### 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

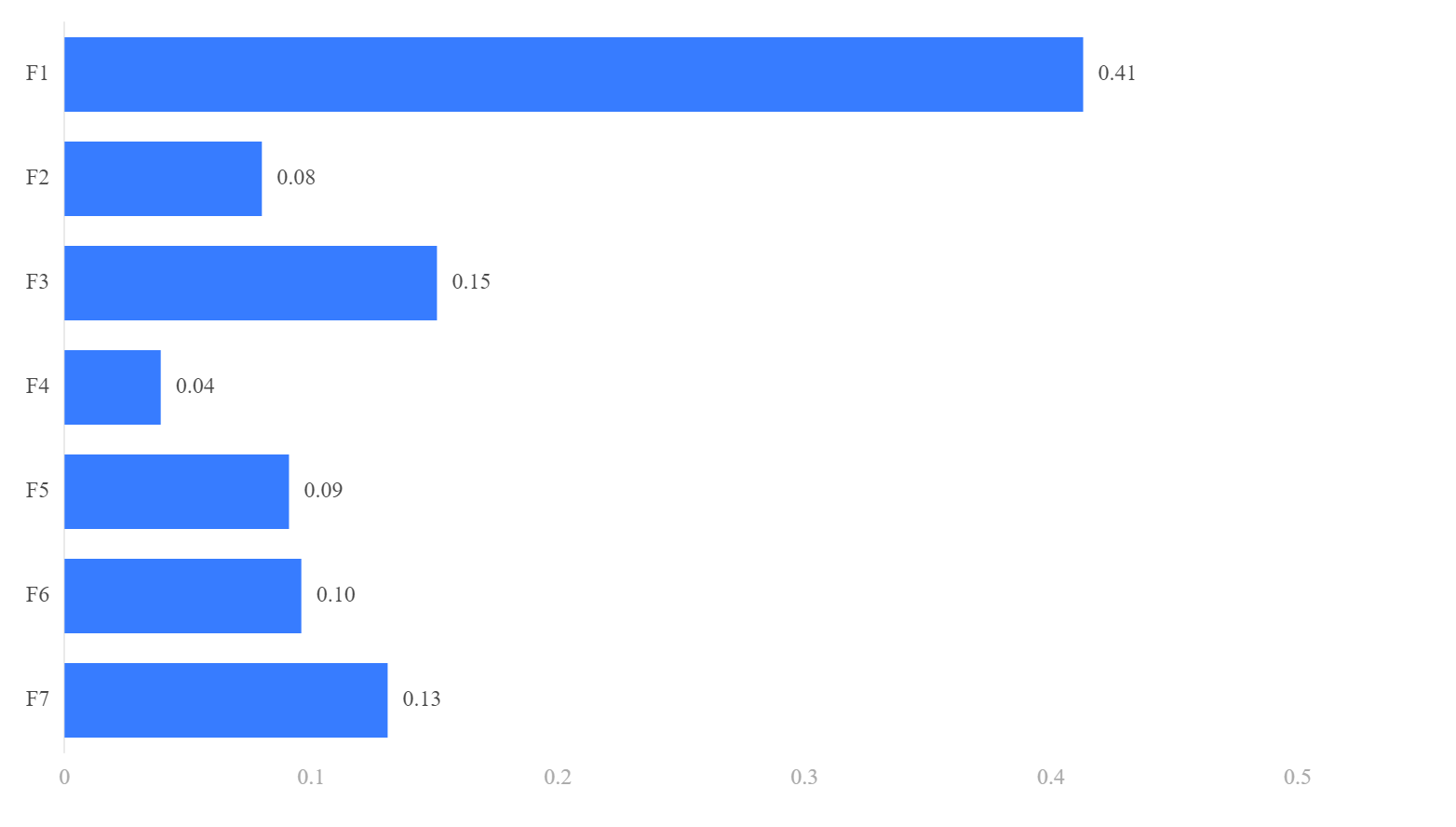
**输出结果1：模型参数**

|  |  |
| --- | --- |
| 参数名 | 参数值 |
| 训练用时 | 0.673s |
| 数据切分 | 0.7 |
| 数据洗牌 | 是 |
| 交叉验证 | 5 |
| 节点分裂评价准则 | entropy |
| 决策树数量 | 99 |
| 有放回采样 | true |
| 袋外数据测试 | true |
| 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 | 0.03690618496574074 |

