HACKATHON CHALLENGE · DATA SCIENCE

เปลี่ยนกลิ่นให้เป็นข้อมูล

คุณคือทีมข้อมูลของหน้างานที่ติดตั้งสถานีตรวจวัดกลิ่น และมีข้อมูลจากเซ็นเซอร์ต่อเนื่องสามเดือนอยู่ในมือ ภารกิจของคุณคือเปลี่ยนข้อมูลนี้ให้เป็นสิ่งที่ช่วยหน้างานจัดการกลิ่นได้จริง พร้อมสื่อสารผลได้อย่างน่าเชื่อถือ

สถานการณ์ SCENARIO

หน้างานแห่งหนึ่ง เช่น โรงงานหรือโรงบำบัดน้ำเสีย ติดตั้งสถานีตรวจวัดกลิ่นในอากาศตลอดเวลา ระยะหลังชุมชนรอบข้างร้องเรียนเรื่องกลิ่นรบกวนบ่อยขึ้น ผู้จัดการไซต์อยากได้เครื่องมือที่ช่วยให้ทีมเข้าใจสถานการณ์กลิ่น ตอบข้อร้องเรียนอย่างมีหลักฐาน และตัดสินใจได้เร็วขึ้นว่าควรลงไปแก้ที่จุดใดเมื่อใด เขาจึงมอบข้อมูลที่เครื่องบันทึกไว้ให้คุณ และขอให้คุณใช้กระบวนการวิทยาศาสตร์ข้อมูลมาช่วยแก้ปัญหานี้

ข้อมูลที่ได้รับ DATASET

ข้อมูลมาจากสถานีตรวจวัดกลิ่น MUI-NOSE ของบริษัท MUI Robotics ซึ่งใช้หลักการ Artificial Olfaction ร่วมกับ Machine Learning ในการตรวจวัดและจำแนกกลิ่น เครื่องบันทึกค่าทุกหนึ่งนาที ต่อเนื่องประมาณ 92 วัน รวมราว 130,000 แถว ข้อมูลนี้เป็นข้อมูลจริงที่ยังไม่ผ่านการทำความสะอาด การสำรวจและทำความเข้าใจข้อมูลก่อนการวิเคราะห์หรือพัฒนาแบบจำลองจึงเป็นขั้นตอนที่สำคัญ

รายละเอียดจากสเปกของเครื่องที่เป็นประโยชน์ต่อการทำงาน ค่าความเข้มกลิ่น D/T คือจำนวนเท่าที่ต้องเจือจางอากาศที่มีกลิ่นด้วยอากาศสะอาด จนกลิ่นจางลงถึงระดับที่คนเริ่มได้กลิ่นพอดี ค่ายิ่งสูงแปลว่ากลิ่นยิ่งแรง มีพิสัยตั้งแต่ 0 ถึง 60 ความละเอียด 0.1 ชุดเซ็นเซอร์ตรวจกลิ่นมีไม่น้อยกว่าแปดชนิดทำงานร่วมกัน และเครื่องจำแนกกลิ่นได้สูงสุดห้าโปรไฟล์ พิสัยที่ถูกต้องของเซ็นเซอร์แต่ละตัวตามสเปกแสดงไว้ในคอลัมน์ขวาของตาราง ซึ่งใช้ตรวจสอบได้ว่าค่าใดผิดปกติ

คอลัมน์	ความหมาย	พิสัยตามสเปก
`Time`	เวลาที่บันทึกข้อมูล	บันทึกทุก 1 นาที
`D/T`	ระดับความเข้มของกลิ่น ค่าสูงหมายถึงกลิ่นแรง	0 ถึง 60 ละเอียด 0.1
`Wind Direction`	ทิศทางลม	0 ถึง 360 องศา
`Wind Speed`	ความเร็วลม	0 ถึง 40 เมตรต่อวินาที
`Temperature`	อุณหภูมิ	-40 ถึง 85 องศาเซลเซียส
`Relative Humidity`	ความชื้นสัมพัทธ์	0 ถึง 100 เปอร์เซ็นต์
`PM 2.5`	ปริมาณฝุ่นละอองขนาดเล็ก	0 ถึง 1000 ไมโครกรัมต่อลูกบาศก์เมตร
`Atmospheric Pressure`	ความดันอากาศ	500 ถึง 1100 เฮกโตปาสกาล
`Sensor 1` ถึง `Sensor 8`	ค่าจากเซ็นเซอร์แก๊สแปดตัว เครื่องใช้รูปแบบรวมของทุกเซ็นเซอร์ในการจำแนกกลิ่น ไม่ได้ระบุชนิดก๊าซของแต่ละตัว	สเปกไม่ระบุรายตัว
`Smell Prediction`	ผลการจำแนกกลิ่นที่เครื่องทำนายออกมา	จำแนกได้สูงสุด 5 โปรไฟล์

โจทย์ CHALLENGE

จากข้อมูลที่ได้ ให้ใช้หลักการ Data Science พัฒนาแนวทางแก้ปัญหาที่เป็นประโยชน์ต่อหน้างาน โดยทำให้ครบกระบวนการ พร้อมอธิบายให้เห็นว่าผลที่ได้มีความหมายต่อหน้างานอย่างไร โจทย์นี้ไม่มีคำตอบเดียวที่ถูกต้อง ทีมมีอิสระในการเลือกโจทย์หรือปัญหา

ตัวอย่างทิศทางที่เป็นไปได้ เช่น การพยากรณ์ความเข้มกลิ่นล่วงหน้าเพื่อเตือนชุมชน การหาช่วงเวลาที่กลิ่นแรงที่สุดในแต่ละวัน การประเมินความเข้มกลิ่นจากค่าของเซ็นเซอร์ การตรวจจับเหตุการณ์ผิดปกติ หรือการดูความสัมพันธ์ระหว่างกลิ่นกับสภาพอากาศ ทั้งหมดนี้เป็นเพียงตัวอย่าง ทีมเลือกมุมอื่นหรือผสมหลายมุมได้

สิ่งที่ควรให้ความสำคัญคือความครบถ้วนของกระบวนการ และคุณภาพของการเลือกใช้วิธีในการแก้ปัญหาของแต่ละกระบวนการ

การใช้ AI AI ALLOWED

โจทย์นี้อนุญาตและสนับสนุนให้ใช้เครื่องมือ AI ได้เต็มที่ เพราะการทำงานร่วมกับ AI เป็นทักษะจริงในยุคปัจจุบัน มีเงื่อนไขเพียงข้อเดียวคือต้องเข้าใจทุกขั้นตอนของงานตัวเอง และอธิบายได้ว่าทำไมจึงทำเช่นนั้น การส่งงานที่ไม่สามารถอธิบายเหตุและผลของวิธีการต่างๆ ถือว่าเป็นขั้นตอนการปฏิบัติที่ไม่ได้มาตรฐาน

กระบวนการทำงาน PROCESS

ขอให้ทุกทีมทำงานตามกระบวนการด้านวิทยาศาสตร์ข้อมูลที่เป็นระบบ ซึ่งมีหลายแบบให้เลือกใช้ เช่น กระบวนการวิทยาศาสตร์ข้อมูลทั่วไป หรือ CRISP-DM เป็นต้น ขอเพียงให้กระบวนการครบถ้วนตั้งแต่การทำความเข้าใจปัญหาไปจนถึงการสรุปผล และในแต่ละช่วงให้อธิบายได้ว่าพบสิ่งใดและเพราะเหตุใด