

# 基于Flask的员工离职数据可视分析系统设计与实现

陈志浩<sup>1</sup>, 李玉曼<sup>2</sup>, 何佳宇<sup>1</sup>, 马萨里沐<sup>1</sup>, 郑诗璇<sup>2</sup>, 孟繁华<sup>2</sup>

<sup>1</sup>北京信息科技大学计算机学院, 北京

<sup>2</sup>北京信息科技大学经济管理学院, 北京

收稿日期: 2023年8月23日; 录用日期: 2023年9月19日; 发布日期: 2023年9月26日

## 摘要

本论文介绍了一个基于可视化分析技术的员工离职数据可视分析平台的开发, 旨在预测员工离职并为企业提供科学直观的管理手段。通过使用随机森林模型进行员工离职预测, 并结合多种图表数据挖掘, 系统能够帮助企业管理者了解员工离职的原因和规律, 为人力资源管理提供决策支持。论文中详细介绍了随机森林模型的原理和优势, 并验证了其在员工离职预测上的高准确率。可视化分析系统的功能需求包括员工离职时间规律观察、员工离职空间规律观察、预测模型可视化和交互体验优化等。系统通过图表展示了员工离职与各个变量之间的关联程度, 包括时间规律、空间规律、员工满意度和绩效对离职的影响。此外, 用户可以通过提交当期人力资源管理数据进行模型训练和预测, 并通过邮件通知功能及时通知各部门领导有离职可能的员工。这个可视化分析平台为企业管理者提供了一个科学、直观且易用的数据分析工具, 帮助他们更好地了解员工流动问题, 提高企业的经济效益和运营稳定性。然而, 系统仍存在一些局限性, 未来的研究可以考虑引入其他更复杂的模型或算法来进一步提高预测准确性和泛化能力。

## 关键词

Flask, 员工离职预测, 可视分析, 随机森林

# Design and Implementation of Employee Turnover Data Visualization and Analysis System Based on Flask

Zhihao Chen<sup>1</sup>, Yuman Li<sup>2</sup>, Jiayu He<sup>1</sup>, Masari Mu<sup>1</sup>, Shixuan Zheng<sup>2</sup>, Fanhua Meng<sup>2</sup>

<sup>1</sup>School of Computer Science, Beijing Information Science and Technology University, Beijing

<sup>2</sup>School of Economics and Management, Beijing Information Science and Technology University, Beijing

Received: Aug. 23<sup>rd</sup>, 2023; accepted: Sep. 19<sup>th</sup>, 2023; published: Sep. 26<sup>th</sup>, 2023

文章引用: 陈志浩, 李玉曼, 何佳宇, 马萨里沐, 郑诗璇, 孟繁华. 基于 Flask 的员工离职数据可视分析系统设计与实现[J]. 计算机科学与应用, 2023, 13(9): 1765-1772. DOI: 10.12677/csa.2023.139175

## Abstract

This paper introduces the development of an employee turnover data visualization and analysis platform based on visual analysis technology, aimed at predicting employee turnover and providing scientific and intuitive management approaches for enterprises. By using the Random Forest model for employee turnover prediction and combining multiple chart data mining techniques, the system assists enterprise managers in understanding the reasons and patterns of employee turnover, thereby providing decision support for human resources management. The paper elaborates on the principles and advantages of the Random Forest model and verifies its high accuracy in employee turnover prediction. The functional requirements of the visualization and analysis system include observing the time and spatial patterns of employee turnover, visualizing the prediction model, and optimizing interactive experience. The system presents the correlations between employee turnover and various variables, including time patterns, spatial patterns, employee satisfaction, and performance impact on turnover, through graphical representation. Additionally, users can submit current human resources management data for model training and prediction, and timely notify department leaders of employees at risk of turnover through email notification. This visualization and analysis platform offers enterprise managers a scientific, intuitive, and user-friendly data analysis tool to better understand employee mobility issues and improve economic efficiency and operational stability. However, the system still has some limitations, and future research could consider introducing more complex models or algorithms to further enhance prediction accuracy and generalization capability.

