ความรู้และความเข้าใจข้อมูลสำหรับการวิเคราะห์ข้อมูล

(Data Literacy for Data Analytics)

นางสาวกานต์มณี จันทร์ขาว นักวิชาการเกษตรชำนาญการ กองเทคโนโลยีชีวภาพทางดิน

ที่มา: อบรมผ่านระบบออนไลน์ วันที่ 3 สิงหาคม 2569 จัดโดย Digital Government Learning Platform

วิทยากร ผศ.ดร. โษฑศ์รัตต ธรรมบุษดี คณะวิศวกรรมศาสตร์ มหาวิทยาลัยมหิดล

ความรู้และความเข้าใจข้อมูลสำหรับการวิเคราะห์ข้อมูล (Data Literacy for Data Analytics) หมายถึง ความรู้และความเข้าใจข้อมูล (Data Literacy) คือ ทักษะการอ่าน การตีความ การใช้งาน และการสื่อสารด้วยข้อมูลอย่างมีประสิทธิภาพ เพื่อนำไปใช้ในการวิเคราะห์ข้อมูล (Data Analytics) ประกอบการตัดสินใจ และแก้ไขปัญหาได้อย่างแม่นยำ สามารถนำข้อมูลมาตัดสินใจอย่างมีประสิทธิภาพ และใช้ข้อมูลเป็นพื้นฐานในการทำงานได้อย่างถูกต้อง โดยมีเนื้อหาดังนี้

1. ข้อมูลและรูปแบบของข้อมูล (Data and Types of Data) ข้อมูล (data) คือ สัญญาณดิบที่ยังไม่มีความหมาย เมื่อนำมาประมวลผลจะกลายเป็นสารสนเทศ (information) ความรู้ (knowledge) ภูมิปัญญา (wisdom) และนำไปสู่การตัดสินใจ (decisions) ประเภทของข้อมูลแบ่งเป็น 2 กลุ่มหลัก ได้แก่

1.1 ข้อมูลเชิงคุณภาพ (Qualitative) คือข้อมูลที่ใช้บอก ลักษณะ ประเภท กลุ่ม หรือคุณสมบัติของสิ่งที่ศึกษา มากกว่าการบอกเป็นค่าปริมาณโดยตรง เช่น เพศ ระดับความพึงพอใจ โดยทั่วไปแบ่งได้เป็น 2 ประเภทหลัก ได้แก่

1.1.1 Nominal data (ข้อมูลนามบัญญัติ) เป็นข้อมูลที่ใช้จำแนกสิ่งต่าง ๆ ออกเป็นกลุ่มหรือประเภท โดย ไม่มีลำดับก่อน–หลัง หรือสูง–ต่ำ เช่น เพศ (ชาย/หญิง) ชนิดพืช (ข้าว/มันสำปะหลัง/อ้อย) ข้อมูลชนิดนี้แม้จะกำหนดตัวเลขแทนกลุ่ม เช่น 1 =ชาย , 2 = หญิง ตัวเลขดังกล่าวเป็นเพียง “รหัส” ไม่สามารถนำมาคำนวณค่าเฉลี่ยหรือบอกว่า 2 มากกว่า 1 ในเชิงปริมาณได้

1.1.2 Ordinal data (ข้อมูลอันดับ) เป็นข้อมูลเชิงคุณภาพที่สามารถ เรียงลำดับระดับหรือจัดอันดับได้ แต่ระยะห่างระหว่างแต่ละระดับไม่จำเป็นต้องเท่ากัน เช่น ระดับความพึงพอใจ (น้อย/ปานกลาง/มาก), ระดับความอุดมสมบูรณ์ของดิน (ต่ำ/ปานกลาง/สูง), ระดับความรุนแรงของโรค (น้อย/ปานกลาง/รุนแรง) หรือระดับการประเมินแบบ Likert scale เช่น 1 = ไม่เห็นด้วยอย่างยิ่ง ถึง 5 = เห็นด้วยอย่างยิ่ง แม้จะเรียงลำดับได้ แต่ไม่ควรตีความว่าความแตกต่างระหว่างระดับ 1 กับ 2 เท่ากับความแตกต่างระหว่างระดับ 4 กับ 5 เสมอไป

