**任务描述**

**描述：**

新冠肺炎自从2020年爆发以来到现在已经过去了四年，直到今天，新冠在我国东南沿海地区仍有抬头的趋势。那么，我们应该如何判断某人感染了新冠，是否有除发病症状外更客观、定量的指标去判断呢？假如不幸感染了新冠，我们也希望知道从哪些身体指标和发病症状入手来降低新冠的死亡率。现在需要你建立模型，根据体检指标判断某人是否感染新冠，如果感染新冠，那么当前的身体指标下，某人死亡的概率是多少？

**数据集：**

1）血液生物标志物水平变化对 COVID-19 患者预后的影响分析

1. **数据集详细信息**

数据集共分为两个sheet页，其中sheet1介绍了COVID-19 患者和非 COVID-19 患者的实验室数据，sheet2介绍了新冠肺炎患者死亡和康复病例数据。

病人的基本信息

|  |  |  |  |
| --- | --- | --- | --- |
| 类型 | 名称 | 含义 | 备注 |
| 新冠病人 | Patient Code COVID | 代码 | 以唯一数字区别 |
| Age Group COVID | 年龄 | 范围值，如40-50years |
| Gender COVID | 性别 | - |
| 对照组 | Patient Code CONTROL | 代码 | 以唯一数字区别 |
| Age Group CONTROL | 年龄 | 范围值，如40-50years |
| Gender CONTROL | 性别 | - |

实验室数据

|  |  |
| --- | --- |
| 名称 | 含义 |
| Patient Code | 病人代码 （P开头为新冠，C开头为对照组） |
| Group | 新冠或对照组 |
| Glucose | 葡萄糖（mg/dL） |
| Urea | 尿素（mg/dL） |
| Creatinine | 肌酐（mg/dL） |
| Sodium | 钠（mEq/L） |
| Potassium | 钾（mmol/L） |
| TB | 总胆红素（mg/dL） |
| DB | 直接胆红素（mg/dL） |
| ALT | 丙氨酸转氨酶（U/L） |
| AST | 天冬氨酸氨基转移酶（U/L） |
| ALP | 碱性磷酸酶（U/L） |
| Total Protein | 总蛋白（g/dL） |
| Albubin | 白蛋白（g/dL） |
| Ferritin | 铁蛋白（ng/mL） |
| Hemoglobin | 血红蛋白（g/dL） |
| WBC | 白细胞（十亿细胞/L） |
| Platelet | 血小板（十亿/L） |
| CRP | C反应蛋白（mg/L） |
| PCT | 降钙素原（ng/ml） |
| IL-6 | 白介素-6（pg/ml） |
| PT | 凝血酶原时间（秒） |
| D-Dimer | D-二聚体（ng/mL） |
| Troponin | 肌钙蛋白 |
| CPK-MB | 肌酸激酶-MB（U/L） |

死亡和康复病例数据

该数据部分仅有新冠病人的数据，主要探究各项指标对于新冠病人死亡的影响。

|  |  |
| --- | --- |
| 名称 | 含义 |
| Patient Code | 病人代码 |
| Deceased | 是否死亡 |
| Glucose | 葡萄糖（mg/dL） |
| Urea | 尿素（mg/dL） |
| Creatinine | 肌酐（mg/dL） |
| Sodium | 钠（mEq/L） |
| Potassium | 钾（mmol/L） |
| TB | 总胆红素（mg/dL） |
| DB | 直接胆红素（mg/dL） |
| ALT | 丙氨酸转氨酶（U/L） |
| AST | 天冬氨酸氨基转移酶（U/L） |
| ALP | 碱性磷酸酶（U/L） |
| Total Protein | 总蛋白（g/dL） |
| Albubin | 白蛋白（g/dL） |
| Ferritin | 铁蛋白（ng/mL） |
| Hemoglobin | 血红蛋白（g/dL） |
| WBC | 白细胞（十亿细胞/L） |
| Platelet | 血小板（十亿/L） |
| CRP | C反应蛋白（mg/L） |
| PCT | 降钙素原（ng/ml） |
| IL-6 | 白介素-6（pg/ml） |
| PT | 凝血酶原时间（秒） |
| D-Dimer | D-二聚体（ng/mL） |
| Troponin | 肌钙蛋白 |
| CPK-MB | 肌酸激酶-MB（U/L） |

