Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2026, 15(2), 445-454
Published Online March 2026 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2026.152043

Hans汉斯

学生成绩等级预测模型研究

刘　洋*，杨博豪，乌　伟

西京学院计算机学院，陕西 西安

收稿日期：2025年12月25日；录用日期：2026年2月24日；发布日期：2026年3月9日

摘　要

学生成绩等级的精准预测是优化学生管理、提升教学指导效能的关键支撑，本研究以Kalboard 360数据集为基础，聚焦学生在线学习平台产生的过程性数据——既涵盖学历背景等基础特征，也包含课堂举手、学习资源访问频次等行为特征(此类数据能够有效反映学生的知识掌握水平)，借助PySpark生态工具链开展建模工作：首先通过pyspark.sql库将原始数据转换为DataFrame格式，并完成数据编码等预处理；随后基于pyspark.ml库中的分类算法，分别构建逻辑回归与随机森林两类学生成绩等级预测模型，经混淆矩阵、ROC曲线等性能指标验证，随机森林模型的预测精度显著优于逻辑回归模型，而该研究的核心价值在于，模型输出结果可支撑面向学生的个性化学习建议制定，同时帮助教师及时识别学生的学习难点与问题，进而实施针对性的教学调整与指导，其应用也有助于推动教育领域智能化、个性化教学模式的落地，最终助力学生的全面发展。

关键词

学生成绩等级预测模型，Pyspark库，分类算法

Research on Student Achievement Level Prediction Model

Yang Liu*, Bohao Yang, Wei Wu

School of Computer Science and Technology, Xijing University, Xi'an Shaanxi

Received: December 25, 2025; accepted: February 24, 2026; published: March 9, 2026

Abstract

Accurate prediction of students' academic performance grades serves as a pivotal underpinning for optimizing student management and enhancing the efficacy of teaching guidance. Based on the Kalboard 360 dataset, this study focuses on the process-oriented data generated by students on online

*通讯作者。

文章引用: 刘洋, 杨博豪, 乌伟. 学生成绩等级预测模型研究[J]. 人工智能与机器人研究, 2026, 15(2): 445-454.
DOI: 10.12677/airr.2026.152043

learning platforms, including both basic attributes such as educational background and behavioral features like in-class hand-raising and learning resource access frequency. Such data can effectively reflect students' level of knowledge mastery. This research leverages the PySpark ecosystem toolkit for model construction. First, the pyspark.sql library is used to convert raw data into DataFrame format and complete preprocessing procedures such as data encoding. Subsequently, based on the classification algorithms in the pyspark.ml library, two types of student academic performance grade prediction models (logistic regression and random forest) are constructed respectively. Verified by performance metrics including confusion matrices and ROC curves, the random forest model demonstrates significantly higher prediction accuracy than the logistic regression model. The core value of this study lies in the fact that the model output can support the formulation of personalized learning recommendations for students, while also helping teachers promptly identify students' learning difficulties and problems, thereby implementing targeted teaching adjustments and guidance. The application of this model is conducive to promoting the implementation of intelligent and personalized teaching models in the field of education, ultimately facilitating the all-round development of students.

Keywords

Student Achievement Grade Prediction Model, Pyspark Library, Classification Algorithms

1. 绪论

1.1. 选题背景

学生成绩数据的分析与挖掘工作，对优化教学管理体系具有重要的实践价值。基于学生学习全流程积淀的多元数据开展成绩预测，不仅能够督促学生落实常态化学习任务，更有助于教师及时识别并解决教学过程中存在的各类问题。传统成绩预测方法往往将学生的教育背景、课堂表现、出勤情况以及课外作业完成质量作为核心参考维度，然而在在线教育普及的背景下，学生依托在线学习平台产生的学习轨迹、答题行为特征等过程性数据，同样蕴含着能够反映其知识掌握水平的关键信息。

随着教育领域的持续发展与革新，实现学生成绩的精准预测愈发凸显其重要性。精准的成绩预测结果，可辅助教师深度掌握学生的学习状态，进而制定差异化教学策略，切实提升教学质量。与此同时，该预测结果也能为学生与家长提供有效参考，助力二者科学规划后续学习路径与职业发展方向。由此可见，学生成绩等级预测模型的研究具备极高的理论与应用价值。

