รายงานการเข้าร่วมโครงการเอพีโอ
24-CP-62-GE-TRC-A
Training Course on Big Data Analytics and Data Visualization for Productivity
ระหว่างวันที่ 26-29 พฤศจิกายน 2567
จัดโดย Department for Productivity Implementations, Ministry of Industry and Technology, Turkiye (NPO Turkiye) และ APO Secretariat ผ่านระบบออนไลน์

จัดทำโดย นาย อินทัช แสงกระจ่าง
วิศวกรคอมพิวเตอร์ การไฟฟ้านครหลวง
วันที่ 17 มกราคม 2568

ส่วนที่ 1 เนื้อหา/องค์ความรู้จากการเข้าร่วมโครงการ

1.1 ที่มาหรือวัตถุประสงค์ของโครงการโดยย่อ

เพื่อให้ความรู้และเน้นย้ำถึงความสำคัญเกี่ยวกับการใช้ข้อมูลในการตัดสินใจขององค์กร และปูพื้นฐานความรู้ด้านการวิเคราะห์และจัดแสดงข้อมูลเพื่อใช้ในการดำเนินธุรกิจ และเสริมความรู้ความสามารถให้กับพนักงานภายในองค์กร เพื่อเพิ่มประสิทธิภาพในการทำงาน เนื่องจากในปัจจุบัน การใช้ข้อมูลมาประกอบการดำเนินธุรกิจถือว่าเป็นส่วนสำคัญในการเปลี่ยนแปลงองค์กรให้เป็นองค์กรดิจิทัล และเทคโนโลยีใหม่ๆ เช่น Generative AI ที่กำลังเป็นประเด็นร้อนในโลกปัจจุบัน ก็ใช้ข้อมูลในการขับเคลื่อนเช่นกัน

1.2 เนื้อหา/องค์ความรู้ที่ได้จากกิจกรรมต่างๆ

หัวข้อที่ 1: Unleash the Power of Data: Case Studies from Turkiye (EIS Database) โดย Mehmet Atas, Senior Expert Researcher, Entrepreneur Information System Department, Ministry of Industry and Technology, Turkiye

Big Data และสถาณการณ์ด้านข้อมูลในปัจจุบัน

- Big Data คือข้อมูลต่างๆ ในทุกรูปแบบ ทั้ง Structured (จัดเก็บในรูปแบบตาราง), Semi-structured (จัดเก็บในรูปแบบโครงสร้างที่ยืดหยุ่นได้) และ Unstructured (จัดเก็บในรูปแบบอื่นๆ) ที่มีคุณสมบัติอย่างน้อย 3Vs ดังนี้
    - Volume มีข้อมูลปริมาณมาก
    - Velocity ข้อมูลเกิดขึ้นด้วยความเร็วสูง
    - Variety ข้อมูลมีความหลากหลาย

  ซึ่งในปัจจุบัน Big data ได้มีการนำไปใช้งานในการวางแผนและดูแลกิจกรรมด้านเศรษฐกิจ ความมั่นคง การให้บริการด้านสาธารณะ และการวิจัย
- ในปี 2024 มีข้อมูลดิจิทัลเกิดขึ้นมากถึงวันละ 328.77 exabytes (328.77 ล้าน terabytes) และมีการคาดการณ์ว่าข้อมูลในโลกทั้งหมดจะมีปริมาณสูงถึง 181 zettabytes (1.81 แสนล้าน terabytes) ภายในปี 2025

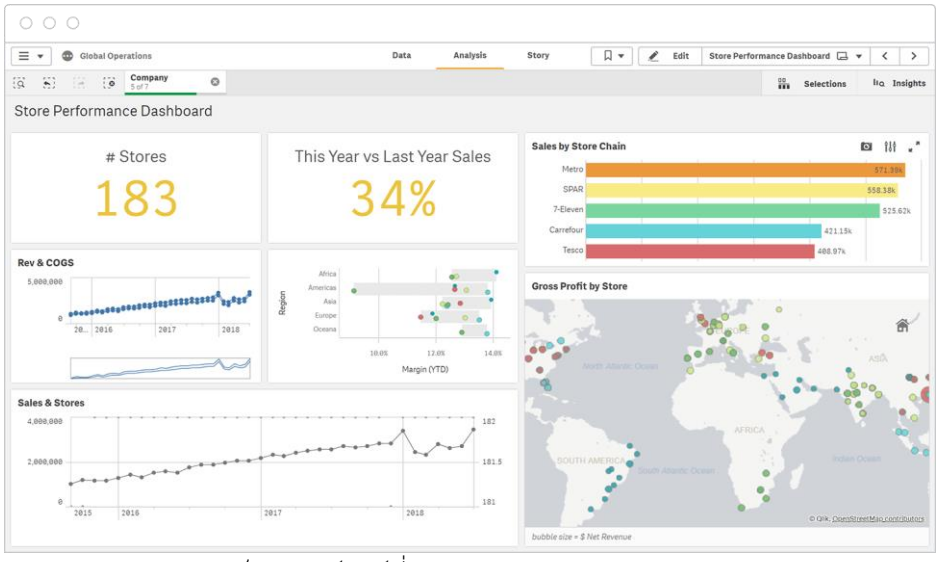
Year	Generated Data Size	Increase Rate
2010	2 Zettabytes	-
2011	5 Zettabytes	150%
2012	6,5 Zettabytes	30%
2013	9 Zettabytes	38%
2014	12,5 Zettabytes	39%
2015	15,5 Zettabytes	24%
2016	18 Zettabytes	16%
2017	26 Zettabytes	44%
2018	33 Zettabytes	27%
2019	41 Zettabytes	24%
2020	64,2 Zettabytes	57%
2021	79 Zettabytes	23%
2022	97 Zettabytes	23%
2023	120 Zettabytes	24%
2024	147 Zettabytes	23%
2025*	181 Zettabytes	23%

(ตารางแสดงปริมาณและอัตราการเติบโตของข้อมูลทั้งหมดตั้งแต่ปี 2010 ถึง 2024 และคาดการณ์ปี 2025)

- ข้อมูลทั้งหมดในโลกกว่า 53% เป็นวิดีโอ ตามมาด้วย social data และ gaming data และในปัจจุบัน ข้อมูลกว่า 70% เป็นข้อมูลที่เกิดจากผู้ใช้งานทั่วไปสร้างขึ้น เช่น โพสต์ สตอรี่ คลิปสั้น และอื่นๆ
- องค์กรในโลกถึง 97% มีการลงทุนทางด้าน Big data คิดเป็นมูลค่ารวม 348.21 พันล้านดอลล่าร์สหรัฐ

การรวบรวมข้อมูล (Data Integration) และเทคโนโลยีที่เกี่ยวข้อง

- ระบบฐานข้อมูลจะถูกแบ่งหลักๆ เป็น 2 แบบ คือ OLTP และ OLAP โดย OLTP จะถูกออกแบบมาเพื่อใช้ในการอ่าน-เขียนข้อมูลในลักษณะ transaction (อ่าน-เขียนข้อมูลปริมาณน้อยๆ เป็นจำนวนมาก) และ OLAP จะถูกออกแบบมาสำหรับการอ่านข้อมูลปริมาณมากๆ ในคราวเดียว เพื่อใช้ในการสรุปและวิเคราะห์ข้อมูลสำหรับใช้ในการตัดสินใจทางธุรกิจ
- ETL เป็นชื่อเรียกวิธีการนำเข้าข้อมูลจากระบบต้นทางสู่ระบบปลายทาง และเป็น 1 ในวิธีการถ่ายข้อมูลจากระบบ OLTP ซึ่งเป็นต้นทาง ไปยัง OLAP ที่ใช้ในการวิเคราะห์ข้อมูล โดยประกอบด้วย 3 ขั้นตอนหลักได้แก่
    - Extract เป็นการดึงข้อมูลจากต้นทาง ไม่ว่าจะเป็น Database, API, ไฟล์ หรือข้อมูลในรูปแบบอื่นๆ โดยข้อมูลที่ดึงมาควรจะเป็นข้อมูลที่จะนำไปใช้ในภายหลังจริงๆ
    - Transform เป็นการแปลงข้อมูลที่ดึงมาให้อยู่ในรูปแบบที่พร้อมจัดเก็บและใช้งานในปลายทาง ซึ่งมีหลายวิธีในการแปลงข้อมูล เช่น Data cleansing, Standardization, Data enrichment และ Data filtering
    - Load เป็นการส่งข้อมูลเพื่อจัดเก็บในระบบปลายทาง ซึ่งสามารถทำได้ในรูปแบบ Full load (จัดเก็บข้อมูลทุกอย่างลงไปในปลายทางทุกครั้งที่มีการทำ ETL) หรือ Incremental load (จัดเก็บเฉพาะสิ่งที่เปลี่ยนแปลงจากข้อมูลที่มีอยู่ในปลายทางแล้ว)
- ตัวอย่างเครื่องมือสำหรับทำ Data Integration พร้อม Data Visualization: Qlik Sense BI Tool



- Qlik Sense เป็นซอฟต์แวร์ที่สามารถทำ ETL, Data Management และ Business Intelligence report ได้ภายในตัวเอง และสามารถใช้ API เพื่อช่วยในการทำกระบวนการทางข้อมูลแบบอัตโนมัติได้ด้วย

