

**2024年《机器学习》**

**工程报告**

**（个人版）**



**课 程：** 机器学习

**姓 名：** 宁子扬

**学 号：** 2022217460

**完成时间：** 5/18

|  |  |  |  |
| --- | --- | --- | --- |
| **“机器学习-大作业”评分细则** | | | |
| **成绩等级** | **具体表现** | **教师评分** | |
| 优秀（100-90] | 报告撰写优秀，题目本身难度大，工作量饱满；Notebook代码清晰准确，实验与可视化优秀 | □ |  |
| 良好（90-80] | 报告撰写良好，题目本身难度较大，工作量较多；Notebook代码较好，实验与可视化良好 | □ |  |
| 中等（80-70] | 报告撰写中等，题目本身难度一般，工作量达标；Notebook代码一般，实验与可视化达标 | □ |  |
| 及格（70-60] | 能完成基本要求，工作量较少，非Notebook代码 | □ |  |
| 不及格（<60） | 未达最低要求或抄袭线上线下资源 | □ |  |

教师签名：

一． 自己所作工作的简介

本报告详细介绍了使用逻辑回归模型对《英雄联盟》比赛结果进行分类的过程。我们使用了从《英雄联盟》比赛中收集的数据，这些数据包含了关于比赛的各种特征，如哪一方先获得一血、哪一方先击杀小龙等。我们的目标是基于这些特征预测比赛的获胜方。

（说明：以下分内容，根据自己做的划分内容，写在下面的对应分项里）

二、研究背景与意义

**研究背景**

电子竞技（Esports）已经成为全球娱乐和竞技活动的重要组成部分。其中，《英雄联盟》（League of Legends）是最受欢迎的电子竞技游戏之一，具有庞大的玩家基础和专业赛事体系。在《英雄联盟》的对局中，两支由五名玩家组成的团队互相对抗，目标是摧毁对方的核心建筑“水晶”（Nexus）。比赛结果受到多种因素的影响，包括团队成员的技能、战术选择、英雄搭配等。

数据科学和机器学习技术在电子竞技分析中的应用逐渐增加，可以帮助团队和玩家从海量数据中提取有价值的信息，从而在比赛中获得优势。通过分析比赛数据，可以识别出影响比赛结果的关键因素，为战术制定和训练提供依据。

**研究意义**

1. 战术优化：通过分析比赛数据，可以识别出哪些战术和策略更可能带来胜利，从而帮助团队优化他们的游戏策略。

2. 个人技能提升：玩家可以通过分析自己的表现和与高水平玩家的对比，找到提升自己技能的方法。

3. 选秀策略：在比赛开始前的英雄选择阶段，了解哪些英雄组合更有可能获胜，对于制定选秀策略至关重要。

4. 预测市场：对于电子竞技赌博和预测市场，准确的比赛结果预测模型可以提供重要的参考。

5. 观众体验：通过提供深入的数据分析，可以提高观众的观看体验，增加电子竞技的观赏性和专业性。

6. 学术研究：在电子竞技领域应用数据科学和机器学习技术，可以为相关学术研究提供新的视角和实验场景。

综上所述，使用逻辑回归对《英雄联盟》比赛结果进行分类的研究具有重要的实际应用价值和学术意义。通过这一研究，我们不仅能够更好地理解游戏机制和比赛策略，还能推动电子竞技数据分析领域的发展。

三、模型方法

**逻辑回归**

四、系统设计

### 1. 数据采集

**数据源**：《英雄联盟》比赛数据可以从官方比赛记录、第三方统计网站或自定义的数据收集工具中获得。

**数据格式**：数据通常以CSV或JSON格式存储，包含比赛的基本信息、玩家统计数据、事件时间戳等。

### 2. 数据预处理

**数据清洗**：去除无效或不完整的记录，处理异常值。

**特征工程**：选择与比赛结果相关的特征，如击杀、助攻、经济差距、塔数等。

**数据转换**：将分类特征转换为数值形式，如将队伍颜色转换为0和1。

**数据分割**：将数据集分割为训练集和测试集，以评估模型的性能。

### 3. 模型训练

**算法选择**：使用逻辑回归算法，因为它适用于二分类问题，并且可以提供概率输出。

**模型训练**：使用训练集数据训练逻辑回归模型。

**参数调优**：通过交叉验证等方法选择最优的模型参数。

### 4. 模型评估

**准确率**：计算模型在测试集上的准确率。

**混淆矩阵**：生成混淆矩阵以评估模型的分类性能。

**分类报告**：提供精确度、召回率、F1分数等详细统计指标。

### 5. 结果可视化

**图表输出**：使用图表库（如Matplotlib、Seaborn）将模型的性能指标以图表形式展示。

**热力图**：使用热力图展示混淆矩阵，以便直观地理解模型的分类效果。

### 6. 部署与应用

**API服务**：将训练好的模型部署为API服务，以便其他系统可以调用。

**用户界面**：设计用户界面，允许用户输入比赛数据并获取预测结果。

**集成**：将系统集成到现有的电子竞技分析平台或工具中。

### 7. 维护与更新

**数据更新**：定期更新数据集，以反映最新的游戏版本和战术变化。

**模型再训练**：随着新数据的积累，定期重新训练模型以保持其准确性和相关性。

通过以上系统设计，我们可以构建一个高效、可扩展的《英雄联盟》比赛结果预测系统，为电子竞技分析提供有价值的见解和工具。

五．实验结果分析、对比和讨论

**实验结果分析**

在本次实验中，我们使用了逻辑回归模型对《英雄联盟》比赛结果进行分类。模型在测试集上的准确率达到了约 90.7%。这一结果表明，模型能够有效地从输入特征中学习到与比赛结果相关的信息，并对新的数据做出相对准确的预测。

