

สรุปบทเรียนจากการพัฒนาความรู้
ความรู้พื้นฐานเพื่อการวิเคราะห์ข้อมูล
สำหรับข้าราชการ
และบุคลากรภาครัฐทุกระดับ

โดย ธนพัฒน์ พรหมทา



Big data คือข้อมูลขนาดใหญ่ มีทั้งแบบโครงสร้างปกติและโครงสร้างข้อมูลที่ไม่มีรูปแบบซึ่งทั้งหมดเป็นข้อมูลที่ใช้ในเชิงธุรกิจ ซึ่งมักถูกใช้กันงานพวกที่ต้องวิเคราะห์ ข้อมูลมีความซับซ้อน และไม่สามารถประเมินขนาดข้อมูลได้

รูปแบบของ Big Data สามารถเป็นได้หลากหลาย ได้แก่

๑. Behavioral Data ได้แก่ ข้อมูลเชิงพฤติกรรมการใช้งานต่างๆ เช่น server log พฤติกรรมการคลิกดูข้อมูล หรือข้อมูลการใช้ ATM
๒. Image & Sounds ตัวอย่างเช่น ภาพถ่าย วิดีโอ รูปจาก Google Street View ภาพถ่ายทางการแพทย์ลายมือ ข้อมูลเสียงที่ถูกบันทึกไว้
๓. languages ตัวอย่าง text message ข้อความที่ถูก tweet เนื้อหาต่างๆในเว็บไซต์
๔. records ตัวอย่างเช่น ข้อมูลทางการแพทย์ ข้อมูลผลสำรวจที่มีขนาดใหญ่ ข้อมูลทางภาษี
๕. sensors ตัวอย่างเช่น ข้อมูลอุณหภูมิ Accelerometer ข้อมูลทางภูมิศาสตร์

คุณลักษณะของ Big Data ประกอบด้วย ๔ ประการ คือ

๑. volume ข้อมูลมีขนาดใหญ่ มีปริมาณข้อมูลมากซึ่งสามารถเป็นได้ทั้งข้อมูลแบบ offline หรือ online
๒. variety ข้อมูลมีความหลากหลาย สามารถเป็นได้ทั้งที่มีโครงสร้างและข้อมูลที่ไม่สามารถจับ Pattern ได้
๓. velocity ข้อมูลมีการเปลี่ยนแปลงตลอดเวลาอย่างรวดเร็ว มีการส่งผ่านข้อมูลอย่างต่อเนื่องในลักษณะ streaming ทำให้การวิเคราะห์ข้อมูลแบบ Manual มีข้อจำกัด
๔. veracity ข้อมูลมีความไม่ชัดเจน (untrusted, uncleaned)

Data lake คืออะไร

๑. การนำเอาข้อมูลจากแหล่งข้อมูลภายนอกองค์กรข้อมูลจากเครือข่ายที่กระจายไปทั่วโลกมาใช้มากขึ้น
๒. ปริมาณข้อมูลจากแหล่งภายนอกเพิ่มขึ้นอย่างต่อเนื่องและมีแนวโน้มที่จะเติบโตแบบก้าวกระโดดมากขึ้น
๓. แก้ไขข้อจำกัดหลายอย่างของ data warehouse ที่ใช้กันมานาน

ลักษณะข้อมูลที่จัดเก็บใน Data lake

๑. ข้อมูลดิบจำนวนมากและมีขนาดใหญ่
๒. ข้อมูลมีรูปแบบไม่แน่นอน
๓. การเข้าถึงข้อมูลไม่สามารถเข้าถึงได้ง่ายต้องใช้ต้องใช้ความสามารถของเจ้าหน้าที่วิเคราะห์ (data scientist)

ความแตกต่างระหว่าง Data lake กับ data warehouse

๑. เก็บข้อมูลทั้งหมด
๒. สนับสนุนข้อมูลทุกชนิด ไม่ใช่เพียงข้อมูลแบบ structure
๓. ผู้ใช้ทุกประเภทสามารถใช้งานได้
๔. ประมวลและวิเคราะห์ข้อมูลได้รวดเร็วกว่า

Big data analytics

การวิเคราะห์ข้อมูล Big data ทำให้มีข้อมูลที่เป็นข้อเท็จจริงซึ่งผ่านการวิเคราะห์อย่างเป็นระบบเพื่อใช้ประกอบการตัดสินใจ โดยระดับการวิเคราะห์ก็เป็นได้หลากหลาย แล้วแต่รูปแบบการนำไปใช้งาน