1. **数据集任务**

根据表格中对照组和新冠组的数据训练一个分类模型和一个预测模型，判断测试样本是否属于新冠患者，如果属于新冠患者，预测他死亡的概率。

2）危重老年患者 COVID-19 的临床特征、危险因素及并发症——病例对照研究

**1.数据集详细信息**

数据集中数据共260条，样本全部为老年新冠患者。

基本信息

|  |  |
| --- | --- |
| 列 | 说明 |
| outcome | 结果，1表示存活, 2表示死亡 |
| Age | 年龄 |
| Residence | 居住地1表示住宅, 2表示养老院 |
| Social Category | 社会类别，1表示高于贫困线的人群(Above Poverty Line，APL),  2表示低于贫困线的人群(Below Poverty Line，BPL) |
| Sex | 性别 |
| Heart Rate | 心跳 |
| RR | 呼吸频率 |
| Systolic Blood Pressure | 收缩压 |
| Diastolic Blood Pressure | 舒张压 |
| spO2 at Room air | 室内空气中的 spO2 |

入院时的症状说明如下:

|  |  |  |
| --- | --- | --- |
| 症状 | 疾病 | 编号 |
| Fever | 发热 | 1=存在, 2=不存在 |
| Cough | 咳嗽 | 1=存在, 2=不存在 |
| Sore throat | 喉咙痛 | 1=存在, 2=不存在 |
| Runny nose | 流鼻涕 | 1=存在, 2=不存在 |
| Wheezing | 喘息 | 1=存在, 2=不存在 |
| Shortness of breath | 气短 | 1=存在, 2=不存在 |
| Lower chest wall indrawing | 胸壁收缩 | 1=存在, 2=不存在 |
| Chest pain | 胸痛 | 1=存在, 2=不存在 |
| Conjunctivitis | 结膜炎 | 1=存在, 2=不存在 |
| Lymphadenopathy | 淋巴结肿大 | 1=存在, 2=不存在 |
| Headache | 头痛 | 1=存在, 2=不存在 |
| Loss of smell (Anosmia) | 嗅觉丧失(嗅觉丧失) | 1=存在, 2=不存在 |
| Loss of taste (Ageusia) | 味觉丧失(味觉丧失) | 1=存在, 2=不存在 |
| Seizures | 癫痫发作 | 1=存在, 2=不存在 |
| Fatigue / Malaise | 疲劳/不适 | 1=存在, 2=不存在 |
| Anorexia | 食欲不振 | 1=存在, 2=不存在 |
| Altered consciousness/confusion | 意识改变/混乱 | 1=存在, 2=不存在 |
| Muscle aches (myalgia) | 肌肉疼痛(肌痛) | 1=存在, 2=不存在 |
| Joint pain (arthralgia) | 关节痛(关节痛) | 1=存在, 2=不存在 |
| Inability to walk | 不能走路 | 1=存在, 2=不存在 |
| Abdominal pain | 腹痛 | 1=存在, 2=不存在 |
| Diarrhoea | 腹泻 | 1=存在, 2=不存在 |
| Vomiting / Nausea | 呕吐/恶心 | 1=存在, 2=不存在 |
| Bleeding (Haemorrhage) | 出血(Haemorrhage) | 1=存在, 2=不存在 |
| If Bleeding Yes | 如果出血是 | 1=存在, 2=不存在 |

危险因素说明如下:

|  |  |  |
| --- | --- | --- |
| 病症 | 疾病 | 编号 |
| Chronic cardiac disease (not hypertension) | 慢性心脏病 (非高血压) | 1 = 存在, 2 = 不存在 |
| Hypertension | 高血压 | 1 = 存在, 2 = 不存在 |
| Chronic pulmonary disease (not asthma) | 慢性肺病 (非哮喘) | 1 = 存在, 2 = 不存在 |
| Asthma (physician diagnosed) | 哮喘 (经医师确诊) | 1 = 存在, 2 = 不存在 |
| Chronic kidney disease | 慢性肾病 | 1 = 存在, 2 = 不存在 |
| Obesity (as defined by clinical staff) | 肥胖 (由医疗人员定义) | 1 = 存在, 2 = 不存在 |
| Moderate or severe liver disease | 中度至重度肝病 | 1 = 存在, 2 = 不存在 |
| Mild liver disease | 轻度肝病 | 1 = 存在, 2 = 不存在 |
| Asplenia | 免疫缺陷综合症 | 1 = 存在, 2 = 不存在 |
| Chronic neurological disorder | 慢性神经系统疾病 | 1 = 存在, 2 = 不存在 |
| Malignant neoplasm | 恶性肿瘤 | 1 = 存在, 2 = 不存在 |
| Chronic hematologic disease | 慢性血液病 | 1 = 存在, 2 = 不存在 |
| Diabetes Mellitus | 糖尿病 | 1 = 存在, 2 = 不存在 |
| Rheumatologic disorder | 类风湿性疾病 | 1 = 存在, 2 = 不存在 |
| Dementia | 痴呆症 | 1 = 存在, 2 = 不存在 |

