17개의 다운로드 대상 자료 중 다운로드 받은 비율	$=(\sum(\text{주별 } Wx_Note_Down, \\ Wx_File_Down)/17)*100$	DownRate	학습노력 (성실도)
기말시험 준비 (0 이상)	· 기말시험(6/13일)을 앞두고 이전 주차부터 학습콘텐츠 열람 횟수 (6/2-6/13일 사이에는 마감일이 지정된 학습콘텐츠는 없으므로 기말고사 준비 목적으로 콘텐츠 를 열람한 것으로 간주)	6월 2일부터 6월 13일까지 영상, 노트, 파일, 링크 열람 횟수의 총합	ExamWeekFreq	학습노력 (성실도)
활동 및 과제 완료(0-5)	· 2-14주 간 총 5개의 활동 및 과제 완료 횟수 (1주차 활동은 제외: 학습자소개)	2, 7, 10, 12, 14주차의 활동/과제 완료 횟수	Total_Task_Completion	학습노력 (성실도)
전체 집계 변수 생성	· 영상: 모든 강의영상에 첫 접속 시점 데이터 중 가장 큰 값과 가장 작은 값 간의 차이 (Diff_Max_Min_VOD), 모든 강의 영상에 대한 첫 접속시점의 평 균(Ave_Diff_VOD), 표준편차 (Std_Diff_VOD)로 수업 전반에서 의 학습 시점의 일관성을 수치 화함	$\text{Diff_Max_Min_VOD} = \\ \text{MAX}(\text{Diff_Wx_VOD}i \\ \text{값들}) - \text{MIN}(\text{Diff_Wx_VOD}i), \\ \text{Ave_Diff_VOD} = \\ (\sum \text{Diff_Wx_VOD}i) / \text{변수갯수}, \\ \text{Std_Diff_VOD} = \\ \sqrt{\sum (\text{Diff_Wx_VOD}i - \text{Ave_Di} \\ \text{ff_VOD})^2 / \text{변수갯수}},$	Diff_Max_Min_VOD, Ave_Diff_VOD, Std_Diff_VOD, Diff_Max_Min_File, Ave_Diff_File, Std_Diff_File, Diff_Max_Min_Link, Ave_Diff_Link, Std_Diff_Link, Diff_Max_Min_Task, Ave_Diff_Task, Std_Diff_Task	시간 관리
전반적 첫 학습시점 규칙성	· 파일: 학습파일 첫 접속시점 데 이터 중 최대 차이, 평균, 표준 편차를 계산하여 파일 학습에 대한 학습행동 일관성 파악 · 링크: 링크의 첫 접속시점 간 최대차이, 평균, 표준편차를 계 산하여 링크 학습에 대한 일관 성 파악 · 활동 및 과제: 활동/과제 첫 제 출 시점 중 최대 차이, 평균, 표 준편차를 구하여 활동/과제 수 행 시점의 일관성 파악	$\text{File, Link, Task 유형도}$ 동일한 방식으로 적용		
기타	· 교과목 수강신청시 P/F 또는 A-F 학점 유형으로 신청했는지 구분	0=P/F 학점 신청, 1=A-F 학점 신청	ScoreOption	학습노력 (성실도)

다운로드 행동은 학습을 위한 노력을 보여주는 것으로, 강의노트와 학습파일 다운로드 횟수를 다운로드 여부 변수로 변환하였다. 다운로드 횟수를 이분화 변수로 변환한 이유는 동영상 열람의 반복 행동은 재복습의 의미로 해석할 수 있으나, 노트와 파일의 다운로드의 한 번의 행동으로 의미 파악이 충분하기 때문이다(Arizmendi et al., 2023). 파일 및 노트를 다운로드한 비율(DownRate), 기말고사를 앞두고 2주간 학습내용에 접속한 횟수(ExamWeekFreq), 활동과 과제 완료 총 횟수(Total_Task_Completion) 변수를 생성하였다. 이 외에 수강신청 정보에 따라 Pass/Fail 학점으로 신청했는지, A-F 학점으로 신청했는지를 구분하는 ScoreOption 변수를 추가하였으며 학점 선택 유형이 학습자의 노력과 수강태도에 관련이 있을 것으로 판단하였기 때문이다. 설명변수는 LMS 로그 데이터로부터 수집 및 생성한 122개 변수와 수강신청 시스템에서 수집한 학점 신청 유형 1개 변수를 포함하여 총 123개를 활용하였다.

다. 반응변수

본 연구에서 반응변수는 해당 교과에 대한 학업성취도로 15주차에 대면으로 치러진 기말고사 시험성적을 수집하였다. 해당 교과목에서 시험은 1회로, 수업에서 다룬 내용을 토대로 30문항이 출제되었다. 채점은 문항당 1점으로, 총 30점이 만점이었고, 평균 14.90, 표준편차는 3.98로 나타났다.

3. 분석 방법

학업성취도 예측 모델링에 elastic net을 활용하였다. Elastic net 기법은 ridge regression과 lasso regression을 결합한 것으로, 다중공선성 문제를 완화하는 ridge regression과 설명변수가 많을 때 간명한 모델을 만드는 lasso regression의 장점을 모두 지닌다(Zou & Hastie, 2005). 또한 결과 해석이 용이하며, 본 연구의 데이터와 같이 사례수가 적은 경우, 과적합을 억제하며 복잡한 모형에 비해 상대적으로 안정적인 결과를 유지한다(Sonnleitner et al., 2025). Elastic Net은 α 와 λ , 두 개의 조절 모수를 가진다. α 는 ridge와 lasso의 혼합비로 0-1사이 값을 가지며, 0에 가까울수록 ridge의 특성이, 1에 가까울수록 lasso의 특성이 강하다. λ 는 회귀계수를 얼마나 축소할지 벌점크기를 결정하는 규제강도로, λ 값이 클수록 회귀계수가 축소된다(Zou & Hastie, 2005).

본 연구에서는 모델의 일반화 성능을 편향 없이 추정하고자 모델 성능 평가와 하이퍼파라미터 최적화 단계로 나누는 중첩 교차타당화(nested cross-validation)를 적용하였다. 모형 성능 평가 단계에서는 훈련용 데이터를 최대한 활용할 수 있는 LOOCV(Leave-One-Out Cross Validation)를 적용하였다. LOOCV는 n 개의 자료가 있을 때, 하나를 검증용으로 빼고, 나머지 $n-1$ 개를 훈련에 사용한다. 이 과정을 n 번 반복하여 모든 데이터를 한 번씩 검증 데이터로 활용하므로 자료의 사례수가 적은 경우에 적절하다(Kohavi, 1995). 하이퍼파라미터 최적화 단계에서는 훈련데이터에 대해

5-fold CV를 5회 반복하여 전처리 과정과 elastic net의 하이퍼파라미터(α , λ)를 결정하였고, 모형 평가는 RMSE(root mean square error)와 MAE(mean absolute error)를 활용하였다. 분석은 R을 사용하였으며, 데이터 전처리(dplyr), elastic net 모델링과 교차검증(glmnet, caret), 시각화(ggplot2) 패키지를 분석에 활용하였다.

