

การป้ายกำกับสำหรับข้อมูล (Data labeling)

Data Labeling คือ การระบุประเภทของข้อมูลให้ชัดเจนว่าเป็นข้อมูลประเภทใด เช่น การระบุประเภทว่า รูปข้อมูลนี้เป็น รูปคน สัตว์ หรือสิ่งของต่าง ๆ หรือ ระบุว่าข้อความนี้มีความหมายไปในเชิงแง่บวกหรือแง่ลบ โดยชุดข้อมูลที่ผ่านการให้ป้ายกำกับหรือผ่านการระบุประเภทมาแล้ว จะถูกนำไปใช้ต่อเพื่อเป็นชุดฝึกสอนให้ระบบเรียนรู้และทำนายข้อมูลที่มีโครงสร้างที่มีความคล้ายกัน โดยคอมพิวเตอร์จะพิจารณาจากหลายมิติตามสูตรคณิตศาสตร์ หาความเชื่อมโยง ซึ่งถ้าหากระบบที่ได้มีความแม่นยำที่สูงก็จะสามารถนำไปประยุกต์เพิ่มประสิทธิภาพให้กับการทำงานได้ในหลากหลายอุตสาหกรรมแทนการใช้แรงงานคน

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 มีการบอกเล่าประสบการณ์การใช้งานส่วนตัวทำให้บทวิจารณ์นี้มีความเป็นเอกลักษณ์ไม่ใช่การสแปม

2. เนื้อหา: (content)

บทวิจารณ์สินค้าที่มีเนื้อหาคือบทวิจารณ์สินค้าที่มีการวิเคราะห์และอธิบายคุณภาพของตัวสินค้า รูปแบบการใช้งานของสินค้า การแพ้สินค้า และบริการของบริษัท เพื่อให้ผู้อ่านได้รับข้อมูลเพื่อช่วยในการตัดสินใจในการซื้อสินค้า

3. การแสดงความรู้สึก: (expression)

บทวิจารณ์สินค้าที่มีการแสดงความรู้สึกคือบทวิจารณ์สินค้าที่เน้นการแสดงอารมณ์และความรู้สึกของผู้ซื้อต่อสินค้าหรือบริการนั้น ว่าเห็นด้วยหรือไม่เห็นด้วยกับสินค้าหรือบริการดังกล่าว โดยมีการใช้ภาษาและรูปแบบที่สร้างความเข้าใจและความไว้วางใจให้กับผู้อ่าน

งาน: อ่านแต่ละรีวิวและให้คะแนนของแต่ละคุณสมบัติ โดยให้คะแนน ดังต่อไปนี้

3 คะแนน	บทวิจารณ์สินค้ากล่าวถึงคุณลักษณะ ชัดเจน
2 คะแนน	บทวิจารณ์สินค้ากล่าวถึงคุณลักษณะ ปานกลาง
1 คะแนน	บทวิจารณ์สินค้ากล่าวถึงคุณลักษณะ ไม่ชัดเจนหรือมีน้อย
0 คะแนน	บทวิจารณ์สินค้าไม่มีการกล่าวถึงคุณลักษณะเลย

คำอธิบายข้อมูล

ข้อมูลบทวิจารณ์สินค้าไอที ประกอบไปด้วย หมวดหมู่ของสินค้าและข้อความที่เป็นของบทวิจารณ์ซึ่งสามารถเป็นไปได้ทั้งภาษาไทยหรือภาษาอังกฤษ

มีรอยเปื้อนนิดหน่อย	1	1	1
ทางร้านดีมากครับแบบเคลมสินค้าให้ด้วยสินค้าใช้ไม่ได้แต่ทางร้านบริการดีมากครับส่งอันใหม่มาให้	2	1	2
สั่งสินค้าเงิน ได้สีเหามา เทอออ	1	1	1
สั่งสีแดงแต่ได้สีเขียว...แต่ไม่เป็นไรรับได้	1	1	0
ได้รับสีโพงตรงตามที่โฆษณา ได้รับสีตรงตามที่เลือก รูปทรงน่ารัก เสียงพอใช้ ฟังก์ชันใช้งานง่าย ตรงมมมีบุบไปบ้างจากการขนส่ง แต่ไม่มาก อยากให้ห่อหุ้มกันกระแทกให้ดี	3	3	2
ดีค่ะ สกกับราคา แคมสามารถใช้ดูโทรศัพท์ได้ค่ะ รวมแล้วคือดีหมดค่ะ แพ็คดีค่ะ รูปแบบสีโพงดังเสียงชัดค่ะ	2	2	3
ถูกใจ ส่งไว บริการดี สินค้าไม่มีตำหนิ สินค้าตรงปก ราคาไม่แพง สินค้ามีคุณภาพดี ส่งไว คนส่งพูดจาดี โทรหาพูดจาดีมาก มาก โดยรวมถือว่าโอเค ประทับใจมาก	3	3	3

ตัวอย่างการทำ Label data