1.2 ข้อมูลเชิงปริมาณ (Quantitative) หมายถึง ข้อมูลที่สามารถแสดงออกมาเป็น ตัวเลขและสามารถนำไปวัด คำนวณ เปรียบเทียบ และวิเคราะห์ทางสถิติได้ เช่น ความสูง น้ำหนัก อุณหภูมิ ข้อมูลประเภทนี้เหมาะสำหรับการวิเคราะห์ทางสถิติ เช่น ค่าเฉลี่ย ค่ามัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน สหสัมพันธ์ การถดถอย และการทดสอบสมมติฐาน สามารถแบ่งออกเป็น 2 ประเภท ได้แก่

1.2.1 Discrete data (ข้อมูลไม่ต่อเนื่อง) เป็นข้อมูลเชิงปริมาณที่มีค่าเป็นจำนวนที่นับได้ เป็นหน่วยๆ โดยทั่วไปมักเป็นจำนวนเต็ม และไม่สามารถมีค่าอยู่ระหว่างหน่วยได้ เช่น จำนวนคน จำนวนต้นไม้ เป็นต้น

1.2.2 Continuous data (ข้อมูลต่อเนื่อง) เป็นข้อมูลที่ได้จากการ วัด และสามารถมีค่าเป็นทศนิยมได้ โดยสามารถมีค่าใด ๆ ภายในช่วงที่กำหนด เช่น ความสูงของต้นไม้ 15.25 เมตร อุณหภูมิ 28.6 ℃ ดังนั้นข้อมูลที่เกิดจากการ “วัด” มักจัดเป็น Continuous data

1.3 การแบ่งข้อมูลตามลักษณะการจัดเก็บนอกจากการแบ่งข้อมูลตามลักษณะของข้อมูลแล้ว ข้อมูลยังสามารถแบ่งตาม รูปแบบและโครงสร้างของการจัดเก็บ ได้เป็น 2 ประเภท ได้แก่

1.3.1 Structured Data (ข้อมูลแบบมีโครงสร้าง)หมายถึง ข้อมูลที่มีการจัดระเบียบและกำหนดโครงสร้างไว้อย่างชัดเจน เช่น มีแถวและคอลัมน์ สามารถจัดเก็บในฐานข้อมูล ตาราง Excel หรือระบบฐานข้อมูลต่างๆ ข้อมูลสามารถค้นหา จัดเรียง คำนวณ และนำไปวิเคราะห์ด้วยโปรแกรมทางสถิติได้

1.3.2 Unstructured Data (ข้อมูลแบบไม่มีโครงสร้าง) หมายถึง ข้อมูลที่ไม่ได้จัดอยู่ในรูปแบบตารางหรือโครงสร้างที่ตายตัว ทำให้การจัดเก็บ ประมวลผล และวิเคราะห์ข้อมูลต้องใช้เทคนิคหรือเครื่องมือเฉพาะ ตัวอย่างเช่น รูปภาพ ภาพถ่าย วิดีโอ ไฟล์เสียง เอกสาร PDF ข้อความจากสื่อสังคมออนไลน์ และภาพจากกล้องจุลทรรศน์ ข้อมูลเหล่านี้อาจมีข้อมูลที่มีคุณค่าอยู่มาก แต่ไม่สามารถนำไปวิเคราะห์ด้วยวิธีเดียวกับข้อมูลตารางได้โดยตรง จำเป็นต้องใช้เทคโนโลยี เช่น Image Processing, Natural Language Processing (NLP), Machine Learning หรือ Artificial Intelligence เพื่อดึงข้อมูลที่สำคัญออกมา