๑. **Descriptive Analytics** เป็นการวิเคราะห์ในระดับที่บอกว่าเกิดอะไรขึ้น จำนวนเท่าไร ถีแค่ไหน เกิดเหตุการณ์สำคัญตอนไหน ตรงไหนบ้าง
๒. **predictive analytics** เป็นการวิเคราะห์ในลักษณะที่ซับซ้อน ขึ้นไปอีกหนึ่งขั้นคือ เป็นการประเมินว่าจะเกิดอะไรขึ้นต่อไป มีการให้ข้อมูลตัวชี้วัดของผลลัพธ์ที่อาจจะเกิดขึ้นถ้าแนวโน้มยังเป็นอย่างนี้ต่อไป
๓. **prescriptive analytics** เป็นรูปแบบการวิเคราะห์ข้อมูลที่มีความซับซ้อนและยากที่สุด เพราะไม่เพียงพยากรณ์หรือทำนายว่าอะไรจะเกิดขึ้นแต่ยังให้คำแนะนำในทางเลือกต่างๆ และผลของทางเลือกต่างๆจะมีผลดีผลเสียอย่างไร โมเดลของ Prescriptive Analytics นั้นจะสามารถปรับเปลี่ยนได้ตามข้อมูลที่เพิ่มเติมเข้ามามากขึ้น และ prescriptive analytics นี้ยังเป็นการใช้ข้อมูลที่มากที่สุด เกี่ยวพันกับเรื่อง big data เป็นอย่างมาก

Method

๑. **data mining** คือ การวิเคราะห์ข้อมูลจากข้อมูลจำนวนมาก (Big data) เพื่อหาความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ โดยทำการจำแนกประเภท รูปแบบ เชื่อมโยงข้อมูลที่มีความสัมพันธ์กัน และหาความน่าจะเป็นที่จะเกิดขึ้นเพื่อให้ได้องค์ความรู้ใหม่ที่สามารถนำไปใช้ประกอบการตัดสินใจในด้านต่างๆ เช่น ตลาดหลักทรัพย์ ทางธุรกิจ ทางด้านการแพทย์ ยุทธศาสตร์ทหาร เป็นต้น
๒. **text mining** คือ เทคนิคเพื่อค้นหารูปแบบ (Pattern) จากข้อความจำนวนมากโดยอัตโนมัติ โดยใช้ขั้นตอนวิธีจากวิชาสถิติ การเรียนรู้ของเครื่องและการรู้จำแบบหรือในอีกนิยามหนึ่ง การทำเหมืองข้อความ คือ กระบวนการกระทำกับข้อความ (โดยส่วนใหญ่จะมีจำนวนมาก) เพื่อค้นหารูปแบบแนวทางและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อความนั้น โดยอาศัยหลักสถิติ การรู้จำการเรียนรู้ของเครื่อง หลัก

คณิตศาสตร์ หลักการประมวลเอกสาร (Document Processing) หลักการประมวลข้อความ (text processing) และหลักการประมวลผลภาษาธรรมชาติ (Natural Language Processing)

๓. **machine Learning** คือศาสตร์ของการสร้างโมเดลคณิตศาสตร์ มุ่งเน้นที่จะสร้างองค์ความรู้จากข้อมูล โดยเริ่มจากการสร้างโมเดลทางคณิตศาสตร์ที่มีความยืดหยุ่นและสามารถปรับตัวเองเข้ากับข้อมูลที่ได้รับ เพื่อที่จะสามารถทำนายอนาคตได้ เทคโนโลยีที่อยู่เบื้องหลังหุ่นยนต์อัจฉริยะต่างๆ หรือเทคโนโลยีการบิน แบบไร้คนขับ หรือระบบวิเคราะห์ภาพ เสียง ภาพเคลื่อนไหวและภาษามนุษย์

Data driven business

๑. สินทรัพย์ในทางธุรกิจ
๒. เข้าถึงลูกค้าได้ดีขึ้น
๓. พัฒนาประสิทธิภาพและการทำงาน
๔. พัฒนาความพึงพอใจของลูกค้าและผลิตภัณฑ์

