

บทที่ 1 บทนำสู่วิทยาศาสตร์ข้อมูลและการวิเคราะห์ข้อมูล

(Introduction to Data Science and Data Analytics)

ในยุคปัจจุบัน ข้อมูลมีอยู่รอบตัวทุกที่ ไม่ว่าจะเป็นในธุรกิจ เทคโนโลยี การตลาด หรือแม้กระทั่งชีวิตประจำวันของ การวิเคราะห์ข้อมูลจึงมีบทบาทสำคัญในการเปลี่ยนข้อมูลดิบเหล่านั้นให้กลายเป็นข้อมูลเชิงลึกที่สามารถนำไปใช้ในการตัดสินใจได้

การวิเคราะห์ข้อมูลแบ่งออกได้หลายประเภท เช่น การวิเคราะห์เชิงสถิติ การทำนายอนาคต (Predictive Analytics) และการหาความสัมพันธ์ในข้อมูล วิธีการเหล่านี้สามารถช่วยให้ค้นพบรูปแบบที่ซ่อนอยู่ในข้อมูล เช่น แนวโน้มในตลาด ความสัมพันธ์ระหว่างตัวแปรต่าง ๆ และการทำนายผลลัพธ์ในอนาคต

Python เป็นภาษาที่ได้รับความนิยมอย่างมากในการวิเคราะห์ข้อมูล มีไลบรารีที่ทรงพลัง เช่น Pandas สำหรับการจัดการข้อมูล, NumPy สำหรับการคำนวณทางคณิตศาสตร์, Matplotlib และ Seaborn สำหรับการสร้างกราฟ และ scikit-learn สำหรับการใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งช่วยให้กระบวนการวิเคราะห์ข้อมูลมีความรวดเร็วและมีประสิทธิภาพมากขึ้น

การวิเคราะห์ข้อมูลไม่เพียงแต่ช่วยในการตัดสินใจ แต่ยังช่วยให้สามารถปรับปรุงกระบวนการธุรกิจ วางกลยุทธ์ที่มีประสิทธิภาพ และคาดการณ์อนาคตได้ดียิ่งขึ้น

1. ความหมายและความสำคัญของการวิเคราะห์ข้อมูล

1.1 ความหมายของการวิเคราะห์ข้อมูล (Data Analytics)

การวิเคราะห์ข้อมูลหมายถึงกระบวนการรวบรวม, จัดระเบียบ, และตรวจสอบข้อมูลเพื่อค้นหาความสัมพันธ์ รูปแบบ และแนวโน้มที่ซ่อนอยู่ในข้อมูล การวิเคราะห์ข้อมูลมีเป้าหมายหลักคือการสร้างข้อมูลเชิงลึก (Insights) ที่สามารถนำไปใช้ในการตัดสินใจทางธุรกิจหรือการแก้ไขปัญหาต่าง ๆ ได้อย่างมีประสิทธิภาพ นอกจากนี้ การวิเคราะห์ข้อมูลยังเป็นเครื่องมือที่ช่วยในการคาดการณ์แนวโน้มในอนาคตจากข้อมูลในอดีต การวิเคราะห์ข้อมูลแบ่งออกได้เป็นหลายประเภท ได้แก่

1.1.1 การวิเคราะห์เชิงพรรณนา (Descriptive Analytics) เป็นการวิเคราะห์ข้อมูลเพื่อให้เข้าใจเหตุการณ์ที่เกิดขึ้นในอดีตหรือปัจจุบัน เช่น การรายงานยอดขายของเดือนที่ผ่านมา หรือการสรุปข้อมูลทางสถิติพื้นฐาน เช่น ค่าเฉลี่ย และค่ามัธยฐาน

1.1.2 การวิเคราะห์เชิงวินิจฉัย (Diagnostic Analytics): เป็นการวิเคราะห์ข้อมูลเพื่อหาเหตุผลหรือปัจจัยที่ทำให้เกิดเหตุการณ์ใดเหตุการณ์หนึ่ง เช่น การวิเคราะห์สาเหตุที่ทำให้ยอดขายตกในบางช่วงเวลา

