

用户使用手册

CDA 大数据实验室

CDA 数据分析师

公司地址 | 北京市海淀区厂洼街 3 号丹龙大厦 B 座 3018 室

目录

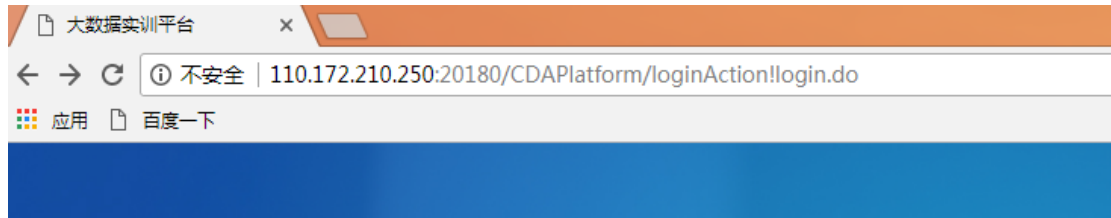
| | |
|--------------------|---|
| 第一部分：学生端操作说明 | 1 |
| 一、登陆 | 1 |
| 二、自主式实训 | 2 |
| 三、向导式实训 | 3 |
| 四、数据采集 | 4 |
| 五、答疑平台 | 6 |
| 六、考试测评 | 7 |
| 七、视频教学 | 9 |

第一部分：学生端操作说明

一、登陆

1、打开 Google 浏览器，输入网址：

<http://110.172.210.250:20180/CDAPatform>



2、输入用户名和密码



3、进入主界面



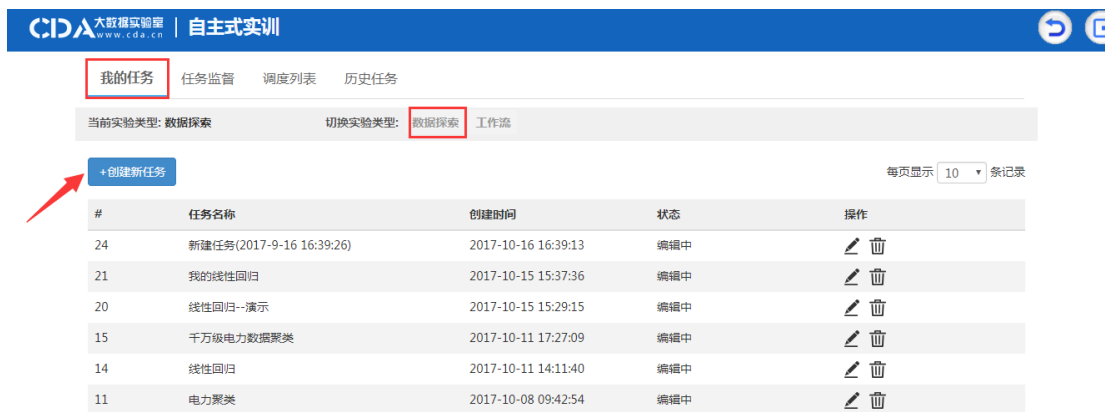
二、自主式实训

自主式培训模块中包括数据探索和工作流 2 种实验类型。该功能可在 windows 系统下实现大数据的多种操作，很大程度上降低了学生自主练习和老师科研的门槛，学生可在没有老师参与的情况下，自主完成整个项目案例操作。

点击自主式实训模块



创建新任务



数据探索包含数据源、数据处理、统计分析、机器学习、数据导出、可视化等功能组件，可实现聚类、分类、回归、推荐等机器学习算法的可视化使用，这些组件都已封装好，可直接使用，有助于老师教学和学生自主练习。各个节点可以单独运行，但必须依赖于上一个节点的输出结果。编程基础薄弱的学生也可以进行大数据实践。

数据探索的使用请参照如下操作说明：

(1) 数据探索操作说明之线性回归

(2) 数据探索操作说明之协同过滤电影推荐

workflow部分可进行自主编程实验，实现诸如【数据获取—>数据处理—>数据建模—>模型评估—>模型应用—>导出结果】等数据挖掘的一个完整流程。包含流程控制、通用大数据处理、导入导出等功能组件，可以实现 Hive、Spark、Map Reduce 等自主式代码编程，以及当前流行的 Sqoop 导入/导出代码实现。

