**REPORT**

**Datasets train:**

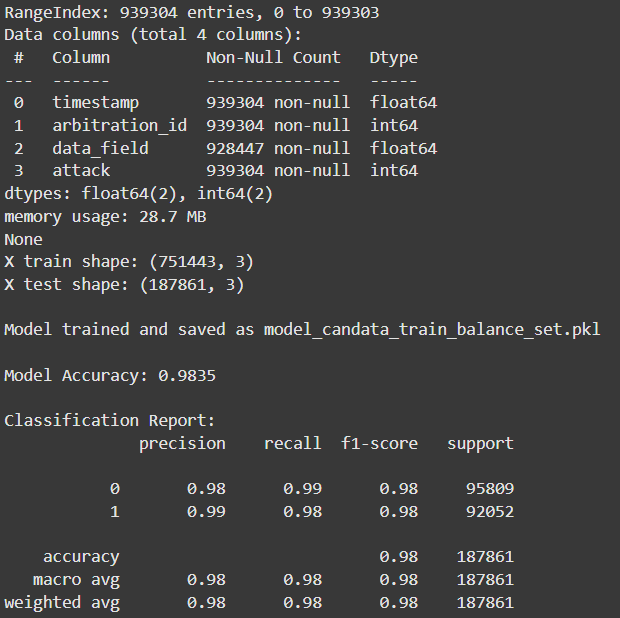
- Là sự kết hợp giữa các dataset bao gồm: set\_01, class 1 set\_02, set\_03, set\_04 và được tiếp tục tiền xử lý theo tỉ lệ class 0: 51%, class 1: 49%. Để có bộ dataset cân bằng tỉ lệ giữa các class.

- Tổng số mẫu: 939304, trong đó lại đựọc chia để phục vụ quá trình train:

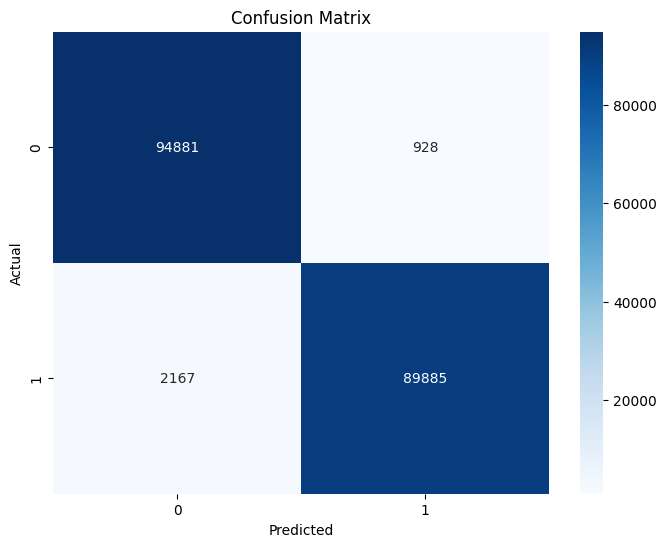
+ train shape:751443 mẫu

+ test shape: 187861 mẫu

**Thông số train:**



-> Độ chính xác : 0.98



-> Confusion matrix (ma trận nhầm lẫn) trong quá trình train và validation

· **TP (True Positive)** = 89885  
→ Dự đoán đúng lớp 1 (có tấn công)

· **TN (True Negative)** = 94881  
→ Dự đoán đúng lớp 0 (bình thường)

· **FP (False Positive)** = 928  
→ Dự đoán sai lớp 1, nhưng thực tế là lớp 0

· **FN (False Negative)** = 2167  
→ Dự đoán sai lớp 0, nhưng thực tế là lớp 1

**Quá trình test thử model**

Set 01/ test 01:

* Tổng mẫu: 5702670
* True 0.980565
* False 0.019435

Set 01/ test 02:

* total: 6447917
* True 0.579509
* False 0.420491

Set 01/ test 03:

* total: 8635275
* True 0.984669
* False 0.015331

Set 01/ test 04:

* total: 13220555
* True 0.838909
* False 0.161091

-> Đảm bảo về chất lượng file .pkl.

Quá trình chuyển đổi model

Tổng số cây: 21

Thông tin quá trình chuyển đổi:

file: tree\_0.bin includes 369 lines and each row is 16 bytes

file: tree\_1.bin includes 535 lines and each row is 16 bytes

file: tree\_2.bin includes 377 lines and each row is 16 bytes

file: tree\_3.bin includes 465 lines and each row is 16 bytes

file: tree\_4.bin includes 491 lines and each row is 16 bytes

file: tree\_5.bin includes 353 lines and each row is 16 bytes

file: tree\_6.bin includes 347 lines and each row is 16 bytes

file: tree\_7.bin includes 325 lines and each row is 16 bytes

file: tree\_8.bin includes 409 lines and each row is 16 bytes

file: tree\_9.bin includes 347 lines and each row is 16 bytes

file: tree\_10.bin includes 491 lines and each row is 16 bytes

file: tree\_11.bin includes 449 lines and each row is 16 bytes

file: tree\_12.bin includes 501 lines and each row is 16 bytes

file: tree\_13.bin includes 433 lines and each row is 16 bytes

file: tree\_14.bin includes 309 lines and each row is 16 bytes

file: tree\_15.bin includes 323 lines and each row is 16 bytes

file: tree\_16.bin includes 521 lines and each row is 16 bytes

file: tree\_17.bin includes 333 lines and each row is 16 bytes

file: tree\_18.bin includes 341 lines and each row is 16 bytes

file: tree\_19.bin includes 519 lines and each row is 16 bytes

file: tree\_20.bin includes 363 lines and each row is 16 bytes

File tree bin có cấu trúc như sau ở mỗi dòng:

**node, feature, threshold, left, right, prediction**

**2bytes,4bytes,4bytes,2bytes,2bytes,1bytes**

**uint16, int32, float32, uint16, uint16, signedChar**

-> Tuy nhiên có padding thêm 1 bytes -> 16 bytes.