

ความรู้และความเข้าใจข้อมูลสำหรับการวิเคราะห์ข้อมูล (Data Literacy for Data Analytics)

โดย นางสาวสุจิตรา ก้อนเรณู
นักวิชาการแผนกที่ภาพถ่ายปฏิบัติการ

วัตถุประสงค์

- เพื่อให้ผู้เรียนเข้าใจความหมายและเห็นความสำคัญของทักษะความสามารถพื้นฐานในเชิงข้อมูล
- เพื่อให้ผู้เรียนมีความรู้ ความเข้าใจ และการเชื่อมโยงของทักษะความสามารถพื้นฐานในเชิงข้อมูล และการวิเคราะห์ข้อมูลได้
- เพื่อให้ผู้เรียนสามารถเข้าใจการทำงาน สร้างความคุ้นเคย และสามารถนำข้อมูลไปใช้ในการทำงานได้อย่างถูกต้อง

สรุปเนื้อหา

๑. ความสำคัญของข้อมูลและรูปแบบของข้อมูล (Data and Types of Data)

ข้อมูล (Data) ถือเป็นทรัพยากรที่มีคุณค่าอย่างยิ่งขององค์กร ไม่ว่าจะอยู่ในรูปแบบกระดาษหรือดิจิทัล เนื่องจากข้อมูลสามารถสะท้อนจุดแข็ง จุดอ่อน และสถานะต่าง ๆ ขององค์กรได้ หากนำข้อมูลมาจัดการอย่างเป็นระบบ ข้อมูลจะมีการพัฒนาไปสู่ระดับที่สูงขึ้น โดยข้อมูลดิบจะถูกแปลงเป็นสารสนเทศ องค์ความรู้ ปัญญา และนำไปสู่การตัดสินใจ ซึ่งมีลำดับขั้นของการแปลงข้อมูล คือ Data → Information → Knowledge → Wisdom → Decision สามารถอธิบายได้ ดังนี้

- ข้อมูล (Data) ข้อเท็จจริงดิบที่ยังไม่ผ่านการประมวลผล
- สารสนเทศเบื้องต้น (Information) ข้อมูลที่ถูกจัดระเบียบ ตรวจสอบความถูกต้อง และจัดให้อยู่ในรูปแบบที่เข้าใจได้
- องค์ความรู้ (Knowledge) ข้อมูลและสารสนเทศที่ถูกวิเคราะห์จนเกิดความเข้าใจ สามารถบ่งชี้รูปแบบหรือความผิดปกติ
- ปัญญา (Wisdom) การประยุกต์ใช้ความรู้เพื่อการตัดสินใจอย่างเหมาะสม
- การตัดสินใจ (Decisions) ผลลัพธ์สุดท้ายที่เกิดจากการใช้ข้อมูล สารสนเทศ และองค์ความรู้ร่วมกัน

วงจรชีวิตของข้อมูล คือกระบวนการตั้งแต่การสร้างหรือเก็บรวบรวมข้อมูล ไปจนถึงการนำไปใช้ และทำลายข้อมูล สามารถแบ่งเป็น ๖ ขั้นตอนหลัก ได้แก่ ๑)การสร้างข้อมูล ๒)การจัดเก็บข้อมูล ๓)การประมวลผลและใช้ข้อมูล ๔)การเผยแพร่ข้อมูล ๕)การจัดเก็บข้อมูลถาวร ๖)การทำลายข้อมูล

มูลค่าของข้อมูล (Data Valuation) ข้อมูลถือเป็นสินทรัพย์ที่สร้างมูลค่าได้ในสองมิติ ประกอบด้วยต้นทุน (Cost) หมายถึงค่าใช้จ่ายในการจัดเก็บ ดูแล และบริหารจัดการข้อมูล และรายได้ (Revenue/Benefit) หมายถึงมูลค่าที่เกิดจากการนำข้อมูลไปใช้ตัดสินใจหรือสร้างความได้เปรียบเชิงการแข่งขัน

รูปแบบของข้อมูล (Data Types) ได้แก่ ๑)ข้อมูลที่ไม่มีโครงสร้าง (Unstructured Data) เช่น ข้อความยาว ๆ ไฟล์เอกสาร วิดีโอ หรือภาพถ่าย มีความยืดหยุ่นสูง แต่ยากต่อการวิเคราะห์ ๒)ข้อมูลที่มีโครงสร้าง (Structured Data) เช่น ข้อมูลตารางใน Excel หรือฐานข้อมูลเชิงสัมพันธ์ (Relational Database) ง่ายต่อการประมวลผล แต่ยืดหยุ่นน้อย และ ๓)ข้อมูลกึ่งมีโครงสร้าง (Semi-structured Data) เช่น XML, JSON มีโครงสร้างกำกับบางส่วน ทำให้จัดการง่ายขึ้นแต่ยังคงความยืดหยุ่น

