### 分析步骤 1. 通过训练集数据来建立随机森林分类模型。 2. 通过建立的随机森林来计算特征重要性。 3. 将建立的随机森林分类模型应用到训练、测试数据，得到模型的分类评估结果。 4. 由于随机森林中具有随机性，每次运算的结果不一样，若保存本次训练模型，后续可以直接上传数据代入到本次训练模型进行计算分类。 5. 注：随机森林无法像传统模型一样得到确定的方程，通常通过测试数据分类效果来对模型进行评价。

### 随机森林分类分析结果

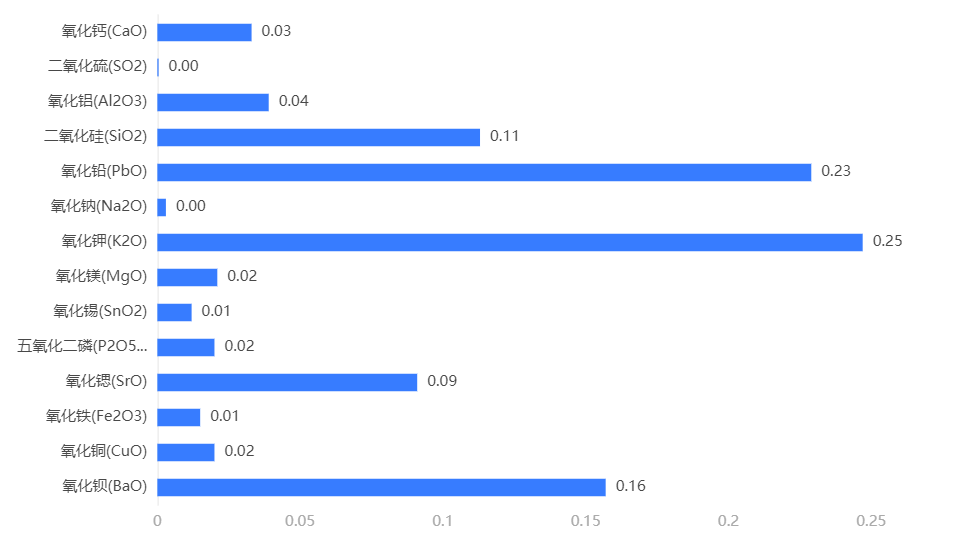
**输出结果1：模型参数**

|  |  |
| --- | --- |
| 参数名 | 参数值 |
| 训练用时 | 0.253s |
| 数据切分 | 0.7 |
| 数据洗牌 | 是 |
| 交叉验证 | 否 |
| 节点分裂评价准则 | gini |
| 决策树数量 | 100 |
| 有放回采样 | true |
| 袋外数据测试 | false |
| 划分时考虑的最大特征比例 | auto |
| 内部节点分裂的最小样本数 | 2 |
| 叶子节点的最小样本数 | 1 |
| 叶子节点中样本的最小权重 | 0 |
| 树的最大深度 | 10 |
| 叶子节点的最大数量 | 50 |
| 节点划分不纯度的阀值 | 0 |

**图表说明：**

上表展示了模型各项参数配置以及模型训练时长。

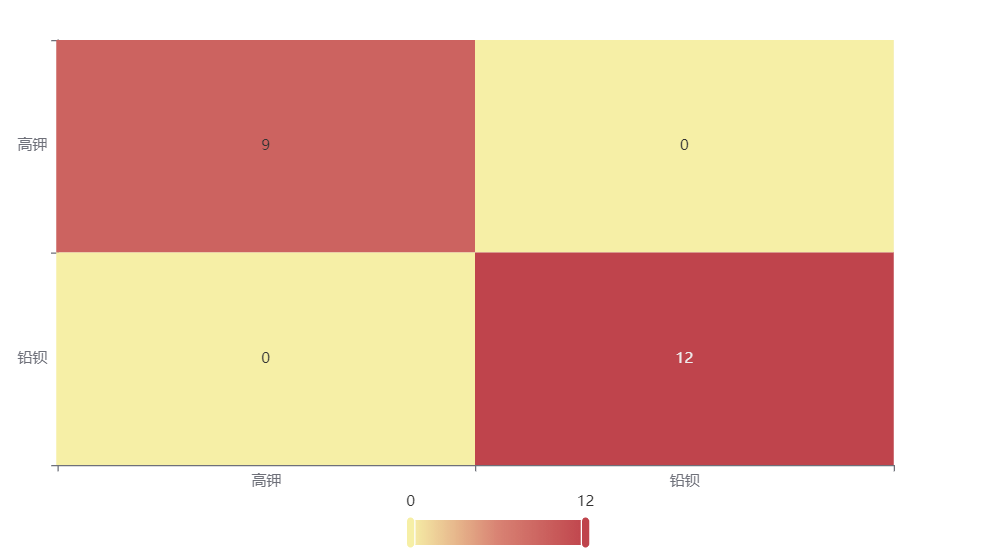
**输出结果2：特征重要性**



**图表说明：**

上柱形图或表格展示了各特征（自变量）的重要性比例。

**输出结果3：混淆矩阵热力图**



**图表说明：**

上表以热力图的形式展示了混淆矩阵。

**输出结果4：模型评估结果**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 准确率 | 召回率 | 精确率 | F1 |
| 训练集 | 1 | 1 | 1 | 1 |
| 测试集 | 1 | 1 | 1 | 1 |

