

## 第 3 部分

### 理论知识复习题

#### 一、单选题

| 序号 | 标题  | 选项 A         | 选项 B           | 选项 C        | 选项 D           |
|----|---|--------------|----------------|-------------|----------------|
| 1  | 关于道德,正确的说法是   | 道德在职业活动中不起作用 | 道德在公共生活中几乎不起作用 | 道德威力巨大,无坚不克 | 道德是调节社会关系的重要手段 |
| 2  | 职业道德的特征是  | 范围上的有限性      | 规范制定上的任意性      | 内容上的多变性     | 形式上的单一性        |
| 3  | 在企业文化中,居于核心地位的是   | 文体活动         | 企业价值观          | 企业礼俗        | 员工服饰           |
| 4  | 职业道德是指从事一定职业劳动的人们,在特定的工作和劳动中以其( )和特殊社会手段来维系的,以善恶进行评价的心理意识、行为原则和行为规范的总和。   | 纪律约束         | 理想目标           | 评价标准        | 内心信念           |
| 5  | 各种职业集体对从业人员的道德要求,总是从本职业的活动和交往的内容和方式出发,适应于本职业活动的客观环境和具体条件。这说明职业道德具有( )的特点。 | 实践性          | 时代性            | 客观性         | 多样性            |
| 6  | 不同的历史时期,有不同的道德标准,职业道德也不例外。这表明了职业道德具有明显的( )特点。                             | 时代性          | 实践性            | 多样性         | 经济性            |

|    |  |                      |                               |                             |                                 |
|----|--|----------------------|-------------------------------|-----------------------------|---------------------------------|
| 7  | 某机械厂的一位领导说：“机械工业工艺复杂，技术密集，工程师在图纸上画得再好、再精确，工人操作中如果差那么一毫米，最终出来的就可能是废品。”这段话主要强调（ ）素质的重要性。 | 专业技能                 | 思想政治                          | 职业道德                        | 身心素质                            |
| 8  | 下列关于职业首先的说法中，你认为正确的是：  | 有职业道德的人一定能胜任工作。      | 没有职业道德的人干不好任何工作。              | 职业道德有时起作用，有时不起作用。           | 职业道德无关紧要，可有可无。                  |
| 9  | 下列有关职业道德的说法，不正确的是：   | 职业道德有利于协调职工与领导之间的关系。 | 职业道德有利于协调职工与企业之间的关系。          | 如果企业职工不遵守企业规章制度，是因为规章制度不合理。 | 职业道德是企业文化的重要组成部分。               |
| 10 | 关于职业道德对企业发展的积极作用，你认为正确的论述是：  | 职业道德是协调同事之间关系的法宝。    | 职业道德只能维系职工和领导的表面关系。           | 与物质激励相比，职业道德的作用较弱。          | 遵守职业道德有助于提高服务水平，但对提高产品的质量作用不明显。 |
| 11 | 职业道德是增加企业凝聚力的：   | 目的                   | 手段                            | 要求                          | 宗旨                              |
| 12 | 以下关于职业道德说法中，你认为正确的是：   | 任何职业道德的使用范围都是普遍的。    | 职业道德与职业责任无论在形式上还是目标上，都是完全一致的。 | 职业道德不具有连续性。                 | 某一特定行业的职业道德只适用于专门从事本职业的人。       |
| 13 | （ ）是社会主义道德最广泛的社会基础。  | 公民道德                 | 社会道德                          | 职业道德                        | 个人道德                            |
| 14 | （ ）是社会主义职业道德的重要规范，是职业道德的基础和基本精神。   | 诚实守信                 | 爱岗敬业                          | 服务群众                        | 奉献社会                            |

|    |   |                                    |                                     |                                     |                                       |
|----|---|------------------------------------|-------------------------------------|-------------------------------------|---------------------------------------|
| 15 | 在社会主义市场经济条件下，集体主义原则要求把社会集体利益与（ ）结合起来，以促进个人与社会的和谐发展。 | 国家利益                               | 个人利益                                | 小集体利益                               | 党的利益                                  |
| 16 | 职业道德的最基本要求是（ ），为社会主义建设服务。                           | 勤政爱民                               | 奉献社会                                | 忠于职守                                | 一心为公                                  |
| 17 | （ ）是贯通社会主义道德和共产主义道德的统一原则。                           | 个人主义原则                             | 集体主义原则                              | 马克思主义原则                             | 为人民服务原则                               |
| 18 | 社会主义职业道德体系的丰富内容，都是围绕着一个核心而展开的，这个核心就是（ ）。            | 为人民服务                              | 马克思主义                               | 社会主义荣辱观                             | 国家和人民利益                               |
| 19 | 办事公道是指职业人员在进行职业活动时，                                 | 原则至上，不徇私情，举贤任能，不避亲疏                | 奉献社会，襟怀坦荡，待人热情，勤俭持家                 | 支持真理，公私分明，公平公正，光明磊落                 | 牺牲自我，助人为乐，邻里和睦，正大光明                   |
| 20 | 下面不属于 python 特性的是（ ）。                               | 简单易学                               | 开源免费的                               | 属于低级语言                              | 高可移植性                                 |
| 21 | Python 中,关于逻辑运算 a or b 的描述错误的是（ ）。                  | 若 a=True<br>b=True 则 a or b ==True | 若 a=True<br>b=False 则 a or b ==True | 若 a=True<br>b=True 则 a or b ==False | 若 a=False<br>b=False 则 a or b ==False |
| 22 | Python 中使用关键字（ ）来创建自定义函数。                           | function                           | func                                | procedure                           | def                                   |
| 23 | （ ）不是用于处理中文的字符编码。                                   | gb2312                             | gbk                                 | big5                                | ASCII                                 |
| 24 | Python 类中定义私有变量的方法为（ ）。                             | 使用 <code>__private</code> 关键字      | 使用 <code>public</code> 关键字          | 使用 <code>__xxx</code> 定义变量名         | 使用 <code>__xxx_</code> 定义变量名          |
| 25 | Python 中,关键字（ ）表示匿名函数，是从数学里的 $\lambda$ 得名，          | lambda                             | map                                 | filter                              | zip                                   |

|    |   |             |   |                                   |                                   |
|----|---|-------------|---|-----------------------------------|-----------------------------------|
| 26 | Python 中, 内置函数 ( ) 可用于接受从键盘输入的字符串。                                      | accept ( )  | input ( )                                     | readline ( )                      | login ( )                         |
| 27 | Python 脚本文件的扩展名为 ( )。   | .python     | .py   | .pty                              | .pgh                              |
| 28 | Python 中, 与 $x > y$ and $y > z$ 语句等价的是 ( )                              | $x > y > z$ | $\text{not } x < y$<br>or $\text{not } y < z$ | $\text{not } x < y$<br>or $y < z$ | $x > y$ or<br>$\text{not } y < z$ |
| 29 | Python 字典每个键值对的键和值之间, 用 ( ) 分隔  | 空格          | :   | ;                                 |                                   |
| 30 | Python 中, 运算符不正确的使用项是: ( )  | $1 + 'a'$   | $[1, 2, 3]$<br>$+ [4, 5, 6]$                  | $3 * 'abc'$                       | $-10\%-3$                         |
| 31 | Python 中, 获得字符串 s 中字符个数的方法是什么? ( )                                      | s.len()     | s.length                                      | len(s)                            | length(s)                         |
| 32 | Python 中, 下列不能使用 for 循环进行遍历的是:  | set('str')  | (123)   | [1, 2, 3, 4, 5]                   | range(0, 10, 5)                   |
| 33 | Python 中, 可以使用函数 math.sqrt(x) 获得 x 的平方根, 如果 x 为负数, 则将产生 ( )             | 都不产生        | 虚数  | 程序崩溃                              | ValueError 错误                     |
| 34 | Python 中, 下列哪项数据类型数据是不可变化的:   | 集合(set)     | 字典(dict)                                      | 元组(tuple)                         | 列表(list)                          |
| 35 | Python 中, 对于字典 $d = \{ 'abc':1, 'qwe':2, 'zxc':3 \}$ , len(d) 的结果为: ( ) | 6           | 3   | 12                                | 9                                 |
| 36 | Python 中, 字符串是一个字符序列, 例如, 字符串 s, 从右侧向左第 2 个字符用什么索引?                     | s[:-2]      | s[-2]   | s[0:-2]                           | s[2]                              |
| 37 | 下面哪个不是 Python 合法的标识符:   | int32       | 4XL   | self                              | __name__                          |

|    |  |                  |                  |                      |                      |
|----|--|------------------|------------------|----------------------|----------------------|
| 38 | 关于 Python 中的字符串，下列说法错误的是：  | 字符应该视为长度为 1 的字符串 | 字符串以 \0 标志字符串的结束 | 既可以用单引号，也可以用双引号创建字符串 | 在三引号字符串中可包含换行回车等特殊字符 |
| 39 | Python 语句<br>print(1 if 'a' in 'ABC' else 2)<br>执行结果是 ( )  | 1                | 2                | 报错                   | 4                    |
| 40 | Python 中，不是 tuple 类型的是 ( )   | (1)              | (1,)             | ([], [1])            | (['a': 1], ['b', 1]) |
| 41 | Python 语句：<br>print((1, 2, '1', '2')[0] > 1)<br>执行结果是：   | True             | False            | 报错                   | 以上都不对                |
| 42 | 以下对 Python 中的数据类型字典的说法错误的是：  | 字典可以为空           | 字典的键不能相同         | 字典键值对中的键不可变          | 字典键值对中的值不可变          |
| 43 | 以下 Python 代码的执行结果是 ( )。<br>a={'name': 'hehe', 'detail': [{'age': 18, 'job': 'tester'}, 'man']}<br>print(a['detail'][1][2]) | 18               | n                | e                    | 以上都不对                |
| 44 | Python 中的集合数据类型，以下说法错误的是：  | 元组的长度可变          | 列表的长度是可变的        | 可以通过索引访问元组           | 可以通过索引访问列表           |
| 45 | 针对 Python 数据类型元组 (1, 2, [1, 2, '1', '2']) 的说法正确的是：   | 长度为 6            | 属于二维元组           | 元素可变                 | 嵌入的列表的值可变            |
| 46 | 计算机中信息处理和信息储存用：  | 二进制代码            | 十进制代码            | 十六进制代码               | ASCII 代码             |

|    |  |                              |                                  |                                  |                                   |
|----|--|------------------------------|----------------------------------|----------------------------------|-----------------------------------|
| 47 | Python 函数如下：<br>def<br>showNumber (numbers<br>):<br>for n in numbers:<br>print (n)<br>下面那些在调用函数<br>时会报错： | showNumber (<br>"1")         | showNumb<br>er([1])              | showNumb<br>er((1,))             | showNumb<br>er((1))               |
| 48 | Python 代码：<br>for i in range(0, 2):<br>print(i, end="")<br>输出结果是：  | 0 1 2                        | 0 1                              | 1 2 3                            | 1 2                               |
| 49 | 下列是 python 合法<br>标识符的是：  | 2v                           | variable1<br>223                 | \$another                        | if                                |
| 50 | Python 中，设有变量赋<br>值 x=3.5；y=4.6；<br>z=5.7，则以下的表达<br>式中值为 True 的是：  | x>y or x>z                   | x!=y                             | z>y+x                            | x<y and<br>not (x<z)              |
| 51 | 在一个 Python 应用程<br>序中定义<br>a=[1, 2, 3, 4, 5, 6, 7, 8,<br>9, 10]，为了打印输出<br>列表 a 的最后一个元<br>素，下面代码正确的<br>是：  | print(a[10])                 | print(a[1<br>en(a)])             | print(a[9<br>)                   | print(a(9<br>)                    |
| 52 | Python 中设 s="Happy<br>New Year"，则 s[3:8]<br>的结果是：  | 'ppy Ne'                     | 'py Ne'                          | '0ppy N'                         | 'py New'                          |
| 53 | Python 中，下面哪一个<br>是以只读模式打开文<br>件 f.txt?  | f=open('f.txt', 'r')         | f=open('f<br>.txt', 'w'<br>)     | f=open('f<br>.txt', 'a'<br>)     | f=open('f<br>.txt', 'b'<br>)      |
| 54 | Python 中，为了读取音<br>频文件 radio.mp3，打<br>开文件的方式正确的<br>是：   | f=open('radi<br>o.mp3', 'r') | f=open('r<br>adio.mp3'<br>, 'w') | f=open('r<br>adio.mp3'<br>, 'a') | f=open('r<br>adio.mp3'<br>, 'rb') |

|    |  |  |   |   |  |
|----|--|--|---|---|--|
| 55 | <p>Python 中代码片段:</p> <pre>def changeList(lst):     lst.append("end")     print("list : ",lst) strs = ['1', '2'] changeList(strs) print("strs : ",strs) </pre> <p>运行结果, 正确的是:</p>                                   | <pre>list : ['1', '2', 'end'] strs : ['1', '2'] </pre> | <pre>list : ['1', '2', 'end'] strs : ['1', '2', 'end'] </pre> | <pre>list : ['1', '2'] strs : ['1', '2'] </pre> | <pre>list : ['1', '2'] strs : ['1', '2', 'end'] </pre> |
| 56 | <p>Python 中代码片段:</p> <pre>def changeInt(number2):     number2 = number2 + 1 </pre> <pre>print("changeInt: number2=", number2) number1 = 2 changeInt(number1) print("number = ", number1) </pre> <p>运行结果, 正确的是:</p> | <pre>changeInt: number2= 3 number: 3 </pre>            | <pre>changeInt : number2= 2 number: 3 </pre>                  | <pre>changeInt : number2= 2 number: 2 </pre>    | <pre>changeInt : number2= 3 number: 2 </pre>           |
| 57 | <p>正确生成 Numpy 一维数组的语句是:</p>  | <pre>np. array[1, 2 , 3]</pre>                         | <pre>np. range ( 10)</pre>                                    | <pre>np. range ( 1, 10)</pre>                   | <pre>np. arange (10)</pre>                             |
| 58 | <p>正确生成 Numpy 二维数组的语句是:</p>  | <pre>np. array[[1, 2], [1, 2]]</pre>                   | <pre>np. array[ [1 2][1 2]]</pre>                             | <pre>np. array ( [1, 2][1, 2 ])</pre>           | <pre>np. array ( [[1, 2][1, 2]])</pre>                 |
| 59 | <p>对于 numpy 二维数组:</p> <pre>arr2d = np. random. randn(4, 3 )</pre> <p>获取数组中所有大于 0 的数, 正确的操作是:</p>   | <pre>arr2d&gt;0</pre>                                  | <pre>arr2d[arr 2d&gt;0]</pre>                                 | <pre>arr2d(arr 2d&gt;0)</pre>                   | <pre>arr2d[if arr2d&gt;0]</pre>                        |

|    |  |                       |   |   |   |
|----|--|-----------------------|---|---|---|
| 60 | 对于 numpy 二维数组：<br>arr2d =<br>np.random.randn(7, 4)<br>获取数组中第 1, 3, 7 行的数据，正确的操作是：    | arr2d[0, 2, 6]        | arr2d[0:2 :6, :]  | arr2d[[0, 2, 6], :]   | arr2d[[0: 2:6], :]  |
| 61 | 对于 numpy 二维数组：<br>arr2d =<br>np.random.randn(7, 4)<br>获取数组中第 1, 3, 5, 7 行的数据，正确的操作是： | arr2d[0, 2, 4, 6]     | idx=[True if i%2==0 else False for i in range(7)]<br>arr2d[idx] | idx=[True if i%2==1 else False for i in range(7)]<br>arr2d[idx] | idx=[True if i%2==1 else False for i in range(7)]<br>arr2d[[idx]] |
| 62 | 对于 numpy 二维数组：<br>arr2d =<br>np.random.randn(7, 4)<br>获取数组中第 1, 2 列的数据，正确的操作是：       | arr2d[:, 2]           | arr2d[:, 2, :]  | arr2d[:, : 2]   | arr2d[:, 2 ]  |
| 63 | Numpy 数组<br>n=np.arange(12).reshape(2, -1, 2)<br>那么，n.shape 的返回值是：                   | (2, -1, 2)            | (2, 2, 3)   | (2, 1, 6)   | (2, 3, 2)   |
| 64 | Numpy 中创建全为 0 的数组，使用：  | zeros                 | ones  | empty   | eye   |
| 65 | Numpy 中创建对角线上全是 1，其他全是 0 的数组，使用：   | zeros                 | ones  | identity  | empty   |
| 66 | Numpy 中把多维数组 arr_nd 转变为一维数组的方法是：   | transpose             | reavel  | arange  | random  |
| 67 | Numpy 中针对数组的统计方法，下列说法错误的是：   | sum 对数组全部，或某轴向的元素，求和。 | mean 对数组全部，或某轴向的元素，求算术平均值。                                      | max 对数组全部，或某轴向的元素，求最大值。   | argmax 对数组全部，或某轴向的元素，求最大值。  |
| 68 | Numpy 读取文本文件的方法是：  | load                  | save  | loadtxt   | genfrom   |



