**漏电电流分析与预测问题**

1. **机器学习算法所需环境：**

**1）sklearn:** pip install sklearn

**2) pandas:**pip install pandas

**3) numpy**：pip install numpy

**4) matplotlib:** pip install matplotlib

1. **漏电问题用数据建模解决的思路：**

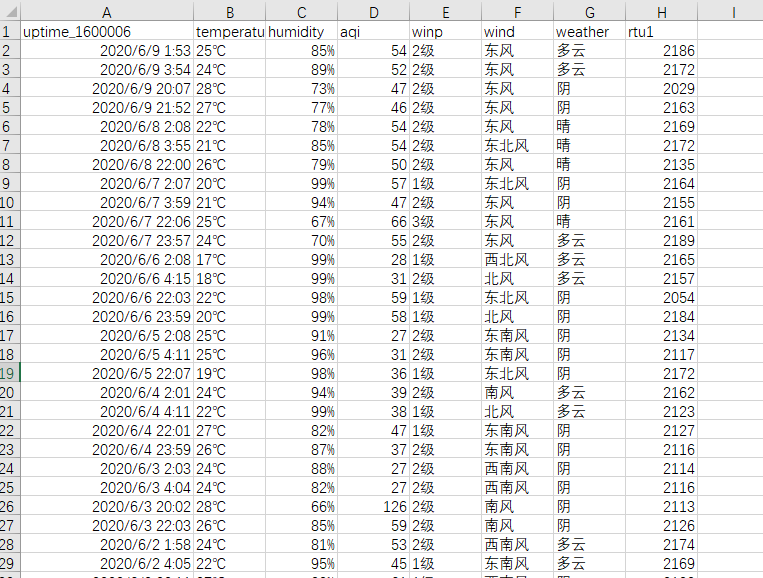
1）先使用**聚类算法**中kmeans算法进行漏电的分级。（按照电流大小分成三级：低风险、中风险、高风险），根据历史数据将该条线路的风险程度告诉用户。

