### 分析流程 数据源： 所给单词的英文分词--英文标注——分数.xlsx 算法配置： 算法： 随机森林分类 分析结果： 随机森林分类基于准确率、召回率、精确率、F1指标对模型进行评价，请看详细结论。

### 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。

### 详细结论

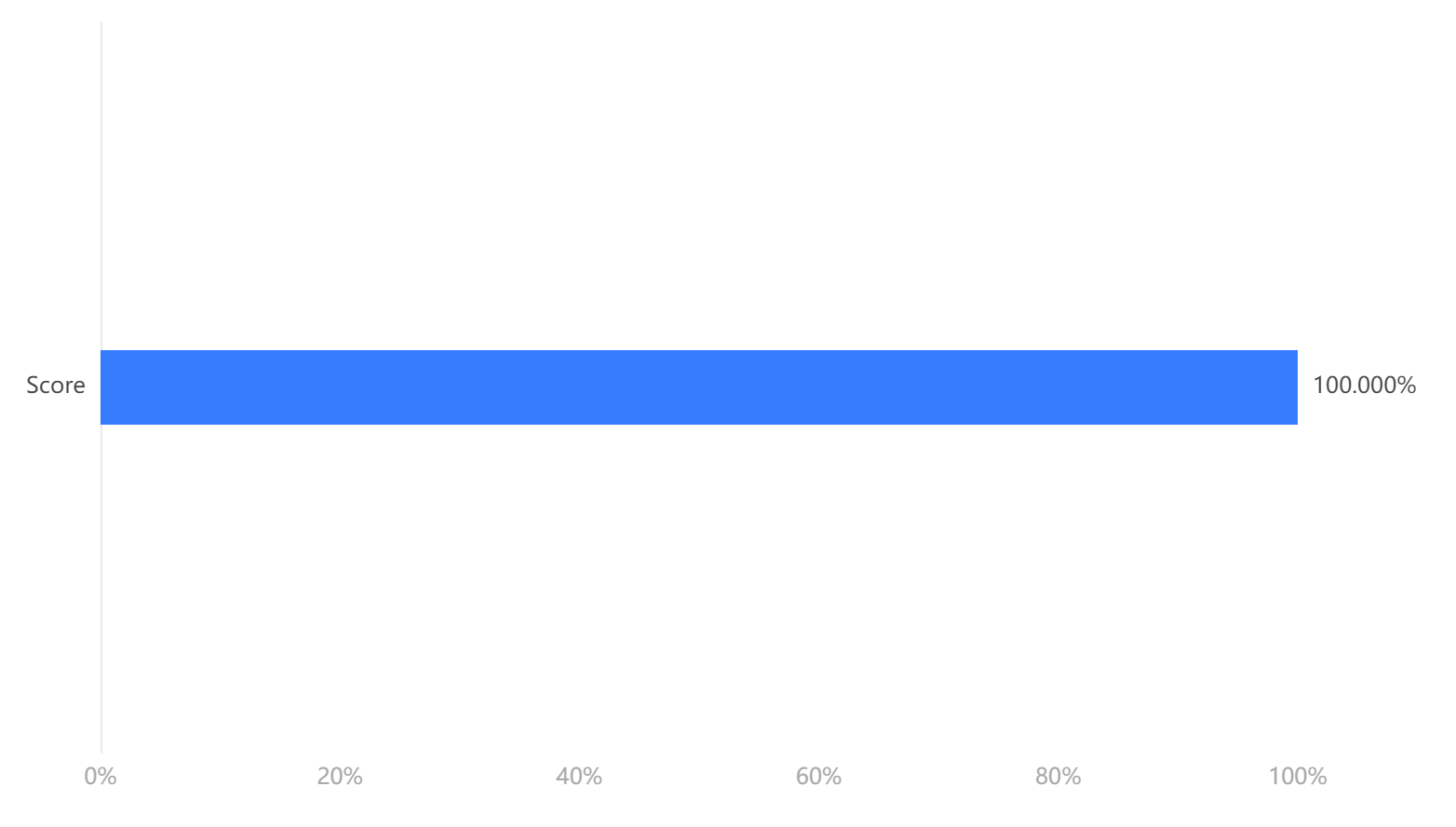
**输出结果1：模型参数**

|  |  |
| --- | --- |
| 参数名 | 参数值 |
| 训练用时 | 1.116s |
| 数据切分 | 0.9 |
| 数据洗牌 | 是 |
| 交叉验证 | 10 |
| 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。

