

## بسمه تعالی

مهلت تحویل: ۱۴۰۲/۰۹/۰۳

دانشگاه صنعتی شاهرود

تمرین سری دوم داده کاوی

### نکات

- ۱- پاسخ تمرینات به صورت تایپ شده در قالب فایل pdf صرفاً در سامانه LMS قرار گیرد.
- ۲- در پاسخ به تمرینات روی دیتاست خودتان، تحلیل شما از نتایج بسیار مهم است. دقت کنید که باید نتایج را در قالب شرح و اسکرین شات از برنامه عملی یا درج کد نوشته شده و خروجی برنامه ذکر کرده، و سپس تحلیل کنید. (دیتاست انتخابی تا انتهای ترم یکسان باقی می ماند)
- ۳- انجام تمرینات تئوری به صورت انفرادی و بخش عملی در قالب گروه‌های دو نفره صورت گیرد.
- ۴- تمرین عملی در زمانی که بعداً اعلام خواهد شد، توسط حل تمرین به صورت حضوری نیز تحویل گرفته می شود.

### بخش تئوری

- ۱- الگوهای مکرر را با ذکر یک مثال کاربردی تعریف کنید. (مثال تان می تواند از دیتاستتان باشد)
- ۲- فروشندگانی می‌خواهد کالاهایش را با توجه به قیمتشان درون ۴ قفسه جداگانه قرار دهد. اگر قیمت کالاها به ترتیب زمان دریافتشان توسط فروشنده بصورت زیر باشد ابتدا یکبار میانگین، میانه و مد را برای قیمت کالاهای قفسه‌بندی نشده و سپس پس از قراردعی کالاها در قفسه‌های مناسب محاسبه کنید.

قیمت کالا = 

4	12	11	18	22	4	3	12	24	24	4	8
---	----	----	----	----	---	---	----	----	----	---	---

- ۳- با توجه به جدول زیر محاسبه کنید که آیا مهارت آشنایی با پایتون و آشنایی با زبان C دو ویژگی مرتبط هستند یا خیر؟ (ابتدا مقادیر مورد انتظار را می‌بایست بدست آورید)

	Knows Python	Doesn't Know Python	Sum(row)
Knows C	70	150	220
Doesn't Know C	500	280	780
Sum(Col)	570	430	1000

۴ - جدول زیر را در نظر بگیرید.

ID	ITEMS
1	A, D
2	A, B, C, D
3	B, C
4	A, B, D
5	A, B, C, D, E

Min Support = 3

الف) الگوهای بسته (Closed-Pattern) و بیشینه (Max-Pattern) را از جدول فوق پیدا کنید.

ب) الگوریتم Apriori را پیاده‌سازی کنید.

ج) الگوریتم fp-growth را پیاده‌سازی و سپس الگوهای مکرر را پیدا کنید.

د) نتایج الگوریتم Apriori و fp-growth را باهم مقایسه کرده و برداشت خود را گزارش کنید.

۵ - پنج کنفرانس و پنج ژورنال مرتبط با داده کاوی را نام ببرید. برای هر ژورنال و کنفرانس ناشر مقالات را مشخص نمایید. از هر کنفرانس نام برده شده یک عنوان مقاله را ذکر کنید (ذکر عنوان کافی است و نیازی به توضیح نیست). سپس مشخص کنید این مقالات چند ارجاع (citation) داشته اند؟

### بخش عملی

۱- در نرم‌افزار وکا، یک فیلد پیوسته دیتاست خود را به دلخواه به nominal تغییر داده و درون ۳ سطل قرار دهید.

۲- در صورت وجود، داده‌های outlier را از دیتاست خود پیدا کرده و سپس حذف کنید. در صورت نبود outlier، مراحل پیدا کردن و حذف را فقط توضیح دهید.

۳ - الگوهای مکرر و قوانین انجمنی دیتاست خود را استخراج کرده و تحلیل خود را از قوانین نمایش داده شده ارائه دهید. مقدار حداقل پشتیبانی را افزایش دهید و سپس نتیجه را گزارش کنید. با کاهش حداقل اطمینان قوانین انجمنی دیتاست شما چه تغییری خواهند کرد؟