**实验一：特征处理**

1. 实验题目：

**我们提供了在双11促销期间一组商人他们相应增长的买家。你的任务是预测未来这些新买家是否会成为这些商家的忠诚客户。换句话说，您需要预测这些新购买者在6个月内再次从同一商家购买商品的可能性**

1. 实验过程
2. 检测异常样本
3. 缺省值处理情况

**缺省值极多**

**非连续特征缺省值适中**

**连续特征缺省值适中**

**缺省较少**

**解决方法如下：**

**均值，众数，中位数等填充办法**

**模型填充办法**

**拉格朗日插值法**

**在保证原有数据样本分布不变情况下进行随机填充**

1. 数据采样

**从负样本中抽取部分样本出来和正样本结合（欠采样，容易造成信息损失）**

**正样本重复若干次**

**代价敏感学习Cost Sensitive Learning**

**SMOTE即合成少数类过采样技术**

1. 实验分析

特征学习就是从数据集中自动抽取特征或者是表示方法，特征工程就是我们人为的对数据进行处理，然后就可以得到我们认为的、适合实验模型的样式。

所以先要从数据集中下载实验所需要的数据，然后再对这些数据进行处理和分析。

特征分析是从数据库中的一组数据中提取出关于这些数据的特征式，这些特征式表达了该数据集的总体特征。如营销人员通过对客户流失因素的特征提取，可以得到导致客户流失的一系列原因和主要特征，利用这些特征可以有效地预防客户的流失