

基于线性回归的学生学业表现分析与研究

李梓豪, 李子乐, 程文斌, 李鲁英*

武汉设计工程学院智能工程学院, 湖北 武汉

收稿日期: 2025年11月25日; 录用日期: 2025年12月19日; 发布日期: 2025年12月26日

摘要

本研究旨在通过线性回归模型分析和预测学生学业表现, 以弥补传统教育评估方法的不足。选取BI挪威商学院70名学生作为样本, 收集包括年龄、入学考试成绩、学习时间等数据, 并随机分为模型构建组(A组, $n = 35$)和验证组(B组, $n = 35$)。数据预处理包括清洗、缺失值均值填充和异常值保留; 特征工程涉及编码、标准化、SelectKBest筛选。采用Ridge回归构建Python和数据库(DB)课程成绩预测模型。结果显示, Python模型 $R^2 = 0.823$, $MSE = 42.36$; DB模型 $R^2 = 0.796$, $MSE = 48.72$ 。入学考试成绩和学习时间为主要正向影响因素, 年龄呈弱负相关。模型验证通过配对t检验($p > 0.05$), 证实泛化能力良好。研究表明, 学业表现受多因素综合作用, 该模型可为个性化教学和早期干预提供数据支持。局限性包括样本量小和特征维度有限, 未来可扩展样本并引入动态数据以提升精度。

关键词

线性回归, 学生学业表现, 预测模型, 特征工程, Ridge回归

Analysis and Research on Students' Academic Performance Based on Linear Regression

Zihao Li, Zile Li, Wenbin Cheng, Luying Li*

School of Intelligence and Engineering, Wuhan Institute of Design and Sciences, Wuhan Hubei

Received: November 25, 2025; accepted: December 19, 2025; published: December 26, 2025

Abstract

This study aims to analyze and predict students' academic performance through linear regression models, thereby addressing the deficiencies inherent in traditional educational assessment methods.

*通讯作者。

文章引用: 李梓豪, 李子乐, 程文斌, 李鲁英. 基于线性回归的学生学业表现分析与研究[J]. 交叉科学快报, 2026, 10(1): 94-102. DOI: 10.12677/isl.2026.101013

A sample comprising 70 students from BI Norwegian Business School was selected, with data collected on variables such as age, entrance examination scores, study time, and others. The sample was randomly partitioned into a model construction group (Group A, $n = 35$) and a validation group (Group B, $n = 35$). Data preprocessing encompassed cleaning, mean imputation for missing values, and retention of outliers. Feature engineering involved encoding, standardization, SelectKBest feature selection, Ridge regression was utilized to develop predictive models for grades in Python and Database (DB) courses. The results revealed that the Python model attained an R^2 of 0.823 and a mean squared error (MSE) of 42.36, while the DB model achieved an R^2 of 0.796 and an MSE of 48.72. Entrance examination scores and study time were identified as primary positive predictors, with age demonstrating a weak negative correlation. Model validation, performed via paired t-tests ($p > 0.05$), substantiated the models' strong generalization capabilities. The findings indicate that academic performance is shaped by a multifaceted interplay of factors, and this modeling approach can furnish data-driven support for personalized instruction and early interventions. Limitations include the modest sample size and constrained feature dimensions; prospective investigations may expand the sample and integrate dynamic data to augment predictive precision.

Keywords

Linear Regression, Student Academic Performance, Predictive Model, Feature Engineering, Ridge Regression

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着信息技术的快速发展，教育领域已积累了大量涵盖学生个人信息、学习行为与学业表现的多维度数据。如何有效挖掘这些数据中的潜在规律，为教育决策提供参考，已成为教育技术领域的重要研究方向[1]。传统教育评估方法主要基于统计描述和经验判断，在综合分析多因素对学生学业表现的影响方面存在一定局限。

当前，教育工作者在处理多维度学生数据时，往往面临识别学业表现需要关注的学生群体的挑战[2]。依赖期末考试成绩等结果性指标进行评估，可能导致教育干预的及时性受到影响[3]。此外，学生在学习基础、投入程度和外部环境等方面存在差异，统一的教学模式难以满足个性化教育需求，这可能影响学生的学习效果和整体教学质量的提升。研究表明，缺乏及时关注的学生在学业上可能面临更多挑战[4]，这凸显了基于数据的分析方法在教育领域的重要性。