2. จริยธรรมด้านข้อมูล หมายถึง หลักการและแนวทางในการเก็บรวบรวม วิเคราะห์ จัดเก็บ และแบ่งปันข้อมูลอย่าง ถูกต้อง เหมาะสม เป็นธรรม โปร่งใส และคำนึงถึงสิทธิของเจ้าของข้อมูล โดยเฉพาะเมื่อข้อมูลนั้นเกี่ยวข้องกับบุคคลหรืออาจส่งผลกระทบต่อบุคคลและสังคม การมีข้อมูลจำนวนมากไม่ได้หมายความว่าสามารถนำข้อมูลเหล่านั้นไปใช้ได้ทุกกรณี ผู้ใช้ข้อมูลจำเป็นต้องพิจารณาว่า ข้อมูลมาจากไหน เก็บมาอย่างไร ใช้เพื่ออะไร ใครสามารถเข้าถึงได้ และการใช้ข้อมูลจะส่งผลกระทบต่อใครบ้าง ดังนั้น จริยธรรมด้านข้อมูลจึงเป็นกรอบที่ช่วยให้การใช้ข้อมูลเกิดประโยชน์ โดยลดความเสี่ยงต่อการละเมิดสิทธิ ความเป็นส่วนตัว และการนำข้อมูลไปใช้ในทางที่ไม่เหมาะสม ดังนั้นการเข้าถึงข้อมูลจึงครอบคลุมมิติต่างๆ เช่น แหล่งข้อมูล ข้อจำกัด การแบ่งปัน บริบทด้านกฎหมายและจริยธรรม ผลกระทบเชิงบวกและลบต่อผู้คน ความโปร่งใส และแผนการดำเนินงาน สิทธิของเจ้าของข้อมูลส่วนบุคคลตามกฎหมาย PDPA ประกอบด้วย 9 สิทธิ ได้แก่ สิทธิในการเข้าถึง แจ้งข้อมูล คัดค้าน ลบ/ทำลาย ระงับ แก้ไข โอนถ่าย ถอนความยินยอม และร้องเรียน

3. การเก็บรวบรวมข้อมูล เป็นกระบวนการรวบรวมข้อเท็จจริง ตัวเลข ข้อความ หรือสารสนเทศจากแหล่งต่างๆ อย่างเป็นระบบ เพื่อนำข้อมูลไปใช้ในการวิเคราะห์ การตัดสินใจ การวางแผน หรือการแก้ไขปัญหา การเก็บข้อมูลที่ดีเริ่มจากการกำหนดวัตถุประสงค์อย่างชัดเจน ดังนั้นการเก็บข้อมูลไม่ได้หมายถึงเพียงการรวบรวมข้อมูลให้ได้จำนวนมาก แต่ต้องให้ความสำคัญกับ คุณภาพ ความถูกต้อง ความน่าเชื่อถือ ความครบถ้วน และความเหมาะสมของข้อมูล โดยขั้นตอนสำคัญในกระบวนการเก็บรวบรวมข้อมูล ได้แก่ การกำหนดวัตถุประสงค์ธุรกิจ แหล่งข้อมูล ประเภทข้อมูล วิธีการเก็บรวบรวม การตรวจสอบสารสนเทศ และกำหนดกรอบเวลา ความถูกต้องของข้อมูล (data validity) ต้องพิจารณา 5 มิติ ได้แก่ ด้านการวัด (measurement) ด้านสารสนเทศ (information) ด้านการจำแนกความแตกต่าง (discriminate different) ด้านการพยากรณ์ (predict) และ ด้านความสอดคล้อง (agreement) เพื่อให้ข้อมูลสามารถนำไปใช้วิเคราะห์และสนับสนุนการตัดสินใจได้อย่างมีประสิทธิภาพ