|    |  |                  |                  |                      |                  |
|----|--|------------------|------------------|----------------------|------------------|
| 69 | Python 中创建了 DataFrame 对象 pd 后, pd.tail() 是用来:  | 用来创建数据           | 用来分析数据           | 用来可视化数据              | 用来显示数据           |
| 70 | Python 中创建了 DataFrame 对象 pd 后, pd.min() 是用来:   | 寻找元素最小值          | 寻找每行最小值          | 寻找每列最小值              | 报错               |
| 71 | Pandas 中最简单的 Series 是由 ( ) 组成的?  | 一个数组             | 二个数组             | 三个数组                 | 四个数组             |
| 72 | Pandas 中, ser = pd.Series([1, 2, 3, 4], index=list("ABCD")) print(ser.C) 输出结果是:  | 1                | 2                | 3                    | 4                |
| 73 | Pandas 中: ser=pd.Series(range(10, 20)) print(ser[ser>17].values) 输出结果是:  | [18 19]          | [18, 19]         | [18 19 20]           | [18, 19, 20]     |
| 74 | Pandas 中: df = pd.DataFrame([[1, 2, 3], [4, 5, 6]], columns=list("ABC")) print(df.B.values) 输出结果是:   | [4 5 6]          | [1 4]            | [2 5]                | [3 6]            |
| 75 | Pandas 中: arr = np.arange(1, 25).reshape(6, 4) df = pd.DataFrame(arr, index=list("abcdef"), columns=list("ABCD")) print(df.iloc[:2, [1, 3]].values) 输出结果是: | [[1 3]<br>[5 7]] | [[1 2]<br>[5 6]] | [[2 3 4]<br>[6 7 8]] | [[2 4]<br>[6 8]] |

|    |  |                         |                       |                     |                               |
|----|--|-------------------------|-----------------------|---------------------|-------------------------------|
| 76 | 国务院在( )年发布的《中国制造 2025》中,就已经提出了“建立国家工业基础数据库,加强企业试验检验数据和计量数据的采集、管理、应用和积累。” | 2015                    | 2016                  | 2017                | 2018                          |
| 77 | 数据采集的英文缩写和含义是:   | DAO, Data Access Object | DAQ, Data acquisition | DB, DataBase        | DBMS, DataBase Manager System |
| 78 | 数据采集获取的数据类型,说法正确的是:  | 只能是结构化的数据               | 只能是结构化,半结构化的数据        | 只能是结构化,半结构化的,准结构化数据 | 可以是结构化,半结构化,准结构化和非结构化的数据      |
| 79 | 2013 年,国际数据管理协会(DAMA International)英国分会提出了数据质量的( )个核心维度。                 | 4                       | 5                     | 6                   | 7                             |
| 80 | 2013 年,国际数据管理协会(DAMA International)英国分会提出了数据质量的核心维度,以下中文含义和英文对应不正确的是:    | 完整性 (Consistency)       | 有效性 (Validity)        | 及时性 (Timeliness)    | 唯一性 (Uniqueness)              |
| 81 | 数据质量的核心维度,其中完整性指数据信息是否存在缺失的状况,下列数据表中常见情况不属于完整性评判的是:                      | 行的缺失                    | 数据记录不规范               | 字段的缺失               | 码值的缺失                         |
| 82 | 数据质量的核心维度,其中有效性一般指范围有效性、日期有效性、形式有效性等主要体现在数据记录的规范和数据是否符合逻辑。逻辑问题如:         | 多项数据间存在着逻辑问题。           | 数据记录不规范               | 数据表中行的缺失            | 数据表中字段的缺失                     |

|    |  |                            |                 |                             |                                 |
|----|--|----------------------------|-----------------|-----------------------------|---------------------------------|
| 83 | 数据质量的核心维度，其中准确性指数据记录的信息是否存在异常或错误。下列情况不属于准确性评判的是：             | 乱码                         | 异常大的数据          | 异常小的数据                      | 多项数据间存在着逻辑问题。                   |
| 84 | 数据质量的核心维度，其中及时性指数据从开始处理到可以查看的时间间隔。需要数据的及时更新和产出，不包括下列的：       | 实时业务大盘数据                   | 多项数据间存在着逻辑问题    | 机动响应特殊突发情况                  | 暴露业务指标的异常波动                     |
| 85 | 数据质量的核心维度，其中一致性相同含义信息在多业务多场景是否具有 consistency。不属于数据实体不一致的情况是： | 数据编码不一致                    | 数据表中字段的缺失       | 命名及含义不一致                    | 生命周期不一致                         |
| 86 | 数据质量的核心维度，其中唯一性指在数据集中数据不重复的程度。下列关于唯一性说法正确的是：                 | 唯一性通过唯一数据条数，和总数据条数的百分比来衡量。 | 唯一性要求数据中不能存在重复。 | 唯一性通过唯一数据条数，和不唯一数据条数的比率来衡量。 | 唯一性通过不唯一数据条数，和总数据条数的百分比来衡量。     |
| 87 | 我国在（ ）年的国标 GB/T38667-2020 中对数据质量做了明确的规范要求。                   | 2018                       | 2019            | 2020                        | 2021                            |
| 88 | GB/T38667-2020 按照数据质量分类的要素，不包括：                              | 数据的准确性                     | 数据的完整性          | 数据的一致性                      | 数据的有效性                          |
| 89 | GB/T38667-2020 按照数据质量分类的要素，不包括：                              | 数据的重复性                     | 数据的唯一性          | 数据的及时性                      | 数据的一致性                          |
| 90 | 数据采集规范通常不涉及下列中的：   | 采集时间                       | 采集范围            | 规范性引用文件                     | 采集要求                            |
| 91 | 自动驾驶车辆道路测试数据采集规范中，规范性引用文件不包括：                                | GB14886 道路交通信号灯设置与安装规范     | GB14887 道路交通信号灯 | GB/T40429-2021 汽车驾驶自动化分级    | GB/T38619-2020 工业物联网数据采集结构化描述规范 |

|    |   |                               |                                      |                              |                                   |
|----|---|-------------------------------|--------------------------------------|------------------------------|-----------------------------------|
| 92 | 自动驾驶车辆道路测试数据采集规范中，术语和定义部分，中英文不匹配的是：                 | 测试车辆坐标系（VUT coordinates）      | 自动驾驶功能（automated driving function）   | 测试主体（testing scenario）       | 自动驾驶模式（automated driving mode）    |
| 93 | 自动驾驶车辆道路测试数据采集规范中，采集技术要求不包括：                        | 数据采集工作人员具备良好的职业道德和沟通能力。       | 采集车辆位置相关数据                           | 采集车辆状态数据                     | 采集路径规划数据                          |
| 94 | 自动驾驶车辆道路测试数据采集规范中，测试场地和环境要求内容中包括：                   | 数据采集工作人员应符合的要求                | 交通信号灯数据                              | 车辆位置相关数据                     | 测试场地要求                            |
| 95 | 自动驾驶数据采集员的岗位职责中，不包括：                                | 根据业务方的实际采集需求，制定外业采集计划和详细行车路线。 | 按照采集规范完成外业数据采集，并对数据进行整理、检查，确保数据的完整性。 | 熟悉网站前后端交互技术和原理，能独立完成前端页面的开发。 | 及时学习并掌握技术变更，工作中保持积极思考，及时提出技术改进建议。 |
| 96 | 自动驾驶数据采集员的岗位职责中，不包括的职责是：                            | 高效完成部门、团队安排的各项工。作。            | 负责配置网站数据抓取规则。                        | 整理每日工作日志，定期输出工作总结和报告。        | 熟悉采集系统各个模块的工作原理，及时的预判和解决设备问题      |
| 97 | 大数据基本呈现出的4V特征不包括：                                   | 体量                            | 多样性                                  | 稳定性                          | 多变性                               |
| 98 | 制造业数据除了具备传统的大数据“4V”共性特点以外，还兼具了体现制造业特点的“3M”特性，3M不包括： | 多来源                           | 多样性                                  | 多维度                          | 多噪声                               |

|     |  |                     |           |                     |           |
|-----|--|---------------------|-----------|---------------------|-----------|
| 99  | 工业企业数据分类维度,来源于研发数据域,数据类型最有可能归类于:                 | 控制信息                | 物流数据      | 系统设备资产信息            | 研发设计数据    |
| 100 | 工业企业数据分类,来源于生产数据域,数据类型最有可能归类于:                   | 系统日志                | 开发测试数据    | 产品售后服务数据            | 产品供应链数据   |
| 101 | 工业企业数据分类维度,来源于生产数据域,数据类型最有可能归类于:                 | 客户与产品信息             | 控制信息      | 物流数据                | 开发测试数据    |
| 102 | 工业企业数据分类维度,来源于运维数据域,数据类型最有可能归类于:                 | 系统设备资产信息            | 系统日志      | 物流数据                | 研发设计数据    |
| 103 | 平台企业数据分类维度,来源于平台运营数据域,数据类型最有可能归类于:               | 物联网采集数据             | 客户数据      | 业务合作数据              | 人事财务数据    |
| 104 | 平台企业数据分类维度,来源于企业管理数据域,数据类型最有可能归类于:               | 物联网采集数据             | 人事财务数据    | 知识库模型库数据            | 研发数据      |
| 105 | 企业资源计划的英文缩写是:                                    | SCM                 | PLM       | ERP                 | CRM       |
| 106 | CRM 的意思是:  | 企业资源计划              | 产品生命周期管理  | 供应链管理               | 客户关系管理    |
| 107 | 根据企业生产阶段划分数据来源,来源于企业内部信息(产业链内部数据),数据类型最有可能归类于:   | 装备、物料及产品加工过程的工况状态参数 | 使用、运营情况数据 | 企业资源计划              | 供应商名单     |
| 108 | 根据企业生产阶段划分数据来源,来源于企业内部信息(产业链内部数据),数据类型最有可能归类于:   | 供应商名单               | 能耗管理系统    | 环境参数                | 外部的互联网    |
| 109 | 根据企业生产阶段划分数据来源,来源于物联网信息(工业生产过程中的数据),数据类型最有可能归类于: | 供应链管理               | 客户关系管理    | 装备、物料及产品加工过程的工况状态参数 | 使用、运营情况数据 |

|     |   |                               |   |                                 |                           |
|-----|---|-------------------------------|---|---------------------------------|---------------------------|
| 110 | 根据企业生产阶段划分数据来源,来源于企业外部信息(产品售后数据),数据类型最有可能归类于: | 供应商名单                         | 装备、物料及产品加工过程的工况状态参数                           | 能耗管理系统                          | 供应链管理                     |
| 111 | 工业数据分为一级、二级、三级3个级别,属于等级一的描述是:                 | 易引发特别重大生产安全事故,对国民经济造成严重影响。    | 易引发较大或重大生产安全事故,给企业造成较大负面影响                    | 对工业控制系统及设备、工业互联网平台等的正常运行影响较小    | 引发的级联效应明显,影响范围涉及多个行业、区域   |
| 112 | 工业数据分为一级、二级、三级3个级别,属于等级二的描述是:                 | 易引发突发环境事件,对公众利益造成严重影响         | 恢复工业数据或消除负面影响所需付出的代价较大。                       | 给企业造成负面影响较小,或直接经济损失较小;          | 恢复工业数据或消除负面影响所需付出的代价较小。   |
| 113 | 工业数据分为一级、二级、三级3个级别,下列描述属于等级二的是:               | 易引发特别重大生产安全事故,对国民经济造成严重影响。    | 易引发突发环境事件,对公众利益造成严重影响                         | 引发的级联效应明显,影响范围涉及多个行业,导致大量个人信息泄露 | 给企业造成负面影响较小,或直接经济损失较小     |
| 114 | 工业数据分为一级、二级、三级3个级别,属于等级三的描述是:                 | 易引发突发环境事件,对公众利益造成严重影响         | 恢复工业数据或消除负面影响所需付出的代价较大。                       | 给企业造成负面影响较小,或直接经济损失较小;          | 恢复工业数据或消除负面影响所需付出的代价较小。   |
| 115 | 依据数据的分类分级原则,数据常规处理应该遵循的规范。在数据整理阶段,不正确的说法是:    | 对数据进行格式和名称的规范,保障后续的归类和汇总的有效性。 | 同一事物使用不同名称会造成重复归类或者汇总误差,不同事物命名相同,则不会造成数据整理问题。 | 变量格式应该具备有效性,能够用于后续的归类和汇总工作。     | 同一事物不能使用不同名称,不同事物的名称不能相同。 |

|     |  |  |  |   |   |
|-----|--|--|--|---|---|
| 116 | 数据常规处理的归类，归类原则可以参考 MECE 原则。说法错误的是：           | 所有的数据都必须涵盖全，不能遗留   | 分类之间不允许重复和交叉   | 因为数据的复杂性，分类之间允许一定程度的重复和交叉。                | 同级次分类的维度要统一，颗粒度要一致。   |
| 117 | 下列操作属于数据常规处理的整合操作的是：                         | 按照类别，对数据进行排列组合，获取有价值的信息。                                       | 把性质或参数相同的数据划分在一起。  | 对数据进行格式和名称的规范                             | 按照归类的类别，把同类数据合并在一起。   |
| 118 | 工业数据采集集中，DTU 是指：                             | 无线远程采集终端   | 机器人数据采集  | 数据传输终端                                    | 数控机床数据采集  |
| 119 | 数据采集流程框架中，起始点是：                              | 搭建设备，场景  | 预采集  | 制订采集方案                                    | 执行采集  |
| 120 | 数据采集流程框架中，终点是：                               | 质量审核   | 数据交付   | 执行采集                                      | 优化采集方案  |
| 121 | 数据采集流程框架中，优化执行方案是紧跟：                         | 制定执行方案之后展开   | 搭建设备，场景之后展开  | 执行采集之后，质量审核不合格展开                          | 预采集之后，发现采集效果不佳展开  |
| 122 | 互联网数据采集流程中，可以执行发起请求的工具具有：                    | BeautifulSoup  | requests   | lxml                                      | 正则表达式   |
| 123 | 互联网数据采集，python 中使用 requests，正确的说法是：          | requests 不是 python 的内置库，也不需要额外安装，只需在使用之前导入：<br>import requests | requests 是 python 的内置库，不需要额外安装，只需在使用之前导入：<br>import requests | requests 是 python 的内置库，不需要额外安装，使用之前也无需导入。 | requests 不是 python 的内置库，需要使用 pip 安装：<br>pip install requests<br>在使用之前需要导入：<br>import requests |
| 124 | 互联网数据采集，如果发出请求后收到的响应 Response 中的状态码是 404，表示： | 请求已成功  | 重定向  | 请求错误                                      | 服务器错误   |

|     |  |   |  |  |  |
|-----|--|---|--|--|--|
| 125 | 互联网数据采集, 如果发出请求后收到的响应 Response 中的状态码是 200, 表示:         | 请求已成功                                     | 重定向  | 请求错误   | 服务器错误  |
| 126 | 互联网数据采集, 如果发出请求后收到的响应 Response 中的状态码是 500, 表示:         | 请求已成功                                     | 重定向  | 请求错误   | 服务器错误  |
| 127 | 互联网数据采集, 发出请求后收到的请求成功的响应, 需要对响应的数据进行解析。常用的解析工具是下列中的:   | requests                                  | Mysql  | lxml   | mongodb  |
| 128 | 互联网数据采集, 采集的数据保存到数据库中, 可以选择的数据库是:                      | BeautifulSoup                             | lxml   | requests   | mongodb  |
| 129 | 互联网数据采集, 关于内容解析工具 lxml, 说法不正确的是:                       | lxml 是第三方库, 使用之前需要安装:<br>pip install lxml | lxml 不是 python 内置库, 使用之前需要导入:<br>import lxml | lxml 是 python 的一个解析库, 支持 HTML 和 XML 的解析, 支持 XPath 解析方式, 而且解析效率非常高。 | lxml 是 python 的一个解析库, 支持 HTML 和 XML 的解析, 解析效率非常高, 但不支持 Xpath 解析方式。 |
| 130 | 互联网数据采集, 采集的数据可以保存到数据库。下列中不属于关系数据库的是:                  | mongodb                                   | Mysql  | sqlite   | oracle   |
| 131 | 互联网数据采集, 常常使用 JSON 进行数据的传输。在 Python 中, 关于 JSON 说法错误的是: | python 内置了 json 库, 在使用前导入:<br>import json | json.dumps 可以把 JSON 字符串解码成 python 对象。        | json.loads 可以把 JSON 字符串解码成 python 对象。                              | Python 中要使用 json 库, 需要通过 pip 进行安装:<br>pip install json             |



