电力窃漏电用户识别

## 一、挖掘目标：

1、归纳出窃漏电用户的关键特征，构建且漏电用户的识别模型

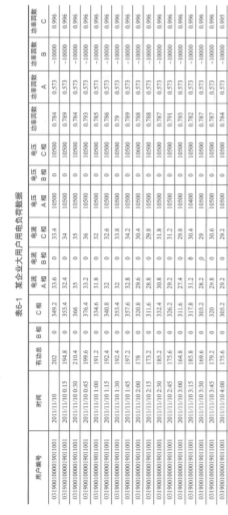
2、利用实时监测数据，调用窃漏电用户识别模型实现实时判断

属于分类与预测的问题

## 二、数据抽取：

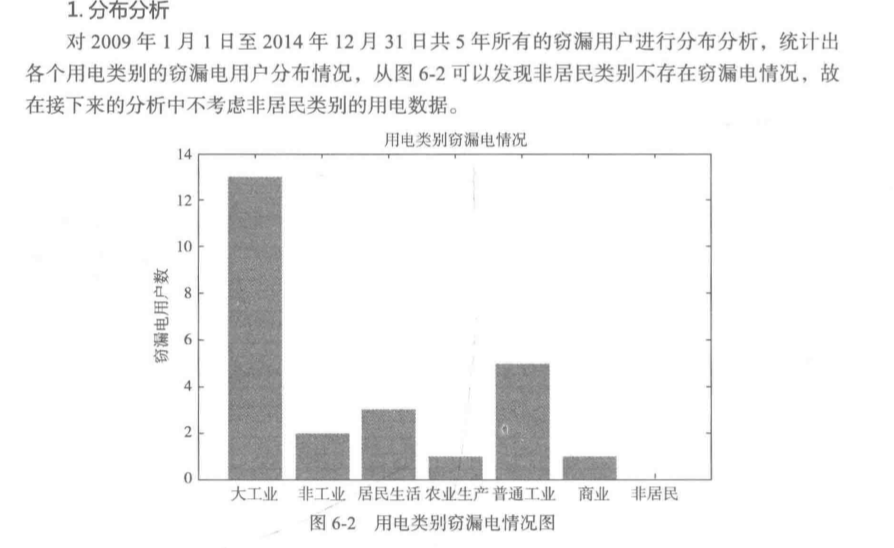
1、从营销系统抽取

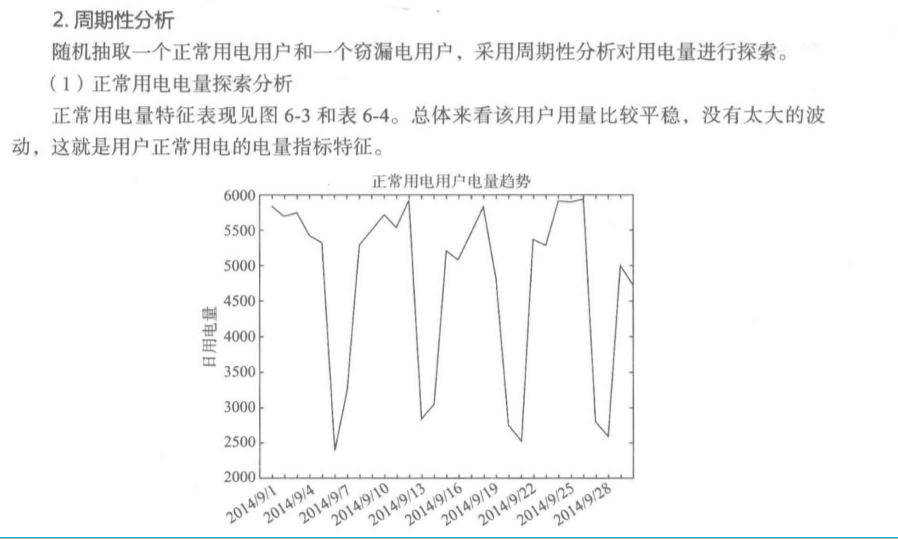
