ข้อสอบมีทั้งหมดจำนวน 1 ข้อ

ข้อที่ 1 : Natural Language Processing (การประมวลผลภาษาธรรมชาติ)

ให้ทำ model muti-label text classifation เพื่อทำนายบทความในแต่ละบทความพูดถึงเนื้อหาด้านใดบ้างจาก 12 ด้านดังนี้

(1 ข้อความสามารถพูดถึงเนื้อหาได้หลายด้าน อย่างน้อยจะเกี่ยวข้อง 1 ด้าน)

{0 : 'การเมือง',

1 : 'สิทธิมนุษยชน',

2 : 'คุณภาพชีวิต' ,

3 : 'ต่างประเทศ',

4 : 'สังคม',

5 : 'สิ่งแวดล้อม',

6 : 'เศรษฐกิจ',

7 : 'วัฒนธรรม',

8 : 'แรงงาน',

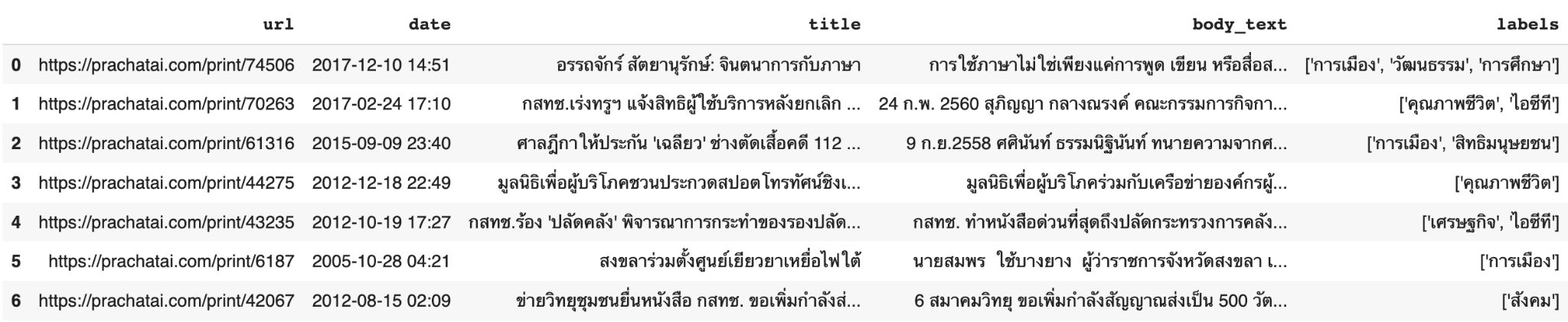
9 : 'ความมั่นคง',

10 : 'ไอซีที',

11 : 'การศึกษา',}

โดยจะไฟล์ข้อมูลให้ 2 ไฟล์ ได้แก่

1. ไฟล์ exam03\_trian.csv เป็นไฟล์ข้อมูลที่ให้ใช้สำหรับการ train โมเดล text classifation



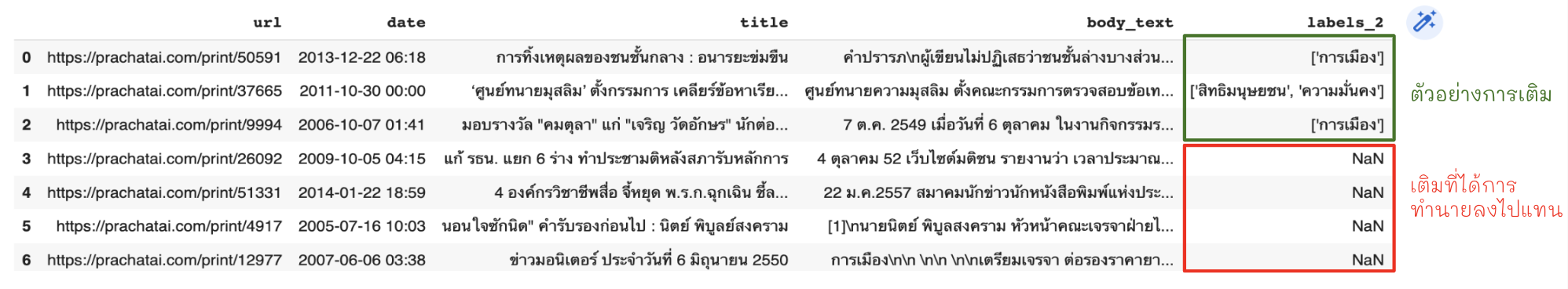
โดยที่ tiltle = ข้อหัวของบทความ

text = เนื้อหาของบทความ

date = วันที่ที่มีการเผยแพร่บทความ

labels = ประเภทเนื้อหาที่มีเกี่ยวข้องกับบทความ

1. ไฟล์ exam03\_test.csv เป็นไฟล์ข้อมูลที่ให้ผู้เข้าสอบส่งผลการทำนายประเภทเนื้อหาของบทความ ลงใน column ที่ชื่อว่า labels



สามารถ โหลด data ได้ที่ : <https://drive.google.com/drive/folders/1dchHpJffRf32Hmi92AURBHINWu-rxHb9?usp=sharing>

เมื่อทำเสร็จแล้วให้ส่ง

1.ไฟล์ผลลัพธ์การทำนาย (exam03\_test.csv )

2.อธิบายกระบวนการจัดการข้อมูลและเหตุผลการเลือกโมเดล (ทำเป็นสไลด์หรือเขียนสรุปมาก็ได้)

3.ไฟล์ notebook (ถ้ามี)

ส่งมาที่ email hr@thebrainstem.com