III. 연구 결과

1. 학습 양상

가. 발생 이벤트 분석

수집된 로그 데이터로부터 분석한 학습 양상은 다음과 같다. 먼저 15주간 학생 개인 1인당 로그 개수의 평균은 380.63($SD=152.08$)이었고, 최소값은 111, 최대값은 818이었다. 발생 이벤트 중 주요 항목을 살펴보면(<표 5> 참조), 학습자료/활동을 클릭함이 5,628건으로 가장 많았다. 학생들이 자료를 다운로드 한 이벤트 횟수는 824건, 게시판의 게시글을 열람한 이벤트 횟수는 955건, 과제물 업로드 완료는 237건, 과제물 제출 상태를 확인하는 이벤트 횟수는 1,612건, 이수현황 보고서를 열람한 이벤트 횟수는 1,210건 등이었다.

<표 5> 학습자 주요 학습행동 이벤트 항목별 분석 결과

주요 학습행동	빈도	1인당 평균
파일 다운로드	824	17.17
게시글을 열람(클릭)	955	19.90
과제물 업로드 완료	237	4.94
과제물 제출 상태 확인함	1,612	33.58
이수현황 보고서 열람(클릭)	1,210	25.2
학습자료/활동을 클릭함	5,628	117.25

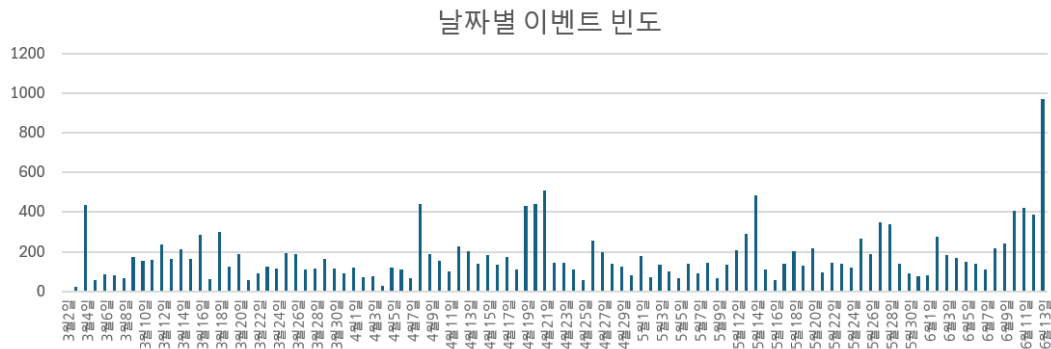
또한 학습콘텐츠를 동영상, 링크, 노트, 파일 유형으로 구분하여 발생한 이벤트를 집계한 결과는 <표 6>과 같다. 진도율로 출석 확인이 반영되는 동영상의 경우 1인당 열람 횟수는 평균 1.88이었고, 최소값은 1.21, 최대값은 2.75로 나타났다. 파일의 1인당 이벤트 평균은 6.30인데, 최소값이 1.58, 최대값이 10.38로 다른 유형에 비해 큰 편차를 보였다. 그 이유는 1주차의 과제 안내, 학습활동지와 같이 필수 내용을 담은 파일에 대한 이벤트 수는 매우 높은 반면, 학습주제에 대한 참고내용을 담은 파일(예: 9주차 파일: 생성형 AI 윤리 가이드라인)의 경우의 이벤트(클릭)

수가 적었기 때문이다. 링크 유형은 1인당 평균이 1.18, 최소값이 0.54, 최대값이 1.63으로 가장 저조한 학습행동 횟수를 보였다. 이러한 결과는 콘텐츠 유형별로 학습자의 다른 학습행동 경향을 보인다고 해석할 수 있다.

<표 6> 콘텐츠 유형별 발생 이벤트 분석 결과

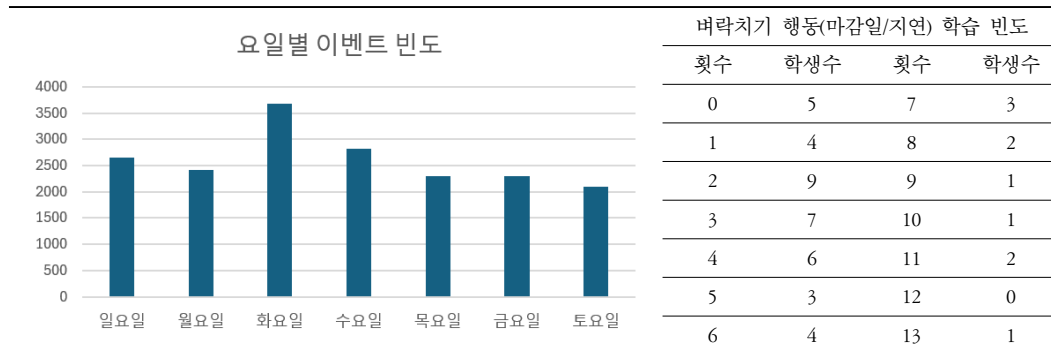
콘텐츠 유형	항목 수	총 이벤트수	구분	평균	최소	최대
동영상	19	1,710	항목당	90	58	132
			1인당	1.88	1.21	2.75
링크	4	228	항목당	57	26	78
			1인당	1.18	.54	1.63
노트	13	1,868	항목당	143.69	63	273
			1인당	2.99	1.31	5.69
파일	4	1,210	항목당	302	76	498
			1인당	6.30	1.58	10.38

학습자들이 접속한 날짜를 기준으로 학습행동 양상을 분석한 결과, 개강 초기, 활동 또는 과제 마감일에 임박한 날짜, 기말고사에 임박한 날짜에서 400건 이상의 높은 학습행동 이벤트가 나타났다(그림 1) 참조).



[그림 1] 15주 간 날짜별 발생 이벤트 빈도

[그림 2]를 살펴보면, 매주 학습내용에 대한 학습기간은 월요일부터 일요일까지로 지정되어 있는데, 학습자의 요일별 접속 빈도를 분석한 결과, 화요일에 발생한 이벤트가 3,685건으로 가장 많았고, 그 다음으로 수요일 2,820건, 일요일이 2,652건이었다. 학습 마감일이 일요일 자정인 점을 고려할 때, 마감일에 임박한 학습이 두드러지지 않고 요일별 접속 및 학습행동이 대체로 고



[그림 2] 요일별 발생 이벤트 빈도와 개인별 벼락치기식 학습행동(CramLateCount) 빈도 그래프

르게 나타났다. 한편, 각 주차에서 생성한 학습시점 변수 중에서 값이 0이하인 경우의 빈도를 집계한 CramLateCount 변수를 분석하였다. 0 또는 작은 값은 마감일 당일 또는 마감일 이후에 학습동영상, 링크, 파일, 활동 링크를 첫 클릭하거나, 과제 제출을 지연 및 미완료한 경우에 해당한다. CramLateCount 변수 값은 0-33의 값을 갖는 변수로, 모든 학습행동을 마감일에 임박하지 않게 학습한 학습자가 5명, 1번 마감일 또는 지연 학습행동을 한 학습자 수는 4명이었다. 33회 중 20%이하(약 7회 이하)로 벼락치기 학습행동을 보인 학습자는 41명으로 수강생의 85%였다. 반면 벼락치기 학습양상을 8회 이상 해당 학습자 수는 총 7명으로, 약 15%에 해당하였고 최댓값은 13회(1명)이었다. CramLateCount 변수의 경향을 파악함으로써 전체 학생의 요일별 분석에서는 잘 드러나지 않았던 개인 학습자별 벼락치기식 학습행동을 관찰할 수 있었다.

