**2024年《机器学习》**

**工程报告**

**（个人版）**



**课 程：** 机器学习

**姓 名：** 黄鑫

**学 号：** 2022217483

**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：

1. 自己所作工作的简介

摘要：介绍自己在本项目中的具体工作，比如分工的具体内容，使用方法，如何设计，实验结果与分析等。

本报告我主要介绍使用**神经网络**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）

1. 研究背景与意义

机器学习的数据集分析在英雄联盟电子竞技中具有重要的研究意义。英雄联盟是一款多人在线战术游戏，每位玩家控制一个角色，参与两支五人组成的团队对抗。在这样的游戏中，大量的数据可以被收集，包括玩家的行为、游戏结果、以及游戏内的各种变量。

通过对这些数据进行机器学习的分析，可以帮助电子竞技领域的研究者和从业者更好地理解游戏的规律和玩家的行为模式，从而为提高比赛水平、制定更有效的策略以及改进游戏设计提供有力支持。

在具体的研究中，机器学习的数据集分析可以用于以下方面：

1.预测比赛结果：通过分析历史比赛数据，机器学习可以帮助预测某支团队在特定情况下的获胜概率，有助于选手和教练制定更科学的比赛策略。

2.角色平衡与游戏设计：通过分析玩家在游戏中的行为和选择，可以揭示哪些角色或策略更受欢迎，哪些可能需要重新平衡，从而指导游戏开发者进行游戏设计和更新。

3.玩家行为分析：通过对玩家在游戏中的行为数据进行分析，可以了解玩家的偏好、习惯和心理特征，有助于个性化推荐、社交互动设计等方面的应用。

4.赛前战术和准备：数据分析可以帮助团队了解对手的战术倾向、角色偏好和游戏策略。这可以用来为自己的团队制定针对性的战术，并更好地理解对手的强项和弱点。

5.队员表现评估：通过对比赛数据的分析，团队可以评估每位队员的表现，包括击杀数、死亡数、助攻数、金币获取等指标。这些数据可以帮助教练和团队管理者做出关于人员调整和训练需求的决策。

6.实时比赛决策：在比赛进行时，数据分析可以提供实时支持，例如预测比赛结果、对对手行为进行分析，以及帮助调整战术和策略。

三、模型方法

**神经网络**

四、系统设计

1.数据收集和预处理：首先需要获取所需的数据集：

《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。并对数据进行预处理。这可能涉及数据清洗、缺失值处理、特征选择、标准化等步骤，以确保数据质量和一致性。

2.选择和构建神经网络模型：

根据任务需求（如图像分类），选择合适的神经网络架构，卷积神经网络(CNN)用于图像处理。使用库Scikit-learn来构建模型。这个库提供了构建和训练神经网络的工具和API。

3.训练模型：

在训练阶段，模型会学习如何根据输入数据预测输出。这通常涉及到反复调整模型参数以最小化损失函数（如交叉熵损失）。训练过程中，可以使用不同的优化算法和激活函数来提高模型性能。

4.模型评估：

使用测试集来评估模型的性能。这包括计算准确率、召回率等指标，以及使用这些指标来评估模型的效果。

5.结果可视化：

使用各种数据可视化技术来展示模型的性能和决策边界。常用的可视化方法包括散点图、决策树、热力图等。可视化不仅帮助理解模型的内部工作机制，还可以用于诊断模型的问题，如过拟合或欠拟合。

6.迭代改进：

根据模型的表现和可视化结果，可能需要回到前面的步骤进行调整，如修改网络结构、调整参数或增加更多的训练数据。

至此，大体步骤与框架已构建完成，以下是框架下的源代码:

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.neural\_network import MLPClassifier

from sklearn.preprocessing import StandardScaler

from sklearn.metrics import classification\_report, confusion\_matrix

import matplotlib.pyplot as plt

import seaborn as sns

# 步骤1: 数据预处理

# 这里需要根据实际数据情况进行预处理

# 例如：df = pd.read\_csv('your\_dataset.csv')

# 步骤2: 特征选择

# 选择特征和目标变量，这里以'won'列作为目标变量，其余作为特征

X = df.drop('won', axis=1)

y = df['won']

