**1. Output:**  
Đối với mỗi sample, output có dạng là vector 6 chiều ứng với 6 khía cạnh, mỗi chiều nhận giá trị nguyên thuộc đoạn [0, 5] tương ứng với 0: không có khía cạnh và 1-5 ứng với số sao đánh giá nếu khía cạnh đó xuất hiện. Như vậy, output bài toán ứng với n samples:



Trong đó:

* a*j*: giá trị review của khía cạnh j.
* (0: khía cạnh j không xuất hiện, 1-5: mức độ biểu cảm).
* n: số samples.
* y*i*: output của sample thứ i.

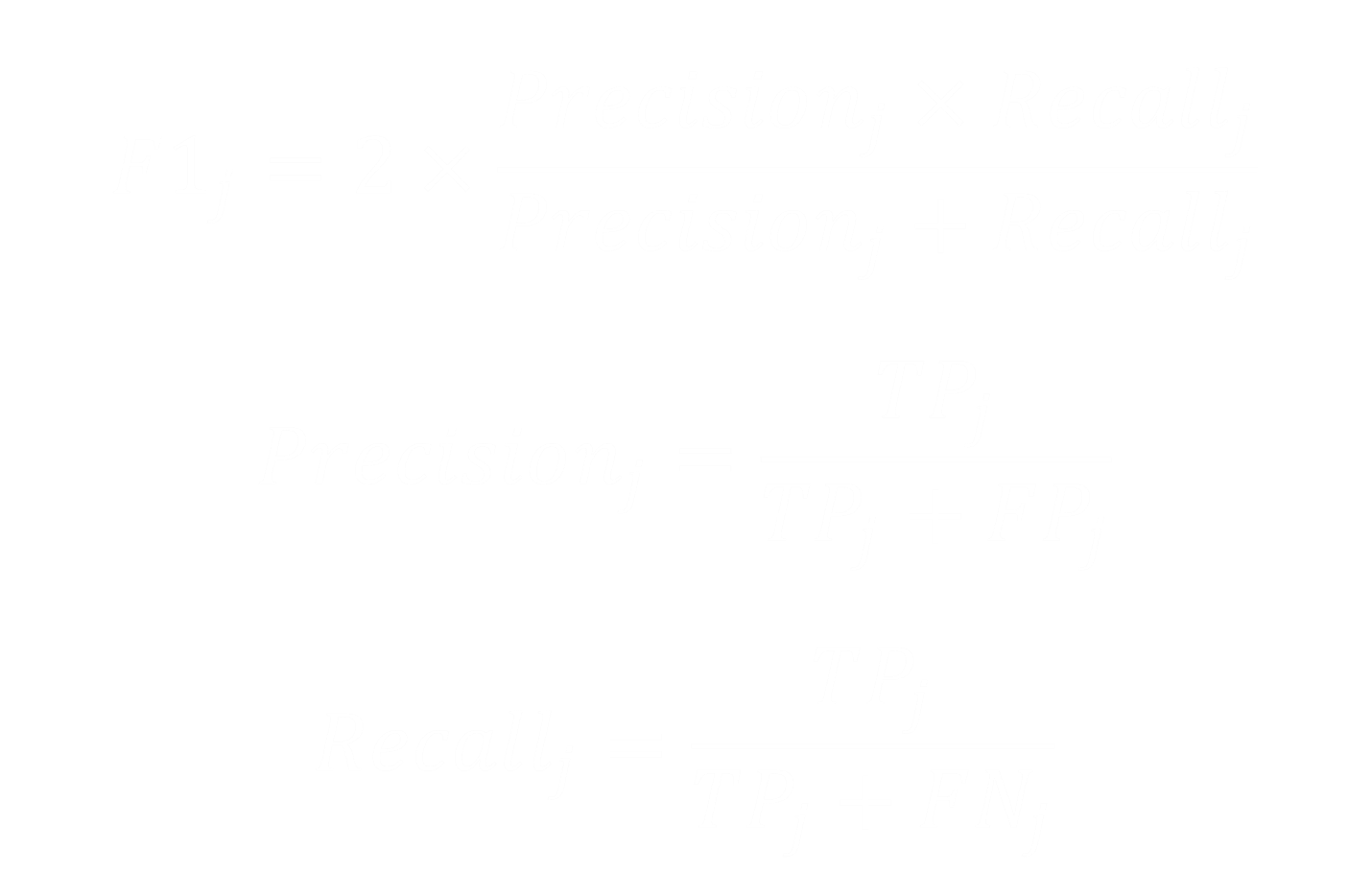
fau

Quy ước: khía cạnh 0-5 lần lượt là:

* a*0*: Dịch vụ vui chơi giải trí.
* a*1*: Dịch vụ lưu trú.
* a*2*: Hệ thống nhà hàng.
* a*3*: Dịch vụ ăn uống.
* a*4*: Dịch vụ di chuyển.
* a*5*: Dịch vụ mua sắm.