|     |  |   |                                   |                                    |                                    |
|-----|--|---|-----------------------------------|------------------------------------|------------------------------------|
| 132 | 互联网数据采集, 要把采集到的图片数据保存到文件 a.jpg 中, 打开文件的方式是:            | <code>f=open("a.jpg", "a")</code>                   | <code>f=open("a.jpg", "r")</code> | <code>f=open("a.jpg", "wb")</code> | <code>f=open("a.jpg", "rb")</code> |
| 133 | 语音采集流程中, 输入语音, 通过拾音器获得的是:                              | 音频文件  | 原始模拟信号                            | 原始数字信号                             | 信号编码                               |
| 134 | 工业相机实时采集图像的流程中, 选择采集模式, 属于内触发采集的是:                     | 连续采集  | 软触发                               | 硬触发                                | 混合触发                               |
| 135 | 音频数据采集, 关于拾音器的优化方案, 正确的是:                              | 场景是教室时, 拾音器考虑安装在教室中央。                               | 场景是汽车内部, 拾音器考虑安装在后排中央的背面位置。       | 场景是汽车内部, 拾音器考虑安装在副驾驶背面位置。          | 场景是教室时, 拾音器考虑安装在讲台附近。              |
| 136 | 数据库结构优化不仅可以提高数据存储效率, 还能有效提高数据查询速度。下列操作中, 不属于数据库结构优化的是: | 把数据表分为实时表和历史表, 并对实时表设定存储时间范围, 通过后台服务程序自动完成实时表迁移到历史表 | 采集端部署大量数据库, 并且考虑负载均衡。             | 精简日志中的字段                           | 使用数据库分区, 减少查询量                     |
| 137 | 聚类后的样本, 同簇内的样本具有 ( ) 的特征                               | 相同  | 相似                                | 一致                                 | 不相关                                |
| 138 | 聚类后样本的簇大小是指簇中样本的 ( )                                   | 密度  | 紧密度                               | 数量                                 | 特征                                 |
| 139 | 以下不属于基于密度的算法的聚类算法的是 ( )                                | DBSCAN 算法   | Optics 算法                         | GMM 模型算法                           | DENCLUE 算法                         |
| 140 | 如果有 m 个点, DBSCAN 在最坏的情况下时间复杂度为 ( )                     | $O(m)$  | $O(m \log m)$                     | $O(m^2)$                           | $O(\log m)$                        |

|     |   |           |          |           |           |
|-----|---|-----------|----------|-----------|-----------|
| 141 | 关于K均值聚类的描述正确的是（ ）<br>① K均值对簇中心初始化非常敏感<br>② 初始化不良会导致收敛速度差<br>③ 初始化不良可能导致整体聚集不良   | ①③        | ①③       | ②③        | ①②③       |
| 142 | 可以用下面哪一种方法来获得和全局最小值有关的K均值算法的良好结果（ ）<br>① 试着运行不同的质心初始化算法<br>② 调整迭代的次数<br>③ 找出最佳的簇数   | ①③        | ①③       | ②③        | ①②③       |
| 143 | 执行聚类分析时，对聚类对象最少要有多少个变量或属性（ ）  | 0         | 1        | 2         | 3         |
| 144 | 一般来说，聚类分析过程包括（ ）<br>① 数据准备<br>② 特征选择<br>③ 特征提取<br>④ 聚类分组<br>⑤ 结果评估  | ①②③④⑤     | ①③④⑤     | ①②④⑤      | ①②③④      |
| 145 | 下面对基于质心的K均值聚类分析算法和基于分布的期望最大化聚类分析算法的描述，哪些是不正确的？<br>① 都从随机初始化开始<br>② 都是可迭代算法<br>③ 两者对数据点的假设很强<br>④ 都对异常值敏感<br>⑤ 期望最大化算法是K均值的特殊情况<br>⑥ 都需要对所需要的簇数有先验知识<br>⑦ 结果是不可再现的 | ①②③④      | ⑤        | ①③④       | ⑥⑦        |
| 146 | 相关系数的取值范围是（ ）   | $[-1, 1]$ | $[0, 1]$ | $(-1, 1)$ | $[-1, 0]$ |

|     |   |        |              |              |           |
|-----|---|--------|--------------|--------------|-----------|
| 147 | 两个随机变量之间线性依赖关系的程度越低，则相关系数<br>( )                    | 越接近-1  | 越接近+1        | 越接近 0        | 越接近 0.5   |
| 148 | 两个随机变量之间线性依赖关系的程度高，则相关系数<br>( )                     | 越接近-1  | 越接近+1        | 越接近 0        | 越接近-1和+1  |
| 149 | 变量之间的相关关系按其形式可分为<br>( )                             | 正或负相关  | 单或复相关        | 线性或非<br>线性相关 | 无相关       |
| 150 | 能够测定变量之间相关关系密切程度的主要方法是<br>( )                       | 相关表    | 相关关系         | 定量分析         | 定性分析      |
| 151 | Pearson 相关系数是最早由英国数学家，生物统计学家 ( ) 设计的统计指标            | 卡尔·皮尔逊 | 查尔斯·爱德华·斯皮尔曼 | 莫里斯·乔治·肯德尔   | 阿尔伯特·爱因斯坦 |
| 152 | Pearson 相关系数是最早由<br>( ) 数学家，生物统计学家卡尔·皮尔逊设计的统计指标     | 美国     | 英国           | 德国           | 法国        |
| 153 | Pearson 相关系数 $\rho$ (x, y) 的范围<br>( )               | [0, 1] | (0, 1]       | [0, 1)       | (0, 1)    |
| 154 | Pearson 相关系数 $\rho$ (x, y) 的越大，则 X, Y 变量的<br>( ) 越强 | 相似性    | 相关性          | 随机性          | 相似度       |
| 155 | spearman 相关系数是以英国理论和实验心理学家 ( ) 而命名的。                | 卡尔·皮尔逊 | 查尔斯·爱德华·斯皮尔曼 | 肯德尔；肯达尔·詹娜   | 阿尔伯特·爱因斯坦 |
| 156 | spearman 相关系数是以 ( ) 理论和实验心理学家查尔斯·爱德华·斯皮尔曼而命名的。      | 美国     | 英国           | 德国           | 法国        |

|     |   |              |              |               |          |
|-----|---|--------------|--------------|---------------|----------|
| 157 | 在实际计算 Spearman 相关系数时,需要先对原始变量数据进行 ( ) 排列                                    | 升序           | 降序           | 乱序            | 随机       |
| 158 | 当变量 X 和 Y 完全单调相关时,Spearman 相关系数的绝对值为 ( )                                    | 0            | 0.5          | 0.8           | 1        |
| 159 | Kendall 相关系数是由 ( ) 数学家,生物统计学家莫里斯·乔治·肯德尔设计的统计指标                              | 美国           | 英国           | 德国            | 法国       |
| 160 | 如果数据具有正态性,此时首选( )   | Pearson 相关系数 | Kendall 相关系数 | Spearman 相关系数 | 都可以      |
| 161 | 当自变量 X 的数值确定时,因变量 Y 的值也随之完全确定,则这种关系属于( )                                    | 相关关系         | 函数关系         | 回归关系          | 随机关系     |
| 162 | 在相关分析中,要求相关的两个变量 ( )  | 都是随机的        | 都不是随机的       | 因变量是随机变量      | 自变量是随机变量 |
| 163 | 在进行变量相关分析过程中,其具体分析过程有以下几个常见步骤 ( ) ① 数据标准化 ② 绘制散点图 ③ 计算相关系数 ④ 显著性检验 ⑤ 进行业务判断 | ①②③④⑤        | ①②③④         | ①②③⑤          | ①②③      |
| 164 | 在进行变量相关分析过程中,数据标准化环节需要将数据中存在的不同度量规模的特征(即变量)进行转化,使之具有相同的 ( )                 | 度量尺度         | 数量           | 顺序            | 维度       |

|     |   |                                |                        |                                |                        |
|-----|---|--------------------------------|------------------------|--------------------------------|------------------------|
| 165 | 在进行变量相关分析过程中，绘制散点图环节中绘制散点图环节作用是（ ）                        | 直观判断数据的各个特征间是否存在某种内在关系         | 美化输出效果                 | 间接判断数据的各个特征间是否存在某内在关系          | 直观判断数据的各个特征间是否存在某种外在关系 |
| 166 | 关联分析 (Association Analysis)，也叫关联规则挖掘，是属于一种（ ）算法           | 无监督                            | 半监督                    | 全监督                            | 机器学习                   |
| 167 | 统计分析案例中经典的啤酒与尿布的属于（ ）                                     | 关联关系                           | 相关关系                   | 聚类关系                           | 线性关系                   |
| 168 | 关联分析的主要目的就是从大量数据中找到不同事物或现象之间的（ ）                          | 关联关系                           | 相关关系                   | 聚类关系                           | 线性关系                   |
| 169 | 支持度用图表示就是两者之间的（ ）   | 并集                             | 交集                     | 子集                             | 补集                     |
| 170 | 假设某超市在 100 次的商品交易中同时出现了啤酒和尿布的次数是 20 次，那么此关联的支持度为（ ）       | 0.2                            | 0.4                    | 0.6                            | 0.8                    |
| 171 | 项集 X 和项集 Y 的关联规则 {X→Y} 置信度计算方法是（ ）                        | 项集 X 与项集 Y 同时出现的概率除以项集 X 出现的概率 | 项集 Y 出现的概率除以项集 X 出现的概率 | 项集 X 与项集 Y 同时出现的概率乘以项集 X 出现的概率 | 项集 Y 出现的概率乘以项集 X 出现的概率 |
| 172 | 某超市中某天共有 10 笔订单，其中购买 A 的次数是 8，同时购买 A 和 B 的次数是 6，则其置信度是（ ） | 0.25                           | 0.33                   | 0.5                            | 0.75                   |
| 173 | 在关联规则 {X→Y} 中，提升度是将其置信度除以 Y 的（ ）                          | 发生概率                           | 不发生的概率                 | 支持度                            | 项集数                    |

|     |  |               |               |             |                |
|-----|--|---------------|---------------|-------------|----------------|
| 174 | 关联分析过程的发现关联规则环节中就是按照（ ）原则来寻找关联规则。  | 支持度大于最小支持度的项集 | 置信度大于最小置信度的规则 | 置信度最小置信度的规则 | 支持度远大于最小支持度的项集 |
| 175 | Apriori 算法是一个经典的（ ）方法  | 关联分析          | 相关分析          | 聚类分析        | 回归分析           |
| 176 | Apriori 算法搜索关联规则的过程，为了提高搜索效率，一般通过逐层剪枝算法将低于（ ）项集剪枝掉                                       | 最小支持度         | 最大支持度         | 最小置信度       | 最大置信度          |
| 177 | Apriori 算法关联分析过程中，加速过程依赖于（ ）   | 抽样            | 剪枝            | 缓冲          | 并行             |
| 178 | Apriori 算法搜索频繁项集的过程是利用（ ），从 1 项集开始不断地查找频繁项集，直到找到最大的 k 项频繁项集。                             | 最小支持度         | 最大支持度         | 最小置信度       | 最大置信度          |
| 179 | 在线性回归分析中，需要通过（ ）求取优化参数   | 最小二乘法         | 最大自然估计法       | 最小梯度下降法     | 最大梯度下降法        |
| 180 | 岭回归的参数寻优可以直接通过（ ）迭代求解  | 最大自然估计法       | 坐标下降法         | 梯度下降法       | 最小二乘法          |
| 181 | 回归分析的一般步骤（ ）① 根据预测目标，确定自变量和因变量 ② 绘制散点图，确定回归模型类型 ③ 估计模型参数，建立回归模型 ④ 对回归模型进行检验 ⑤ 利用回归模型进行预测 | ①②③④⑤         | ①③④⑤          | ①②③④        | ①②③⑤           |

|     |   |             |         |            |           |
|-----|---|-------------|---------|------------|-----------|
| 182 | 为了提高数据集的质量、确保数据标注的准确性、推动数据标注任务的顺利验收，通常会对标注后的数据进行（ ） | 质量检验        | 数量检验    | 顺序检验       | 优劣检验      |
| 183 | 相关数据显示，数据标注质量达到 98% 的时候，机器学习的训练效果为（ ）。              | 0.6         | 0.7     | 0.8        | 0.9       |
| 184 | 机器学习或者深度学习的算法，在很大程度上依赖数据集的（ ）                       | 质量          | 充裕度     | 简单度        | 复杂度       |
| 185 | 质量检验方法不包括（ ）  | 逐条抽取检查      | 抽样检查    | 机器检查       | 人工检查      |
| 186 | 下列不属于质量检验方法的是（ ）                                    | 随机检查        | 抽样检查    | 机器检查       | 逐条抽取检查    |
| 187 | 在半人工标注方案中，特征工程或自编码技术对标注数据是有损的，这样不可避免地引入（ ）          | 检验误差        | 工程误差    | 标注误差       | 主观误差      |
| 188 | 由于不同的标注人员和标注团体存在人工主观判定尺度不一致问题，从而导致标注数据（ ）难以保证。      | 统一性         | 完整性     | 主观性        | 客观性       |
| 189 | 一些特定行业的专业数据，标注技术门槛高，如果存在（ ）情况下很容易出现标注误差甚至标注错误。      | 人工主观判定尺度不一致 | 有损的标注数据 | 数据的准确性难以控制 | 相关专业储备不充足 |
| 190 | 按照是否人工参与，标注数据质量检验可以分为（ ）和自动质量检验。                    | 半自动重量检验     | 全样质量检验  | 人工质量检验     | 抽样质量检验    |

|     |   |            |        |        |        |
|-----|---|------------|--------|--------|--------|
| 191 | 人工质量检验是指由数据审核人员，按照数据标注要求和规范，对标注人员标注的（ ）进行人工核对和校验          | 所有数据       | 部分数据   | 抽样数据   | 少量数据   |
| 192 | （ ）是通过相关软件平台自带的校验工具，对数据审核的校验逻辑和标准进行人工设定后，再通过算法对数据进行智能审核   | 抽样质量检验     | 自动质量检验 | 全样质量检验 | 人工质量检验 |
| 193 | 自动质量检验是通过相关（ ）自带的校验工具，对数据审核的校验逻辑和标准进行人工设定后，再通过算法对数据进行智能审核 | 软件平台       | 信息平台   | 人工平台   | 应用平台   |
| 194 | 自动质量检验方法通过算法对数据进行智能审核时会根据不同的数据类型，调用相应的算法对人工标注的数据进行（ ）检查   | 全自动或者半自动质量 | 半自动质量  | 全自动    | 人工质量   |
| 195 | 为了提高检查的效率，在实际项目中往往会引入（ ）                                  | 逐条抽取检查     | 抽样检查   | 机器检查   | 随机检查   |
| 196 | 全样质量检验是对所有标注数据进行检验，它需要质检员对已完成标注的数据集，严格按照数据标注的（ ）进行检验。     | 数量标准       | 质量标准   | 顺序标准   | 重量标准   |
| 197 | 在全样质量检验过程中，需要标注员对于不合格的数据标注进行（ ）。                          | 返工重新检验     | 返工重新录入 | 返工重新标注 | 直接删除   |



|     |   |          |          |          |          |
|-----|---|----------|----------|----------|----------|
| 198 | 特别是对于许多数据量比较大的标注项目进行质量检验而言，（ ）是常用方法                   | 随机质量检验   | 全样质量检验   | 人工质量检验   | 抽样质量检验   |
| 199 | 下列不属于抽样检查的是（ ）  | 简单抽样     | 系统抽样     | 分批抽样     | 分层抽样     |
| 200 | 抽样检查不包括（ ）  | 简单抽样     | 系统抽样     | 随机抽样     | 分层抽样     |
| 201 | 在实际项目中，简单抽样检验的抽样概率与数量往往来自（ ）。                         | 客户要求     | 抽样人员要求   | 项目难度要求   | 数量规模要求   |
| 202 | 系统抽样一般要求每隔一段时间进行检验，然后再从抽取的每个时间间隔的数据样本中进行（ ）。          | 简单抽样     | 随机抽样     | 全样质量检验   | 复杂抽样     |
| 203 | 当数据标注任务需要采用实时检验方法，但质检员与标注员比例失衡，质检员数量不够时，可以采用（ ）辅助实时检验 | 多重抽样检验方法 | 单重抽样检验方法 | 辅助抽样检验方法 | 人工抽样检验方法 |
| 204 | 通过多重抽样检查辅助实时检验，可以让质检员重点检验那些（ ）的标注员                    | 合格率低     | 合格率高     | 优秀       | 平凡       |
| 205 | 在多轮次质量检查体系的自检和质检环节中，质检员可分设（ ）级                        | 1        | 2        | 3        | 4        |
| 206 | 在自检和质检环节，设3级质检员不包括（ ）                                 | 低级质检员    | 初级质检员    | 中级质检员    | 高级质检员    |
| 207 | 自检轮次环节由（ ）来完成质检工作                                     | 低级和中级质检员 | 低级和初级质检员 | 初级和中级质检员 | 中级和高级质检员 |
| 208 | 在自检轮次质检环节中，标注员完成标注任务后提交到（ ）                           | 低级质检员    | 初级质检员    | 中级质检员    | 高级质检员    |

|     |   |         |          |          |       |
|-----|---|---------|----------|----------|-------|
| 209 | 在多轮次质量检查体系中，项目质检组是由（ ）组成                            | 低级质检员   | 初级质检员    | 中级质检员    | 高级质检员 |
| 210 | 100%准确度的图像标注是要求标注像素点与标注物的边缘像素点存在（ ）个像素以内的误差         | 0       | 1        | 2        | 3     |
| 211 | 下面哪些不是图像标注框类的质检点（ ）                                 | 目标框是否贴合 | 目标类别是否正确 | 目标属性是否正确 | 顺序    |
| 212 | 下面哪个不是图像标注框类的质检点（ ）                                 | 目标框是否贴合 | 数量       | 目标属性是否正确 | 关联一致性 |
| 213 | 下面哪个不是图像标注框类的质检点（ ）                                 | 目标框是否贴合 | 目标类别是否正确 | 颜色       | 关联一致性 |
| 214 | 在框标注任务中，同一人头部标注框和身体标注框的（ ）要保持一致                     | 高度      | 宽度       | 尺寸编码     | 对象编码  |
| 215 | 在对图像数据进行标注时，需要对标注后的图像数据进行质量检测，下列哪个不属于质量检测项（ ）       | 目标框是否贴合 | 目标类别是否正确 | 目标属性是否正确 | 目标的顺序 |
| 216 | 在图像标注过程中，标注像素点越接近于标注物的（ ），标注的质量就越高，算法的训练效果也就越好      | 边缘位置    | 几何中心位置   | 重心位置     | 质心位置  |
| 217 | 对图像标注中，对车辆颜色的标注，属于（ ）标注                             | 属性标注    | 关键点标注    | 框选标注     | 区域标注  |
| 218 | 在图像标注过程中标注出人脸面部关键区域位置，包括眉毛、眼睛、鼻子、嘴脸部轮廓等，这个标注任务属于（ ） | 属性标注    | 关键点标注    | 框选标注     | 图像理解  |