Data Warehousing

- Data Warehouse เป็นฐานข้อมูลแบบ OLAP ที่ออกแบบมาเพื่อใช้งานสำหรับการวิเคราะห์ข้อมูล โดยรองรับการจัดเก็บข้อมูลจากหลายๆ แหล่ง ทั้ง RDBMS, ไฟล์เอกสาร และอื่นๆ
- Data Warehouse ช่วยลดโอกาสการเกิด Data silo หรือการทับถมกันของข้อมูลโดยไม่ถูกใช้งาน และช่วยให้ข้อมูลมีความน่าเชื่อถือมากขึ้น เนื่องจากเป็นการรวบรวมข้อมูลจากหลายๆ แหล่งมาไว้ในที่เดียว
- Data Warehouse นิยมใช้ในการเก็บข้อมูลย้อนหลังเป็นเวลานานเพื่อใช้ในการวิเคราะห์แนวโน้มการเปลี่ยนแปลง, หา pattern ของข้อมูล และตรวจจับความผิดปกติของข้อมูล
- โครงสร้างของฐานข้อมูลแบบ OLAP ที่ใช้ใน Data Warehouse จะอยู่ในรูปแบบ Fact และ Dimension table
    - Fact table เป็นตารางที่เก็บข้อมูลที่ใช้ในการวิเคราะห์ ซึ่งมักจะเป็นตัวเลข เช่น ยอดขาย
    - Dimension Table เป็นตารางที่เก็บข้อมูลที่ใช้ประกอบข้อมูลจาก fact table เช่น วันที่ รายละเอียดลูกค้า

ระบบ Entrepreneur Information System (EIS)

- ระบบรวบรวมข้อมูลกลางของตุรกี EIS เป็นระบบที่รวบรวมข้อมูลรายงานจากองค์กรต่างๆ ในตุรกีเข้าไว้ด้วยกันในแพลตฟอร์มเดียว โดยมีเป้าหมายเพื่อเป็นศูนย์กลางการให้บริการข้อมูลที่แม่นยำให้กับผู้ที่ต้องใช้งานข้อมูล เช่น สถาบันวิจัย ผู้เชี่ยวชาญ ผู้บริการระดับสูง มหาวิทยาลัย และองค์กรต่างๆ เพื่อใช้ในการวิเคราะห์และประเมินเศรษฐกิจ และวางแผนนโยบาย โดย goal ของระบบ EIS มีดังนี้
    - เชื่อถือได้
    - มีเป้าหมายชัดเจน
    - ยืดหยุ่นได้
    - เข้าถึงข้อมูลได้
    - มีการควบคุมข้อมูลลับ
    - มีความคล่องตัว
    - สามารถใช้งานร่วมกันกับข้อมูลอื่นๆ ได้

- ในปัจจุบัน ระบบ EIS มีข้อมูลจาก 13 องค์กรภาครัฐของตุรกี โดยมีการรวบรวมข้อมูลเกี่ยวกับผลิตภัณฑ์ภายในประเทศ, ข้อมูลบริษัทจดทะเบียนและข้อมูลด้านการเงิน, ข้อมูลแรงงาน, ข้อมูลสิทธิบัตร การวิจัย เป็นต้น โดยในระบบ EIS มีข้อมูลย้อนหลังไปตั้งแต่ปี 2006 ถึงปัจจุบัน และมีข้อมูลรวมกันมากกว่า 100 ชุดข้อมูล
- ข้อมูลในระบบ EIS จะถูกแบ่งกลุ่มด้วยหมวดหมู่ต่างๆ หลายรูปแบบ เช่น
    - ลักษณะของธุรกิจ และ sectoral code ตามมาตรฐานต่างๆ (NACE, PRODTR, GTIP, ISIC Rev.4, BEC)
    - พื้นที่ตามภูมิศาสตร์
    - ขนาดของธุรกิจ เล็ก กลาง ใหญ่
    - Technology Level สำหรับอุตสาหกรรมการผลิต เช่น Low level สำหรับอุตสาหกรรมอาหาร และ High level สำหรับอุตสาหกรรมคอมพิวเตอร์และเวชกรรม
    - พื้นที่การแบ่งเขตอุตสาหกรรม เช่น พื้นที่ธุรกิจเทคโนโลยี พื้นที่การวิจัยและพัฒนา พื้นที่อุตสาหกรรม
- ระบบ EIS มีการใช้ Tool ชื่อ Pentaho Data Integration (PDI) ในการนำเข้าข้อมูลมายังระบบ EIS
    - PDI เป็น Open-source ETL tool ที่ใช้สร้างกระบวนการนำเข้าข้อมูลทั้งแบบ no code และแบบเขียนโค้ดในภาษา Java, JS และ Python ที่เปิดตัวเมื่อปี 2004 และมีการเพิ่มฟังก์ชันด้าน Business Intelligence ในปี 2005 โดยรองรับการนำเข้าข้อมูลจากไฟล์ excel, json, xml และการเชื่อมต่อฐานข้อมูลโดยตรง
    - ระบบ EIS มีการใช้ PDI .ใน 5 component หลักๆ ดังนี้
        - Kettle: ชุดระบบสำหรับนำเข้าข้อมูล
        - Spoon: User Interface สำหรับสร้าง ETL process
        - Kitchen: ระบบที่ใช้รันงานที่สร้างจาก Spoon
        - Pan: Command-line เพื่อใช้สั่งงานระบบของ PDI
        - Carte: Web container เพื่อใช้ในการตั้ง ETL server
- ระบบ EIS มีการควบคุมการเข้าถึงข้อมูลที่ไม่ใช้ข้อมูลสาธารณะในระบบด้วยการกำหนดสถานที่ที่สามารถดูข้อมูลได้ คือ SRC Study Environment
    - มีการทำระบบ Access control ด้วยลายนิ้วมือ
    - เครื่องคอมพิวเตอร์ที่เข้าถึงระบบ EIS ได้ภายในอาคารจะไม่มีการเชื่อมต่อเครือข่ายข้างนอก
    - ระงับการนำอุปกรณ์อิเล็กทรอนิกส์อื่นๆ เข้าไปในอาคาร
    - มีการติงตั้งกล้องวงจรปิดเพื่อตรวจสอบการเคลื่อนไหวภายในอาคาร
    - มีซอฟต์แวร์สำหรับวิเคราะห์และใช้งานข้อมูลติดตั้งไว้ในเครื่องคอมพิวเตอร์ภายในอาคาร
- ข้อมูลที่เผยแพร่สู่สาธารณะในระบบ EIS เพื่อใช้ในงานวิจัยและการใช้งานอื่นๆ ของบุคคลทั่วไป จะเป็นข้อมูลที่สังเคราะห์ขึ้นมา โดยให้มีคุณลักษณะเหมือนกับข้อมูลจริงๆ แต่จะไม่สามารถสืบหาข้อมูลส่วนบุคคลจากข้อมูลที่สังเคราะห์ขึ้นมาได้ โดยโครงการนี้มีชื่อว่า e-SRC (e-Study and Research Center)
- เป้าหมายในอนาคตของ EIS
    - รองรับการวิเคราะห์ข้อมูลภูมิศาสตร์
    - รองรับการใช้งานจากภายนอก (remote working)
    - ย้ายระบบจัดเก็บข้อมูลไปใช้ซอฟต์แวร์ Open-source
    - พัฒนา Data mart สำหรับบุคคลทั่วไปให้เข้ามาใช้งาน พร้อม storytelling ให้กับข้อมูล
    - Migrate ให้รองรับ Big data platform tool เช่น Hadoop, Spark

ความท้าทายและอุปสรรคในการประมวลผลข้อมูล

- การรักษาคุณภาพข้อมูลให้เป็นมาตรฐานเดียวกันจากข้อมูลหลายๆ แหล่งที่มา
    - ใช้ระบบตรวจสอบคุณภาพข้อมูลแบบอัตโนมัติร่วมกับการใช้คนในการเช็ค

- การจัดการข้อมูลปริมาณมากๆ ซึ่งใช้เวลาและทรัพยากรสูง
    - จัดกลุ่มข้อมูลในการประมวลผล
    - Parallel processing
- การจัดการความปลอดภัยของข้อมูลและการจัดการข้อมูลส่วนบุคคล
- การประมวลผลข้อมูลที่มีหลายหลายประเภท ซึ่งข้อมูลแต่ละแบบมี tool ที่เหมาะสมแตกต่างกัน
- การสร้าง report ที่ไม่เกิดประโยชน์
    - ศึกษากลุ่มเป้าหมายที่จะดู report ก่อนสร้าง
    - กำหนด objective ของ report ให้ชัดเจน
    - ทำ report ให้เรียบง่าย

หัวข้อที่ 2: Data Ethics and Information Security โดย Aytunc Ayhan, Industry and Technology Specialist, Entrepreneur Information System Department, Ministry of Industry and Technology, Turkiye

Data Ethics (จริยธรรมข้อมูล) เป็นเรื่องเกี่ยวกับการจัดการข้อมูลโดยเคารพความเป็นส่วนตัวของแต่ละบุคคลและเปิดเผยการทำงานกับข้อมูลให้โปร่งใส (เก็บข้อมูลนี้ดีไหม เอามาทำอะไร จัดเก็บอย่างไร)

