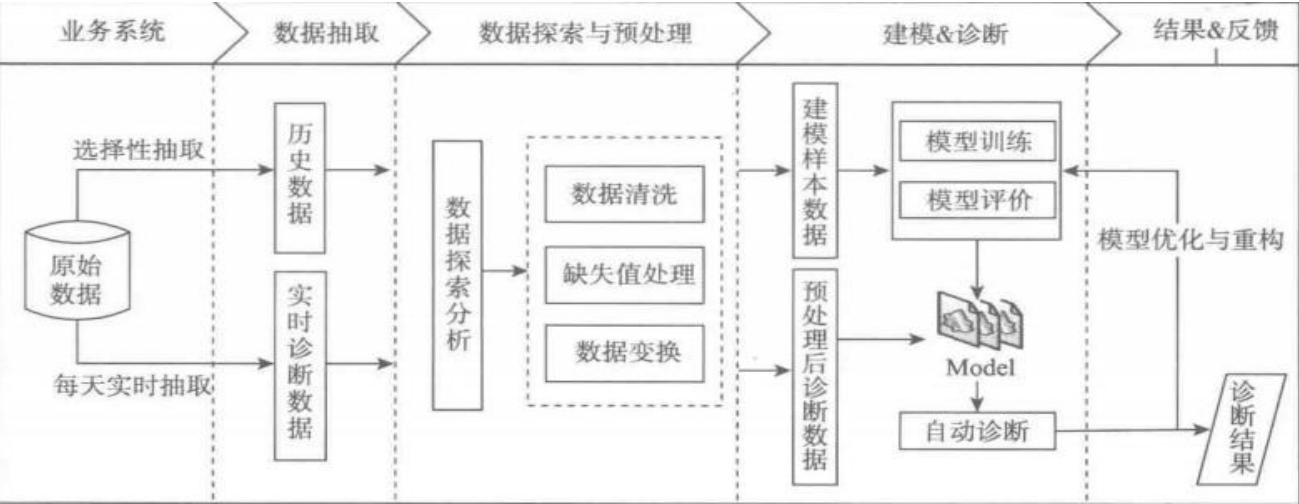


电力窃漏电用户自动识别

用电负荷不能直接体现出用户的窃电行为，故需要进行数据探索和预处理，总结窃漏电用户的行为规则，再从数据中提炼出描述的特征。最后结合历史窃漏电用户信息，整理出识别模型的样本数据集，再进一步构建分类模型，实现用户的自动识别。



数据抽取

数据探索分析

对数据进行初步研究，发现数据的内在规律特征，有助于选择合适的数据预处理和数据分析下技术。

- **分布分析：**统计获得各个用电类别的窃漏电用户分布情况。
- **周期性分析：**随机抽取一个正常用电用户和窃漏电用户，采用周期性分析对用电量进行探索。

可以发现正常用电到窃漏电过程事用电量持续下降的过程，并且持续下降，这就是用户窃漏电所表现出来的重要特征。

数据预处理

为了尽可能全面覆盖各种窃漏电方式，建模样本要包含不同用电类别的所有窃漏电用户及部分正常用户。窃漏电用户的窃漏电开始时间和结束时间是表征其窃漏电的关键时间节点，在这些时间节点上，用电负荷和终端报警等数据也会有一定的特征变化，故样本数据抽取时务必要包含关键时间节点前后一定范围的数据，并通过用户的负荷数据计算出当天的用电量，公式如下。

$$f_l = 0.25 \sum_{m_i \in l \text{天}} m_i \quad (6-1)$$

其中， f_l 为第 l 天的用电量， m_i 为第 l 天每隔 15 分钟的总有功功率，对其累加求和得到当天用电量。

数据清洗

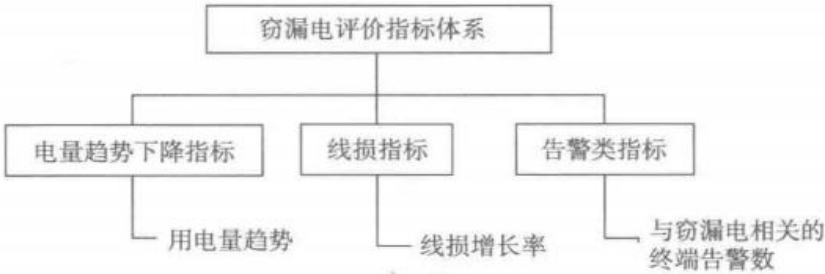
数据清洗的主要目的是从业务以及建模的相关需要方面考虑，筛选出需要的数据。由于原始数据中并不是所有的数据都需要进行分析，因此需要在数据处理时，将赘余的数据进行过滤。

- 1) 通过数据的探索分析，发现在用电类别中，非居民用电类别不可能存在漏电窃电的现象，需要将非居民用电类别的用电数据过滤掉。
- 2) 结合本案例的业务，节假日用电量与工作日相比，会明显偏低。为了尽可能达到较好数据效果，过滤节假日的用电数据。

缺失值处理

在原始计量数据，特别是用户电量抽取过程中，发现存在缺失的现象。若将这些值抛弃掉，会严重影响供出电量的计算结果，最终导致日线损率数据误差很大。为了达到较好的建模效果，需要对缺失值处理。

拉格朗日插值法、数据变换。



构建专家样本

模型构建

构建窃漏电用户识别模型

在专家样本准备完成后，需要划分测试样本和训练样本，随机选取 20% 作为测试样本，剩下的作为训练样本。窃漏电用户识别可通过构建分类预测模型来实现，比较常用的分类预测模型有 LM 神经网络和 CART 决策树，各个模型都有各自的优点，故采用这两种方法构建窃漏电用户识别，并从中选择最优的分类模型。构建 LM 神经网络和 CART 决策树模型时输入项包括电量趋势下降指标、线损类指标和告警类指标，输出项为窃漏电标识。

数据划分
LM 神经网络
CART 决策树

模型评价

采用 ROC 曲线评价方法进行模型评估，模型的 ROC 曲线越靠近左上角泛化性能越好。

进行窃漏电诊断