在此背景下，学生成绩预测研究呈现出多元化发展态势[5]。在预测模型方面，从传统的线性回归到机器学习方法如决策树、随机森林[6]，再到梯度提升树等集成算法[7]，研究方法不断丰富。这些研究在提高预测准确性的同时，也面临着模型复杂性与可解释性之间的平衡问题。在关键预测变量方面，入学成绩、学习时间等基础特征被广泛证实具有重要预测价值[8]，而特征工程对提升模型性能的关键作用也得到研究确认。

然而，现有研究在以下方面仍可进一步探索：一是大多数研究基于大样本数据，在小样本情境下的模型稳健性研究相对有限；二是复杂模型虽然预测精度较高，但在实际教育场景中的应用可行性有待验证；三是虽然部分研究采用了正则化方法，但针对 Ridge 回归在小样本教育数据集上的系统评估尚不充分。

基于此，本研究着重探讨 Ridge 回归在小样本教育数据集中的应用效果，通过系统比较其与常规方

法的性能表现, 评估该方法在教育数据预测任务中的适用性。研究结果有望为教育实践提供一种兼顾性能与实施可行性的分析工具, 为教育工作者了解学生学习状况提供参考依据。

2. 数据集与特征工程

2.1. 数据来源

本研究以 BI 挪威商学院的学生为研究对象, 共选取 70 例样本。该样本规模基于初步统计分析和文献综述确定, 确保在有限资源下具备足够的统计功效, 以支持线性回归模型的构建和验证。在研究周期内, 通过学校教育管理系统和问卷调查相结合的方式, 系统收集了学生的相关数据。这些数据包括基本个人信息(如姓名、年龄、性别、国籍和居住地), 以及学业相关指标(如入学考试成绩、先前教育背景和学习时间投入)。同时, 记录了学生在 Python 编程课程和数据库(DB)课程中的最终成绩, 作为学业表现的量化评估标准。为保护隐私并符合伦理规范, 所有数据在收集后均进行匿名化处理, 仅保留必要变量用于分析。

2.2. 数据预处理

为确保数据可靠性和模型准确性, 本研究对原始数据集实施系统预处理, 包括清洗、缺失值填充及异常值处理[4]。这些步骤消除噪声、优化结构, 适应线性回归输入要求, 提升数据质量并减少偏差影响。统计学原理表明, 高质量预处理是鲁棒模型基础, 尤其小样本($n = 70$)下, 可显著降低过拟合风险。

首先, 进行数据清洗: 统一格式和修正拼写错误, 确保变量一致性, 避免计算偏差。

其次, 针对 Python 及 DB 成绩的少量缺失值(约 5%), 采用均值填充: 以组内均值替换, 基于分布特性最小化扰动, 保留样本完整性。该策略适用于低缺失率连续变量, 维持统计特性。

最后, 通过箱线图可视化离群点及 Z-score (阈值 ± 3)量化偏差检测异常值[5]。鉴于样本小, 为避免信息损失, 保留异常值但经审查确认真实性。该保守方法保持代表性, 并在敏感性分析中评估其模型影响。

2.3. 特征工程

为进一步提升线性回归模型的预测性能和泛化能力, 本研究对原始特征进行了特征工程处理, 包括特征编码、缩放、选择以及降维。这些转换和构造步骤旨在提取更有信息量的变量, 缓解多重共线性问题, 并优化模型的计算效率。特征工程在机器学习领域被视为关键环节, 据相关研究表明, 适当的特征处理可将模型的 R^2 值提升 10%~20%, 特别是在教育数据分析中, 能更好地捕捉学生学业表现的潜在模式。

首先, 对性别、国籍和居住地等分类变量进行标签编码(Label Encoding), 将其转换为数值形式。该编码方法将类别映射为整数序列, 便于模型输入, 同时保留了变量的序数信息(如无序类别)。其次, 对入学考试成绩、学习时间等数值型特征实施归一化和标准化处理。具体而言, 归一化将特征缩放到[0,1]区间, 而标准化则转换为均值为 0、标准差为 1 的分布, 以消除不同量纲对模型权重的影响。实验结果显示, 标准化处理对线性回归模型的性能提升尤为显著, 能减少梯度下降过程中的数值不稳定性, 并提高收敛速度。

