

Workshop : การประเมินราคาเสื้อผ้ามือสอง

เวิร์คช็อปนี้มุ่งเน้นไปที่การประเมินราคาเสื้อผ้ามือสอง โดยน้องๆ จะได้ออกแบบโมเดลเพื่อแก้ไขปัญหาใน 2 สัปดาห์ ข้อมูลสำหรับสัปดาห์แรกประกอบด้วยรูปภาพและเสียง ซึ่งน้องๆ สามารถนำไปใช้ทำนายคุณสมบัติต่างๆ เช่น ประเภท, สี, สภาพ, กลิ่น, และการเป็นขุย ข้อมูลที่โมเดลของน้องๆ ทำนายได้นี้จะถูกนำไปใช้ต่อในเวิร์คช็อปสัปดาห์ที่ 2 เพื่อประเมินราคาเสื้อผ้ามือสอง

เวิร์คช็อปสัปดาห์ที่ 1: การจำแนกประเภท (Classification Model)

กำหนดการ: 16 – 22 มิถุนายน

วัตถุประสงค์: ออกแบบโมเดลเพื่อประเมินคุณสมบัติของเสื้อผ้ามือสองจากข้อมูลรูปภาพและเสียง

ชุดข้อมูล (Dataset):

- ไฟล์ `.zip` จะประกอบด้วยชุดข้อมูลสำหรับฝึก (`train.csv`) และทดสอบ (`test.csv`) ในรูปแบบ CSV พร้อมไฟล์เดอร์สำหรับรูปภาพ (`images`) ประกอบไปด้วยรูปภาพของเสื้อผ้า และไฟล์เสียง (`audio`) ที่ประกอบไปด้วยเสียงบรรยายคุณลักษณะของเสื้อผ้าแต่ละชิ้น
- ไฟล์ `train.csv` : จะมีชื่อไฟล์รูปภาพในคอลัมน์แรกที่สุดอดคล้องกับชื่อไฟล์เสียง (โดยนามสกุลจะเปลี่ยนจาก `.jpg` เป็น `.wav`) และมี Label ของแต่ละคุณสมบัติ (`Type, Color, Condition, Smell, Pilling`) กำหนดให้
- ไฟล์ `test.csv` : จะมีคอลัมน์แรกเป็นชื่อไฟล์รูป ซึ่งน้องๆ จะต้องใช้โมเดลของตนเองทำนายในแต่ละคอลัมน์ที่เหลือได้แก่ `Type, Color, Condition, Smell, Pilling` และส่งผลในรูปแบบ CSV เพื่อส่ง submission เพื่อวัดประสิทธิภาพ

ตัวอย่าง submission csv

image_name	condition	pilling	smell	color	type
80...c.jpg	2	5	TRUE	White	top
b8...4.jpg	2	3	FALSE	Green	bottom

- ข้อควรทราบ:** น้องๆ ต้องใช้ไฟล์ CSV ในการระบุรูปภาพและไฟล์เสียงใดที่ใช้สำหรับฝึกหรือทดสอบเพื่อส่ง submission

Task ที่ต้องทำนาย (Multi-task):

- ประเภท (Type) : `['bottom' 'top']`
- สี (Color) : `['Black' 'Blue' 'Brown' 'Gray' 'Green' 'Orange' 'Pink' 'Purple' 'Red' 'White' 'Yellow']`
- สภาพ (Condition) : `[1 2 3 4 5]`
- กลิ่น (Smell) : `[False True]`
- การเป็นขุย (Pilling) : `[1 2 3 4 5]`

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): **Weighted F-beta Score**

- คำนวณจากค่าเฉลี่ยถ่วงน้ำหนักของ F-beta score ของทั้ง 5 tasks โดยมีค่า β เฉพาะสำหรับแต่ละ tasks ดังนี้:

$$\text{Weighted F-beta Score} = (0.19 \cdot F_{\beta_{\text{Condition}}}) + (0.20 \cdot F_{\beta_{\text{Pilling}}}) + (0.23 \cdot F_{\beta_{\text{Smell}}}) + (0.20 \cdot F_{\beta_{\text{Color}}}) + (0.18 \cdot F_{\beta_{\text{Type}}})$$

โดยค่า β สำหรับแต่ละคุณสมบัติคือ:

- $\beta_{\text{Condition}} = 1.5$
- $\beta_{\text{Pilling}} = 1.5$
- $\beta_{\text{Smell}} = 2.0$
- $\beta_{\text{Color}} = 1.0$
- $\beta_{\text{Type}} = 1.0$

เกณฑ์การให้คะแนนการวิเคราะห์ข้อมูล:

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

เงื่อนไขการส่งผลงานใน Challenge:

- วันเริ่มต้น: 16 มิถุนายน (10.00 น.)
- วันสิ้นสุด: 22 มิถุนายน (23.59 น.)
- จำนวนครั้งที่ส่งได้ต่อวัน: 5 ครั้ง
- เวลาที่ระบบปรับเพิ่มจำนวนครั้งการส่ง: 22.00 น. ของทุกวัน (ยกเว้นวันที่ 22 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`)