**2. Metrics for Aspect Classification:**

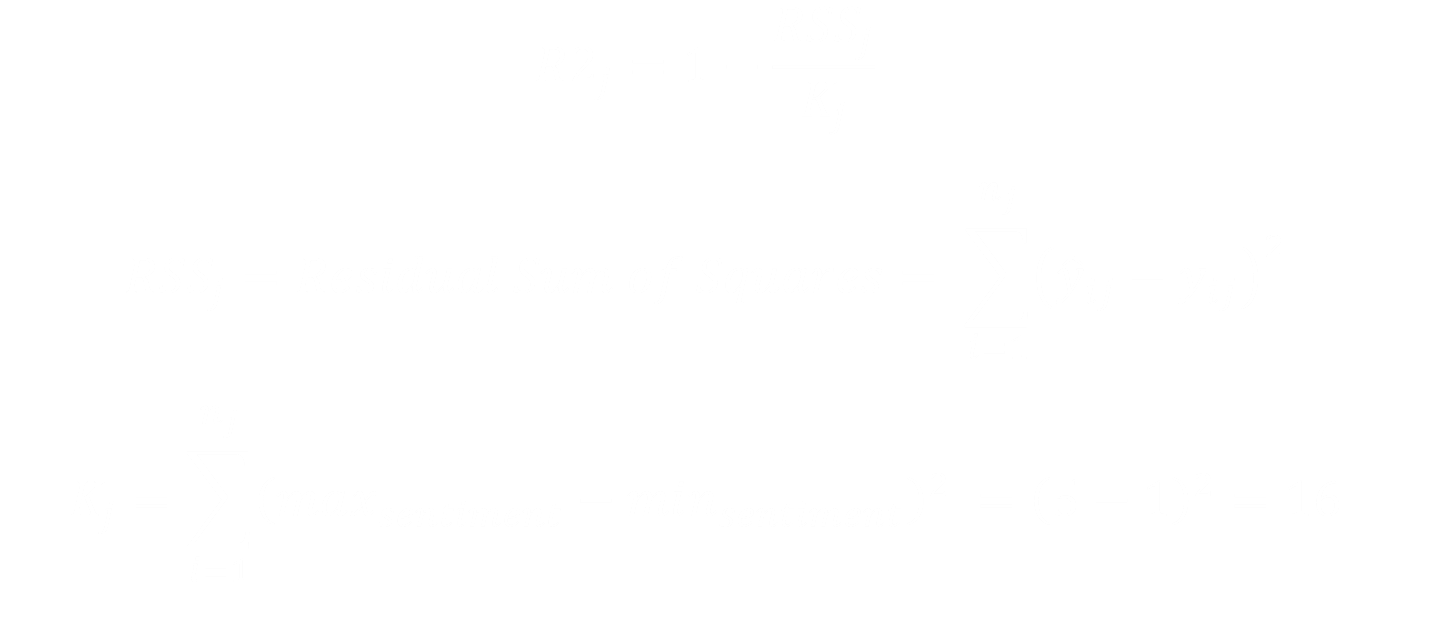
* Tính F1-score cho từng khía cạnh j:



**Lưu ý:** Khi tính F1-score thì các giá trị 1-5 đều xem là 1 (chỉ quan tâm có khía cạnh hay không, không quan tâm sentiment), giá trị 0 vẫn giữ nguyên như output của bài toán Classification thông thường nhận giá trị đầu ra {0, 1}.

