

Họ tên: Nguyễn Mạnh Đức

MSSV: 20521196

Môn học: Thu thập và tiền xử lý dữ liệu

Lớp: DS103.M21

BÀI LÀM

1. Tổng quan về bộ dữ liệu:

Số lượng dữ liệu: 206

Số nhãn: 206

2. Thống kê số lượng nhãn theo từng annotator

		Người 1		
		0	1	2
Người 2	0	107	8	6
	1	5	8	23
	2	10	7	32

3. Tính toán độ đồng thuận. Ghi rõ số liệu tính toán

Xác suất đồng thuận của 2 người:

$$\Pr(a) = (107+8+32)/206 = 147/206 \sim 0.7136$$

Xác suất giả định đồng thuận của 2 annotator trên 3 nhãn:

$$\begin{aligned}\Pr(0) &= \Pr(\text{Người 1}|0) \times \Pr(\text{Người 2}|0) \\ &= (122/206) \times (121/206) = 0.3479\end{aligned}$$

$$\begin{aligned}\Pr(1) &= \Pr(\text{Người 1}|1) \times \Pr(\text{Người 2}|1) \\ &= (23/206) \times (36/206) = 0.0195\end{aligned}$$

$$\begin{aligned}\Pr(2) &= \Pr(\text{Người 1}|2) \times \Pr(\text{Người 2}|2) \\ &= (61/206) \times (49/206) = 0.0704\end{aligned}$$

$$\Rightarrow \Pr(e) = \Pr(0) + \Pr(1) + \Pr(2) = 0.3479+0.0195+0.0704 = 0.4378$$

Độ đồng thuận giữa 2 người:

$$k = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)} = \frac{0.7136 - 0.4378}{1 - 0.4376} \sim 0.49$$

4. Kết luận về độ đồng thuận theo thang đo của Landis và Koch (1977)

$k = 0.49$ nằm trong khoảng $0.41 - 0.60$

\Rightarrow Độ đồng thuận thuộc: moderate