2）再使用**决策树算法**将漏电电流进行和温度、湿度等因素关联。

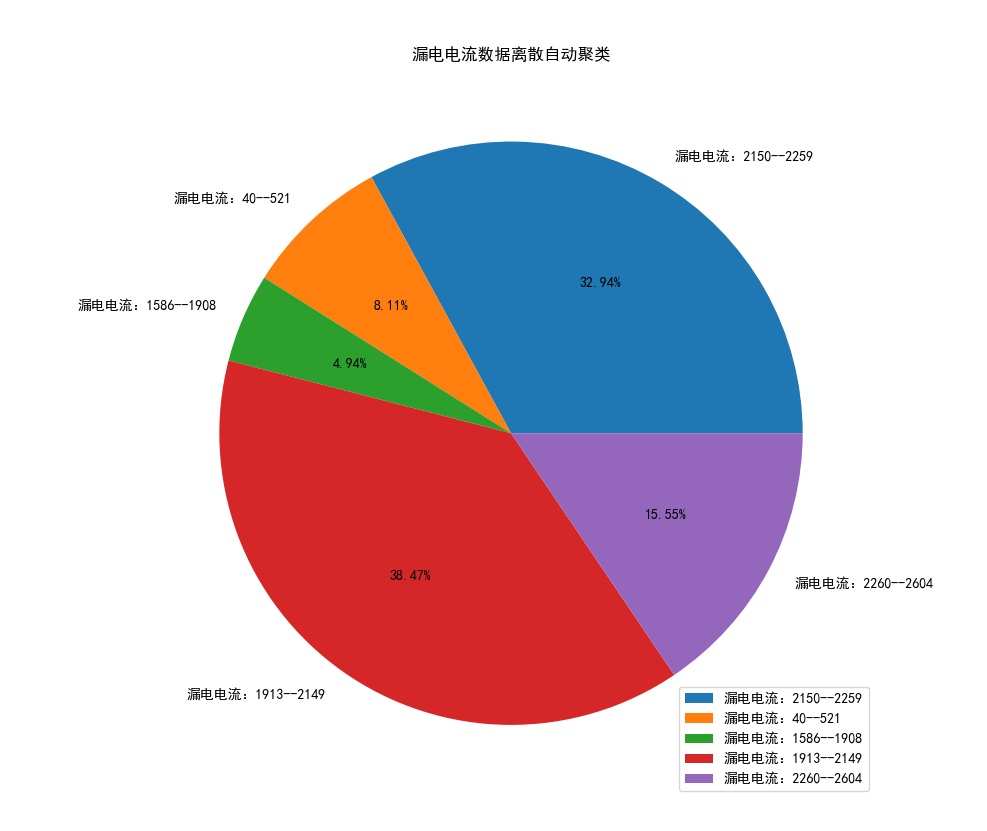
**分为三个功能函数：**

1. **数据预处理模块：**

**输入：历史漏电数据表**



**输出:漏电数据分析，扇形图展示**



漏电电流的历史最小值为： 40

漏电电流的历史最大值为： 2604

该条线路历史最大漏电电流高于2000mA，风险较高，请抓紧排查问题

1. **K-means漏电电流聚类模块**

**输入：上一模块的漏电数据数组**

loudian1=[max0\_dianliu,max1\_dianliu,max2\_dianliu,max3\_dianliu,max4\_dianliu]

loudian2=[min0\_dianliu,min1\_dianliu,min2\_dianliu,min3\_dianliu,min4\_dianliu]

**输出：漏电风险等级**

根据本条回路的历史漏电数据进行聚类建模分析，该条回路可分为五个不同等级：

等级1： 40 -- 521

等级2： 1586 -- 1908

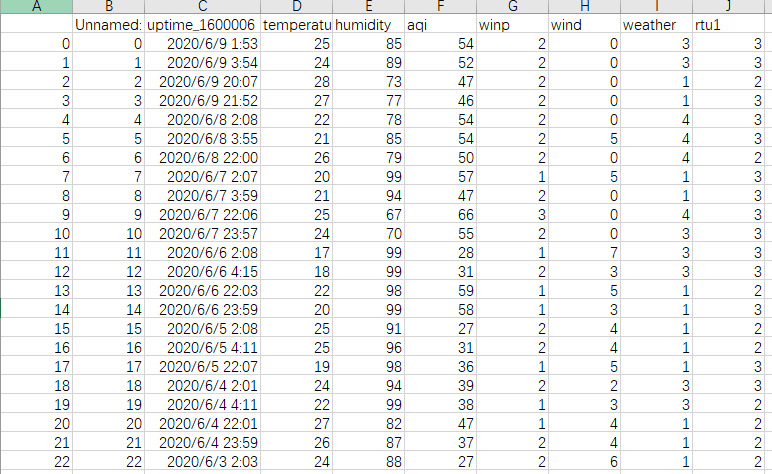
等级3： 1913 -- 2149

等级4： 2150 -- 2259

等级5： 2260 -- 2604

1. **基于湿度等和漏电电流的决策树分类模块**

**输入：预处理后的漏电和天气数据**

****

**输出: 模型和天气等状况的拟合度**

建立漏电电流和湿度、天气之间的关系时使用的数据量为： 4744

输出训练准确率： 0.9957839262187088

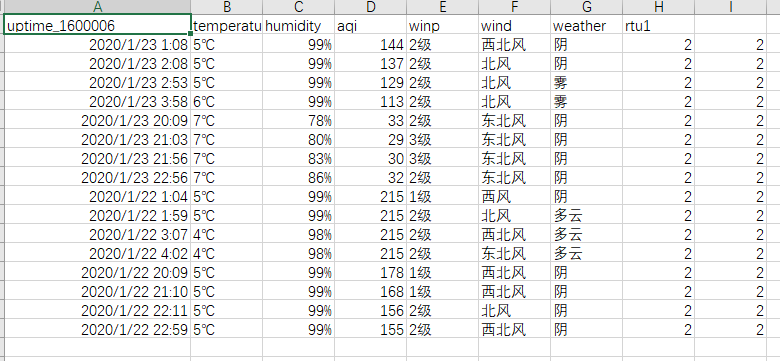
输出验证准确率： 0.541622760800843

k折交叉验证的平均准确率： 0.3294713562046976

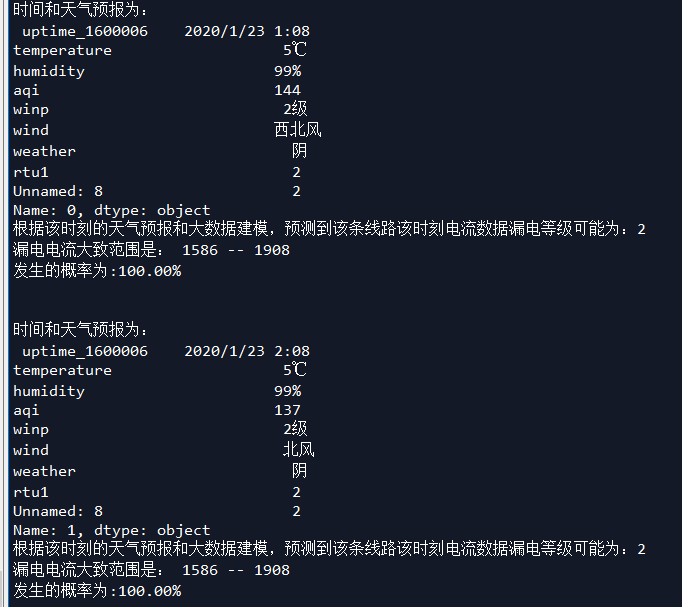
该线路漏电情况和天气等因素相关性为54.16%,还需要关注其他因素导致的电流漏电问题