1.2. 国内外研究与发展现状

近年来，学生成绩等级预测[1]模型成为教育领域与数据挖掘交叉研究的热点，国内外学者围绕多元技术路径与应用场景开展了大量深入探索。国内研究多聚焦传统统计方法、深度学习与集成学习技术的实践应用，以学生历史数据为核心，为教学管理优化提供支撑。马玉玲在山东大学的博士研究中，基于高校学生课程成绩、出勤记录等历史数据，对比逻辑回归、支持向量机等算法，构建了适配高校场景的成绩预测优化模型，可提前识别学业风险学生，弥补传统学业预警的滞后性。马丹在吉林大学的研究采用传统统计分析与决策树算法结合[2]的方式，挖掘学生考试成绩、课堂表现等多维度数据，明确影响学业表现的关键变量，其开发的分析系统能自动生成报告，为教师调整教学提供数据参考。深度学习领域，

林梦楠、李金辉提出基于自适应差分进化的神经网络模型，以学习轨迹数据为输入，预测准确率较传统BP神经网络提升约5% [3]；陆鑫赟、王兴芬构建的双隐层BP神经网络模型，虽聚焦创新能力预测，但多因素协同的设计思路为成绩预测模型[4]融入非学业特征提供了借鉴。

国外研究更注重多元数据融合，广泛纳入学习行为、社交网络、情感状态等信息，采用多种算法构建精准预测模型。Amoo M. Adewale等学者以尼日利亚中学生的课堂参与度、家庭背景等数据为研究对象[5]，通过逻辑回归筛选关键特征后，结合人工神经网络构建模型，准确率达82%，为识别学业困难学生、优化资源分配提供依据。部分国外学者采用AdaBoost算法集成多个弱分类器，以学业历史数据、社交网络互动数据为特征构建多因素融合模型，对成绩极端群体识别准确率超85%，为高校制定差异化教育政策提供支撑。此外，NGUYEN K T [6]等学者将情绪智力、学习环境等因素纳入预测框架，结合传统学业数据构建模型，发现情绪智力对成绩的影响贡献率达18%，推动模型向多维度、人性化方向发展。这些国内外研究均围绕“数据驱动–算法建模–决策支持”核心逻辑，显著提升了成绩预测精准度，推动教育体系从“经验驱动”向“数据驱动”转型，为个性化教学、资源优化配置提供重要支撑，持续助力教育质量提升。

2. 相关理论知识

2.1. Pyspark库

在信息时代，数据极为宝贵，Python和Pyspark成为数据处理和分析的利器。Pyspark.ml提供机器学习工具和算法，支持模型训练和参数调优，适用于大规模数据集。Pyspark.sql以DataFrame为数据结构，支持SQL查询和分布式数据处理。DataFrame将数据转换为适合Spark处理的格式，优化了大数据分析。Pyspark是数据科学家和Spark工程师的重要技能。

2.2. 分类算法介绍

2.2.1. 逻辑回归算法

逻辑回归(Logistic Regression)算法是一类专门用于解决二分类任务的统计学习方法。尽管其名称中带有“回归”字样，但该算法本质上属于分类算法范畴，而非回归算法。逻辑回归的核心原理是，将线性回归模型的输出结果(即对数几率，也可称为对数–几率函数)代入逻辑函数(又称sigmoid函数)中，把连续的预测值映射至0到1的区间内，以此实现二分类的目标。该算法的数学表达式详见公式(2-1)。

$$P\left(Y=1\middle|X\right)=\frac{1}{1+e^{\left(\beta 0+\beta 1X1+\beta 2X2+\cdots+\beta nXn\right)}} \tag{2-1}$$

其中，$P\left(Y=1\middle|X\right)$是给定输入X条件下观测值属于类别1的概率，e是自然对数的底；$\beta 0$，$\beta 1$，…，βn是模型参数，需要通过训练数据来学习。

2.2.2. 混淆矩阵

混淆矩阵是评估分类模型性能的表格，包含真阳性、真阴性、假阳性和假阴性四个关键值，用于计算性能指标，其结构如表1所示。

Table 1. Confusion matrix
表1. 混淆矩阵

	实际 Positive	实际 Negative
预测 Positive	True Positives	False Positives
预测 Negative	False Negatives	True Negatives