2、从计量自动化系统采集、电量、负荷等

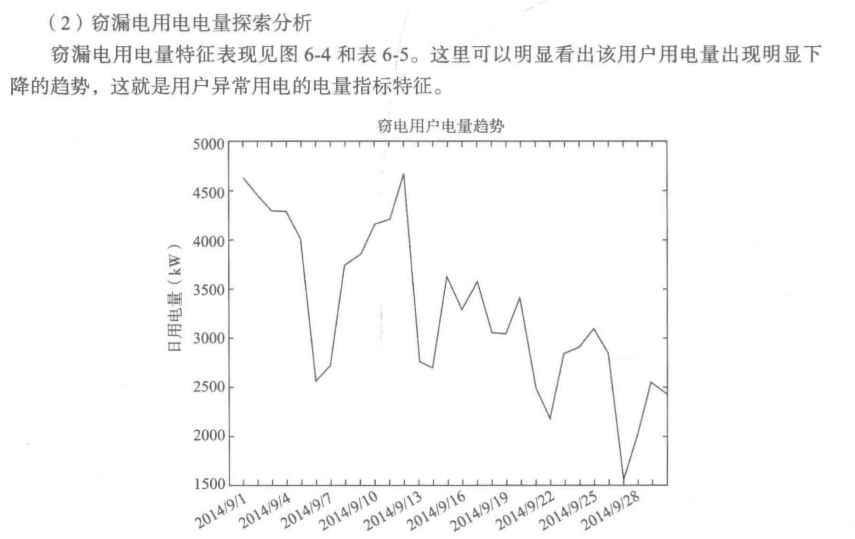


## 三、数据探索：分布分析+周期性分析

**探索与预测无关的数据！！**







分析结论：正常用电到窃漏电过程是用电量持续下降的过程

## 四、数据预处理

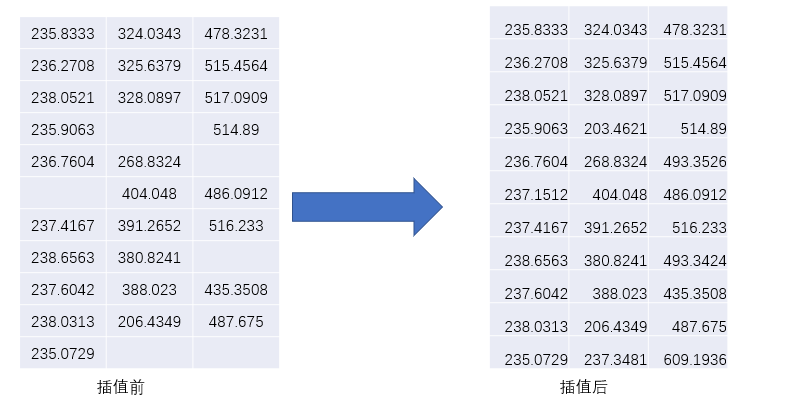
1、缺失值可以删除也可以插补、插补的方法很多，处理请参考书《Python数据分析与挖掘实战》Page 61页