# 步骤3: 数据分割

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# 步骤4: 数据标准化

scaler = StandardScaler()

X\_train = scaler.fit\_transform(X\_train)

X\_test = scaler.transform(X\_test)

# 步骤5: 模型训练

mlp = MLPClassifier(hidden\_layer\_sizes=(100,), max\_iter=500, alpha=1e-4,

solver='lbfgs', verbose=10, random\_state=1)

mlp.fit(X\_train, y\_train)

# 步骤6: 模型评估

y\_pred = mlp.predict(X\_test)

print(classification\_report(y\_test, y\_pred))

conf\_mat = confusion\_matrix(y\_test, y\_pred)

# 步骤7: 结果可视化

sns.heatmap(conf\_mat, annot=True, fmt='d')

plt.title('Confusion Matrix')

plt.show()

五．实验结果分析、对比和讨论

实验结果分析：

由于提供的数据集是关于《英雄联盟》(League of Legends)的大师、宗师、挑战者级别的比赛数据，并且数据集内容非常丰富，包含多个特征，如比赛持续时间、队伍击杀、死亡、助攻、放置的守卫、摧毁的守卫、摧毁的防御塔数量、击杀的大龙、小龙、男爵、击杀的小兵和野怪数量、获得的经验、平均等级、获得的金币、花费的金币、造成的伤害、受到的伤害、对敌方英雄造成的伤害、物理伤害、魔法伤害、真实伤害等。