混淆矩阵及其计算得到的性能指标，可为分类模型性能评估提供全面信息，尤其适用于二分类问题。面对多类别问题时，混淆矩阵的结构会相应扩展，以适配各类别的组合情况。

2.2.3. ROC 曲线

ROC 曲线是一类用于评估二分类模型性能的可视化分析工具，其核心作用在于直观呈现真正例率(True Positive Rate，亦称为敏感性)与假正例率(False Positive Rate)二者之间的关联关系。该曲线以假正例率(False Positive Rate, FPR)为横轴，以真正例率(True Positive Rate，TPR，又名敏感性或召回率)为纵轴，两个指标的具体计算公式分别对应公式(2-3)与公式(2-4)。

$$\mathrm{FPR} = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}} \tag{2-3}$$

$$\mathrm{TPR} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{2-4}$$

3. 数据探索与预处理

3.1. 数据探索

本研究的实验数据来源于“Kalboard 360”学习管理系统，该数据集总计收录 480 条学生相关记录，数据采集周期覆盖两个学期：其中第一学期采集 245 条学生记录，第二学期采集 235 条学生记录。该数据集包含 17 个维度的字段信息，并依据总成绩将学生划分为三个等级，具体为：低 0~69、中：70~89、高：90~100。各个数据项的含义如表 2 所示。

Table 2. Explanation of data item meanings
表 2. 数据项含义说明

数据项	数据项含义
gender	学生性别(Male or Female)
NationallTy	学生国籍
Place of Birth	学生的出生地
StageID	受教育水平(lowerlevel) (MiddleSchool, HighSchool)
GradeID	年级(G-01, G-02, G-03, G-04, G-05, G-06, G-07, G-08, G-09, G-10, G-11, G-12)
SectionID	隶属的教室(A, B, C)
Topic	课程名
Semester	学校的学期 (First, Second)
Relation	监护学生的家长(mom, father)
Ralsedhands	学生在教室中举手次数(0~100)
VIsITedResources	学生访问在线课程次数(0~100)
AnnouncementsVIew	学生检查新公告的次数(0~100)
Discussion	学生参加讨论组的次数(0~100)
ParentAnsweringSurvey	家长是否回答了学校提供的调查问卷(Yes, No)
ParentschoolSatisfaction	家长对学校的满意度(Yes, No)
StudentAbsenceDays	每个学生的缺勤天数(above-7, under-7)
Class	根据学生的总成绩分为三个等级(低分：0~69，中等分数：70~89，高分：90~100)

xAPI-Edu-Data.csv 数据集共计包含 480 条样本记录，每条记录对应 17 项特征信息。本研究通过 Pandas 库中的 read_csv()函数完成数据集的读取与初步检视。该数据集涵盖 480 行记录与 17 个字段，为满足数据挖掘模型的输入规范，需逐一核查各字段的数据类型并开展深度分析。

3.2. 数据预处理

3.2.1. 数据转换

DataFrame 具备分布式特性，这一特性使其能够充分发挥 Spark 的并行计算优势，大幅提升数据处理任务的执行效率。将原始数据转换为 DataFrame 格式后，可实现更灵活、高效的数据处理与分析操作，为后续特征工程、模型训练等核心环节提供了便捷的数据基础。该数据格式转换过程是构建大规模数据处理与数据挖掘应用的关键步骤，不仅为数据科学家和分析师提供了高效的分析工具，还显著降低了大数据环境下复杂数据操作的实施难度。本研究以 Kalboard 360 数据集为研究对象，完成了数据加载及向 DataFrame 格式的转换工作。

3.2.2. 数值编码

数据集预处理需将字符型数据转换为数值型，以满足模型输入需求。通过数值编码，字符型特征得以有效利用，为特征工程和建模打下基础。无序分类特征需进行独热编码，将其转换为二进制向量，便于模型处理。

3.2.3. 特征整合

为提升特征处理效率并为后续模型训练奠定基础，本研究采用 VectorAssembler 方法整合全部特征，将其归并为一个独立的特征列。该特征整合方式能够将输入特征以更紧凑、结构化的形式呈现，便于模型对特征进行解析与处理。通过将分散的各特征列合并为单一向量列，不仅简化了特征管理流程，还能为模型提供更高效的输入数据格式。