入院时实验数据说明如下:

|  |  |
| --- | --- |
| 指标 | 说明 |
| WBC count (x109/L) | 白细胞计数 (x10^9/L) |
| Lymphocyte count (109/L) | 淋巴细胞计数 (x10^9/L) |
| Neutrophil count (109/L) | 中性粒细胞计数 (x10^9/L) |
| HbA1c (%) | 血红蛋白A1c (%) |
| Creatinine (μmol/L) | 肌酐 (μmol/L) |
| Sodium (mmol/L) | 钠 (mmol/L) |
| Potassium (mmol/L) | 钾 (mmol/L) |
| Procalcitonin (ng/mL) | 前降钙素原 (ng/mL) |
| CRP (mg/L) | C反应蛋白 (mg/L) |
| LDH (U/L) | 乳酸脱氢酶 (U/L) |
| D-dimer (mg/L) | D-二聚体 (mg/L) |
| Ferritin (ng/mL) | 铁蛋白 (ng/mL) |
| IL-6 (pg/mL) | IL-6 (pg/mL) |

并发症说明如下：

|  |  |  |
| --- | --- | --- |
| 疾病 | 说明 | 编号 |
| Viral pneumonia/pneumonitis | 病毒肺炎/肺炎 | 1 = 存在, 2 = 不存在 |
| Bacterial pneumonia | 细菌性肺炎 | 1 = 存在, 2 = 不存在 |
| Acute Respiratory Distress Syndrome | 急性呼吸窘迫综合症 | 1 = 存在, 2 = 不存在 |
| Pneumothorax | 气胸 | 1 = 存在, 2 = 不存在 |
| Pleural effusion | 胸腔积液 | 1 = 存在, 2 = 不存在 |
| Cryptogenic organizing pneumonia (COP) | 密闭性肺炎 (COP) | 1 = 存在, 2 = 不存在 |
| Bronchiolitis | 支气管炎 | 1 = 存在, 2 = 不存在 |
| Cardiac arrest | 心脏骤停 | 1 = 存在, 2 = 不存在 |
| Myocardial infarction | 心肌梗死 | 1 = 存在, 2 = 不存在 |
| Cardiac ischaemia | 心肌缺血 | 1 = 存在, 2 = 不存在 |
| Cardiac arrhythmia | 心律失常 | 1 = 存在, 2 = 不存在 |
| Myocarditis / Pericarditis | 心包炎/心包炎 | 1 = 存在, 2 = 不存在 |
| Endocarditis | 心内膜炎 | 1 = 存在, 2 = 不存在 |
| Cardiomyopathy | 心肌病 | 1 = 存在, 2 = 不存在 |
| Congestive heart failure | 心力衰竭 | 1 = 存在, 2 = 不存在 |
| Seizure | 抽搐 | 1 = 存在, 2 = 不存在 |
| Stroke / Cerebrovascular accident | 中风 / 脑血管意外 | 1 = 存在, 2 = 不存在 |
| Meningitis / Encephalitis | 脑炎/脑膜炎 | 1 = 存在, 2 = 不存在 |
| Bacteremia | 细菌感染 | 1 = 存在, 2 = 不存在 |
| Coagulation disorder / DIC | 凝血障碍/DIC | 1 = 存在, 2 = 不存在 |
| Pulmonary embolism | 肺栓塞 | 1 = 存在, 2 = 不存在 |
| Anemia | 贫血 | 1 = 存在, 2 = 不存在 |
| Rhabdomyolysis / Myositis | 肌肉溶解/肌炎 | 1 = 存在, 2 = 不存在 |
| Acute renal injury/ Acute renal failure | 急性肾损伤/急性肾衰竭 | 1 = 存在, 2 = 不存在 |
| Gastrointestinal haemorrhage | 胃肠道出血 | 1 = 存在, 2 = 不存在 |
| Pancreatitis | 胰腺炎 | 1 = 存在, 2 = 不存在 |
| Liver dysfunction | 肝功能异常 | 1 = 存在, 2 = 不存在 |
| Hyperglycemia | 高血糖 | 1 = 存在, 2 = 不存在 |
| Hypoglycemia | 低血糖 | 1 = 存在, 2 = 不存在 |
| Respiratory failure | 呼吸衰竭 | 1 = 存在, 2 = 不存在 |
| Septic shock | 休克 | 1 = 存在, 2 = 不存在 |