**图表说明：**

上表中展示了训练集和测试集的分类评价指标，通过量化指标来衡量决策树对训练、测试数据的分类效果。  
● 准确率：预测正确样本占总样本的比例，准确率越大越好。  
● 召回率：实际为正样本的结果中，预测为正样本的比例，召回率越大越好。  
● 精确率：预测出来为正样本的结果中，实际为正样本的比例，精确率越大越好。  
● F1：精确率和召回率的调和平均，精确率和召回率是互相影响的，虽然两者都高是一种期望的理想情况，然而实际中常常是精确率高、召回率就低，或者召回率低、但精确率高。若需要兼顾两者，那么就可以用F1指标。  
● oob\_score：对于分类问题，oob\_score是袋外数据的准确率。若在建立树过程中选择有放回抽样时，大约1/3的记录没有被抽取。没有被抽取的自然形成一个对照数据集，可用于模型的验证。所以随机森林不需要另外预留部分数据做交叉验证，其本身的算法类似交叉验证，而且袋外误差是对预测误差的无偏估计。  
（当算法参数选择了“袋外测试数据”后，才会通过oob\_score来检验模型的泛化能力）

**输出结果5：测试数据预测评估结果**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 预测结果Y | 类型 | 预测结果概率\_铅钡 | 预测结果概率\_高钾 | 氧化钙(CaO) | 二氧化硫(SO2) | 氧化铝(Al2O3) | 二氧化硅(SiO2) | 氧化铅(PbO) | 氧化钠(Na2O) | 氧化钾(K2O) | 氧化镁(MgO) | 氧化锡(SnO2) | 五氧化二磷(P2O5) | 氧化锶(SrO) |
| 高钾 | 高钾 | 0.26 | 0.74 | 0.62 | 0 | 1.32 | 95.02 | 0 | 0 | 0.59 | 0 | 0 | 0.35 | 0 |
| 高钾 | 高钾 | 0.21 | 0.79 | 0.72 | 0 | 1.46 | 94.29 | 0 | 0 | 1.01 | 0 | 0 | 0.15 | 0 |
| 铅钡 | 铅钡 | 0.97 | 0.03 | 1.13 | 0 | 1.45 | 49.01 | 32.92 | 2.71 | 0 | 0 | 0 | 0.35 | 0 |
| 高钾 | 高钾 | 0.17 | 0.83 | 8.23 | 0 | 9.23 | 62.47 | 1.62 | 3.38 | 12.28 | 0.66 | 0 | 0.16 | 0 |
| 高钾 | 高钾 | 0.07 | 0.93 | 8.7 | 0 | 6.16 | 59.01 | 0 | 2.86 | 12.53 | 0 | 0 | 1.27 | 0 |
| 铅钡 | 铅钡 | 1 | 0 | 3.19 | 0 | 1.87 | 17.98 | 44 | 0 | 0 | 0.47 | 0 | 6.34 | 0.66 |
| 铅钡 | 铅钡 | 0.92 | 0.08 | 0.38 | 0 | 1.44 | 65.91 | 22.05 | 0 | 0 | 0 | 0 | 0.42 | 0 |
| 高钾 | 高钾 | 0.41 | 0.59 | 4.71 | 0 | 6.19 | 76.68 | 1 | 0 | 0 | 1.22 | 0 | 1.1 | 0 |
| 高钾 | 高钾 | 0.08 | 0.92 | 8.27 | 0 | 6.18 | 65.18 | 0.11 | 2.1 | 14.52 | 0.52 | 0 | 0 | 0.04 |
| 铅钡 | 铅钡 | 0.92 | 0.08 | 0.46 | 0 | 2.36 | 69.71 | 19.76 | 0 | 0.21 | 0 | 0 | 0.17 | 0 |
| 铅钡 | 铅钡 | 0.98 | 0.02 | 1.6 | 0 | 3.11 | 65.91 | 16.55 | 0 | 0 | 0.89 | 0 | 1.62 | 0.3 |
| 铅钡 | 铅钡 | 1 | 0 | 3.51 | 0 | 2.69 | 33.59 | 25.39 | 0 | 0.21 | 0.71 | 0 | 9.38 | 0.37 |
| 高钾 | 高钾 | 0.03 | 0.97 | 7.35 | 0.47 | 7.5 | 61.58 | 0 | 0 | 10.95 | 1.77 | 0 | 0.94 | 0.06 |
| 铅钡 | 铅钡 | 0.97 | 0.03 | 0.89 | 3.66 | 2.72 | 60.12 | 17.24 | 0 | 0.23 | 0 | 0 | 1.46 | 0.31 |
| 高钾 | 高钾 | 0.29 | 0.71 | 1.07 | 0 | 1.98 | 92.63 | 0 | 0 | 0 | 0 | 0 | 0.61 | 0 |

**图表说明：**

上表格为预览结果，只显示部分数据，全部数据请点击下载按钮导出。  
上表展示了随机森林模型对测试数据的分类结果，分类结果值是拥有最大预测概率的分类组别。

**输出结果6：模型预测与应用**

**图表说明：**

● SPSSPRO会自动保存模型，需要注意的是：在机器学习中的随机森林算法保存的模型是非常复杂的，不是类似于线性回归那样可以用一个公式保存，SPSSPRO以二进制文件方式进行序列化保存；  
● 由于随机森林具有随机性，每次训练的模型可能不一致，若保存本次训练模型，后续可以直接上传数据代入到本次训练模型进行计算预测；  
● 若删除本分析报告将会直接删除模型的缓存。