4. 模型构建与训练

4.1. 数据集划分

为实现数据集的科学划分，本研究采用 randomSplit()方法，将完成编码预处理的数据集按照 7:3 的比例切分为训练集与测试集。其中参数[0.7, 0.3]对应训练集与测试集的划分占比，设置 seed = 1 旨在保证数据集划分结果的可复现性。该步骤的核心目的在于：构建学生成绩等级预测模型时，利用训练集完成模型的拟合训练，同时依托测试集开展模型性能验证。此种划分策略能够保障训练集与测试集均具备数据代表性，有效提升模型在未知数据上的泛化能力。

4.2. 逻辑回归模型

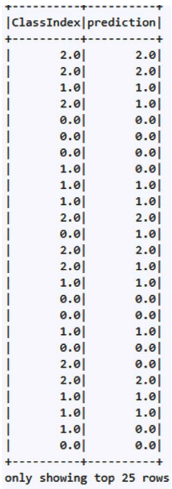
4.2.1. 逻辑回归模型构建

在逻辑回归模型的构建阶段，本研究基于 pyspark 库中的 LogisticRegression 类[7]实例化逻辑回归模型，并设定核心参数：最大迭代次数为 10，正则化系数取值 0.1，将“ClassIndex”字段指定为目标变量。逻辑回归作为经典的二分类算法，其核心原理是通过学习训练数据中特征与标签的关联关系，实现对目标变量两个类别归属的预测。本研究构建该模型的核心目标，是使其能够在学生成绩等级预测任务中完成高效的特征学习与类别划分。

4.2.2. 模型训练及预测

本研究调用 fit()方法对逻辑回归模型(lrmodel) [8]开展训练，以训练集作为数据输入让模型完成特征

学习；随后借助 transform()方法，将训练完成的模型部署至测试集，以此生成测试集的预测结果。该流程的核心目的在于验证模型针对新数据的适配性，进而评估其泛化能力与预测准确率。模型训练与预测是构建学生成绩等级预测模型的核心环节，通过这一系列操作可实现对学生成绩等级的有效预测，具体预测结果详见图 1。



ClassIndex	prediction
2.0	2.0
2.0	2.0
1.0	1.0
2.0	1.0
0.0	0.0
0.0	0.0
0.0	0.0
1.0	0.0
1.0	1.0
1.0	1.0
2.0	2.0
0.0	1.0
2.0	2.0
2.0	1.0
1.0	1.0
0.0	0.0
0.0	0.0
1.0	1.0
0.0	0.0
2.0	0.0
2.0	2.0
1.0	1.0
1.0	1.0
1.0	0.0
0.0	0.0

only showing top 25 rows

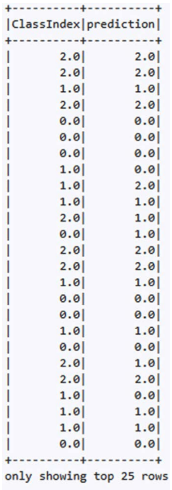
Figure 1. Test set prediction results
图 1. 测试集预测结果

4.3. 随机森林模型

4.3.1. 随机森林模型构建

本研究使用 pyspark 库中的 RandomForestClassifier 类[9]创建了一个随机森林模型。该模型的参数包括目标列为“ClassIndex”，特征列为“features”，以及设定了随机森林中决策树的个数为 10。

4.3.2. 模型训练及预测



ClassIndex	prediction
2.0	2.0
2.0	2.0
1.0	1.0
2.0	2.0
0.0	0.0
0.0	0.0
0.0	0.0
1.0	0.0
1.0	2.0
1.0	1.0
2.0	1.0
0.0	1.0
2.0	2.0
2.0	2.0
1.0	1.0
0.0	0.0
0.0	0.0
1.0	1.0
0.0	0.0
2.0	1.0
2.0	2.0
1.0	0.0
1.0	1.0
1.0	1.0
0.0	0.0

only showing top 25 rows

Figure 2. Test set prediction results
图 2. 测试集预测结果

本研究使用 fit()方法对训练集进行训练，得到了随机森林模型。接着，通过 transform()方法，将该模

型应用于测试集 test [10]，得到了对测试集的预测结果。这个过程是为了评估随机森林模型在新数据上的预测效果，以便确定模型的性能和泛化能力。

测试集预测结果如图 2 所示。

5. 模型评估与优化

5.1. 逻辑回归模型

5.1.1. 性能评估

使用 MulticlassClassificationEvaluator 进行模型评估。本次模型评估设定的核心参数为：以“prediction” [11]列为预测结果列，以“ClassIndex”列为目标变量列。评估过程选取准确率作为核心性能指标，通过调用 evaluate()方法测算逻辑回归模型在测试集上的预测准确率。经计算可得，该模型在测试集上的准确率达到 0.724，这一数值表明模型能够对 72.4%的学生成绩等级实现准确预测。

