

# 大数据竞赛题 – 高职组

## 赛题背景

随着大数据技术在各行各业的渗透，通过大数据技术进行人力资源管理，成为众多大型企业的首选。现如今，每一家公司都面临着人力资源的多重挑战，企业在培养优秀员工过程中，将会耗费巨大的人力、物力与财力，如何才能在多重因素的权衡中，招聘到更加优秀且能持久发展的员工，成为众多企业研究的热点问题。在本项目中，你要对多地超过万名在职人员进行实际调查，通过对数据的综合分析和建模，掌握影响员工离职的重要因素，并建立预测模型，为企业人力资源管理提供更加科学且合理的依据。

## 赛题任务

**任务一：**分析员工的工作经验与企业对员工的培训时长对员工离职意向的变化趋势，并探索培训时长对员工离职意向的影响因素。

**任务二：**协助企业建立预测模型，预测员工离职意向，并评价模型的预测效果。

**任务三：**根据分析结果，为企业提供人才招聘的建议。

## 答案提交标准

请就本项目完成一份分析报告，在报告中应选择合适的统计模型或机器学习算法对各任务做出解答。解答应包含完整的分析过程，并合理使用表格和图表来呈现分析结果。选手提交的文档应符合报告的一般形式，并以 PDF 格式保存，文件名为选手代码（例如：“DS12119999.pdf”）。

## 赛题数据

竞赛数据集为：多地员工离职意向调查数据集，数据的描述如表 1 所示，包含 10109 个样本，13 个变量，各变量的名称、解释及取值说明如表 2 所示。

表 1. 数据简要说明

数据名称	样本量	变量数量
多地员工离职意向调查数据集	10109	13

表 2. 变量说明

变量名称	变量解释	取值
X0	受调查人员 ID	受调查人员唯一标识，数值
X1	城市代码	城市代码标识
X2	城市发展指数	城市规模标识，数值
X3	性别	性别标识
X4	相关经验	受调查人员工作前相关经验，
X5	注册大学类型	受调查人员已注册大学类型
X6	学历	受调查人员学历
X7	工作经验	受调查人员相关工作经验，数值，单位：年
X8	所在公司规模	受调查人员目前所在公司规模，数值，单位：人
X9	所在公司类型	受调查人员所在公司类型
X10	上一份工作和这份工作之间间隔时长	受调查人员上一份工作和当前工作之间间隔的市场，数值，单位：年
X11	培训时长	受调查人员在当前公司受培训时长，数值，单位：小时
X12	离职意向	受调查人员离职意向，布尔值，1 表示计划离职，0 表示无离职计划

大赛作品上传说明

最终提交的文件，为 1 个 pdf 文档，以选手代码作为文件名（例如“DS12119999.pdf”），并按以下要求进行提交。

- 1. 作品中不可以包含参赛选手本人姓名和学校名称，不可与其他人作品雷同。
- 2. 打开浏览器，在地址栏内输入下面的地址，然后在打开的网页上输入登录信息，然后点登录按钮。

上传网站: <http://59.111.61.168>

用 户 名: upload

密 码: 123456



3. 登陆后，点击网页中部文件浏览器的工具栏上，点击【上传】按钮。



4. 点击选择文件-选择自己做好的作品，然后点击【上传】按钮。



5. 等待上传作品。



6. 待上传文件框消失后，点击【刷新】按钮。



7. 最后请查看列表中是否有您上传的文件，确认无误后，**请截图或者用手机拍照留存（作品在列表中的图片）**，点击右上角**【作品上传】**的位置，点击注销，然后退出浏览器即可。



8. 如果上传遇到问题，请及时联系赛务组负责老师解决。