

1) • با کمک کتابخانه های `matplotlib`, `seaborn`, `pandas` انجام شد.

• این دیتاست هیچ گونه `Missing Value` ندارد.

• اگر مقدار ویژگی برابر یا بزرگتر از میانه باشد، عدد 1 و در غیر این صورت عدد 0 را به آن نسبت می دهیم. این روش باعث میشود که تمام داده ها باینری شوند و مناسب برای الگوریتم های کشف الگوهای پرتکرار باشند.

2) به طور کلی الگوریتم `FP-Growth` سریع تر اجرا میشود و نیازی به تولید کاندیدا ها ندارد اما ممکن است `Apriori` برای داده های کوچک مناسب باشد ولی در داده هایی با حجم زیاد `FP-Growth` گزینه ی بهتری است.

3) برای استخراج قوانین انجمنی با استفاده از آیتمستهای پرتکرار از مجموعه داده ها، میتوان از الگوریتم هایی مانند `Apriori` یا `FP-Growth` استفاده کرد. این روشها به یافتن ترکیبهای پرتکرار ویژگیها کمک میکنند تا روابط مهم بین آنها را کشف کنیم.

• گام های کلی آنها عبارت است از:

• پیش پردازش داده ها

• استخراج آیتم ست های پرتکرار

• تولید قوانین انجمنی

• تحلیل و تفسیر نتایج

• برای ارزیابی قوانین انجمنی استخراج شده از مجموعه داده ها، سه معیار `Support`، `Confidence` و `Lift` محاسبه میشوند. این معیارها به ما کمک میکنند تا میزان تأثیرگذاری و اهمیت قوانین استخراج شده را تحلیل کنیم.

4) در این فرآیند، داده های مربوط به بیماران دیابتی بررسی شدند تا روابط پنهان بین ویژگی ها کشف شوند. این تحلیل شامل پیشپردازش داده ها، استخراج الگوهای پرتکرار و شناسایی قوانین انجمنی بود.

• با بررسی داده های آپلود شده، میتوان الگوهای پرتکرار را شناسایی و قوانین انجمنی را استخراج کرد. داده ها شامل ویژگی هایی مانند سطح گلوکز، فشار خون، `BMI`، تعداد بارداری ها، سطح انسولین و نتیجه ابتلا به دیابت (`Outcome`) هستند.

■ این فایل خلاصه ای از گزارش تمرین دوم بود و شما میتوانید بقیه گزارش را در قالب کد در فایل فشرده مشاهده نمایید.