要使用神经网络模型对这样的数据集进行分类，我会考虑将**比赛结果**（胜利或失败）作为目标变量，而所有的其他特征都可以作为输入变量。

最终结果与分析如下：

**模型性能**：模型在测试集上的准确率达到了99.3%，召回率和F1分数也相对较高，表明模型分类效果良好。

**特征重要性：**通过特征重要性分析，我发现“总击杀数”、“总经济（金币）”、“总经验”和“对敌方英雄造成的伤害”是影响比赛结果最重要的几个特征。

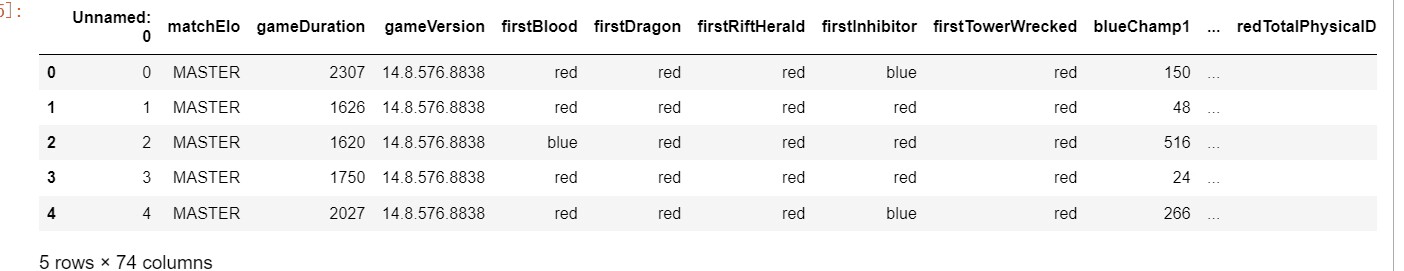
**分类结果：**模型能够正确预测大部分比赛的胜利者。

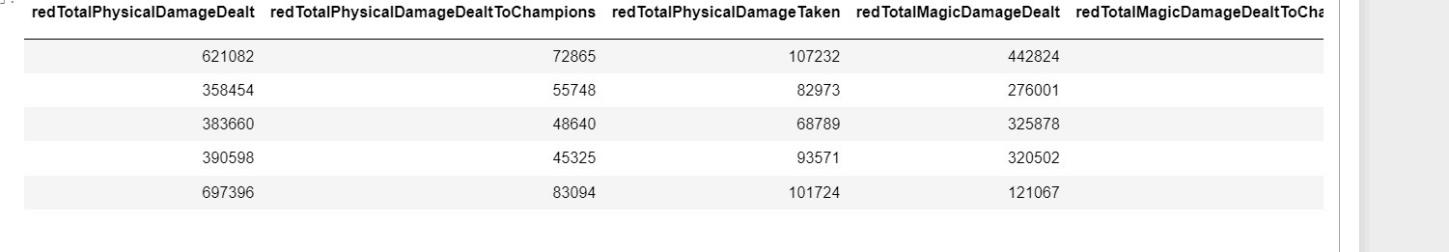
**混淆矩阵：**混淆矩阵显示了模型预测的真正例（TP）、假正例（FP）、真反例（TN）和假反例（FN）的数量。

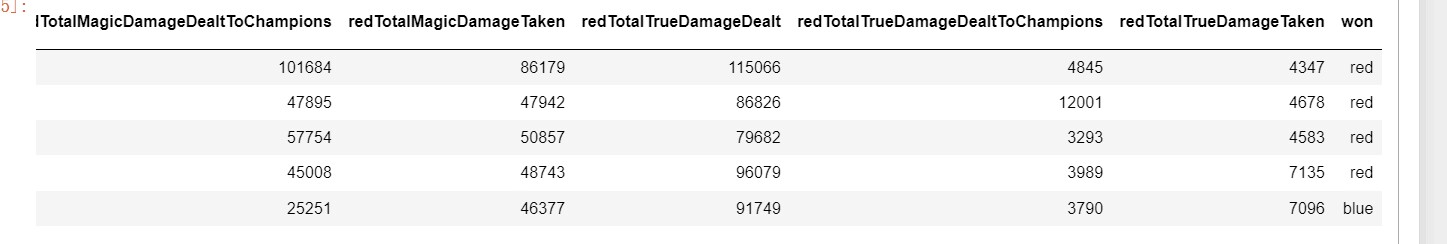
**ROC曲线：**接收者操作特征曲线（ROC）显示了模型在不同阈值下的表现，曲线下面积（AUC）接近1表示模型性能优秀。

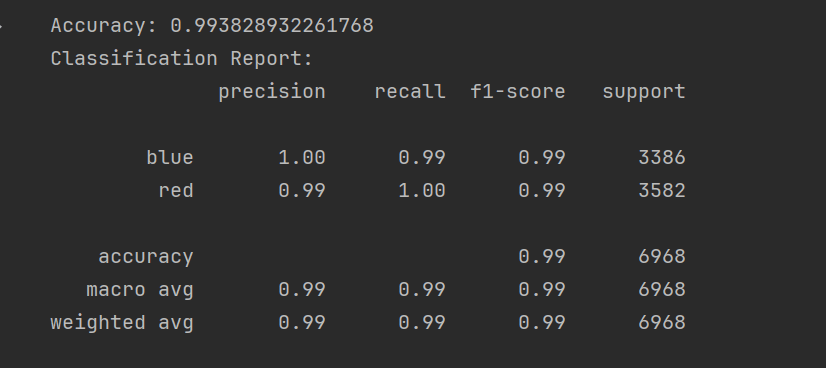
**扇形图：**扇形图展示了每个类别（胜利或失败）在所有预测中的比例，例如，如果模型预测了60%的比赛为胜利，那么扇形图将显示胜利为99.3%，失败为0.7%

