

第八课 机器学习简单介绍

第15节 实战：通过移动设备行为数据预测性别和年龄

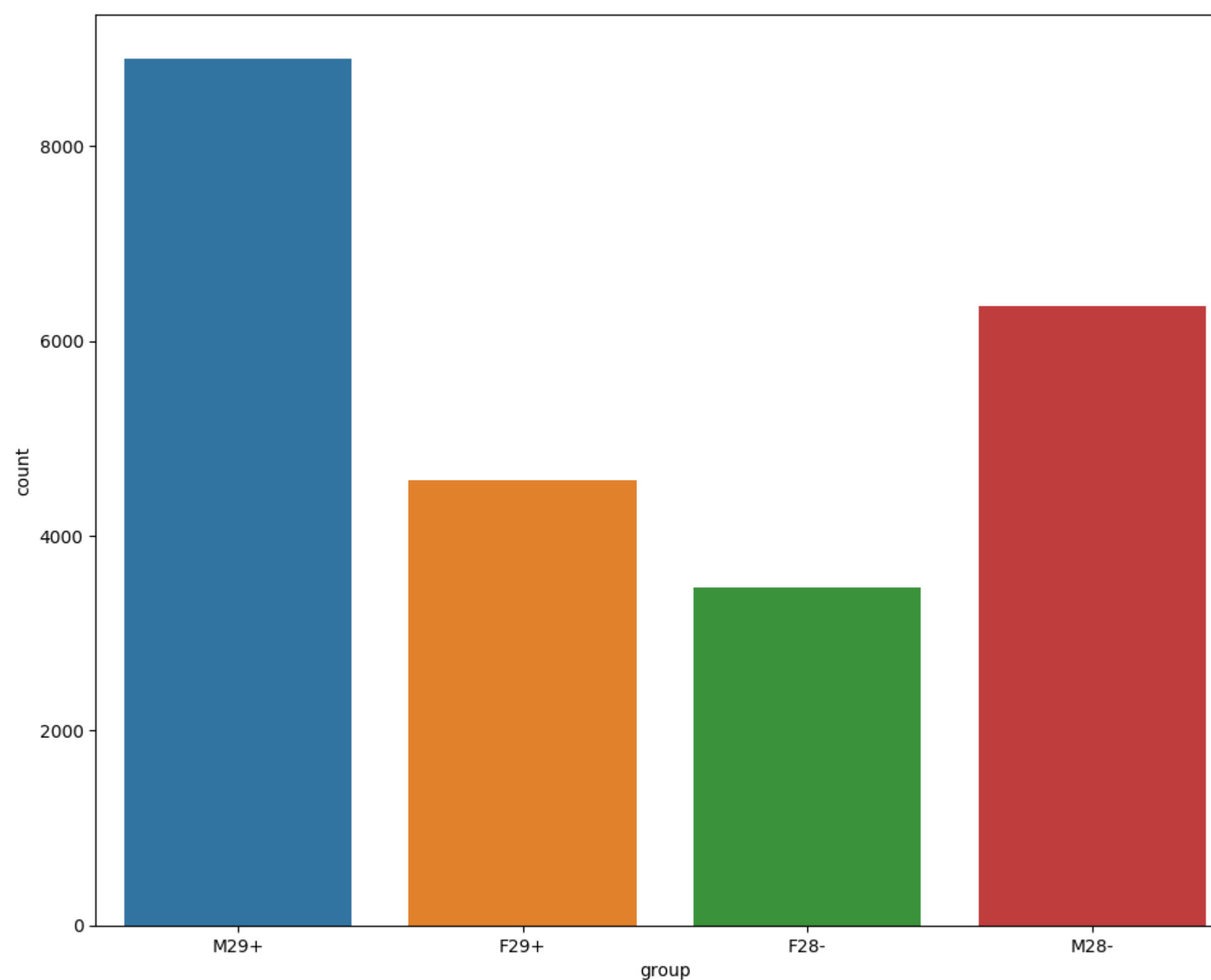
主 讲：Robin

通过移动设备行为数据预测性别和年龄

- <https://www.kaggle.com/c/talkingdata-mobile-user-demographics>
- 通过使用APP的行为对移动用户属性（年龄+性别）进行预测
- 数据及包含~20万用户数据，分成4组，同时提供了用户行为属性，如：手机品牌、型号、APP的使用情况等

