

การให้ป้ายกำกับสำหรับข้อมูล (Data labeling)

Data Labeling คือ การระบุประเภทของข้อมูลให้ชัดเจนว่าเป็นข้อมูลประเภทใด เช่น การระบุประเภทว่า รูปข้อมูลนี้ เป็น รูปคน สัตว์ หรือสิ่งของต่าง ๆ หรือ ระบุว่าข้อความนี้มีความหมายไปในเชิงแง่บวกหรือแง่ลบ โดยชุดข้อมูลผ่านการให้ป้ายกำกับหรือผ่านการระบุประเภทมาแล้ว จะถูกนำไปใช้ต่อไปเพื่อเป็นชุดฝึกสอนให้ระบบเรียนรู้และทำนายข้อมูลที่มีโครงสร้างที่มีความคล้ายกัน โดยคอมพิวเตอร์จะพิจารณาจากหลายมิติตามสูตรคณิตศาสตร์ หาความเชื่อมโยง ซึ่งถ้าหากระบวนการที่ได้มีความแม่นยำที่สูงก็จะสามารถนำไปประยุกต์ เพิ่มประสิทธิภาพให้กับการทำงานได้ในหลากหลายอุตสาหกรรมแทนการใช้แรงงานคน

วัตถุประสงค์ของงานนี้ คือ การหาบทวิจารณ์สินค้าที่มีประโยชน์ต่อผู้อ่าน เพื่อให้ได้ชุดฝึกสอนที่สามารถนำไปฝึกระบบ ให้สามารถช่วยลดเวลาในการตัดสินใจซื้อสินค้าให้กับผู้ใช้งานในเว็บไซต์ **E-commerce** ที่มีการซื้อขายของออนไลน์ แนะนำเฉพาะบทวิจารณ์ที่เป็นประโยชน์ให้กับผู้อ่าน แต่ในการระบุประเภทข้อมูลที่เป็นข้อความภาษาในโซเชียลมีเดีย การระบุประเภทของบทวิจารณ์สินค้าที่มีเพียง 2 กลุ่ม ได้แก่ มีประโยชน์ และ ไม่มีประโยชน์ อาจมีความคลาดเคลื่อนที่สูงและทำให้ข้อมูลฝึกสอนไม่มีคุณภาพ ดังนั้น จึงจำเป็นต้องมีการแบ่งการระบุข้อมูลตามคุณสมบัติของบทวิจารณ์สินค้าที่ดี แบ่งออกเป็น 3 ปัจจัย ได้แก่ ความน่าเชื่อถือ (credibility), เนื้อหา (content) และการแสดง (expression)

1. ความน่าเชื่อถือ: (credibility)

บทวิจารณ์สินค้าที่มีความเชื่อถือได้คือบทวิจารณ์ที่ผู้อ่านมีความเชื่อว่าเป็นข้อมูลที่มีประโยชน์และช่วยในการตัดสินใจซื้อสินค้าได้ มีข้อมูลเกี่ยวข้องกับตัวสินค้าที่ถูกต้องและชัดเจน มีการบอกเล่าประสบการณ์การใช้งานส่วนตัวทำให้บทวิจารณ์นี้มีความเป็นเอกลักษณ์ ทำให้ผู้อ่านรู้สึกได้ว่าไม่เป็นเท็จ ไม่ใช่การสแปม และไม่ทำให้ผู้อ่านสับสน เช่น “ขอลองใช้ดูก่อน”, “ยังไม่ได้ลองใช้งาน”

2. เนื้อหา: (content)

บทวิจารณ์สินค้าที่มีเนื้อหาคือบทวิจารณ์สินค้าที่มีการวิเคราะห์และอธิบายคุณภาพของตัวสินค้า รูปแบบการใช้งานของสินค้า การแพ้คัดสินค้า และบริการของบริษัท เพื่อให้ผู้อ่านได้รับข้อมูลเพื่อช่วยในการตัดสินใจในการซื้อสินค้า โดยเนื้อหาที่ดีควรกล่าวถึงคุณสมบัติสินค้าและบริการอย่างชัดเจน เช่น “เสียงดี”, “น้ำหนักเบา”, “การขนส่งรวดเร็ว” ไม่ควรพูดถึงภาพรวมที่อาจทำให้ผู้อ่านสับสน เช่น “คุณภาพดี”, “ประสิทธิภาพปานกลาง”, “การใช้งานโอเค”