이 외에 생성한 변수의 기술통계는 다음과 같다. 기말고사를 앞둔 기간(6/2-6/13)에 학습콘텐츠 열람 빈도(ExamWeekFreq)의 평균은 7.17, 표준편차는 6.71이었고, 최소값은 0, 최댓값은 20으로 나타나 학생 간 차이가 큰 편이었다. 또한 강의노트와 학습자료 파일을 다운로드한 비율(DownRate)의 평균은 53.68%, 표준편차는 30.03으로 분석되었다. 최소값은 5.88, 최댓값은 94.12로 다운로드 행동의 학습자 간 편차도 큰 것으로 해석할 수 있다. 활동 및 과제를 완료한 총 횟수(Total_Task_Completion)의 평균은 4.73으로 대부분의 학생이 활동 및 과제를 완료한 것으로 나타났다. 총 5회 중 4회 제출한 학생이 8명, 한 번도 제출하지 않은 학생이 1명이었다. 학점유형 선택(ScoreOption)은 Pass/Fail로 신청한 학생이 25명(52.1%), A-F학점으로 신청한 학생이 23명(47.9%)이었다.

2. 기계학습 기법으로 학업성취도 예측 변수 탐색

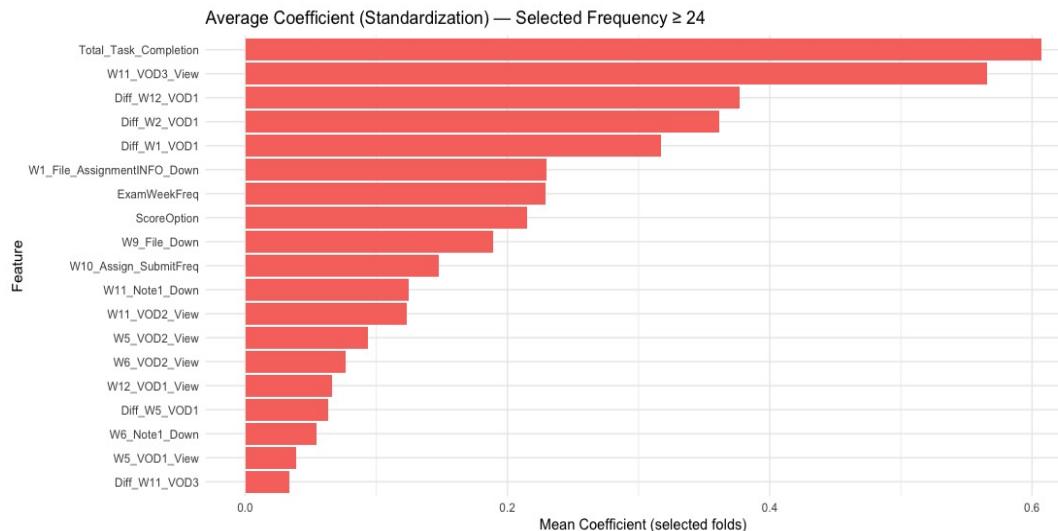
총 123개의 설명변수를 활용하여 elastic net으로 학업성취도 예측 모델을 구축하였으며, 조수 모수 α 는 0.2, λ 는 2.33으로 결정되었고, $RMSE = 3.62$, $MAE = 2.91$ 로 분석되었다. 학업성취도 예

<표 7> 최종모형에서 반복 선택되는 변수의 빈도(상위 30개) 및 평균 계수

No	변수	변수 내용	선택 빈도	평균 계수
1	W11_VOD3_View	11주차 3번째 강의영상(AI윤리#3) 열람 횟수	48	0.565
2	Diff_W12_VOD1	12주차 첫 번째 강의영상(최신 생성형 AI 모델)의 첫 열람일과 마감일 간 차이	48	0.377
3	Diff_W2_VOD1	2주차 강의영상(디지털 교육플랫폼과 콘텐츠)의 첫 열람일과 마감일 간 차이	48	0.361
4	Diff_W1_VOD1	1주차 오리엔테이션 영상의 첫 열람일과 마감일 간 차이	48	0.317
5	W1_File_AssignmentInfo_Down	1주차 파일(과제 상세 소개)의 다운로드 여부	48	0.230
6	ExamWeekFreq	기말고사 시험 준비 기간 콘텐츠 접속 빈도	48	0.230
7	ScoreOption	학점 이수 선택 유형 (Pass/Fail, A-F 학점)	48	0.215
8	W9_File_Down	9주차 파일(생성형 AI 윤리가이드북)의 다운로드 여부	48	0.189
9	W11_Note1_Down	11주차 강의노트(AI 윤리)의 다운로드 여부	48	0.125
10	W11_VOD2_View	11주차 두 번째 강의영상(AI 윤리#2)의 열람 횟수	48	0.123
11	Total_Task_Completion	학습활동 및 과제 완료 총합	47	0.607
12	W10_Assign_SubmitFreq	10주차 과제제출 횟수	47	0.147
13	W5_VOD2_View	5주차 두 번째 강의영상(기계학습3: ML의 주요 알고리즘)의 열람 횟수	47	0.094
14	W12_VOD1_View	12주차 첫 번째 강의영상(최신 생성형 AI 모델)의 열람 횟수	46	0.066
15	Diff_W5_VOD1	5주차 첫 번째 강의영상(기계학습2: ML 처리과정)의 첫 열람일과 마감일 간 차이	46	0.063
16	W6_VOD2_View	6주차 두 번째 영상(딥러닝2)의 열람 횟수	44	0.077
17	W6_Note1_Down	6주차 강의노트(딥러닝)의 다운로드 여부	39	0.055
18	Diff_W11_VOD3	11주차 3번째 강의영상(AI 윤리#3)의 첫 열람일과 마감일 간 차이	30	0.034
19	W5_VOD1_View	5주차 첫 번째 강의영상(기계학습2: ML 처리과정)의 열람 횟수	27	0.039
20	W2_File_Down	2주차 파일(새로운 혁신 교육모델 사례) 다운로드	19	0.067
21	Diff_W4_VOD2	4주차 두 번째 강의영상(AI발전역사)의 첫 열람일과 마감일 간 차이	8	0.020
22	W6_VOD1_View	6주차 첫 번째 강의영상(딥러닝1) 열람 횟수	5	0.056
23	W2_Activity_Complete	2주차 학습활동 완료 여부	4	0.083
24	W11_Note2_Down	11주차 두 번째 강의노트(AI 윤리#2) 다운로드	3	0.032
25	Diff_W2_Activity	2주차 학습활동 첫 접속일과 마감일 간 차이	2	0.037
26	W2_Activity_View	2주차 학습활동 접속 횟수	2	0.021
27	W7_File_Down	7주차 파일(학습활동지 과제) 다운로드 여부	1	0.134
28	W7_AssignmentFolder_View	7주차 과제함 열람 횟수	1	0.099
29	W9_File_View	9주차 파일 열람 횟수	1	0.059
30	W10_AssignmentFolder_View	10주차 과제함 열람 횟수	1	0.053

측에 중요 변수 파악은 선택되는 변수의 안정성을 고려하고자 LOOCV로 48개의 최종모형(48 fold)에 반복적으로 해당 변수가 선택되는 빈도를 산출하였다. <표 7>에 선택 빈도가 높은 순으로 상위 30개를 평균 계수와 함께 제시하였다.