|     |  |   |                |                |                   |
|-----|--|---|----------------|----------------|-------------------|
| 219 | 在图像标注中，对于属性类别标注（如车型）的评价指标是（ ）                  | 所选实体与事实是否一致   | 标注类别与实际类别是否一致  | 标注点数量与标注要求是否一致 | 颜色数量与区域数量是否一致     |
| 220 | 视频属性标注内容不包括（ ）                                 | 标注物位置   | 状态变化           | 视频长度           | 属性判断              |
| 221 | 以下不属于图像框选标注的标注内容（ ）                            | 框选数量  | 标注点位置          | 框选位置           | 框选目标              |
| 222 | 在图像标注中，对关键点数量的质量评价指标是（ ）                       | 所选实体与事实是否一致   | 标注类别与实际类别是否一致  | 标注点数量与标注要求是否一致 | 颜色数量与区域数量是否一致     |
| 223 | 在对人眼进行关键打点标注时，标注后的数据质量检验评价指标是（ ）               | 标注点位置和正确位置质检的位置差  | 标注点数量与标注要求是否一致 | 标注类别与实际类别是否一致  | 所选实体与事实是否一致       |
| 224 | 利用 2D 或 3D 框对图像中目标物体进行框选标注师，标注后的数据质量检验评价指标是（ ） | 所画框与要求是否一致  | 所选实体与事实是否一致    | 标注点数量与标注要求是否一致 | 颜色数量与区域数量是否一致     |
| 225 | 在对病灶。特定类别车辆、特殊年龄行人等目标进行框选标注时，标注质量侧重的评价指标是（ ）   | 框选物体数量与实际数量是否一致<br>+H108:J108H108:M108H108:L109H10H108:K108 | 所画框与要求是否一致     | 框选目标与要求目标是否一致  | 框选的误差情况，多框了多少无关像素 |
| 226 | 在对图像进行区域标注时，利用不同颜色标注不同区域目标，标注后的数据质量检验评价指标是（ ）  | 分割的数量与目标类别是否一致  | 所选实体与事实是否一致    | 分割误差           | 颜色数量与区域数量是否一致     |
| 227 | 在对视频关键帧标注过程中，标注后的数据质量检验评价指标是（ ）                | 标注类别与实际类别是否一致   | 对每一帧的类别标注是否正确  | 标注点数量与标注要求是否一致 | 分割误差是否满足要求        |
| 228 | 智能语音处理技术方向不包括（ ）                               | 身份识别  | 语种识别           | 情感识别           | 图片识别              |

|     |   |                 |                     |                   |                   |
|-----|---|-----------------|---------------------|-------------------|-------------------|
| 229 | 下列属于智能语音处理技术方向的是( ):<br>①身份识别 ②语音分离 ③情感识别 ④语种识别 ⑤语音合成 | ①②③④⑤           | ②③④⑤                | ②④⑤               | ①③④⑤              |
| 230 | 下列不属于语音数据标注项目的错误类型的是( )                               | 噪声干扰错误          | 截取错误                | 文本错误              | 图像错误              |
| 231 | 下列不属于语音合成任务的标注内容的是( )                                 | 文本校对            | 文本分词                | 音素标记              | 语音切分              |
| 232 | 在语音识别任务中,对语音进行切分标注,标注后的数据质量检验评价指标是( )                 | 切分后的内容与实际内容是否一致 | 切分后的音频是否包含信息        | 切分后的内容与说话着的语气是否相似 | 切分后的内容与说话着的语调是否相似 |
| 233 | 文本标注的对象是( )   | 自然语言文本          | 计算机程序               | 图片                | 自然语言语音            |
| 234 | 在文本数据标注中,首先是对( )进行筛选检验。                               | 文本内容            | 语料                  | 关键词               | 文本情感              |
| 235 | 在文本情感标注任务中,对情感类别标注内容的的数据质量检验评价指标是( )                  | 标注内容与真是内容是否一致   | 标注类别与真是类别是否一致       | 标注内容和正确内容的相似度     | 打分与真实情感分数的差异      |
| 236 | 数据标注质量的衡量尺是( )  | 数据标注规则          | 数据标注                | 数据采集              | 数据分类              |
| 237 | 以下关于知识整理的重要性说法错误的是( )。                                | 整理分类的知识便于记忆处理   | 整理知识能锻炼逻辑思维和结构化思维能力 | 知识整理最终目的只是为了整理    | 整理过的知识更便于查找和提取    |
| 238 | 以下关于知识整理的重要性说法正确的是( )。                                | 整理分类的知识便于记忆处理   | 整理知识能锻炼逻辑思维和结构化思维能力 | 经过整理后的知识会延缓遗忘     | 全部正确              |

|     |   |                                      |                      |                       |                     |
|-----|---|--------------------------------------|----------------------|-----------------------|---------------------|
| 239 | 以下对知识整理描述正确的是（ ）。                           | 把零散、碎片化的知识围绕某个主题结构化，建立起知识框架和知识体系的过程。 | 将知识从一个领域转移到另一个领域的过程。 | 通过学习获得新的知识。           | 将知识进行压缩和简化，使之更易于理解。 |
| 240 | 以下属于知识整理常见方法的是（ ）。                          | 图表式整理                                | 对比式整理                | 知识框架整理                | 以上都是                |
| 241 | 以下不属于知识整理的常用方法的是（ ）。                        | 图表式整理                                | 关键词整理                | 代码式整理                 | 要点式整理               |
| 242 | 知识整理的常见方法中不包括（ ）。                           | 图表式整理                                | 关键词整理                | 要点式整理                 | 分类式整理               |
| 243 | （ ）整理的主要优点是直观形象，提升记忆效率，稳固记忆效果。              | 图表式                                  | 对比式                  | 知识框架                  | 要点式                 |
| 244 | （ ）整理是以图表的形式对某一知识进行整理。                      | 图表式                                  | 对比式                  | 知识框架                  | 要点式                 |
| 245 | 当我们梳理一些框架和流程时，使用（ ）的整理方法会更容易理解和记忆，同时也更方便交流。 | 图表式                                  | 问题式                  | 知识框架                  | 要点式                 |
| 246 | 以下对于图表式整理的优缺点和适用范围描述正确的是（ ）。                | 图表式整理适用于任何知识的整理和表达                   | 图表式整理的缺点是可能会影响表达的准确性 | 图表式整理适用于对某一具体知识的识记、理解 | 图表式整理的优点是可以提升表达的条理性 |
| 247 | 对于相关性强、相似程度高的知识的理解可以使用（ ）整理。                | 问题式                                  | 对比式                  | 图表式                   | 要点式                 |
| 248 | （ ）整理是采取“比照”的形式，分清容易混淆的知识点的关系（区别与联系）。       | 问题式                                  | 对比式                  | 图表式                   | 要点式                 |

|     |  |     |     |     |      |
|-----|--|-----|-----|-----|------|
| 249 | 对于相关性强、相似程度高的知识可以选用（ ）的整理方式。                           | 图表式 | 问题式 | 对比式 | 要点式  |
| 250 | （ ）整理是提炼知识中的“观点性”语言，按知识既定框架作“去粗存精”式的整理。                | 问题式 | 对比式 | 图表式 | 要点式  |
| 251 | 对框题知识的及时掌握，比较合适的知识整理方法是（ ）整理。                          | 问题式 | 对比式 | 要点式 | 图表式  |
| 252 | （ ）整理的优点是容易操作，方便记忆，但知识的系统性、条理性不强。                      | 问题式 | 对比式 | 要点式 | 图表式  |
| 253 | （ ）整理的优点是一问一答，条理清楚，知识清晰，有利于培养主观题的答案习惯。                 | 问题式 | 对比式 | 要点式 | 图表式  |
| 254 | 可以使用像思维导图类的工具来进行知识整理，从而对单元知识点的系统掌握，该种知识整理方法为（ ）整理。     | 问题式 | 对比式 | 要点式 | 知识框架 |
| 255 | 打破课本单元、课题、框题、目题的设置，将相关性强的知识点以框架的形式进行整理，该种知识整理方法为（ ）整理。 | 问题式 | 对比式 | 要点式 | 知识框架 |
| 256 | （ ）整理相比于要点式整理，其知识的系统性、条理性更强，但对知识点的理解程度要求较高。            | 问题式 | 对比式 | 要点式 | 知识框架 |

|     |   |                 |                 |               |             |
|-----|---|-----------------|-----------------|---------------|-------------|
| 257 | 比如我们从某一文章中快速整理出核心关键词，通过这些关键词可以快速的对整篇文章有所了解。称之为（ ）整理。          | 图表式             | 对比式             | 关键词           | 知识框架        |
| 258 | （ ）整理是指以某一出现频率较高的关键词为线索，对分散于不同地方的相关知识进行整理。                    | 问题式             | 对比式             | 关键词           | 知识框架        |
| 259 | （ ）整理，能快速知识点的全盘掌握，但比较耗时。                                      | 图表式             | 对比式             | 关键词           | 知识框架        |
| 260 | 如写一时事的评论文章时可以围绕某一热点，从经济、文化、哲学、法律、道德等多角度、全方位进行解读。这个就是所谓的（ ）整理。 | 图表式             | 专题              | 对比式           | 要点式         |
| 261 | （ ）整理是指围绕某一专题将能够运用到该事件的相关知识进行整理。                              | 图表式             | 专题              | 对比式           | 要点式         |
| 262 | 某公司针对某一项产品对其所有相关的资料按照时间进行汇总和整理。该方法属于知识整理中的（ ）知识。              | 文档              | 图表              | 问答            | 图谱          |
| 263 | 以智能客服在线机器人为例，AI 训练师拿到收集好的知识后，一般通过（ ）存储形态来完成知识的整理。             | 文档知识            | 图谱型知识           | 问答知识          | 以上都是        |
| 264 | 以下哪个选项最符合知识整理工具的作用？（ ）  | 帮助用户记录个人信息和日程安排 | 收集和整理互联网上的文章和资源 | 整合和归纳各种知识点和经验 | 用于团队协作和项目管理 |

|     |  |              |                     |                 |               |
|-----|--|--------------|---------------------|-----------------|---------------|
| 265 | ( )是对调查、观察、实验等研究活动中所搜集到的资料进行检验、归类编码和数字编码的过程。                                 | 数据审核         | 数据整理                | 数据分析            | 数据验收          |
| 266 | ( )是根据统计研究的任务和要求,对统计调查搜集到的大量原始资料进行审核、分组、汇总,使之条理化、系统化,得出能够反映总体综合特征的统计资料的工作过程。 | 数据审核         | 数据整理                | 数据分析            | 数据验收          |
| 267 | 数据整理是统计工作的( )环节。   | 最早           | 中间                  | 后期              | 审查            |
| 268 | ( )是在统计调查的基础上进行的,也是积累历史资料的必要手段。  | 数据整理         | 数据审核                | 数据分析            | 数据验收          |
| 269 | 数据整理的主要作用是( )。   | 对原始数据进行统计分析  | 对原始数据进行检验、归类编码和数字编码 | 帮助人们更好地理解和使用数据  | 对已整理的资料进行再加工  |
| 270 | 对于调查得到的数据资料,在进行了审核与整理之后,调研人员就可以着手建立数据库了,这是数据整理步骤中的( )。                       | 根据研究目的设计整理方案 | 统计数据的审核与检查          | 数据分组和汇总,并计算各项指标 | 统计资料的积累、保管和公布 |
| 271 | 可能通过经过训练的算法模型快速而且准确地识别出来当前的图片中哪些是猫,哪些是狗。使用的是数据整理中( )技术。                      | 群集           | 分类                  | 聚类              | 预测            |
| 272 | 分类就是根据事物的属性取值将事物分成不同的类。( )技术就是指定对象,以确定集合。                                    | 群集           | 分类                  | 聚类              | 预测            |



|     |  |          |             |            |            |
|-----|--|----------|-------------|------------|------------|
| 273 | 以下属于数据整理的方法的是（ ）。  | 简单地查看数据  | 通过描述性统计分析数据 | 理解数据属性的相关性 | 以上都是       |
| 274 | （ ）不是常见的数据整理方法。  | 数据清洗和预处理 | 数据可视化和探索性分析 | 数据建模和算法应用  | 数据库设计和查询语言 |
| 275 | （ ）数据整理工具成为企业级数据集成及商业智能套件 Pentaho 的主要组成部分。                           | Excel    | Python      | Kettle     | Eclipse    |
| 276 | 以下属于数据整理工具的是（ ）。   | Excel    | Python      | Kettle     | 以上都是       |
| 277 | 智能应用中不包括（ ）。   | 云计算技术    | 人工智能技术      | 大数据技术      | 传统的手工操作技术  |
| 278 | （ ）基础是人工智能应用场景开发落地的前提。   | 功能       | 用户          | 产品         | 技术         |
| 279 | 实现人工智能目前最主流的方法则是机器学习和（ ）。  | 正强化学习    | 负强化学习       | 深度学习       | 消退学习       |
| 280 | 通过对已有技术的判断及重点需求的研判生成市场需求，为后续的人工智能应用场景设计做准备。属于人工智能应用场景的开发设计步骤中的场景（ ）。 | 发掘       | 生成          | 实现         | 应用         |
| 281 | 人工智能应用场景由（ ）、人、技术三要素组成。  | 企业       | 研究机构        | 教育机构       | 环境         |
| 282 | 为智能产品适配应用场景的设计思路要清楚人工智能场景还具备智能化、替代性发挥人的主体性、交互性以及（ ）的特点和能力。           | 集成化      | 前提化         | 标准化        | 效果化        |

|     |  |               |                       |                |                 |
|-----|--|---------------|-----------------------|----------------|-----------------|
| 283 | 需求被技术实现后，需要通过技术优化和条件改良（包括设备、环境等物质条件）实现工程化。属于人工智能应用场景的开发设计步骤中的场景（ ）。      | 发掘            | 生成                    | 实现             | 应用              |
| 284 | 智能数据平台是由数据流程和（ ）流程两大主体共同构成的解决方案。   | 数字            | 技术                    | 业务             | 管理              |
| 285 | 数据在应用层面的呈现方式（ ）。   | 取决于数据本身的价值    | 取决于业务和技术视角的不同         | 取决于数据的来源和采集方式  | 取决于数据处理和分析的技术手段 |
| 286 | 智能数据平台的一般功能包含（ ）。  | 分布式文件系统       | 分布式计算引擎               | 数据整合           | 以上都是            |
| 287 | 以下属于智能数据平台的一般功能的是（ ）。  | 分布式文件系统(HDFS) | 插件式的分布式计算资源管理平台(YARN) | 分布式计算引擎(Spark) | 以上都是            |
| 288 | 智能数据平台中（ ）是把在不同数据源的数据收集、整理、清洗，转换后（有点像ETL）加载到一个新的数据源，为消费者提供统一数据视图的数据集成方式。 | 安全管理          | 数据整合                  | 运维管理           | 分布式计算引擎         |
| 289 | 从智能数据平台技术角度出发，（ ）面向应用而生，为了保证数据的普遍适用性及拓展性，该阶段的关键是“分层、建模”。                 | 数据采集          | 数据清洗                  | 数据仓库建立         | 数据计算            |
| 290 | 以下属于智能数据平台中数据流程的是（ ）。  | 数据产生          | 数据采集                  | 数据计算           | 以上都是            |

|     |  |                              |                   |                |                   |
|-----|--|------------------------------|-------------------|----------------|-------------------|
| 291 | 数据从生产到应用的整体流程是任何一个数据从业者都要面对的问题，要展开讨论数据流程，需要解决的问题不包括（ ）。            | 数据供给                         | 数据迁移              | 数据产出           | 过程管理              |
| 292 | 智能数据平台从业务视角思考问题描述错误的是（ ）。  | 沿着技术视角开展数据流链路工作时，无需关注业务本身的情况 | 为什么需要数据团队解决       | 需要解决的问题是什么     | 通过什么方式解决          |
| 293 | 从智能数据平台业务角度出发，（ ）具体策略是对关键核心指标进行维度拆解，寻找出影响核心指标波动中不同维值的“贡献度”，最终定位问题。 | 实时监测                         | 智能分析              | 警告             | 数据清洗              |
| 294 | 智能数据平台架构主要由（ ）三个部分组成。  | 数据采集、数据处理、数据销售               | 数据采集、数据处理、数据输出与展示 | 数据采集、数据处理、数据备份 | 数据采集、数据挖掘、数据输出与展示 |
| 295 | 商业企业每天进购商品的情况、商品进价、库存管理、销售情况等。属于业务（ ）。                             | 指标                           | 绩效                | 数据             | 标准                |
| 296 | 为了得到更准确的结果，必须理解（ ）的特征、分布情况，以及需要解决的问题，以便建立和优化算法模型。                  | 数据                           | 数字                | 文本             | 图片                |
| 297 | 用户数据指标中那些会时不时地光顾下网站，并为网站带来一些价值的用户属于（ ）用户。                          | 留存                           | 活跃                | 新增             | 流失                |