## Keywords

Flask, Employee Turnover Prediction, Visual Analysis, Random Forest

Copyright © 2023 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

在现代企业中，员工离职已成为管理中的一个重要问题。员工离职不仅会带来人力资源的损失，还会对企业的经济效益和运营稳定性造成影响。因此，预测员工离职已成为企业管理中的重要课题之一。全球知名咨询公司麦肯锡曾提出“大数据”时代的到来，认为数据已经渗透到当今每一个行业和业务职能领域，成为重要的生产因素[1]。在这个背景下，基于可视化分析技术的员工离职数据可视分析平台的开发，充分利用了海量数据的挖掘和运用，为企业提供了科学直观的员工离职预测和管理手段。这不仅可以提高企业的经济效益和运营稳定性，还能为企业提供员工管理方面的意见和建议，帮助企业更好地管理员工流动问题。

随着科技的不断进步，数据采集技术和数据分析技术得到了广泛的应用，可视化分析技术作为一种新型的数据分析手段，以其直观性和易用性，逐渐成为数据分析的重要工具。基于可视化分析技术的员工离职数据可视分析平台，可以通过对历史员工离职数据的分析和挖掘，预测员工离职的可能性，为企业提供决策支持。

本文提出了一个基于 Flask 的员工离职数据可视分析平台，利用随机森林算法实现员工离职预测，并通过可视化分析技术展示各因素与员工离职间的关系。该平台不仅可以帮助企业预测员工离职，还能直

观反映出员工离职与哪些因素有关，为企业提供员工管理方面的意见和建议。该平台的开发基于数据分析、可视化理论和 Web 技术，是多学科多理论融合的数据分析方法的典型应用。

本文的主要贡献在于提供了一种基于可视化分析技术的员工离职预测方法和数据分析平台，为企业管理者提供了一种科学、直观和易用的数据分析手段。该平台的应用可以帮助企业更好地管理员工流动问题，提高企业的经济效益和运营稳定性。

## 2. 基于随机森林模型的员工离职预测

由于企业人力资源管理数据的保密性，本文使用 Kaggle 平台上的 HR\_comma\_sep.csv 数据集模拟企业中真实的人力资源管理数据。随机森林是一种集成学习方法，可以用于分类和回归。相比于单个决策树模型，随机森林具有准确性高、可处理大量特征、可处理缺失值和不平衡数据、可解释性强、适用于大规模数据集等优点。随机森林可以通过多个决策树模型的集成来减少过拟合，提高模型的准确性，同时可以提供特征重要性的评估，帮助我们理解哪些特征对模型的性能影响最大。随机森林还可以在包含缺失值和不平衡数据的情况下进行训练，不需要对数据进行额外的处理，可以在高维数据集上表现良好，并可以使用并行化算法，在大规模数据集上高效地进行训练和预测。将数据集中的 Salary、Sales 等字段进行编码，变量列表如表 1 所示：

Table 1. Variable list

表 1. 变量列表

数量	质量	排序	稿件	件数
1	3	4	7	8
2	4	3	5	9
3	7	6	7	8
2	4	3	5	9
2	4	3	5	9
2	4	3	5	9
2	4	3	5	9

### 2.1. 模型建立

本文首先对原始数据集中的缺失值进行填充、剔除、特征值提取、标准化等数据预处理工作，然后将员工满意度、上次评估得分、参与的项目数、平均每月工作小时数、在公司工作年限等作为模型的输入，预测未来某一时刻的员工离职情况并将其作为输出建立模型。

### 2.2. 模型准确率检验

本文使用预处理后的 HR\_comma\_sep.csv 作为数据集进行训练和测试，将样本数据的 70% 作为训练集，剩余 30% 作为测试集。在训练和测试前，对人力资源数据中的薪资水平变量、所在部门变量进行了编码处理。为了评估模型的性能，使用 ROC 曲线进行可视化展示。为了绘制 ROC 曲线，需要计算不同阈值下的真正率(True Positive Rate, TPR)和假正率(False Positive Rate, FPR)，然后使用 Python 中的 matplotlib 库的 plot 函数绘制 FPR 对应的横坐标值和 TPR 对应的纵坐标值，从而得到 ROC 曲线。该曲线能够有效地表示模型的分类性能。如图 1 所示。

