数据挖掘可视化系统-接口文档 V1.0

/api-DataFrame

ENGLISH:

The api to analyze the csv format language.

- **METHOD:** POST
- **Args:**
  - dataSet: the data set originally loaded to submit,
  - sep: the separate character
- **Return:**
  - DataFrame as Json shape like:
  - ["id" : ["1","2", "3", "4"], "feature1" : ["abc", "12fw", "nmd", "wtm", "haode1"]]

中文:

用以解释 csv 文法的接口

- **提交方法:** POST
- **参数:**
  - dataSet: 读取未经处理的 csv 字符串
  - sep: 分割符号
- **返回值:**
  - Json 格式的 DataFrame，如:
  - ["id" : ["1","2", "3", "4"], "feature1" : ["abc", "12fw", "nmd", "wtm", "haode1"]]

/api-pretreatment

ENGLISH:

The api to pretreat the data set.

- **METHOD**: POST
- **Args:**
  - dataSet: the data set to submit, must be shaped like :
    - ["id" : ["1","2", "3", "4"], "feature1" : ["abc", "12fw", "nmd", "wtm", "haode1"]]
  - dropColumns: the columns to be dropped like :
    - ["id", "feature1"]
  - discreteColumns: the columns with discrete elements value like:
    - ["feature1"]
  - textColumn: the columns with Text like:
    - "text"
- **Return:**
  - The hash key to pull data set and a status code (200 for OK, 403 for error)

中文:

用以对数据集进行预处理的

- 提交方法: POST
- **参数:**
  - dataSet: 用以提交的数据集， 形如 :
    - ["id" : ["1","2", "3", "4"], "feature1" : ["abc", "12fw", "nmd", "wtm", "haode1"]]
  - dropColumns: 用以 drop 掉的特征或者列， 形如 :
    - ["id", "feature1"]

- discreteColumns: 标记为离散值的特征或者列，形如:
  - ["feature1"]
- textColumn: 存储文段的特征或者列，形如:
  - "text"
- **返回值:**
- 用以拉取数据的哈希键以及状态码 (200 意为可以, 403 意为错误)

/api-fit

ENGLISH:

The api to fit the model.

- **METHOD:** POST
- **Args:**
- hashKey: the hash key to pull data set
- target: the target our model should predict and we call "label" as term
- model: the model chosen to fit the data set
  - 1: Naive Bayes
  - 2: KNN
  - 3: SVM
  - 4: Linear Regression
  - 5: Logistic Regression
  - 6: Decision Tree
- Return:(in JSON shape)
  - hashKey: the key to pull trained Model
  - cv_score: the score spawned by cross_validation
  - images: viusalization of the trained models (base64)
  - status_code: 200 for OK, 403 for ERROR

中文：

用以训练模型的 API

- **提交方法:** POST
- **参数:**
  - hashKey: 数据集的哈希键
  - target: 需要进行预测的特征，或者我们说的“标签”
  - model: 选择不同的模型
    - 1: 朴素贝叶斯（文字版）
    - 2: KNN
    - 3: 支持向量机
    - 4: 线性回归
    - 5: 对数几率回归
    - 6: 决策树
- 返回值:(JSON 格式)
  - hashKey: 与训练好的模型唯一对应的哈希键
  - cv_score: 通过交叉验证法生成的分数
  - images: 可视化图像（base64）
  - status_code: 200 意为 OK, 403 意为 错误

/api-predict

ENGLISH:

The api to predict.

- **METHOD:** POST
- **Args:**
  - hashKeyl: the hash key to pull data set

- hashKeyII: the hash key to pull trained model

- **Return:**

- The result in JSON shape and a status code (200 for OK, 403 for error)

中文:

用以预测的 API。

- **提交方法:** POST

- **参数:**

- hashKeyI: 存储数据的哈希键

- hashKeyII: 存储训练好的模型的哈希键

- **返回值:**

- JSON 形式的 DataFrame 训练结果以及状态码 (200 意为可以, 403 意为错误)

/api-getDataSet

ENGLISH:

The api to pull default data set

- **METHOD:** POST

- **Args:**different data sets

- 0: spam_ham_dataset

- 1: tree

- 2: cancer

- 3: house

- 4: iris

- 5: forest fire

-

- **Return:**

- DataFrame as Json shape like:
- ["id" : ["1","2", "3", "4"], "feature1" : ["abc", "12fw", "nmd", "wtm", "haode1"]]

中文：

用以抓取默认数据集

- **提交方法:** POST
- **Args:**different data sets
  - 0: 垃圾邮件集
  - 1: tree 集
  - 2: 乳腺癌集
  - 3: 房价集
  - 4: 简化鸢尾花集
  - 5: 森林火灾集
- **返回值:**
  - Json 格式的 DataFrame，如：
  - ["id" : ["1","2", "3", "4"], "feature1" : ["abc", "12fw", "nmd", "wtm", "haode1"]]

How to explain the Pool Manager?

调度系统是如何作业的?

All in this Image.

一图流。