<표 7>을 보면, 48회 모든 모형에 포함된 변수는 총 10개, 47회 선택된 변수는 3개, 46회 선택된 변수는 2개, 44회 선택된 변수는 한 개였다. 안정적인 변수 선정을 위해 48개 모형에 최소 50%(24회) 이상 반복적으로 선택된 변수를 선별한 결과, 19개로 나타났다. 변수의 평균 계수 크기로 학업성취도에 가장 크게 기여하는 변수는 활동 및 과제 제출완료 횟수(Total_Task_Completion, $\beta=.61$)였고, 11주차 3번째 강의동영상 열람을 많이 할수록($\beta=.56$), 12주차, 2주차, 1주차의 동영상 학습을 일찍 시작할수록 학업성취도를 높게 예측하였다. 안정적으로 선택된 19개 변수들의 평균 계수를 시각화하여 [그림 3]에 제시하였다.



[그림 3] 최종 모형에 50%(24회) 이상 반복적으로 선택된 변수와 평균 계수

19개 변수를 변수 특징으로 분류한 결과(<표 8> 참조), 시간관리 행동에 관련된 학습시점 변수가 5개, 성실한 학습행동에 관련된 변수들에는 강의영상 열람 횟수 변수가 6개, 파일 및 노트의 다운로드 변수가 4개, 과제 관련 변수 2개, 그 외에 기말고사 준비기간에 학습콘텐츠 접속 횟수, 교과목 수강신청시 선택하는 학점신청 유형 변수로 분류되었다.

강의 영상 열람 횟수는 5, 6, 11, 12주차에 속한 강의 동영상으로 머신러닝 처리과정, 다양한 기계학습 알고리즘, 딥러닝과 같은 개념과 원리를 다루는 주제와 인공지능 윤리, 생성형 인공지능의 활용과 이슈 등을 다루는 주제였다. 이 동영상의 열람 빈도가 높을수록 학업성취도 성취에 긍정적 관계를 보였고, 특히 11주차 세 번째 영상(W11_VOD3_View)의 열람 횟수가 높을수록 학

<표 8> 선정된 19개 중요 변수 구분

해석	변수 유형과 관련 시기	변수 수	선정된 변수
시간관리	학습시점 관련 변수 (1, 2, 5, 11, 12주차)	5	Diff_W1_VOD1, Diff_W2_VOD1, Diff_W5_VOD1, Diff_W11_VOD3, Diff_W12_VOD1
	강의 영상 열람 횟수 변수 (5, 6, 11, 12주차)	6	W5_VOD1_View, W5_VOD2_View, W6_VOD2_View, W11_VOD2_View, W11_VOD3_View, W12_VOD1_View
학습노력행동 (성실도)	파일/노트 다운로드 여부 변수 (File: 1, 9주차, Note: 6, 11주차)	4	W1_File_AssignmentInfo_Down, W9_File_Down, W6_Note1_Down, W11_Note1_Down
	과제 관련 변수 (10주차, 전체 완료)	2	W10_Assign_SubmitFreq, Total_Task_Completion
	기타 (시험준비 기간, 수강신청시)	2	ExamWeekFreq, ScoreOption,

업성취도가 높게 예측되었다($\beta=.56$). 또한 5개의 학습시점 변수는 1, 2, 5, 11, 12주차의 동영상 학습과 관련된 것으로, 해당 주차의 동영상 학습시점이 빠를수록 학업성취도 상승이 예측된다고 해석할 수 있다. 특히 1, 2, 12주차의 학습시점 변수는 48회 모두 선정된 변수이면서 평균 계수가 .32-.38으로, 19개 선택 변수의 평균 계수 크기순으로 상위를 차지하여(3-5번째) 시간관리 행동의 중요성을 뒷받침하였다.

파일 및 노트 다운로드 변수 중 선택된 1주차 파일 다운로드(W1_File_AssignmentInfo_Down)는 교과목의 과제에 대한 상세 안내를 담고 있는 자료로, 이 자료의 다운로드 여부는 학업성취도와 정적 예측 관계를 보였으며($\beta=.23$) 다운로드한 학습자의 성취도가 더 높게 예측되었다. 9주차에 제공된 파일은 생성형 인공지능 활용시 알고 있어야 할 저작권, 개인정보 등 중요한 윤리에 대한 가이드라인을 제시한 자료로, 파일 다운로드 행동이 학업성취도 상승과 긍정적 관계를 보였다. 10주차의 과제 제출함 열람 횟수가 많을수록, 과제 제출 완료 횟수가 클수록 학업성취도가 높게 예측되었다. 특히 교과목에서 요구하는 활동 및 과제에 대해 완료 횟수는 기말고사 점수인 학업성취도 예측에 가장 큰 관계를 보였다. 이는 학습자들이 배운 지식을 심화하고 확장하는 활동과 과제 참여가 교과목의 학습내용 이해에 긍정적이라고 유추할 수 있다. 기말고사 시험 준비 행동을 보여주는 ExamWeekFreq 변수는 48회 모두 선택된 변수로, 시험 준비기간 학습활동 빈도가 높을수록 학업성취도가 높게 예측되었다($\beta=.23$). 마지막으로 교과목의 학점 신청 유형 선택은 A-F 학점으로 신청한 학습자의 학업성취도가 Pass/Fail 학점으로 신청한 학습자보다 더 높게 예측되었다. 학업성취도 예측에 선정된 중요 지표 결과를 종합해볼 때, 교과에서 요구되는 활동 및 과제를 완료하고, 수강 초기에 중요한 학습 안내 확인과 후반의 강의영상 학습 빈도, 수강 초와 수강 후반의 첫 학습일이 빠를수록 더 높은 학업성취도가 예측되었다고 해석할 수 있다.

IV. 논의 및 결론

본 연구는 대학에서 온라인으로 운영된 교과목의 LMS 로그 데이터를 수집하여 학습자의 학습 양상을 살펴보고, 학업성취도를 예측하는데 중요 변수를 탐색하고자 수행되었다. 연구문제에 따른 결과와 논의는 다음과 같다.

