特征处理

1. 数据预处理：

1.检测异常样本：除去一些极端数据

2．缺省值处理：用中位数，众数，平均数的方法

模型填充，拉格朗日定理等方法。

数据采样：

1．从负样本中抽取部分样本出来和正样本结合

2．正样本重复采用多次

1. 特征工程
2. 特征分类：有数据型，类别型，时间型，文本型，统计型，组合特征型，
3. 特征处理与分析：特征归一化、离散化，特征降维，特征选择。
4. 特征选择：目标是寻找最优特征子集。特征选择能剔除不相关或冗余的特征，从而达到减少特征个数，提高模型精确度，减少运行时间的目的。另一方面，选取出真正相关的特征简化模型，协助理解数据产生的过程。