|     |   |                    |                      |                    |                                       |
|-----|---|--------------------|----------------------|--------------------|---------------------------------------|
| 298 | ( ) 衡量营销推广效果的最基础指标。   | 留存                 | 活跃                   | 新增                 | 流失                                    |
| 299 | 在购物平台上, 统计某产品的用户浏览次数以及转发数。属于业务数据中的( )数据。                      | 用户                 | 行为                   | 产品                 | 流失                                    |
| 300 | 以下属于行为数据指标的是( )。  | 访问次数和访问人数          | 用户满意度                | 平均响应时间             | 转化率                                   |
| 301 | ( ) 又被称为问题树、演绎树和分解树, 它不但是界定问题与议题之间联系的纽带, 而且还能在解决问题的过程中建立一种共识。 | 二叉树                | 红黑树                  | 逻辑树                | 决策树                                   |
| 302 | 北极星指标也叫( ) 关键指标。  | 唯一                 | 规范                   | 统一                 | 准确                                    |
| 303 | 业务数据指标体系中( ) 指标是产品现阶段最关键的指标。                                  | 用户满意度              | 留存用户                 | 北极星                | 流失用户                                  |
| 304 | 业务数据指标体系的搭建流程包括( )。   | 确定北极星指标            | 拆分子指标                | 拆分过程指标             | 以上都是                                  |
| 305 | 以下北极星指标描述最为准确的是( )  | 北极星指标是所有指标中最重要的一个。 | 北极星指标是所有指标中最容易衡量的一个。 | 北极星指标是公司制定的长期发展目标。 | 北极星指标是产品当前发展阶段最关键的指标, 能够反映出产品核心价值的追求。 |
| 306 | 以下关于数据分析报告要求描述错误的是( )。  | 无需确定报告受众和分析目的      | 让图表传达更加直接            | 保障数据准确             | 框架、思路清晰                               |
| 307 | ( ) 是数据分析报告中必须包含的要素。  | 数据来源和数据采集方法        | 研究背景和分析目的            | 结论和建议              | 以上都是                                  |
| 308 | 数据分析的通用流程包括( )。   | 明确需求               | 数据获取                 | 数据整理               | 全部都是                                  |

|     |  |                  |                 |                  |                  |
|-----|--|------------------|-----------------|------------------|------------------|
| 309 | 下列哪个选项最好地解释了数据的分类？<br>( )                              | 数据可以是连续的或离散的。    | 数据可以是数字的或非数字的。  | 数据可以是可视化的或非可视化的。 | 数据可以是结构化的或非结构化的。 |
| 310 | AARRR 模型将数据分析分为了五个部分，以下拆分含义错误的是<br>( )。                | 获取用户             | 激活用户            | 提高留存             | 流失比例             |
| 311 | 某购物网站去掉了购物时反复的页面跳转，提高了用户的转化率。体现了 AARRR 用户运营分析中的 ( )。   | 获取用户             | 提高留存            | 获取收入             | 激活用户             |
| 312 | PEST 模型分析法中的“S”指的是 ( )。                                | 政治               | 经济              | 社会               | 技术               |
| 313 | PEST 模型分析法中的“E”指的是 ( )。                                | 政治               | 经济              | 社会               | 技术               |
| 314 | PEST 模型分析法中的“P”指的是 ( )。                                | 经济               | 政治              | 社会               | 技术               |
| 315 | ( ) 模型是衡量客户价值和客户潜在价值的重要工具和手段，此模型对于运营、销售、财务、市场来说都比较重要。  | AARRR            | PEST            | RFM              | SWOT             |
| 316 | RFM 分析法中的“F”指的是 ( )。                                   | 客户距离最近的一次采购时间的间隔 | 客户在限定的期间内所购买的次数 | 客户的消费能力          | 客户的自然特征          |
| 317 | ( ) 分析实际上是对企业内外部条件各方面进行综合和概括，进而分析组织的优劣势，面临的机会和威胁的一种方法。 | AARRR            | PEST            | RFM              | SWOT             |
| 318 | SWOT 模型分析法中的“W”指的是 ( )。                                | 劣势               | 优势              | 机会               | 威胁               |

|     |   |                 |                     |                                  |                |
|-----|---|-----------------|---------------------|----------------------------------|----------------|
| 319 | 5W1H 分析法中的“H”指的是（ ）。  | 人               | 目标                  | 方法                               | 原因             |
| 320 | 5W2H 分析中，（ ）是整个业务流程中所有涉及最重要部分。                                  | What            | Where               | Who                              | When           |
| 321 | 以下属于常见的数据分析工具的是（ ）。   | Power BI        | Tableau             | FineBI                           | 以上都是           |
| 322 | （ ）是软件服务、应用和连接器的集合，它们协同工作以将相关数据来源转换为连贯的视觉逼真的交互式见解。              | Excel           | Power BI            | Tableau                          | FineBI         |
| 323 | 以下属于 Tableau 主要功能的是（ ）。   | 在数分钟内完成数据连接和可视化 | 集合多个数据视图，进行更丰富的深入分析 | 多数人都可以使用直观明了地拖放产品分析数据，无需编程即可深入分析 | 以上都是           |
| 324 | 界面上的数据越容易操控，公司对自己在所在业务领域里的所作所为到底是正确还是错误，就能了解得越透彻，是（ ）数据分析工具的理念。 | Power BI        | Tableau             | Excel                            | ReportServer   |
| 325 | 以下对 Python 语言的特点描述错误的是（ ）。                                      | 简单，易学，易读        | 收费，不开源              | 丰富的库                             | 可扩展性、可扩充性，面向对象 |
| 326 | （ ）是帆软公司推出的一款商业智能产品。  | Excel           | Power BI            | Tableau                          | FineBI         |
| 327 | 在聚类分析的数据准备环节中，主要针对数据的特征进行（ ）                                    | 标准化和降维          | 筛选                  | 除掉异常数据                           | 展开             |

|     |   |               |                |                |               |
|-----|---|---------------|----------------|----------------|---------------|
| 328 | 以下哪项可能成为 K-means 的终止条件。<br>①对固定数量的迭代<br>② 在局部最小值不是特别差的情况下, 在迭代中对簇观测值的分配不发生变化 ③ 在连续迭代中质心不发生变化 ④ 当 RRS 下降到阈值以下时终止 | ①②③           | ②③④            | ①③④            | ①②③④          |
| 329 | 按照是否人工参与, 标注数据质量检验可以分为 ( )  | 人工质量检验和自动质量检验 | 人工质量检验和半自动质量检验 | 半自动质量检验和自动质量检验 | 自动质量检验和抽样质量检验 |
| 330 | Pearson 相关系数的前提条件是要两个变量满足 ( )。这就要求在计算相关系数前, 要作该项要求检验。   | 近似正态分布        | 完全正态分布         | 近似高斯分布         | 完全高斯分布        |
| 331 | 相关分析研究的是变量间的 ( )。   | 相关关系          | 函数关系           | 回归关系           | 随机关系          |
| 332 | 对于一个含 m 个对象的频繁项集来说, 可能的规则有 ( ) 个  | $m(m-1)$      | m              | m-1            | $m\%m-1$      |

## 二、多选题

| 序号 | 标题                 | 选项 A       | 选项 B         | 选项 C | 选项 D   | 选项 E         |
|----|--------------------|------------|--------------|------|--------|--------------|
| 1  | 从业人员与企业的关系是        | 利益共同体      | 相互依赖         | 对立   | 双赢     | 各取所需         |
| 2  | 符合爱岗敬业要求的是         | 聚精会神, 专心致志 | 做好自己的事, 不管其他 | 不怠慢  | 对工作有感情 | 按时上下班        |
| 3  | 在职业活动中, 要做到公正公平就必须 | 按原则办事      | 随机应变         | 不循私情 | 坚持按劳分配 | 不惧权势, 不计个人得失 |

|    |                                |  |   |   |   |   |
|----|--------------------------------|--|---|---|---|---|
| 4  | 职业道德的价值在于                      | 有利于企业利益  | 有利于企业提高产品和服务的质量   | 可以降低成本、提高劳动生产率和经济效益                                     | 有利于协调职工之间及职工与领导之间的关系  | 有利于企业树立良好形象, 创造著名品牌   |
| 5  | 爱岗敬业的具体要求是                     | 抓住择业机遇   | 增加工资待遇  | 树立职业理想  | 强化职业责任  | 提高职业技能  |
| 6  | 下列行为中符合为人民服务要求的是               | 追求合理社会效益   | 克服困难满足顾客的需要   | 提高商品的质量   | 以个人利益为中心做事  | 实现自我价值  |
| 7  | 坚持办事公道, 必须做到                   | 坚持真理   | 自我牺牲  | 舍己为人  | 光明磊落  | 个人奉献  |
| 8  | 关于诚实守信的说法, 正确的是:               | 只有在大家都诚实守信的环境下, 才需要自己诚实守信。                         | 诚实守信是市场经济法则   | 诚实守信是企业的无形资产  | 诚实守信是为人之本   | 奉行诚实守信的原则, 在市场中难以立足。  |
| 9  | 描述 Python 的特性, 正确的是:           | 简单易学   | 开源免费  | 高可移植性   | 不容易入门   | 应用范围窄   |
| 10 | Python 中使用 pip 安装第三方包, 说法正确的是: | 安装 xxx 包:<br>pip get xxx                           | 安装 xxx 包:<br>pip install xxx                            | 卸载 xxx 包:<br>pip delete xxx                             | 卸载 xxx 包:<br>pip uninstall xxx                              | 查看已安装的包:<br>pip list  |
| 11 | 关于 Python 中的变量, 说法正确的是:        | abc#ABC 是有效的变量名。                                   | _123 是有效的变量名。   | 给变量 a, b 赋值:<br>a=2, b=4                                | 给变量 a, b 赋值:<br>a, b = 2, 4                                 | a=100<br>b=100<br>id(a) 和 id(b) 是一样的。                       |
| 12 | 关于 Python 中的字符串, 说法正确的是:       | a="""<br>Hello<br>World<br>"""<br>变量 a 的类型是字符串 str | a =<br>"abc"<br>b =<br>"abc"<br>表达式<br>a is b 的结果是 True | a =<br>"AbC"<br>b =<br>"abc"<br>表达式<br>a is b 的结果是 True | a = "ABc"<br>b = "abC"<br>表达式<br>a.upper()<br>==b 的结果是 True | a = "ABc"<br>b = "abc"<br>表达式<br>a.lower()<br>==b 的结果是 True |



|    |                              |  |  |  |   |   |
|----|------------------------------|--|--|--|---|---|
| 13 | Python 中的字符串, 下列说法正确的是:      | <code>a="abcd"</code><br>表达式<br><code>a[-2]=='b'</code><br>的结果是<br>True  | <code>a="abcd"</code><br>表达式<br><code>a[2]=='b'</code><br>的结果是<br>True   | <code>a="abcd"</code><br>表达式<br><code>a[2]==a[-2]</code><br>的结果是<br>True                                       | <code>a="abcd"</code><br>表达式<br><code>a[1:2]==a[1]</code><br>是<br>True  | <code>a="abcd"</code><br>表达式<br><code>a[1:2]==a[2]</code><br>是<br>True  |
| 14 | Python 中的元组 tuple, 下列说法正确的是: | 表达式<br><code>(1,2)*2</code><br>会报错   | 表达式<br><code>(1,2)*2</code><br>的结果是<br><code>(1,2,1,2)</code>  | 表达式<br><code>(1,2)*2</code><br>的结果是<br><code>(2,4)</code>  | <code>a = (1, 2, 3)</code><br><code>a[-1] += 1</code><br>会报错  | <code>a = (1, 2, 3)</code><br><code>b = (4, 5)</code><br>表达式<br><code>a+b</code> 的结果是<br><code>(1, 2, 3, 4, 5)</code>                 |
| 15 | Python 中的列表 list, 下列说法正确的是:  | <code>a = [1, 3, 2]</code><br><code>a.sort()</code><br><code>print(a)</code><br>结果显示<br><code>[1, 2, 3]</code> | <code>a = [1, 3, 2]</code><br><code>a.sort()</code><br><code>print(a)</code><br>结果显示<br><code>[1, 3, 2]</code> | <code>a = [1, 3, 2]</code><br><code>a.sort()</code><br><code>print(a)</code><br>结果显示<br><code>[3, 2, 1]</code> | <code>a = [1, 3, 2]</code><br><code>sorted(a)</code><br><code>print(a)</code><br>结果显示<br><code>[1, 3, 2]</code> | <code>a = [1, 3, 2]</code><br><code>a = sorted(a)</code><br><code>print(a)</code><br>结果显示<br><code>[1, 2, 3]</code>                   |
| 16 | Python 中的列表 list, 下列说法正确的是:  | <code>a = [1, 3, 3, 6, 4, 2, 4, 3, 5, 7]</code><br><code>print(a.count(7)==1)</code><br>结果显示<br>True           | <code>a = [1, 3, 3, 6, 4, 2, 4, 3, 5, 7]</code><br><code>print(a.index(3)==1)</code><br>结果显示<br>True           | <code>a = [1, 3, 3, 6, 4, 2, 4, 3, 5, 7]</code><br><code>print(a.index(3,2)==1)</code><br>结果显示<br>True         | <code>a = [1, 3, 3, 6, 4, 2, 4, 3, 5, 7]</code><br><code>print(a.count(3)==3)</code><br>结果显示<br>True            | <code>a = [1, 3, 3, 6, 4, 2, 4, 3, 5, 7]</code><br><code>b = list(set(a))</code><br><code>print(a.count(3)==3)</code><br>结果显示<br>True |
| 17 | 创建 Numpy 中的一维数组, 正确的是:       | <code>np.array([1, 2, 3])</code>   | <code>np.array[1, 2, 3]</code>   | <code>np.asarray([1, 2, 3])</code>   | <code>np.array([1, 2, 3])</code>  | <code>np.arange(1, 4)</code>  |
| 18 | 创建 Numpy 中的二维数组, 正确的是:       | <code>np.array([1, 2, 3, [1, 2, 3]])</code>  | <code>np.array([[1, 2, 3], [1, 2, 3]])</code>  | <code>np.array(range(10), range(10))</code>  | <code>np.array([range(10), range(10)])</code>   | <code>np.arange(2, 10)</code>   |

|    |  |  |  |  |  |  |
|----|--|--|--|--|--|--|
| 19 | Numpy 中的二维数组, 和 Python 的嵌套二维列表, 说法正确的是:  | 可以通过 np.array() 方法把列表转换为 Numpy 的数组。比如: arr2d=np.array([[1,2],[1,2]]) | 可以通过 np.array 把列表转换为 Numpy 的数组。比如: arr2d=np.array([[1,2],[1,2]]) | 可以通过 python 的 list() 函数把 Numpy 的二维数组转换为嵌套的二维列表: a_list = list(arr2d) | 可以通过 Numpy 的二维数组对象的 tolist() 方法, 转换为嵌套的二维列表: a_list = arr2d.tolist() | 二维数组和嵌套的二维列表, 都可以使用 for 循环进行元素遍历。            |
| 20 | Numpy 中, 生成特殊数组的说法, 正确的是:  | 生成一个全 0 的二维 (3,4) 数组: np.zeros(3,4)                                  | 生成一个全 0 的二维 (3,4) 数组: np.zeros((3,4))                            | 生成一个全 1 的二维数组: np.ones((2,3))  | 生成一个对角线全 1, 其他全 0 的二维数组: np.ones((3,3))                              | 生成一个对角线全 1, 其他全 0 的二维数组: np.eye(3)           |
| 21 | Numpy 中生成两个数组:<br>arr1 = np.array([[1,2],[3,4]])<br>arr2 = np.array([[5,6],[7,8]])<br>两个数组的运算, 正确的是: | arr1+arr2 的结果是:<br>[[6 8]<br>[10 12]]                                | arr1-arr2 的结果是:<br>[[4 4]<br>[4 4]]                              | arr1*arr2 的结果是:<br>[[ 5 12]<br>[21 32]]                              | arr1*arr2 的结果是:<br>[[ 21 32]<br>[5 12]]                              | arr1.dot(arr2) 的结果是:<br>[[19 22]<br>[43 50]] |

