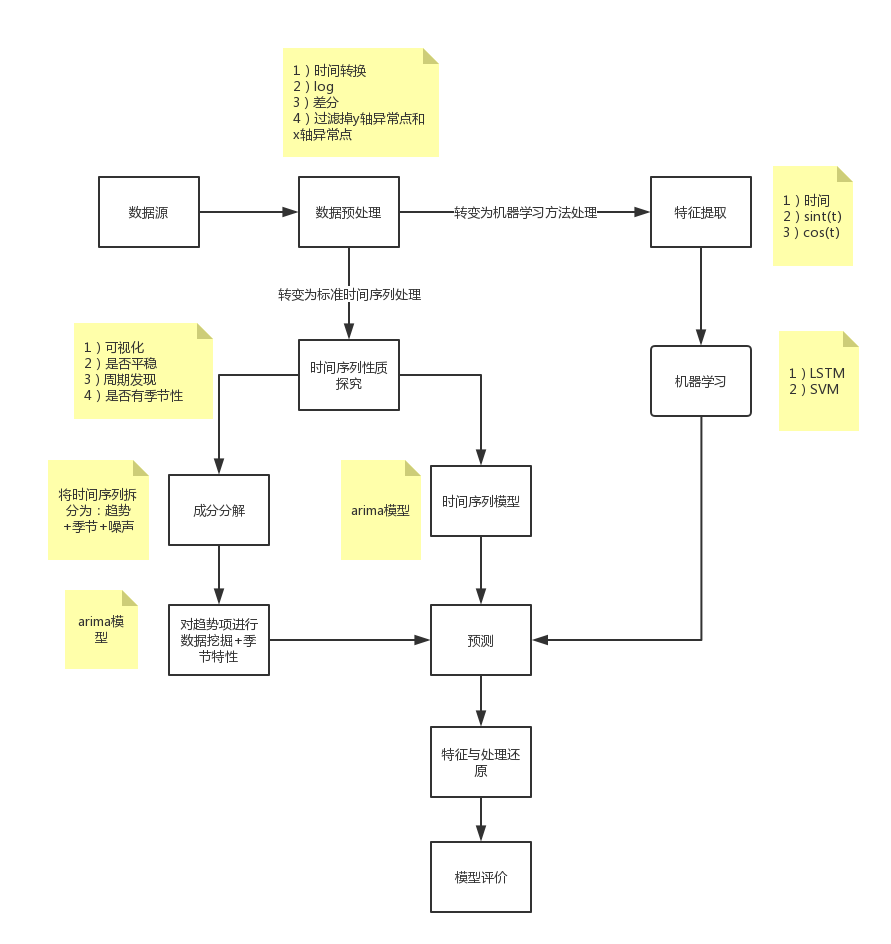
时序数据挖掘

# 时序数据挖掘示意图



# 流程注意事项

## 数据预处理

### 时间字符串转时间戳 + 排序

SQL算子就可以做到，如果是毫秒时间则需要写一个udf

### 时间分箱 + SQL过滤频数过低的箱子

### 过滤极端值

根据分位数，过滤掉前后若干百分比的数据。

### 数量数据预处理

功能简介：将输入数据进行预处理并输出

输入：DataFrame，假定时间列已经被转为了毫秒级的时间戳，且已经过滤了极端值

功能点：

1）log变换

2）diff

3）模型需要有变换和反向变换两种操作