

CSDN新首页上线啦，邀请你来立即体验！(http://blog.csdn.net/)

立即体



博客 (//blog.csdn.net/Default.html) 学院 (//edu.csdn.net?ref=toolbar)

下载 (//download.csdn.net?ref=toolbar) GitChat (//gitbook.cn/?ref=csdn)

更多 ▾

Q

🔍

📁

登录 (https://passport.csdn.net/mobile/register?ref=toolbar&action=mobileRegister)

注册 (http://passport.csdn.net/account/mobile/register?ref=toolbar&source=csdnblog1)

Bike Sharing Demand （Kaggle）

原创

2015年03月09日 19:18:03

2128

Forecast use of a city bikeshare system

Bike sharing systems are a means of renting bicycles where the process of obtaining membership, rental, and bike return is automated via a network of kiosk locations throughout a city. Using these systems, people are able rent a bike from a one location and return it to a different place on an as-needed basis. Currently, there are over 500 bike-sharing programs around the world.

The data generated by these systems makes them attractive for researchers because the duration of travel, departure location, arrival location, and time elapsed is explicitly recorded. Bike sharing systems therefore function as a sensor network, which can be used for studying mobility in a city. In this competition, participants are asked to combine historical usage patterns with weather data in order to forecast bike rental demand in the Capital Bikeshare program in Washington, D.C.



给出一个城市的自行车租借系统的历史租借数据，要求预测自行车租借数量。

Data Fields

- datetime** - hourly date + timestamp
- season** - 1 = spring, 2 = summer, 3 = fall, 4 = winter
- holiday** - whether the day is considered a holiday
- workingday** - whether the day is neither a weekend nor holiday
- weather** - 1: Clear, Few clouds, Partly cloudy, Partly cloudy
2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
- temp** - temperature in Celsius
- atemp** - "feels like" temperature in Celsius
- humidity** - relative humidity
- windspeed** - wind speed
- casual** - number of non-registered user rentals initiated
- registered** - number of registered user rentals initiated
- count** - number of total rentals

wiking_acm (http://bl...)

+ 关注

(http://blog.csdn.net/wiking_acm)

原创

粉丝

喜欢

未开通 (https://gi utm_sourc

- 他的最新文章
- 更多文章 (http://blog.csdn.net/wiking_acm)
- 工业界怎样评估一个问题是否适合用NL P解决 (http://blog.csdn.net/Wiking_acm/article/details/78019573)
- UTAustinX: UT.5.04x LAFF: Linear Al gebra - Foundations to Frontiers 课程笔记 (http://blog.csdn.net/Wiking_acm/article/details/77995828)
- 《人类简史》读书笔记 (http://blog.csdn.net/Wiking_acm/article/details/77995592)
- Stanford CS224n: Natural Language Processing with Deep Learning 课程笔记 (http://blog.csdn.net/Wiking_acm/article/details/72851005)
- Writing in the Sciences 课程笔记（ 论文的组织 and 写作 ）(http://blog.csdn.net/Wiking_acm/article/details/5299 9169)

- 相关推荐
- Kaggle Bike Sharing项目报告 (http://blog.csdn.net/u013795429/article/details/52092659)
- kaggle入门-Bike Sharing Demand自行车需求预测 (http://blog.csdn.net/qq_36330643/article/details/78276233)
- kaggle 入门 Bike sharing Demand prediction (http://blog.csdn.net/onepiecehuiyu/article/details/51519349)
- 大数据竞赛平台——Kaggle 入门篇 (http://blog.csdn.net/qq_36330643/article/details/78284364)

