

ความรู้และความเข้าใจข้อมูลสำหรับการวิเคราะห์ข้อมูล

โดย นายเจนภพ สีแดง
นักวิชาการแผนกภาพถ่ายปฏิบัติการ

วัตถุประสงค์

- เพื่อให้เข้าใจความหมายและเห็นความสำคัญของข้อมูล
- เพื่อให้สามารถแก้ไขปัญหาหรือทำงานเกี่ยวกับข้อมูล รวมไปถึงการทำความเข้าใจเกี่ยวกับรูปแบบของการทำงานด้านข้อมูลได้
- เพื่อให้มีความรู้ ความเข้าใจ และการเชื่อมโยงของทักษะความสามารถพื้นฐานในเชิงข้อมูล และการวิเคราะห์ข้อมูลได้

สรุปเนื้อหา

การขับเคลื่อนองค์กรด้วยข้อมูลที่มีประสิทธิภาพ อาศัยรากฐานสำคัญจากความเข้าใจในโครงสร้างและรูปแบบของข้อมูล เพื่อเปลี่ยนข้อมูลดิบให้กลายเป็นสารสนเทศและภูมิปัญญาที่มีคุณค่า ควบคู่ไปกับการยึดมั่นในจริยธรรมด้านข้อมูล เพื่อคุ้มครองความเป็นส่วนตัวและปฏิบัติตามกฎหมายอย่างถูกต้อง ในกระบวนการทำงาน เริ่มต้นจากการเก็บรวบรวมข้อมูล ให้ครอบคลุม ทันเวลา และได้มาตรฐาน จากนั้นนำไปผ่านการบริหารจัดการข้อมูล และธรรมาภิบาลข้อมูล เพื่อจัดสรรระบบจัดเก็บ กำหนดสิทธิ์การเข้าถึง และควบคุมคุณภาพข้อมูล ให้มีความน่าเชื่อถือตรงกันทั้งองค์กร เมื่อข้อมูลพร้อมจะเข้าสู่ขั้นตอนการเตรียมข้อมูลเพื่อใช้งาน โดยผ่านการทำความสะอาดข้อมูลเพื่อตรวจสอบความถูกต้องและเป็นการคัดกรองข้อมูล รวมไปถึงแปลงข้อมูลให้อยู่ในรูปแบบที่พร้อมประมวลผล ก่อนจะเข้าสู่กระบวนการวิเคราะห์ข้อมูลเบื้องต้น เพื่อค้นหาความเชื่อมโยง คาดการณ์แนวโน้ม และแปลงข้อมูลให้เป็นข้อสรุปเชิงกลยุทธ์ที่นำไปใช้ตัดสินใจได้อย่างแม่นยำ

๑. ความรู้พื้นฐานเกี่ยวกับข้อมูลและวงจรชีวิตของข้อมูล

๑.๑ นิยามและความหมายของข้อมูล

ข้อมูล (Data) หมายถึง ข้อเท็จจริง สัญญาณ ตัวเลข ข้อความ ภาพหรือเสียง ที่อยู่ในรูปของข้อมูลดิบ (Raw Data) ซึ่งถูกบันทึกและเก็บรวบรวมไว้จากเหตุการณ์ สภาวะการณ์หรือการดำเนินงานต่าง ๆ โดยที่ข้อมูลเหล่านั้นยังไม่ผ่านกระบวนการประมวลผล การจัดหมวดหมู่หรือการใส่บริบทใด ๆ

สารสนเทศ (Information) หมายถึง ข้อมูลดิบที่ผ่านกระบวนการจัดการ ประมวลผล จัดหมวดหมู่ เรียบเรียงหรือคำนวณทางสถิติ พร้อมทั้งถูกใส่บริบทเข้าไป จนทำให้ข้อมูลนั้น มีความหมาย มีโครงสร้าง และอยู่ในรูปแบบที่ผู้ใช้งานสามารถเข้าใจและนำไปใช้ประโยชน์ในการทำงาน การวางแผนหรือการตัดสินใจได้อย่างทันที

