

Goal

เขียนโปรแกรมที่รับเอา order_id ทั้งหมด 5 ชุด ไปคำนวณเพื่อทำนายว่าสินค้าอะไรที่เคยซื้อมาแล้ว จะถูกซื้อใน order นี้ โดยโปรแกรมจะต้องโหลดไฟล์ order_ids_x.csv ทั้งหมด 5 ชุด (ในโฟลเดอร์ dataset) และบันทึกผลลัพธ์การทำนายของแต่ละชุดไว้ในไฟล์โดยตั้งชื่อว่า submission_x.csv ทั้งหมด 5 ชุด (ในโฟลเดอร์ submissions)

รูปแบบของไฟล์ submission_x.csv จะเป็นดังนี้

```
order_id,products
29,1 2
99,None
323,4
```

products คือ product_id ที่เรากำหนดว่าจะมีการซื้อใน order นี้

*****ย้ำว่า product นี้จะนับเฉพาะที่ user คนนี้ เคยซื้อมาก่อนเท่านั้น ถ้าทำนายออกมาว่าไม่มี product ที่เคยซื้อมาก่อนเลย (มีแต่ product ใหม่) ให้ส่งคำตอบเป็น 'None'**

*****dataset นี้มีการย่อขนาดข้อมูลลง เพื่อประหยัดเวลาในการคำนวณ เพราะฉะนั้นจะมี order จาก user เพียง 10000 คนเท่านั้น**

Evaluation

การวัดประสิทธิภาพของ model ในโจทย์นี้เราใช้ตัวที่ชื่อว่า F1 Score

ทำการรัน evaluate.py เพื่อทำการ evaluate

โปรแกรมจะโหลด submission_x.csv ทั้งหมด 5 ไฟล์ ที่อยู่ในโฟลเดอร์ submission แล้วนำไปตรวจคำตอบกับไฟล์ dataset/test_set.csv

ตัวอย่าง ผลลัพธ์การรัน evaluate.py

```
#1 set: 0.38057905798
#2 set: 0.375602553014
#3 set: 0.380567793281
#4 set: 0.361075611868
#5 set: 0.372231120906
Total Average: 0.37401122741
```

***** F1 Score อ้างอิงจาก Leaderboard ของ kaggle**

- 0.35 พอใช้
- 0.38 ดี
- 0.4 ดีมาก