

Summary Note จากหลักสูตร

Practical Data Mining with RapidMiner Studio

Chapter 1

- Big Data ประกอบด้วย 3V ได้แก่
 - **Volume** คือ มีขนาดใหญ่
 - **Velocity** คือ เกิดขึ้นอย่างรวดเร็ว
 - **Variety** คือ มีความหลากหลาย
- การหาความสัมพันธ์ระหว่างสภาพอากาศกับยอดขายทำให้ช่วยลดของที่เหลือลงได้
- แต่ละวันเราสร้างข้อมูลขึ้นมาเยอะมาก ไม่ว่าจะเป็นแบบ online และ offline
- เราสามารถแบ่งข้อมูลตามแหล่งที่อยู่ได้ 2 แบบ คือ
 - **ข้อมูลภายในองค์กร** (Internal Data Sources) ซึ่งส่วนใหญ่จะเป็นข้อมูลในรูปแบบของตาราง
 - **ข้อมูลภายนอกองค์กร** (External Data Sources) ซึ่งส่วนใหญ่จะเป็นข้อมูลในรูปแบบข้อความ รูปภาพ
 - การเก็บข้อมูลแบบ **Excel ทำได้ง่าย** แต่มี**ข้อเสียบ้าง** เช่น **ข้อมูลซ้ำซ้อน** กับ **ข้อมูลไม่ตรงกัน**

facebook.com/datacube.th

Summary Note จากหลักสูตร

Practical Data Mining with RapidMiner Studio

Chapter 1

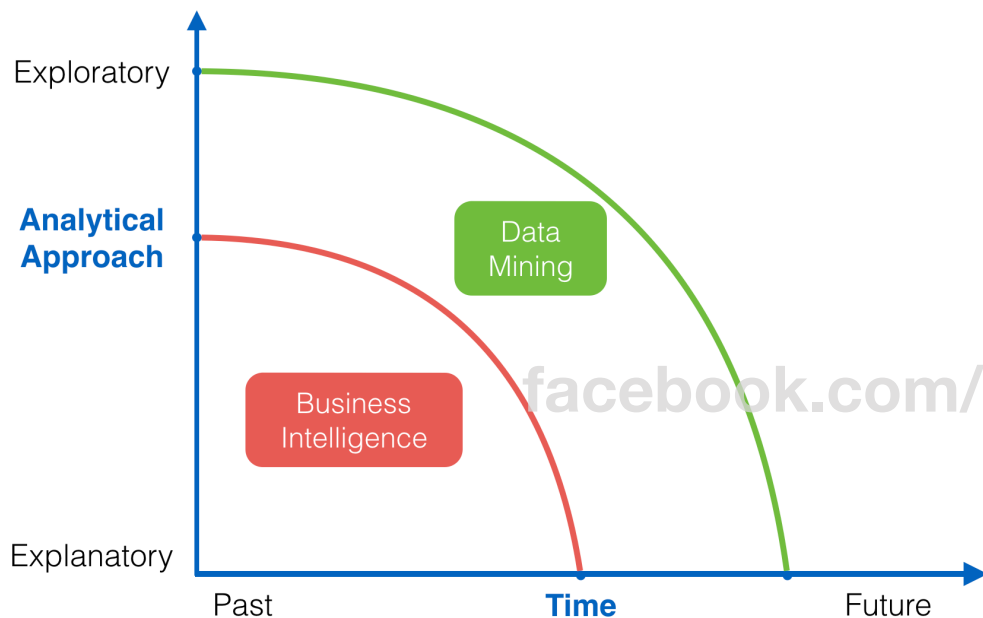
- ข้อมูลแต่ละฝ่ายจึงควรจัดเก็บเป็น **ฐานข้อมูล** เพื่อใช้ตอบโจทย์ของฝ่ายตัวเอง
- แต่จะยุ่งยากในการดึงข้อมูลมาสร้าง **dashboard** หรือ **รายงาน** ให้ผู้บริหาร **ดูภาพรวมขององค์กร**
- มี concept ของ **คลังข้อมูล** (data warehouse) ที่ **รวบรวมข้อมูลจากหลายๆ ฐานข้อมูล** มารวมกัน
- ข้อมูลต้องทำ **ETL (Extract, Transform, Load)** ก่อนนำเข้ามาใส่ในคลังข้อมูล
- สามารถดึงข้อมูลจากคลังข้อมูลออกมาวิเคราะห์ในรูปแบบของ **Business Intelligence (BI)**
- และสามารถวิเคราะห์ในเชิงลึก เช่น **Data Mining (DM)** หรือ **Machine Learning (ML)** ได้
- เพื่อให้แสดงผลการวิเคราะห์ได้ง่ายขึ้นหลังจากวิเคราะห์ด้วยเทคนิค DM หรือ ML แล้ว **ควรนำผลลัพธ์มาแสดงผลในระบบ BI** ด้วย

Summary Note จากหลักสูตร

Practical Data Mining with RapidMiner Studio

Chapter 1

- ความแตกต่างของ BI และ DM/ML คือ
- BI จะเป็น**การสรุปสิ่งที่เกิดขึ้นในอดีต**
เช่น ยอดขายของแต่ละสาขาในเดือนที่ผ่านมา
- DM/ML จะสามารถ**คาดการณ์เหตุการณ์ในอนาคตได้** โดยการใช้ข้อมูลในอดีตที่ผ่านมา