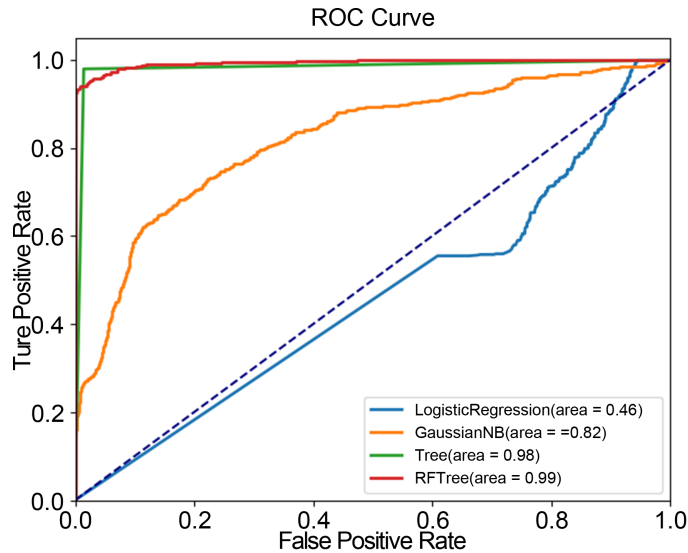


Figure 1. ROC curve  
图 1. ROC 曲线

ROC 曲线显示，模型性能越好，其曲线越靠近左上角，接近于 1 分类器。从图 2 的结果可以得知该模型准确率较高。本文采用了随机森林模型对未来员工离职情况进行预测，并通过使用 Python 来训练该模型后，在模型训练集上与原始数据进行对比，结果显示准确率高达 97.66%。这表明该随机森林模型对员工离职情况的预测效果表现良好。通过这个高准确率的结果，我们可以得出结论，该随机森林模型在员工离职预测上具有较高的精确性。这意味着我们可以有信心使用该模型来对未来员工离职的发生进行预测。这个发现对企业预防员工流失和人才储备具有重要意义。利用这种准确的预测模型，企业决策者和相关管理者可以更有效地实施员工管理和资源分配，以减少员工离职的发生，减少企业的资源损失。然而，在实际应用中，仍需谨慎对待模型的局限性。例如，随机森林模型在处理非线性关系方面表现优秀，但仍需综合考虑其他因素和可能的影响因素，以确保预测结果的全面性和准确性。

### 3. 基于 Flask 的员工离职数据可视化分析系统需求分析

随着员工离职数据的不断积累，公司保存了大量有关员工流动情况的数据信息。为了让用户能够轻松地通过 Web 可视系统获取这些历史员工离职数据中有价值的规律，同时注重用户与图表数据的交互体验，本文提出了以下可视分析技术的具体需求：

**员工离职时间规律观察：**员工离职可能受到多种因素影响，如公司业务状况、季节、年度周期等。因此，该系统需要通过高效的数据处理手段，获取员工离职时间上的信息数据，并以简洁清晰的可视化方式展示，使用户能够观察到员工离职随时间变化的趋势，是否存在周期性的概率特征以及离职与其他因素的关联。

**员工离职空间规律观察：**员工离职可能在公司的不同部门或地区发生集中，这些区域性特点对于了解员工流动的整体情况具有重要意义。因此，系统应该能够研究员工离职的重点区域和特定部门，将局部和整体结合起来进行研究，从而更加清晰直观地了解员工离职的空间模式。

**预测模型可视化：**系统需要展示基于随机森林模型的员工离职预测结果。用户应该能够通过点击按钮或交互操作，实时查看不同参数和特征对预测结果的影响，从而深入理解模型的性能和预测准确性。

**交互体验优化：**在整个数据可视化分析过程中，系统应注重用户与图表数据的交互体验。用户应能够灵活选择感兴趣的数据维度、时间范围或其他关键因素，以便针对性地探索员工离职的规律和趋势[2]。

## 4. 系统架构与模块设计

### 4.1. 系统架构设计

该系统采用 Browser/Server 架构，后端使用 Flask 框架搭建服务，并通过 mysql-connector-python 访问 MySQL 数据库，用于读取存储在公司人力资源管理系统的数据。前端利用 Bootstrap 框架结合 HTML5、CSS 和 JavaScript 开发用户界面和交互功能，并使用 PyEcharts 函数库实现数据可视化功能。随后将当期的人力资源管理数据加载到模型中，返回预测文件，通过将预测文件载入发送页面，及时通知各部门领导访谈有离职可能的员工。这款员工离职数据可视化分析系统可以为企业提供更直观的数据展示和深入的数据分析，为人力资源管理带来更多的洞察和决策支持，助力企业在竞争激烈的市场中取得更大的成功。

