**机器学习作业1-PUBG排名预测**

**郭远帆，唐荣俊，罗乙然**

目录

[任务介绍 1](#_Toc527654737)

[实现方案 1](#_Toc527654738)

[实验结果 1](#_Toc527654739)

[分析与后续 1](#_Toc527654740)

# 任务介绍

## 任务来源

本次任务来源于Kaggle数据科学竞赛网站提供的比赛项目:**PUBG Finish Placement Prediction （绝地求生:大逃杀玩家排名预测）**

比赛说明网站: <https://www.kaggle.com/c/pubg-finish-placement-prediction> 网站提供了该任务的详细说明，输出要求以及数据（包括训练集和测试集）

这是一个非常有趣的项目，竞赛网站与PUBG制作商蓝洞公司合作，获取了超过65，000场游戏的玩家数据，要求竞赛参加者根据这些数据预测玩家在比赛中的排名。

## 输出要求

## 任务目标与规划

我们小组以此项目作为机器学习第一次作业，该问题属于回归问题。而由于该项目本身是数据科学竞赛项目，其中必然存在数据分析与特征提取的过程（但并非主要的），我们将任务划分为几个阶段:

第一阶段: 数据缺失处理，进行简单的数据可视化分析

第二阶段: 参考竞赛网站其它竞赛参加者提供的特征工程方案，提取简单特征

第三阶段: 构建不同的机器学习模型训练，与测试集进行比较

第四阶段: 采用集成学习技术将数个模型整合，构成最终模型并进行分析比较

其中第三第四阶段为主要的工作

# 实现方案

## 编程环境

## 特征工程

# 实验结果

## 线性回归模型

在完成一轮特征工程后，数据集中留下来的特征拥有较好的独立性，并且对于初步模型来说降低了数据的复杂度，因此也降低了拟合的难度。首先我们使用scikit-learn 中的线性模型进行了第一次预测，计算了预测的平均绝对值误差。

实验结果:

模型在测试集上的表现为 M = 0.1266 ，M为平均绝对值误差，通过预测结果的取值范围得，线性回归模型平均将引入12%的误差，结果不太理想。

实际上，线性回归模型的准确性十分依赖于特征提取，因此结果的不理想性可以理解。由于是初步搭建模型，线性回归仅作测试用，接下来将使用更加合适的模型并进行改进。

## 回归SVM模型

# 分析与后续