- ข้อมูลที่ได้จากผู้ใช้งานมีปริมาณเยอะมาก ซึ่งสามารถนำไปสู้การตัดสินใจที่ชาญฉลาด ปรับแต่งการบริการลูกค้าให้เข้ากับแต่ละคน และพัฒนาคุณภาพชีวิตโดยรวม ดังนั้นจึงต้องมีการกำกับดูแลข้อมูลที่ทำให้ผู้ใช้งานระบบเกิดความเชื่อถือว่าข้อมูลของเขาจะมีความปลอดภัย

ภาพรวมของการจัดการข้อมูลอย่างมีจริยธรรม

- กำกับความรับผิดชอบในการเก็บ ประมวลผล และใช้ข้อมูลที่มีผลกับชีวิตของบุคคลอื่น
- สร้างความน่าเชื่อถือของผู้เก็บข้อมูลกับผู้ใช้งาน
- วางแผนกำกับควบคุมทั้งเรื่องเทคนิค กฎหมาย สังคม และความเชื่อของแต่ละคน
- ระวังเรื่องความลำเอียงของข้อมูลและการเปิดเผยข้อมูลส่วนบุคคล เพราะจะทำให้ทำให้ข้อมูลขาดความน่าเชื่อถือ
- สร้างความสมดุลระหว่างนวัตกรรมและความรับผิดชอบต่อข้อมูล

พื้นฐานของหลักการทำให้ระบบมีจริยธรรมด้านข้อมูล

- ปกป้องข้อมูลส่วนบุคคลไม่ให้ถูกเข้าถึงโดยไม่ได้รับอนุญาต
- องค์กรต้องทำให้เห็นชัดเจนว่าองค์กรมีการรวบรวม เก็บข้อมูล และใช้ข้อมูลอย่างไรบ้าง เพื่อสร้างความน่าเชื่อถือ
- ข้อมูลต้องไม่มีความลำเอียง ที่นำไปสู่การตัดสินใจที่ไม่เป็นธรรม
- องค์กรต้องแสดงความรับผิดชอบถ้าเกิดปัญหาจากข้อมูลหรือวิธีการจัดการข้อมูล
- ให้เจ้าของข้อมูลเป็นผู้ตัดสินใจในการเก็บข้อมูล

Information Security (InfoSec) เป็นเรื่องเกี่ยวกับการป้องกันข้อมูลที่เราเก็บไว้ไม่ให้ถูกใช้งานในทางที่ผิด หรือถูกนำออกไปข้างนอกโดยไม่ได้รับอนุญาต หรือข้อมูลถูกรวบกวนหรือข้อมูลเสียหาย โดยมีหลักการสำคัญ 3 ข้อ (CIA Triad) ดังนี้

- Confidentiality: คนที่ได้รับอนุญาตเท่านั้นที่จะเข้าถึงข้อมูลได้
- Integrity: ข้อมูลมีความถูกต้อง และเหมือนกันในทุกขั้นตอน และป้องกันการแก้ไขที่ไม่ได้รับอนุญาต
- Availability: ข้อมูลสามารถเข้าถึงได้กับคนที่ได้รับอนุญาตเมื่อต้องการ

ประเภทของข้อมูลที่อยู่ในองค์กร

- Personal Data: ข้อมูลที่ระบุถึงตัวตน เช่น รายละเอียดพนักงานหรือลูกค้า
- Financial Data: ข้อมูลเกี่ยวกับการเงินขององค์กร เช่น งบประมาณ รายได้ ค่าใช้จ่าย
- Operational Data: ข้อมูลที่ใช้ในการดำเนินงานรายวัน เช่น ข้อมูลสินค่าในคลัง ข้อมูลนัดหมาย
- Intellectual Property Data: ข้อมูลสิ่งที่เป็นเอกลักษณ์ขององค์กร เช่น สิทธิบัตร โปรแกรมที่พัฒนาเอง ดีไซน์ของสินค้า
- Customer Data: ข้อมูลที่เก็บจากลูกค้า เช่น ประวัติการซื้อของ feedback ความชอบ

วิธีการเก็บรวบรวมข้อมูลอย่างมีจริยธรรม

- Transparency มีการชี้แจงว่าเก็บข้อมูลเมื่อไหร่ อย่างไร และเก็บไปทำไม
- Consent ใช้การขอความยินยอมที่ชัดเจน และให้ user ขอให้เก็บข้อมูล แทนที่จะให้ user ขอให้ไม่เก็บข้อมูล
- Data Minimization เก็บเท่าที่ต้องการใช้เท่านั้น
    - เช่น เว็บ Online Shopping ให้เก็บข้อมูลเฉพาะประวัติการซื้อของเท่านั้น เพื่อใช้ในการแนะนำสินค้า ไม่เก็บข้อมูลลูกค้าและข้อมูลการท่องเว็บ
    - ลดความเสี่ยงข้อมูลรั่วไหล และเพิ่มความเชื่อใจของผู้ใช้งาน
- Security ป้องกันข้อมูลที่เก็บมาอย่างปลอดภัย เช่น encryption

ความถูกต้องและเป็นหนึ่งเดียวกันของข้อมูล

- ข้อมูลที่แม่นยำช่วยให้การตัดสินใจมีความยุติธรรมและมีเหตุมีผลจากข้อมูล ลดความเสี่ยงและความผิดพลาด
- มีการตรวจสอบข้อมูลว่าถูกต้องและทำตามมาตรฐานที่วางไว้
- ข้อมูลที่ไม่ถูกต้องทำให้การตัดสินใจจากข้อมูลผิดพลาด ทำให้ได้ผลลัพธ์ที่ไม่ถึงประสงค์ และลดความน่าเชื่อถือและความยุติธรรมลง
- องค์กรที่เก็บข้อมูลต้องรับผิดชอบในการรักษาข้อมูลด้วยนโยบายและการติดตาม

วิธีการจัดเก็บข้อมูลอย่างมีจริยธรรม

- ใช้ข้อมูลในจุดประสงค์ที่ขอ consent มาเท่านั้น ถ้าจะใช้ทำอย่างอื่นต้องขอ consent ใหม่
- ข้อมูลต้องถูกต้องและ up to date
- Storage มีการเข้ารหัสและกำหนดการเข้าถึงข้อมูลให้ผู้ที่ได้รับอนุญาตเท่านั้น
- เก็บข้อมูลไว้นานแค่เท่าที่ใช้งาน หลังจากนั้นให้ทำลายข้อมูลด้วยวิธีที่ปลอดภัย ไม่สามารถกู้คืนได้
- ข้อมูลต้องสามารถเข้าถึงได้ในเวลาที่ต้องการใช้ข้อมูล โดยการใช้ระบบสำรอง มีการตั้ง server มากกว่า 1 แห่ง มีการ backup ข้อมูล เพื่อรับประกันว่าข้อมูลจะสามารถใช้งานได้ตลอดเวลา

วิธีการแชร์ข้อมูลกับบุคคลภายนอก

- ข้อมูลที่แชร์ต้องถูกต้องแม่นยำ
- มีช่องทางและขั้นตอนการแชร์ข้อมูลที่ปลอดภัย
- ตั้งนโยบายให้ผู้ที่จะใช้ข้อมูลทำตามให้สอดคล้องกับนโยบายการใช้ข้อมูลขององค์กร
- มีการตั้งขอบเขตความรับผิดชอบกรณีเปิดปัญหากับข้อมูล
- มีการ audit และ monitor การใช้งานข้อมูลของบุคคลภายนอกให้เป็นไปตามมาตรฐานและข้อกำหนดที่ตั้งไว้

วิธีการสร้างความปลอดภัยให้กับข้อมูล

- เข้ารหัสข้อมูล ใช้ในการรับ-ส่งข้อมูล
    - สามารถปลดล็อคข้อมูลให้อ่านได้ถ้ามี key เท่านั้น โดย key มี 3 รูปแบบ
        - Symmetric ใช้ Key เดียวในการล็อกและปลดล็อก
        - Asymmetric key ใช้ public key ในการเข้ารหัส และ private key ในการถอดรหัส
        - Hybrid
    - การเข้ารหัสข้อมูลสามารถทำได้ 2 จุด
        - At rest (เข้ารหัสข้อมูลที่เก็บไว้)
        - In transit (เข้ารหัสข้อมูลก่อนรับ-ส่ง)
- Multi Factor Authentication ใช้รหัสผ่านมากกว่า 1 รูปแบบ เช่น Password + OTP เพื่อเข้าถึงข้อมูล
- Access Control กำหนดการเข้าถึงข้อมูลด้วยตำแหน่ง เช่น ข้อมูลพนักงานให้เข้าถึงได้เฉพาะทีม HR
- Audit, Monitoring อย่างสม่ำเสมอและต่อเนื่อง เพื่อให้สามารถตรวจเจอปัญหาได้รวดเร็ว