逻辑回归模型的混淆矩阵[12]与 ROC 曲线分析结果如图 3 所示。图 3 呈现的混淆矩阵是分类模型性能评估的常用工具，其核心作用在于清晰呈现模型预测类别与样本真实标签的对应关系。本研究中的混淆矩阵涵盖 0、1、2 三个类别，分别对应学生成绩等级的高、中、低三个预测层级。

类别 1 识别效果佳，正确预测 49 个，错误分类 15 个。类别 0 和类别 2 识别较差，类别 2 易误分为类别 1。模型在区分 0 和 1 类别上优于区分 0 和 2、1 和 2。混淆矩阵揭示了模型性能和易混淆类别。ROC 曲线图 4 显示，类别 0 的 ROC 面积 0.93 表现最佳，类别 1 为 0.76 需改进，类别 2 为 0.87 表现次之。模型整体表现良好，尤其在类别 0 和 2 上。

5.1.2. 参数调优

利用 pyspark 的 tune 包和 GridSearch 对逻辑回归进行参数优化，调整 maxIter 和 Elastic net。通过交叉验证评估，测试准确率从 0.724 增至 0.731，优化了学生成绩预测。

调优后逻辑回归混淆矩阵与 ROC 曲线如图 3 与图 4 所示。

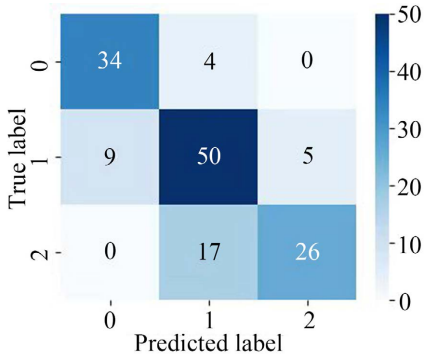


Figure 3. Confusion matrix of logistic regression after tuning
图 3. 调优后逻辑回归混淆矩阵

图 3 展示了调优后的混淆矩阵，包含三个成绩等级类别。类别 0 正确预测 33 个，5 个误判为类别 1。类别 1 正确预测 46 个，8 个误判为类别 0，10 个误判为类别 2。类别 2 正确预测 27 个，16 个误判为类别 1。类别 1 识别率高，类别 2 最不稳定，易误判为类别 1。整体上，模型对类别 1 预测最准，类别 2 最不准确。

图 4 的 ROC 曲线显示，类别 0 的识别最准确(AUC = 0.96)，类别 1 次之(AUC = 0.79)，类别 2 表现较高(AUC = 0.92)。Grid Search 调优后，模型整体表现良好，类别 0 最佳。