การเก็บข้อมูล (Data Storage) แบ่งออกได้เป็น ๒ ประเภทหลัก คือ ฐานข้อมูล (Database / Relational File) จัดเก็บข้อมูลในรูปแบบตาราง เชื่อมโยงกันด้วยคีย์ ใช้ในระบบสารสนเทศขนาดใหญ่ และไฟล์ (Flat File) เช่น Excel, CSV ข้อดีคือใช้งานง่ายและสามารถนำข้อมูลไปใช้ได้ทันที แต่ไม่เหมาะกับข้อมูลปริมาณมากหรือซับซ้อน

๒. จริยธรรมด้านข้อมูล (Data Ethics)

การใช้ข้อมูลต้องคำนึงถึงสิทธิส่วนบุคคล ความโปร่งใส ความปลอดภัย และการไม่บิดเบือนข้อมูล เพื่อสร้างความน่าเชื่อถือและความยั่งยืนในการใช้ข้อมูล สิ่งสำคัญคือต้องระบุวัตถุประสงค์ของการเก็บข้อมูล นั้นๆ ด้วย โดยหลักการจัดการข้อมูลในระดับสากลอ้างอิงตามแนวทางของ OECD ขณะที่ประเทศไทยกำหนดกรอบกฎหมายที่เกี่ยวข้อง ได้แก่ พระราชบัญญัติข้อมูลข่าวสารของราชการ พ.ศ. ๒๕๔๐ และพระราชบัญญัติคุ้มครองข้อมูลส่วนบุคคล พ.ศ. ๒๕๖๒ (PDPA)

๓. การเก็บรวบรวมข้อมูล (Data Collection)

การเก็บรวบรวมข้อมูล (Data Collection) หมายถึง กระบวนการรวบรวมข้อเท็จจริง ตัวเลข หรือหลักฐานจากแหล่งต่าง ๆ เพื่อนำมาใช้ในการวิเคราะห์ ประเมินผล และสนับสนุนการตัดสินใจ โดยวิธีการเก็บรวบรวม มีดังนี้ ๑)แบบสอบถาม (Questionnaire) เป็นวิธีการเก็บข้อมูลที่สะดวก แต่ต้องทราบว่ากลุ่มตัวอย่างคือใคร และคำถามต้องละเมิดความเป็นส่วนตัวให้น้อยสุด ๒)แบบสัมภาษณ์ (Interview) เป็นวิธีที่ตรงไปตรงมามากที่สุด ข้อดีคือได้ข้อมูลเชิงลึก ข้อเสียคือใช้เวลานาน ๓)การสังเกตการณ์ (Observation) สังเกตวิธีการทำงาน หรือดูว่าเขาทำงานกันอย่างไร เช่น การลงสำรวจชุมชน ซึ่งต้องวางตัวเป็นกลางว่าสิ่งที่สังเกตได้คือเรื่องจริง เป็นต้น

๔. การบริหารจัดการข้อมูล (Data Management)

การบริหารจัดการข้อมูล (Data Management) คือ กระบวนการวางแผน จัดเก็บ ใช้งาน ควบคุม ปกป้อง และพัฒนาข้อมูลในองค์กรอย่างเป็นระบบ เพื่อให้ข้อมูลมีคุณภาพ น่าเชื่อถือ และพร้อมใช้งานสำหรับการวิเคราะห์และการตัดสินใจ โดยถือเป็นรากฐานสำคัญของการทำงานด้าน Data Analytics, Machine Learning, และ Digital Transformation

๕. ธรรมาภิบาลข้อมูล (Data Governance)

ธรรมาภิบาลข้อมูล (Data Governance) คือ กรอบแนวทาง (Framework) ที่กำหนด นโยบาย มาตรฐาน บทบาทหน้าที่ และกระบวนการ ที่เกี่ยวข้องกับการจัดการข้อมูล เพื่อให้มั่นใจว่าข้อมูลมีคุณภาพ มีความถูกต้อง ปลอดภัย โปร่งใส และสามารถนำไปใช้ประโยชน์ได้อย่างมีประสิทธิภาพ โดยมีผลลัพธ์ คือ ต้องการข้อมูลที่มีคุณภาพ ทะเบียนบัญชีรายชื่อข้อมูลภาครัฐ การรักษาความปลอดภัยและคุ้มครองข้อมูล การเปิดเผยข้อมูลภาครัฐ