ความรู้ (Knowledge) หมายถึง การนำสารสนเทศมาผ่านกระบวนการสังเคราะห์ ร่วมกับประสบการณ์ บริบทแวดล้อม และกรอบความคิด จนเกิดเป็นความเข้าใจเชิงรูปแบบ

ภูมิปัญญา (Wisdom) หมายถึง ขั้นสูงสุดของการประยุกต์ใช้ความรู้ ร่วมกับวิสัยทัศน์ จริยธรรม และการคาดการณ์อนาคต เพื่อกำหนดทิศทางยุทธศาสตร์หรือแก้ปัญหาในระดับนโยบาย

๑.๒ วงจรชีวิตของข้อมูล ประกอบด้วย ๖ ขั้นตอนหลัก ดังนี้

๑. การสร้างข้อมูล (Create) เป็นการสร้างข้อมูลขึ้นมาใหม่หรือการปรับปรุงแก้ไขข้อมูลที่มีอยู่เดิม โดยมีวิธีการบันทึกผ่านทางบุคคลหรือการบันทึกข้อมูลแบบอัตโนมัติด้วยอุปกรณ์อิเล็กทรอนิกส์

๒. การจัดเก็บข้อมูล (Store) เป็นการจัดเก็บข้อมูลที่เกิดขึ้นจากกระบวนการสร้างข้อมูล หรือข้อมูลที่ได้รับจากการเชื่อมโยงและแลกเปลี่ยนข้อมูลร่วมกับหน่วยงานอื่น

๓. การประมวลผลและการใช้ข้อมูล (Use) เป็นการนำข้อมูลที่ได้ทำการจัดเก็บไว้มาประมวลผลเพื่อการใช้งาน

๔. การเผยแพร่ข้อมูล (Publish) เป็นการนำข้อมูลที่อยู่ในความครอบครองของหน่วยงานมาเผยแพร่ผ่านช่องทางต่าง ๆ อย่างเหมาะสม

๕. การจัดเก็บข้อมูลถาวร (Archive) เป็นการย้ายข้อมูลทั้งหมดช่วงอายุการใช้งานหรือข้อมูลที่ไม่ถูกนำมาใช้งานแล้วไปเก็บรักษาไว้ในรูปแบบถาวร โดยข้อมูลดังกล่าวจะไม่มี การลบ ปรับปรุงหรือแก้ไขอีก และสามารถดึงกลับมาใช้งานได้เมื่อมีความต้องการ

๖. การทำลายข้อมูล (Destroy) เป็นการทำลายข้อมูลที่มีการจัดเก็บถาวรมาเป็นระยะเวลานานหรือเกินกว่าระยะเวลาที่กำหนดไว้ตามมาตรฐานการจัดเก็บข้อมูล



ภาพที่ ๑ วงจรชีวิตของข้อมูล

ที่มา : สถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล สำนักงานพัฒนารัฐบาลดิจิทัล (องค์การมหาชน)

๑.๓ ประเภทของข้อมูล แบ่งตามระดับความเป็นระเบียบของโครงสร้างออกเป็น ๓ ประเภท

๑. ข้อมูลแบบมีโครงสร้าง ข้อมูลที่มีรูปแบบจัดเก็บแน่นอน ชัดเจน อยู่ในรูปตารางแถวและคอลัมน์ ประมวลผลและคิวรีได้ง่ายด้วยภาษา SQL ตัวอย่างเช่น ฐานข้อมูลลูกค้า รายการธุรกรรมทางการเงิน ไฟล์ CSV หรือ Excel เป็นต้น

๒. ข้อมูลแบบกึ่งมีโครงสร้าง ข้อมูลที่ไม่ได้อยู่ในรูปตารางแบบดั้งเดิม แต่มีป้ายกำกับ (Tags) หรือตัวค้น (Metadata) ที่อธิบายโครงสร้างภายใน ทำให้ระบบคอมพิวเตอร์สามารถตีความได้ ตัวอย่างเช่น ไฟล์ JSON, XML, YAML หรือเอกสาร NoSQL เป็นต้น

