



## تحلیل سبد خرید و recommender بر اساس ARM

**Deadline 1404.09.30**

### شرح مسئله

در این تمرین شما باید تحلیل الگوهای مشتریان در فروشگاه instacart را انجام داده و قوانین Association Rule Mining را استخراج کنید.

دیتاست را از این [لینک](#) می‌توانید دانلود کنید.

### Task 1 - آماده سازی دیتاست

مانند تمرین‌های قبل پیش‌پردازش داده‌ها را انجام دهید:

- حذف مقادیر null
- حذف quality منفی (کالای بازگشتی)
- حذف قیمت‌های صفر یا منفی
- حذف Invoice هایی که فقط یک قلم کالا دارند (برای تحلیل سبد لازم است حداقل دو کالا وجود داشته باشد)
- و هر preprocess که تشخیص می‌دهید باید انجام شود.

در این تمرین به بخشی از این دیتاست احتیاج دارید. (نیازی به استفاده از کل دیتاها نیست)، به طور مثال:

- یا سفارش‌های 20,000 کاربر را انتخاب کنید.
- یا سفارش‌هایی از یک ماه خاص یا یک هفته‌ی خاص را انتخاب کنید (تعداد آن قابل قبول باشد، کمتر از 10,000 نمونه نباشد).

زیرا کار را برای استخراج قوانین ARM راحت‌تر می‌کند.

## Task 2 - ساخت سبد خرید

در این مرحله شما باید با استفاده از step های زیر سبد خرید کاربران را ایجاد کنید:

1. داده‌ها را بر اساس InvoiceNo گروه بندی کنید.
2. برای هر Invoice یک لیست از کالاها بسازید.
3. یک جدول One-Hot Encoding برای آیتم‌ها بسازید (سبد خرید دودویی).

## Task 3 - اجرای الگوریتم Apriori

- از `mlxtend.frequent_patterns.apriori` استفاده کنید. (استفاده از توابع کتابخانه‌ی `mlxtend.frequent_patterns`)
- حداقل حمایت (`min_support`) را 0.01 امتحان کنید.
- سپس `min_support = 0.05` امتحان کنید.
- تعداد و اندازه‌ی `frequent itemsets` را مقایسه کنید.
- به این دو سوال پاسخ دهید:
  - در نهایت مقدار `min_support` باید چقدر باشد؟
  - چرا با `mini_support` بزرگتر تعداد `item_set` ها کمتر می‌شود؟

## Task 4 - استخراج قوانین ارتباطی (Association Rules)

1. از تابع `association_rules` استفاده کنید. (استفاده از توابع کتابخانه‌ی `mlxtend.frequent_patterns`)
2. معیارهای زیر را تحلیل کنید:
  - a. Support
  - b. Confidence
  - c. Lift
3. سه قانون با بیشترین `lift` را پیدا کنید.

## Task 5 - تحلیل تجاری (Business Interpretation)

فروشگاه اینترنتی چگونه می‌تواند از این قوانین در عمل استفاده کند؟

---

۱- متن گزارش شما باید شامل موارد زیر باشد:

- ❖ توضیح کامل مراحل
- ❖ نمودار اقلام پرتکرار
- ❖ جدول Frequent Itemsets
- ❖ جدول Association Rules
- ❖ تحلیل کسب و کار

۲- پاسخ به سوالات تشریحی در متن گزارش:

- ❖ نقش lift چیست؟
- ❖ چرا الگوریتم Apriori برای فروشگاه‌های اینترنتی مهم است؟

۳- و Jupyter Notebook (یا هر code editor یا نوت‌بوکی که راحت‌تر هستید) که شامل:

- ❖ کدهای کامل پیش‌پردازش
- ❖ اجرای Apriori
- ❖ تولید قوانین
- ❖ یافتن بهترین قوانین

در نهایت فایل‌های نهایی را به صورت یک فایل zip فشرده و به صورت name\_familyname\_stdID نام گذاری کرده و به آدرس [quera](mailto:quera) ارسال کنید.

در صورت تمایل به ارسال ایمیل، لطفاً subject را "e-commerce hw3" قرار دهید. (مهلت ارسال تمرین ۳۰ آذر ماه می‌باشد)

[yadollahisamin4@gmail.com](mailto:yadollahisamin4@gmail.com)

(در صورتی که کد را بر روی گیت‌هاب قرار داده‌اید؛ کافی ست آدرس پروژه را در فایل گزارش نوشته و فقط فایل گزارش را ارسال کنید.)

\* نیازی به ارسال دیتاست نمی‌باشد.

\* نیازی به بارگذاری دیتاست در ریپو نمی‌باشد.

موفق باشید