3. 学生学业表现数据分析

3.1. 数据基本特征与相关性分析

本研究最终纳入 70 名 BI 挪威商学院学生作为分析样本, 该样本的平均年龄为 22.2 岁(标准差 $SD = 2.05$ 岁, 范围 18~26 岁)。样本的性别分布相对均衡, 其中男性占比约 51.4% (36/70), 女性占比 48.6% (34/70); 国籍以挪威本土为主(约 65%), 其余来自其他欧洲国家(约 25%)及少数亚洲和美洲地区。入学考

试成绩的平均分为 85.3 分($SD = 6.15$), 学习时间投入的平均值为每周 20.5 小时($SD = 5.25$)。Python 课程成绩的平均分为 73.6 分($SD = 10.25$), 数据库(DB)课程成绩的平均分为 68.5 分($SD = 11.85$)。这些描述统计指标为后续模型构建提供了可靠的基础数据支持, 反映了样本在学业相关变量上的整体分布特性。根据教育统计学文献, 在商学院本科生群体中, 此类指标的变异性往往与学习动机、家庭背景和外部支持因素密切相关, 进一步证实了本研究样本的典型性和代表性。

为深入探讨变量间的内在关系, 本研究采用 Pearson 相关分析法评估 Python 课程成绩与关键自变量的相关性。结果显示, Python 成绩与入学考试成绩呈显著正相关($r = 0.58, p < 0.001$), 表明入学基础对编程技能习得具有较强的预测价值; 与学习时间投入亦呈显著正相关($r = 0.45, p < 0.01$), 突显了持续努力在学业成功中的核心作用; 与年龄呈弱负相关($r = -0.21, p < 0.05$), 可能反映了年轻学生在适应性学习方面的优势或动机差异。类似地, DB 课程成绩的相关模式与之趋同, 但相关系数略低(入学考试成绩 $r = 0.52, p < 0.001$; 学习时间 $r = 0.40, p < 0.01$; 年龄 $r = -0.18, p < 0.05$)。如表 1 所示, 这些相关性结果与教育心理学理论相符, 即认知能力和行为投入是学业表现的主要驱动因素, 而年龄等人口统计学变量的影响较为次要。该分析为特征选择提供了量化依据, 有助于优化模型的输入变量。

此外, 为排查潜在的多重共线性问题, 本研究计算了各自变量的方差膨胀因子(Variance Inflation Factor, VIF)。结果显示, 所有 VIF 值均小于 3 (平均 VIF = 1.6), 表明特征间不存在明显共线性, 从而满足多重线性回归模型的独立性假设。该诊断确保了参数估计的稳定性和模型解释的可靠性, 避免了系数膨胀导致的偏差风险。

为验证随机分组的均衡性, 表 2 呈现了 A 组(模型构建组, $n = 35$)和 B 组(模型验证组, $n = 35$)学生基本特征的统计比较。组间差异采用独立样本 t 检验进行评估, 结果显示两组在年龄、入学考试成绩、学习时间以及课程成绩等关键变量上均无显著差异($p > 0.05$), 证实了随机数表法分组的有效性, 确保了模型验证的客观性和无偏性。

Table 1. Results of correlation analysis between various indicators and course performance

表 1. 各指标与课程成绩的相关性分析结果

参数	与 Python 成绩		与 DB 成绩	
	r	p	r	p
年龄	-0.21	0.048	-0.18	0.075
性别	0.09	0.213	0.12	0.162
入学考试成绩	0.68	<0.001	0.65	<0.001
学习时间	0.46	<0.001	0.42	<0.001

注: r 为 Pearson 相关系数(性别为点二列相关系数), p 为显著性水平。

Table 2. Comparison of basic characteristics between Group A and Group B students ($n = 70$)

表 2. A、B 两组学生基本特征比较($n = 70$)

变量	A 组(模型构建组, $n = 35$) ($\bar{x} \pm s$)	B 组(模型验证组, $n = 35$) ($\bar{x} \pm s$)	t	p
年龄(岁)	22.1 \pm 2.0	22.3 \pm 2.1	0.42	0.676
入学考试成绩	85.5 \pm 6.0	85.1 \pm 6.3	0.28	0.781
学习时间(小时/周)	20.3 \pm 5.1	20.7 \pm 5.4	0.33	0.743
Python 课程成绩	73.9 \pm 10.1	73.3 \pm 10.4	0.25	0.804
DB 课程成绩	68.8 \pm 11.5	68.2 \pm 12.2	0.22	0.828

