

数据说明：

| | | | | | |
|--------------------|--------------------|---------------------|--------------------|---------------------|---------------------|
| actc2 活动码 | dbtm钻头深度（测量） | dbtv钻头深度（垂深） | dmea井眼深度（测量） | dver井眼深度（垂深） | bpos游车位置 |
| ropa（机械钻速（平均）） | hkla大钩负荷（平均） | hklx大钩负荷（最大） | woba钻压（地面，平均） | wobx钻压（地面，最大） | tqa转盘扭矩（地面，平均） |
| txq转盘扭矩（地面，最大） | rpma转盘速度（地面，平均） | sppa立管压力（平均） | chkp套管压力（平均） | spm1 1#泵冲速 | spm2 2#泵冲速 |
| spm3 3#泵冲速 | tva 泥浆池体积（活动） | tvca 泥浆池体积变化（活动） | mfop 泥浆流量% | mfoa 出口泥浆流量（平均） | mfia 入口泥浆流量（平均） |
| mdoa 出口泥浆密度（平均） | mdia 入口泥浆密度（平均） | mtoa 出口泥浆温度（平均） | mtia 入口泥浆温度（平均） | mcoa 出口泥浆电阻率（平均） | mcia 入口泥浆电阻率（平均） |
| stkc 泵冲数（总计） | lstk 滞后泵冲 | drtm 迟到井深（测量） | gasa 气（平均） | spr1 备用（1-9） | spr2 |

3倍标准差检验

判断异常值需要字段为数值型，先将所有数值型字段摘取出来，单独形成一个数据框t。数据框的iloc()方法用于对数据框进行行和列的摘取。abs(t-t.mean()) > 3*t.std()中的t指某一行中某一个值，t.mean()指某一列的均值，t.std()指某一列的均值。

注释：返回结果为NaN,即缺失值，表明数据不是异常值，那么如果删除整行全是NaN的记录,则剩下的就是包含异常值的记录（结果在3倍标准差检验）修改异常点的判断规则为其他倍标准差可调节异常点数量

对于异常值可以删除，也可以当作缺失值，采用各种方法进行填充。

机器学习检测结果，图表见报告

标签为true识别为异常数据，算法采用

pca+md及Autoencoder