|    |   |   |  |   |  |  |
|----|---|---|--|---|--|--|
| 22 | Numpy 中两个二维数组 arr1_2d, arr2_2d 的乘法,说法正确的是:                                    | 如果两个数组的 shape 不一样,是不能进行 arr1_2d.dot(arr2_2d) 运算的。 | 如果两个二维数组的 shape 不一样,是不能进行 arr1_2d*arr2_2d 运算的。 | arr1_2d*arr2_2d 和 arr2_2d*arr1_2d 的运算结果一样。  | arr1_2d.dot(arr2_2d)能运算的条件是 arr1_2d 的行和 arr2_2d 的列要一样多的元素。 | arr1_2d.dot(arr2_2d)和 arr2_2d.dot(arr1_2d)的运算结果是一样的。 |
| 23 | Numpy 中数组的属性 shape 表示数组的维度和每个维度上的数值。数组对象的 reshape() 方法可以改变数组的维度,下列代码片段正确的是:   | arr = np.arange(10)<br>arr.reshape(3, 4)          | arr = np.arange(12)<br>arr.reshape(3, 4)       | arr = np.arange(12)<br>arr.reshape(3, 2, 2) | arr = np.arange(12)<br>arr.reshape(3, 2, -1)               | arr = np.arange(12)<br>arr.reshape(3, -1, -1)        |
| 24 | 生成 Pandas 的 Series 的正确语句是:  | pd.Series([1, 2, 3, 4])                           | pd.Series([1, 2, 3, 4])                        | pd.Series(range(10, 20))                    | pd.Series(range(10))                                       | pd.Series(range(10))                                 |
| 25 | 已知有 Numpy 二维数组:<br>arr = np.random.randn(5, 4)<br>通过 arr 生成 DataFrame 的正确语句有: | pd.DataFrame(arr)                                 | pd.DataFrame(arr, index=range(1, 6))           | pd.DataFrame(arr, index=range(1, 7))        | pd.DataFrame(arr, columns=list("ABCDE"))                   | pd.DataFrame(arr, columns=list("ACDB"))              |

|    |   |  |   |   |   |  |
|----|---|--|---|---|---|--|
| 26 | <p>对于 pandas 的 Series 对象：<br/> <code>ser = pd.Series(np.arange(10))</code><br/>         关于索引，说法正确的是：</p>  | <code>ser.index</code><br>得到 ser 的索引对象   | <code>ser.index()</code><br>得到 ser 的索引对象  | <code>ser.index = list("ABCD")</code> 报错的原因是索引的个数和 ser 元素个数不符合。 | <code>ser.index = list("ABCDEFGHIJ")</code><br><code>print(ser['B'])</code><br>显示的结果是 2 | <code>ser.index = list("ABCDEFGHIJ")</code><br><code>print(ser.B)</code><br>显示的结果是 1 |
| 27 | <p>对于 pandas 的 DataFrame 对象：<br/> <code>df=pd.DataFrame(np.random.randn(5,4), index=range(1,6), columns=list("ACDB"))</code><br/>         关于 df 的正确说法有：</p> | <code>print(df.index[0])</code><br>结果是 0 | <code>print(df.index[-1])</code><br>结果是 6 | <code>print(df.index[1])</code><br>结果是 2                        | <code>print(df.columns[1])</code><br>结果是 'C'  | <code>print(df.columns[1])</code><br>结果 'B'。   |
| 28 | <p>对于 pandas 的 DataFrame 对象：<br/> <code>df=pd.DataFrame(np.random.randn(5,4), index=range(1,6), columns=list("ACDB"))</code><br/>         下列语句正确的是：</p>     | <code>df.loc(1)</code>                   | <code>df.loc[1]</code>                    | <code>df.loc[:2]</code>   | <code>df.loc['A']</code>  | <code>df.loc[:, ['A', 'B']]</code>   |

|    |  |   |   |   |   |   |
|----|--|---|---|---|---|---|
| 29 | <p>对于 pandas 的 DataFrame 对象：<br/>df=pd.DataFrame(np.random.randn(5,4), index=range(1,6), columns=list("ACDB"))<br/>下列正确的是：</p>   | <pre>ser = df.A&gt;0</pre> <p>A 列数据中大于 0 的数据被挑选出来保存在 Series 对象 ser 中。</p> | <pre>ser = df.A[df.A&gt;0]</pre> <p>A 列数据中大于 0 的数据被挑选出来保存在 Series 对象 ser 中。</p> | <pre>ser = df[df&gt;0]</pre> <p>df 中所有大于 0 的数被筛选出来，结果保存在一个 Series 对象 ser 中。</p> | <pre>ser=df[df&gt;0]</pre> <p>df 中所有大于 0 的数被筛选出来，结果保存在一个 Series 对象 ser 中。</p> | <pre>ser=df[df&gt;0]</pre> <p>df 中所有大于 0 的数保留，小于等于 0 的数被遮掩为 NaN，结果仍是一个 DataFrame 对象</p> |
| 30 | <p>针对 Pandas 的 DataFrame 对象 df：<br/>dic_data = {'name': ['tom', 'bob', 'dog'], 'phone': [123, 444, 789], 'year': 2019, 'xuehao': pd.Series([2, 3, 4])}<br/>df = pd.DataFrame(dic_data)<br/>正确的操作是：</p> | <p>获取列名是 name 的数据：<br/>df.name<br/>或<br/>df['name']</p>                   | <p>获取列名是 name 的数据：<br/>df.loc['name']</p>                                       | <p>获取列名是 name 和 xuehao 的两列数据：<br/>df.loc[:, 'name', 'xuehao']</p>               | <p>获取列名是 name 和 xuehao 的两列数据：<br/>df.loc[:, ['name', 'xuehao']]</p>           | <p>获取列名是 name 和 xuehao 的两列数据：<br/>df.loc[:, ['name', 'xuehao']]</p>                     |

|    |   |   |  |  |  |  |
|----|---|---|--|--|--|--|
| 31 | <p>针对 Pandas 的 DataFrame 对象 df:</p> <pre>dic_data = {'name': ['tom', 'bob', 'dog'], 'phone': [123, 444, 789], 'year': 2019, 'xuehao': pd.Series([2, 3, 4])} df = pd.DataFrame(dic_data) 正确的操作是:</pre> | <p>给 xuehao 这里重新赋值:<br/>df.xuehao = np.arange(1, 4)</p> | <p>给 xuehao 这里重新赋值:<br/>df[xuehao] = np.arange(1, 4)</p> | <p>给第一行重新赋值:<br/>df[0] = ['cat', '321', 2018, 4]</p> | <p>新加一行假设数据:<br/>df.loc[3] = ['tom', '123', 2018, 1]</p> | <p>新加一列。df 中新增一列 class, 每行初始 class1:<br/>df.class = 'class1'</p> |
| 32 | <p>数据采集又称数据获取, 是通过下面哪些手段获取的?</p>  | 数据库   | 射频识别 (RFID)  | 条码扫描器  | 传感器  | 打印机  |
| 33 | <p>国际数据管理协会 (DAMA International) 英国分会提出的数据质量的六个核心维度, 包含下面的:</p>   | 实时性   | 可靠性  | 有效性  | 及时性  | 唯一性  |

|    |   |      |        |        |         |        |
|----|---|------|--------|--------|---------|--------|
| 34 | 我国在 2020 年的国标中对数据质量做了明确的规范要求,按照数据质量分类的要素包括: | 准确性  | 完整性    | 可维护性   | 及时性     | 可靠性    |
| 35 | 数据采集规范内容通常涉及有:                              | 范围   | 时间     | 地点     | 规范性引用文件 | 采集技术要求 |
| 36 | 原始数据采集后,需要依据数据的分类分级原则,对数据进行规范化的处理,包括:       | 数据存储 | 数据整理   | 归类     | 整合      | 维护     |
| 37 | 随着数据采集技术的发展,大数据基本呈现出四个特性,即通常说的 4V 特征。4V 包括: | 体量   | 多样性    | 多变性    | 稳定性     | 结构性    |
| 38 | 工业企业数据分类,数据来源于研发数据域,数据类型通常可归类于:             | 控制信息 | 研发设计数据 | 开发测试数据 | 产品供应链数据 | 物流数据   |

|    |  |          |         |          |          |          |
|----|--|----------|---------|----------|----------|----------|
| 39 | 工业企业数据分类,数据来源于生产数据域,数据类型通常可归类于:                  | 控制信息     | 工况状态    | 开发测试数据   | 工艺参数     | 物流数据     |
| 40 | 工业企业数据分类,数据来源于运维数据域,数据类型通常可归类于:                  | 产品运行状态数据 | 客户与产品信息 | 物流数据     | 工艺参数     | 产品售后服务数据 |
| 41 | 工业企业数据分类,数据来源于管理数据域,数据类型通常可归类于:                  | 产品供应链数据  | 业务统计数据  | 开发测试数据   | 产品售后服务数据 | 客户与产品信息  |
| 42 | 工业企业数据分类,依据平台企业数据分类维度划分来源,来源于平台运营数据域,数据类型通常可归类于: | 客户数据     | 物联网采集数据 | 知识库模型库数据 | 研发数据     | 人事财务数据   |



|    |  |            |                                     |                                     |            |            |
|----|--|------------|-------------------------------------|-------------------------------------|------------|------------|
| 43 | 工业企业数据分类,依据平台企业数据分类维度划分来源,来源于企业管理数据域,数据类型通常可归类于:         | 客户数据       | 知识库<br>模型库<br>数据                    | 研发数据                                | 业务合作<br>数据 | 人事财务<br>数据 |
| 44 | 工业企业数据分类,根据企业生产阶段划分数据来源,来源于物联网信息(工业生产过程中的数据),数据类型通常可归类于: | 客户关系<br>管理 | 装备、物<br>料及产品加<br>工过程的<br>工况状<br>态参数 | 装备、物<br>料及产品加<br>工过程的<br>环境参<br>数   | 客户名单       | 供应商名<br>单  |
| 45 | 工业企业数据分类,根据企业生产阶段划分数据来源,来源于企业外部信息(产品售后数据),数据类型通常可归类于:    | 外部的互<br>联网 | 使用、运<br>营情况<br>数据                   | 装备、物<br>料及产品加<br>工过程的<br>工况状<br>态参数 | 能耗管理<br>系统 | 供应商名<br>单  |

|    |   |                               |                                |                             |                                   |   |
|----|---|-------------------------------|--------------------------------|-----------------------------|-----------------------------------|---|
| 46 | <p>根据不同类别工业数据遭篡改、破坏、泄露或非法利用后,可能对工业生产、经济效益等带来的潜在影响,将工业数据分为一级、二级、三级等 3 个级别。属于等级三的有:</p> | <p>给企业造成负面影响较小,或直接经济损失较小;</p> | <p>恢复工业数据或消除负面影响所需付出的代价较小。</p> | <p>易引发特别重大生产安全事故,突发环境事件</p> | <p>易引发特别重大生产安全事故,造成直接经济损失特别巨大</p> | <p>对国民经济、行业发展、公众利益、社会秩序乃至国家安全造成严重影响</p> |
| 47 | <p>互联网数据采集是按照某种规则,通过程序或脚本自动抓取万维网信息,采集的数据有:</p>  | <p>数控机床</p>                   | <p>文本</p>                      | <p>视频</p>                   | <p>网络流量</p>                       | <p>DTU (数据传输终端)</p>                     |
| 48 | <p>工业数据采集是基于工业设备的采集,采集的数据有:</p>   | <p>数控机床</p>                   | <p>PLC</p>                     | <p>RTU</p>                  | <p>音频</p>                         | <p>网络流量</p>                             |
| 49 | <p>互联网数据采集流程中,获得服务器响应 Response 的内容,类型可能是:</p>   | <p>HTML</p>                   | <p>机器人数据采集</p>                 | <p>Json 字符串</p>             | <p>RTU (无线远程采集终端)</p>             | <p>二进制数据 (图片或者视频)</p>                   |

|    |                              |                    |                                  |                     |  |            |
|----|------------------------------|--------------------|----------------------------------|---------------------|--|------------|
| 50 | 互联网数据采集流程中,保存数据形式多样,可能是:     | 关系型数据库             | NoSQL 数据库                        | Python              | BeautifulSoup  | 特定格式的文件    |
| 51 | 工业相机实时采集图像的流程中,选择采集模式,可以是:   | 静态采集模式             | 内触发采集模式                          | 动态采集模式              | 外触发采集模式  | 混合采集模式     |
| 52 | 工业相机实时采集图像的流程中,取流抓图方式可以是:    | 静态取流抓图             | 动态取流抓图                           | 主动取流抓图              | 回调取流抓图   | 被动取流抓图     |
| 53 | 自动驾驶汽车场景下,图像及视频数据采集比较好的优化策略: | 传感器选择优化,安装位置固定。    | 传感器选择不变,安置位置优化                   | 传感器的选择、安装位置的优化。     | 传感器校准优化。   | 传感器校准始终固定。 |
| 54 | 聚类后的样本簇可以用 ( ) 来表示           | 簇中心                | 簇大小                              | 簇密度                 | 簇描述  |            |
| 55 | 下列关于聚类说法正确的是                 | K 均值聚类初次选取聚类中心是随机的 | 密度聚类和层次聚类都不用先设定簇的数量, k 均值聚类则需要先设 | K 均值聚类计算距离一般采用曼哈顿距离 | 层次聚类是首先将每个数据点视为一个单一的簇,然后计算所有簇之间的距离来合并簇,知道所有的簇聚合成为一个簇为止 |            |

|    |                                    |              |              |            |                 |
|----|------------------------------------|--------------|--------------|------------|-----------------|
| 56 | 聚类后样本的簇密度是指簇中样本的( )                | 紧密程度         | 一个度量单位内样本的多少 | 数量         | 中心              |
| 57 | 以下属于基于划分的算法的聚类算法的是( )              | K-means 算法   | K-medoids 算法 | DBSCAN 算法  | K-prototypes 算法 |
| 58 | 基于层次的算法的聚类算法主要包括( )两种聚类算法。         | 自上而下         | 自下而上         | 有概率模型聚类算法  | 神经网络模型聚类算法      |
| 59 | K 均值聚类分析在下面哪种情况下无法得出好的结果。          | 具有异常值的数据点    | 具有不同密度的数据点   | 具有非环形的数据点  | 具有非凹形的数据点       |
| 60 | 变量之间的相关关系按其相关程度可分为( )              | 正或负相关        | 完全相关         | 不完全相关      | 不相关             |
| 61 | 常用的相关系数( )                         | Pearson 相关系数 | Spearman 系数  | Kendall 系数 | 克朗巴哈系数          |
| 62 | 如果数据不服从正态分布或近似正态分布, 此时进行相关分析时可选( ) | Pearson 相关系数 | Spearman 系数  | Kendall 系数 | 都可以             |

|    |                                  |              |             |            |              |
|----|----------------------------------|--------------|-------------|------------|--------------|
| 63 | 回归分析方法种类繁多,其分类的度量标准主要有( )。       | 自变量的个数       | 因变量的类       | 回归线的形状     | 复杂度          |
| 64 | 回归分析方法按自变量多少分类可以分为( )            | 有一元回归        | 二元回归        | 三元回归       | 多元回归         |
| 65 | 回归分析方法按因变量的类型可以分为( )             | 逻辑回归         | 非逻辑回归       | 有线性回归      | 非线性回归        |
| 66 | 多元线性回归还存在多重共线性问题,因此自变量必须选择( )的变量 | 最重要          | 最有代表性       | 最易求解       | 变化最小         |
| 67 | 对标注后的数据进行质量检验的目的是( )             | 发现数据标注中的错误   | 提高数据标注的准确率  | 纠正数据标注中的错误 | 改善数据标注质量     |
| 68 | 在数据标注任务中,由于以下原因会出现标注数据质量欠佳的情况( ) | 标注数据的准确性难以控制 | 标注数据统一性难以保证 | 标注数据的数据量太大 | 某些标注服务的技术门槛高 |

|    |                                    |                       |                  |               |               |
|----|------------------------------------|-----------------------|------------------|---------------|---------------|
| 69 | 按照是否审核所有数据,标注数据质量检验可以分为<br>( )     | 全样质量检验                | 抽样质量检验           | 人工质量检验        | 全自动或者半自动质量检验  |
| 70 | 以下属于全样检验方法的优点是<br>( )              | 能够对数据集做到无遗漏检验         | 最大程度地保证标注数据的质量   | 可以对数据集进行准确率评估 | 可以对数据集进行精确的评估 |
| 71 | 全样检验的缺点也很明显,主要体现在<br>( )           | 该方法项目执行时间长,要耗费更多的检验时间 | 无法应用在一些紧急交付的项目中  | 无法对数据集做到无遗漏检验 | 无法对数据集进行准确率评估 |
| 72 | 按照是否在标注过程中进行审核,标注数据质量检验可以分为<br>( ) | 实时质量检验                | 非实时质量检验          | 随机质量检验        | 抽样质量检验        |
| 73 | 实时检验方法的优点主要有<br>( )                | 及时发现问题并解决问题           | 有效减少标注过程中错误的重复出现 | 保证整体标注任务的流畅性  | 实时掌握数据标注任务的进度 |

|    |   |                |                |                |                   |
|----|---|----------------|----------------|----------------|-------------------|
| 74 | 如果标注员（ ）抽样检验存在不合格的数据标注,则认定该标注员的数据标注为合格,该标注员只需改正检验中发现的不合格标注即可。 | 没有             | 只有一轮           | 少于两轮           | 两轮或两轮以上           |
| 75 | 多重抽样检验方法的优点主要有（ ）   | 能够合理调配质检员的工作重心 | 有效地弥补其他检验方法的疏漏 | 提高数据标注质量检验的准确性 | 提高数据标注质量检验的效率     |
| 76 | 多轮次质量检查体系包含的轮次为（ ）  | 自检             | 互检             | 质检             | 验收                |
| 77 | 质检流程一般分为（ ）   | 项目标注质检点确认      | 项目质量总结报告       | 目质检人员培训        | 输出批次数据质量报告及质量问题解决 |
| 78 | 在对图像数据进行标注时,需要对标注后的图像数据进行质量检测,下列哪些属于质量检测项（ ）                  | 关联一致性          | 目标类别是否正确       | 目标属性是否正确       | 目标框是否贴合           |

