

«به نام خدا»

تکلیف ششم - سوال چهارم - مرضیه علیدادی - 9631983
(کد های مربوط، در دو فرمت py و ipynb. ضمیمه شده اند.)

4.

(b)

:

	item_set
0	Lassi Coffee Powder Butter Yougurt Ghee Cheese
1	Ghee Coffee Powder
2	Lassi Tea Powder Butter Cheese
3	Cheese Tea Powder Panner Coffee Powder Butter ...
4	Cheese Yougurt Coffee Powder Sugar Butter Sweet
...	...
12521	Panner Sugar Bread Milk Cheese
12522	Sugar Bread Coffee Powder Cheese
12523	Bread Cheese Yougurt Milk
12524	Ghee Bread Yougurt Sugar Cheese
12525	Panner Yougurt Bread

12526 rows × 1 columns

(c) هر سطر از خروجی شامل دو بخش است که در بخش اول، item set های پرتکرار را نشان داده است. و در بخش دوم، عدد support نظیر هر یک از آن item set های پرتکرار را نمایش می دهد. support، سهم item set نظیر را نسبت به تعداد کل item set ها نمایش میدهد

(d) الگوریتم Apriori، item set های کاندید را تولید می کند و چک می کند که آیا پرتکرار هستند یا خیر. ولی الگوریتم fp_growth، از تکنیک رشد pattern fragment ها استفاده می کند تا item set های پرتکرار را استخراج کند. یعنی در واقع از پترن های کوچک تر به سمت بزرگ کردن آن ها که پرتکرار هستند، حرکت می کند. و برای ذخیره اطلاعات مهم و فشرده شده در مورد پترن های پرتکرار، از یک ساختار درخت ایندکس گذاری شده و extend شده استفاده میکند.