2、处理的数据集missing\_data.xls，处理结束的数据集为missing\_data\_processed.xls

拉格朗日插值算法实现



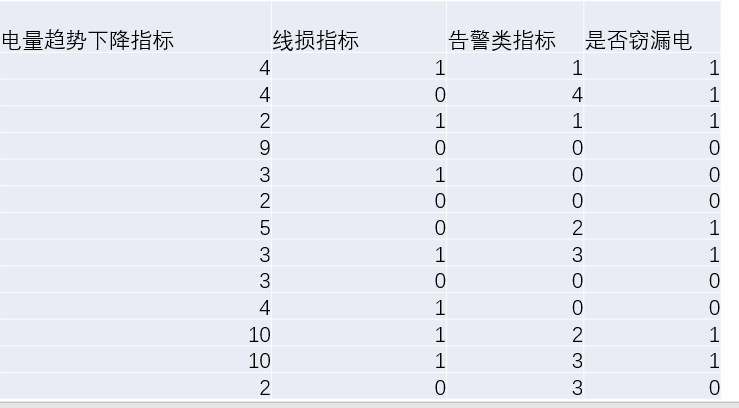
拉格朗日插补法结果对比图



根据数据变换得到的数据取专家样本数据（291个样本）

我们就是对如上的样本数据进行模型构建，第四列为目标分类

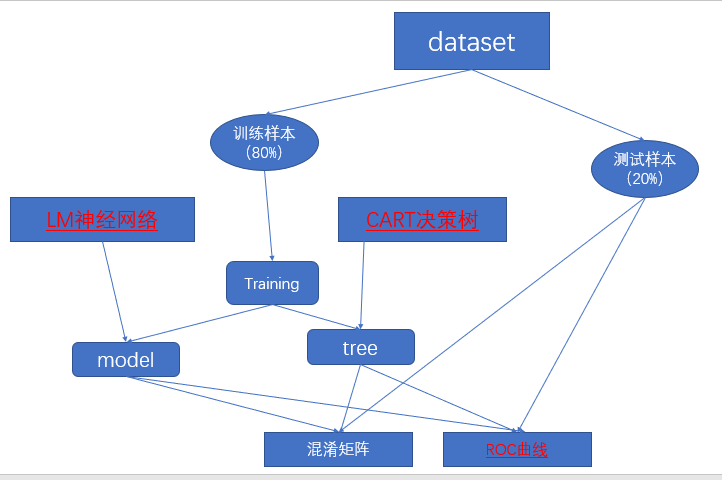
数据为model.xls



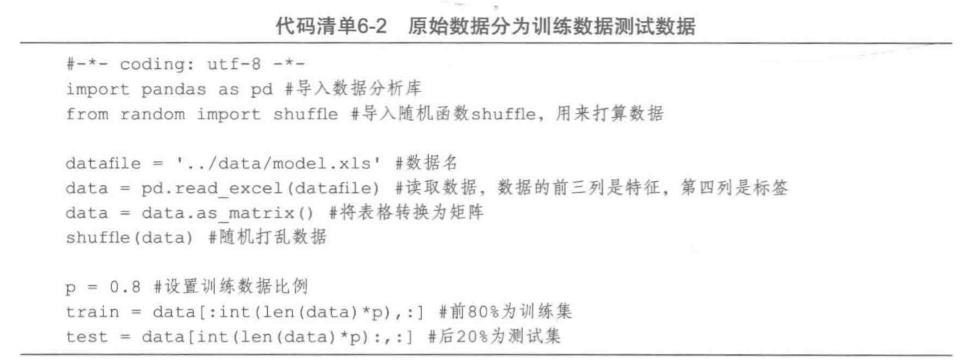
## 五、模型构建

具体的[LM神经网络算法](LM神经网络算法(BP).docx)和[CART决策树算法](CART决策树算法(ID3).docx)详解请参照连接的ppt

[ROC曲线](ROC(混淆矩阵).docx)的介绍也请参照详细的ppt



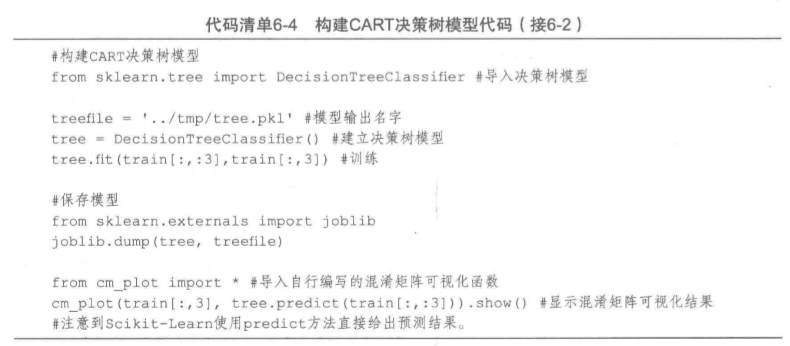
#### 1、数据划分代码实现：



#### 2、LM神经网络算法实现：



#### 3、CART决策树算法实现：



## 六、模型评价

结论：选用LM神经网络模型