1. **基于未来两天天气状况的漏电电流预测**

**输入：未来两天的天气数据，上一步的训练模型**

****

**输出：未来两天晚上每个时刻的漏电电流预测结果。**

****

**3、使用的核心算法：**

**1）数据预处理算法：**

a.去掉漏电电流为0的数据

b. 天气等因素进行编码，转换成计算机能够处理的编码

data.loc[data['weather'] == '雨', 'weather'] = 0

data.loc[data['weather'] == '阵雨', 'weather'] = 0

data.loc[data['weather'] == '阴', 'weather'] = 1

data.loc[data['weather'] == '雾', 'weather'] = 2

data.loc[data['weather'] == '多云', 'weather'] = 3

data.loc[data['weather'] == '晴', 'weather'] = 4

data.loc[data['weather'] == '阵雪', 'weather'] = 5

data.loc[data['weather'] == '雪', 'weather'] = 5

data.loc[data['weather'] == '雨夹雪', 'weather'] = 6

data.loc[data['weather'] == '扬沙', 'weather'] = 7

#风向编码，转化为算法可以处理的编码

data.loc[data['wind'] == '东风', 'wind'] = 0

data.loc[data['wind'] == '西风', 'wind'] = 1

data.loc[data['wind'] == '南风', 'wind'] = 2

data.loc[data['wind'] == '北风', 'wind'] = 3

data.loc[data['wind'] == '东南风', 'wind'] = 4

data.loc[data['wind'] == '东北风', 'wind'] = 5

data.loc[data['wind'] == '西南风', 'wind'] = 6

data.loc[data['wind'] == '西北风', 'wind'] = 7

data.loc[data['wind'] == '无持续风向', 'wind'] = 8

data.loc[data['wind'] == '风', 'wind'] = 9

df=data

df['temperature']= df['temperature'].str.replace("℃","")

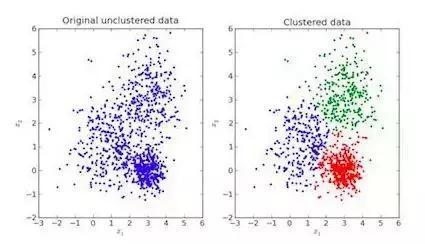
df['humidity']=df['humidity'].str.replace("%","")

df['winp']=df['winp'].str.replace("级","")

df.dropna(inplace=True) df[['temperature','humidity','winp']]=df[['temperature','humidity','winp']].astype('int')

**2）k-means算法：用于漏电数据分析**

**无监督学习算法—不需要打标签**



K-Means是一种**迭代聚类算法**，目的：每次迭代中找到局部最大值。首先，选择所需数量的聚类n。现在，随机将三个点（输入）分成n个聚类。基于每个点之间的质心距离，下一个给定的输入被分为所需的聚类。然后，重新计算所有聚类的质心。聚类的每个质心是特征值的集合，定义生成的组。检查质心特征权重可以定性地解释每个聚类代表什么类型的组。

从sklearn库导入K-Means模型，拟合特征并进行预测.

算法过程：

1. 加载数据集

2. 数据初始化

2.1 创建随机质心点

2.2 穿件保存结果的各个矩阵/数组

3. 多次迭代 （判断所有点的分类是否发生变化）

3.1 计算所有点的分类

3.2 根据3.1分类结果，重新计算质心点（用属于当前类的数据取平均作为新的质心点）

4. 返回数据

**3）决策树算法：用于漏电数据和天气的关联预测**

基于**信息熵**选择最佳测试属性，选择当前样本中具有**最大信息增益值**的属性作为测试属性，信息增益越大，不确定性越小。

**总信息熵：**

根据**属性A划分样本的信息熵值**为：

用属性A划分样本集S的**信息增益（Gain）**为：  
Gain(A)=I(S1,S2,…,Sm)-E(A)

