# 程序说明

## 基本功能说明

本程序提供了csv文件解析工具，Web端数据可视化模块。

### 目录格式说明

data：存放待处理的原始数据；

templates：存放前端HTML UI页面；

utils：存放工具类csv.py，csv.py为自主实现的读取csv格式文件的工具。

run.py：后端flask文件，用于后端路由和utils工具包的调用。

## 预处理关键部分代码

# 本函数高度封装了csv文件的解析方法，具有复用性

# file\_path参数可以是文件，也可以是目录（处理目录下的所有文件）

# title参数标识了csv格式文件的第一行是否是title。如果不是需要传入title数组用以在表头显示

**def** parser(file\_path, title=None):  
  
 result = dict()  
 result[**'data'**] = []  
  
 **if** os.path.isdir(file\_path):  
 result[**'title'**] = title  
 **for** files **in** os.listdir(file\_path):  
 **with** open(file\_path + **'/'** + files) **as** f:  
  
 **for** line **in** f:  
 result[**'data'**].append(Parser.parser\_str(line))  
  
 **elif** os.path.isfile(file\_path):  
 **with** open(file\_path) **as** f:  
  
 **if not** title:  
 title = Parser.parser\_str(f.readline())  
 result[**'title'**] = title  
  
 **for** line **in** f:  
 result[**'data'**].append(Parser.parser\_str(line))  
 **else**:  
 **raise** IOError, **'The file path is wrong!'  
  
 return** result

## 运行截图









# 数据集的格式

## 数据集1 Forest-Fires

每行代表一条数据，分隔符为逗号，结尾有回车符。

其字段含义如下：

1. X - x-axis spatial coordinate within the Montesinho park map: 1 to 9

2. Y - y-axis spatial coordinate within the Montesinho park map: 2 to 9

3. month - month of the year: "jan" to "dec"

4. day - day of the week: "mon" to "sun"

5. FFMC - FFMC index from the FWI system: 18.7 to 96.20

6. DMC - DMC index from the FWI system: 1.1 to 291.3

7. DC - DC index from the FWI system: 7.9 to 860.6

8. ISI - ISI index from the FWI system: 0.0 to 56.10

9. temp - temperature in Celsius degrees: 2.2 to 33.30

10. RH - relative humidity in %: 15.0 to 100

11. wind - wind speed in km/h: 0.40 to 9.40

12. rain - outside rain in mm/m2 : 0.0 to 6.4

13. area - the burned area of the forest (in ha): 0.00 to 1090.84

## 数据集4 Credit + Approval

每行代表一条数据，用逗号作为分隔符，结尾有回车符。

其字段含义如下：

A1: b, a.

A2: continuous.

A3: continuous.

A4: u, y, l, t.

A5: g, p, gg.

A6: c, d, cc, i, j, k, m, r, q, w, x, e, aa, ff.

A7: v, h, bb, j, n, z, dd, ff, o.

A8: continuous.

A9: t, f.

A10: t, f.

A11: continuous.

A12: t, f.

A13: g, p, s.

A14: continuous.

A15: continuous.

A16: +,- (class attribute)

## 数据集5 Iris

每行代表一条数据，用逗号作为分隔符，结尾有回车符。

其字段含义如下：

1. sepal length in cm

2. sepal width in cm

3. petal length in cm

4. petal width in cm

5. class:

-- Iris Setosa

-- Iris Versicolour

-- Iris Virginica

## 数据集 附加 出租车GPS

共有四个文件，每行代表一条数据。逗号为分隔符，结尾有回车符和换行符。

其字段含义如下：

1车辆标识6个字符

2触发事件0=变空车，1=变载客，2=设防，3=撤防，4=其它

3运营状态0=空车，1=载客，2=驻车，3=停运，4=其它

4 GPS时间 格式yyyymmddhhnnss，北京时间

5 GPS经度 格式ddd.ddddddd，以度为单位

6 GPS纬度 格式dd.ddddddd，以度为单位

7 GPS速度 格式ddd，取值000-255内整数，以公里/小时为单位

8 GPS方位 格式ddd，取值000-360内整数，以度为单位

9 GPS状态 0=无效，1=有效

# 可执行程序

本程序展示部分在Web端。使用Python web框架进行开发，使用前请在本机使用pip install flask指令安装后台框架。然后使用python运行run.py，就可以访问<http://localhost:8080> 查看UI。

为了展示方便，本程序已经在http://118.89.230.54:8080部署。但是由于个人云服务带宽限制（1M），在HTML页面展示所有数据会导致网络访问特别缓慢。所以在展示的时候，只展示前100条数据用以表明自己成功读取。