实验三：基于随机森林的气温预测

此实验需要完成三项任务：

任务一：使用随机森林算法完成基本建模任务，基本任务需要处理数据，观察特征，完成建模并进行可视化展示分析

任务二：观察数据量与特征个数对结果影响，在保证算法一致的前提下，加大数据个数，观察结果变换。重新考虑特征工程，引入新特征后观察结果走势。

任务三：对随机森林算法进行调参，找到最合适的参数，掌握机器学习中两种经典调参方法，对当前模型进行调节

任务流程：

任务一：

1. 读取数据，统计指标
2. 处理数据、数据展示、独热编码
3. 划分测试集与训练集
4. 建立一个基础的随机森林模型
5. 测试与可视化展示树
6. 剪枝处理

Q1：画出特征重要性的示意图，讨论各个特征的重要性。

Q2：画出测试值与真实值的差异对比图，讨论为什么存在差异？

任务二：

1. 导入更多的数据
2. 数据处理
3. 分别查看新数据和老数据的训练结果
4. 加入新特征再训练一次
5. 对比精度和训练时间
6. 绘制误差、精度和运行时间对比图

Q3：写出新老数据的准确率，讨论：只增大数据量的话，结果会提升吗？

Q4：画出特征重要性累加图，讨论用几个特征比较好？

Q5：附上采用全部特征和只采用高价值特征的精度-时间表，两者之间的精度差了多少？所相对时间效率差了多少？这说明了什么？

任务三：

1. 导入数据、处理数据
2. 建立随机森林模型
3. 使用RandomizedSearchCV()函数调节参数
4. 评估模型
5. 使用GridSearchCV()再进行微调
6. 对最终模型进行性能评估

Q6：结合整个调参过程中模型精度的变化，讨论调参的意义。