วิธีการปกป้องข้อมูลความลับ

- Data anonymization ลบข้อมูลที่สามารถระบุตัวตนได้ออกไปก่อนนำไปใช้งาน
- Role-based access control กำหนดการเข้าถึงข้อมูลด้วยตำแหน่ง เช่น ข้อมูลพนักงานให้เข้าถึงได้เฉพาะทีม HR
- Secure Communication channel ส่งข้อมูลด้วยช่องทางที่เข้ารหัสข้อมูล เช่น secure email เพื่อป้องกันการดักข้อมูลระหว่างส่ง

เครื่องมือที่ช่วยในการปกป้องข้อมูล

- Data masking tool: เครื่องมือสำหรับซ่อนข้อมูลส่วนบุคคล
- Data Loss Prevention: เครื่องมือป้องกันการสูญหายของข้อมูล
- SFTP: เครื่องมือรับ-ส่งข้อมูลแบบปลอดภัย
- Privacy impact assessment: เครื่องมือประเมินความเสี่ยงที่จะทำให้ข้อมูลส่วนบุคคลรั่วไหล
- Data Governance Platform: เครื่องมือจัดการควบคุมการใช้ข้อมูลตามนโยบายที่กำหนด

ประเภทของภัยทางไซเบอร์

- มัลแวร์ คือซอฟต์แวร์ที่ออกแบบมาสำหรับเจาะระบบหรือทำลายระบบ
    - Ransomware เข้ารหัสไฟล์เพื่อเรียกค่าไถ่
- Phishing ปลอมแปลงระบบเพื่อหลอกให้ผู้ใช้งานกรอกข้อมูลโดยคิดว่าเป็นระบบจริง
- DDoS โจมตีระบบให้ล่มด้วยการยิง packet จำนวนมาก
- ผู้ใช้งานต้องได้รับความรู้และวิธีการรับมือกับภัยทางไซเบอร์เพื่อป้องกันการรั่วไหลของข้อมูล

วิธีการรับมือเหตุการณ์ข้อมูลรั่วไหล

- แยกแยะว่าข้อมูลรั่วไหลจากวิธี/ช่องทางไหน
    - Unauthorize access
    - ถูกขโมยข้อมูล
    - Ransomware
    - malware
- มีแผนการรับมือระยะสั้น
- ควบคุมความเสียหายให้อยู่ในวงแคบที่สุด
- สืบหาว่าจุดเริ่มต้นของการรั่วไหลข้อมูลมาจากที่ไหน และเกิดจากสาเหตุอะไร
- วางแผนการป้องกันในอนาคต และตรวจสอบอีกครั้ง

จริยธรรมสำหรับ AI

- ทำให้การตัดสินใจของ AI เห็นได้ชัดเจนว่าเกิดมาจากอะไร
- ใช้แหล่งข้อมูลที่มีความหลากหลายเพื่อป้องกัน Bias
    - เช่น การ train AI ด้วยข้อมูลที่มีเพศ/สีผิวแบบใดแบบหนึ่งเยอะเป็นพิเศษ ทำให้ AI เกิด bias ได้
- ระมัดระวังการนำข้อมูลส่วนตัวของผู้ใช้งานไป train AI
- วางแผนการและผู้รับผิดชอบต่อสิ่งที่ AI ตัดสินใจ

วิธีการกำหนดนโยบายปกป้องข้อมูล

- กำหนดนโยบายเรื่องการเข้าถึง, จัดเก็บ, ใช้งานข้อมูลให้ชัดเจน
- จัดการอบรมเรื่องวิธีการใช้ข้อมูลอย่างปลอดภัยอย่างสม่ำเสมอ
- กำหนด Data Protection Officer เพื่อกำหนดและตรวจสอบนโยบายเรื่องความเป็นส่วนตัวต่างๆ ว่าใช้งานได้จริง
- Audit และตรวจสอบความเสี่ยงอย่างสม่ำเสมอ

วิธีการสร้างวัฒนธรรมองค์กรให้มีจริยธรรมข้อมูล

- ผู้นำองค์กรต้องเป็นตัวอย่างการใช้งานจริยธรรมข้อมูลที่ดี
- มีนโยบายที่ชัดเจน และจัดการสอนให้พนักงานในองค์กร

- สนับสนุนให้พนักงานรับผิดชอบต่อข้อมูล และมีการพูดคุยทันทีที่เห็นความเสี่ยงของการรั่วไหลข้อมูล
- พูดคุยเรื่องจริยธรรมข้อมูลอย่างเปิดเผยภายในทีม
- ชื่นชมพนักงานที่ปฏิบัติตามนโยบายจริยธรรมข้อมูล

แนวโน้มในอนาคตของ Data Security และ Data Ethic

- เทคโนโลยีในอนาคตต้องการข้อมูลมากขึ้นเรื่อยๆ ดังนั้นต้องค้นหาสมดุลระหว่างนวัตกรรมและความรับผิดชอบต่อข้อมูล
- พัฒนาวิธีการเข้ารหัสและการป้องกันข้อมูลใหม่ๆ เพื่อรองรับกับเทคโนโลยีใหม่ๆ โดยเฉพาะ Quantum computer ที่จะทำให้วิธีการเข้ารหัสในปัจจุบันใช้งานไม่ได้
    - Quantum cryptography ที่ใช้หลักการอื่นนอกเหนือจากการเข้ารหัสผ่านจำนวนเฉพาะจะช่วยลดความเสี่ยงจากการโดนเจาะโดย Quantum computer
- กฎหมายเกี่ยวกับข้อมูลส่วนบุคคลใหม่ๆ จะให้ผู้ใช้งานมีสิทธิ์ในข้อมูลของตนเองมากขึ้น และทำให้ผู้ใช้งานข้อมูลต้องพัฒนาระบบให้ผู้ใช้งานควบคุมข้อมูลของตนเองได้มากขึ้น
- ลงทุนและพัฒนา security protocol ใหม่ๆ เพื่อตามให้ทันเทคโนโลยีในอนาคต

ตัวอย่างการนำ Data Ethic และ Data Security ไปใช้งานในระบบ EIS ของตุรกี

- มีการควบคุมการนำเข้าและรวบรวมข้อมูลที่ระดับย่อย โดยมีการกำหนด standard เป็นแบบพื้นฐานและใช้งานกับข้อมูลที่นำเข้ามาจากหลายๆ แหล่ง
- วางระบบให้เป็นแหล่งข้อมูลที่น่าเชื่อถือสำหรับผู้ใช้ข้อมูล
- กำหนดการเข้าถึงข้อมูลรายผู้ใช้งาน และมีการตรวจสอบการใช้งานข้อมูลของแต่ละผู้ใช้
- มีการตรวจสอบความปลอดภัยของระบบอย่าสม่ำเสมอ
- ควบคุมสภาพแวดล้อมของจุดที่เข้าถึงข้อมูลภายในได้
    - ไม่มีการต่อ internet, ควบคุมการเข้า-ออกอาคาร, ไม่อนุญาตอุปกรณ์อิเล็กทรอนิกส์จากภายนอกในอาคาร
    - มีการตรวจสอบข้อมูลที่เผยแพร่ต่อบุคคลภายนอกว่าไม่มีข้อมูลส่วนบุคคล

หัวข้อที่ 3: Unlocking Productivity Gains for Small and Medium Enterprises โดย Dr. Murphy Choy, Chief Executive Officer, Alionova Consulting, Singapore

Data Science มีส่วนช่วยในการพัฒนา ตัดสินใจ และปรับปรุงกระบวนงานขององค์กรในช่วงที่ผ่านมา โดยเมื่อก่อน Data Science ถูกนำไปใช้กันแค่ในองค์กรใหญ่ๆ แต่ในช่วง 2-3 ปีที่ผ่านมา SME เริ่มมีการใช้ Data science กันมากขึ้น ซึ่งช่วยให้บริษัทขนาดเล็กมาก (2-3 คน) สามารถทำงานได้เทียบเท่ากับบริษัทที่มีขนาดใหญ่กว่า (15 คน) เพราะ process เกือบทั้งหมดถูกผลักดันด้วย data และ AI และสามารถดำเนินการได้โดยอัตโนมัติ

ความท้าทายในการนำ Data Science เข้าไปใช้งานกับ SME

- Lack of resources
    - งบไม่พอในการจัดหา tool ด้าน DS
        - มองหา open source tool, ใช้ tool ที่ซื้อไว้ทำอย่างอื่นอยู่แล้ว เช่น MS365
    - คอมพิวเตอร์ไม่สามารถประมวลผลงาน DS ได้
        - upgrade ระบบ, deploy ระบบบน cloud
    - Skill gap
        - จัด workshop ให้พนักงาน
        - ให้พนักงานอ่านเอกสารด้าน Data Science
    - พนักงานไม่มีเวลาศึกษาอย่างจริงจัง เพราะต้องทำหลายๆ อย่างพร้อมกัน
        - เลือกใช้เวลาให้คุ้มค่า
- Data silo: ข้อมูลที่มีไม่สามารถนำมาใช้ได้ เพราะคุณภาพต่ำ (ข้อมูลหาย, ข้อมูลไม่มีมาตรฐานกลาง, อื่นๆ)

