

北京大学临床数据科学家 Python 数据分析研修班

项目编号：2023-81(校)

招生简章



【项目内容】

本课程聚焦面向健康医疗数据的 Python 语言编程，通过实践加深对相关算法原理的认知，提升解决实际问题的动手操作能力，从而提高学员的数据科学素养，旨在培训学员：

- 掌握 Python 语言编程基础和编程思想，熟悉基础语法，能够使用编程工具进行 Python 编程；
- 能够运用 numpy 进行科学计算；运用 pandas 进行表格处理；运用 statsmodels 进行统计分析；运用 matplotlib 进行作图；运用 sklearn 进行机器学习建模；
- 培养学员结合健康数据科学理论与实践的能力，通过真实健康医疗数据分析的实际项目，深入了解编程和机器学习的相关原理，以及锻炼解决问题的实际动手能力。

【办学单位】

北京大学健康医疗大数据国家研究院

【招生对象】

本次课程为希望提升数据分析能力的临床工作者、医疗健康其他领域工作者量身打造，也适用于健康数据科学相关从业者。包括但不限于：

- 临床和非临床专业的医学研究生：为自己的研究课题增加新内容，运用 Python 补充新的数据分析结果和新发现，丰富毕业论文；
- 医院的科研人员：进一步发掘真实世界健康医疗大数据的隐藏价值，运用 Python 对医院现有的数据进行处理、分析、建模、可视化和预测，产出高质量创新科研成果；
- 掌握了统计分析方法的数据分析师：进一步了解人工智能、机器学习的新工具，使用 Python 建立预测模型，变传统关联性分析为预测；
- 科研管理人员：了解数据驱动的研究新思想，了解人工智能从业人员的工作流程，了解医学与人工智能合作的挑战和机遇。

【培训地点】

腾讯会议线上授课+北大医学部线下授课

【培训时间】

2023 年 11 月 24 日-12 月 2 日。

晚 18:30-21:30，仅 12 月 2 日周六全天线下课程，共计 20 学时。

见具体课程安排。

【课程安排】

时间	主题	案例分析/操作练习
第 1 次课线上 11 月 24 日周五 18:30-21:30	<ul style="list-style-type: none"> Python 的基础语法、运算符、语句和数据类型 <ul style="list-style-type: none"> Python 语言及生态圈 环境搭建：Anaconda 的下载与安装，Jupyter、Spyder 等界面及功能介绍 基础语法、运算符和语句 成功运行第一个程序 数据类型：字符串、列表、元组、字典、时间日期 	<ul style="list-style-type: none"> 编程环境配置 第一个程序 Hello World!
第 2 次课线上 11 月 26 日周日 18:30-21:30	<ul style="list-style-type: none"> Python 的函数、模块和文件输入输出 <ul style="list-style-type: none"> 函数：定义与调用，参数及作用域 模块：定义与调用 文件输入输出 读写电子健康档案数据：表格、图像、信号、文本、图/网络 	<ul style="list-style-type: none"> 读写电子健康档案数据
第 3 次课线上 11 月 28 日周二 18:30-21:30	<ul style="list-style-type: none"> Python 的 numpy 科学计算、pandas 表格处理、statsmodels 统计分析与 matplotlib 数据可视化分析 <ul style="list-style-type: none"> 利用 numpy 创建、操作数组，并进行数学和统计运算 利用 pandas 创建、操作表格，并进行表格 	<ul style="list-style-type: none"> 电子健康档案数据预处理、统计分析与可视化分析

时间	主题	案例分析/操作练习
	清洗、查询和整合 <ul style="list-style-type: none"> ■ 利用 statsmodels 进行统计检验 ■ 利用 matplotlib 作图，美化论文插图 	
第 4 次课线上 11 月 30 日周四 18:30-21:30	<ul style="list-style-type: none"> ● Python 的 sklearn 机器学习 <ul style="list-style-type: none"> ■ 机器学习简介 ■ 利用 sklearn 建立回归、分类、聚类模型，并进行模型筛选和评估 ■ 基于电子健康档案数据建立风险预测模型 	<ul style="list-style-type: none"> ● 基于电子健康档案数据建立风险预测模型  <pre>## 逻辑回归模型建立 lr_model = LogisticRegression(max_iter=100) lr_model.fit(X_train, y_train)</pre>
第 5 次课线下 12 月 2 日周六 8:30-17:00	<ul style="list-style-type: none"> ● Python 健康数据分析实战训练 <ul style="list-style-type: none"> ■ 课程回顾与答疑 ■ 临床医学公开数据库介绍 (PhysioNet、MIMIC、PIC) ■ 如何利用公开数据库开展医学研究 ■ 利用 Python 和公开数据库建立临床预测模型 (代码量 300 行以内) 	<ul style="list-style-type: none"> ● 利用 Python 和公开数据库建立临床预测模型 

【主讲教师】

洪申达 北京大学健康医疗大数据国家研究院 副研究员等

【培训证书】

学员完成全部课程学习，考评合格，将颁发北京大学结业证书，加盖北京大学钢印。

【报名缴费】

1. 费用：3600 元/人

注：学员完成全部课程学习，考评合格，将颁发北京大学结业证书，加盖北京大学钢印。

2. 报名缴费流程：

1) 报名：扫描下方二维码填写报名信息

2) 缴费：

【路径一】：报名后，北京大学医学部缴费平台发送缴费链接到参会代表注册邮箱，请点击缴费链接交费；

【路径二】：报名后，将培训费用于截止日期前全额对公汇入以下账户：

账户名称：北京大学医学部

开户银行：工商银行北京东升路支行

账 号：0200 0062 0908 9112 565

用途（汇款时请备注）：健康数据科研+学员姓名

注：将缴费凭证（转账截图/汇款凭证照片）通过微信或邮件发送给报名联系人崔老师（联系方式见下），确认报名成功。

3. **报名缴费截止时间：2023 年 11 月 22 日。**

注：小班授课，人数有限，额满即止（以缴费顺序为准）。



（扫描二维码，填写报名信息）

【联系方式】

报名联系人：崔老师

电话：010-82806531, 13810687742 (微信号：16759657)

邮箱：pku_dshm@bjmu.edu.cn

地址：北京市海淀区学院路 38 号北京大学医学部校内北大医学科技楼 (100191)

2023 年 10 月