**E(A)越小，Gain(A)越大**，属性A对分类提供的信息越大，选择A之后分类的不确定性越小。

在决策树的各级节点上都使用**信息增益**作为判断标准进行**属性选择**，使每个非叶节点进行测试时，都能获得最大的类别分类增益，**使分类后数据集的熵最小**。

**算法步骤：**

1. 对当前样本集合，计算所有属性的信息增益；
2. 选择信息增益最大的属性作为测试属性，测试属性取值相同的样本呢划分为同一个子样本集；
3. 若子样本集只有单个属性，则分支为叶子节点，判断其属性值并标上相应符号后返回调回处；否则对子样本集递归调用算法。

**注意：**

如果有属性的取值是**非数值型**的，在进行数据预处理时，应该合理应用Pandas的功能，让数据被模型接受。

**4、算法过程和产生的结果：**

以1600006\_rtu1回路为例

1）漏电分析功能

除去漏电电流为0的数据，数据集使用的漏电数据条数为： 1424

聚类后每个类别的样本数量

0 704

4 325

1 210

3 154

2 31

dtype: int64

每个聚类类别所占数据集总量的比例为：

0.4943820224719101

0.14747191011235955

0.021769662921348316

0.10814606741573034

0.22823033707865167

**漏电电流的历史最小值为： 10**

**漏电电流的历史最大值为： 6552**

**该条线路历史最大漏电电流高于2000mA，风险较高，请抓紧排查问题**

最后使用的漏电分类数为： 4

**根据本条回路的历史漏电数据进行聚类建模分析，该条回路可分为五个不同等级：**

**等级1： 10 -- 1051**

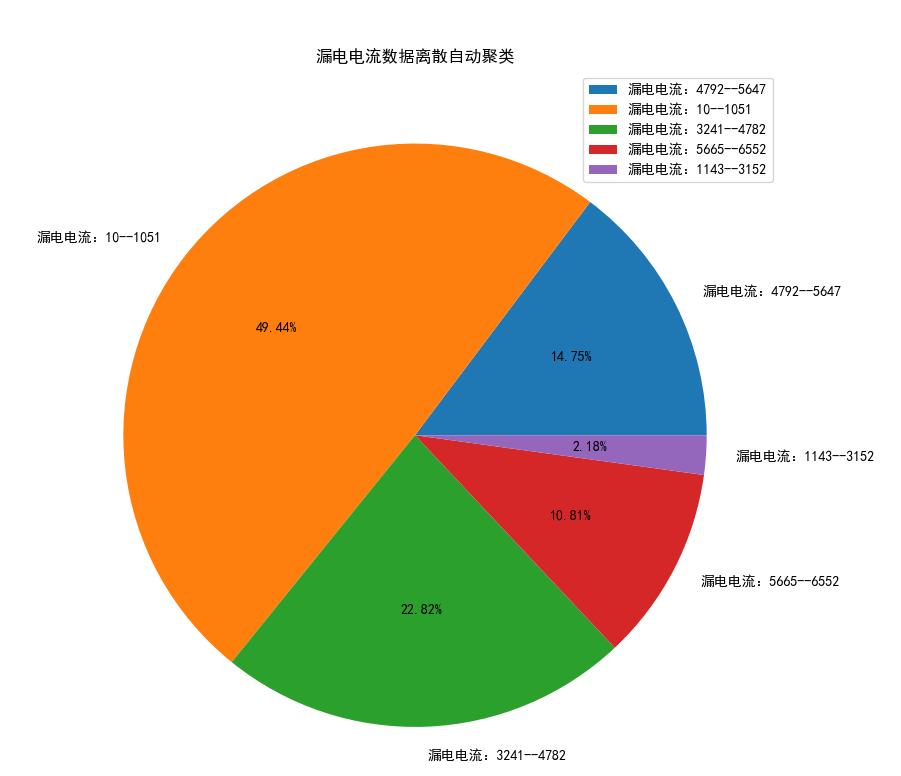
**等级2： 1143 -- 3152**

**等级3： 3241 -- 4782**

**等级4： 4792 -- 5647**

**等级5： 5665 – 6552**

**使用图例展示：**

****

2）与天气等数据的关联性分析

建立漏电电流和湿度、天气之间的关系时使用的数据量为： 4744

输出训练准确率： 0.9957839262187088

输出验证准确率： 0.5374077976817703

k折交叉验证的平均准确率： 0.3260968888864012

该线路漏电情况和天气等因素相关性为53.74%,还需要关注其他因素导致的电流漏电问题

3）根据明后两天的天气预报预测明后两天的漏电电流情况