**3. Metrics for 5-Stars Scale Rating:**

* Xem như bài toán regression.
* Tính R2 score cho từng khía cạnh j:

****

Trong đó:

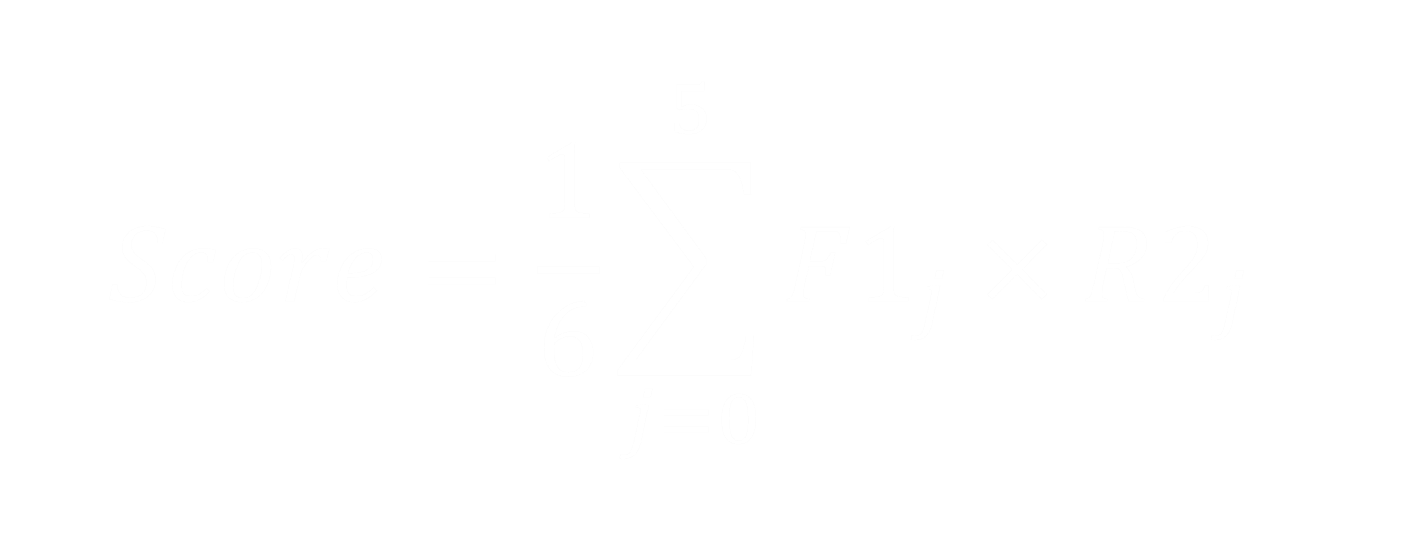
* + y*ij^*: giá trị dự đoán của khía cạnh j thuộc samples i.
  + y*ij*: giá trị đúng của khía cạnh j thuộc sample i.
  + n*j*: số samples dự đoán đúng tại khía cạnh j.
  + max\_sentiment = 5 (5 sao)
  + min\_sentiment = 1 (1 sao)

***Lưu ý:***

* Nếu nj = 0 (tức là không có sample nào đúng tại khía cạnh j) thì gán R2 = 1, vì khi đó F1\_score = 0, giá trị R2 bao nhiêu cũng không ảnh hưởng.
* Chỉ tính R2-score đối với khía cạnh của samples mà ở đó giá trị dự đoán và giá trị thực tế đều khác 0.
  + trường hợp dự đoán sai khía cạnh đã được tính ở bước F1\_score.
  + trường hợp cà 2 đều = 0: MSE = 0 => R2 = 1 (sai ý nghĩa => loại)

