**人工智能训练师（三级）操作技能考核**

**试题单**

准考证号：

试题代码：

试题名称：智能农业系统中的业务数据采集和处理流程设计

考核时间：30min

**1.**场地设备要求

人工智能训练师主机：CPU（intel i5 及以上）、内存（不少于 16GB）、操作系统（windows10）、支持深度学习训练；

**2.**工作任务

某农业公司计划引入智能农业系统，通过安装在农田中的各种传感器（如温度传感器、湿度传感器、土壤传感器等）实时监控农田环境，收集数据并进行分析，以优化作物管理和提高产量。为此，公司需要设计并实现一套数据采集和处理流程，确保数据的高效采集、传输和处理，为智能分析提供可靠的数据支持。

我们提供一个传感器数据集（sensor\_data.csv），包含以下字段：

* SensorID: 传感器ID
* Timestamp: 时间戳
* SensorType: 传感器类型（Temperature温度, Humidity湿度, SoilMoisture土壤水分, SoilPH土壤酸碱度, Light光传感器）
* Value: 传感器读数
* Location: 传感器安装位置：

你作为智能农业系统的人工智能训练师，根据提供的sensor\_data.csv数据集和Python代码框架（1.1.2.ipynb），完成以下数据的采集和处理任务，为智能农业系统提供可靠的数据支持。请按照以下要求完成任务，确保结果准确并保存相应的截图。

（1）传感器数据统计：

通过补全并运行Python代码（1.1.2.ipynb）分别统计每种传感器的数据数量和平均值。将上述统计结果截图以JPG的格式保存，命名为“1.1.2-1”。

（2）按位置统计温度和湿度数据：

通过补全并运行Python代码（1.1.2.ipynb）统计每个位置的温度和湿度传感器数据的平均值。将上述统计结果截图以JPG的格式保存，命名为“1.1.2-2”。

（3）数据清洗和异常值处理：

通过补全并运行Python代码（1.1.2.ipynb）对数据进行清洗，处理异常值。具体要求如下：

* 将明显异常的温度（< -10 或 > 50）和湿度（< 0 或 > 100）数据进行标记并统计。
* 对缺失值使用前面数据的值（如果前面值没有采用后面数据的值）进行填补。
* 将清洗后的数据保存为新文件cleaned\_sensor\_data.csv。

所有结果文件储存在桌面新建的考生文件夹中，文件夹命名为“准考证号+身份证号后六位”。

**3.**技能要求

（1）能结合人工智能技术要求和业务特征，设计整套业务数据采集流程

（2）能结合人工智能技术要求和业务特征，设计整套业务数据处理流程

**4.**质量指标

（1）设计出的业务数据底层逻辑清晰，有效合理。

（2）数据完整性：每个传感器数据记录数应完整，缺失值尽量少。

（3）数据准确性：传感器数据应合理，与参考数据偏差小。

**人工智能训练师（三级）操作技能考核**

**试题评分表**

准考证号：

试题代码：

试题名称：智能农业系统中的业务数据采集和处理流程设计

考核时间：30min

测量分评分表

| 细则编号 | 配分 | 评分细则描述 | 规定或  标称值 | 结果或  实际值 | 得分 |
| --- | --- | --- | --- | --- | --- |
| M1 | 2 | 1.1.2-1.JPG中数据正确得2分； | 根据数据 |  |  |
| M2 | 2 | 1.1.2-2.JPG中数据正确得2分； | 根据数据 |  |  |
| M3 | 1 | cleaned\_sensor\_data.csv文件中数据正确得1分； | 根据数据 |  |  |
| M4 | 2 | 读取数据集代码正确得2分； | 根据数据 |  |  |
| M5 | 3 | 对传感器类型进行分组，并计算每个组的数据数量和平均值代码正确，每空1分，总计得3分； | 根据数据 |  |  |
| M6 | 2 | 筛选出温度和湿度数据，然后按位置和传感器类型分组，计算每个组的平均值代码正确得2分； | 根据数据 |  |  |
| M7 | 3 | 标记异常值代码正确，每空1分，总计得3分； | 根据数据 |  |  |
| M8 | 2 | 输出异常值数量代码正确得2分； | 根据数据 |  |  |
| M9 | 4 | 使用前向填充和后向填充的方法填补缺失值代码正确，每空1分，总计得4分； | 根据数据 |  |  |
| M10 | 4 | 删除用于标记异常值的列，并将清洗后的数据保存到新的CSV文件中代码正确，每空1分，总计得4分； | 根据数据 |  |  |
| 合计配分 | 25 | 合计得分 | |  |  |

**参考答案**：

Python参考代码（1.1.2.ipynb）：

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

# 读取数据集

data = pd.read\_csv('sensor\_data.csv')

# 1. 传感器数据统计

# 对传感器类型进行分组，并计算每个组的数据数量和平均值

sensor\_stats = data.groupby('SensorType')['Value'].agg(['count', 'mean'])

# 输出结果

print("传感器数据数量和平均值:")

print(sensor\_stats)

# 2. 按位置统计温度和湿度数据

# 筛选出温度和湿度数据，然后按位置和传感器类型分组，计算每个组的平均值

location\_stats = data[data['SensorType'].isin(['Temperature', 'Humidity'])].groupby(['Location', 'SensorType'])['Value'].mean().unstack()

# 输出结果

print("每个位置的温度和湿度数据平均值:")

print(location\_stats)

# 3. 数据清洗和异常值处理

# 标记异常值

data['is\_abnormal'] = np.where(

((data['SensorType'] == 'Temperature') & ((data['Value'] < -10) | (data['Value'] > 50))) |

((data['SensorType'] == 'Humidity') & ((data['Value'] < 0) | (data['Value'] > 100))),

True, False

)

# 输出异常值数量

print("异常值数量:", data['is\_abnormal'].sum())

# 填补缺失值

# 使用前向填充和后向填充的方法填补缺失值

data['Value'].fillna(method='ffill', inplace=True)

data['Value'].fillna(method='bfill', inplace=True)

# 保存清洗后的数据

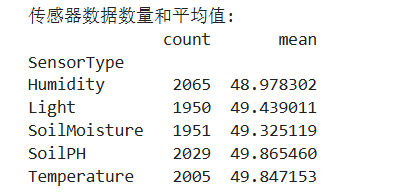
# 删除用于标记异常值的列，并将清洗后的数据保存到新的CSV文件中

cleaned\_data = data.drop(columns=['is\_abnormal'])

cleaned\_data.to\_csv('cleaned\_sensor\_data.csv', index=False)

print("数据清洗完成，已保存为 'cleaned\_sensor\_data.csv'")

1.1.2-1.jpg



1.1.2-2.jpg