첫째, 15주간 운영된 온라인 기반 수업에서 48명의 학습자가 LMS와의 상호작용으로 발생한 이벤트는 18,270건에 달했으며, 이 중 강의동영상, 강의노트, 학습자료, 학습활동 등에 접속하는 열람 이벤트가 가장 높은 비중을 차지하였고, 게시글 열람, 과제물 제출, 파일 다운로드, 이수 현황 확인 등 다양한 학습자 행동 빈도를 파악할 수 있었다. 또한 온라인 수업은 일부 선행연구(Neiberding & Heckler, 2021; Theobald et al., 2021)에서 학습자가 마감일 압박해서 벼락치기 학습 행동을 보인다는 우려가 종종 제기되는 만큼, 온라인 학습 양상을 시간관리 측면에서 파악하였다. 먼저 한 학기 동안 날짜별 학습 양상에서는 개강일, 과제 제출일과 시험일 등 특정 이벤트가 있는 날짜 근처에서 쉽게 예상되듯이 학습이벤트 발생 규모가 현저히 크게 나타났다. 주차별 학습에서 매주 일요일 자정이 학습마감일임을 고려하여 요일별로 학습행동을 살펴보았으나 특별히 마감일 쏠림 현상은 두드러지지 않았다. 다음은 개인별로 마감일 당일 또는 이후에 첫 학습활동을 시작하는 벼락치기식 학습행동 정도를 분석하였는데, 9명(18.8%)이 미루기가 거의 없이 한 학기 학습을 완료하였고, 다수의 학습자들은 몇 번의 벼락치기 학습 경험이 있지만 우려할 수준의 문제적 행동을 보이지는 않았다. 수강생의 65%(31명)가 수집된 33개의 학습시점 데이터 중 4회 이하로 마감일 압박 학습행동을 보였다. 그러나 약 15%정도의 학습자는 8-13회의 벼락치기 학습행동을 보인 것으로 나타났다. 시간 관리 분석 결과를 종합해볼 때, 전반적으로 우려할 정도의 미루기 또는 벼락치기 학습이 온라인 학습에서 만연한 것이 아니었다. 이는 수강생이 교육대학원의 예비교사라는 특성을 지니고 있어, 학부생보다는 상대적으로 시간관리 전략을 잘 활용하는 것으로 추측할 수 있다. 하지만, 지속적으로 시간관리에 어려움을 겪는 학습자가 존재하는 것으로 나타나 향후 수업에서 학습자에게 스스로 학습일정을 조절할 수 있도록 리마인드 알림 세팅 기능을 활용하도록 추천하거나 학기 초반에 주차별 첫 학습시점 패턴을 파악하여 미루기 학습이 반복되는 학생들을 선별하여 맞춤 시간관리 지원 제공을 고려할 수 있다.

또한 학습콘텐츠 유형별로 발생한 이벤트 수를 분석한 결과로부터 학습자가 학습콘텐츠 유형에 따라 다른 학습행동 경향을 보인다는 점에 주목할 필요가 있다. 동영상 기반의 온라인 수업이나 플립러닝 수업에서 학습자의 학습참여를 파악할 수 있는 대표적인 지표는 동영상 진도율이다. 일반적으로 마감일까지 학습한 동영상 진도율에 따라 주차별 출결에 반영하고, 학습자들도 LMS 메뉴에서 자신의 학습현황을 확인할 수 있다. 하지만, 교수자는 수업 설계시 강의 동영상 이외에도 중요 내용을 담은 파일, 참고자료, 외부 콘텐츠 링크 등 추가 학습자원을 제공하지만, 학습자 입장에서는 이러한 자원에 대해서는 출결에 반영되지 않으므로 이 자원들에 대한 학

습 우선순위가 강의동영상과는 다를 수 있다. 현재 LMS에서는 이러한 추가 자원에 대한 학습자 열람 정도를 한눈에 파악할 수 있는 교수자 메뉴가 없으나, LMS 로그 데이터 분석으로 학습자가 강의노트와 학습파일을 열람하는지, 다운로드를 하는지, 또 외부 학습자원 링크를 클릭했는지까지 파악함으로써 학습자 행동을 면밀히 관찰할 수 있었다. 본 연구에서 유형별 분석 결과, 외부 학습자원에 대한 링크는 다른 콘텐츠 유형에 비해 학습빈도가 가장 저조하게 나타났고, 파일의 경우에는 파일 내용별로 열람 및 다운로드 이벤트 수에 현격한 차이는 것을 실증적으로 확인할 수 있었다. 이는 학습유형과 자료 내용에 따라 학습자가 자료에 대한 중요도나 필요성을 다르게 인식하고 학습동기에 영향을 주는 것으로 해석할 수 있다. 따라서 온라인 수업 설계시 학습콘텐츠의 유형 중 학습자료 파일이나 외부 콘텐츠 링크는 그 중요도를 고려하여야 함을 시사한다. 즉, 해당 자료가 학습에 필수적인 중요정보를 담고 있다면, 관련 자료에 대한 학습 이해도를 파악할 수 있는 학습활동을 추가하여 학습자의 자료 접속과 학습을 가시화하는 설계로 보완할 수 있다.

둘째, 본 연구는 학업성취도 예측에 LMS 로그 데이터로부터 자기조절학습 기반의 온라인 학습행동을 반영할 수 있는 123개의 설명변수를 활용하여 예측에 중요한 19개 변수를 선별하였다. 변수 선정에는 학습을 위한 노력과 성실한 학습행동을 대리할 수 있는 변수와 시간관리와 연관된 학습행동 변수에 초점을 두었으며, 기존 연구에서 주로 활용하는 빈도 유형의 지표, 학습시점의 규칙성 이외에 마감일 대비 첫 학습시점, 마감일 임박 벼락치기 학습행동, 시험 대비 준비 행동, 자료의 다운로드 비율 등의 설명변수를 생성하여 연구에 활용하였다. Elastic net을 활용한 분석결과, 학업성취도 예측에 중요한 19개의 변수를 선별하였으며, 학습자의 성실한 노력을 보여주는 지표가 14개, 시간관리 관련 지표가 5개가 선정되었다.

선택된 지표를 학습 관점에서 해석해보면 다음과 같이 논의할 수 있다. 총 과제 완료 횟수는 가장 큰 크기의 계수로 학업성취도를 예측하는 것으로 나타났는데, 이는 과제 요구사항에 따라 자원을 탐색하고 재구성하면서 마감일을 지켜 학습산출물을 제출하는 행동이 성실한 학습과 노력을 반영하는 것으로 추론할 수 있다. 또한 학습자가 주차별 학습자원을 클릭하고 기말고사를 앞둔 기간에 학습 콘텐츠에 접속한 빈도는 학습을 성실히 수행하고 반복하는 학습행동으로 해석할 수 있다. 학습 빈도는 기존 연구에서도 학습성취 예측에 주요 요인으로 나타나 유사한 결과로 볼 수 있으며(김향미, 전광호, 2025; Chen & Cui, 2020), 기말고사 준비 학습콘텐츠 열람 변수는 계수 크기가 7위에 랭크될 정도의 중요한 변수로 나타났다. 본 연구의 대상 교과목의 경우, 기말시험을 앞둔 약 2주간에는 과제를 수행하고 제출하는 14주차와 전체 자율학습으로 15주차를 운영함에 따라 14, 15주차에는 의무적으로 학습해야 하는 학습콘텐츠는 없었다. 따라서 이 기간 동안에 학습자가 지난 학습콘텐츠를 열람하는 것은 복습행동으로 간주할 수 있으며, 학습자 스스로 학습을 위한 노력 행동이 학습성취를 향상시키는 것으로 해석할 수 있다.