**2.数据集任务**

根据表格中新冠患者的数据训练一个预测模型，预测他死亡的概率。

1. 新冠肺炎感染者的临床数据

**1.数据说明**

数据分析了符合纳入条件的患者的电子病历，获取了以下数据：人口统计学特征、合并症、年龄调整后的Charlson指数、首发症状及时间、入院情况、入院后一周内肺部CT扫描或胸部X光检查结果、诊断、严重程度（重症、危重）、实验室检查（包括COVID-19核酸检测CT值）、处方和用药记录（包括抗病毒药物Paxlovid、类固醇、抗凝剂、免疫治疗（Baricitinib或Tocilizumab））、ICU住院信息均有记录。

·人口统计学特征：年龄范围。

·基础疾病：每个数字代表一种特定的基础疾病或合并症情况，罕见的潜在影响疾病已单独注释。

·初期症状：代表COVID-19患者初期症状影响的系统。

·体温：确诊感染新冠肺炎第一天的体温，以摄氏度为单位。

·发热天数：发病第一天至第七天体温高于正常范围的天数。

·严重程度（D3）：发病第三天的疾病严重程度。世卫组织对重症和危重病例的定义如下：重症病例定义为符合以下任何一项标准：①室内空气中氧饱和度<90%；②肺炎征兆；③严重呼吸窘迫。危重病例定义为符合以下任何一项标准：①需要生命支持治疗；②急性呼吸窘迫综合征；③脓毒症；④感染性休克。

·肺部CT状况：“CT1-D1”表示第一次胸部CT扫描与确诊当天之间的时间差（以天为单位），“CT2-D1”可以类似方式推断。肺CT采用CT严重程度评分（CTSS）进行评估，根据肺叶受累程度将病情严重程度分为5组（1：<5%、2：5%~25%、3：26%~50%、4：51%~75%、5：>75%，范围：1~5），将5个肺叶的评分相加得到总评分。

·病毒核酸检测：新冠病毒核酸检测的CT值可以反映病毒拷贝数，N基因和ORF基因是目前新冠病毒核酸检测的靶基因，是新冠病毒核酸的特定片段。time1-D1表示从核酸检测到确诊第一天的天数，以此推算。阴性转阴描述患者新冠病毒检测结果转阴的情况。

·抗病毒药物：记录药物种类、用药时间（以确诊后天数表示）和用药天数，以及类固醇和抗凝剂的使用情况。此外，我们还详细记录抗凝剂的每日剂量。

**2.数据集任务**

根据表格中新冠患者的数据训练一个预测模型，预测他死亡的概率。

**评价指标：**

* + - * 1. 分类任务——判断某人是否属于新冠患者：

准确率（Acc）：衡量模型预测正确的比例。

F1分数：分类模型在精确率和召回率之间的调和平均值，尤其适用于数据不均衡的情况。

* + - * 1. 预测死亡率任务——预测新冠患者的死亡率：

平均绝对误差值（MAE）：预测值与实际值之间差值的绝对值的平均值

均方根误差（RMSE）：预测值与实际值之间差值的平方的均值的平方根

**排行榜规则**

1. 分类任务

分类任务分数计算：

分类任务排名规则：按照分类任务分数降序排名

1. 预测死亡率任务

预测死亡率任务分数计算：

预测死亡率任务排名规则：按照预测死亡率任务分数升序排名

3）总排名

队伍总排名=

例如，如果甲队伍在分类任务中排名为1，在预测任务1中排名为2，在预测任务2中排名为3，在预测任务3中排名为2，那么甲队伍最终排名为8/4 = 2

注：如果队伍排名相同，采取并列排名。

**其他要求：**

最终每支队伍需要提交代码和实验报告