工作流的使用请参照如下操作说明：

(1) 工作流操作说明之电力聚类

(2) 工作流操作说明之电影推荐

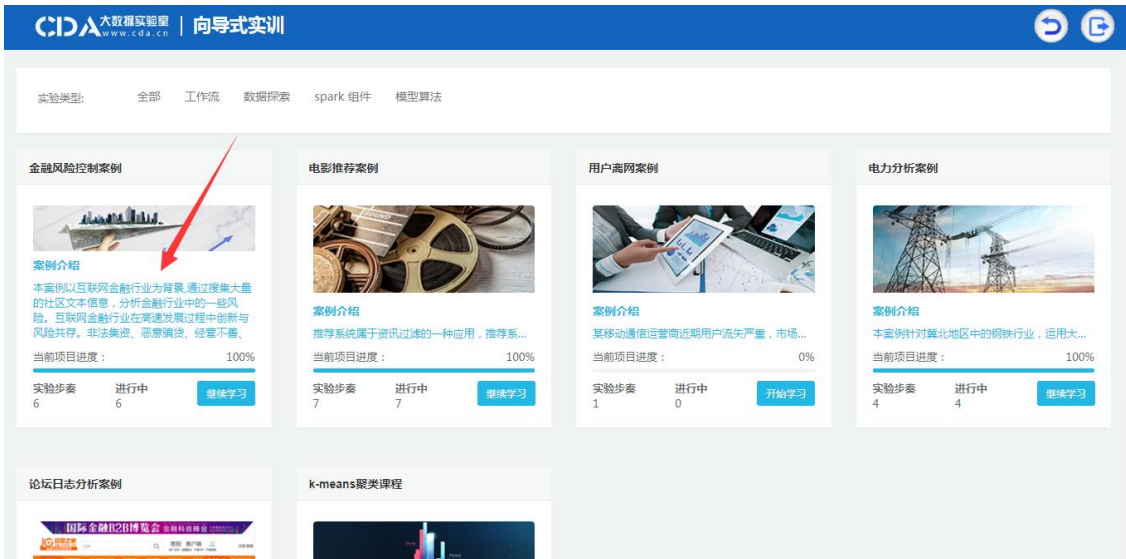
三、向导式实训

该模块包含大量企业真实大数据项目案例，我们将这些案例按步骤拆开，代码内置于模块中，并配备了详细的实验前导、教学内容、操作步骤说明。自动生成实验报告，导出报告，老师可对学生的实验报告进行评分。学生可以在本模块中独自学习大数据案例，学习知识点的使用以及完整的大数据分析流程，感受不同行业的大数据应用环境。学生可以对实验的操作过程进行修改，因为设有重置实验功能，随时可以还原任务。并且随着和学校的不断合作，这些企业案例库会越来越丰富。

点击向导式实训模块



师生可以根据需要选择案例进行学习。



以金融风险控制案例为例，开始实验。

案例的使用请参照：向导式案例操作说明之金融风险控制

四、数据采集

勾选想要爬取的淘宝、微博、聚划算等数据源，填写搜索关键词和采集数量，点击开始采集，即可实现关键词的定向爬取，获得想要的的数据，数据可以下载。爬虫功能可以解决高校老师或学生手头缺乏可用的有效数据的问题，同时爬虫功能还可以激发学生的学习兴趣，使得学生独立进行学习和提升。

点击数据采集模块，进入采集界面。



选择“电商”中的“淘宝下拉词”、“淘宝宝贝属性”选项，设定采集条数为 1000，编辑关键词为“手套”，然后点击“开始采集”。

数据源

电商: ☐ 淘宝下拉词 ☐ 淘宝宝贝属性 ☐ 淘宝宝贝评价 ☐ 聚划算-量贩团 ☐ 淘抢购微博: ☐ 新浪微博

采集条数: 1000

关键词: 请输入搜索关键词.....