다운로드 변수의 경우, 1주차에 제공한 과제 상세 안내 자료의 다운로드 여부가 학습성취 예

측에 중요한 변수로 선택되었을 뿐 아니라, 계수의 크기 측면에서도 높은 순위를 보인 점에서 의미를 찾을 수 있다. 이는 다운로드 행동 중에서도 다른 자료의 다운로드 행동과는 중요도가 다를 수 있음을 시사한다. 이 결과는 You(2016)에서 온라인 교과목 운영에 대한 지침을 읽었는지 여부가 학습성취 예측에 중요한 요인으로 분석된 결과와 유사한 맥락으로 사료된다. 비대면 학습상황에서 한 학기 교과목에 대한 중요 정보를 숙지하는 것은 한 학기 동안의 수강 태도와 노력에 영향을 미칠 수 있는 결정 요소로 학습성취 예측에 중요하다고 추론할 수 있다. 이 외에 주차별 강의노트와 참고자료의 다운로드 행동은 학습에 필요한 정보를 획득하려는 기본적인 탐색 활동으로 추후 학습을 위해 관련 자료를 충실히 확보하려는 행동이자 성실한 학습태도를 반영하는 것으로 해석할 수 있다. 수강 학점 유형이 중요 변수에 선택된 것은 학습자가 스스로 설정하는 학습성취 목표에 영향을 미치기 때문으로 해석할 수 있다. Pass/Fail을 선택한 학습자는 Fail 기준을 초과하는 정도의 낮은 성취목표를 세움으로써 A-F학점을 신청한 학습자에 비해 수강과정에서 보이는 노력이 부족할 것으로 추론할 수 있다.

또한 본 연구에서는 매주 제공된 학습에 대하여 첫 학습시점을 마감일 대비 얼마나 빠르게 혹은 베퉼치기 학습을 하는지의 차이를 반영하는 지표로 생성하여 활용하였고, 총 5개의 지표가 최종 예측모형에 선정되었다. 마감일 관련 선행연구에서는 지연학습과 학습성취 예측에 부정적인 관계를 보고하고 있는데(You, 2016), 대부분 마감일을 넘겼는지 아닌지에 따라 지연학습을 정의하고 있다. 하지만 본 연구에서는 주차별로 월요일부터 공개되어 마감일인 일요일까지 일주일 중 언제 첫 학습행동을 보이는지 시점을 파악하여 첫 학습시점 날짜와 마감일 간의 일수 차이로 값을 계산하였다. 이는 학습 시간에 대한 조절 행동으로 조기에 학습을 시작하는 학습자와, 마감일에 임박한 학습자, 마감일을 지난 학습자 등 첫 학습시점의 차이를 반영하려는 의도였다. 대부분 시간관리의 중요성을 밝힌 선행연구(박지우, 임규현, 2023; 조일현, 김운미, 2013)에서는 전체 수강기간 동안 학습접속 간격과 시점규칙성 변수로 학습 접속 시점 간 평균값과 표준편차값을 활용하고 있다. 본 연구에서도 이들 변수를 설명변수에 포함하였으나, 새롭게 생성한 마감일 대비 첫 학습시점 변수가 중요변수로 최종 선정된 점을 주목할 필요가 있다. 이는 시간관리 측면에서의 자기조절학습을 대리하는 새로운 변수 발굴을 시사하며, 투입한 123개 중 최종 선정된 19개 변수에 5개의 마감일 대비 첫 학습시점 변수가 포함되었다는 점에서 그 잠재성과 유용성을 찾아볼 수 있다.

한편, 본 연구에서 대부분 선택된 변수는 특정 주차에서의 학습빈도, 학습시점, 학습자료 다운로드 변수였다. 따라서 이 변수들은 운영한 교과에서의 주차별 내용 및 활동 등 학습맥락을 고려하여 해석함으로써 교과 설계 및 운영에 참고할 필요가 있다. 강의영상 열람횟수는 5, 6, 11, 12주차의 강의동영상이 중요 변수로 나타났는데 이것은 학습주제와 관련성이 있다고 추측된다. 5, 6주차에서는 인공지능 리터러시 내용 중 머신러닝의 처리과정, 주요 알고리즘(KNN, SVM, Random Forest, K-Means Clustering 등) 원리와 딥러닝에 대한 내용을 다루고 있어, 비전공자 예비

교사들에게 생소하고 다소 어려운 학습내용으로 사료된다. 실제 다른 주차에 비해 수업영상 열람 횟수가 많았으며, 이 주제에 대한 반복 학습이 학업성취도 예측에 중요했을 것으로 추론할 수 있다. 한편 11주차에는 인공지능 윤리와 윤리적 이슈 사례, Moral Machine 활동 등을 다루었는데 해당 영상의 열람 빈도는 다른 동영상보다 낮게 나타났다. 학습내용 자체는 그리 어려운 내용이 아니어서 영상 열람 빈도가 낮을 수 있겠으나, 인공지능 윤리는 교과목에서 중요한 주제이며, 무엇보다 강의 후반부에 배치된 만큼 꾸준히 집중력을 가지고 온라인 학습을 성실히 수행하는지가 예측 변수 선택에 반영되었을 것으로 추측된다. 12주차는 최신 생성형 인공지능의 특징, 다양한 멀티모달 생성형 인공지능 도구를 소개하고 이를 학생들이 직접 사용하면서 교육용 콘텐츠 개발에 활용 가능성을 살펴보는 과제 수행으로 구성되었다. 최신 생성형 인공지능 모델 정보를 제공하고 강의 영상과 연계된 과제 활동이 제시되어 12주차 강의동영상 활용 빈도가 높았고, 강의 후반에 배치된 수업으로 중요 변수로 선택된 것으로 추측해 볼 수 있다. 학습시점 변수는 1, 2, 5, 11, 12주차에서 선택되었는데 1, 2주차는 교과목 수강 초기로 학습에 대한 루틴을 만들고 학습에 대한 시간관리가 중요한 시기라 할 수 있다. 또한 강의동영상 콘텐츠가 마지막으로 12주차에 제공된 점을 고려한다면, 11, 12주차는 수업 후반에 해당한다. 따라서 수강 초기, 중반, 후반에서의 학습시점이 중요 변수에 선택된 것으로, 이는 온라인 학습 전반에 걸쳐 마감일에 촉박한 학습을 지양하고 주차별 첫 학습을 일찍 시작하는 시간 관리 행동이 학업성취도 예측에 중요하다고 해석할 수 있다. 이상의 결과를 토대로 교과 운영 개선에 적용해 보면, 학습자가 수강 초기에 교과목 운영지침과 과제 내용을 숙지할 수 있도록 강조하고, 각자의 시간관리 루틴을 설정할 수 있도록 안내하며, 특히 수강 후반에 끝까지 집중력을 잃지 않고 시간관리를 하면서 학습하도록 독려할 것을 제안할 수 있다. 또한 제공하는 학습자료 유형에서 필수 자료의 경우 동영상 강의에서 다루거나 학습자료에 대한 학습여부를 가시화할 수 있는 간단한 토론 활동 등을 고려할 수 있다. 학습자들이 어려워하는 주차의 주제는 이해를 높이기 위해 강의자료와 설명을 보완하고, 학습자가 스스로 이해도를 점검할 수 있는 퀴즈를 추가하는 등의 방안도 고려할 수 있다.