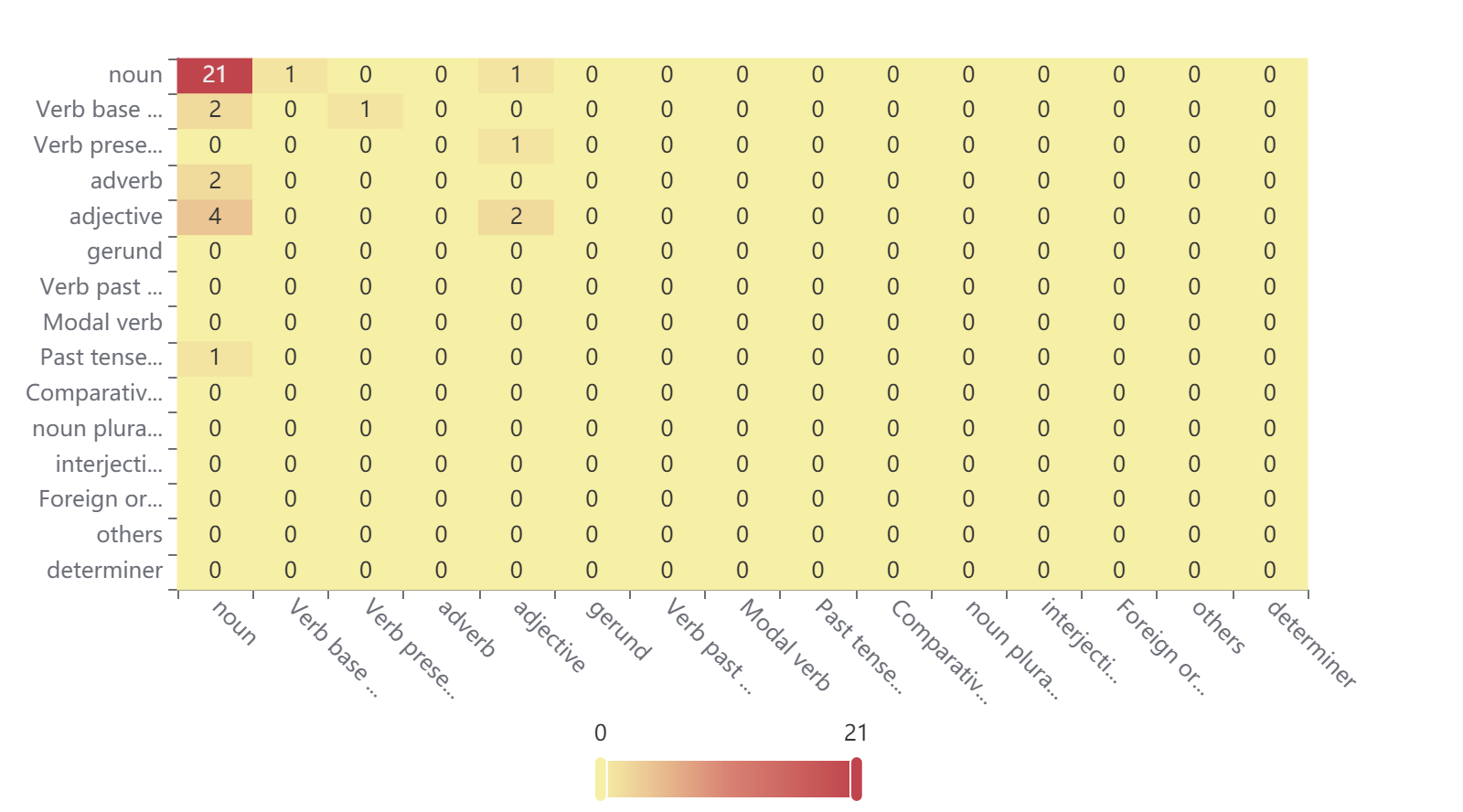
**输出结果2：特征重要性**



**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 | 0.672 | 0.672 | 0.674 | 0.612 |
| 交叉验证集 | 0.466 | 0.466 | 0.354 | 0.393 |
| 测试集 | 0.639 | 0.639 | 0.531 | 0.573 |