### 4.2. 系统架构设计

系统主要功能模块如图 2 所示。

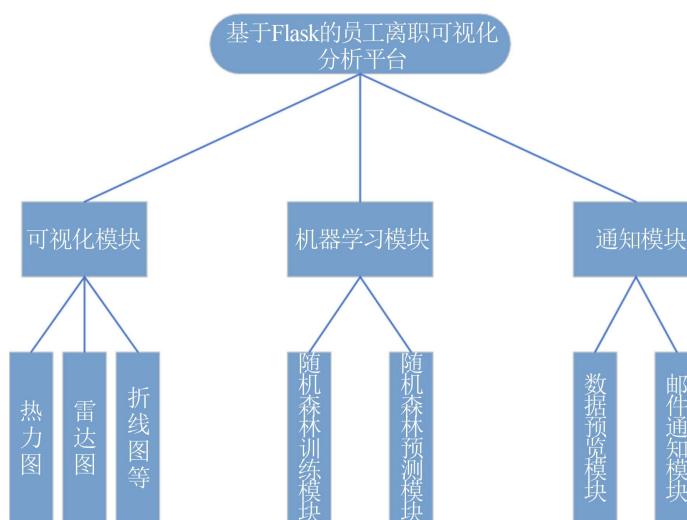


Figure 2. System function design  
图 2. 系统功能设计

如图 2 所，在首页的可视化模块中，用户可以获得各种类型的图表，以便直观地了解与员工离职相关的关键变量。以下是首页图表的示例：热力图、柱状图和折线图、分布图、箱线图等。机器学习模块，该模块允许用户提交过去的人力资源管理系统中的人事数据，用于训练随机森林模型。随机森林是一种强大的集成学习算法，它可以用于分类和回归任务。系统会提供模型的评估指标，例如准确率、召回率、F1 分数等，以帮助用户了解模型的性能。同时用户也可以选择提供当期的人力资源管理数据，让系统进行分析，最后导出预测结果。一旦用户获得随机森林的预测结果，他们可以将结果提交至通知模块，并选择需要通知的部门的邮箱。

## 5. 系统展示与案例分析

### 5.1. 系统界面展示

首页展示。首页包含热力图分析、未离职员工与离职员工项目数雷达图、未离职员工与离职员工各满意度条形图、未离职员工与离职员工各满意度折线图、已经离职员工绩效区间玫瑰图、未离职员工绩效区间饼状图、未离职员工平均月工作时间饼状图。如图 3 所示。



机器学习模块与通知模块。用户进入提交页面通过选择当期的人力资源管理资料进行上传，随后显示模型的预测结果。随后进入邮箱账号填写页面，随后点击发送邮件，即可实现邮件发送。