=> Chỉ tính R2 khi cả 2 đều khác 0

**4. Final metrics:**

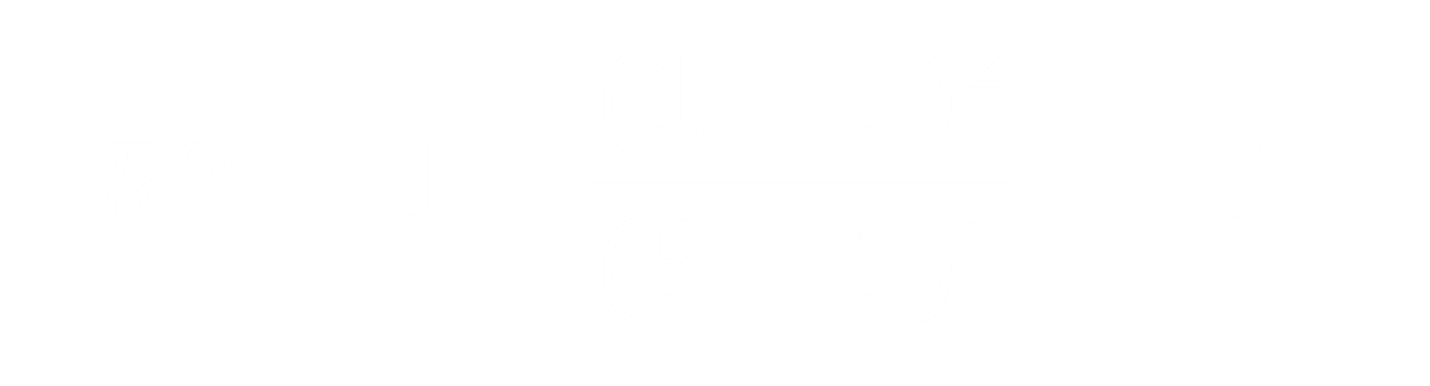


**5. Example:**

pred = [[1, 0, 3, 4, 2, 0], [0, 2, 2, 5, 1, 0]]  
gt = [[1, 1, 3, 5, 3, 1], [0, 0, 3, 4, 0, 0]]

Xét từng khía cạnh:

giai\_tri: pred: [1, 0]; gt: [1, 0].



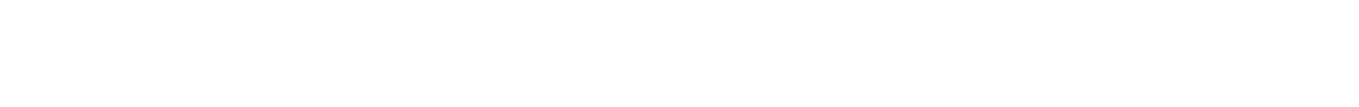
luu\_tru: pred: [0, 2] => format: [0, 1]; gt: [1, 0]

F1\_score = 0.

R2\_score = 1 (nj = 0: ko có khía cạnh nào đúng => R2 = 1 như đề cập ở trên)

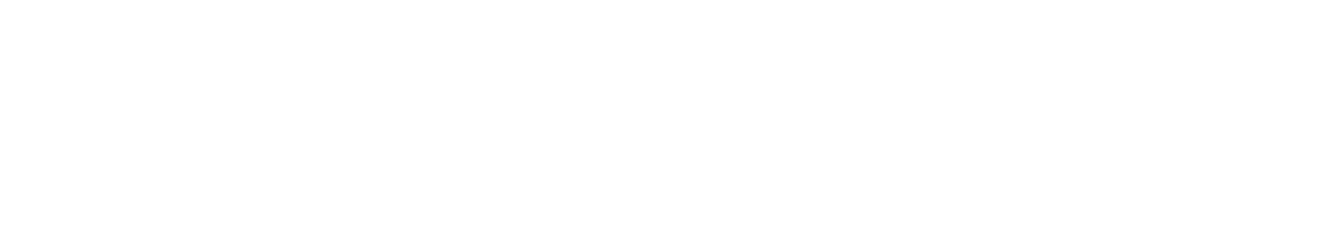
nha\_hang: pred: [3, 2] => format: [1, 1]; gt: [3, 3] => format: [1, 1]

F1\_score = 1.



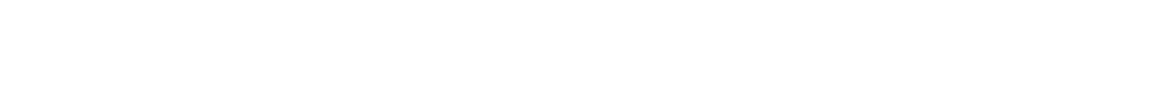
an\_uong: pred: [4,5] => format: [1,1]; gt: [5, 4] => format: [1, 1].

F1\_score = 1



di\_chuyen: pred: [2,1] => format: [1,1]; gt: [3,0] => format: [1,0]

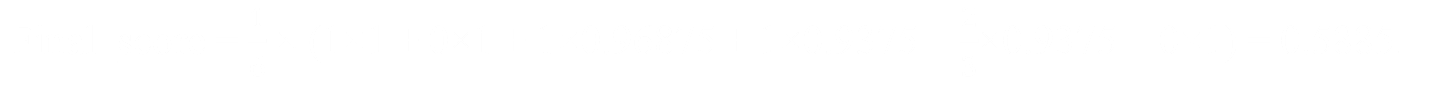
F1\_score = 2/3



mua\_sam: pred: [0,0]; gt: [1, 0].

F1\_score = 0.

R2\_score = 1.



Các đội thi tải evaluate code tại đây