

۱. در فرایند کشف دانش از پایگاه داده، پس از مرحله پیش پردازش معمولاً کدام مرحله انجام می شود؟

الف) ارزیابی الگوها

ب) تبدیل داده ها

ج) استقرار دانش

د) تفسیر نتایج

پاسخ: ب

۲. در چرخه رایج CRISP-DM، پس از ساخت مدل کدام مرحله قرار دارد؟

الف) درک داده ها

ب) آماده سازی داده ها

ج) ارزیابی مدل

د) درک کسب و کار

پاسخ: ج

۳. کدام عبارت درباره نسبت داده کاوی و کشف دانش صحیح است؟

الف) داده کاوی تمام مراحل کشف دانش را دربرمی گیرد.

ب) کشف دانش تنها به ساخت مدل طبقه بندی محدود است.

ج) داده کاوی یکی از مراحل فرایند کشف دانش است.

د) کشف دانش پس از استقرار مدل پایان می یابد.

پاسخ: ج

۴. کدام فعالیت، بیش از همه با پردازش تحلیلی برخط ارتباط دارد؟

الف) کشف الگوهای ناشناخته در داده ها

ب) بررسی داده ها در ابعاد و سطوح تجمیع

ج) پیش بینی برچسب نمونه های جدید

د) تشخیص خودکار نقاط پرت

پاسخ: ب

۵. کدام مقیاس اندازه گیری فاقد ترتیب ذاتی میان مقادیر است؟

الف) اسمی

ب) رتبه ای

ج) فاصله ای

د) نسبی

پاسخ: الف

۶. در کدام مقیاس، ترتیب مقادیر مشخص است اما فاصله میان رتبه ها الزاماً برابر نیست؟

الف) اسمی

ب) رتبه‌ای

ج) فاصله‌ای

د) نسبتی

پاسخ: ب

۷. وجود صفر حقیقی و امکان مقایسه نسبت مقادیر، ویژگی کدام مقیاس است؟

الف) اسمی

ب) رتبه‌ای

ج) فاصله‌ای

د) نسبتی

پاسخ: د

۸. تعداد درخواست‌های ثبت‌شده در یک روز، از نظر ماهیت داده معمولاً کدام نوع است؟

الف) پیوسته

ب) گسسته

ج) رتبه‌ای

د) متنی

پاسخ: ب

۹. در نمایش سبد خرید، هر رکورد معمولاً بیانگر کدام مورد است؟

الف) ویژگی عددی یک کالا

ب) مجموعه اقلام یک تراکنش

ج) برچسب یک طبقه از کالاها

د) ترتیب زمانی مراحل خرید

پاسخ: ب

۱۰. کدام روش، در حالت معمول، رویکردی توصیفی محسوب می‌شود؟

الف) پیش‌بینی میزان فروش

ب) برآورد احتمال نکول

ج) خوشه‌بندی مشتریان

د) طبقه‌بندی درخواست‌ها

پاسخ: ج

۱۱. داده‌های گمشده از نوع کاملاً تصادفی، چه ویژگی‌ای دارند؟

الف) گمشدگی فقط به مقدار همان متغیر وابسته است.

- (ب) گمشدگی به مقدار متغیرهای مشاهده شده وابسته است.
(ج) گمشدگی به مقادیر مشاهده شده و مشاهده نشده وابسته است.
(د) احتمال گمشدگی مستقل از مقادیر داده هاست.
پاسخ: د

۱۲. جایگزینی مقادیر گمشده با میانگین، معمولاً چه اثری دارد؟
(الف) افزایش پراکندگی مقادیر متغیر
(ب) کاهش پراکندگی مقادیر متغیر
(ج) حذف رابطه میان متغیرها
(د) افزایش تعداد سطوح متغیر
پاسخ: ب

۱۳. برای جایگزینی مقدار گمشده در یک متغیر عددی دارای چولگی و نقاط پرت، کدام شاخص معمولاً مقاوم تر است؟
(الف) میانگین حسابی
(ب) میانه
(ج) واریانس
(د) دامنه تغییرات
پاسخ: ب

۱۴. در مقیاس بندی کمینه - بیشینه، تبدیل استاندارد معمولاً مقادیر را به کدام بازه می برد؟
(الف) صفر تا یک
(ب) منفی یک تا صفر
(ج) منفی یک تا یک
(د) منفی بی نهایت تا مثبت بی نهایت
پاسخ: الف

۱۵. در استاندارد سازی با نمره استاندارد، میانگین و انحراف معیار داده های تبدیل شده معمولاً به ترتیب کدام اند؟
(الف) صفر و یک
(ب) یک و صفر
(ج) صفر و صفر
(د) یک و یک
پاسخ: الف

۱۶. برای مقیاس بندی مقاوم در برابر نقاط پرت، کدام ترکیب معمولاً به کار می رود؟
(الف) میانگین و دامنه
(ب) میانه و فاصله میان چارکی
(ج) نما و واریانس