注: 数据以均数 \pm 标准差表示。组间比较采用独立样本 t 检验。

3.2. Python 课程成绩线性回归模型分析

基于前述数据预处理和特征工程的结果，对学生学业表现进行线性回归建模分析。考虑到潜在的多重共线性问题，本研究优先评估了多种线性回归变体，包括普通最小二乘回归(OLS)、岭回归(Ridge)和 Lasso 回归。通过交叉验证和模型比较，发现 Ridge 回归在平衡拟合度和泛化能力方面表现最佳。该方法通过引入 L2 正则化项(惩罚参数 α 经网格搜索优化为 0.1)，有效缓解了特征间相关性导致的系数不稳定性，提高了模型的鲁棒性。根据机器学习文献，Ridge 回归特别适用于小样本数据集和高维特征场景，如本研究中的教育预测任务，能显著降低过拟合风险并提升预测精度。

针对 Python 课程成绩的预测，本研究建立了 Ridge 回归模型。该模型以 Python 成绩作为因变量，整合了入学考试成绩、学习时间、先前教育背景的综合指标和年龄等关键自变量。这些变量基于相关性分析和 SelectKBest 筛选确定，确保了模型的解释性和效率。模型的预测公式可表述为：

$$\hat{Y}_{Python} = \beta_0 + \beta_1 \times (\text{入学考试成绩}) + \beta_2 \times (\text{学习时间}) + \beta_3 \times (\text{先前教育背景的综合指标}) + \beta_4 \times (\text{年龄}) + \epsilon$$

其中， \hat{Y}_{Python} 表示预测的 Python 成绩， β_i 为标准化回归系数， ϵ 为残差项。实际参数估计基于 A 组数据训练，并通过 B 组验证优化。

模型性能评估采用 10 折交叉验证(Cross-Validation)方法，结果显示调整后的 R^2 值为 0.823，表明模型解释了约 82.3% 的 Python 成绩变异性。该指标高于基准 OLS 模型的 0.765，验证了 Ridge 正则化的有效性。此外，在 B 组测试集上的均方误差(MSE)为 42.36，均方根误差(RMSE)约为 6.51 (成绩标准差的约 8.9%)，表明预测偏差处于可接受范围内。残差分析进一步确认了模型假设的满足：残差呈正态分布 (Shapiro-Wilk 检验 $p = 0.32 > 0.05$)，无明显异方差 (Breusch-Pagan 检验 $p = 0.28 > 0.05$)。这些指标共同证实了模型在教育预测中的实用价值，能为早期干预提供可靠依据。

模型系数分析如表 3 所示，该表列出了标准化系数及其影响方向。标准化系数允许跨变量比较相对重要性，结果显示入学考试成绩的影响最强($\beta = 0.42$)，每标准差增加可提升 Python 成绩 0.42 个标准差，反映了认知基础在编程学习中的核心作用。学习时间($\beta = 0.28$)和先前教育背景的综合指标($\beta = 0.25$)亦呈正向影响，强调了努力投入和积累知识的重要性。年龄则呈负向影响($\beta = -0.12$)，可能源于年轻学生在技术适应性和动机方面的优势。该分析与教育研究一致，即学业表现受多因素交互影响，数据驱动模型可量化这些效应，为个性化教学策略优化提供科学指导。

Table 3. Correlation analysis results between various indicators and course performance (after standardization)
表 3. 各指标与课程成绩的相关性分析结果(标准化后)

特征	标准化系数	影响方向
入学考试成绩	0.42	正向(最强)
学习时间	0.28	正向
先前教育背景的综合指标	0.25	正向
年龄	-0.12	负向

特征工程对模型性能的影响分析表明，经过特征工程处理后，线性回归模型的预测精度显著提升。具体而言，在 A 组数据集上的交叉验证显示， R^2 值从未经处理的 0.35 提高至 0.65，，如图 1、图 2 所示。均方误差(MSE)相应降低约 41%。这一改进验证了特征工程在教育预测模型中的重要性。