4. การบริหารจัดการและธรรมาภิบาลข้อมูล เป็นแนวคิดสำคัญในการจัดการข้อมูลขององค์กรให้สามารถนำไปใช้ได้อย่างมีประสิทธิภาพ มีคุณภาพ ปลอดภัย และสอดคล้องกับกฎหมายและจริยธรรม โดยทั้งสองแนวคิดมีความเกี่ยวข้องกัน มีหน้าที่แตกต่างกัน ครอบคลุมข้อมูล 6 ขั้นตอน การสร้าง จัดเก็บ นำไปใช้ เผยแพร่ จัดเก็บระยะยาว และการทำลายข้อมูล โดยมีฐานสำคัญ ได้แก่ สารานุกรมข้อมูล (data catalog) ที่ช่วยให้ค้นหาและเข้าถึงข้อมูลได้ง่าย ความมั่นคงปลอดภัยและความเป็นส่วนตัวที่รองรับกฎหมาย เช่น PDPA รวมถึงการดูแลคุณภาพข้อมูลให้ถูกต้องและเชื่อถือได้ ทั้งนี้ การบริหารข้อมูลที่ดียังต้องยึดหลักธรรมาภิบาล 7 ประการ ได้แก่ ความถูกต้องแม่นยำ ความครบถ้วนสมบูรณ์ ความทันสมัย ความมั่นคงปลอดภัย ความเป็นส่วนตัว ความเชื่อมโยงระหว่างระบบ และความคุ้มค่า เพื่อให้ข้อมูลสามารถสร้างประโยชน์สูงสุดต่อองค์กรและสนับสนุนการตัดสินใจได้อย่างมีประสิทธิภาพ

5. การวิเคราะห์ข้อมูลและการตัดสินใจ เป็นกระบวนการนำข้อมูลที่เก็บรวบรวมและผ่านการตรวจสอบคุณภาพแล้ว มาจัดระเบียบ ประมวลผล วิเคราะห์ และตีความ เพื่อค้นหา รูปแบบ แนวโน้ม ความสัมพันธ์ ความแตกต่าง หรือสารสนเทศที่มีความหมาย แล้วนำผลลัพธ์ที่ได้ไปใช้ประกอบการตัดสินใจ โดยใช้เทคนิคการวิเคราะห์ 5 ระดับ ได้แก่

5.1 การวิเคราะห์เชิงพรรณนา (อธิบายสิ่งที่เกิดขึ้น) การสรุปลักษณะของข้อมูลที่มีอยู่ เช่น ค่าเฉลี่ย ร้อยละ จำนวน หรือแนวโน้ม

5.2 เชิงวินิจฉัย (หาสาเหตุ) การค้นหาสาเหตุหรือปัจจัยที่เกี่ยวข้องกับผลลัพธ์ที่เกิดขึ้น

5.3 เชิงพยากรณ์ (คาดการณ์อนาคต) การใช้ข้อมูลในอดีตและปัจจุบันเพื่อสร้างแบบจำลองสำหรับคาดการณ์แนวโน้มในอนาคต

5.4 เชิงสั่งการ (เสนอแนวทางดำเนินงาน) การนำผลจากการวิเคราะห์มาพิจารณาทางเลือกและผลกระทบ เพื่อสนับสนุนการเลือกแนวทางดำเนินงาน

5.5 การวิเคราะห์ด้วย AI สามารถเสริมหรือทดแทนการตัดสินใจของมนุษย์ได้ในบางกรณี ทั้งนี้ กระบวนการตัดสินใจโดยใช้ข้อมูล (data-driven decision making) ประกอบด้วย 4 ขั้นตอน คือ การกำหนดปัญหา การวิเคราะห์ข้อมูล การตีความผลลัพธ์ และการสื่อสารผล ซึ่งสอดคล้องกับวงจรการตัดสินใจที่ครอบคลุมการกำหนดเกณฑ์ การระบุความต้องการข้อมูล และการประเมินผล เพื่อให้การตัดสินใจมีประสิทธิภาพและอ้างอิงข้อมูลที่น่าเชื่อถือ