

# 数据挖掘课程期末项目介绍

## 项目概览

本项目旨在通过实践应用数据挖掘技术,培养学生的研究、编程、报告撰写和公开演示能力。每个小组将选择一个自选的题目进行**可行性调研、编程实现、报告撰写和课堂展示**。项目的评分要点包括解决方法的**完整性、创新性、工程复杂度、结果分析的逻辑性**,以及**最终报告与展示的质量**。具体项目题目不限,但每个小组可以参考以下几个方向给出的参考提示。可以自行搜索并下载合适的数据集(例如阿里云天池、Kaggle等),鼓励使用多个的数据集进行更广泛的分析。请注意,本项目严禁抄袭,验收时会进行代码和报告的查重与检测。

## 项目代码与报告提交

每个小组通过 Blackboard 提交一份代码实现的 Jupyter Notebook (或其他常用的数据挖掘工具)和一份项目总结报告。代码实现要尽量详尽,包含运行结果,代码量不限;报告使用 ACM sigconf 模板(即以 KDD 会议论文的形式,双栏),不超过 7 页。

报告提交的截止日期是第 16 周星期三(6 月 5 日),23:59 PM。报告不接受任何理由的补交或缓交,请同学们注意尽早完成并提交。Blackboard 上有 3 次提交机会,希望同学们提交前能好好检查完再上传,如果真的有需求重新上传,请联系助教处理。逾期不收。

## 课堂展示要求

**形式:** 小组在课堂上进行展示。

**时长:** 每组 10 分钟。

**日期:** 第 16 周星期三(6 月 5 日)讲座和实验课程期间。

**顺序:** 按照 <https://docs.qq.com/sheet/DSHdvWWIWUnVmY0V1?tab=BB08J2> 中的小组 1 至小组 15 的顺序进行。

**注意事项:** 将展示文件发送到课程的微信群中的 TAs,文件将被上传到讲台电脑。除此之外需要的任何设置时间将从您的分配展示时间中扣除。

**协助:** 课程助教 Kaiyi Huang、Tao Sun 和 Wei Hu 会解答关于展示任务的任何疑问。

## 参考题目

### 方向一:快递员到达检测系统(考察重点:传感器数据分析;事件检测;设计优化)

**参考数据集:** aBeacon: Alibaba Beacon System of Couriers' Arrival Detection (<https://tianchi.aliyun.com/dataset/76359>)

**做法参考提示:** 对数据集进行必要的清洗和预处理。通过分析快递员的多种特征(如手机信号 RSSI, GPS 信息),对快递员的到达和离开事件进行检测。根据快递员轨迹信息,进行聚类分析和目的地预测。分析不同移动设备对到达和离开检测的影响以及优势、劣势。这个项目将帮助理解如何应用数据分析来改善实际的快递业务流程。

### 方向二:智慧交通流量与时间预测(考察重点:流量预测;模式识别;事件检测)

**参考数据集:** 智慧交通预测数据集 (<https://tianchi.aliyun.com/dataset/1079>)

做法参考提示：探索相关数据集，以分析和预测城市交通流量和通行时间。对数据集进行必要的清洗和预处理。通过分析交通流量数据和相关时间信息，发现交通流量的关键趋势和模式，并尝试构建模型来精准预测各关键路段在某个时段的通行时间，实现对交通状态波动起伏的预判。这个项目将帮助理解如何应用数据分析来改善交通管理和规划。

### **方向三：广告点击行为分析与投放策略优化（考察重点：行为分析；模型训练；策略优化）**

参考数据集：Avazu Click-Through Rate Prediction Dataset  
(<https://tianchi.aliyun.com/dataset/147617>)

做法参考提示：探索广告点击流数据相关数据集。对数据集进行必要的清洗和预处理。分析广告展示的时间、用户的设备类型、应用类型等数据，以识别影响点击率的关键因素。通过构建多种统计模型或机器学习算法，预测用户对特定广告的点击行为，从而评估广告表现和优化广告投放策略。这个项目会帮助理解并应用数据分析在在线广告领域的实际应用。

### **方向四：工业生产趋势分析与预测挑战（考察重点：时间序列分析；需求预测；可视化）**

参考数据集：FRED Economic Data (<https://fred.stlouisfed.org/>)

做法参考提示：对多个相关的数据进行预处理，包括清洗、标准化等。通过使用时间序列分析方法，将分析数据中的趋势、季节性和周期性成分。此外，尝试构建多种数据挖掘模型，来预测未来的工业生产水平。探讨如何将这些预测应用于经济政策制定和市场策略。最终，解释模型的选择、预测结果的准确性以及模型可能的改进方向。这个项目将帮助深入了解经济数据分析，并掌握预测经济指标的实用技能。