**实验结果对比**

为了更好地理解模型的表现，我们可以将逻辑回归模型的性能与其他分类算法进行比较。常见的分类算法包括决策树、随机森林、支持向量机等。这些算法在不同类型的数据集上可能有不同的表现，因此进行对比可以帮助我们找到最适合当前问题的算法。

**实验结果讨论**

1. 模型解释性：逻辑回归模型的一个重要优点是其结果的可解释性。每个特征对预测结果的贡献可以通过模型的系数来衡量。这有助于我们理解哪些因素对比赛结果最重要。

2. 特征重要性：在未来的工作中，可以进一步分析模型系数，以识别出对比赛结果影响最大的特征。这可以为团队提供战术上的指导，比如在比赛中应该重点关注的对象或策略。

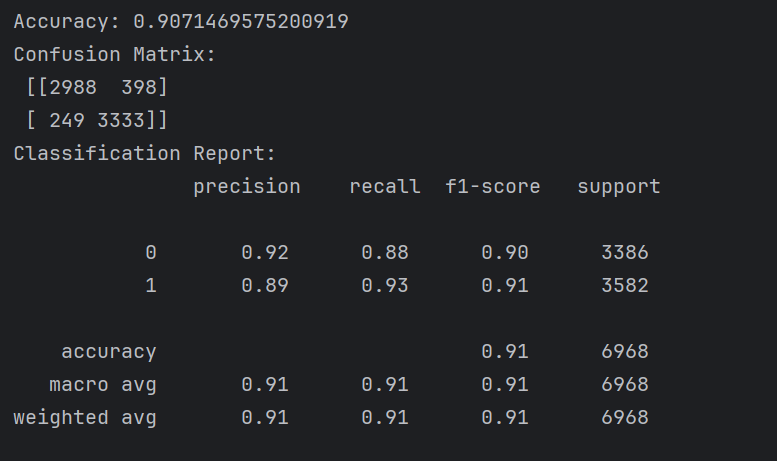
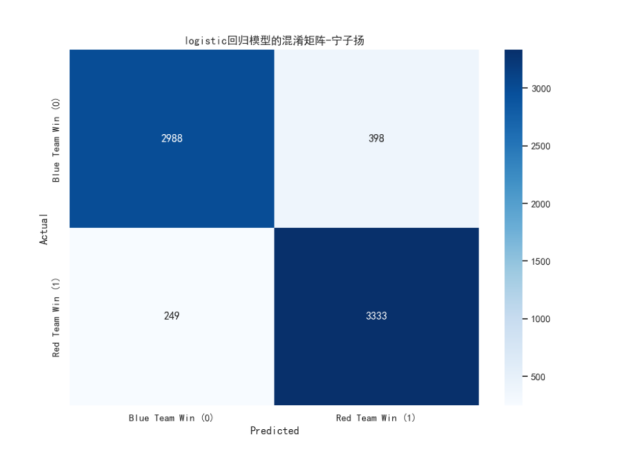
3. 模型改进：虽然逻辑回归模型在本次实验中表现良好，但仍有改进空间。可以考虑引入更多的特征，如玩家个人技能评分、团队协作评分等，以提高模型的预测能力。

4. 过拟合风险：逻辑回归模型在特征数量相对于样本数量过多时，可能会出现过拟合现象。在实际应用中，应通过交叉验证等方法来避免过拟合，并确保模型具有良好的泛化能力。

5. 实时预测：在电子竞技领域，实时数据分析和预测具有重要价值。可以将模型部署为实时预测系统，为观众提供即时的比赛结果预测，增加观赛体验。

6. 版本更新和战术变化：《英雄联盟》游戏会定期更新，每次更新都可能引入新的英雄、改变游戏机制或战术。因此，模型需要定期更新以适应这些变化。

通过本次实验，我们不仅验证了逻辑回归模型在《英雄联盟》比赛结果预测中的有效性，还为进一步的研究提供了方向和思路。未来的工作可以围绕模型改进、特征工程和实时预测系统等方面展开。



注意：代码输出的所有可视化结果图中必须加中文标题，并在标题结尾加上“实现该代码的学生姓名”，如“**多种算法ROC曲线对比图-张三**”。

六．其他

上面五部分之外的内容，写在这里。

七．对本门课的感想、意见和建议

机器学习是一门高度实用的学科，它不仅提供了一系列强大的工具和算法来处理和分析数据，还能够帮助我理解和模拟复杂的现象和行为。这门课程涵盖了从数据预处理、特征工程、模型选择、训练、评估到部署的全过程，使我能够系统地掌握机器学习的基本原理和应用技巧。

在机器学习课程中，我将学习如何使用各种算法来识别数据中的模式和趋势，以及如何通过这些模式和趋势来做出预测或决策。这些算法包括但不限于线性回归、决策树、随机森林、支持向量机、神经网络等，每种算法都有其适用的场景和优缺点。

通过学习机器学习，不仅能够提高我对数据和模型的理解能力，还能够培养批判性思维和问题解决能力。这有助于我们学生在未来的职业发展中更好地应对复杂的问题和挑战。此外，机器学习还与许多其他学科密切相关，如统计学、计算机科学、人工智能等，因此学习机器学习也有助于学生建立跨学科的知识体系。

总的来说，机器学习是一门非常实用的学科，它为我们提供了强大的工具和算法来处理和分析数据，并帮助我们在未来的职业发展中更好地应对复杂的问题和挑战。