1.1.3 การวิเคราะห์เชิงพยากรณ์ (Predictive Analytics): ใช้ข้อมูลดีในการทำนายอนาคต โดยอาศัยโมเดลทางคณิตศาสตร์และการเรียนรู้ของเครื่อง (Machine Learning) เช่น การพยากรณ์ยอดขายของสินค้าภายในไตรมาสถัดไป

1.1.4 การวิเคราะห์เชิงกำหนด (Prescriptive Analytics): เป็นการวิเคราะห์เพื่อหาแนวทางที่ดีที่สุดในการดำเนินการ โดยใช้ผลการวิเคราะห์ข้อมูลร่วมกับแบบจำลองการตัดสินใจ (Decision Models) เพื่อให้ได้คำแนะนำที่สามารถนำไปใช้ได้จริง

1.2 ความสำคัญของการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลมีบทบาทสำคัญในโลกที่เต็มไปด้วยข้อมูลอย่างปัจจุบัน องค์กรและธุรกิจต่าง ๆ ต้องพึ่งพาการวิเคราะห์ข้อมูลเพื่อสร้างความได้เปรียบในการแข่งขัน ความสำคัญของการวิเคราะห์ข้อมูลสามารถอธิบายได้ในแง่ต่าง ๆ ดังนี้

1.2.1 การตัดสินใจที่แม่นยำและมีประสิทธิภาพมากขึ้น (Data-Driven Decision Making) การวิเคราะห์ข้อมูลช่วยให้ผู้บริหารสามารถตัดสินใจโดยอาศัยข้อเท็จจริงที่ได้จากข้อมูลจริง แทนที่จะใช้การคาดเดาหรือการตัดสินใจจากประสบการณ์เพียงอย่างเดียว เช่น การวิเคราะห์ยอดขายของผลิตภัณฑ์เพื่อกำหนดกลยุทธ์ทางการตลาดใหม่ ๆ หรือการวางแผนการผลิตสินค้า

1.2.2 การเพิ่มประสิทธิภาพของกระบวนการทำงาน (Optimizing Operations) การวิเคราะห์ข้อมูลช่วยให้องค์กรสามารถปรับปรุงกระบวนการทำงานได้อย่างมีประสิทธิภาพ โดยการวิเคราะห์ข้อมูลการผลิตหรือข้อมูลการดำเนินงาน เช่น การค้นหาความล่าช้าในการผลิต และการหาทางแก้ไขปัญหที่เกิดขึ้น

1.2.3 การคาดการณ์แนวโน้มในอนาคต (Predicting Future Trends) การวิเคราะห์เชิงพยากรณ์ช่วยให้องค์กรสามารถคาดการณ์เหตุการณ์ที่อาจเกิดขึ้นในอนาคต เช่น การพยากรณ์ความต้องการของลูกค้า การทำนายความเปลี่ยนแปลงในตลาด หรือการคาดการณ์ผลกระทบจากการ ซึ่งช่วยให้องค์กรสามารถวางแผนและเตรียมการได้ล่วงหน้า

1.2.4 การค้นพบโอกาสใหม่ ๆ (Discovering New Opportunities) การวิเคราะห์ข้อมูลช่วยให้องค์กรสามารถค้นพบโอกาสใหม่ ๆ ทางธุรกิจ เช่น การหาตลาดที่ยังไม่ถูกเจาะ การค้นพบสินค้าที่มีแนวโน้มขายดี หรือการวิเคราะห์พฤติกรรมลูกค้าเพื่อสร้างผลิตภัณฑ์ที่ตอบสนองความต้องการได้มากขึ้น

1.2.5 การปรับปรุงความพึงพอใจของลูกค้า (Improving Customer Satisfaction) ข้อมูลเกี่ยวกับลูกค้า เช่น พฤติกรรมการซื้อ ความคิดเห็นเกี่ยวกับผลิตภัณฑ์ สามารถนำมาใช้วิเคราะห์เพื่อพัฒนาการบริการและผลิตภัณฑ์ให้ตรงตามความต้องการของลูกค้ามากขึ้น ส่งผลให้ลูกค้ามีความพึงพอใจสูงขึ้น