开始采集

筛选: 爬虫状态

数据来源

任务提交时间

关键词搜索

确定

| <input type="checkbox"/> 数据来源 | 关键词 | 采集条数 | 开始时间 | 任务进度 |
|----------------------------------|-----------|------|---------------------|------------------------|
| <input type="checkbox"/> 淘抢购 | 淘抢购监控 | 100 | 2017-11-15 14:41:52 | <div><div></div></div> |
| <input type="checkbox"/> 聚划算-量贩团 | 聚划算-量贩团监控 | 100 | 2017-11-15 14:41:52 | <div><div></div></div> |
| <input type="checkbox"/> 淘宝宝贝评价 | 手套 | 100 | 2017-11-15 14:41:52 | <div><div></div></div> |

刷新即可看到采集的进度，采集完成后点击关键词下的“手套”，即可看到从网上爬取的关于手套的信息，并且可以点击“下载 CSV”将采集的数据下载到本地。

数据源

电商: ☐ 淘宝下拉词 ☐ 淘宝宝贝属性 ☐ 淘宝宝贝评价 ☐ 聚划算-量贩团 ☐ 淘抢购微博: ☐ 新浪微博

采集条数: 1000

关键词: 请输入搜索关键词.....

开始采集

筛选: 爬虫状态

数据来源

任务提交时间

关键词搜索

确定

| <input type="checkbox"/> 数据来源 | 关键词 | 采集条数 | 开始时间 | 任务进度 |
|---------------------------------|-------|------|---------------------|------------------------|
| <input type="checkbox"/> 淘宝宝贝属性 | 手套 | 1000 | 2017-11-15 14:47:12 | <div><div></div></div> |
| <input type="checkbox"/> 淘宝下拉词 | 手套 | 1000 | 2017-11-15 14:47:12 | <div><div></div></div> |
| <input type="checkbox"/> 淘抢购 | 淘抢购监控 | 100 | 2017-11-15 14:41:52 | <div><div></div></div> |

数据源: 淘宝宝贝属性 关键词: 手套 采集条数: 1001 采集时间: 2017-11-15 14:47:12

下载 CSV

| 宝贝ID | 主图 | 标题 | 价格 | 销量 | 所在地 | 旺旺 | 店铺类型 | 信誉 |
|--------------|---|--------------------------------|------|-------|--------|---------|------|-----|
| 541470942175 |  | 情侣触屏毛线手套男女冬季针织加绒保暖韩版学生骑行五指户外防滑 | 19.8 | 83552 | 浙江 杭州 | 富柜旗舰店 | 天猫 | 1金冠 |
| 559718112096 |  | 南极人手套男冬季保暖加厚防风触屏皮手套男士户外开车骑行摩托车 | 29.9 | 49804 | 浙江 杭州 | 小龙服饰专营店 | 天猫 | 3皇冠 |
| 40763353135 |  | 手套女冬季骑行摩托车保暖加绒加厚滑雪棉手套男冬防风骑车皮手套 | 19.9 | 48752 | 浙江 杭州 | 宏文服饰专营店 | 天猫 | 1金冠 |
| 558829580336 |  | 韩版触屏毛线手套男女冬季开车骑车保暖加绒加厚情侣学生全指防滑 | 19.8 | 47576 | 上海 | 卡兰薇旗舰店 | 天猫 | 3金冠 |
| 41357695556 |  | 触屏男士手套冬季皮手套男加绒加厚保暖防风防水骑摩托车骑行手套 | 15.6 | 45237 | 河北 石家庄 | 卓幻美冬旗舰店 | 天猫 | 1金冠 |

五、答疑平台

该模块和我管之家论坛相连接，CDA 数据分析交流社区在国内最大的经管在线社区经管之家的基础上成立，涵盖计量经济学与统计软件、商业数据分析、大数据技术、机器学习技术、数据交流中心以及 IT 基础等领域内的常用软件和学科，汇聚了国内外成千上万名数据分析领域专业人才，并留下了数百万条分享和讨论贴，成为国内最活跃的数据分析学习资源分享和专业问答社区之一。并经常组织线下沙龙分享和俱乐部交流活动。

