

Workshop : การประเมินราคาเสื้อผ้ามือสอง สัปดาห์ที่ 2

กำหนดการ: 23 – 29 มิถุนายน

วัตถุประสงค์: ออกแบบโมเดลเพื่อประเมินราคาเสื้อผ้ามือสอง

ชุดข้อมูล (Dataset):

- ไฟล์ `.zip` จะประกอบด้วยชุดข้อมูลสำหรับฝึก (`train.csv`) และทดสอบ (`test.csv`) ในรูปแบบ CSV พร้อมไฟล์สำหรับรูปภาพ (`images`) ประกอบไปด้วยรูปภาพของเสื้อผ้า และไฟล์เสียง (`audio`) ที่ประกอบไปด้วยเสียงบรรยายคุณลักษณะของเสื้อผ้าแต่ละชิ้น
- ไฟล์ `train.csv` : จะมีชื่อไฟล์รูปภาพในคอลัมน์แรกที่สอดคล้องกับชื่อไฟล์เสียง (โดยนามสกุลจะเปลี่ยนจาก `.jpg` เป็น `.wav`) และมี Label ของแต่ละคุณสมบัติ (`Type, Color, Condition, Smell, Pilling, size, material, cut, season, usage, price`) กำหนดให้ น้องๆจะต้องใช้โมเดลจากสัปดาห์แรก ทำนายผลของคอลัมน์ `Type, Color, Condition, Smell, Pilling` เพื่อนำมาใช้เป็นข้อมูลในการทำโจทย์ สัปดาห์ที่ 2
- ไฟล์ `test.csv` : จะมีคอลัมน์แรกเป็นชื่อไฟล์รูป ซึ่งน้องๆ จะต้องใช้โมเดลจากสัปดาห์แรกของตนเองทำนายในแต่ละคอลัมน์ `Type, Color, Condition, Smell, Pilling` เพื่อเติมข้อมูลในตารางให้เต็ม และใช้โมเดลของสัปดาห์ที่ 2 ในการทำนายราคา และส่งผลในรูปแบบ CSV เพื่อส่ง submission ตามตัวอย่างไฟล์ `sample_submission.csv` เพื่อวัดประสิทธิภาพ

image_file	price
de5a64cd-8fab-4aa8-99bf-cf8052377141.jpg	10
f45dccd5-bf53-4610-b0a6-938d008bb547.jpg	23.5
106f939f-ab62-4b9d-83d6-099a6b084478.jpg	49.4
8a058855-b7a0-4c93-86f3-9e38bb6ea073.jpg	23.5
188ff116-3d74-457c-94af-a498a9d58330.jpg	234.5
f9e16468-943a-4805-a677-cae34bfa9bdd.jpg	23.5

Task ที่ต้องทำนาย (Single task):

1. การทำนายราคา (Price Prediction)

เกณฑ์การให้คะแนนโมเดล:

- **ความคิดสร้างสรรค์ในการออกแบบโมเดล (Model Creative Design)**
- **ขนาดโมเดล (Model Size):** โมเดลขนาดเล็กจะได้คะแนนสูง (Small size: High score)
- **การฝึกโมเดล:** สร้างและฝึกโมเดลเองทั้งหมด (Train from scratch) **ไม่อนุญาตให้ใช้ Transfer Learning**
- **การแสดงผล Feature Map (Feature Map Visualization)**
- **ประสิทธิภาพ (Performance):** คะแนน Public และคะแนน Private
 - **ตัวชี้วัด (Metric):** ค่า Exponential Negative Weighted MSE

$$\text{Exponential Negative Weighted MSE} = 1.5 \cdot e^{(-k \cdot \text{wMSE})}$$

$$\text{Weighted MSE} = \text{wMSE} = w \cdot \text{MSE}$$

- โดย $k = 5$, $w = 10^{-4}$
- $\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_{\text{true}} - y_{\text{predict}})^2$

เกณฑ์การให้คะแนนการวิเคราะห์ข้อมูล:

- ลักษณะทางสถิติ (Statistical Characteristics)
- ข้อมูลผิดปกติ (Outlier)
- การเตรียมและทำความสะอาดข้อมูล (Data Cleansing and Preparation)
- การเพิ่มข้อมูล (Data Augmentation)

เงื่อนไขการส่งผลงานใน Challenge:

- **วันเริ่มต้น:** 23 มิถุนายน (10.00 น.)
- **วันสิ้นสุด:** 29 มิถุนายน (23.59 น.)
- **จำนวนครั้งที่ส่งได้ต่อวัน:** 5 ครั้ง
- **เวลาที่ระบบปรับเพิ่มจำนวนครั้งการส่ง:** 22.00 น. ของทุกวัน (ยกเว้นวันที่ 29 มิถุนายน)
- **ข้อควรระวัง:** ควรตรวจสอบให้มั่นใจก่อนส่ง จะไม่มีการเคลียร์ผล CSV check หรือ submission

สิ่งที่ต้องส่ง:

- **ในเว็บไซต์ Exercise:** อธิบายผลการทดลองในรูปแบบไฟล์ PDF ประกอบด้วย:

- แนวคิดการออกแบบโมเดล (Model Design Idea)
- วิธีการฝึกโมเดลและพารามิเตอร์ (Training method and parameters)
- การเตรียมข้อมูล (EDA, Augmentation, Cleansing)
- การวิเคราะห์ผลการทำนาย (Prediction Result Analysis)
- **ในเว็บไซต์ Challenge:** สำหรับ Check และ Submission:
 - ไฟล์ CSV (ผลการทำนาย)
 - ซอร์สโค้ด (ไฟล์ `.ipynb`)
 - ไฟล์โมเดล (`.pth` หรือ `.pt`)