본 연구의 시사점과 의의는 다음과 같다. 첫째, 온라인 학습성취 예측에 중요한 변수를 LMS 로그 데이터로부터 탐색하는데 학습자의 다양한 학습행동을 포착할 수 있는 123개의 설명변수를 활용하였다. LMS 로그 데이터가 교육적으로 활용되려면 의미 있는 변수 추출이 전제되어야 하는데, 본 연구는 마감일 대비 첫 학습시작 시점, 마감일 임박 벼락치기 학습행동 빈도, 시험준비행동 등 LMS 로그 데이터로부터 학습자의 노력과 시간관리를 대리하는 변수를 추출하는데 아이디어를 제공했다는 점에서 의의가 있다. 특히 예측에 중요한 최종 19개 변수 중 강의동영상 열람 횟수, 과제 제출 빈도를 제외하면 대부분 전처리 과정을 거쳐 생성된 변수가 선택되었고, 생성된 변수는 단순 로그로 잘 드러나지 않는 학습 양상을 보여줌으로써 학습자의 온라인 학습 행동을 이해하는데 기여하였다. 하지만, 의미 있는 변수 활용의 중요성은 현재 단순 관리용 수

준의 로그 데이터를 제공하는 LMS의 개선 필요성도 시사한다. 현재 LMS 로그 데이터에는 학생 링크, 로그 이름, 접속 일시, 접속 날짜, 이벤트 문맥, 구성요소, 발생 이벤트 정도의 정보를 담고 있기 때문에, 의미 있는 변수를 발굴하는데 한계가 따른다. 따라서 교수학습 측면에서 의미 있는 데이터를 제공하고 학습 촉진을 지원하는 시스템으로써의 LMS로 역할과 제공 기능을 새롭게 확장할 필요가 있다.

둘째, LMS 로그데이터를 기계학습으로 분석하여, 기계학습 분석이 지닌 다양한 장점을 보여준 학습분석 연구 사례라는 점에 의의가 있다. 기존 전통적인 회귀모형으로 분석하기 어려운 123개의 설명변수를 활용하여 학업성취도를 예측하였고, 설명변수 수에 비해 적은 48명의 표본에서 18,270개의 로그 데이터를 수집하여 분석을 수행함으로써 수강규모에 제약 없이 수업에서 학습 분석 활용 가능성을 시사하였다. 또한 기계학습 분석을 적용함으로써 분석에 투입되는 데이터의 세밀한 정도(granularity)를 결정할 수 있다. 전통적 분석방법에서는 수업 전체를 집계하는 대표 변수를 쓸 수밖에 없었으나, 기계학습 방법 사용으로 변수 개수 제약에서 벗어나 주차별로 접속 빈도, 학습시점, 자료 다운로드, 과제 제출 등 세분화된 데이터를 분석에 활용할 수 있었다. 특히 본 연구에서는 주차별 변수 외에도 교과목 전반에 걸친 학습 규칙성을 파악하고자 학습시점 데이터의 최대-최소 차이, 평균, 표준편차 변수를 생성하였고, 노트 및 파일 다운로드 전체 비율, 벼락치기 학습양상 등 전체 학습행동을 반영하는 집계 변수를 생성하여 분석에 함께 투입하였다. 그러나 전체 집계 변수 중 학업성취도 예측에 선택된 것은 총 과제 제출 횟수와 기말고사 준비기간에 학습콘텐츠 접속 횟수 2개뿐으로, 주차별 세분화된 변수의 상대적 중요도를 보여주는 것이다. 이러한 논의는 Arizmendi와 동료들(2023)의 연구와 일맥상통하는 것으로, 저자들은 전체 콘텐츠 열람 횟수, 전체 다운로드 비율, 평균 접속 횟수, 평균 학습일 등 집계 변수를 사용할 경우 교과목 운영 시점에 따른 학습자 행동 변화를 포착할 수 없다고 지적하면서 세분화된 지표 활용을 추천한 바 있다. 따라서 세분화된 변수 사용은 특정 학습시점에서의 학습행동을 파악하고 교과목 맥락과 함께 해석함으로써 향후 교과목 운영과 학습촉진 전략 수립에 참고할 수 있는 구체적인 정보를 제공한다는 점에서 단순히 학습성취 예측의 중요 변수 탐색 이상의 의미를 지닌다.

본 연구는 다음과 같은 제한점이 있으며 이를 보완하는 후속 연구를 제안한다. 첫째, 학습분석은 수집되는 데이터의 맥락 변화에 민감한 특징으로 인해 일반화가 어렵다. 또한 기계학습으로 분석한 결과는 인과관계가 아니며, 반드시 교육적 이론으로 해석되지 않는다는 점도 학습분석 분야 연구가 지닌 제약이다(Arizmendi et al., 2023). 따라서 본 연구도 이러한 제약으로부터 자유롭지 못하므로 연구결과의 해석과 일반화에 주의가 필요하다. 하지만 예측 모델링 연구에서 지나친 일반화를 추구하기보다는 특정 맥락에 적용하고 해석할 때 특정 교과목에 대한 구체적인 개선안이나 개입 아이디어를 얻을 수 있다는 Gašević과 동료들(2016)의 논의를 상기할 필요가 있다. 따라서 후속 연구를 통하여 새롭게 발굴한 지표의 유용성을 재검증하고, 해당 교과목 맥

락에서 보완된 수업 운영으로 그 결과를 비교분석하고 학습분석적 개입 효과를 살펴볼 수 있다.

둘째, 예측력 향상을 위해 활용 변수 측면에서 보완을 고려할 수 있다. 기존 연구에서는 LMS 로그 데이터 이외에 학습자의 배경 변인(성별, 연령, 사전 학점 등)을 활용하여 예측하는 사례도 다수 나타나지만, 이러한 요소들은 학습자의 고정적인 특성으로 변화 되지 않고, 연구 대상마다 달라지며, 연구결과에서 LMS의 학습행동 데이터에 비해 기여도도 낮다고 밝혀진바 있다 (Jovanović et al., 2021). 본 연구에서는 학습자의 배경요인을 활용하지 않고 운영한 수업에 국한하여 학습자의 온라인 학습행동 데이터를 활용하였다. 그러나 학업성취도로 활용한 기말고사 이외에는 중간에 퀴즈나 중간고사 시험점수가 없어 평가데이터를 전혀 활용하지 못하였다. 향후 연구에서는 주차별 자율 퀴즈(self quiz)와 같은 콘텐츠를 활용함으로써 학습자의 학습을 촉진하는 동시에 학습분석의 데이터로도 활용할 것을 제안한다.