    - สร้างจุดเก็บข้อมูลกลางขึ้นมา 1 จุดเพื่อเป็น Single source of truth แล้วใช้งานข้อมูลจากจุดนี้
    - ใช้แหล่งเก็บข้อมูลแบบง่ายๆ เช่น Excel เพื่อก่อให้เกิดการสร้างข้อมูล
- ไม่มีประสบการณ์การใช้งาน Data Science เพราะไม่มีงบให้พนักงานไปเรียน หรือไม่มีงานมากพอที่จะส่งพนักงานไปเรียนโดยเฉพาะ
    - ไม่ต้องจ้างผู้เชี่ยวชาญเฉพาะทาง แต่จ้างคนมาสอนให้พนักงานประจำแทน
    - สอนให้เอา data science ไปใช้กับงานที่กำลังทำอยู่ แบบ hands-on

ประโยชน์ของ Data Science ต่อ SME

- Productivity
    - เคสตัวอย่าง บริษัทซอฟต์แวร์ ใช้ GenAI และ DS ในระบบอีเมล
    - ทำนายความสำคัญและความด่วนของอีเมล, แผนกที่ควรส่งต่อ เมื่อมี inbox เข้ามา
    - ใช้ GenAI ช่วยร่างอีเมลตอบกลับ และมีการคำนวณความน่าจะเป็นที่ลูกค้าจะ OK กับเมลที่ตอบไป
- Efficiency
    - จากปกติการตอบอีเมลใช้เวลา 1.30 ชม (หาข้อมูล และเขียนเนื้อหาอีเมล) เหลือเพียง 3 นาที (review ครั้งสุดท้ายก่อนส่งเท่านั้น)
- Growth
    - ใช้ AI ช่วยในการหาช่องทางการเติบโตของ บ.

การใช้ข้อมูลช่วยในการตัดสินใจ (Data-driven Decision Making)

- ใช้งานกับธุรกิจแนว operational ได้ดี
- เคสตัวอย่าง: supermarket ขนาดเล็กจากจีนมาเปิดในสิงคโปร์ ใช้ข้อมูลในการช่วยตัดสินใจ ขยายสาขาจาก 2 สาขาเป็น 80 สาขาได้รวดเร็ว
    - แยกประเภทลูกค้า และดูว่าลูกค้าของตัวเองซื้ออะไร
        - พบว่าลูกค้าส่วนใหญ่เป็นคนจีน และเข้ามาเพื่อซื้อของแบรนด์จีน (เนื่องจากเป็น supermarket จากจีน)
        - จากข้อมูลดังกล่าว บริษัทเลยเน้นการนำเข้าของรูปแบบใหม่ๆ เข้ามาจากจีน ทำให้สามารถสร้างกลุ่มลูกค้าเฉพาะทางที่ต้องการของแบรนด์จีนมากขึ้น ทำให้แข่งกับ supermarket ขนาดใหญ่ได้
- Streamline operation ทำให้กระบวนการทำงานมีประสิทธิภาพมากขึ้น

Data Science Methodologies

- Exploratory Data Analysis
    - เรียนรู้ข้อมูลที่มีอยู่ หา pattern ในข้อมูลและทำการตั้งสมมุติฐาน
    - เป็นจุดเริ่มต้นที่ควรทำก่อนที่จะใช้ data science
    - เคสตัวอย่าง: บริษัทขายส่งอยากรู้ว่าเวลาส่วนลดมีการใช้งาน เกิดจาก salesman รีเปล่า
        - จากการสำรวจข้อมูล พบว่า salesman ไม่ได้ใส่ส่วนลดให้ตลอดเวลา ทำให้บางคนได้ของเยอะกว่าในราคาเท่ากัน
        - นอกจากนี้ ข้อมูลบางส่วนมีคุณภาพไม่ดี ทำให้ต้องเก็บข้อมูลเพิ่มเติม ทำให้ต้องแก้ไข data quality ก่อน
- Data visualization
    - Tool ที่แนะนำ: PowerBI (ถ้ามี MS365), Tableau Public
    - แดชบอร์ดทำให้เห็น insight จากข้อมูลได้ง่ายขึ้น
    - เป็นขั้นตอนหลังจากสำรวจข้อมูล
    - SME ส่วนใหญ่ทำถึงขั้นนี้ก็พอแล้ว
- Predictive Modeling

    - ทำนายอนาคตจากข้อมูลที่มี
    - SME ขนส่งส่วนใหญ่มีโมเดลไว้ทำนายสิ่งต่างๆ เช่น demand, โอกาสเกิดอุบัติเหตุ
- Prescriptive Analytics
    - ทำนายว่าขั้นตอนต่อไปควรทำอะไรถึงจะได้ผลลัพธ์ที่ดีที่สุด
    - ส่วนใหญ่ใช้กับบริษัทขนาดกลาง-ใหญ่ ในหลายๆ อุตสาหกรรม เช่น อาหาร รถยนต์
- Business Process Optimization
    - ปรับปรุงกระบวนการทางธุรกิจเพื่อผลลัพธ์ที่ดีขึ้น

วิธีการนำ Data Science มาใช้ในองค์กร SME: ปรับองค์กรให้วางแผนการดำเนินงานด้วยข้อมูล

- เช็คข้อมูลที่มีก่อนว่าสามารถใช้งานได้หรือไม่
    - เช็ค data source, data quality และเข้าใจ data flow
- พิจารณาว่า use case ไหนที่มีประโยชน์จากข้อมูลที่มี
    - ตั้งจุดประสงค์ให้เข้ากับ objective หลักของ บ และข้อมูลที่มี
        - อาจต้องปรับให้เข้ากับข้อมูล
    - วิเคราะห์ข้อมูลที่มีอยู่
    - เน้น use case เป็นหลัก
- วางแผนข้อมูลเพื่อขยาย use case เพิ่มเติม
- ปรับปรุงแผนไปเรื่อยๆ ระหว่างดำเนินการ
    - พยายามทำให้ข้อมูลน่าสนใจเพื่อผลักดันการพัฒนาต่อไปเรื่อยๆ

วิธีการสร้างวัฒนธรรมองค์กรให้เป็น data driven

- Commitment: ทำให้ผู้นำองค์กรยืนยันที่จะผลักดัน data driven
- Communication: ทำให้คนในองค์กรพูดคุยกันเรื่องข้อมูล ข้อมูลมาจากไหน เอาไปใช้ยังไงได้บ้าง
- Collaborate: ลดการสร้าง data silo ที่ตัวบุคคลในองค์กร และส่งเสริมการแบ่งปันข้อมูลภายในองค์กร

วิธีการใช้ความเชี่ยวชาญภายนอกมาช่วยผลักดัน Data science ในองค์กร

- Consultant: หาคำปรึกษาจากที่ปรึกษาภายนอก
- Partnering: หาคู่ค้าที่มีความเชี่ยวชาญด้าน data science
- Upskilling: จัดอบรมเพื่อพัฒนาสกิลด้านข้อมูลให้กับบุคลากรภายใน

ความท้าทายในการปรับเปลี่ยนวัฒนธรรมภายในองค์กร

- มักเกิดขึ้นกับองค์กรขนาดกลาง-ใหญ่
- ผู้เกี่ยวข้อง (stakeholder) บางส่วนไม่อยากเปลี่ยนแปลงตัวเอง และเล่นการเมืองใส่ ทำให้ผลักดันการเปลี่ยนแปลงได้ยากขึ้น
    - ต้องทำความเข้าใจว่าผู้เกี่ยวข้องแต่ละคนต้องการอะไร
    - ถ้าเจอผู้เกี่ยวข้องที่ไม่ยอมเปลี่ยนแปลง มีวิธีรับมือ 2 แบบ
        - ไปคุยกับผู้เกี่ยวข้องที่สนใจที่จะผลักดัน data science ให้เขาเห็นถึงประโยชน์
        - กดดันให้ผู้ที่ไม่ยอมเปลี่ยนแปลง ให้ยอมเปลี่ยน และทำให้เห็นผลลัพธ์จากการใช้ data science
- อบรมคนในองค์กรให้มีทักษะที่จำเป็น
- ค่อยๆ ปรับเปลี่ยนวัฒนธรรมเป็น data driven อย่าเปลี่ยนแบบรวดเร็วเกินไป
- Incremental implementation ค่อยๆ เปลี่ยนจากส่วนเล็กๆ ง่ายๆ ก่อน แล้วค่อยๆ ขยายผลจากข้อมูลที่สร้างมาและมีคุณภาพแล้ว

วิธีการวางแผนระบบ Data Science ที่สามารถขยายขนาดได้อย่างมั่นคง

- Deploy ระบบในที่ๆ ขยายได้ เช่น public cloud service
- Talent development สร้างผู้ใช้งานที่มีความสามารถในการทำ data science
- Continuous improvement พัฒนาโมเดลให้ดีขึ้นเรื่อยๆ ตามข้อมูล

- วางแผนเรื่องการจัดการเรื่องความเป็นส่วนตัว และความปลอดภัยของข้อมูลและโมเดล
- สร้าง AI ที่สามารถอธิบายขั้นตอนการคิดได้ เพื่อป้องกันการ hallucinate (คิดไปเอง)

หัวข้อที่ 4: Data-driven Decision-making and Business Models โดย Dr. Chintan Amrit, Senior Associate Professor, University of Amsterdam, the Netherlands

