SEMMA

一、Sample

二 、数据准备

Modify

1. iv:分类与连续分别筛选
2. 信息冗余

分类变量：（1）概化（10类，对于每一个新类，）原来共有54类，现在分为10类，每一类的样本数要求基本相等 （2）WOE

连续变量：

Eplore:

1 描述性统计

2、异常值（众数-均值）/四分位距 越大，越有可能有异常

3、缺失值检测与填充（变量的中位数）

三、Model

四、Access