๓. ข้อมูลแบบไม่มีโครงสร้าง ข้อมูลที่ไม่มีโครงสร้างทางคณิตศาสตร์หรือตารางที่แน่นอน คิดเป็นสัดส่วนมากกว่า ๘๐% ของข้อมูลทั้งหมดในองค์กรปัจจุบัน เช่น ข้อความในอีเมล รูปภาพ ไฟล์เสียง วิดีโอหรือโพสต์บนโซเชียลมีเดีย เป็นต้น

๒. จริยธรรมด้านข้อมูลและการคุ้มครองข้อมูลส่วนบุคคล

๒.๑ กรอบจริยธรรมข้อมูล กรอบจริยธรรมข้อมูลครอบคลุมประเด็นหลัก ๔ ประการ ได้แก่

๑. สิทธิในความเป็นเจ้าของ คือการยอมรับว่าผู้ให้ข้อมูลคือเจ้าของข้อมูลที่แท้จริง ไม่ใช่ผู้เก็บรวบรวมข้อมูล

๒. ความโปร่งใส คือการที่ผู้ใช้ต้องรับรู้และเข้าใจว่าข้อมูลของตนถูกนำไปใช้ทำอะไร และระบบอัลกอริทึมมีกระบวนการตัดสินใจอย่างไร

๓. ความเป็นส่วนตัว คือการปกป้องข้อมูลส่วนบุคคลไม่ให้ถูกเข้าถึงโดยไม่ได้รับอนุญาต

๔. ความเป็นธรรม คือการป้องกันไม่ให้อัลกอริทึมสร้างความลำเอียงหรือเลือกปฏิบัติต่อบุคคลหรือกลุ่มคน

๒.๒ กฎหมายสำคัญเกี่ยวกับข้อมูลในประเทศไทย

- พ.ร.บ. ข้อมูลข่าวสารของราชการ พ.ศ. ๒๕๔๐ เป็นกฎหมายกำกับข้อมูลฝั่งภาครัฐ ยึดหลัก "เปิดเผยเป็นหลัก ปกปิดเป็นข้อยกเว้น" เพื่อความโปร่งใสเป็นพระราชบัญญัติที่รับรองสิทธิของประชาชนในการเข้าถึงข้อมูลข่าวสารของราชการ เพื่อส่งเสริมความโปร่งใสและการมีส่วนร่วม โดยกำหนดให้หน่วยงานรัฐนำข้อมูลสำคัญลงพิมพ์ในราชกิจจานุเบกษา จัดไว้ให้ประชาชนตรวจดูหรือเปิดเผยตามคำขอ ทั้งนี้ยกเว้นข้อมูลบางประเภทที่อาจกระทบต่อความมั่นคงหรือข้อมูลข่าวสารส่วนบุคคล ที่ต้องได้รับการคุ้มครองสิทธิความเป็นส่วนตัว

- พ.ร.บ. คุ้มครองข้อมูลส่วนบุคคล พ.ศ. ๒๕๖๒ (PDPA) เป็นกฎหมายคุ้มครองประชาชนไม่ให้ถูกแอบอ้าง นำข้อมูลส่วนตัวไปใช้โดยไม่ได้รับอนุญาตเป็นกฎหมายที่กำหนดหลักเกณฑ์การเก็บรวบรวม ใช้หรือเปิดเผยข้อมูลส่วนบุคคล โดยต้องมี ฐานทางกฎหมาย เช่น ความยินยอม การปฏิบัติตามสัญญาหรือเหตุอื่นที่กฎหมายให้อำนาจ เพื่อสร้างมาตรฐานการคุ้มครองสิทธิขั้นพื้นฐานของประชาชนจากการละเมิด ความเป็นส่วนตัว

๓. การเก็บรวบรวมข้อมูล

กระบวนการเก็บรวบรวมข้อมูลที่มีประสิทธิภาพจำเป็นต้องมีระบบขั้นตอนที่ชัดเจน ตั้งแต่การกำหนดวัตถุประสงค์เพื่อตอบโจทย์เป้าหมายการใช้งาน การระบุแหล่งข้อมูลที่ต้องการ และการเลือกประเภทข้อมูลทั้งเชิงปริมาณ เชิงคุณภาพหรือข้อมูลส่วนบุคคลอย่างเหมาะสม จากนั้นจึงวางแผนเลือกเครื่องมือ ตรวจสอบความถูกต้องและความน่าเชื่อถือของข้อมูล ตลอดจนการกำหนดกรอบเวลาที่ชัดเจนเพื่อให้ได้ข้อมูลที่สมบูรณ์และทันต่อการใช้งานจริง