Data Science ในทางธุรกิจ คือการใช้ความรู้ทั้งด้าน analytics, business และ computer science ร่วมกันเพื่อสร้าง model ขึ้นมา

- ขั้นตอนในการสร้าง data science model จากข้อมูลมีดังนี้ (Knowledge Discovery Process)
    - นำเข้าข้อมูลดิบไปยัง Data Warehouse
    - ทำความสะอาดและแปลงรูปข้อมูลให้พร้อม train
    - หากฎหรือ pattern ของข้อมูล (data mining)
    - ตีความข้อมูลจากสิ่งที่ได้ และทำความเข้าใจ
- โดยปกติ database ต้นทางมักจะไม่ได้เก็บข้อมูลเป็นเวลานาน เพราะถูกออกแบบมาใช้งานในรูปแบบ transaction ถ้าต้องการใช้ข้อมูลปริมาณมากเพื่อนำไปวิเคราะห์ ควรใช้จาก data warehouse ที่มีการเก็บข้อมูล archive และออกแบบมาสำหรับงานวิเคราะห์ข้อมูลแทน

ความแตกต่างระหว่าง Statistics, Machine Learning และ Data Mining

- Statistics เน้นเรื่องการทดสอบสมมุติฐานตามทฤษฎี
- Machine Learning เน้นเรื่องการปรับปรุงโมเดลที่เกิดจากการเรียนรู้ของคอมพิวเตอร์
- Data Mining เน้นการสร้าง knowledge ทั้ง stack ตั้งแต่ cleaning จน visualize
    - มีการใช้ทั้ง stats และ ml

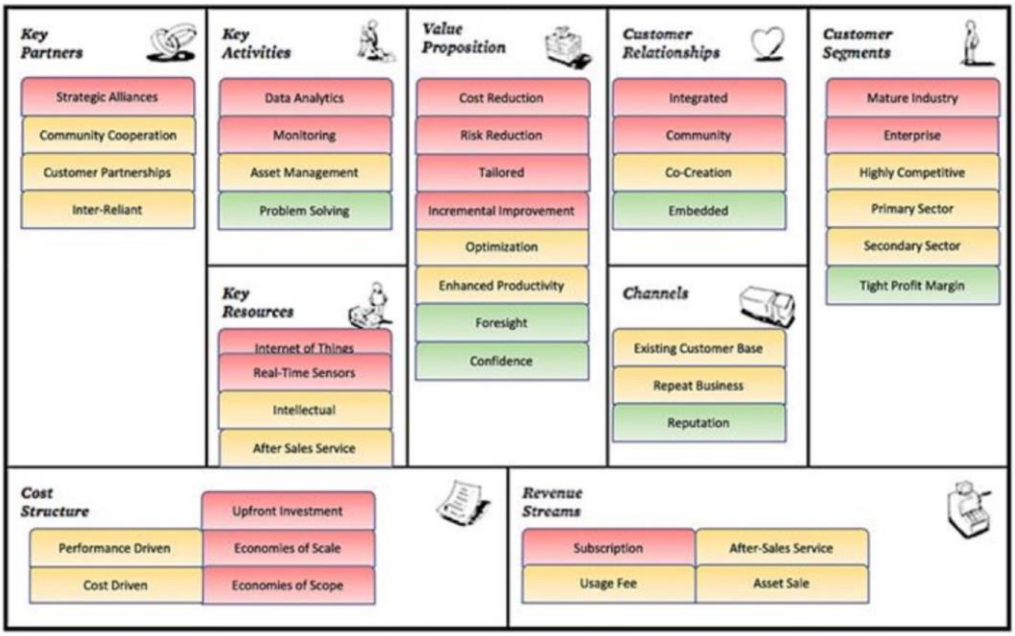
ประเภทของ Machine Learning

- Supervised มี input+label และ output เป็นการ mapping input เข้ากับ label
    - Label = สิ่งที่ต้องการให้ model predict ออกมา
- Unsupervised มีแค่ input และ output เป็นการ grouping
- Semi supervised มีการระบุ Label หลังจากทำการ grouping แล้ว
- Reinforcement มี input เป็น state และ action และ output เป็น state+action แล้วมีการ feedback กลับว่าผลลัพธ์ของ action ว่าดี-ไม่ดี เพื่อนำไป retrain
    - สอนรถให้ขับเองหรือสอนหุ่นยนต์ให้เดิน
    - ส่วนใหญ่ AI จะหมายถึงอันนี้

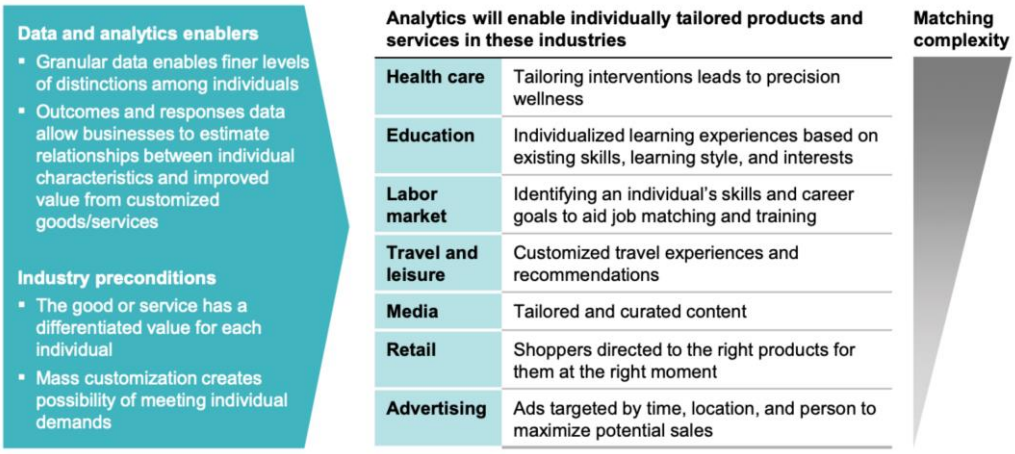
ประเภทของข้อมูลที่ใช้ในงาน data science

- Structured data ข้อมูลลักษณะตาราง
- Unstructured data ข้อมูลที่ไม่มี pattern ชัดเจน (รูป ข้อความ เสียง อื่นๆ)

The Business Model Canvas



- ตารางสำหรับวางแผนการเริ่มต้นธุรกิจ ใช้ร่างว่าธุรกิจของเรามีส่วนประกอบสำคัญอะไรบ้าง เช่น คู่ค้า กิจกรรมหลัก กลุ่มลูกค้า ช่องทางการขาย เป็นต้น แต่ในตารางดังกล่าวจะไม่มีส่วนที่เกี่ยวข้องกับ data science โดยตรง ดังนั้นในส่วนสีแดงจึงมีการเพิ่มสิ่งที่เกี่ยวข้องกับ data science เข้าไป เป็นตัวอย่างถึงวิธีการเพิ่ม data science เข้าไปในแก่นหลักของธุรกิจ



- ตารางแสดงความซับซ้อนถึงการนำ Data analytics และ data science เข้ามาใช้ในธุรกิจรูปแบบต่างๆ รวมถึงตัวอย่างการนำ data science มาใช้ในสาขาดังกล่าว เช่น โฆษณาแบบเจาะกลุ่มบุคคล แนะนำสื่อบันเทิง/แผนท่องเที่ยวตามความชอบขอบลูกค้า เป็นต้น

Data and analytics underpin six disruptive models, and certain characteristics make individual domains susceptible

Indicators of potential for disruption:

- Assets are underutilized due to inefficient signaling
- Supply/demand mismatch
- Dependence on large amounts of personalized data
- Data is siloed or fragmented
- Large value in combining data from multiple sources
- R&D is core to the business model
- Decision making is subject to human biases
- Speed of decision making limited by human constraints
- Large value associated with improving accuracy of prediction

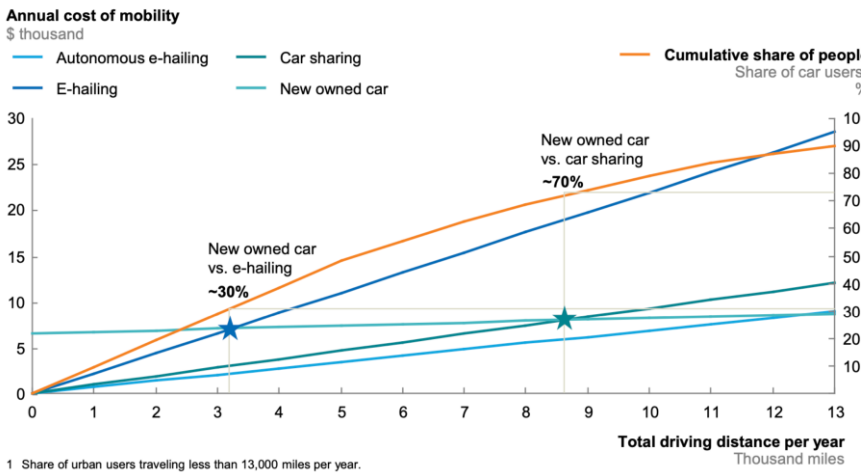
Archetype of disruption	Domains that could be disrupted
Business models enabled by orthogonal data	▪ Insurance ▪ Health care ▪ Human capital/talent
Hyperscale, real-time matching	▪ Transportation and logistics ▪ Automotive ▪ Smart cities and infrastructure
Radical personalization	▪ Health care ▪ Retail ▪ Media ▪ Education
Massive data integration capabilities	▪ Banking ▪ Insurance ▪ Public sector ▪ Human capital/talent
Data-driven discovery	▪ Life sciences and pharmaceuticals ▪ Material sciences ▪ Technology
Enhanced decision making	▪ Smart cities ▪ Health care ▪ Insurance ▪ Human capital/talent