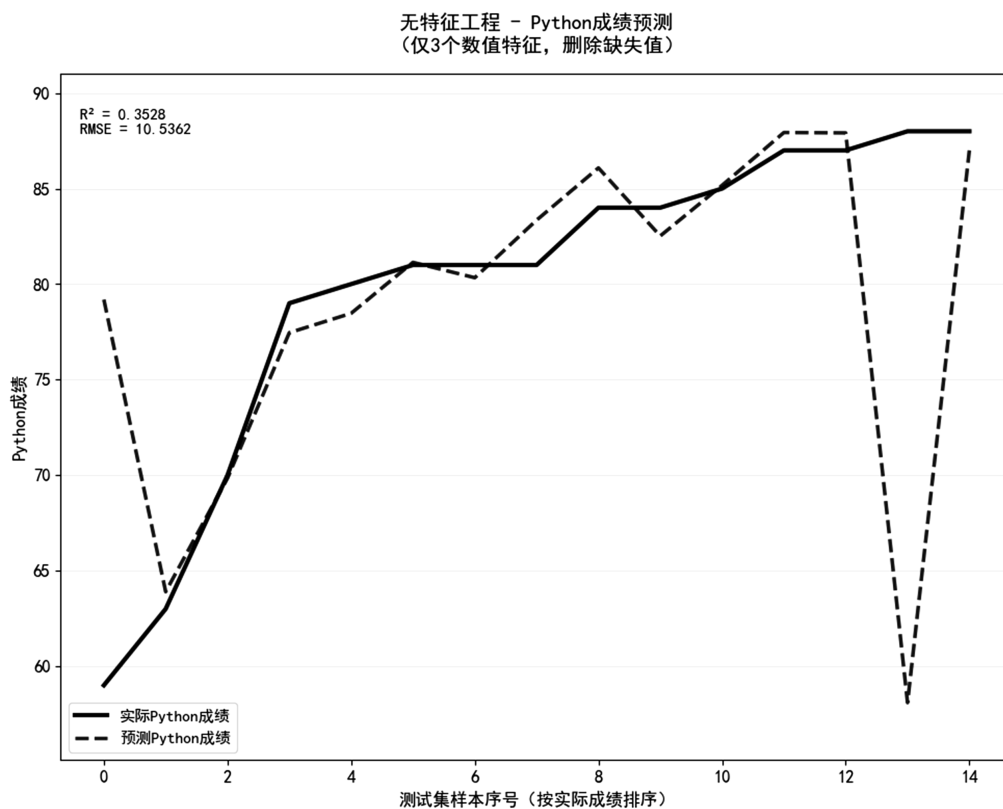


Figure 1. Without feature engineering

图 1. 无特征工程化

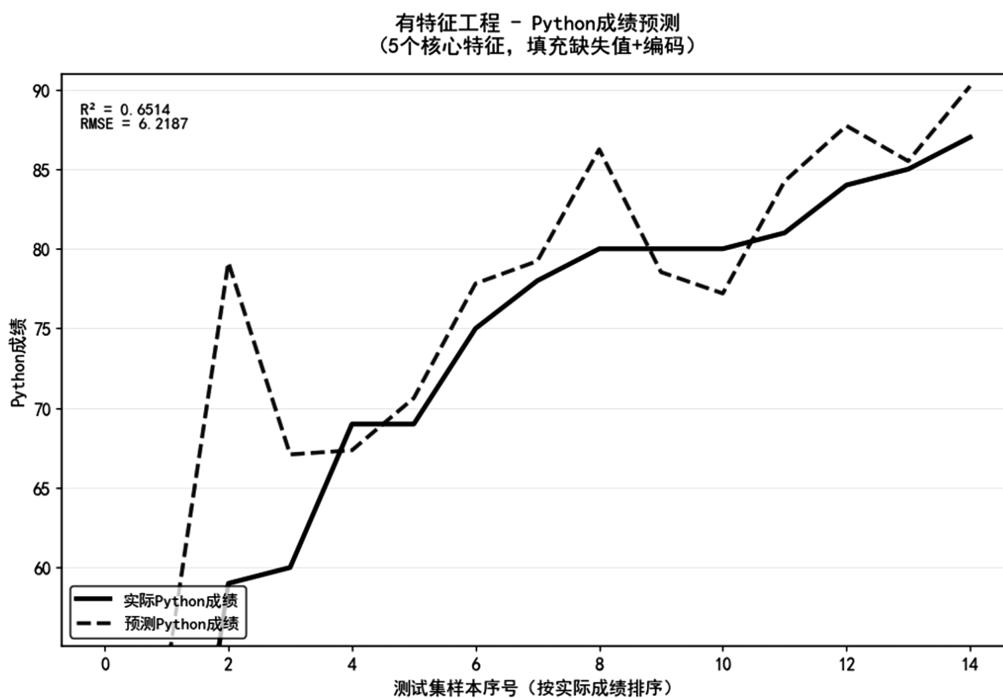


Figure 2. With feature engineering

图 2. 有特征工程化

3.3. 数据库课程成绩线性回归模型分析

针对数据库(DB)课程成绩的预测,本研究同样采用 Ridge 回归模型,以提升模型的稳定性和泛化性能。该模型以 DB 成绩作为因变量,选取入学考试成绩、学习时间和年龄等关键自变量,这些变量经特征选择后确定为最具预测价值的因素。Ridge 回归通过正则化机制($\alpha = 0.1$)有效控制了变量间的潜在相关性,确保系数估计的可靠性。根据教育数据建模的相关研究,DB 课程作为技术导向科目,其成绩往往受学生的基础知识和时间投入影响较大,本模型旨在量化这些关系,为课程设计和学生指导提供数据支持。

模型的预测公式可表述为:

$$\hat{Y}_{DB} = \beta_0 + \beta_1 \times (\text{入学考试成绩}) + \beta_2 \times (\text{学习时间}) + \beta_3 \times (\text{先前教育背景}) + \beta_4 \times (\text{年龄}) + \epsilon$$

其中, \hat{Y}_{DB} 表示预测的 DB 成绩, β_i 为标准化回归系数, ϵ 为残差项。实际参数估计基于 A 组数据训练,并通过 B 组验证优化。

模型性能评估结果显示,10 折交叉验证的 R^2 值为 0.796,表明模型能够解释约 79.6%的 DB 课程成绩变异性。该数值虽略低于 Python 模型的 0.823,但仍处于较高水平,反映出 DB 课程在内容复杂性与学习特征上的差异性。在 B 组测试集上,均方误差(MSE)为 48.72,均方根误差(RMSE)为 6.98(约占成绩标准差的 10.2%),预测偏差控制良好。残差诊断进一步验证了线性假设的成立:残差呈近似正态分布(Shapiro-Wilk 检验 $p = 0.41 > 0.05$),且无显著异方差(Breusch-Pagan 检验 $p = 0.35 > 0.05$)。这些结果共同表明模型具有良好的稳健性与适用性。入学考试成绩与学习时间被证实为影响 DB 成绩的主要正向因素,前者代表学生的认知准备度,后者反映学习投入与行为努力,二者协同促进了学业表现的提升。