ในการลงมือปฏิบัติ สามารถเลือกใช้เทคนิคการเก็บรวบรวมข้อมูลได้หลากหลายตามบริบท ได้แก่ วิธีการดั้งเดิมอย่างแบบสอบถาม การสัมภาษณ์ การสังเกตการณ์ การวิเคราะห์เอกสาร การสนทนากลุ่ม และการบอกเล่าประสบการณ์หรือเลือกใช้การวิจัยแบบผสมผสานเพื่อเพิ่มความแม่นยำ รวมถึงการประยุกต์ใช้เทคโนโลยียุคดิจิทัล เช่น การติดตามพฤติกรรมออนไลน์ และการเฝ้าสังเกตเทรนด์บนสื่อสังคมออนไลน์ เป็นต้น ซึ่งช่วยให้ได้ข้อมูลที่ลึกซึ้ง ครอบคลุม และสะท้อนพฤติกรรมที่แท้จริงของผู้ให้ข้อมูล

๔. การบริหารจัดการข้อมูล

การบริหารจัดการข้อมูลอย่างมีประสิทธิภาพเริ่มต้นจากการจำแนกประเภทและการจัดระดับชั้นของข้อมูลอย่างเป็นระบบ โดยแบ่งข้อมูลออกเป็น ๕ หมวดหมู่หลัก ได้แก่ ข้อมูลสาธารณะ ที่เปิดเผยได้ทั่วไป ข้อมูลใช้ภายใน สำหรับการดำเนินงานในองค์กร ข้อมูลส่วนบุคคล ที่ระบุตัวตนและได้รับความคุ้มครองตามกฎหมาย ข้อมูลข่าวสารลับ ตามกฎหมายข้อมูลข่าวสาร และ ข้อมูลความมั่นคง ที่เกี่ยวข้องกับความปลอดภัยของประเทศ จากนั้นนำหมวดหมู่เหล่านี้มาจัดเป็น ๕ ระดับชั้นความลับ เพื่อกำหนดสิทธิ์

การเข้าถึง ได้แก่ ระดับเปิดเผย ระดับเผยแพร่ภายในองค์กร ระดับลับ ระดับลับมาก และระดับลับที่สุด เพื่อควบคุมและปกป้องข้อมูลให้ปลอดภัยตามความสำคัญอย่างเหมาะสม

๕. ธรรมาภิบาลข้อมูล

ธรรมาภิบาลข้อมูล (Data Governance) คือกรอบการกำกับดูแลและบริหารจัดการข้อมูลตลอดวงจรชีวิต ตั้งแต่การสร้าง จัดเก็บ ใช้งาน เผยแพร่ จัดเก็บระยะยาว ไปจนถึงการทำลายข้อมูล โดยเน้นการจัดทำบัญชีข้อมูล การคุ้มครองความเป็นส่วนตัว ความปลอดภัย และการรักษาคุณภาพข้อมูลให้ถูกต้อง ปลอดภัย และเชื่อมโยงกันได้ การนำธรรมาภิบาลข้อมูลมาใช้จะช่วยแก้ปัญหาข้อมูลกระจัดกระจาย ซ้ำซ้อน การนำธรรมาภิบาลข้อมูลภาครัฐเข้ามาประยุกต์ใช้ จึงเข้ามาช่วยแก้ปัญหาผ่าน ๕ เสาหลักสำคัญ ได้แก่ การกำหนดนโยบายและมาตรฐาน การดูแลความเป็นส่วนตัวและการปฏิบัติตามกฎหมาย การกำหนดบทบาทหน้าที่ การวางกระบวนการทำงาน และการมีแนวปฏิบัติที่ชัดเจน ซึ่งผลลัพธ์จะช่วยสร้างระบบข้อมูลที่มีคุณภาพ ปลอดภัย เกิดการเปิดเผยข้อมูลอย่างเหมาะสม และมีมาตรฐานการแลกเปลี่ยนข้อมูลระหว่างหน่วยงานได้อย่างมีประสิทธิภาพ

