

## داده کاوی - نیمسال اول ۱۴۰۱-۱۴۰۰

پروژه پایانی - موعده تحویل: ۱۴۰۰/۱۰/۱۰

**نحوه ارسال:** صرفاً از طریق صفحه درس در سامانه درس افزار شریف. از ارسال پروژه این درس با استفاده از ایمیل، جداً خودداری نمایید.

**نحوه انجام:** انجام پروژه پایانی درس داده کاوی پیشرفته، بایستی به صورت جداگانه توسط هر یک از دانشجویان درس، صورت گیرد.

**موارد ارسالی:** دانشجویان بایستی موارد زیر را به عنوان خروجی پروژه نهایی درس، ارسال نمایند:

- گزارش شرح انجام پروژه و نتایج حاصل از اجرای گام‌های مندرج در شرح پروژه
- کد برنامه پیاده سازی شده به انضمام فایل توضیحات بخش‌های کد، نحوه عملکرد هر بخش و نیز ملاحظات ضروری جهت اجرای کد

**شرح پروژه:** در این پروژه در نظر است مجموعه داده منتشر شده توسط شرکت دیجی کالا از منظر تحلیل سفارش‌های ثبت شده مشتریان، مورد بررسی قرار گیرد. برای این کار، ابتدا این مجموعه داده را از [اینجا](#)، دانلود نمایید و از فایل orders.csv برای این منظور، استفاده کنید. در این فایل، اطلاعات مربوط به سوابق خریدهای مشتریان، درج شده است. نخست، ستون داده با عنوان DateTime\_CartFinalize را انتخاب کنید و با انجام راست کلیک و انتخاب گزینه Format Cells...، در پنجره ظاهر شده از بخش Category، مقدار Time را انتخاب نمایید و دکمه OK را بفشارید و فایل نتیجه را ذخیره کنید. با استفاده از ستون‌های داده با عنوان ID\_Customer و ID\_Item می‌توانید به شناسه مشتریان و شناسه کالاهای خریداری شده توسط آنها، دست بیابید و با توجه به ستون داده DateTime\_CartFinalize، اقلام خریداری شده توسط مشتریان در هر نوبت خرید را مشخص نمایید. حالا برنامه‌ای بنویسید که با استفاده از دو الگوریتم شناسایی الگوهای مکرر معرفی شده در درس، یعنی:

• الگوریتم Apriori، و

• الگوریتم FPGrowth

مجموعه الگوهای مکرر را در سفارش‌های مشتریان بدست آورد و شاخص‌های Support و Confidence را به ازای الگوهای کشف شده، محاسبه نماید. سپس بر مبنای شاخص‌های سنجش میزان اهمیت و جذابیت الگوهای کشف شده نظیر:

• شاخص Lift

- شاخص  $\chi^2$

- شاخص Imbalance Ratio

از بین الگوهای مکرر شناسایی شده، الگوهای حائز اهمیت بیشتر، تعیین شوند.

نهایتاً در بخش پایانی پروژه، تحلیل‌های فوق را با لحاظ نمودن اطلاعات محل جغرافیایی سکونت مشتریان (ستون city\_name\_fa) و دسته‌بندی الگوهای مهم کشف شده در بخش قبلی بر مبنای محل شهر و استان محل سکونت مشتریان، تکرار نمایید.

تندرست و پیروز باشید.

کیهانی پور