**系统使用说明书**

胡剑桥 201830660420

1. **软件概况**

使用python库的sklearn进行机器学习模型的训练。使用pandas库进行数据的读取与写入CSV文件。使用boto3库对于生成CSV文件上传到S3数据库中进行保存以便于之后的使用。

使用的数据有sklearn中自带的波士顿数据集和kaggle上面下载的皮马印第安人糖尿病数据集。

本软件使用了代码版本管理，托管在码云，地址是https://gitee.com/masaikk/kgsweek4.git

1. **文件代码**

根目录下面的pass文件夹中放的是本项目的文档。

notebook文件夹中是使用jupyter，其中使用了线性回归对于波士顿房价数据集进行预测。

src文件夹下面的是对于波士顿房价的线性回归的python文件。

LR文件夹下面的是使用随机森林对于皮马印第安人糖尿病数据集进行预测。

请先按照根目录下面的requirements.txt文件进行依赖的安装。

使用anaconda，运行命令conda install --yes --file requirements.txt

1. **代码运行**

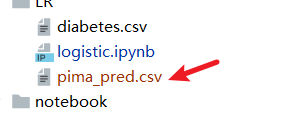
notebook文件夹下面运行main.ipynb

LR文件夹下面运行logistic.ipynb

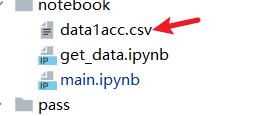
在运行完之后，本地会生成两个csv文件表示着数据保存下来了。

如下所示：

这里是皮马印第安人糖尿病数据集的预测，最后一列是预测的结果。

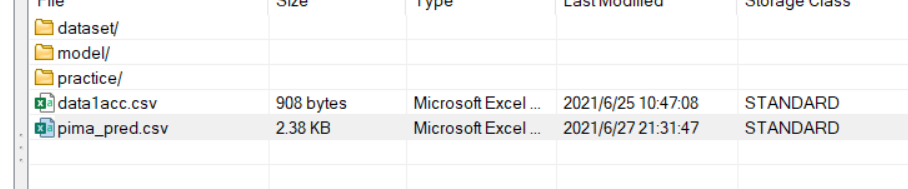


这里是波士顿房价的预测，最后添加了预测的房价和真实数据与预测数据的偏差值。



最后会自动将这两个文件上传到S3数据库中。

1. **上传结果**



这两个csv文件的数据如下所示

