文本

描述已自动生成

# 基于Scikit-learn基本算法的机器学习使用说明

姓名：李琰朕

学号：201830661250

班级：软件三班

年级：2018级

指导教师：汤峰

报告时间：2021年6月24日

## 环境配置：

* Python版本大于等于3.6
* 已经安装Scikit-learn的依赖
* 已经安装boto3的依赖
* 在S3对应的bucket内新建一个result文件夹用于存放上传的预测结果

## 软件配置：

1. 将dataset文件夹下的dataset.zip文件解压得到train.csv和test.csv文件，此时dataset文件下应有如下三个csv文件：

图形用户界面, 应用程序

描述已自动生成

1. 随后用PyCharm打开项目，在code文件夹下修改Upload.py中的S3的个人配置，包括access\_key以及secret\_key，如下图所示：

文本

描述已自动生成

1. 随后可以打开Main.py，点击运行该python文件即可，之后程序会自动运行，并给出如下输出：

可以看到文本

描述已自动生成，程序会先使用四种sklearn中给的基本算法，包括KNN，SVM，高斯分布，多项式算法来进行训练和预测；随后还会跟官方给出的bench\_mark来进行比较得出错误率进行评估；最后会将得到的结果上传至S3的result文件夹内，完成程序的运行。

1. 最后我们可以打开S3的result文件夹下，即可看到里面已经上传了四种算法给出的预测结果：

图片包含 形状

描述已自动生成

## 报错及可能原因：

1. 显示未能连接至S3，或S3没有指定位置，请检查是否按照软件配置的要求进行S3的相关配置
2. 运行后显示dataset下没有指定文件，请检查是否将所要求的数据集放入该文件夹下