

# ۱. الگوریتم های استخراج Frequent Itemset

A. با توجه به الگوریتم FP-Growth ابتدا FP-tree را برای مجموعه تراکنش زیر رسم کرده و سپس با استفاده از آن،

جدول Conditional Pattern Base را ایجاد و درنهایت از روی جدول فوق Frequent Itemset ها را استخراج

کنید. (min\_support = ۱)

TID	ITEMS
T۱۰۰	۱۱,۱۲
T۲۰۰	۱۲,۱۴
T۳۰۰	۱۲,۱۳
T۴۰۰	۱۱,۱۲,۱۴
T۵۰۰	۱۱,۱۳
T۶۰۰	۱۲,۱۳
T۷۰۰	۱۱,۱۳
T۸۰۰	۱۱,۱۲

B. این بار با استفاده از الگوریتم ECLAT مجموعه آیتم های پرتکرار استخراج کنید. (min\_support=۳)

C. این دو روش چه مزایا و معایبی نسبت به یکدیگر دارند؟

## تحلیل و بررسی جذابیت قوانین انجمنی و همبستگی بین آیتم های تراکنش

جدول احتمالات (contingency) زیر خلاصه ای از داده های تراکنش های یک رستوران است:

	<i>hot dogs</i>	$\overline{hot\ dogs}$	$\Sigma_{row}$
<i>hamburgers</i>	2000	500	2500
$\overline{hamburgers}$	1000	1500	2500
$\Sigma_{col}$	3000	2000	5000

A. فرض کنید قانون انجمنی "hot dogs  $\Rightarrow$  hamburgers" استخراج شده باشد. بررسی کنید که آیا این قانون قوی

است؟ (min\_support = ۲۵٪, min\_confidence = ۵۰٪)

B. بر اساس جدول داده شده یکبار با استفاده از معیار lift و بار دیگر با استفاده از معیار  $\chi^2$  نوع

همبستگی (correlation relationship) بین این دو آیتم را بررسی کنید. (انواع همبستگی: مثبت-مستقل-منفی)