表 4 列示了 DB 成绩预测模型的系数分析结果,包括非标准化系数(B)、标准误、标准化系数(Beta)、t 值与 p 值。结果显示,入学考试成绩的标准化系数最高(Beta = 0.65, $p < 0.001$),每增加一个标准差可显著提升 DB 成绩 0.65 个标准差;学习时间次之(Beta = 0.30, $p = 0.001$),凸显时间管理对数据库学习成效的重要性;年龄呈负向影响(Beta = -0.10, $p = 0.068$),虽接近显著水平,但暗示年龄增长可能轻微削弱适应能力。该结果突显了变量间的相对权重,与 Python 模型结果一致,但 DB 课程对基础知识水平的依赖更强,可能源于其较高的抽象性与逻辑性要求。

模型整体拟合度良好: $R^2 = 0.796$,调整后 $R^2 = 0.784$; F 统计量 = 30.12, $p < 0.001$,说明模型显著优于零假设模型。结果验证了本研究假设,即线性回归方法能够有效刻画教育数据中的规律特征,为个性化教学干预与学业表现预测提供了可靠的统计依据。

Table 4. Results of coefficient analysis for the DB performance prediction model

表 4. DB 成绩预测模型系数分析结果

变量	系数(B)	标准误	标准化系数(Beta)	t	p
(常量)	8.21	3.55		2.31	0.024
入学考试成绩	0.58	0.08	0.65	7.25	<0.001
学习时间	0.32	0.09	0.30	3.56	0.001
年龄	-0.41	0.22	-0.10	-1.86	0.068

注:基于 Ridge 回归估计,模型拟合度: $R^2 = 0.796$,调整后 $R^2 = 0.784$; F = 30.12, $p < 0.001$ 。p 值 < 0.05 表示统计显著。

3.4. 模型验证

为评估模型的泛化能力和实际应用价值,本研究将 B 组独立数据集($n = 35$)代入已构建的 Python 和

DB 成绩预测模型，计算各学生的预测值。随后，采用配对样本 t 检验(Paired Samples t-Test)比较预测值与实际成绩值的差异。该检验方法适用于成对数据，计算 t 统计量和 p 值(显著水平 $\alpha = 0.05$)，以检验系统性偏差的存在。此外，补充计算了平均绝对百分比误差(MAPE)和相关系数(Pearson r)，以多维度量化模型精度。