3. การแสดงความรู้สึก: (expression)

บทวิจารณ์สินค้าที่มีการแสดงความรู้สึกคือบทวิจารณ์สินค้าที่เน้นการแสดงอารมณ์และความรู้สึกของผู้ซื้อต่อสินค้าหรือบริการนั้น ว่าเห็นด้วยหรือไม่เห็นด้วยกับสินค้าหรือบริการดังกล่าว อาจใช้ภาษาในเชิงของการชักชวนให้คล้ายตามให้กับผู้อ่าน บทวิจารณ์ที่แสดงออกถึงความรู้สึกอย่างชัดเจน เช่น “ประทับใจมาก”, “ร้านนี้ดีกว่าร้านอื่น”, “รอบหน้าจะไม่ซื้ออีก”

คำอธิบายข้อมูล

ข้อมูลบทวิจารณ์สินค้าไอที ประกอบไปด้วย 5 หมวดหมู่สินค้า รวมกันจำนวน 1,000 บทวิจารณ์ ข้อความของบทวิจารณ์ซึ่งสามารถเป็นไปได้อีกทั้งภาษาไทยหรือภาษาอังกฤษ

งาน: อ่านแต่ละรีวิวและให้คะแนนของแต่ละคุณสมบัติ โดยให้คะแนน ดังต่อไปนี้

3 คะแนน	บทวิจารณ์สินค้ากล่าวถึงคุณลักษณะ ชัดเจน - ความน่าเชื่อถือ มีคุณสมบัติของ 2 คะแนน มีรูปประกอบชัดเจน บทวิจารณ์มีความยาวที่เหมาะสม - เนื้อหา มีการอธิบายการใช้งานของสินค้า ข้อดีหรือข้อเสีย รวมไปถึงส่วนบริการอื่นๆ - ความรู้สึก มีการใช้คำชักชวนให้ซื้อหรือไม่ซื้อ แสดงถึงความรู้สึกชอบไปในทิศทางหนึ่งที่ชัดเจน
2 คะแนน	บทวิจารณ์สินค้ากล่าวถึงคุณลักษณะ ปานกลาง - ความน่าเชื่อถือ มีเอกลักษณ์ แสดงถึงประสบการณ์ที่ได้รับและประสบการณ์ส่วนตัวจากการซื้อ/ใช้งาน - เนื้อหา อธิบายคุณลักษณะสินค้า การบริการ หรือ การขนส่ง ไม่ครบแต่มีเนื้อหาครอบคลุม - ความรู้สึก แสดงความรู้สึกชอบของตัวสินค้าไปในทิศทางหนึ่งที่ชัดเจน เช่น สินค้าดี บริการดี
1 คะแนน	บทวิจารณ์สินค้ากล่าวถึงคุณลักษณะ ไม่ชัดเจนหรือมีน้อย - ความน่าเชื่อถือ มีประโยชน์สั้น มีเนื้อหาน้อยแต่ยังสามารถจับใจความสำคัญที่มีต่อสินค้าได้ - เนื้อหา อธิบายคุณลักษณะสินค้า การบริการ หรือ การขนส่ง อย่างใดอย่างหนึ่งเพียงอย่างเดียว - ความรู้สึก แสดงถึงความรู้สึกที่ไม่ชัดเจน แสดงในทิศทางที่ดีบ้างและไม่ดีบ้าง
0 คะแนน	บทวิจารณ์สินค้าไม่มีการกล่าวถึงคุณลักษณะเลย หรือคาดว่าอาจเป็นสแปม

Note: ในลิงก์ที่ให้จะมีอยู่ 3 ไฟล์ครับ

- data.xlsx ไฟล์ที่จะให้ทำ (สามารถแก้ไขในตัวไฟล์ได้เลยครับ)
- Example.xlsx ไฟล์ตัวอย่าง ดูประกอบการทำได้
- Instruction.pdf