|    |   |   |            |               |                    |
|----|---|---|------------|---------------|--------------------|
| 79 | 在图像标注中, 关键点标注内容主要有 ( )  | 关键点数量   | 标注点位置      | 标注点属性         | 标注点大小              |
| 80 | 在对图像进行区域标注时, 主要涉及到得标注内容有 ( )                                    | 区域数量  | 颜色数量       | 标注点位置         | 区域贴合度              |
| 81 | 在连续视帧标注任务中, 主要的标注内容有 ( )  | 目标追踪  | 关键帧标注      | 情感标注          | 图像理解               |
| 82 | 在一幅包含多辆车的图像中, 需要对该图像中的所有车辆进行框选标注, 如果对标注进行质量检验, 主要依据下列哪些评价指标 ( ) | 框选物体数量与实际数量是否一致<br>+H108:J10<br>8H108:M10<br>8H108:L10<br>9H10H108:K108 | 所画框与要求是否一致 | 框选目标与要求目标是否一致 | 框选的误差情况, 多框了多少无关像素 |
| 83 | 下列属于语音无效的语音数据标注错误的是 ( )   | 口齿不清或结巴   | 无法理解的方言    | 片段丢失或跳帧       | 音量波动、语音失真          |
| 84 | 对于语音数据标注项目的截取错误主要体现在 ( )。                                       | 截取音频过长  | 截取音频过短     | 截取音频不当        | 突发噪声               |



|    |   |             |           |                   |              |
|----|---|-------------|-----------|-------------------|--------------|
| 85 | 以下属于语音数据标注项目的噪声干扰错误问题的有<br>( )                | 突发噪声        | 持续噪声      | 信号干扰              | 回声干扰         |
| 86 | 语音数据标注项目的文本错误主要是语音内容错误,主要体现在<br>( )           | 少读多写        | 逻辑错误      | 语法错误              | 多读少写         |
| 87 | 语音数据标注规则规定有效的语音需要满足<br>( )                    | 必须是人发出的有效声音 | 音频中语气词要撰写 | 前后截音需要严丝合缝,切记漏音多截 | 要根据语义转写正确的词语 |
| 88 | 下列属于语音识别任务的标注内容的是<br>( )                      | 语音发出者       | 语音内容的转写   | 语音切分              | 音素标记         |
| 89 | 下列属于语音合成任务的标注内容的是<br>( )                      | 文本校对        | 文本分词      | 音素标记              | 音标标记         |
| 90 | 在语音识别任务中,对语音发出者进行相应的标注,标注后的数据质量检验评价指标为<br>( ) | 人名是否正确      | 编号是否合理    | 内容是否顺畅            | 语音韵律是否合理     |

|    |                                       |                 |            |            |              |
|----|---------------------------------------|-----------------|------------|------------|--------------|
| 91 | 在语音合成任务中,对文本进行校对标注,标注后的数据质量检验评价指标是( ) | 校对后的内容与实际内容的相似度 | 校对后的内容是否顺畅 | 校对后的逻辑是否正确 | 校对后的语音韵律是否合理 |
| 92 | 在文本情感标注任务中,下列属于文本情感类别的是( )            | 快乐              | 悲伤         | 激动         | 兴奋           |

### 三、判断题

| 序号 | 标题  | 对/错 |
|----|---|-----|
| 1  | 语音标注是指将听到的音频进行转写,并适当打上一些标签。其性质与翻译类似。      |     |
| 2  | Kendall 相关分析是基于样本数据对之间的关系来进行相关系数的强弱的分析    |     |
| 3  | 谁讲职业道德谁吃亏。                                |     |
| 4  | 加强职业道德只是经营之外的事务,往往会加大企业的负担。               |     |
| 5  | 加强职业道德能改变人们的思想观念,但与推进技术进步无关。              |     |
| 6  | 加强职业道德会增加企业投入,但从总体上看可降低企业的经营成本。           |     |
| 7  | 员工的文化素质是在上学期间习得的,与企业文化无关。                 |     |
| 8  | 职业活动内在的道德准则是忠诚、秩序、勤勉。                     |     |
| 9  | 职业道德规范——“诚信”的特征包括通识性、止损性、智慧和资质性。          |     |
| 10 | 办事公道是指从业人员在进行职业活动时,要做到助人为乐,有求必应。          |     |
| 11 | 企业文化对企业具有整合的功能。                           |     |
| 12 | 创新既不能墨守成规,也不能标新立异。                        |     |
| 13 | 勤劳节俭虽然有利于节省资源,但不能促进企业的发展。                 |     |
| 14 | if 不是 Python 语言的关键字。                      |     |
| 15 | Python 中被单引号('...')或者双引号(""...")包围的都是字符串。 |     |

|    |  |  |
|----|--|--|
| 16 | Python 中的 for 语句可以在任意序列上进行迭代访问，例如列表、字符串和元组。                  |  |
| 17 | 机器语言要通过编译才能被计算机接受。   |  |
| 18 | 编写软件使用的算法，代表数学的计算公式。   |  |
| 19 | Python 中的列表数据结构必须按顺序插入元素。                                    |  |
| 20 | 现在人们普遍使用机器语言编写计算机程序。   |  |
| 21 | Python 中空列表(list)对象的布尔值是 False 。                             |  |
| 22 | Python 中 char 不是 Python 支持的数据类型。                             |  |
| 23 | Python 中打开文件的 open() 函数只接收两个参数。                              |  |
| 24 | Python 中，math.trunc(x) 可以去掉 x 的小数部分。                         |  |
| 25 | Python 打开文件的 open() 函数中访问模式参数 'w+' 表示只读。                     |  |
| 26 | Python 中，数据类型字典的“键”必须是不可变的数据类型。                              |  |
| 27 | 为了让代码更加紧凑，编写 Python 程序时应尽量避免加入空格和空行。                         |  |
| 28 | 执行语句 from math import sin 之后，可以直接使用 sin() 函数，例如 sin(3)。      |  |
| 29 | python 中，假设 x 为列表对象，那么 x.pop() 和 x.pop(-1) 的作用是一样的。          |  |
| 30 | Python 表达是 'a'+1 的结果是 'b'                                    |  |
| 31 | Python 中，已知 x={1:1, 2:2}，那么语句 x[3]=3 无法正常执行。                 |  |
| 32 | Python 中，已知 x 是一个列表对象，那么执行语句 y=x 之后，对 y 所做的任何操作，都同样会作用到 x 上。 |  |
| 33 | Python 表达是 {1, 2}*2 的结果是 {1, 2, 1, 2}                        |  |
| 34 | Python 中，集合(set)不支持使用下标访问其中的元素。                              |  |
| 35 | Python 表达式：<br>3//5<br>的结果是 3。                               |  |
| 36 | 如果需要使用 numpy, 则必须安装 numpy, 因为 numpy 不是 python 的内置库。          |  |
| 37 | numpy 数组中的所有元素，数据类型必须一致。                                     |  |
| 38 | numpy 中的二维数组(arr)转置，可以缩写为 arr.T                              |  |

|    |  |  |
|----|--|--|
| 39 | numpy 中的二维数组的乘法, arr1.dot(arr2) 和 arr2.dot(arr1) 的结果一致。  |  |
| 40 | numpy 中, 二维数组的乘法 arr.dot(arr.T) 的结果, 一定是一个方阵。  |  |
| 41 | numpy 中的二维数组 arr, arr[arr>2] 得到的新数组中, 每个元素都大于 2。   |  |
| 42 | numpy 中的二维数组 arr, 经过 arr[arr>2], arr 的每个元素都大于 2。   |  |
| 43 | numpy 中的二维数组 arr2d, arr2d[1] 得到数组的第一行数据。   |  |
| 44 | Pandas 中的 Series 对象的索引, 只能从 0 开始。  |  |
| 45 | Pandas 中的 DataFrame 对象 df 的列索引由 list("ABCD") 产生, 则可以通过 df.B 访问第二列数据。   |  |
| 46 | Pandas 中读入 csv 文件非常简单, 比如需要读入 a.csv 文件, 只需要 pd.read_csv("a.csv") 即可得到 DataFrame 对象, q 其中包含了 a.csv 文件的行, 列内容。 |  |
| 47 | 原始数据采集的质量决定了数据预处理的难度和工作量, 影响着生产过程的效率和效果, 最终影响产品的质量。  |  |
| 48 | 数据采集是获得有效数据的重要途径, 同时也是工业大数据分析和应用的前提, 需要在大数据分析之前准备好。  |  |
| 49 | 数据质量的核心维度, 其中完整性是数据质量的最基础的一项评估标准。  |  |
| 50 | 数据质量的核心维度, 其中有效性主要体现在数据记录的规范, 不包括数据是否符合逻辑。   |  |
| 51 | 数据质量的核心维度, 其中有效性一般指范围有效性、日期有效性、形式有效性等。   |  |
| 52 | 数据质量的核心维度, 其中数量级记录错误而引起的准确性问题可以使用最大值和最小值的统计量去审核。   |  |
| 53 | 数据质量的核心维度, 其中及时性对于数据分析本身的影响并不大, 但如果数据建立的时间过长, 就无法及时进行数据分析, 可能导致分析得出的结论失去了借鉴意义。                               |  |
| 54 | 数据质量的核心维度, 其中及时性指数据记录的信息是否存在异常或错误。   |  |

|    |  |  |
|----|--|--|
| 55 | 数据质量的核心维度，其中一致性问题包括多源数据的数据模型不一致和数据记录的信息存在异常。           |  |
| 56 | 数据质量的核心维度，其中唯一性通过唯一数据条数，和不唯一数据条数的比率来衡量。                |  |
| 57 | 唯一性通过不唯一数据条数，和总数据条数的百分比来衡量。                            |  |
| 58 | GB/T38667-2020 按照数据质量分类的要素，数据的准确性指数据是否存在缺失及缺失程度。       |  |
| 59 | GB/T38667-2020 按照数据质量分类的要素，数据的一致性指数据内容是否遵循统一规范。        |  |
| 60 | GB/T38667-2020 按照数据质量分类的要素，数据的及时性指数据内容是否遵循统一规范。        |  |
| 61 | GB/T38667-2020 按照数据质量分类的要素，数据的重复性指是否存在大量重复数据。          |  |
| 62 | 大数据 4V 特征中的体量是指数据规模，大型工业企业的数据集将达到 TB 级甚至 PB 级别。        |  |
| 63 | 大数据 4V 特征中的多样性是指数据类型多样，但数据来源往往限制在具体的目标，不具备广泛性。         |  |
| 64 | 体现制造业特点的 3M 特性中的多来源特性，是指制造业数据来源广泛，数据覆盖了整个产品全生命周期各个环节。  |  |
| 65 | 数据分级原则，描述为“受影响的用户和企业数量较少、生产生活区域范围较小、持续时间较短”，数据分级为三级。   |  |
| 66 | 数据分级原则，描述为“对国民经济、行业发展、公众利益、社会秩序乃至国家安全造成严重影响。”，数据分级为三级。 |  |
| 67 | 企业针对三级数据采取的防护措施，应能抵御来自国家级敌对组织的大规模恶意攻击                  |  |
| 68 | 针对一级数据采取的防护措施，应能抵御来自国家级敌对组织的大规模恶意攻击                    |  |
| 69 | 企业针对三级数据采取的防护措施，应能抵御一般恶意攻击                             |  |
| 70 | 数据常规处理中的数据整理，是对数据进行格式和名称的规范，保障后续的归类和汇总的有效性。            |  |
| 71 | 数据常规处理中的数据整合，是对数据进行排列组合，获取有价值的信息。                      |  |

|    |   |  |
|----|---|--|
| 72 | 流行的大数据汇总工具有 Excel、BI 、python 语言、SPSS 等。             |  |
| 73 | 数据处理方法的优化包括数据采集流程的优化和数据储存流程的优化。                     |  |
| 74 | 数据采集流程框架起始于制定采集方案。                                  |  |
| 75 | 数据采集流程框架中，在搭建设备，场景之后，就着手优化采集方案。                     |  |
| 76 | 数据采集流程框架中，质量审核不过关的情况下，需要重新制定采集方案。                   |  |
| 77 | 互联网数据采集流程中，获取响应后就可以保存响应中的数据了。                       |  |
| 78 | 工业相机实时采集图像的流程中，可以直接关闭相机，无需提前关闭采集流。                  |  |
| 79 | 大数据采集过程，成千上万的用户同时进行访问和操作引起高并发数,因此需要考虑数据库的负载均衡。      |  |
| 80 | 数据采集和处理流程中，数据库结构的优化可以提高数据存储效率,但会降低据查询速度。            |  |
| 81 | 数据采集和处理流程中，数据库结构的优化可以考虑使用数据库分区，这样会增加查询量             |  |
| 82 | 图像拼接是结构化数据整合的一种方式，它是将同一场景的多个重叠图像拼接成较大的图像的一种方法。      |  |
| 83 | 在音频采集过程，拾音器的安装原则是要距离监听声源尽量近，尽量靠近主要的谈话区域，才能达到预期的拾音效果 |  |
| 84 | 聚类后样本的簇中心是指一个簇中所有样本质心点的平均和                          |  |
| 85 | 由于聚类结果是未知的，所以不同的聚类方法可能得到不同的分类结果。                    |  |
| 86 | 由于聚类结果是未知的，所以相同的聚类方法但是所分析的变量不同，也会得到不同的聚类结果。         |  |
| 87 | K-均值聚类算法计算过程中，所选取 K 个样本作为初始聚类中心是随机选取的               |  |
| 88 | k-均值算法中初始聚类中心的选取也对算法结果影响很大，不同的初始中心可能会导致不同的聚类结果。     |  |
| 89 | K-均值聚类算法计算过程中，计算每个样本与初始 聚类中心的距离中马氏距离又称为广义欧氏距离       |  |
| 90 | 聚类分析是根据在数据中发现的描述对象及其关系的信息，将数据对象分组                   |  |
| 91 | 在聚类分析过程中，组内相似性越小，组间差距越小，说明聚类效果越好                    |  |

|     |  |  |
|-----|--|--|
| 92  | 聚类的目标是得到较高的簇内相似度和较低的簇间相似度，使得簇间的距离尽可能大，簇内样本与簇中心的距离尽可能小。         |  |
| 93  | 聚类分析就是其中一种典型的探索性分析方法，它属于一种无监督机器学习。                             |  |
| 94  | 在聚类分析中，聚类指标越多，聚类结果就越客观。  |  |
| 95  | 在聚类分析中，聚类指标越多，聚类结果不一定越客观。                                      |  |
| 96  | 特征降维是从原始数据中提取有用的特征，以减少数据的维度，从而提高算法的性能。                         |  |
| 97  | 运行过两次的 K 均值聚类，可以得到相同的聚类结果                                      |  |
| 98  | 聚类分析是以基于距离来度量个案间或变量间的亲疏程度，在选择聚类变量或特征时，变量或特征的聚类指标间不能存在高度相关。     |  |
| 99  | 聚类分析是以基于距离来度量个案间或变量间的亲疏程度，在选择聚类变量或特征时，变量或特征的聚类指标间可以高度相关。       |  |
| 100 | 聚类分析过程的特征选择环节中，选择最最明显的特征进行向量化处理                                |  |
| 101 | 聚类分析过程的特征选择环节中，选择最适合有效的特征进行向量化处理                               |  |
| 102 | 聚类分析过程的特征提取环节中，主要完成对所选特征进行线性变换，构成新的特征。                         |  |
| 103 | 只有相关系数接近+1 时，才能说明两个变量之间存在高度相关关系                                |  |
| 104 | 在任何相关条件下，都可以用相关系数说明变量之间的密切程度。                                  |  |
| 105 | 完全相关即为函数关系，其相关系数为±1  |  |
| 106 | Pearson 相关系数适合于度量两个呈正态分布的连续成对变量之间的线性关系。                        |  |
| 107 | Pearson 相关系数适合于度量两个呈高斯分布的连续成对变量之间的线性关系。                        |  |
| 108 | 当变量 X 和 Y 的均值的变化不会影响两者之间的 Pearson 相关系数                         |  |
| 109 | 当变量 X 和 Y 的均值的变化不会影响两者之间的 Pearson 相关系数，说明 Pearson 相关系数具有位移不变性。 |  |
| 110 | 如果样本数据点精确的落在直线上，或者双变量分布完全在直线上，则 Pearson 相关系数等于 1 或-1           |  |

