

แบบทดสอบ

Data Scientist

Background

ICD10 คือ บัญชีจำแนกโรค และสาเหตุความเจ็บป่วย ซึ่งมีทั้งหมด 35,000 codes

ปัจจุบัน Medical coder เป็นผู้แปลงการวินิจฉัยของแพทย์ ที่เป็น free text ทั้งการพิมพ์แบบตัวเต็มและ ตัวย่อ ให้เป็นรหัส ICD10 เราจึงอยากลองสร้างกระบวนการบางอย่างมาช่วยลดภาระงานที่สามารถ แปลง การวินิจฉัยแพทย์เป็น ICD10 เนื่องจาก 35,000 codes นั้นมากเกินไป เราจึงสุ่มมาจำนวนหนึ่งเพื่อใช้เป็นแบบทดสอบ

โดยในแบบทดสอบนี้มีข้อมูลมาให้ ประกอบไปด้วย

dataset.tsv ข้อมูล ประกอบด้วย column code = ICD10 และ column term = การวินิจฉัยของแพทย์

icd10_choose.json = รายละเอียดของ ICD10 ที่อยู่ใน column code สามารถใช้ประกอบการแก้ไขปัญหาได้

โจทย์

1. Propose your ideas to solve this problem

Note 1: should be more than one approach.

Note 2: not necessary to be a machine learning.

Note 3: you can use any existed algorithm, pretrained models, or external data to solve this problem.

2. Explain something about the given dataset

3. Pick one (or more) idea(s) to show us how to solve the problem in Python language

Note: We have the secret validation set. Please prepare your code to receive this validation dataset.

Note: you can write in Jupyter Notebook or raw Python file.

4. (Extra score/Optional) deploy it as API. Optionally build a Docker image.

การส่งแบบทดสอบ

ให้อัพโหลดลงใน Github และส่งมาที่อีเมล Contact@sati.co.th ภายในวันที่ 12 มิถุนายน 2565 ก่อนเวลา 12.00 น.เท่านั้น