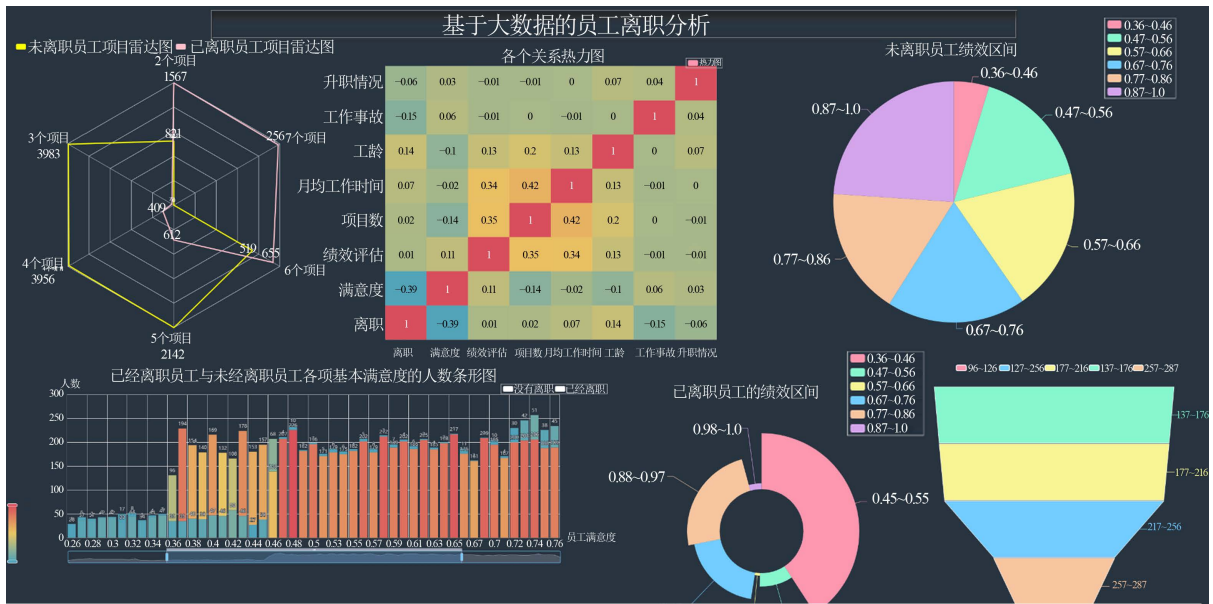


Figure 3. Homepage display  
图 3. 首页展示

### 5.2. 案例分析——图表数据挖掘

图 4 展示了各数据间关联程度的热力图，通过色彩的深浅和数字标识，直观地揭示了不同变量之间的相关性。从中我们可以得出一些重要的结论，为进一步的分析指明了明确的方向[3]。

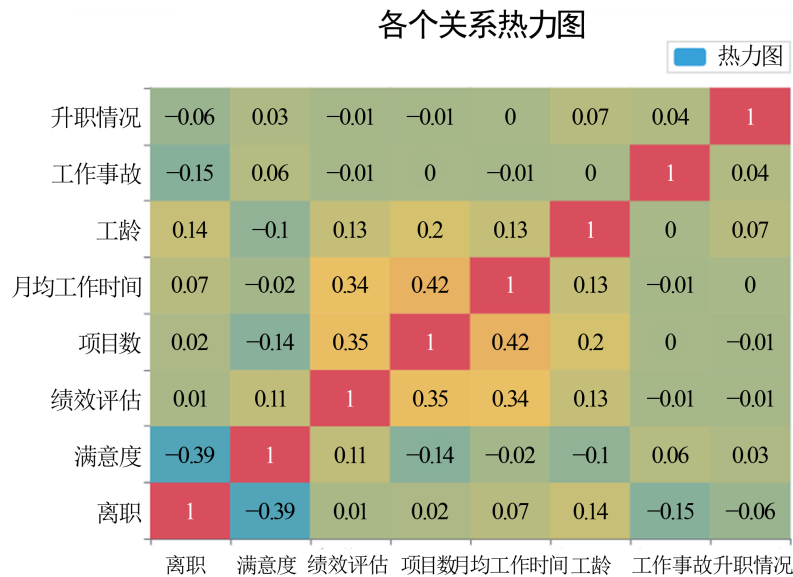


Figure 4. Heat map  
图 4. 热力图

首先，值得注意的“项目数”与“平均月工作时间”以及“绩效评估”之间存在一定程度的正相关

关系。这意味着员工参与更多项目的情况下，通常也倾向于有更长的月工作时间，并且通常表现出较好的绩效。这一发现可以引导关注项目数的管理与分配，以及确保员工在参与项目的同时保持合理的工作负荷，以提高绩效和减少离职率。

其次，我们还可以观察到“离职”与“满意度”之间存在一定程度的负相关关系，同时“离职”与“工作事故”也呈现出负相关。这意味着员工的离职倾向与他们的满意度程度以及工作事故之间存在联系。这为我们提供了一个有价值的见解，即员工满意度的提高可能有助于减少离职率，而减少工作事故可能也会对员工离职率产生积极的影响。

如图 5 所示，我们可以观察到离职和满意度之间的关系。在满意度低于 0.2 的员工中，离职人数达到了峰值，这表明这部分员工离开公司可能是因为对公司不满意。他们的低满意度促使他们寻找更好的工作机会，从而选择了离职。而在满意度在 0.4 左右和 0.8 左右的员工中，离职人数也达到了峰值。这暗示着这些员工对公司感到相对满意，但他们离职的原因可能是因为他们找到了更具吸引力的职业机会。

