**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：智能医疗系统中的业务数据处理流程设计

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

某医疗机构计划引入智能医疗系统，以提升诊断效率和准确性。通过分析患者的历史数据，使用机器学习算法预测患者的健康风险，从而辅助医生进行诊断和治疗。为此，该机构需要设计一套全面的业务数据处理流程，确保数据处理的高效性和准确性，为人工智能模型提供可靠的输入数据。

我们提供一个患者数据集（patient\_data.csv），包含以下字段：

* PatientID: 患者ID
* Age: 年龄
* BMI: 体重指数
* BloodPressure: 血压
* Cholesterol: 胆固醇水平
* DaysInHospital: 住院天数

你作为人工智能训练师，根据提供的数据集和Python代码框架（1.1.1.ipynb），完成以下数据的统计和分析，为智能医疗系统提供可靠的数据支持。

（1）通过补全并运行Python代码（1.1.1.ipynb）分别统计住院天数超过7天的患者数量以及其占比。这类患者被定义为高风险患者，反之为低风险患者。将上述统计结果截图以JPG的格式保存，命名为“1.1.1-1”。

（2）通过补全并运行Python代码（1.1.1. ipynb）统计不同BMI区间中高风险患者的比例和统计不同BMI区间中的患者数。BMI区间分类设置为：偏瘦（低于18.5），正常（18.5～23.9），超重（24.0～27.9），肥胖（28.0及以上），将上述统计结果截图以JPG的格式保存，命名为“1.1.1-2”。

（3）通过补全并运行Python代码（1.1.1. ipynb）统计不同年龄区间中高风险患者的比例和统计不同年龄区间中的患者数。年龄区间分类设置为：年龄区间分类设置为：≤25岁，26-35岁，36-45岁，46-55岁，56-65岁，＞65岁，将上述统计结果截图以JPG的格式保存，命名为“1.1.1-3”。

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。

**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：智能医疗系统中的业务数据处理流程设计

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 | 1 | 1.1.1-1.JPG中数据正确得1分； | 根据数据 |  |  |
| M2 | 1 | 1.1.1-2.JPG中数据正确得1分； | 根据数据 |  |  |
| M3 | 1 | 1.1.1-3.JPG中数据正确得1分； | 根据数据 |  |  |
| M4 | 1 | 读取数据集代码正确得1分； | 根据数据 |  |  |
| M5 | 3 | 创建新列'RiskLevel'，根据住院天数判断风险等级代码正确得3分； | 根据数据 |  |  |
| M6 | 2 | 统计不同风险等级的患者数量代码正确得2分； | 根据数据 |  |  |
| M7 | 1 | 计算高风险患者占比代码正确得1分； | 根据数据 |  |  |
| M8 | 1 | 计算低风险患者占比代码正确得1分； | 根据数据 |  |  |
| M9 | 4 | 根据年龄值划分指定区间代码正确得4分； | 根据数据 |  |  |
| M10 | 2 | 计算每个BMI区间中高风险患者的比例代码正确得2分； | 根据数据 |  |  |
| M11 | 1 | 统计每个BMI区间的患者数量代码正确得1分； | 根据数据 |  |  |
| M12 | 4 | 根据年龄值分配年龄区间标签代码正确得4分； | 根据数据 |  |  |
| M13 | 2 | 算每个年龄区间中高风险患者的比例代码正确得2分； | 根据数据 |  |  |
| M14 | 1 | 统计每个年龄区间的患者数量代码正确得1分； | 根据数据 |  |  |
| 合计配分 | 25 | 合计得分 | |  |  |

**参考答案**：

Python参考代码（1.1.1.ipynb）：

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

# 读取数据集

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

# 1. 统计住院天数超过7天的患者数量及其占比

# 创建新列'RiskLevel'，根据住院天数判断风险等级 3分

data['RiskLevel'] = np.where(data['DaysInHospital'] > 7, '高风险患者', '低风险患者')

# 统计不同风险等级的患者数量 2分

risk\_counts = data['RiskLevel'].value\_counts()

# 计算高风险患者占比 1分

high\_risk\_ratio = risk\_counts['高风险患者'] / len(data)

# 计算低风险患者占比 1分

low\_risk\_ratio = risk\_counts['低风险患者'] / len(data)

# 输出结果

print("高风险患者数量:", risk\_counts['高风险患者'])

print("低风险患者数量:", risk\_counts['低风险患者'])

print("高风险患者占比:", high\_risk\_ratio)

print("低风险患者占比:", low\_risk\_ratio)

# 2. 统计不同BMI区间中高风险患者的比例和统计不同BMI区间中的患者数

# 定义BMI区间和标签bmi\_bins = [0, 18.5, 24, 28, np.inf]

bmi\_labels = ['偏瘦', '正常', '超重', '肥胖']

# 根据BMI值划分指定区间 4分

data['BMIRange'] = pd.cut(data['BMI'], bins=bmi\_bins, labels=bmi\_labels, right=False) # 使用左闭右开区间

# 计算每个BMI区间中高风险患者的比例 2分

bmi\_risk\_rate = data.groupby('BMIRange')['RiskLevel'].apply(lambda x: (x == '高风险患者').mean())

# 统计每个BMI区间的患者数量 1分

bmi\_patient\_count = data['BMIRange'].value\_counts()

# 输出结果

print("BMI区间中高风险患者的比例和患者数:")

print(bmi\_risk\_rate)

print(bmi\_patient\_count)

# 3. 统计不同年龄区间中高风险患者的比例和统计不同年龄区间中的患者数

# 定义年龄区间和标签

age\_bins = [0, 26, 36, 46, 56, 66, np.inf]

age\_labels = ['≤25岁', '26-35岁', '36-45岁', '46-55岁', '56-65岁', '＞65岁']

# 根据年龄值划分指定区间 4分

data['AgeRange'] = pd.cut(data['Age'], bins=age\_bins, labels=age\_labels,right=False) # 使用左闭右开区间)

# 计算每个年龄区间中高风险患者的比例 2分

age\_risk\_rate = data.groupby('AgeRange')['RiskLevel'].apply(lambda x: (x == '高风险患者').mean())

# 统计每个年龄区间的患者数量 1分

age\_patient\_count = data['AgeRange'].value\_counts()

# 输出结果

print("年龄区间中高风险患者的比例和患者数:")

print(age\_risk\_rate)

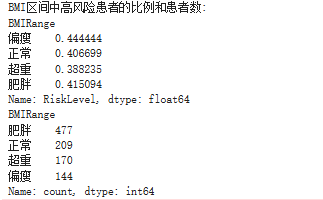
print(age\_patient\_count)

1.1.1-1.jpg

A group of black and white text

AI-generated content may be incorrect.

1.1.1-2.jpg



1.1.1-3.jpg