广告

他的热门文章

登录

注册

✕

主要的特征工程在于datetime上，很容易想到凌晨2点和下午两点的租借数是很不一样的。用的是RandomForestRegressor，很容易的得到了一个排名在1/3位置的结果。

```
[python]
1. import csv
2. import pandas as pd
3. from math import log, sqrt
4. from sklearn.ensemble import RandomForestRegressor
5. from sklearn.cross_validation import train_test_split
6.
7. data_file = pd.read_csv("train.csv", header=0)
8. test_file = pd.read_csv("test.csv", header=0)
9.
10. data_file["hour"] = pd.factorize([i.split()[1][:2] for i in data_file["datetime"]])[0]
11. data_file = data_file.drop("casual", axis=1)
12. data_file = data_file.drop("registered", axis=1)
13. data_file["rent"] = data_file["count"]
14. data_file = data_file.drop("count", axis=1)
15. test_file["hour"] = pd.factorize([i.split()[1][:2] for i in test_file["datetime"]])[0]
16.
17. train_x, test_x, train_y, test_y = train_test_split(
18.     data_file.values[:,1:-1], data_file.values[:, -1], test_size=.01, random_state=0)
19. test_data = test_file.values[:, 1:]
20. ids = test_file.values[:, 0]
21.
22. def rmsle(pred, real):
23.     n = len(pred)
24.     return sqrt(sum([(log(pred[i]+1)-log(real[i]+1))*2 for i in range(n)])/n)
25.
26. reg = RandomForestRegressor(n_estimators=100, random_state=0)
27. reg.fit(train_x, train_y)
28. print reg.feature_importances_
29. print rmsle(reg.predict(test_x), test_y)
30. print rmsle(reg.predict(train_x), train_y)
31. output = reg.predict(test_data).astype(int)
32.
33. predictions_file = open("RandomForestRegssor.csv", "wb")
34. open_file_object = csv.writer(predictions_file)
35. open_file_object.writerow(["datetime", "count"])
36. open_file_object.writerows(zip(ids, output))
37. predictions_file.close()
38. print "done."
```

版权声明：本文为博主原创文章，未经博主允许不得转载。



相关文章推荐

Kaggle Bike Sharing项目报告 (http://blog.csdn.net/u013795429/article/details/520...

Kaggle Bike Sharing项目报告标签：ML项目定义项目预览项目地址 https://www.kaggle.com/c/bike-sharing-demand
项目简介： ...

u013795429 (http://blog.csdn.net/u013795429) 2016年08月03日 14:04 1282

kaggle入门-Bike Sharing Demand自行车需求预测 (http://blog.csdn.net/qq_36330643...

接触机器学习断断续续有一年了，一直没有真正做点什么事，今天终于开始想刷刷kaggle的问题了，慢慢熟悉和理解机器学习
加入CSDN享受更精准的内容推荐(与5000万程序员共同成长) 问题， ...

《机器学习（周志华）》西瓜数据集3.0spc
(http://blog.csdn.net/wiking_acm/article/details/50971461)

7910

uva 工具网站 (http://blog.csdn.net/wiking_acm/article/details/8254330)

5549

Titanic: Machine Learning from Disast
er (Kaggle 数据挖掘竞赛) (http://blo
g.csdn.net/wiking_acm/article/detail
s/42742961)

5077

Deep Learning for Image Recognition
a Bengio and Alex C. Courville 读书笔记

4194

CNN 识别图形验证码 (http://blog.csdn.
net/wiking_acm/article/details/51145
701)


3775

内容举报


返回顶部

登录

注册


 qq_36330643 (http://blog.csdn.net/qq_36330643)

2017年10月18日 18:42


 227

kaggle 入门 Bike sharing Demand prediction (http://blog.csdn.net/onepiecehuiyu/...

根据天气等因素对自行车租赁数量进行预测 利用xgboost进行预测 import csv from math import log, exp import numpy
author =...


 onepiecehuiyu (http://blog.csdn.net/onepiecehuiyu)

2016年05月27日 20:10


 1040

大数据竞赛平台——Kaggle 入门篇 (http://blog.csdn.net/qq_36330643/article/details/...

这篇文章适合那些刚接触Kaggle、想尽快熟悉Kaggle并且独立完成一个竞赛项目的网友，对于已经在Kaggle上参赛过的网友来说，大可不必耗费时间阅读本文。本文主要介绍Kaggle，如何解决一个竞赛...


 qq_36330643 (http://blog.csdn.net/qq_36330643)

2017年10月19日 14:21


 98

七月机器学习项目实战之特征工程6城市自行车共享系统使用状况 (http://blog.csdn.net/qiuji...

特征工程小案例Kaggle上有这样一个比赛：城市自行车共享系统使用状况。提供的数据为2年内按小时做的自行车租赁数据，其中训练集由每个月的前19天组成，测试集由20号之后的时间组成。本项目功能:数据清理...


 qiujiyahao123 (http://blog.csdn.net/qiujiyahao123)

2017年03月31日 17:38


 1338

Kaggle之Grupo Bimbo Inventory Demand (http://blog.csdn.net/lijingpengchina/ar...

Grupo Bimbo Inventory Demandkaggle比赛解决方案集合 Grupo Bimbo Inventory Demand 在这个比赛中，我们需要预测某个产品在某个销售点每周的需求...

 lijingpengchina (http://blog.csdn.net/lijingpengchina)

2016年09月09日 20:54

 593



elect romotion bike (http://download.csdn.net/download/bmagui/...