๖. การเตรียมข้อมูลเพื่อใช้งาน

การเตรียมข้อมูลเพื่อใช้งาน เริ่มต้นจากการทำความสะอาดข้อมูลโดยมีขั้นตอนทั้งหมด ๘ ขั้นตอน คือการนำเข้าข้อมูล การรวมชุดข้อมูล การเติมข้อมูลที่ขาดหาย การปรับมาตรฐาน การจัดรูปแบบ การลบข้อมูลซ้ำ การตรวจสอบความถูกต้อง และส่งออกข้อมูลพร้อมใช้ โดยในการปฏิบัติงานจริงจะดำเนินการกระบวนการกำจัดข้อมูลซ้ำ การข้อมูลที่ไม่เกี่ยวข้องและข้อมูลที่ผิดพลาด การแก้ไขโครงสร้าง การจัดทำมาตรฐาน และการตรวจสอบค่าโดเมนผิดปกติ พร้อมทั้งจัดการข้อมูลที่สูญหาย เพื่อให้ได้ข้อมูลที่สมบูรณ์ แม่นยำ และพร้อมนำไปประมวลผลวิเคราะห์ได้อย่างมีประสิทธิภาพ

๗. การวิเคราะห์ข้อมูลเบื้องต้น

การวิเคราะห์ข้อมูลเบื้องต้นโดยการนำข้อมูลมาใช้ประโยชน์ในองค์กรเริ่มต้นจากการวิเคราะห์ข้อมูลตามลำดับความยากง่าย ตั้งแต่การดูผลสรุปในอดีต การเจาะลึกหาสาเหตุของปัญหา การพยากรณ์สิ่งที่เกิดขึ้นในอนาคต การเสนอแนะทางเลือกที่ดีที่สุด ไปจนถึงการใช้ระบบอัตโนมัติหรือปัญญาประดิษฐ์เข้ามาช่วยตัดสินใจแทนมนุษย์ ซึ่งในการวิเคราะห์ระดับสูงจะมีการนำระบบคอมพิวเตอร์ที่เรียนรู้ได้ด้วยตัวเองเข้ามาช่วยจัดหมวดหมู่ ทำนายผลหรือค้นหารูปแบบพฤติกรรมที่ซ่อนอยู่ อย่างไรก็ตาม การจะเปลี่ยนผลวิเคราะห์ไปสู่การใช้งานจริงได้นั้น ต้องผ่านกระบวนการตัดสินใจอย่างเป็นระบบ เริ่มจากการมองหาปัญหาให้ชัดเจน การลงมือวิเคราะห์ การแปลความหมายโดยระวังไม่ให้ความลำเอียง แล้วจึงนำมาสื่อสารให้ผู้อื่นเข้าใจ ซึ่งการสื่อสารที่มีประสิทธิภาพจะต้องเข้าใจความต้องการของผู้ฟัง เลือกใช้รูปภาพหรือแผนภูมิให้ตรงกับวัตถุประสงค์ ใช้สีหรือขนาดเพื่อเน้นจุดสำคัญ ลบสิ่งรบกวนสายตาออก ตั้งชื่อหัวข้อให้เห็นข้อสรุปชัดเจน และจบด้วยข้อเสนอแนะที่บอกได้ทันทีว่าต้องทำอะไรต่อ ทั้งนี้ ความสำเร็จของการวิเคราะห์ข้อมูลในภาพรวมยังต้องอาศัยการสนับสนุนจากองค์กร ทั้งการวางแผนงานให้ตรงกับเป้าหมาย การมีฐานข้อมูลที่ต้องนำเชื่อถือ ตลอดจนการจัดเตรียมระบบเทคโนโลยีที่ปลอดภัย ยุติธรรม และปฏิบัติตามกฎหมายคุ้มครองความเป็นส่วนตัวอย่างเคร่งครัด