**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随机森林对训练、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 | Class | 预测结果概率\_Comparative adjective | 预测结果概率\_Foreign original word | 预测结果概率\_Modal verb | 预测结果概率\_Past tense of verb | 预测结果概率\_Verb base form | 预测结果概率\_Verb past participle | 预测结果概率\_Verb present singular | 预测结果概率\_adjective | 预测结果概率\_adverb | 预测结果概率\_determiner | 预测结果概率\_gerund | 预测结果概率\_interjection | 预测结果概率\_noun |
| noun | noun | 0.0024818453532713686 | 0.00010973084886128364 | 0.010313504204741792 | 0.016031651504813706 | 0.1648987442636664 | 0.0012475586632258295 | 0.0004404102893608141 | 0.07142476413162929 | 0.009065055420182652 | 0 | 0.00009523809523809524 | 0.0002127659574468085 | 0.7232195322956191 |
| adjective | Verb present singular | 0.0005800865800865801 | 0.00010973084886128364 | 0.0006290692066554135 | 0.0013849039516205936 | 0.05519097283128398 | 0.052683386105178226 | 0.0007073821564755967 | 0.5596383720612907 | 0.003163718134549021 | 0 | 0.0001913919413919414 | 0 | 0.32500618837911577 |
| noun | noun | 0.000170995670995671 | 0.013968650251371388 | 0 | 0.0012064839712013263 | 0.17919937932636681 | 0.0028688909297786315 | 0.01640910867515153 | 0.316289630433779 | 0.0156151580139561 | 0.00012195121951219512 | 0.0037987301821388085 | 0 | 0.44994782030951375 |
| noun | noun | 0.010131518784336222 | 0.00004761904761904762 | 0.01120740298758581 | 0.0012712889797629342 | 0.2411331847312899 | 0.0009462807307634893 | 0.00031618668687634203 | 0.0378347907961672 | 0.014071558991842986 | 0 | 0.00009523809523809524 | 0.0002127659574468085 | 0.6826270738990843 |
| noun | noun | 0.00009523809523809524 | 0.00022493517975188907 | 0.0003260389036251105 | 0.0005368647359343188 | 0.050434174568191804 | 0.1358162964832867 | 0.0035504588784144777 | 0.25473537793645235 | 0.020370594126303376 | 0 | 0.0009427828888613203 | 0 | 0.5110469122781575 |
| noun | noun | 0.000170995670995671 | 0.001795022061082136 | 0.00011494252873563218 | 0.0006693201580850095 | 0.30394114065766875 | 0.0048519590562372365 | 0.003247184150947013 | 0.22057661124757977 | 0.007610493398483553 | 0 | 0.00661804101871605 | 0 | 0.4494103119076678 |
| noun | Past tense of verb | 0.0003982683982683983 | 0.00022493517975188907 | 0.0006290692066554135 | 0.0008398950389646218 | 0.25116853475141354 | 0.011443697105576656 | 0.0016510478543804617 | 0.13902268375456373 | 0.008210372089261786 | 0 | 0.0006567290733957401 | 0 | 0.5846989715733857 |
| noun | adverb | 0.0003225806451612903 | 0 | 0 | 0.014294328767913717 | 0.0182648071063168 | 0.00012195121951219512 | 0.012447938598942903 | 0.13203189293081455 | 0.0004878048780487805 | 0.028739963188567804 | 0.0027128138673008393 | 0 | 0.6680715897930921 |
| noun | noun | 0.000170995670995671 | 0.0061356096056051305 | 0 | 0.0009625815321769361 | 0.15293596998661552 | 0.004315176967750386 | 0.006019387738282822 | 0.2405512528920727 | 0.010412464289655978 | 0 | 0.006763678668942674 | 0 | 0.5711373739393598 |
| noun | adjective | 0 | 0 | 0.00032327586206896546 | 0.003826665016773909 | 0.025767800565081184 | 0.00028735632183908046 | 0.04720031852224657 | 0.03717182412942313 | 0.0008938615798483738 | 0 | 0 | 0.0013238770685579195 | 0.8753732180073034 |
| noun | noun | 0.0003225806451612903 | 0 | 0 | 0.009794328767913718 | 0.014341730183239875 | 0.00012195121951219512 | 0.0544479385989429 | 0.0325447134436351 | 0.0004878048780487805 | 0.007522014470619089 | 0.0019435830980700699 | 0 | 0.8663023590238617 |
| noun | noun | 0.0003225806451612903 | 0 | 0 | 0.009794328767913718 | 0.015591730183239876 | 0.00012195121951219512 | 0.06944793859894291 | 0.03379471344363509 | 0.0004878048780487805 | 0.007522014470619089 | 0.0019435830980700699 | 0 | 0.8488023590238616 |
| noun | noun | 0.000170995670995671 | 0.006988911768083215 | 0 | 0.0012064839712013263 | 0.3382542331195969 | 0.00338228276669988 | 0.007444841676655323 | 0.1445165400433077 | 0.010848545428885092 | 0.00012195121951219512 | 0.0055464817870116865 | 0 | 0.4811155315318162 |
| noun | noun | 0.29098649192575865 | 0.0007926916221033868 | 0 | 0.0009530345767279565 | 0.057420745759117715 | 0.0006501563477173233 | 0.0018809034745988086 | 0.32020763471110486 | 0.001846751813466304 | 0.0008291472492888701 | 0.0031722987459107998 | 0 | 0.32126014377420536 |
| Verb present singular | Verb base form | 0 | 0 | 0 | 0.006121854900317898 | 0.022471249496866196 | 0 | 0.45978736949325183 | 0.059340099969606604 | 0 | 0.00282258064516129 | 0.0009108159392789373 | 0 | 0.4398036053130929 |

**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：**

● 系统会自动保存模型，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，系统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。

### 参考文献 [1] Scientific Platform Serving for Statistics Professional 2021. SPSSPRO. (Version 1.0.11)[Online Application Software]. Retrieved from https://www.spsspro.com. [2] 周志华. 机器学习[M]. 清华大学出版社, 2016.