# 《创新实践》课程项目介绍

## 个人简介：

王然，男，本科就读于浙江大学竺可桢学院混合班，博士毕业于浙江大学理学院数学系概率论与数理统计专业。先后就职于杭州电子科技大学经济学院统计系、计算机学院智能所。目前主要研究方向：统计大样本理论、数据挖掘、机器学习。在杭电任教期间，担任过统计学、推断统计学、随机过程、非参数统计、统计计算、R语言程序设计、C语言程序设计、数据结构等多门统计、计算机专业的核心专业课程，并获得2014年微课校级二等奖、2017年青年教师技能竞赛校一等奖。目前承担浙江省重点研发计划一项，参与国家级和省部级项目若干项。曾指导学生参加天池大数据竞赛并获得广东公共交通大数据竞赛优秀导师奖项。

## 项目简介：

目前，实验室的研究方向主要聚焦于数据挖掘在垂直行业内的应用。我们致力于打通数据分析流程的所有环节，培养学生在数据分析项目的研发过程中，能够掌握从数据采集、数据清洗，到统计建模、模型评估等分析方法，从而具备数据挖掘工程师的实战能力。

开发环境主要是：Python、R等语言，也会用到部分其他语言。项目来源主要有两个方面：一是实验室在做的纵向和横向项目，二是互联网平台上的大数据竞赛。我们会结合学生的学习情况，进行相应的项目安排。

以下是部分项目的内容简介：

1. 基于数据挖掘的信贷审批预测
2. 调研银行、第三方征信机构的基本信贷审批模型，熟悉信贷模型建立的发展历史和基本要素。
3. 分析用户的信贷历史数据，从性别、婚姻状况、受教育情况、亲属信息、收入、贷款申请额度和周期等信息中，提取有效的用户信用区分度信息。
4. 构建决策树等模型预测用户的借贷是否会被审批，将信贷审批从人工主动判别转向程序自动判别。
5. 基于数据挖掘的商品销量预测
6. 对商品的原始销售数据进行清洗和补全。变量有：商品标识、重量、脂肪含量、展位曝光率、商品类别、价格、店铺ID、店铺成立年份、店铺大小、店铺位置、店铺类型、商品销量。
7. 对变量进行相关性分析，使用多种数据挖掘模型提取关键变量，并构建统计模型，对不同的模型进行评估。
8. 构建集成方法等模型预测商品可能的销量，通过模型调参，提高准确率。
9. 基于面部图像的年龄预测方法研究
10. 学习如何提取面部图像特征。
11. 利用机器学习方法构建基于面部图像的年龄预测方法。
12. 图像识别、定位等相关方法的研究。

邮箱：[wangran@hdu.edu.cn](mailto:wangran@hdu.edu.cn)

QQ：1907303