참고문헌

- 김향미, 전광호 (2025). 온라인 학습자의 학습행동과 성과에 대한 탐색적 연구: LMS 로그데이터 기반. *비즈니스융합연구*, 10(2), 93-100.
- (Translated in English) Kim, H., & Jeon, K. (2023). An exploratory study of online learners' behavior and performance: Evidence from LMS log data. *Journal of Business Convergence*, 10(2), 93-100.
- 김혜빈, 표은서, 신안나, 조일현 (2023). 국내 디지털 교수학습 환경에서의 학습분석 연구 및 데이터 활용에 대한 체계적 문헌 고찰. *교육공학연구*, 39(2), 595-638.
- (Translated in English) Kim, H., Pyo, E., Shin, A., & Jo, I. (2023). A systematic literature review of learning analytics and data utilization in a digital teaching and learning environment in Korea. *Journal of Educational Technology*, 39(2), 595-638.
- 박연정, 조일현 (2014). 학습 분석학 기반 대시보드의 설계와 적용. *교육정보미디어연구*, 20(2), 191-216.
- (Translated in English) Park, Y., & Jo, I. (2014). Design and application of visual dashboard based on learning analytics. *The Journal of Educational Information and Media*, 20(2), 191-216.
- 박지우, 임규연 (2023). 대학 온라인 학습에서 LMS 로그데이터를 활용한 자기조절학습 예측 요인 탐색: 학습 시간 관리를 중심으로. *교육정보미디어연구*, 29(2), 485-510.
- (Translated in English) Park, J. W., & Lim, K. Y. (2023). Exploring the predictors for self-regulated learning using LMS log data in online higher education. *Journal of Educational Information and Media*, 29(2), 485-510.
- 서진선, 이재경 (2020). LMS 로그 데이터 분석을 통한 온라인 협업 성취 예측모형에 관한 사례 연구. *현장수업연구*, 1(2), 121-154.
- (Translated in English) Seo, J. S., & Lee, J. K. (2020). A case study of on-line academic achievement prediction model by analyzing LMS log data. *Journal of Field-based Lesson Studies*, 1(2), 121-154.
- 신종호, 최재원, 박수영, 오상희 (2018). 대학 수업 지원을 위한 학습분석 기반 교수자 대시보드 개발. *교육정보미디어연구*, 24(3), 489-512.
- (Translated in English) Shin, J., Choi, J., Park, S., & Oh, S. (2018). Development of the dashboard based on learning analytics for teaching support in higher education. *Journal of Educational Information and Media*, 24(3), 489-512.
- 유진은 (2020). 기계학습 기법을 활용한 플립러닝 강좌의 LMS 로그파일 분석 사례 연구. *열린교육연구*, 28(5), 79-102.
- (Translated in English) Yoo, J. E. (2020). Learning management system log data analysis via machine learning: A case study from flipped learning. *The Journal of Yeolin Education*, 28(5), 79-102.

- 이성혜, 이유경, 박소영, 함은혜 (2021). 자기조절학습 행동데이터에 대한 탐색적 연구: 해외 학습 분석 연구를 중심으로. *교육공학연구*, 37(2), 191-236.
- (Translated in English) Lee, S., Lee, Y., Park, S., & Ham, E. (2021). An Exploration of self-regulated learning behavior data: International research on learning analytics. *Journal of Educational Technology*, 37(2), 191-236.
- 조일현, 김윤미 (2013). 이러닝에서 학습자의 시간관리 전략이 학업성취도에 미치는 영향: 학습분석학적 접근. *교육정보미디어연구*, 19(1), 83-107.
- (Translated in English) Jo, I., & Kim, Y. (2013). Impact of learner's time management strategies on achievement in an e-learning environment: A learning analytics approach. *The Journal of Educational Information and Media*, 19(1), 83-107.
- 홍유정, 함은혜, 이성혜 (2025). 시계열 온라인 학습행동의 유형화: 학습동기 및 성취목표와의 관계 탐색. *교육정보미디어연구*, 31(2), 319-346.
- (Translated in English) Hong, Y., Jam, E. H., & Lee, S. (2025). Typology of time-series online learning behaviors: Exploring the relationship with learning motivation and achievement goals. *The Journal of Educational Information and Media*, 31(2), 319-346.
- Arizmendi, C. J., Bernacki, M. L., Raković, M., Plumley, R. D., Urban, C. J., Panter, A. T., Greene, J. A., & Gates, K. M. (2023). Predicting student outcomes using digital logs of learning behaviors: Review, current standards, and suggestions for future work. *Behavior Research Methods*, 55(6), 3026-3054.
- Arnold, K., & Pistilli, M. (2012). Course signals at Purdue: Using learning analytics to increase student success. *ACM International Conference Proceeding Series*, 267-270.
- Bernacki, M., Chavez, M. M., & Uesbeck, P. M. (2020). Predicting achievement and providing support before STEM majors begin to fail. *Computers & Education*, 158, 103999.
- Brooks, C., & Thompson, C. (2017). Predictive modelling in teaching and learning. In C. Lang, G. Siemens, A. F. Wise, & D. Gašević, (Eds.), *The handbook of learning analytics* (pp. 61-68). Society for Learning Analytics Research (SoLAR).
- Bzdok, D., Altman, M., & Krzywinski, M. (2018). Statistics versus machine learning. *Nature Methods*, 15, 223-234.
- Chen, F., & Cui, Y. (2020). Utilizing student time series behaviour in learning management systems for early prediction of course performance. *Journal of Learning Analytics*, 7(2), 1-17.
- Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies*, 10(1), 17-29.

- Dabbagh, N., & Kitsantas, A. (2004). Supporting self-regulation in student-centered web-based learning environments. *International Journal on E-Learning*, 3, 40-47.
- Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28, 68-84.
- Howard, E., Meehan, M., & Parnell, A. (2018). Contrasting prediction methods for early warning systems at undergraduate level. *The Internet and Higher Education*, 37, 66-75.
- Jovanović, J., Saqr, M., Joksimović, S., & Gašević, D. (2021). Students matter the most in learning analytics: The effects of internal and instructional conditions in predicting academic success. *Computer & Education*, 172, 104251.
- Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computer & Education: Artificial Intelligence*, 3, 100074.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *International Joint Conference on Artificial Intelligence*, 14(2), 1137-1145.
- Long, P., & Siemens, G. (2011, September). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 31-40.
- Nieberding, M., & Heckler, A. F. (2021). Patterns in assignment submission times: Procrastination, gender, grades, and grade components. *Physical Review Physics Education Research*, 17, 013106.
- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385-407.
- Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, 21(1), 135-146.
- SoLAR (Society for Learning Analytics Research). (2025, May). *Reimagining learning analytics*.
<https://www.solaresearch.org/wp-content/uploads/2025/06/Reimagining-Learning-Analytics-V3-002.pdf>
- Sonnleitner, B., Madou, T., Deceuninck, M., Theodosiou, F., & Sagaert, Y. R. (2025). Evaluation of early student performance prediction given concept drift. *Computer & Education: Artificial Intelligence*, 8, 100369.
- Theobald, M., Ellhauser, H., & Imhof, M. (2021). Deadlines don't prevent cramming: Course instruction and individual differences predict learning strategy use and exam performance. *Learning and Individual Difference*, 87, 101994.
- Tiukhova, E., Vemuri, P., Flores, N. L., Islind, A. S., Óskarsdóttir, M., Poelmans, S., Baesens, B., &

- Snoeck, M. (2024). Explainable learning analytics: Assessing the stability of student success prediction models by means of explainable AI. *Decision Support Systems*, 182, 114229.
- Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & Education*, 143, 103676.
- Warnock, S., Bingham, K., Driscoll, D., Fromal, J., & Rouse, N. (2012). Early participation in asynchronous writing environments and course success. *Journal of Asynchronous Learning Networks*, 16(1), 35-47.
- Winne, P. H. (2022). Learning analytics for self-regulated learning. In C. Lang, G. Siemens, A. Wise & D. Gašević, & Merceron, A. (Eds.), *Handbook of learning analytics* (2nd ed., pp. 78-85). SoLAR, Society for Learning Analytics Research.
- You, J. W. (2016). Identifying significant indicators using LMS data to predict course achievement in online learning. *The Internet and Higher Education*, 29, 23-30.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of Royal Statistical Society Series B*, 67(2), 301-320.