การนำองค์ความรู้ไปปรับใช้ในการปฏิบัติงาน

การประยุกต์ใช้องค์ความรู้ด้านข้อมูลในการปฏิบัติงาน ครอบคลุมการนำข้อมูลมาใช้สร้างคุณค่าทางราชการ ควบคู่ไปกับการดูแลความปลอดภัยและความเป็นส่วนตัวของข้อมูลให้ถูกต้องตามกฎหมาย ในส่วนของการทำงานจริง เน้นการจัดเก็บ คัดกรอง และทำความสะอาดข้อมูลให้ถูกต้องแม่นยำ พร้อมทั้ง จัดหมวดหมู่และกำหนดสิทธิ์การเข้าถึงอย่างเป็นระบบ เพื่อให้ข้อมูลมีความสมบูรณ์พร้อมนำไปวิเคราะห์เชิงลึกสำหรับการหาสาเหตุ คาดการณ์แนวโน้ม และช่วยในการตัดสินใจได้อย่างตรงจุด ตลอดจนการถ่ายทอดข้อสรุปผ่านการนำเสนอที่ชัดเจน เข้าใจง่าย เพื่อให้ผู้บริหารและทีมงานสามารถนำผลลัพธ์ไปกำหนดทิศทาง แก้ไขปัญหา และขับเคลื่อนงานในองค์กรได้อย่างมีประสิทธิภาพสูงสุด

ประโยชน์ต่อตนเอง

๑. ช่วยให้มีความรู้และความเข้าใจเกี่ยวกับข้อมูลมากขึ้น
๒. สามารถจัดเก็บ คัดกรอง และทำความสะอาดข้อมูลได้อย่างถูกต้อง ช่วยลดข้อผิดพลาด และทำงานได้รวดเร็วขึ้น
๓. ช่วยให้มีกระบวนการคิดวิเคราะห์อย่างมีเหตุผล โดยใช้ข้อมูลจริงมาสนับสนุนการตัดสินใจแทนการคาดเดาหรือใช้ความรู้สึก

ประโยชน์ต่อองค์กร

๑. ช่วยให้การจัดหมวดหมู่ข้อมูลและกำหนดสิทธิ์การเข้าถึงอย่างชัดเจน ช่วยปกป้องข้อมูลสำคัญ ป้องกันข้อมูลรั่วไหล และปฏิบัติตามกฎหมายอย่าง PDPA หรือข้อบังคับต่าง ๆ ได้อย่างถูกต้อง
๒. เปลี่ยนจากการตัดสินใจด้วยความรู้สึกไปสู่การใช้ข้อมูลจริง ช่วยลดความเสี่ยง ผิดพลาดน้อยลง และคาดการณ์แนวโน้มในอนาคตได้อย่างแม่นยำ
๓. ช่วยให้ทุกฝ่ายในองค์กรเข้าใจภาพรวมตรงกัน สื่อสารเป้าหมายได้ชัดเจน และผลักดันนโยบาย หรือกลยุทธ์ต่าง ๆ ไปสู่การลงมือปฏิบัติได้อย่างรวดเร็วและเป็นรูปธรรม

ประโยชน์ต่อสาธารณะ

๑. การบริหารจัดการข้อมูลอย่างเป็นระบบช่วยลดความซ้ำซ้อน ลดขั้นตอนที่สิ้นเปลือง และป้องกันการรั่วไหลของงบประมาณ ทำให้นโยบายหรือทรัพยากรสาธารณะไปใช้พัฒนาสังคมได้อย่างคุ้มค่าสูงสุด
๒. การจัดชั้นความลับและปฏิบัติตามกฎหมายคุ้มครองข้อมูลส่วนบุคคล (PDPA) อย่างเข้มงวด ช่วยป้องกันไม่ให้ข้อมูลสำคัญของประชาชนรั่วไหล นำไปสู่สังคมดิจิทัลที่ปลอดภัยและน่าเชื่อถือ

แหล่งที่มา

หลักสูตร : ความรู้และความเข้าใจข้อมูลสำหรับการวิเคราะห์ข้อมูล

บรรยายโดย : ผู้ช่วยศาสตราจารย์ ดร.โชติศรีรัต ธรรมบุษดี

สถาบัน: สถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล (Thailand Digital Government Academy)

รูปแบบหลักสูตร : การเรียนออนไลน์ TDGA e-Learning

ช่วงเวลาการฝึกอบรม : พฤษภาคม ๒๕๖๔