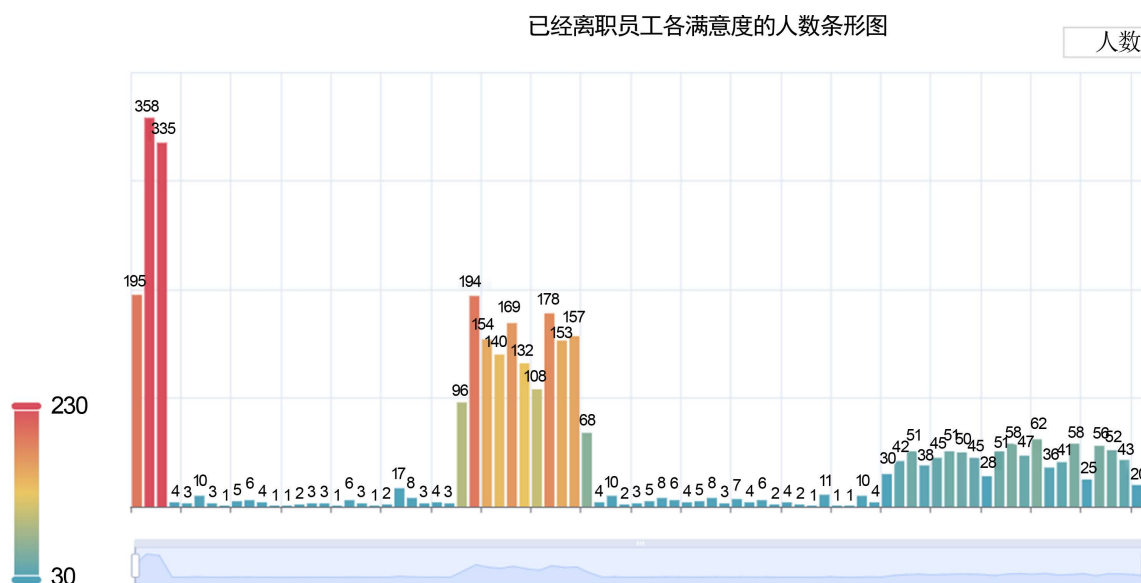


Figure 5. Bar chart of satisfaction levels for former employees  
图 5. 已经离职员工各满意度的人数条形图

## 6. 结语

本篇论文设计了一个基于 Flask 的员工离职数据可视分析平台，旨在通过数据分析和可视化技术来预测员工离职，并为企业提供科学直观的离职预测和管理手段。通过构建基于随机森林模型的离职预测，以及多种图表数据挖掘，系统能够帮助企业管理者更好地了解员工离职的原因和规律，从而为人力资源管理提供决策支持。

在论文中，我们深入介绍了随机森林模型的原理，并说明其在员工离职预测中的优势，通过对 HR\_comma\_sep.csv 数据集进行训练和测试，证明了该模型的高准确率，对员工离职预测有着较好的性能。

此外，我们提出了可视化分析系统的功能需求，通过数据处理和多种图表的展示，使用户可以直观地了解员工离职与各个变量之间的关联程度，包括时间规律、空间规律以及员工满意度和绩效对离职的影响。同时，用户还可以通过提交当期人力资源管理数据，进行模型训练和预测，并通过邮件通知功能，及时通知各部门领导有离职可能的员工，从而采取相应的人力资源管理措施。

通过对系统功能模块的设计，我们构建了一款可视化分析平台，为企业管理者提供了一个科学、直

观且易用的数据分析工具，帮助他们更好地了解员工流动问题，提高企业的经济效益和运营稳定性[4]。然而，系统仍存在一些局限性。例如，随机森林模型虽然性能优秀，但在处理高维数据和大规模数据时仍可能面临一些挑战。因此，未来的研究可以考虑引入其他更加复杂的模型或算法来进一步提高预测准确性和泛化能力[5]。

## 基金项目

由北京信息科技大学大学生创新创业训练计划项目——经济管理学院支持。

## 参考文献

- [1] 李一苇. 大数据时代的人力资源管理: 挑战与创新[J]. 湖北工程学院学报, 2021, 41(1): 91-95.
- [2] 蔡庆秋, 王双园, 白国振, 等. 基于树莓派的电池管理系统设计与实现[J]. 软件导刊, 2022, 21(11): 93-98.
- [3] 丁晓进, 朱晓荣. 基于微信小程序的云计算学习平台开发[J/OL]. 软件导刊, 2023: 1-5.  
<http://kns.cnki.net/kcms/detail/42.1671.TP.20230625.1743.038.html>
- [4] 平欣, 宋育斌, 孔维宾, 等. 基于 GPS 与微信小程序的智能宠物管理系统设计[J]. 软件导刊, 2023, 22(2): 121-126.
- [5] 张同硕, 廖明军, 张荣华, 等. 基于 Django 的交通事故数据可视分析系统设计与实现[J]. 软件导刊, 2023, 22(7): 112-117.