๖. การเตรียมข้อมูลเพื่อใช้งาน (Data Provisioning)

ข้อมูลสามารถจัดประเภทเพื่อความเข้าใจและการวิเคราะห์ได้ ๒ ประเภท ได้แก่ ๑)ข้อมูลเชิงคุณภาพ (Qualitative Data) เป็นข้อมูลที่ไม่ใช่ตัวเลข เช่น สีเขียว สีเหลือง แบ่งย่อยเป็น Nominal คือข้อมูลที่ไม่สามารถเรียงลำดับได้ Ordinal คือข้อมูลที่ไม่ใช่ตัวเลขแต่เรียงลำดับได้ เช่น อุณหภูมิ ร้อนอุ่น หนาว และ ๒)ข้อมูลเชิงปริมาณ (Quantitative Data) คือข้อมูลตัวเลข ได้แก่ จำนวนเต็ม เช่น จำนวนนักเรียนในห้องเรียน และเลขทศนิยม เช่น ส่วนสูง น้ำหนัก

๗. การวิเคราะห์ข้อมูลเบื้องต้น (Data Analysis)

เป็นกระบวนการตรวจสอบ ทำความเข้าใจ และสรุปข้อมูลดิบเพื่อนำไปสู่สารสนเทศ (Information) และองค์ความรู้ (Knowledge) ที่สามารถสนับสนุนการตัดสินใจและการวางแผนได้อย่างมีประสิทธิภาพ มีรูปแบบของการวิเคราะห์ข้อมูล ดังนี้

๑) การวิเคราะห์เชิงพรรณนา (Descriptive Analysis) เป็นการสรุปและบรรยายข้อมูลปัจจุบันหรืออดีตเพื่อให้เข้าใจสถานการณ์ เพื่อให้เห็นภาพรวมของข้อมูล เช่น แนวโน้ม ความถี่ ค่าเฉลี่ย และรูปแบบเบื้องต้น

๒) การวิเคราะห์เชิงวินิจฉัย (Diagnostic Analysis) เป็นการวิเคราะห์เพื่อหาสาเหตุหรือปัจจัยที่ส่งผลต่อเหตุการณ์ที่เกิดขึ้น เพื่อให้เข้าใจว่า “ทำไม” สิ่งนั้นเกิดขึ้น

๓) การวิเคราะห์เชิงคาดการณ์ (Predictive Analysis) เป็นการใช้ข้อมูลอดีตและเทคนิคทางสถิติ/ Machine Learning (ML) เพื่อคาดการณ์เหตุการณ์ในอนาคต เพื่อตัดสินใจเชิงกลยุทธ์โดยอิงการคาดการณ์

๔) การวิเคราะห์เชิงเสนอแนะ (Prescriptive Analysis) เป็นการวิเคราะห์เพื่อแนะนำทางเลือกหรือแนวทางการตัดสินใจ เพื่อให้คำตอบว่า “ควรทำอย่างไร”

หลักการนำเสนอข้อมูลหรือการสื่อสารด้วยข้อมูล

๑) รู้จักผู้ฟัง วิเคราะห์ระดับความรู้ของผู้ฟัง สามารถจัดกลุ่มผู้ฟัง เช่น Executives / Business Managers / Analysts เพื่อปรับความซับซ้อนของเนื้อหา

๒) วัตถุประสงค์และรูปแบบการนำเสนอ ชัดเจนว่าอยากให้ผู้ฟัง ตัดสินใจเรื่องอะไร เลือกการจัดวางกราฟ สี และชนิดกราฟให้เหมาะสม

๓) เน้นจุดสำคัญ สื่อสารว่าควรให้ความสนใจกราฟ/ตัวเลขใดเป็นหลัก

๔) โครงสร้างการนำเสนอแบบ Top-Down เริ่มจากภาพรวม (ภาพใหญ่) → รายละเอียด (ภาพเล็ก) ใช้การสื่อสารแบบสองทาง (๒-way communication) เพื่อสอบถามและรับฟังความคิดเห็นผู้ฟัง

Machine Learning (ML) คือ สาขาหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence: AI) ที่ทำให้คอมพิวเตอร์สามารถเรียนรู้จากข้อมูล (Data) และปรับปรุงประสิทธิภาพการทำงานโดยไม่ต้องเขียนโปรแกรมแบบกำหนดขั้นตอนตายตัว แบ่งเป็น ๓ ประเภท ได้แก่

๑) Supervised Learning (การเรียนรู้แบบมีผู้สอน) ใช้ข้อมูลที่มี Label หรือ “คำตอบ” อยู่แล้ว เช่น การทำนายราคาบ้านจากข้อมูลทำเล ขนาด และสิ่งอำนวยความสะดวก ตัวอย่างอัลกอริทึม เช่น Linear Regression, Decision Trees, Support Vector Machines (SVM)