(http://download.csdn.net/download/bmagui/...


2007年12月13日 11:27

18KB


 下载

深度揭秘android摄像头的autoFocus-----循环自动聚焦的实现（Android Camera AutoFo...

要实现android摄像头的autoFocus，并不难，但要实现自动连续的自动聚焦就要费点劲。这里将我的研究思路和解决方法记录于此。实现自动聚焦，方法有如下： 第一：最常见的实现---利用按键或...


 qq_15037231 (http://blog.csdn.net/qq_15037231)

2016年04月18日 11:46


 328

【nginx-rtmp】10、指令（Directives）之 视频点播（Video on demand）(http://blog...

六、视频点播（Video on demand） 1、play 语法：play dir|http://loc [dir|http://loc]* 上下文：rtmp, server, appl...


 envon123 (http://blog.csdn.net/envon123)

2017年08月03日 09:50


 138

单例模式之Initialization Demand Holder (IoDH) (http://blog.csdn.net/guanyue520/...

一种更好的单例实现方法 饿汉式单例类不能实现延迟加载，不管将来用不用始终占据内存；懒汉式单例类线程安全控制烦琐，而且性能受影响。可见，无论是饿汉式单例还是懒汉式单例都存在这样那样的...


 guanyue520 (http://blog.csdn.net/guanyue520)

2016年09月29日 17:48


 87

Palace Museum（故宫）faces funds demand (http://blog.csdn.net/easyjsf/article/...

BEIJING - China's Palace Museum, also known as the Forbidden City, has been urged to disclose its ac...

 easyjsf (http://blog.csdn.net/easyjsf)

2011年05月27日 20:28

 455



Bike英语文本播放机 (<http://download.csdn.net/download/heqinggu...>)

[/http://downloa](#)

2006年06月02日 00:00

0B

下载



Bike英语文本播放机 (<http://download.csdn.net/download/heqinggu...>)

[/http://downloa](#)


2006年05月27日 00:00

0B

下载

深度揭秘android摄像头的autoFocus-----循环自动聚焦的实现（Android Camera AutoFo...

要实现android摄像头的autoFocus，并不难，但要实现自动连续的自动聚焦就要费点劲。这里将我的研究思路和解决方法记录于此。实现自动聚焦，方法有如下： 第一：最常见的实现---利用按键或拍照...



yanzi1225627 (<http://blog.csdn.net/yanzi1225627>)

2013年02月09日 19:16

42559



Renesas R8c/2k 120 degree E_bike Solution (<http://download.csdn....>)

[/http://downloa](#)

2010年10月29日 21:53

111KB

下载



Why You Should Wear A Motor Bike Jacket (<http://download.csdn....>)

[/http://downloa](#)


2011年04月05日 16:40

29KB

下载

Oracle On Demand(OOD)项目中EBS打补丁的详细步骤 (<http://blog.csdn.net/t0nsha/ar...>)

EBS patching steps in Oracle On Demand(OOD) projects#check if patch applied & count of invalid objec...




t0nsha (<http://blog.csdn.net/t0nsha>)

2012年07月13日 10:41

1313

Visual Studio Macro: Locate Item in Solution Explorer on Demand (<http://blog.csd...>)

From: <http://dvanderboom.wordpress.com/2008/03/21/visual-studio-macro-track-item-in-solution-explo...>




pendle (<http://blog.csdn.net/pendle>)

2012年08月24日 00:55

1369

Demand Side Platform (<http://blog.csdn.net/longshenlmj/article/details/9137131>)

DSP特点： DSP不是从网络媒体那里包买广告位，也不是采用CPD(Cost Per Day)的方式获得广告位；而是从广告交易平台(AdExchange)来通过实时竞价的方式获得对广告进行曝光的机会...



longshenlmj (<http://blog.csdn.net/longshenlmj>)

2013年06月20日 15:00

869

【转载】搜索广告与广告网络Demand技术 (<http://blog.csdn.net/zhangmengjlu/article...>)

搜索广告与广告网络Demand技术 原作者：刘鹏 @北冥乘海生 编辑：Koala++/屈伟 视频课程：师徒网 流式计算平台-Stor m 我们以Sto...



zhangmengjlu (<http://blog.csdn.net/zhangmengjlu>)

2013年10月22日 02:31

426