- ตารางตัวอย่างธุรกิจรูปแบบใหม่ๆ ที่สามารถเกิดขึ้นได้ด้วยความรู้ด้าน data science

เพื่อผลักดันธุรกิจแบบใหม่ๆ เช่น smart city ต้องมีการเก็บและแบ่งปันข้อมูลที่มากเพียงพอที่จะสามารถนำมาสร้างโมเดลสำหรับ use case ดังกล่าวได้

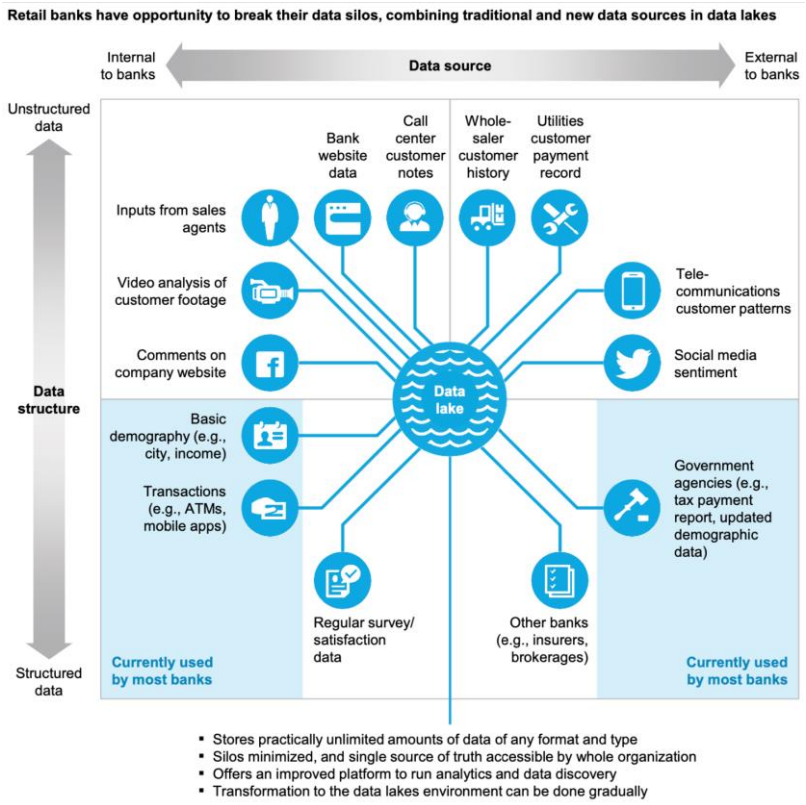
Study case: การหากลุ่มเป้าหมายสำหรับ e-hailing app ที่สหรัฐอเมริกา โดยการใช้กราฟเปรียบเทียบค่าใช้จ่ายเทียบกับระยะทางที่เดินทางทั้งหมดใน 1 ปี

Mobility services could drastically change the calculus of car ownership



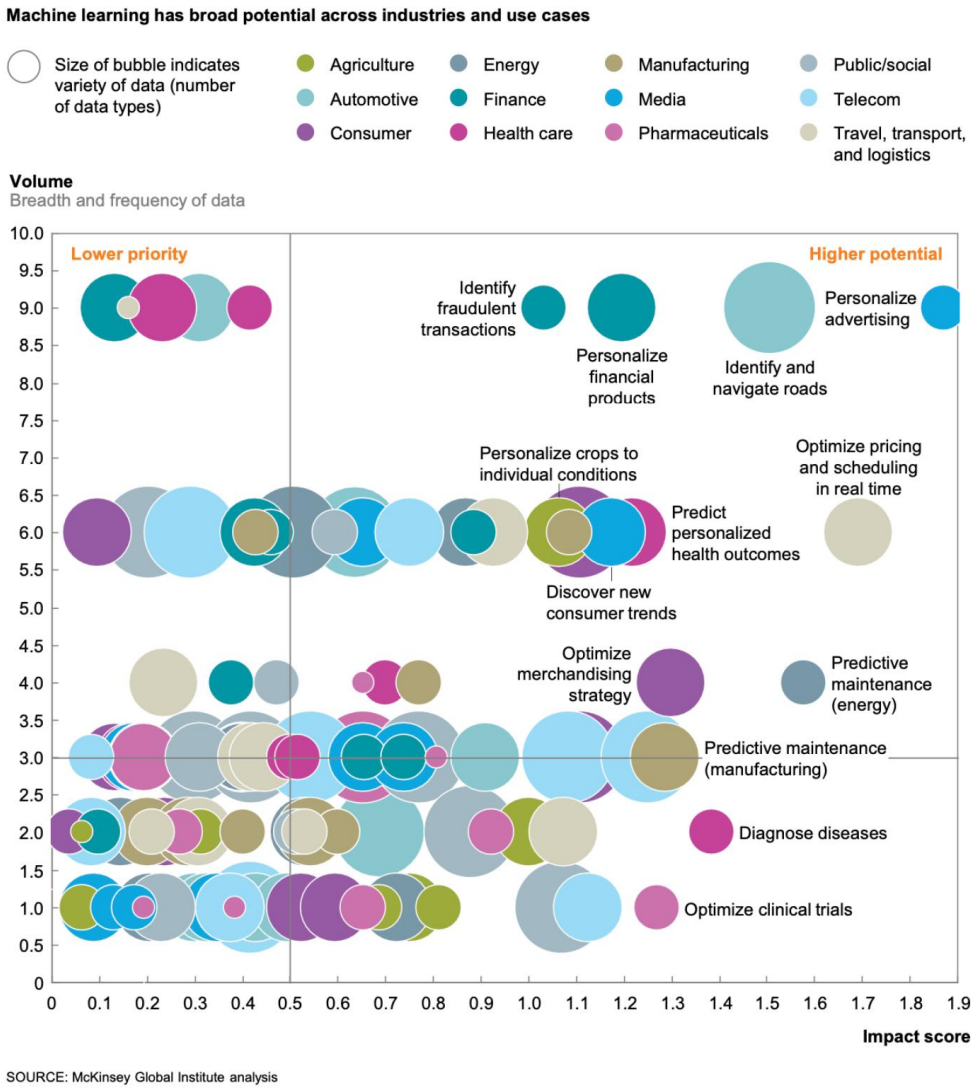
1 Share of urban users traveling less than 13,000 miles per year.

จากกราฟข้างต้น พบว่า ผู้ที่เดินทางน้อยกว่าปีละ 3,200 ไมล์ต่อปี ซึ่งคิดเป็น 30% ของกลุ่มตัวอย่าง จะได้ประโยชน์จากการใช้ e-hailing app มากกว่าการซื้อรถใช้เอง ดังนั้น e-hailing app จะสามารถวางแผนการตลาดให้ดึงดูดกลุ่มคนที่เดินทางน้อยเหล่านี้ได้



ตารางแสดงตัวอย่างข้อมูลสำหรับธุรกิจธนาคารที่สามารถใช้งานได้ในการทำ data science โดยธนาคารที่ยังใช้ฐานข้อมูลแบบเดิมจะสามารถเก็บข้อมูลที่สามารถใช้ประโยชน์ได้แค่เพียงบางส่วนเท่านั้น (structured data)

- ถ้าต้องการสร้างตลาดใหม่ ต้องเกิดการแบ่งปันข้อมูลกับคนอื่นๆ เพื่อสร้างมุมมองเพิ่มเติม แต่ต้องระวังเรื่องความปลอดภัยด้วยเช่นกัน



กราฟแสดงกลุ่มตลาดที่มีอยู่ในปัจจุบัน แบ่งตามปริมาณข้อมูลที่มีและผลกระทบต่อผู้ใช้

เคสตัวอย่างการใช้งาน data science

- ทำนายเวลาที่รถบรรทุกใช้ในการเดินทางไปยังจุดหมาย เพื่อลดจำนวนคิวรับ-ส่งของ
- ทำนายการทำร้ายร่างกายและทอดทิ้งเด็กจากข้อมูลสุขภาพของเด็ก
- วิเคราะห์การหาเสียงทางออนไลน์ และตรวจสอบการทุจริตเลือกตั้ง
- ทำนายเวลาเดินทางของเรือเดินสมุทร

หัวข้อที่ 5 (session 5-6): Hands-on Exercise: Data Visualization โดย Dr. Chintan Amrit Supported by Dr. Murphy Choy

เป็น session ที่แนะนำการใช้ Business Intelligent Tool (Tableau) เพื่อสร้างแดชบอร์ดสำหรับนำเสนอข้อมูล