**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。

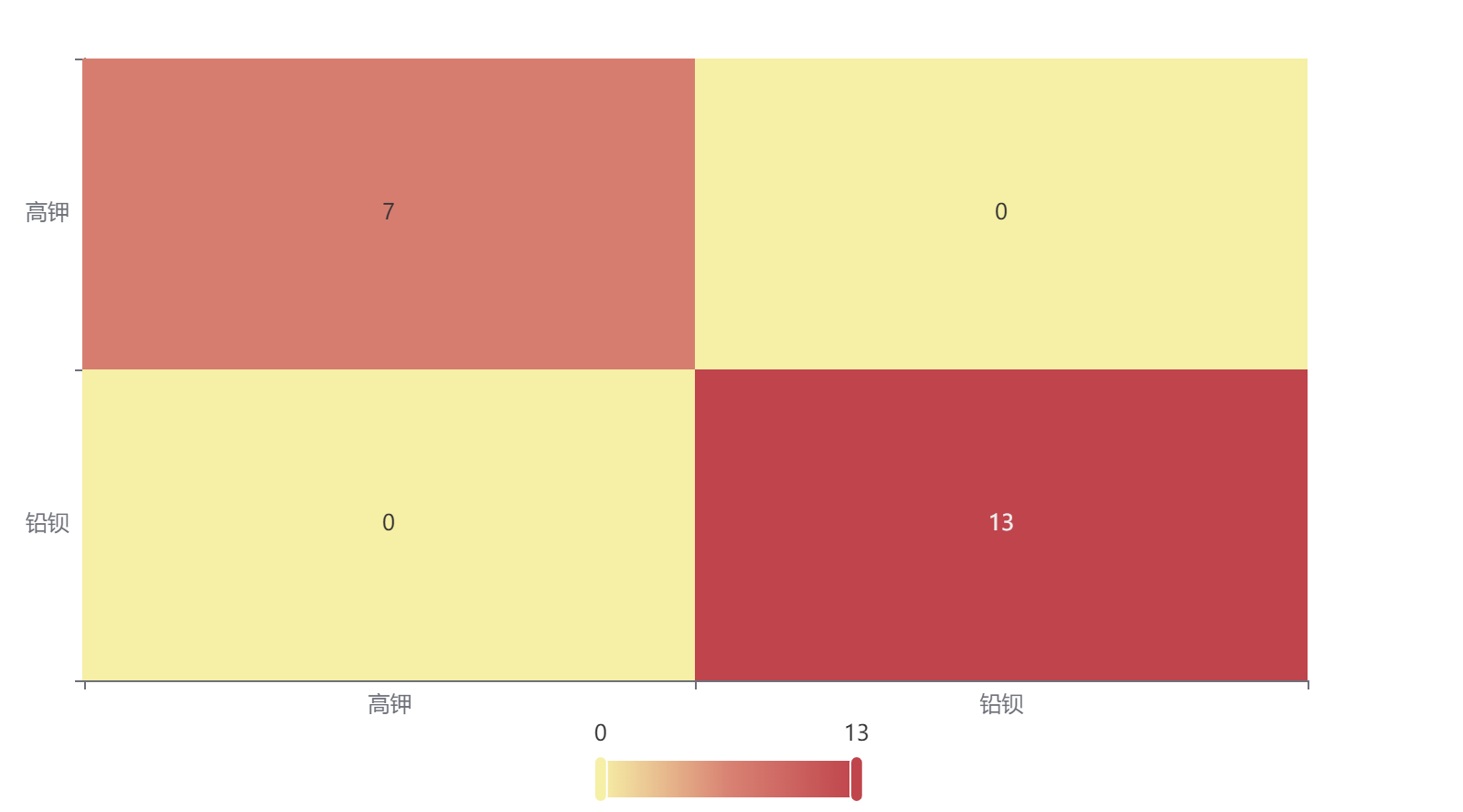
**输出结果2：特征重要性**



**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 | obb\_score |
| 训练集 | 1 | 1 | 1 | 1 | 1 |
| 交叉验证集 | 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 | F1 | F2 | F3 | F4 | F5 | F6 | F7 |
| 铅钡 | 铅钡 | 1 | 0 | -10.12908 | -10.94016 | -3.966629999999999 | 3.30342 | 3.477690000000001 | -7.4567 | -1.58023 |
| 高钾 | 高钾 | 0.13131313131313133 | 0.8686868686868687 | -14.41126 | -7.326549999999997 | 3.38134 | 1.11616 | 4.894990000000002 | -5.927210000000001 | 1.775600000000001 |
| 铅钡 | 铅钡 | 0.8686868686868687 | 0.13131313131313133 | 12.90475 | 3.85216 | 7.251209999999999 | 4.20062 | 5.532580000000001 | 8.49811 | 3.280889999999998 |
| 铅钡 | 铅钡 | 0.9595959595959596 | 0.04040404040404041 | -5.159860000000001 | -4.25407 | -5.74428 | 2.1549 | 1.448569999999999 | -4.899729999999999 | -0.8945399999999999 |
| 铅钡 | 铅钡 | 0.9797979797979798 | 0.020202020202020204 | -1.089310000000002 | -2.036590000000003 | -6.30524 | 2.33956 | -0.5759099999999993 | -5.387630000000001 | -0.26505 |
| 高钾 | 高钾 | 0.21212121212121213 | 0.7878787878787878 | -12.99167 | -5.53295 | 4.51735 | 1.10786 | 4.481070000000002 | -3.679380000000002 | 0.4091500000000004 |