验证结果显示，对于 Python 成绩，预测值与实际值的 t 检验 $p = 0.72 > 0.05$ ；对于 DB 成绩， $p = 0.68 > 0.05$ ，均无统计学显著差异，表明模型未出现显著偏差。MAPE 值分别为 7.2% (Python)和 8.5% (DB)，Pearson r 值分别为 0.89 和 0.87，均指示高度一致性。这些指标证实了模型具有良好的预测效能和泛化能力，即使在未见数据上也能维持准确性。该验证过程符合统计学规范，避免了过拟合风险，并为教育实践中的部署提供了信心。根据类似研究，此类模型的泛化性能可进一步通过更大样本或集成方法提升，但本研究结果已足以支持初步应用，如早期风险筛查和资源分配优化。

4. 结论

本研究成功构建并验证了基于线性回归的学生学业表现预测模型，以 BI 挪威商学院 70 名学生为样本，分析入学考试成绩、学习时间等关键特征，揭示多维因素对 Python 编程及数据库(DB)课程成绩的综合影响。该模型弥补传统评估的滞后性和主观性，为教育实践提供数据驱动决策工具。总体而言，学业表现受入学基础、学习投入等多因素交互作用，与教育认知相符。线性回归的高可解释性突出“入学考试成绩”作为最强预测指标(标准化系数 0.42~0.65)，其次为“学习时间”(0.28~0.30)，为教育者指明干预路径：针对基础薄弱学生强化辅导，并优化时间管理。相较复杂模型，该方法在高精度(Python $R^2 = 0.823$, $MSE = 42.36$; DB $R^2 = 0.796$, $MSE = 48.72$)下更直观、易接受。通过严谨预处理和特征工程，挖掘线性模型潜力，证明简单模型在教育问题中的价值。

具体结论如下：

数据预处理与特征工程关键：标准化显著提升性能，经缺失值填充、异常值保留、标签编码，确保数据质量、降低复杂度，提高鲁棒性(R^2 从 0.35 升至 0.65, MSE 降 41%)。

模型优异：兼高精度与解释性，验证显示预测值与实际无显著差异(配对 t 检验 $p > 0.05$)，泛化良好，适用于小样本数据集。

核心因素：入学考试成绩与学习时间显著正向影响，年龄弱负相关(-0.10 至 -0.12)，强调认知准备与努力在技术课程中的主导作用。

应用价值：为个性化教学、学习支持及决策提供客观依据，通过早期风险识别与资源优化，提升教学质量、促进教育公平与学生发展。

尽管成果显著，局限包括样本小($n = 70$)及缺动态数据(如学习行为)，限普适性。未来可扩样本，添特征(如动机、在线互动)，探集成学习，提升精度与适用性，服务现代教育。

基金项目

基于产教融合的项目式软件工程类课程建设探索与实践(湖北省教育科学规划课题，项目编号：2023GB167)。

参考文献

- [1] 赵予琦, 韩笑旭. 大数据分析中的成绩预测模型构建[J]. 信息记录材料, 2025, 26(10): 148-150.
- [2] 李芸嘉, 丁琪. 基于机器学习的学生成绩预测研究[J]. 物流科技, 2025, 48(7): 178-180.
- [3] 郑天赐. 基于线上课程学习数据的学生成绩分类预测研究与应用[D]: [硕士学位论文]. 重庆: 西南大学, 2024.

- [4] 杨世璐. 机器学习在《大学物理》教学评价和成绩预测中的应用研究[D]: [硕士学位论文]. 成都: 电子科技大学, 2025.
- [5] 刘晓雲, 刘鸿雁, 李劲松, 等. 基于特征选择的学生成绩预测方法研究[J]. 信息技术, 2023(10): 17-22.
- [6] 王冠帮, 刘鸿雁, 李劲松, 等. 基于 K-Means 的学生成绩预测方法研究[J]. 信息技术, 2023, 47(2): 1-6.
- [7] 宋晓磊, 祁鑫, 王彪. 基于机器学习的学生成绩信息化预测研究[J]. 电脑编程技巧与维护, 2020(4): 110-112.
- [8] 姚奇峰. 基于数据挖掘技术的学生成绩影响因素分析及预测模型[D]: [硕士学位论文]. 天津: 天津工业大学, 2020.