- Excel ไม่เหมาะกับข้อมูลขนาดใหญ่ เนื่องจาก excel ไม่สามารถรองรับข้อมูลขนาดใหญ่ หรือข้อมูลที่มี column เยอะๆ ได้ โดย data warehouse จะรองรับข้อมูลขนาดใหญ่มากๆ ได้ดีกว่า
- ทำตาม Tableau Getting Started Guide https://help.tableau.com/current/guides/get-started-tutorial/en-us/get-started-tutorial-home.htm โดยรายละเอียดคร่าวๆ ได้แก่
    - นำเข้าข้อมูลจาก Superstore Sample
    - สร้าง Chart แสดงผลยอดขาย

- แบ่ง Chart แยกประเภทตามประเภทสินค้าและ Region
- ใส่สีใน Chart เพื่อให้เห็นข้อมูลในมิติอื่นๆ เพิ่มขึ้น (กำไร/ขาดทุน)
- สร้าง Map เพื่อแสดงข้อมูลยอดขายและกำไรตามตำแหน่ง location
- Drill down เพื่อมองไปถึง insight ของข้อมูล
- สร้าง Dashboard จาก sheet ที่สร้างไว้
- การตั้งค่าให้ element จาก sheet ใน dashboard สามารถ interact กับ sheet อื่นได้
- สร้าง Story เพื่อนำเสนอข้อมูลในรูปแบบและลำดับที่ต้องการ
- Publish Dashboard ที่ทำไว้ขึ้น Server Public หรือ Server ขององค์กร

หัวข้อที่ 6: Hands-on Exercise: Data Collection and Processing โดย Dr. Murphy Choy Supported by Dr. Chintan Amrit

เป็น session ที่แนะนำเรื่องวิธีการการนำเข้าข้อมูลและการเตรียมข้อมูลเพื่อนำไปใช้ในการวิเคราะห์

- ผู้ที่สร้าง data science model ไม่จำเป็นว่าจะต้องเป็นผู้ที่ความเข้าใจในข้อมูลที่นำมาทำโมเดลเสมอไป ดังนั้น ก่อนที่จะสร้างโมเดล ควรมีการพูดคุยกับผู้ที่เข้าใจในข้อมูลว่า ถ้าเราทำการสร้างโมเดลขึ้นมา ผลลัพธ์ที่คาดหวังควรมีลักษณะเป็นอย่างไร และใช้ความคาดหวังนี้เป็นจุดเริ่มต้นในการสร้างโมเดล
- การทำ data science สามารถแบ่งออกได้เป็น 2 ส่วน (data+science)
    - Data: data collection + data cleansing
    - Science: data analysis + model building
- Process cycle ของการทำ data science model
    - กำหนด Goal
    - เก็บรวบรวมข้อมูล
        - การเก็บข้อมูลที่มีประสิทธิภาพมีความสำคัญมาก เพราะความรวดเร็วของ model ขึ้นกับความเร็วในการเก็บข้อมูล
        - ควร automate ขั้นตอนการเก็บข้อมูลถ้าเป็นไปได้เพื่อความรวดเร็วและป้องกัน error ในข้อมูลที่เกิดจากคน (manual error)
    - สำรวจข้อมูลและเตรียมข้อมูล
    - สร้างและ train model
    - Evaluate และ deploy
    - 2 ขั้นตอนแรกมักจะใช้เวลา 60% ของเวลาทั้งหมด ขั้นตอนสำรวจข้อมูลและเตรียมข้อมูลใช้เวลาประมาณ 20% อีกสองขั้นตอนใช้เวลาขั้นตอนละประมาณ 10%
- องค์กรส่วนใหญ่มีโมเดล, บุคลากร และ software พร้อมแล้ว จุดต่างของ model ที่สามารถใช้งานได้สำเร็จคือความเร็วในการเข้าถึง insight

ตัวอย่าง data source และวิธีการในการทำ data collection (เก็บรวบรวมข้อมูล)

- web scraping ดึงข้อมูลจากสิ่งที่แสดงบนหน้าเว็บ
    - Parsehub: โปรแกรมสำหรับทำ web scraping โดยไม่ต้องเขียนโค้ด
- API ดึงข้อมูลจาก API
- ดึงข้อมูลตรงจาก database, data warehouse
    - Orange Data Mining tool
        - สามารถใช้งานเป็น ETL Tools ได้
            - Import from data source (sqlite3)
            - Data cleaning
        - ทำ Feature Selection

        - สร้าง ML Model (Segmentation)
- Sensor, IoT device

หัวข้อที่ 7: Hands-on Exercise: Leveraging Big Data and Analytics for Data-driven Operations โดย Dr. Murphy Choy Supported by Dr. Chintan Amrit

เป็น session ที่แนะนำการจัดเตรียมและวิเคราะห์ข้อมูล และการตีความผลลัพธ์ของการวิเคราะห์ข้อมูลดังกล่าวเพื่อนำไปใช้ในการตัดสินใจ

- โครงการที่เกี่ยวข้องกับ Big data มีเพียงแค่ 10% เท่านั้นที่ถือว่าเป็นโครงการด้าน Big data จริง โดยส่วนใหญ่จะเป็นโครงการ Data Analytic, Data Science ที่ใช้คำว่า Big data เสียมากกว่า
    - โครงการ Big data ไม่ใช่ว่าแค่มีข้อมูลเยอะๆ แล้วจะเป็น Big data เท่านั้น แต่จะต้องมีพลังในการประมวลผลและเกิดการตัดสินใจจากข้อมูลดังกล่าวด้วย จึงจะเป็นโครงการ Big data ที่แท้จริง

มุมมองในการวิเคราะห์ข้อมูล 4 ระดับ

- Descriptive: รายงานผลจากสิ่งที่เกิดขึ้นมาแล้ว
- Diagnostic: หาต้นเหตุของสิ่งที่เกิดขึ้น
- Predictive: ทำนายสิ่งที่จะเกิดขึ้นในอนาคต (เป็นความน่าจะเป็น)
- Prescriptive: แนะนำสิ่งที่ควรจะทำเพื่อผลลัพธ์ที่ต้องการ

ตัวอย่างการทำ prediction จากข้อมูล

- ทำนายแนวโน้มตลาดในอนาคต
- ทำนายความเสี่ยงและโอกาสที่อาจจะเกิดขึ้น
- ทำนายความต้องการลูกค้าเพื่อการบริการที่ตรงจุดมากขึ้น
- มองหาแนวทางในการปรับปรุงกระบวนงานให้มีประสิทธิภาพ
    - Track การใช้เวลาของพนักงานที่ work from home
- ทำนายสุขภาพและโอกาสที่จะเกิดโรคของบุคคล

ตัวอย่างการให้คำแนะนำจากข้อมูล

- ให้คำแนะนำในการปฏิบัติการกับทหารและประเทศต่างๆ ในช่วงสงคราม โดยใช้ข้อมูลจากทหารที่อยู่ในสงคราม
- Chatbot แบบมีสติปัญญาเพื่อลดการใช้พนักงานตอบกลับ

ส่วนที่ 2 ประโยชน์ที่ได้รับและการขยายผลจากการเข้าร่วมโครงการ

โปรดระบุประโยชน์ที่ได้รับจากการเข้าร่วมโครงการ โดยแบ่งเป็น

- ประโยชน์ต่อตนเอง
  - ทำให้เห็นภาพว่าการทำงานกับข้อมูลในมุมมองต่างๆ ที่นอกเหนือจากงานประจำที่ทำอยู่มีการนำไปใช้อย่างไร ช่วยให้มุมมองการมองประโยชน์จากข้อมูลที่เรามีเพิ่มมากขึ้น
- ประโยชน์ต่อหน่วยงานต้นสังกัด
  - ช่วยให้ผู้บริหารและผู้ที่เกี่ยวข้องกับการวิเคราะห์ข้อมูลได้รู้ถึง trend ของการวิเคราะห์ข้อมูลของประเทศต่างๆ ในปัจจุบัน
- ประโยชน์ต่อสายงานหรือวงการวิชาชีพในหัวข้อนั้นๆ
  - สามารถนำทักษะด้านข้อมูลที่เรียนรู้มาไปประยุกต์ใช้ในการทำงานจริงในสายงานที่ไม่ได้เกี่ยวข้องกับ IT โดยตรงได้
- กิจกรรมการขยายผลที่ได้ดำเนินการภายในระยะเวลา 60 วันนับจากวันสุดท้ายของโครงการ
  - จัดการอบรมแนว Lecture ให้กับคนในทีม Data ขององค์กรเพื่อถ่ายทอดความรู้ที่ได้
- กิจกรรมการขยายผลที่จะดำเนินการภายใน 6 เดือนหลังเข้าร่วมโครงการ
  - สร้าง Dashboard และ Machine Learning Model จากข้อมูลที่มีในองค์กรเพื่อนำไปใช้ต่อยอดประโยชน์ของข้อมูลที่จัดเก็บ

ส่วนที่ 3 เอกสารแนบ

- รายชื่อผู้เข้าร่วมโครงการและประเทศที่เข้าร่วมโครงการ
- กำหนดการฉบับล่าสุด (Program)
- เอกสารประกอบการประชุม/สัมมนา (Training Materials)