



กนกวรรณ กรองแก้ว นักวิทยาศาสตร์
สำนักพัฒนาศักยภาพนักวิทยาศาสตร์ปฏิบัติการ

Big Data คือ ?

ปัจจุบันที่โลกถูกขับเคลื่อนด้วยข้อมูลผู้คนที่ต้องพึ่งพิงข้อมูลในการตัดสินใจต่างๆกระบวนการคิดของคนทั่วไปหรือแม้แต่องค์กรต่างๆ ในทุกวันนี้ส่วนใหญ่อยู่บนฐานของการรวบรวมข้อมูลประมวลผลเพื่อประกอบการตัดสินใจทำอะไรบางอย่าง แต่ด้วยความที่ข้อมูลในปัจจุบันมีอยู่มากมายและกระจุกกระจายคำถามสำคัญ คือ เราจะเข้าถึงข้อมูลมหาศาลเหล่านั้นและใช้ข้อมูลเหล่านั้นให้เป็นประโยชน์ได้อย่างไรจากคำถามดังกล่าวจึงทำให้คนหันมาสนใจสิ่งที่เรียกว่า Big data กันมากขึ้นจนกลายเป็นกระแสสังคมในปัจจุบัน

What is big data?

Big data คือ ข้อมูลที่ประกอบด้วยคุณลักษณะ 4 อย่างคือ

1. Volume—size ของข้อมูลมีขนาดใหญ่ มีปริมาณข้อมูลมาก ซึ่งสามารถเป็นได้ทั้งข้อมูลแบบ offline หรือ online
2. Variety—ข้อมูลมีความหลากหลาย สามารถเป็นได้ทั้งที่มีโครงสร้างและข้อมูลที่ไม่สามารถจับ pattern ได้
3. Velocity—ข้อมูลมีการเปลี่ยนแปลงตลอดเวลาอย่างรวดเร็ว มีการส่งผ่านข้อมูลอย่างต่อเนื่องในลักษณะ streaming ทำให้การวิเคราะห์ข้อมูลแบบ manual มีข้อจำกัด
4. Veracity—ข้อมูลมีความไม่ชัดเจน (untrusted, uncleaned)

รูปแบบของข้อมูลของ big data สามารถเป็นไปได้หลากหลายตั้งแต่

1. Behavioral data: ข้อมูลเชิงพฤติกรรมกรรมการใช้งานต่างๆ เช่น server log, พฤติกรรมการคลิกดูข้อมูล, ข้อมูลการใช้ ATM เป็นต้น

2. Image & sounds: ภาพถ่าย, วิดีโอ, รูปจาก google street view, ภาพถ่ายทางการแพทย์, ลายมือ, ข้อมูลเสียงที่ถูกรับบันทึกไว้ เป็นต้น

3. Languages: text message, ข้อความที่ถูกรับ tweet, เนื้อหาต่างๆในเว็บไซต์ เป็นต้น

4. Records: ข้อมูลทางการแพทย์, ข้อมูลผลสำรวจ ที่มีขนาดใหญ่, ข้อมูลทางภาษี เป็นต้น

5. Sensors: ข้อมูลอุณหภูมิ, accelerometer, ข้อมูลทางภูมิศาสตร์ เป็นต้น

ให้ข้อมูลตัวชี้วัดของผลลัพธ์ที่อาจจะเกิดขึ้นถ้าแนวโน้มยังเป็นอยู่เช่นนี้ต่อไป โดยผลการวิเคราะห์นี้อาจออกมาในรูปแบบของ

- Statistical analysis: “ทำไมถึงเกิดเหตุการณ์นี้”
- Randomized testing: “จะเกิดอะไรขึ้นถ้าเราทดลองทำวิธีการนี้”
- Predictive modeling: “จะเกิดอะไรขึ้นต่อไป”
- Optimization: “อะไรคือสถานการณ์ที่ดีที่สุดที่จะเกิดขึ้น”

Big data analytics

การวิเคราะห์ข้อมูล Big data อาศัยหลักการพื้นฐานบางอย่างเพื่อพัฒนาเป็นเทคนิคในการดึงข้อมูลสำคัญออกจากชุดฐานข้อมูลขนาดใหญ่ที่เต็มไปด้วยข้อมูลหลากหลายรูปแบบ เพื่อนำมาหา pattern ของข้อมูลที่ไม่เห็นด้วยตาเปล่า หารูปแบบความสัมพันธ์ที่ซ่อนอยู่ หาแนวโน้ม การตลาด เทรนด์ความชอบของลูกค้า และข้อมูลอื่นๆที่เป็นประโยชน์ทางธุรกิจผลจากการวิเคราะห์ข้อมูล Big data ทำให้มีข้อมูลที่เป็นข้อเท็จจริงซึ่งผ่านการวิเคราะห์อย่างเป็นระบบเพื่อใช้ประกอบการตัดสินใจ โดยระดับของการวิเคราะห์ที่กันได้หลากหลาย แล้วแต่รูปแบบการนำไปใช้งาน

1. Descriptive analytics

เป็นการวิเคราะห์ในระดับที่บอกว่าเกิดอะไรขึ้นจำนวนเท่าไร ถัดแค่นั้น เกิดเหตุการณ์สำคัญๆตอนไหนตรงไหนบ้าง เป็นต้น โดยสามารถทำในรูปแบบของ

- Standard report: “เกิดอะไรขึ้น”
- Ad hoc report: “จำนวนเท่าไร บ่อยแค่ไหนที่ไหน”
- Query: “อะไรคือปัญหาที่แท้จริง”
- Alerts: “ต้องเกิด action อะไร”

2. Predictive analytics

เป็นการวิเคราะห์ในลักษณะที่ซับซ้อนขึ้นไปอีกขั้นหนึ่งคือ เป็นการประเมินว่าจะเกิดอะไรขึ้นต่อไป มีการ

3. Prescriptive analytics

เป็นการสังเคราะห์ข้อมูลเพื่อวิเคราะห์แนวโน้มและเสนอทางเลือกในการตัดสินใจที่เหมาะสมกับการคาดการณ์บนฐานของข้อมูล



เอกสารอ้างอิง

- เข้าถึงจาก <https://blog.goodfactory.co/big-data>
- เข้าถึงจาก <http://www.brandquarterly.com/big-analytics-automation-evolution-big-data>