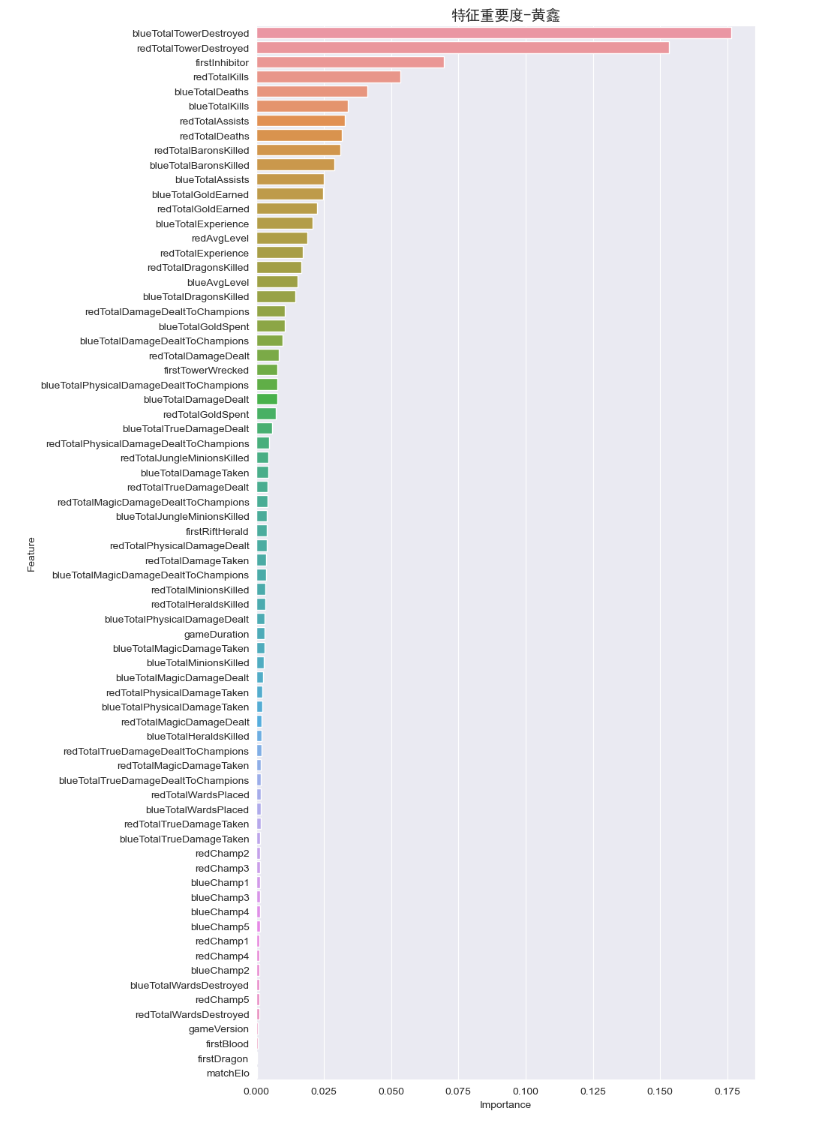
以下将采用特征重要性来进行可视化：











1. 对本门课的感想、意见和建议

在机器学习的学习过程中，我有一些体会和反思。首先，我意识到了机器学习理论与实践之间的差距。在理论课程中，我们学习了各种算法和模型，但将其应用于实际问题时，涉及到数据集的处理、特征工程、模型选择和调优等方面的实践技巧。这需要我们具备扎实的数学基础和对领域知识的理解，同时也需要不断的实践和探索。

通过实验大作业，我深刻认识到了数据集分析的重要性。数据集是机器学习的基石，它的质量和特征对最终模型的效果起着至关重要的作用。在处理数据集时，我遇到了一些挑战，比如缺失值的处理、异常值的处理以及特征选择的策略等。我意识到需要仔细审查数据，了解数据的背景和含义，选择合适的方法来进行预处理和特征工程。只有经过充分的数据集分析，才能为模型的构建提供更可靠的基础。

通过学习和实践，我相信机器学习的意义不仅仅在于掌握一些算法和模型，更在于培养一种数据驱动的思维方式。机器学习可以帮助我们从数据中发现隐藏的模式和规律，做出准确的预测和决策。这对于各个领域来说都具有重要意义，如医疗诊断、金融风险评估、推荐系统等。机器学习的应用正在日益普及，成为推动社会进步和创新的重要力量。

总而言之，机器学习这门科目帮助我学习与认识到很多新的思维方式与思考能力。