| 高钾 | 高钾 | 0.09090909090909091 | 0.9090909090909091 | -17.27277 | -15.34874 | -1.31158 | 0.04533999999999996 | 8.585039999999998 | -8.038419999999999 | -0.17628 |
| 高钾 | 高钾 | 0.23232323232323232 | 0.7676767676767676 | -13.17335 | -3.455890000000001 | 1.60226 | -1.01676 | 5.55754 | -3.246140000000001 | 0.0120600000000001 |
| 铅钡 | 铅钡 | 1 | 0 | 7.855639999999998 | 3.5494 | -6.26581 | 3.2966 | -4.819189999999999 | -2.689510000000001 | -2.06316 |
| 铅钡 | 铅钡 | 0.98989898989899 | 0.010101010101010102 | 6.39878 | -0.2129000000000006 | -4.26715 | 2.31588 | -5.480469999999999 | -1.88222 | -3.75042 |
| 高钾 | 高钾 | 0.42424242424242425 | 0.5757575757575758 | -15.68593 | -11.08055 | -1.42907 | 5.67586 | 7.59451 | -5.839400000000002 | 1.34093 |
| 高钾 | 高钾 | 0.2222222222222222 | 0.7777777777777778 | -14.04604 | -7.2341 | 0.6570299999999999 | -0.8233500000000002 | 7.109510000000001 | -2.347610000000001 | -1.3709 |
| 铅钡 | 铅钡 | 1 | 0 | 7.30072 | 3.95703 | -9.170550000000002 | 0.6106899999999998 | -8.890019999999998 | -4.719980000000001 | -1.32602 |
| 铅钡 | 铅钡 | 0.98989898989899 | 0.010101010101010102 | -2.283600000000002 | -5.564290000000002 | -4.870489999999999 | 1.67217 | -2.88324 | -4.925860000000002 | -2.378480000000001 |
| 高钾 | 高钾 | 0.1414141414141414 | 0.8585858585858586 | -16.84608 | -13.4025 | 0.1411100000000001 | 1.07761 | 7.18656 | -7.53589 | 0.81853 |

**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。