- 任务：

1. 数据集准备
2. 特征处理
3. 建模及调参



通过移动设备行为数据预测性别和年龄

步骤:

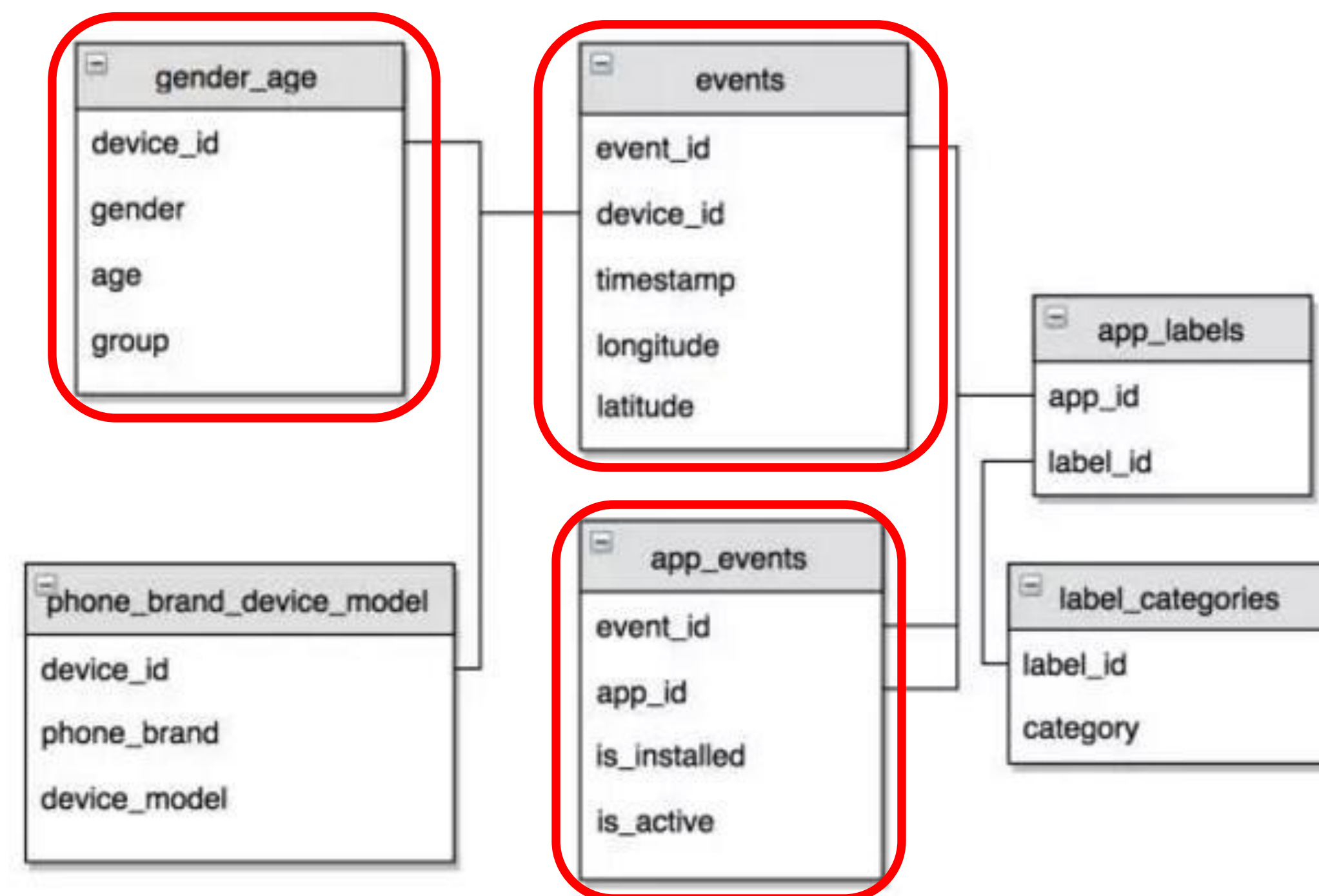
1. 数据解读
2. 特征提取
3. 模型构建及调参
4. 结果比较

通过移动设备行为数据预测性别和年龄

步骤:

1. 解读数据:

- 每个用户对应一个device_id
- 每个用户的行为被记录为一系列的events
- 每个event里面的信息包括该行为发生的时间、地理坐标信息，安装的APP类型等

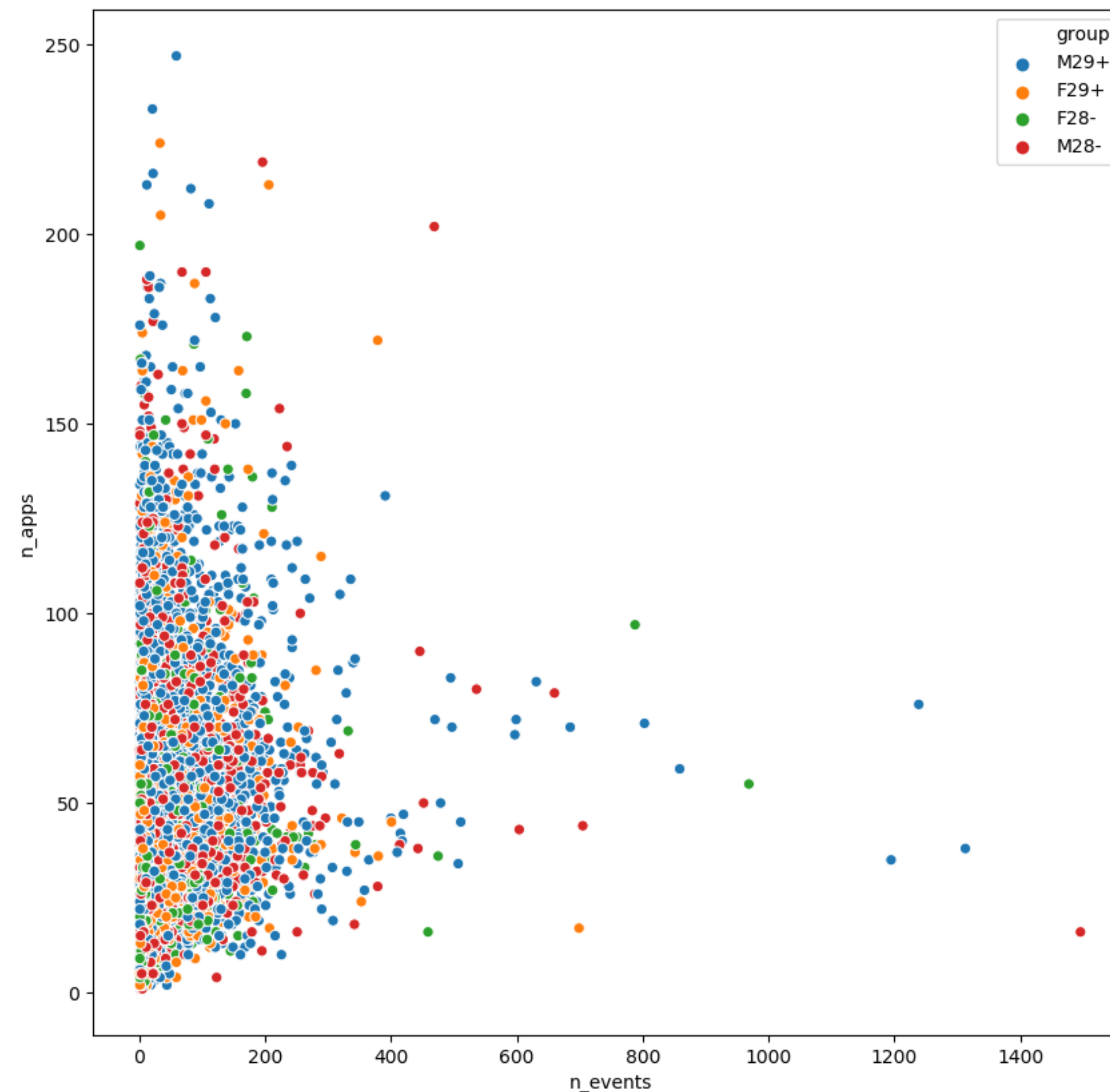


通过移动设备行为数据预测性别和年龄

步骤:

2. 特征提取:

- 根据用户的行为统计
 - event类型的个数
 - 安装app的个数
- 对特征归一化处理



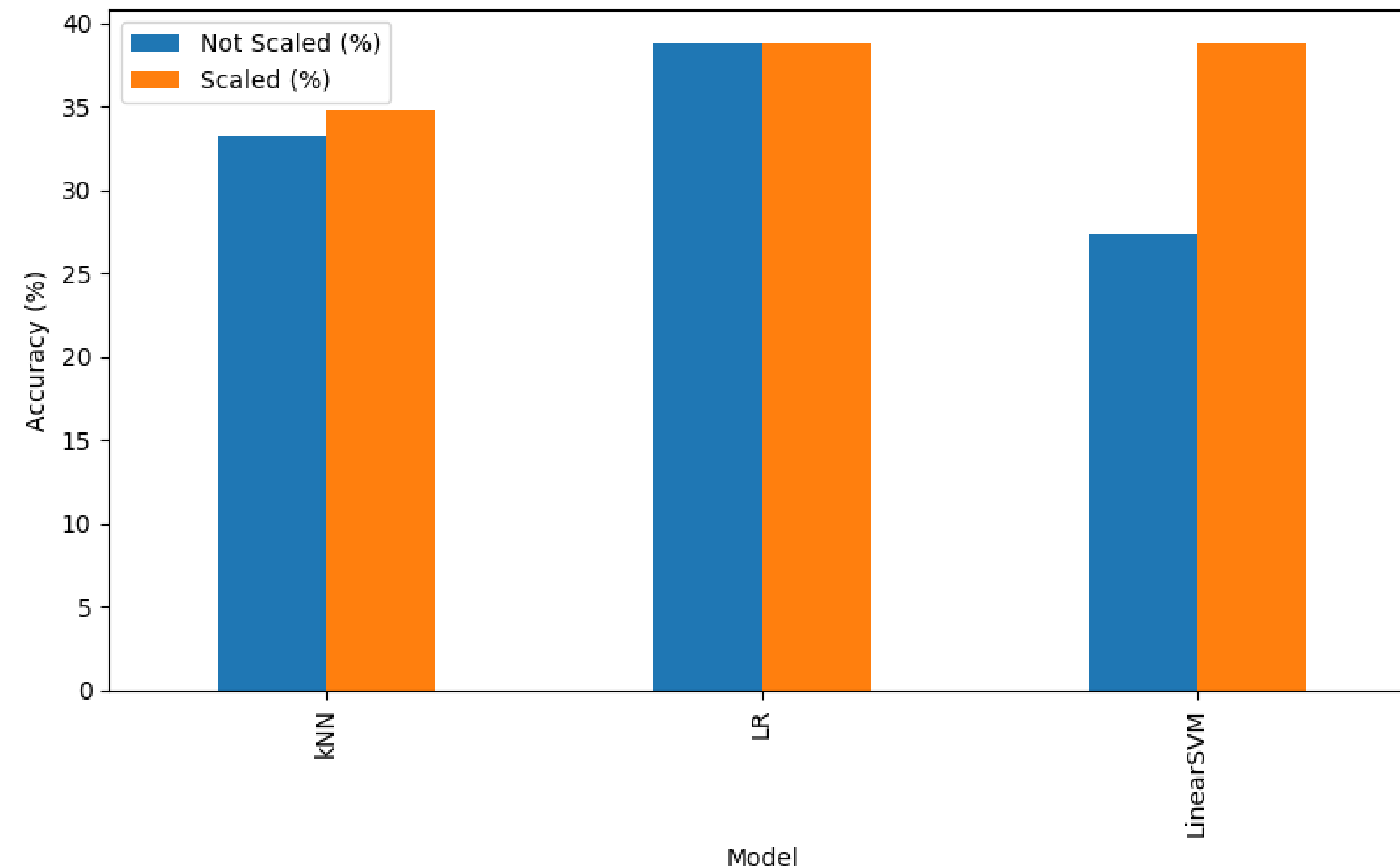
通过移动设备行为数据预测性别和年龄

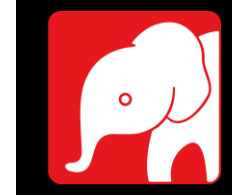
步骤:

3. 模型构建及调参:

- 使用3个模型
 - K近邻距离, kNN
 - 线性SVM, LinearSVC
 - 逻辑回归, LogisticRegression

4. 结果比较





只为遇见明天更优秀的你！