- Data Mining คือ การ**ค้นหารูปแบบที่สำคัญ**จากฐานข้อมูลขนาดใหญ่
- Machine Learning คือ การ**เรียนรู้จากข้อมูลในอดีตเพื่อสร้างแบบจำลอง (Model) คาดการณ์**สิ่งที่จะเกิดขึ้นในอนาคต
- ตัวอย่างการประยุกต์ใช้ เช่น
 - การ**แนะนำสินค้าให้กับลูกค้าที่ตั้ง****ครรภ์**ของห้าง **Target** โดยการดูจากพฤติกรรมการณ์ซื้อสินค้าที่ผ่านมา

Summary Note จากหลักสูตร

Practical Data Mining with RapidMiner Studio

Chapter 1

- ตัวอย่างการประยุกต์ใช้ เช่น
 - การ**แนะนำสินค้าของเว็บไซต์** **amazon.com** เพื่อต้องการเพิ่มยอดขาย
 - การ**แนะนำภาพยนตร์หรือซีรีส์ที่น่าสนใจ**ให้กับสมาชิกใน Netflix
 - การทำ **sentiment analysis** เพื่อ**หาทัศนคติจากข้อความออนไลน์** เพื่อบอกว่าคนเขียนชอบหรือไม่ชอบ ตัวอย่างระบบเช่น Mandala, Zocial Eye, Zanroo facebook.com/datacube.th
- ตัวอย่างการประยุกต์ใช้ เช่น
 - งานวิจัยเรื่อง “Widespread Worry and the Stock Market”
 - เป็นการหา**ความสัมพันธ์ระหว่างโลก online กับโลก offline** โดยการ**ดึงข้อมูลข้อความจากเว็บไซต์ blogger** แล้วหาว่าในวันนั้นคนมี**ความเครียดมากน้อย**แค่ไหน

Summary Note จากหลักสูตร

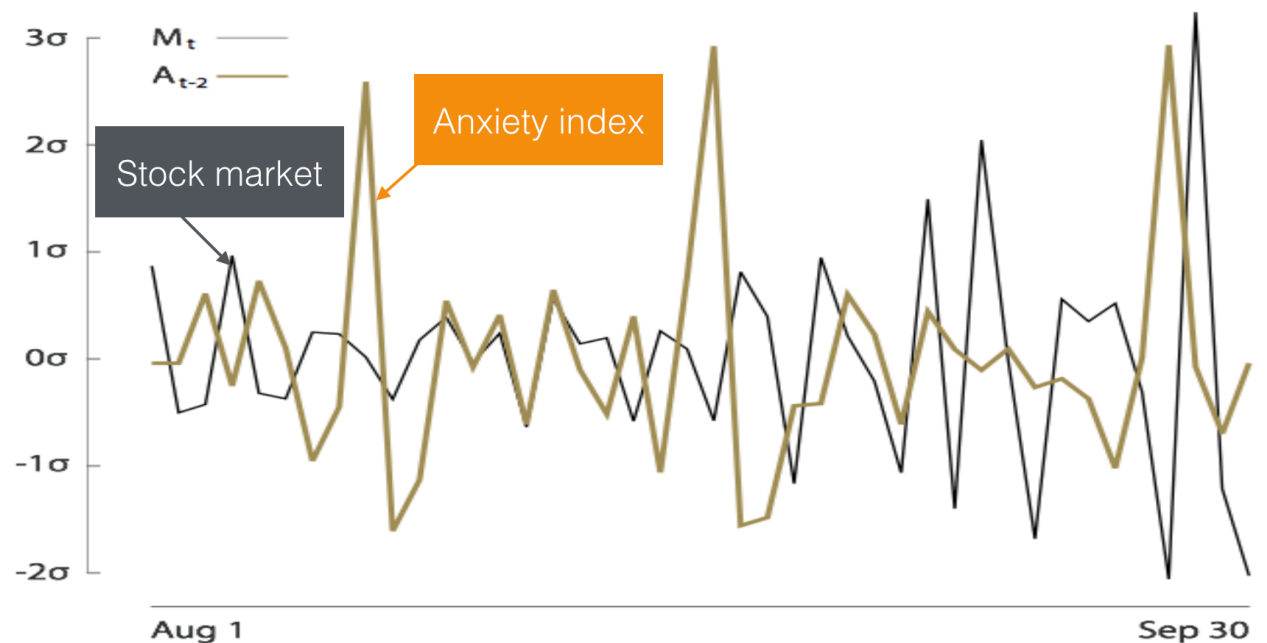
Practical Data Mining with RapidMiner Studio

Chapter 1

- ตัวอย่างการประยุกต์ใช้ เช่น
 - งานวิจัยเรื่อง “Widespread Worry and the Stock Market”
 - งานวิจัยพบว่า **ถ้าคนในโลก**

online เครียดมากอีกประมาณ

2-3 วัน หุ่นจะตกลงมา



Reference: Eric Gilbert, Karrie Karahalios: Widespread Worry and the Stock Market. ICWSM 2010

facebook.com/datacube.th