|     |  |  |
|-----|--|--|
| 111 | Pearson 相关系数是对称的，即满足 $\text{Corr}(X, Y) = \text{Corr}(Y, X)$                             |  |
| 112 | 当 Pearson 相关系数的值为 1 意味着变量 X 和变量 Y 可以很好的由直线方程来描述  |  |
| 113 | Pearson 相关系数只能用来描述两个变量之间的线性关系，。  |  |
| 114 | 当 Pearson 相关系数的值为 1 时表示两个变量方向变化相同，即一个变量增加，另一个变量也增加。                                      |  |
| 115 | Pearson 相关系数的前提条件是要两个变量满足近似正态分布。这要求在计算相关系数前，要作正态性检验。                                     |  |
| 116 | Spearman 相关系数也被称为等级变量之间的 Pearson 相关系数，即在计算过程中，原始数据被转换成等级数据，依据其在总体数据中平均的降序位置，被分配了一个相应的等级。 |  |
| 117 | 如果数据中没有重复值，并且当两个变量完全单调相关时，斯皮尔曼相关系数则为+1 或-1   |  |
| 118 | Spearman 相关系数为零表明当自变量 X 增加时，因变量 Y 没有任何趋向性。   |  |
| 119 | 当变量 X 和 Y 完全单调相关时，斯皮尔曼相关系数的绝对值为 1  |  |
| 120 | Kendall 相关系数基于数据对象的秩 (rank) 来进行两个 (随机变量) 之间的相关关系 (强弱和方向) 的评估                             |  |
| 121 | 在计算两个随机变量相关系数时，需要根据变量是否连续，呈什么分布形态等 特征来选择合适的计算公式  |  |
| 122 | 如果数据具有正态性，此时首选 Pearson 相关系数  |  |
| 123 | 相关分析能帮助我们捕捉到数据之间的关联效应。   |  |
| 124 | 在相关分析中，要求相关的两个变量都是随机变量   |  |
| 125 | 相关分析就是通过计算随机变量间的相关系数，来判断它们之间相关程度的强弱。   |  |
| 126 | 相关分析就是通过计算随机变量间的相关系数，来判断它们之间线性关系程度的强弱。   |  |
| 127 | 相关分析就是通过计算随机变量间的相关系数，来判断它们之间非线性关系程度的强弱。  |  |
| 128 | 相关分析研究的是变量间的相关关系，不是因果关系。   |  |



|     |  |  |
|-----|--|--|
| 129 | 相关分析研究的是变量间的因果关系，不是相关关系。                         |  |
| 130 | 在进行变量相关分析过程中，显著性检验环节主要是分析计算相关系数环节中的结果是否正确。       |  |
| 131 | 在进行变量相关分析过程中，业务判断环节是利用显著性检验得出的结论对实际业务相关解释或预测。    |  |
| 132 | 关联分析（Association Analysis），也叫关联规则挖掘，是属于一种无监督算法   |  |
| 133 | 关联分析（Association Analysis），也叫关联规则挖掘，是属于一种半监督算法   |  |
| 134 | 支持度体现的是某特定项集的频繁程度，表示数据集中包含某特定项的概率                |  |
| 135 | 对于一个含 $m$ 个对象的频繁项集来说，可能的规则有 $m(m-1)$ 个           |  |
| 136 | 在关联分析中，只有当某项集的支持度达到一定程度时，才有被研究的价值                |  |
| 137 | 在实际业务中，关联分析的过程即通过剪枝发现频繁项集和关联规则。                  |  |
| 138 | 在实际业务中，关联分析的过程即通过剪枝发现频繁项集。                       |  |
| 139 | Apriori 算法搜索频繁项集的过程，若发现某个项集是频繁的，那么它的所有子集也是频繁的    |  |
| 140 | Apriori 算法搜索频繁项集的过程，若发现某个项集是非频繁的，那么它的所有子集不一定是频繁的 |  |
| 141 | Apriori 算法搜索频繁项集的过程，若发现某个项集是频繁的，则其所有子集也是非频繁的     |  |
| 142 | 回归分析是一种预测性的建模技术，研究的是因变量（目标）和自变量（预测器）之间的关系        |  |
| 143 | 相关分析是回归分析的基础和前提，回归分析是相关分析的深入和延续。                 |  |
| 144 | 按自变量多少回归分析可分为一元回归和多元回归。                          |  |
| 145 | 线性回归是一种简单直接的建模技术，通过最小二乘法用一条直线（回归线）拟合因变量与自变量之间的关系 |  |
| 146 | 一元线性回归：自变量只有一个的线性回归。                             |  |
| 147 | 多元线性回归指的自变量有两个或两个以上的线性回归                         |  |
| 148 | 逻辑回归是一种广义的线性回归分析模型。                              |  |

|     |   |  |
|-----|---|--|
| 149 | 因变量为二分类的分类变量或某事件的发生率，并且是数值型变量，都可以使用逻辑回归分析                       |  |
| 150 | Lasso 回归源于解决多元线性回归中存在的多元共性问题，希望能在众多自变量中选择少量的关键变量，提高算法的稳定性和可解释性。 |  |
| 151 | Lasso 回归的优化参数可以通过梯度下降法求得。                                       |  |
| 152 | 岭回归实质上是一种改良的最小二乘估计法。  |  |
| 153 | 岭回归的参数寻优要比 Lasso 回归容易一些，可以直接通过梯度下降法迭代求解。                        |  |
| 154 | 用回归分析法进行预测首先要对各个自变量做出预测。  |  |
| 155 | 相关数据显示，数据标注质量达到 98%的时候，机器学习的训练效果为 80%，且后续提升就没有之前那么明显了。          |  |
| 156 | 质检员的作用主要是提高数据标注的质量和合格率  |  |
| 157 | 质量检验方法包括:逐条抽取检查、抽样检查、人工检查                                       |  |
| 158 | 质量检验方法逐条抽取检查是对整个标注项目中所包含的任务逐条检查并确认                              |  |
| 159 | 质量检验逐条抽取方法需要充足的人员配备，同时对于完成时间有着严格要求。                             |  |
| 160 | 为了提高检查的效率，在实际项目中往往会引入机器检查                                       |  |
| 161 | 一般来说，机器检查输出的准确率并不能完全代表数据的准确率                                    |  |
| 162 | 一般来说，机器检查输出的准确率并不能完全代表数据的准确率，机器检查后仍然需要人工最后把关，确定是否需要返工标注。        |  |
| 163 | 一般来说，机器检查输出的准确率虽然并不能完全代表数据的准确率，机器检查后可以直接输出审核报告。                 |  |
| 164 | 抽样质量检验是在不同类型项目中，尤其是大规模数据标注项目使用最广泛的检验方法                          |  |
| 165 | 根据抽样对象的类型不同，抽样检验可以再分为简单抽样、系统抽样和分层抽样                             |  |
| 166 | 根据抽样对象的类型不同，抽样检验可以再分为简单抽样、复杂抽样和系统抽样                             |  |
| 167 | 简单抽样检验是要求抽样人员客观地、随机地并且按照一定概率抽取一定数量的样本进行质量检验。                    |  |

|     |   |  |
|-----|---|--|
| 168 | 分层抽样要求先将整体标注后的数据按某种特征分为若干层级数据，然后再从每一层级进行单纯随机抽样，组成一个样本。        |  |
| 169 | 分层抽样比单纯随机抽样所得到的结果准确性更高。                                       |  |
| 170 | 将抽样检验方式叠加到实时检验，就是辅助实时检验                                       |  |
| 171 | 多重抽样检验方法辅助全样检验，是在全样检验完成后的一种补充检验方法，主要作用是减少全样检验中的疏漏，增加数据标注的准确率。 |  |
| 172 | 按照质量检验的人员组织形式，标注数据质量检验可以分为相互协作质量检验和多轮次质量检验。                   |  |
| 173 | 多轮次质量检查体系分为自检-质检-验收这三个轮次                                      |  |
| 174 | 机器检查输出的准确率并不能完全代表数据的准确率，机器检查后仍然需要人工进行质检                       |  |
| 175 | 机器学习训练图像识别是根据像素点进行的，所以对于图像标注的质量标准也是根据像素点位判定                   |  |
| 176 | 100%准确度的图像标注是要求标注像素点与标注物的边缘像素点存在 1 个像素以内的误差                   |  |
| 177 | 在图像区域标注任务中，图像标注质量每一个边缘像素点进行检验，误差在 1 个像素点以内。                   |  |
| 178 | 图像标注质量的好坏取决于像素点标注的好坏。   |  |
| 179 | 图像标注质量的好坏取决于图像质量的好坏。  |  |
| 180 | 在图像标注过程中，标注像素点越接近于标注物的中心点,标注的质量就越高，算法的训练效果也就越好。               |  |
| 181 | 在图像标注过程中，标注像素点越接近于标注物的重心点,标注的质量就越高，算法的训练效果也就越好。               |  |
| 182 | 由于视频由多帧连续图像所组成，所以用于图像的标注像素点的判断指标也是视频数据质量检验的判断指标之一             |  |
| 183 | 常见的图像数据标注类型包括关键点标注、矩形框标注、图像分割、3D 框标注、属性标注等。                   |  |
| 184 | 关键点标注模板最大的应用即是对脸部的关键点进行标注                                     |  |

|     |  |  |
|-----|--|--|
| 185 | 属性标注是用一个或多个标签标注目标物的属性。   |  |
| 186 | 图像区域标注需同时满足均匀性、连通性、封闭性的条件。   |  |
| 187 | 区域标注是指将图像分成各具特性的区域并提取出感兴趣部分的过程。  |  |
| 188 | 视频数据标注是用机器自动生成自然语言文字来描述视频内容的过程。  |  |
| 189 | 关键点标注一般用于人体脸部轮廓、五官定位、身体部位和动物头像等。                                       |  |
| 190 | 矩形框标注是一种对目标对象进行目标检测框标注的简单处理方式，常用于标注自动驾驶下的人、车、物等。                       |  |
| 191 | 在对图像标框标注进行质检时，如果标框的上下左右边缘边框均与目标标注物最边缘像素点误差在1个像素以内，则是一张合格的标框标注图片，否则不合格。 |  |
| 192 | 噪声干扰错误是指不符合语音数据标注规范的无效数据被当作有效数据来处理。                                    |  |
| 193 | 一条语音中噪音过大是属于文本错误类型   |  |
| 194 | 对于语音数据标注项目的截取错误的处理方法为重新截取。   |  |
| 195 | 对于语音数据标注项目的所有噪声干扰错误的处理方法为舍去。   |  |
| 196 | 知识整理就是要把学到的知识结构化。（ ）   |  |
| 197 | 知识整理对于个人和组织的学习和成长都非常重要。（ ）   |  |
| 198 | 图表式整理的主要优点是直观形象，提升记忆效率，稳固记忆效果。（ ）                                      |  |
| 199 | 对比式整理可以强化理解, 预防误区。（ ）  |  |
| 200 | 要点式整理容易操作，方便记忆。（ ）   |  |
| 201 | 知识整理常见方法之一是要点式整理。（ ）   |  |
| 202 | 对框题知识的及时掌握，比较合适的知识整理方法是要点式整理。（ ）                                       |  |
| 203 | 对单元、全书知识点的针对性掌握，比较合适的知识整理方法是问题式整理。（ ）                                  |  |
| 204 | 问题式整理是指，打破课本单元、课题、框题、目题的设置，将相关性强的知识点以框架的形式进行整理。（ ）                     |  |
| 205 | 关键词式整理虽然较为耗时，但能让我们快速对知识点进行全盘掌握。（ ）                                     |  |

|     |  |  |
|-----|--|--|
| 206 | 专题整理可以快速全面了解某个专题知识内容。（ ）   |  |
| 207 | 专题整理是指围绕某一专题将能够运用到该事件的相关知识进行整理。（ ）   |  |
| 208 | 思维导图可以用于知识整理。（ ）   |  |
| 209 | 在线协作文档可以进行多人协作的知识整理工作。（ ）  |  |
| 210 | 可以利用 Excel 来进行数据的录入、整理和统计等工作。（ ）   |  |
| 211 | 数据审核是根据统计研究的任务和要求，对统计调查搜集到的大量原始资料进行审核、分组、汇总，使之条理化、系统化，得出能够反映总体综合特征的统计资料的工作过程。（ ） |  |
| 212 | 数据整理是在统计调查的基础上进行的，也是积累历史资料的必要手段。（ ）  |  |
| 213 | 数据整理的第一步要先根据研究目的设计整理方案才能继续后续的整理步骤。（ ）  |  |
| 214 | 数据整理的步骤通常包括数据收集、数据清洗、数据转换、数据整合和数据存储等。（ ）   |  |
| 215 | 在无序的方式下集中信息。能够把数据分成很多组，并分析这些组的特性。使用的是数据整理中群集技术。（ ）                               |  |
| 216 | 数据整理技术可以分成群集、分类和预测三大类。（ ）  |  |
| 217 | 演绎推理是由一般到特殊的推理方法。（ ）   |  |
| 218 | 归纳推理是一种由个别到一般的推理。（ ）   |  |
| 219 | ERP 软件属于数据整理工具。（ ）   |  |
| 220 | 智能应用是指以人工智能应用为主，以大数据智能化为引领的智能化技术与管理的应用。（ ）                                       |  |
| 221 | 人工智能应用的主流方法包括机器学习、深度学习、自然语言处理等。（ ）   |  |
| 222 | 实现人工智能商业化需要同时有优秀的产品团队和技术团队，两者是缺一不可的。（ ）  |  |
| 223 | 数据源于业务又作用于业务，智能平台基于数据也服务于数据。（ ）  |  |
| 224 | 智能数据平台的一般功能包括分布式文件系统(HDFS)。（ ）   |  |
| 225 | 数据清洗的关键是“建立清洗规则”，即建立符合业务需要的数据清洗方案。（ ）  |  |

|     |   |  |
|-----|---|--|
| 226 | 在智能客服领域中问答知识包含了客户与客服之间的问答对话。（ ）                       |  |
| 227 | 智能数据平台架构主要由数据采集、数据处理、数据输出与展示三个部分组成。（ ）                |  |
| 228 | 智能数据平台架构通常包括数据采集、数据存储、数据处理、数据分析和数据可视化等组成部分。（ ）        |  |
| 229 | 智能数据平台是由数据流程和业务流程两大主体共同构成的解决方案，两大主体相辅相成、互相依赖、密不可分。（ ） |  |
| 230 | 数据处理是大数据存储与计算的核心，数据同步系统导入的数据只能存储在 MySQL 中。（ ）         |  |
| 231 | 数据一般可来源于数据库、第三方数据统计工具、专业的调研机构的统计年鉴或报告、市场调查。（ ）        |  |
| 232 | 用户新增指标是衡量营销推广效果的最基础指标。（ ）                             |  |
| 233 | 业务指标是由单个数据或多个数据运算构成的，所以在应用和设立业务指标之前无需理解数据含义。（ ）       |  |
| 234 | 对于数据指标的选择上首先要明确业务目标，确定需要达成什么样的效果，然后再根据业务目标选择相应的指标。（ ） |  |
| 235 | 数据分析报告切记不要让数据分析报告出现长篇大论的情况，应该有意识地抓住数据中出现的核心问题。（ ）     |  |
| 236 | 数据分析报告的开篇要求精简干练，既要表现分析主题，还要激发读者阅读兴趣。（ ）               |  |
| 237 | 撰写数据分析报告时要事事详细、具体、充分、全面。（ ）                           |  |
| 238 | 数据分析的通用流程中包括数据整理。（ ）                                  |  |
| 239 | AARRR 是一种营销模型，包括四个关键阶段。（ ）                            |  |
| 240 | PEST 模型分析法中的“E”指的是经济。（ ）                              |  |
| 241 | RFM 是一种常见的客户分群模型，它将客户按照近期购买行为、购买频率以及购买金额进行划分。（ ）      |  |
| 242 | RFM 分析法中的“F”指的是客户在限定的期间内所购买的次数。（ ）                    |  |
| 243 | SWOT 分析实际上是对企业内外部条件各方面进行综合和概括。（ ）                     |  |

|     |  |  |
|-----|--|--|
| 244 | SWOT 模型分析法中的“O”指的是威胁。（ ）   |  |
| 245 | 5W2H 主要用于用户行为分析、业务问题专题分析、营销活动等。（ ）                                       |  |
| 246 | 大量数据收集上来以后，是可以直接用来做分析的。（ ）   |  |
| 247 | 常见的数据分析工具包括 Excel、SPSS、Python、Tableau、Power BI 等。（ ）                     |  |
| 248 | 数据分析工具的种类比较多，一款好的数据分析工具，可以帮助我们提高学习工作的效率。（ ）                              |  |
| 249 | 用户可以通过 Power BI 将数据转化为可视化的图表、仪表盘、报表和地图等形式，帮助用户更好地理解和分析数据，从而支持决策和业务发展。（ ） |  |
| 250 | Power BI 可以创建交互式仪表板和报告，帮助用户更好地理解数据。（ ）                                   |  |
| 251 | 生成报表并进行发布，供组织在 Web 和移动设备上使用，是 Power BI 数据分析工具的功能之一。（ ）                   |  |
| 252 | Tableau 数据分析工具的理念是，界面上的数据越容易操控，公司对自己在所在业务领域里的所作所为到底是正确还是错误，就能了解得越透彻。（ ）  |  |
| 253 | Python 是由荷兰数学和计算机科学研究学会的吉多·范罗苏姆于 1990 年代初设计。（ ）                          |  |
| 254 | R 是属于 GNU 系统的一个自由、免费、源代码开放的软件，它是一个用于统计计算和统计制图的优秀工具。（ ）                   |  |
| 255 | Python 有许多开源库可以用于数据处理、数据分析、机器学习、可视化等领域。（ ）                               |  |