

ใบความรู้

หน่วยที่ 1 หลักการพื้นฐานในการวิเคราะห์ข้อมูล (Data Analytics)

การวิเคราะห์ข้อมูล (Data Analytics) เป็นกระบวนการของการตรวจสอบ การทำความสะอาด การแปลง และการสร้างแบบจำลองข้อมูล เพื่อค้นพบข้อมูลที่เป็นประโยชน์ การสรุปผล และสนับสนุนการตัดสินใจ ขั้นตอนพื้นฐานประกอบด้วย 7 ส่วนหลักดังนี้

1.1 เลือกข้อมูลที่เหมาะสม

ขั้นตอนนี้คือการระบุและรวบรวม ข้อมูลดิบ (Raw Data) ที่มีความเกี่ยวข้องและจำเป็นต่อวัตถุประสงค์ของการวิเคราะห์ ข้อมูลที่ดีควรมีคุณสมบัติดังนี้:

- ความเกี่ยวข้อง (Relevance): ข้อมูลต้องสัมพันธ์โดยตรงกับคำถามหรือปัญหาที่ต้องการหาคำตอบ
- ความสมบูรณ์ (Completeness): ข้อมูลควรมีองค์ประกอบที่จำเป็นครบถ้วน ไม่ขาดหาย
- ความถูกต้อง (Accuracy): ข้อมูลต้องเป็นจริงและเชื่อถือได้
- ความทันสมัย (Timeliness): ข้อมูลควรเป็นปัจจุบันเพียงพอต่อการใช้งาน

1.2 ทำความเข้าใจข้อมูล

หลังจากได้ข้อมูลมาแล้ว ต้องทำการสำรวจ โครงสร้าง (Structure) และ คุณภาพ (Quality) ของข้อมูล เพื่อทราบลักษณะเบื้องต้นและปัญหาที่อาจเกิดขึ้น:

- การสำรวจข้อมูล (Exploratory Data Analysis - EDA): ใช้สถิติพื้นฐาน (เช่น ค่าเฉลี่ย มัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน) และการสร้างภาพข้อมูล (เช่น ฮิสโตแกรม แผนภาพกล่อง) เพื่อทำความเข้าใจการกระจายตัว แนวโน้ม และความผิดปกติของข้อมูล
- การระบุปัญหา: ตรวจสอบหา ค่าที่ขาดหายไป (Missing Values), ค่าผิดปกติ (Outliers), ความไม่สอดคล้องกัน (Inconsistencies), และ รูปแบบข้อมูลที่ไม่ถูกต้อง (Incorrect Data Types)

1.3 การทำความเข้าใจความสัมพันธ์ระหว่างข้อมูล

ขั้นตอนนี้มุ่งเน้นการค้นหา ความสัมพันธ์ (Relationships) และ รูปแบบ (Patterns) ในข้อมูล ซึ่งเป็นหัวใจสำคัญของการวิเคราะห์:

- ความสัมพันธ์เชิงสาเหตุ (Causation): การระบุว่าการเปลี่ยนแปลงในตัวแปรหนึ่ง เป็นสาเหตุ ให้เกิด การเปลี่ยนแปลงในตัวแปรอื่น

- ความสัมพันธ์ร่วม (Correlation): การวัดความเข้มแข็งและทิศทางของความสัมพันธ์ระหว่างสองตัวแปร (เช่น สหสัมพันธ์แบบเพียร์สัน - Pearson Correlation)
- การจัดกลุ่ม (Clustering): การแบ่งข้อมูลออกเป็นกลุ่มย่อยที่มีลักษณะคล้ายกัน (เช่น กลุ่มลูกค้าที่มีพฤติกรรมคล้ายกัน)
- การจัดประเภท (Classification): การทำนายประเภทหรือหมวดหมู่ของข้อมูลใหม่โดยอิงจากข้อมูลที่มีอยู่

1.4 การใช้เครื่องมือและเทคนิคทางสถิติและคณิตศาสตร์

การประยุกต์ใช้หลักการทางสถิติและคณิตศาสตร์เพื่อการวิเคราะห์ที่ลึกซึ้งและแม่นยำ:

ประเภทเทคนิค	ตัวอย่างเทคนิคและเครื่องมือ	วัตถุประสงค์
สถิติเชิงพรรณนา (Descriptive Stats)	ค่าเฉลี่ย, มัธยฐาน, ส่วนเบี่ยงเบนมาตรฐาน, ความถี่	สรุปและอธิบายลักษณะของชุดข้อมูล
สถิติเชิงอนุมาน (Inferential Stats)	การทดสอบสมมติฐาน (Hypothesis Testing), การวิเคราะห์ความแปรปรวน (ANOVA)	อนุมานผลจากกลุ่มตัวอย่างไปยังประชากรทั้งหมด
การสร้างแบบจำลอง (Modeling)	การถดถอยเชิงเส้น (Linear Regression), การถดถอยโลจิสติก (Logistic Regression)	สร้างแบบจำลองเพื่อทำนายหรืออธิบายความสัมพันธ์
คณิตศาสตร์	พีชคณิตเชิงเส้น (Linear Algebra), แคลคูลัส (Calculus)	รองรับการทำงานของอัลกอริทึมการเรียนรู้ของเครื่อง (Machine Learning) และการปรับค่าที่เหมาะสมที่สุด (Optimization)

1.5 การใช้เทคโนโลยีที่เหมาะสม

การเลือกใช้เครื่องมือและเทคโนโลยีเพื่อจัดการข้อมูลขนาดใหญ่และทำการวิเคราะห์ได้อย่างมีประสิทธิภาพ:

- ภาษาโปรแกรม: Python (ใช้ไลบรารีเช่น Pandas, NumPy, Scikit-learn) และ R (เน้นการวิเคราะห์ทางสถิติ)
- ซอฟต์แวร์สถิติ/สเปรดชีต: Microsoft Excel, SPSS, SAS

- ฐานข้อมูล: SQL (สำหรับฐานข้อมูลเชิงสัมพันธ์ - Relational Databases), NoSQL (สำหรับข้อมูลที่ไม่เป็นโครงสร้าง)
 - เครื่องมือแสดงผลข้อมูล (Data Visualization Tools): Tableau, Power BI, Google Data Studio
 - แพลตฟอร์มบิกดาต้า: Hadoop, Spark (สำหรับข้อมูลขนาดใหญ่)
-

1.6 การตีความและการสร้างความเข้าใจ

การเปลี่ยนผลลัพธ์ทางสถิติและตัวเลขให้เป็น ความรู้ (Knowledge) และ ข้อมูลเชิงลึก (Insights) ที่นำไปปฏิบัติได้:

- การเชื่อมโยงกับวัตถุประสงค์: ตีความผลลัพธ์โดยเชื่อมโยงกลับไปคำถามเริ่มต้นของการวิเคราะห์
 - การระบุผลกระทบ: ทำความเข้าใจว่าข้อมูลเชิงลึกที่ได้จะส่งผลกระทบต่อการตัดสินใจหรือธุรกิจอย่างไร
 - การตรวจสอบความน่าเชื่อถือ: พิจารณาข้อจำกัดของข้อมูลและแบบจำลอง (เช่น ข้อมูลเพียงพอหรือไม่, แบบจำลองมีความถูกต้องแค่ไหน)
-

1.7 การนำเสนอผลลัพธ์

การสื่อสารข้อมูลเชิงลึกที่ค้นพบอย่างชัดเจน มีประสิทธิภาพ และน่าสนใจแก่ผู้ที่เกี่ยวข้อง:

- กลุ่มเป้าหมาย: ปรับรูปแบบการนำเสนอให้เหมาะสมกับความรู้และระดับของผู้ฟัง (เช่น ผู้บริหารอาจต้องการสรุปสั้นๆ ในขณะที่นักวิเคราะห์ต้องการรายละเอียดทางเทคนิค)
- การเล่าเรื่องด้วยข้อมูล (Data Storytelling): จัดเรียงข้อมูลเชิงลึกให้เป็นเรื่องราวที่น่าติดตาม มีจุดเริ่มต้น (ปัญหา), เนื้อหา (การวิเคราะห์), และบทสรุป (ข้อเสนอแนะ)
- การแสดงภาพข้อมูล (Visualization): ใช้แผนภูมิ (Charts), กราฟ (Graphs), และแดชบอร์ด (Dashboards) ที่เหมาะสม เพื่อให้ข้อมูลซับซ้อนเข้าใจง่ายขึ้น (เช่น แผนภูมิแท่ง, แผนภูมิวงกลม, แผนภาพกระจาย)