ตัวอย่างการใช้งาน Big data

การเลือกตั้ง

- จัดเก็บข้อมูลของผู้มีสิทธิ์เลือกตั้ง ในรูปแบบ profile ต่างๆ
- เก็บข้อมูลเกี่ยวกับการเปลี่ยนแปลงที่อยู่ของผู้มีสิทธิ์เลือกตั้ง
- เก็บข้อมูลเกี่ยวกับ profile ของผู้สมัครเลือกตั้ง
- ข้อมูลเกี่ยวกับรายงานการทุจริตการเลือกตั้งที่ส่งเข้ามา
- ข้อมูลความคิดเห็นของผู้มีสิทธิ์เลือกตั้งเกี่ยวกับการจัดการเลือกตั้ง สถานที่และความสะดวก
- วิเคราะห์แนวโน้มและความต้องการความสะดวกในการเดินทางของผู้มีสิทธิ์เลือกตั้งโดยดูจากข้อมูลภูมิศาสตร์ จาก GPS เพื่อกำหนดสถานที่เลือกตั้งที่สะดวก
- ให้บริการตรวจสอบข้อมูลการเลือกตั้ง สถานที่จัดเลือกตั้ง ทางอินเทอร์เน็ต

งานราชการ

- การตรวจสอบการทุจริต
- การตรวจสอบภัยคุกคาม
- การรักษาความปลอดภัยในระบบ IT
- การปฏิบัติตามกฎระเบียบและการวิเคราะห์
- การปฏิบัติตามกฎหมายอาญา-การทุจริตและการละเมิด
- การพยากรณ์แนวโน้มอาชญากรรมและการป้องกัน

Big data analytic

เป็นเครื่องมือสำคัญและกำลังได้รับความนิยมเป็นอย่างสูง ที่จะคอยดูแลความเคลื่อนไหวธุรกิจองค์กรที่เกิดขึ้นบนโลกออนไลน์ มีความจำเป็นต่อภาครัฐในทุกประเทศ ในการขับเคลื่อนการบริหารราชการแผ่นดินอย่างมีประสิทธิภาพและขับเคลื่อนเศรษฐกิจให้ทันต่อการเปลี่ยนแปลงทางเทคโนโลยี อย่างก้าวกระโดดในศตวรรษที่ ๒๑

๑. **data visualization** การแสดงผลข้อมูลในรูปแบบที่เข้าใจได้ง่าย
๒. **real time monitoring** ไม่ใช้การรวบรวมเป็นรายงานสรุปแต่แสดงผลออกมาแบบ real time
๓. **quality data** นอกจากจะบอกปริมาณแล้ว ยังบอกทิศทางและรายละเอียดว่าคนกล่าวถึงธุรกิจในแง่มุมบวกหรือลบ

Big data analytic กับ การบริหารภาครัฐ

๑. รับฟังความคิดเห็น รวบรวมข้อมูลและปรึกษากับผู้มีส่วนได้ส่วนเสีย
๒. วางแผนการลงทุนในการจัดโครงสร้าง
๓. มีความเข้าใจและมีทักษะทางธุรกิจและทักษะทางเทคนิค
๔. เตรียมความพร้อมภายใต้การเปลี่ยนแปลงของเทคโนโลยี
๕. เจ้าหน้าที่ภาครัฐจะต้องปรับ mindset ในการเข้าร่วมกับทุกภาคส่วน
๖. ปรับปรุงวิธีคิดและกระบวนการเพื่อทำให้เกิดการแลกเปลี่ยนข้อมูลและการใช้ข้อมูลร่วมกันระหว่างหน่วยงานภาครัฐ
๗. กำหนดแนวและการบริการให้คำปรึกษาในด้าน big data analytics ให้แก่ทุกภาคส่วน

แหล่งที่มา

สรุปบทเรียนจากการเรียนรู้ผ่านสื่ออิเล็กทรอนิกส์ (e-Learning)

สถาบันพัฒนาบุคลากรภาครัฐด้านดิจิทัล (TDGA)

หลักสูตร : ความรู้พื้นฐานเพื่อการวิเคราะห์ข้อมูลสำหรับข้าราชการและบุคลากรภาครัฐทุกระดับ

บรรยายโดย : รองศาสตราจารย์ ดร.วสุชัย ร่มสายหยุด สาขาวิชาวิทยาศาสตร์และเทคโนโลยี

มหาวิทยาลัยสุโขทัยธรรมมาธิราช