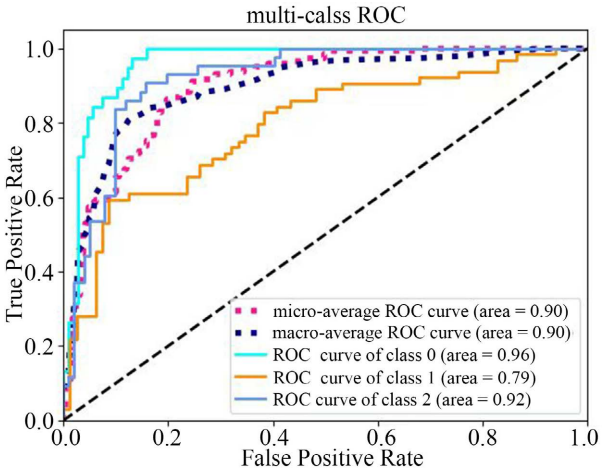


Figure 4. ROC curve of the tuned logistic regression
图 4. 调优后逻辑回归 ROC 曲线

通过调优前后对比模型准确率与混淆矩阵与 ROC 曲线结果可得逻辑回归在初始状态下的准确率为0.724，进行参数调优后，准确率提升至 0.731。

5.2. 随机森林模型

5.2.1. 性能评估

本研究选取准确率作为核心评价指标，对随机森林模型的性能进行系统性验证。实验环节中，通过调用 evaluate()方法定量计算模型在测试集上的准确率；该指标是分类模型性能评估的经典标准，其物理内涵为测试数据中模型正确预测样本数与总样本数的比值。输出数据显示，随机森林模型在测试集上的准确率达到 0.759，这一结果表明模型具有良好的预测效能。

模型混淆矩阵，显示类别 1 预测最准，50 个实例正确分类。类别 0 预测较准，但有误分类。类别 2 表现较差，多误分为类别 1。随机森林模型分类表现不均，需改进类别 2 预测。

模型 ROC 曲线显示：类别 0 区分度最高(AUC = 0.96)，类别 1 较低(AUC = 0.81)，类别 2 良好(AUC = 0.93)。类别 1 识别性能较弱，可能因数据不平衡或特征不明显。理想 ROC 曲线应接近左上角。

5.2.2. 参数调优

用 GridSearch 对随机森林模型进行超参调优[13]。调优了建树个数参数，通过交叉验证评估，测试集准确率从 0.759 提升至 0.79，优化了学生成绩预测。

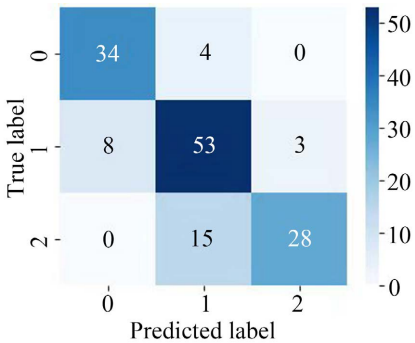


Figure 5. Confusion matrix of the optimized random forest
图 5. 优化后随机森林混淆矩阵

此外，超参数调优后随机森林模型对应的混淆矩阵、ROC 曲线，以及学生成绩等级预测结果，分别如图 5~7 所示。

图 5 显示调优后混淆矩阵，类别 2 识别率高，类别 0 次之，类别 1 最差。总体上，随机森林模型经 Grid Search 调优后，类别 1 预测最准，类别 2 最不稳。图 6 的 ROC 曲线显示类别 0 识别最佳(AUC = 0.97)，类别 2 次之(AUC = 0.96) [14]，类别 1 最弱(AUC = 0.86)。逻辑回归模型经 pyspark 调优后，整体表现提升，类别 0 最优。

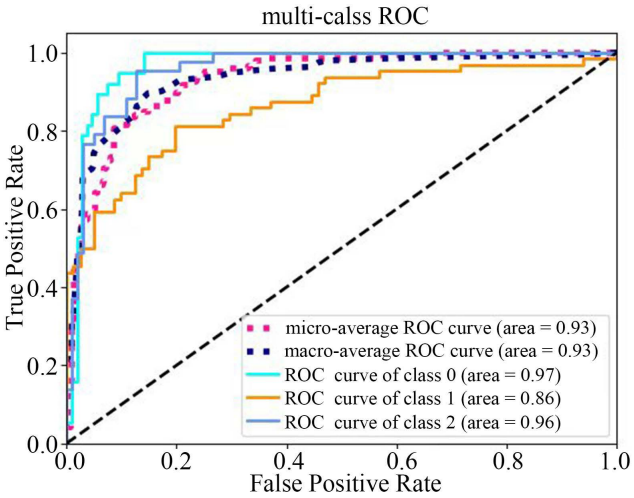


Figure 6. ROC curve of the tuned random forest
图 6. 调优后随机森林 ROC 曲线

综上所述，随机森林模型在未经过参数优化的初始状态下，预测准确率为 0.759。为进一步提升模型性能，本研究对其开展超参数调优以探寻最优参数组合[15]，调优完成后，该模型的预测准确率提升至 0.79。