๒) Unsupervised Learning (การเรียนรู้แบบไม่มีผู้สอน) ใช้ข้อมูลที่ไม่ได้มี Label เพื่อค้นหารูปแบบหรือโครงสร้าง เช่น การจัดกลุ่มลูกค้าตามพฤติกรรมการซื้อ ตัวอย่างอัลกอริทึม เช่น K-Means Clustering

๓) Reinforcement Learning (การเรียนรู้แบบเสริมแรง) เครื่องเรียนรู้จาก การลองผิดลองถูก และ การได้รับรางวัล (Reward) เช่น หุ่นยนต์เรียนรู้การเดิน, AI เล่นเกมชกมนุษย์ ตัวอย่าง เช่น Q-Learning, Deep Reinforcement Learning

๘. หลักเกณฑ์การนำข้อมูลมาช่วยตัดสินใจ

๑) Decision Criteria: การกำหนดเกณฑ์การตัดสินใจ การวิเคราะห์ข้อมูลเพื่อการตัดสินใจต้องเริ่มจากการระบุ “เกณฑ์” ที่ใช้ตัดสินใจ เช่น ผลกระทบต่อเป้าหมายองค์กร ความคุ้มค่าเชิงเศรษฐกิจ ความเสี่ยง และความเป็นไปได้ เป็นต้น ซึ่งการมีเกณฑ์ที่ชัดเจน จะช่วยให้สามารถใช้ข้อมูลมาประเมินได้ตรงตามประเด็นสำคัญและไม่หลงประเด็น

๒) Data Requirement: การระบุความต้องการข้อมูล หลังจากกำหนดเกณฑ์การตัดสินใจแล้ว ต้องพิจารณาว่าต้องใช้ข้อมูลใดบ้างเพื่อสนับสนุนการประเมิน เช่น ชนิดของข้อมูลที่จำเป็น แหล่งข้อมูลที่เชื่อถือได้ ช่วยให้ได้อ้างอิงข้อมูลที่สอดคล้องกับเกณฑ์การตัดสินใจ ลดการใช้ข้อมูลที่ไม่เกี่ยวข้อง

๓) Decision Evaluation: การประเมินทางเลือก ซึ่งจะช่วยให้องค์กรเลือกทางเลือกที่เหมาะสมที่สุด ลดความเสี่ยง และเพิ่มโอกาสความสำเร็จของการตัดสินใจ

การนำองค์ความรู้ไปปรับใช้ในการปฏิบัติงาน

ช่วยให้เข้าใจข้อมูล จัดการ วิเคราะห์ สื่อสาร และนำไปใช้ตัดสินใจอย่างมีประสิทธิภาพ ลดข้อผิดพลาดของงานได้

ประโยชน์ที่ได้รับ

- ต่อตนเอง: เพิ่มทักษะและองค์ความรู้ เพิ่มประสิทธิภาพการทำงาน สามารถตัดสินใจแม่นยำขึ้น ด้วยการใช้ข้อมูลสนับสนุนการตัดสินใจ ลดความเสี่ยงจากความเชื่อหรือความรู้สึกส่วนตัว
- ต่อองค์กร: ลดข้อผิดพลาดและเพิ่มประสิทธิภาพของงาน และมีข้อมูลชัดเจนสำหรับวางแผน และพัฒนาองค์กร เพื่อสนับสนุนการตัดสินใจเชิงกลยุทธ์
- ต่อสาธารณะ: มีข้อมูลที่ชัดเจน ช่วยให้ประชาชนเข้าใจผลการดำเนินงานขององค์กร เพิ่มความโปร่งใสและความน่าเชื่อถือ รวมถึงประชาชนและชุมชนสามารถเข้าถึงข้อมูลเพื่อใช้ประโยชน์ในชีวิตและงานของตนได้

แหล่งที่มา

หลักสูตร : ความรู้และความเข้าใจข้อมูลสำหรับการวิเคราะห์ข้อมูล (Data Literacy for Data Analytics)

บรรยายโดย : ผศ.ดร. โชฬศรีรัตต ธรรมบุษดี อาจารย์ประจำกลุ่มสาขาวิชาเทคโนโลยีการจัดการระบบสารสนเทศ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยมหิดล

สถาบัน/หน่วยงาน/ระบบ : สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)

รูปแบบหลักสูตร : TDGA e-Learning

ช่วงเวลาการฝึกอบรม : พฤษภาคม ๒๕๖๘