Bài tập thêm \_ Câu 1

Trong các phương pháp tìm kiếm luật kết hợp trên ta sử dụng hai giá trị min\_sup và min\_conf để đánh giá các luật tìm được. Tuy nhiên trong thực tế, nếu chỉ sử dụng hai giá trị này thì mô hình vẫn có thể sinh ra một số luật phi lí. Vì thế để giới hạn vấn đề này ta có thể bổ sung thêm một giá trị để đánh giá luật kết hợp đó là tính tương quan giữa hai vế của luật.

Sinh viên tìm hiểu hai phương pháp phân tích tính tương quan giữa hai vế của luật sử dụng giá trị **Lift** và 𝝌𝟐 và sử dụng để đánh giá các luật tìm được ở phần thực hành.

* Lift:
* Là một hệ số đo lường cho sự xuất hiện của A và B là những sự kiện Độc lậphay là liên quan đến nhau**.** Sự xuất hiện của A được gọi là Độc lập nếu như sự xuất hiện của vật B là: P(A ∪ B) = P(A)P(B).
* Định nghĩa này cũng có thể được dùng trong 2 thuộc tính với nhau.

lift(A, B) = P(A ∪B) / P(A)P(B)

* Nếu mức lift = 1 thì cho biết sự xuất hiện của A và B là độc lập với nhau
* Nếu mức lift > 1, điều đó cho biết sự xuất hiện của A thường sẽ dẫn đến B xuất hiện. Các quy tắc đó có khả năng hữu ích để dự đoán trong các tập dữ liệu trong tương lai.
* Nếu mức lift < 1, điều đó cho biết sự xuất hiện của A thường sẽ dẫn đến sự không xuất hiện của B và ngược lại.
* X2:
* Là chi-square, cho ta biết về mối tương quan giữa 2 thuộc tính, ta cũng có thể sử dụng.
* Chi-square để đánh giá các luật tìm ra được, nhờ đó loại bỏ một số luật phi lý. Thường kết luận của Chi-square sẽ giống với Lift nên ta có thể sử dụng 2 cách này để loại bỏ các luật phi lý.