ClassIndex	prediction
2.0	2.0
2.0	2.0
1.0	1.0
2.0	2.0
0.0	0.0
0.0	0.0
0.0	0.0
1.0	0.0
1.0	1.0
1.0	1.0
2.0	1.0
0.0	1.0
2.0	1.0
2.0	2.0
1.0	1.0
0.0	0.0
0.0	0.0
1.0	1.0
0.0	0.0
2.0	1.0
2.0	2.0
1.0	1.0
1.0	1.0
1.0	1.0
0.0	0.0

only showing top 25 rows

Figure 7. Prediction results of student grade levels
图 7. 学生成绩等级预测结果

此外，图 7 展示了随机选取的 25 名学生的成绩等级预测结果，其中以 2.0 代表低等级、1.0 代表中等级、0.0 代表高等级。基于该预测结果，教师可针对性地为不同层次的学生制定差异化教学方案。

6. 总结

本研究使用 Kalboard 360 数据集，构建逻辑回归和随机森林模型预测学生成绩等级，旨在为教育管理提供高效工具。模型基于在线学习过程性数据，通过 pyspark.ml 构建，并使用 DataFrame 进行数据处理。混淆矩阵和 ROC 曲线评估显示模型性能良好。为进一步提高预测精度，采用 pyspark 的 tune 包和 Grid Search 进行超参数调优，将准确率提升至 0.731 和 0.79，达到最佳预测状态。

参考文献

[1] 秦亚杰, 刘梦赤, 胡婕, 冯嘉美. 基于认知诊断与 XGBoost 的学生表现预测研究[J]. 华南师范大学学报(自然科学版), 2023, 55(1): 55-64.

[2] 马丹. 基于数据挖掘技术的学生成绩分析系统的设计与实现[D]: [硕士学位论文]. 长春: 吉林大学, 2015.

[3] 林梦楠, 李金辉. 基于自适应差分进化的学生成绩等级预测神经网络模型[J]. 现代电子技术, 2022, 45(3): 130-134.

[4] 陆鑫赟, 王兴芬. 双隐层 BP 神经网络大学生创新能力预估模型[J]. 中国科技论文, 2018, 13(8): 926-932.

[5] Amoo, M.A., Alaba, O.B. and Usman, O.L. (2018) Predictive Modelling and Analysis of Academic Performance of Secondary School Students: Artificial Neural Network Approach. *International Journal of Science and Technology Education Research*, **9**, 1-8. https://doi.org/10.5897/ijster2017.0415

[6] Nguyen, K.T., Duong, T.M., Tran, N.Y., *et al.* (2020) The Impact of Emotional Intelligence on Performance: A Closer Look at Individual and Environmental Factors. *The Journal of Asian Finance, Economics and Business*, **7**, 183-193.

[7] 孟卓, 袁梅宇. 教育数据挖掘发展现状及研究规律的分析[J]. 教育导刊, 2015(2): 29-33.

[8] 张燕南. 大数据的教育领域应用之研究[D]: [博士学位论文]. 上海: 华东师范大学, 2016.

[9] 高秀梅. 当代大学生学习动机的特征及其对学业成绩的影响[J]. 高教探索, 2020(1): 43-47.

[10] 马玉玲. 基于机器学习的高校学生成绩预测方法研究[D]: [博士学位论文]. 济南: 山东大学, 2020.

[11] 王艳晓. 基于流程性教育数据挖掘的学生成绩预测方法研究[D]: [硕士学位论文]. 青岛: 山东科技大学, 2018.

[12] Kazumali, E. and Kalinga, E. (2017) Neural Network Model for Predicting Students' Achievement in Blended Courses at the University of Dar Es Salaam. *International Journal of Artificial Intelligence & Applications*, **8**, 23-35. https://doi.org/10.5121/ijaia.2017.8203

[13] 张文奇, 王海瑞, 朱贵富. 基于因果推断和多头自注意力机制的学生成绩预测[J]. 现代电子技术, 2023, 46(17): 111-116.

[14] Barman, H., Dutta, M.K. and Nath, H.K. (2018) The Telecommunications Divide among Indian States. *Telecommunications Policy*, **42**, 530-551. https://doi.org/10.1016/j.telpol.2018.05.003

[15] Hu, L.Q. and Zhao, G. (2021) Research on Influencing Factors of Machine Learning Algorithm on Student Achievement Based on Data Mining. *Journal of Nanchang Hangkong University* (*Natural Science Edition*), **35**, 43-48, 97.