

بسمه تعالی

امتحان میان ترم درس داده کاوی پیشرفته

مدت زمان جهت پاسخگویی: ۹۰ دقیقه

شنبه ۱۴۰۱/۰۲/۳۱

۱- به سوالات ذیل پاسخ دهید.

- ۱-۱) مفهوم داده کاوی و نفرین ابعاد را بیان کنید. (۵, ۰ نمره)
- ۱-۲) تفاوت بین برچسب‌های دودویی متقارن و نامتقارن را بیان کنید. (۵, ۰ نمره)
- ۱-۳) انواع برچسب‌های عددی را نام ببرید و توضیح دهید. (۷۵, ۰ نمره)
- ۱-۴) مراحل داده کاوی را نام ببرید. (۵, ۰ نمره)
- ۱-۵) سه مدل Datawarehouse را بیان کنید. (۷۵, ۰ نمره)
- ۱-۶) داده‌های پرت چه داده‌هایی هستند؟ دو روش برای آشکار کردن آنها نام ببرید. (۷۵, ۰ نمره)
- ۱-۷) دو روش کاهش بعد را نام ببرید. (۵, ۰ نمره)
- ۱-۸) سه عمل در OLAP را نام ببرید و توضیح دهید. (۷۵, ۰ نمره)

۲- نمرات درس مربوط به دانشجویان یک دانشگاه مطابق با جدول ذیل است.

الف) میانه تخمینی نمرات را بدست آورید. (۳ نمره)

ب) IQR تخمینی را بدست آورید. (۳ نمره)

ردیف	نمره	فراوانی
۱	۱ تا ۹	۵
۲	۱۰ تا ۲۵	۸
۳	۲۶ تا ۵۰	۱۲
۴	۵۱ تا ۶۵	۱۵
۵	۶۶ تا ۷۵	۳۰
۶	۷۶ تا ۸۰	۲۵
۷	۸۱ تا ۹۰	۱۰
۸	۹۱ تا ۱۰۰	۱۵

۳- مجموعه A به صورت ذیل تعریف شده است.

$$A = \{12, 38, 9, 3, 18, 15, 64, 69, 55, 5, 44, 48, 36, 24, 30\}$$

الف) داده‌های مجموعه A را با روش Min-Max نرمالیزه کنید. (۲ نمره)

ب) داده‌های مجموعه A را با روش equal-depth به سه بخش تقسیم کنید و سپس آن را با روش bin-means یکنواخت کنید. (۲ نمره)

۴- مجموعه داده‌های زیر را با فرض $\text{min-sup}=65\%$ و $\text{min-conf}=75\%$ در نظر بگیرید.

TID	Items
T100	A, C, F, D
T200	F, D, B, E
T300	E, C, A, F
T400	A, C, D, E
T500	A, F, C, B, D
T600	C, E, F, A, D

الف) با استفاده از الگوریتم Apriori داده‌های مکرر (Frequent itemsets) را بدست آورید. (۳ نمره)

ب) قوانین انجمنی قوی که منطبق بر عبارت زیر باشد را تعیین کنید. (۲ نمره)

$$item1 \wedge item2 \Rightarrow item3$$