点击答疑平台模块。



根据问题类型，选择相应的版块，可以查询到相应的答案，也可随时发帖留言进行咨询，会有专门的助教来给您解答。





公告: “论坛贵宾帐号” 享受贵宾服务, 性价比超高, 让下载更轻松

2011-9-9



增加论坛币的方法 论坛制度一览表 ... 2 3 4 5 6 .. 3534

admin: 2008-12-18 发表 | 最后回复: nihaoyaxiaohu 18 分钟前

8793100 35336



CDA数据分析就业学院——塑造辉煌人生的摇篮!!! ... 2 3 4

人脉引爆点: 2016-8-11 发表 | 最后回复: joemok1991 2 小时前

38339 39



经管之家求数据分析助教(全职or实习)一枚!!! ... 2 3

上官燕503: 2017-10-24 发表 | 最后回复: Yvette_Pan 4 小时前

2968 21



【万元征集优秀案例】大数据、商业数据分析、机器学习与深度学习领域应用案例 ... 2 3 4 5 6 .. 23

客初: 2017-7-18 发表 | 最后回复: 客初 8 小时前

28458 221



为什么要学习 R 语言_R语言计量经济学应用零基础入门 ... 2 3 4 5 6 .. 10

资料狂人: 2017-8-2 发表 | 最后回复: 店小二 上菜 1 小时前

5874 95

六、考试测评

为了加强学生的学习效果, 针对各个知识点, 我们专为学生开发了考试测评系统, 其中包括 SPSS、SAS、R、Python、Hadoop、Excel、数据挖掘、统计概率等。学生可以选择题型、涉及的相关软件、理论 or 应用、课程分类、学科分类、行业及应用等方向, 再选择试卷的难易程度(易、中、难), 便可以进行专项知识点的练习和学习。

点击考试测评模块。



选择想要测评的题型方向, 例如点击 R 版块。



SPSS



SAS



R



Python



hadoop



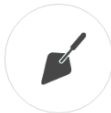
MATLAB



EXCEL



统计概率



数据挖掘



行业



CDA LEVEL1



CDA LEVEL2

选择试题的类型和难度，点击“开始答题”，即可开始 R 相关知识的考试测评。

题型

单项选择题 多项选择

软件

Excel SPSS SAS

理论OR应用

理论 软件应用 行业应用

课程分类

CDA一级 CDA二级(建模)

学科分类

机器学习_数据挖掘_算法 大数据相关

行业及应用

金融(银行、互联网金融)

已选条件：

| 试卷名称 | 来源 | 试题编号 |
|-------------|----|------|
| 试卷-R软件（易） | | 15 |
| 试卷-R软件（中） | | 17 |
| 试卷-R软件（难） | | 18 |
| 试卷-R基础试题（一） | | 112 |
| 试卷-R基础试题（二） | | 114 |

题库 > R软件（易）

试卷-R软件（易）

答题时间：60分钟

试卷信息

题目总数：50道
试卷分类：R
满分：250分

注意事项

- 1、请在规定时间完成试卷内全部题目。
- 2、所有题目可通过答题卡返回修改，点击提前交卷后提交，将无法继续答题，请谨慎提交。
- 3、请认真答题，独立完成。

开始答题

考试结束后，点击提交后有“试卷评估”和“答案解析”，可查看出错题的相关答案解析。

CDA 大数据实训平台 | 考试测评

试卷评估 | 答案解析

R软件（易）

1

在R软件中，运行代码：seq(1,10,length=3)，所得结果是？：

正确答案：D,你的答案：B；

☐ A、1 10

☐ B、1 4 7 10

☐ C、1 5 10

☐ D、1.0 5.5 10.0

提问

错题反馈

七、视频教学

学生在课下可以独自在该模块进行学习，也可以将这些视频设置为通识选修课来辅助教师进行教学，充分调动学生的积极性，增强学生对相关知识点的学习和理解。

点击视频教学模块



选择相关课程进行学习



【视频】CDA大数据分析师前
导视频

3714人学习



【公开课】走进人工智能世界
——机器学习与深度学习

4292人学习



python编程

8719人学习



大数据试听视频

4918人学习



CDA : SPSS软件及应用

7989人学习



【就业班独家发布】数据分
析师学习之路

9390人学习



DEA模型与MAXDEA软件应
用

14025人学习



MySQL初级视频教程

4752人学习



数据挖掘视频教程

8079人学习



SQL Server 2008 高级教程

2370人学习



SQL Server 2008 初中级教程

2941人学习



SAS初级教程

14152人学习



R语言高级教程 (在线体验部
分章节)

11429人学习



R语言初级教程 (在线体验部
分章节)

22473人学习



SPSS初级教程

26790人学习