**使用说明指南**

**项目介绍**

本项目可通过名为“nba\_spider”与“nba\_spider2”的Python爬虫获取来自NBA官网的任意年度的球队数据，保存在本地文件夹后，更改名为“NBA-nwz”的代码中的路径folder为数据文件路径，同时需要按照课程设计文件中的要求更改好csv文件中的数据字段等格式。如下图所示。

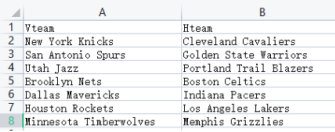


图1.设计报告截图

处理好数据文件之后，即可运行NBA代码，导入球队各类数据而后进行特征向量、逻辑回归、球队的EloScore计算等机器学习，最终将预测的比赛结果输出到特定路径下的格式为.csv的文件查看比赛预测结果。

**使用说明**

1、为了方便，已提前在文件夹中已经包含且处理好了实验所需的5个数据文件，点击打开名为NBA.py的代码。更改代码中的folder路径，设置路径为文件夹所在的路径即可。

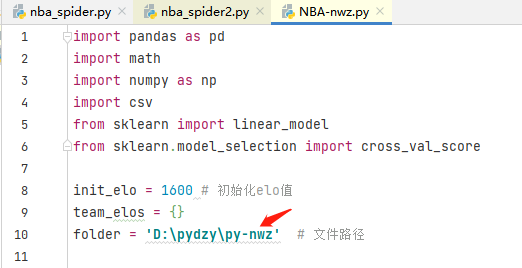


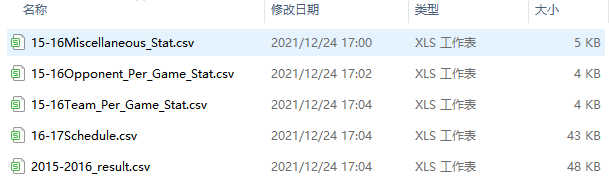
图2.更改folder路径

图3.实验所需5份数据文件

1. 右键点击运行即可。随后将在设置好的folder文件夹下出现名为 prediction of 2016-2017.csv的文件。此文件即为预测胜率文件。

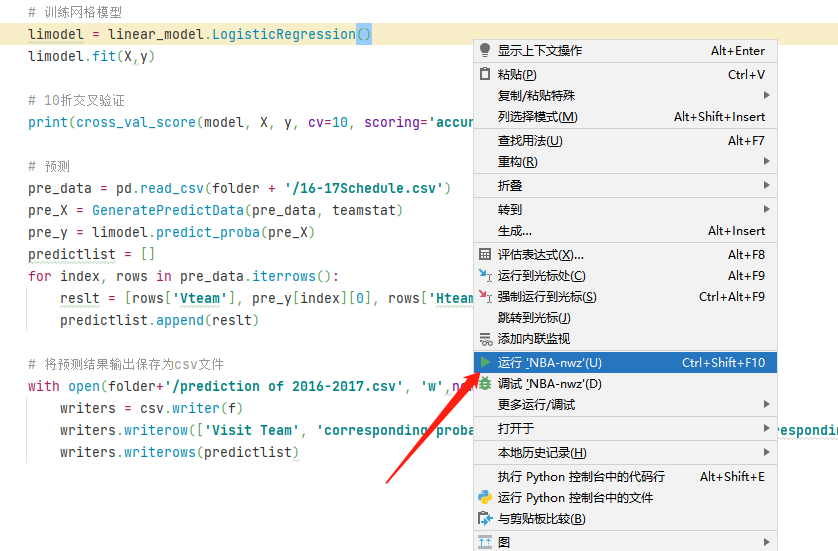


图4.运行NBA预测代码

1. 通过爬虫获取数据文件，在代码“nba\_spider”与“nba\_spider2”中只需要更改好NBA网站对应的url地址，即可一键爬取实验所需的数据文件。

nba\_spider代码：是用于爬取5个表中的Team Per Game Stat表、Opponent Per Game Stat表。

nba\_spider2代码：用于爬取NBA Schedule文件。（但Schedule文件爬取之后需要手动处理数据。）

nba\_spider3代码：用于爬取Miscellaneous\_Stat文件。

关于文件的介绍可以查看课程设计报告。

NBA网站地址链接：

https://www.basketball-reference.com/leagues/NBA\_2016\_games.html

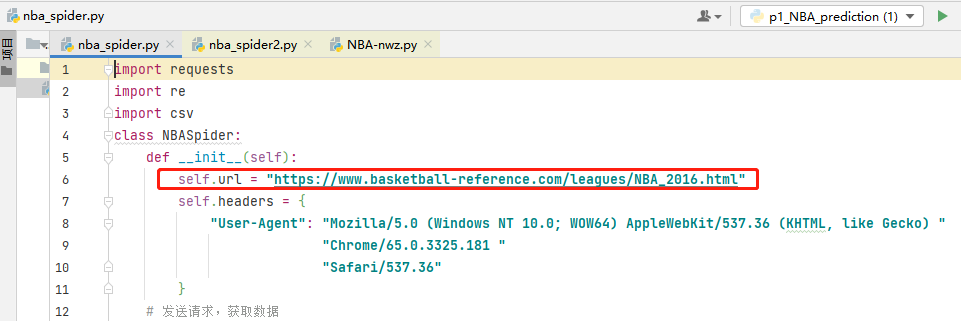


图6.nba\_spider代码截图



图7.nba\_spider2代码截图

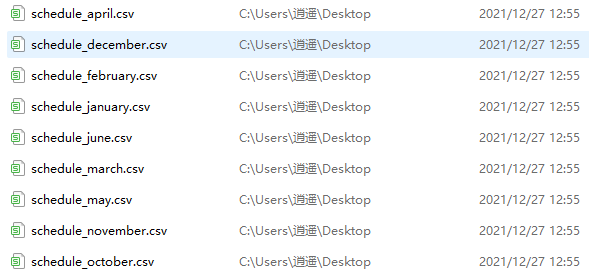


图8.爬取的数据文件示意图