1.2.6 การจัดการความเสี่ยง (Managing Risks) การวิเคราะห์ข้อมูลสามารถใช้ในการตรวจจับและคาดการณ์ความเสี่ยงที่อาจเกิดขึ้น เช่น การตรวจจับการฉ้อโกงในระบบการเงิน หรือการวิเคราะห์ข้อมูลการทำงานเพื่อคาดการณ์ปัญหาทางเทคนิคในระบบก่อนที่จะเกิดขึ้น

1.3 ตัวอย่างของการวิเคราะห์ข้อมูลในภาคธุรกิจ

1.3.1 การตลาด (Marketing) การวิเคราะห์ข้อมูลการซื้อสินค้าของลูกค้าเพื่อหาแนวโน้มในการเลือกซื้อสินค้า และใช้ข้อมูลเหล่านี้ในการพัฒนากลยุทธ์การตลาดที่ตอบสนองความต้องการของลูกค้าได้ดีขึ้น

1.3.2 การแพทย์ (Healthcare) การวิเคราะห์ข้อมูลผู้ป่วยเพื่อทำนายความเสี่ยงของโรคต่าง ๆ เช่น การใช้ข้อมูลทางการแพทย์เพื่อพยากรณ์ว่าผู้ป่วยมีแนวโน้มที่จะเป็นโรคหัวใจหรือไม่

1.3.3 การเงิน (Finance) การวิเคราะห์ข้อมูลทางการเงินเพื่อช่วยในการตัดสินใจลงทุน เช่น การวิเคราะห์แนวโน้มราคาหุ้น และการตรวจจับการทุจริตในระบบการเงิน

2. วงจรการวิเคราะห์ข้อมูล (Data Analytics Lifecycle)

วงจรการวิเคราะห์ข้อมูล (Data Analytics Lifecycle) เป็นกระบวนการที่อธิบายขั้นตอนต่าง ๆ ที่เกี่ยวข้องกับการทำงานกับข้อมูล ตั้งแต่การเก็บรวบรวม จัดการ วิเคราะห์ จนถึงการนำผลลัพธ์ไปใช้ในการตัดสินใจ และการปรับปรุง วงจรนี้เป็นโครงสร้างที่ช่วยให้จัดการกับข้อมูลได้อย่างเป็นระบบ และช่วยให้การวิเคราะห์มีประสิทธิภาพมากขึ้น วงจรการวิเคราะห์ข้อมูลมักแบ่งออกเป็น 6 ขั้นตอนหลัก ดังนี้

2.1 การกำหนดปัญหา (Discovery)

2.1.1 รายละเอียด

ขั้นตอนนี้เป็นการเริ่มต้นโดยการระบุและกำหนดเป้าหมายของการวิเคราะห์ข้อมูล ซึ่งรวมถึงการทำความเข้าใจปัญหาหรือคำถามทางธุรกิจที่ต้องการแก้ไข และการกำหนดวัตถุประสงค์ที่ชัดเจนว่าจะใช้ข้อมูลอย่างไรในการตอบคำถามเหล่านั้น

2.1.2 กิจกรรม

- 1) ทำความเข้าใจกับปัญหาหรือโอกาสทางธุรกิจที่ต้องการแก้ไข
- 2) กำหนดคำถามที่ต้องการคำตอบจากการวิเคราะห์ข้อมูล
- 3) ระบุทรัพยากรที่จำเป็น เช่น ทีมงาน, เครื่องมือ, และข้อมูล

2.1.3 ผลลัพธ์

ได้เป้าหมายการวิเคราะห์ที่ชัดเจนและแนวทางในการรวบรวมข้อมูลที่ต้องการ

2.2 การเก็บรวบรวมข้อมูล (Data Collection)

2.2.1 รายละเอียด

ขั้นตอนนี้เกี่ยวข้องกับการรวบรวมข้อมูลจากแหล่งต่าง ๆ ที่เกี่ยวข้องกับปัญหาที่ต้องการแก้ไข ข้อมูลอาจมาจากระบบภายในองค์กร ข้อมูลออนไลน์ ข้อมูลการสำรวจ หรือข้อมูลสาธารณะ

2.2.2 กิจกรรม

- 1) รวบรวมข้อมูลที่เกี่ยวข้องจากแหล่งต่าง ๆ
- 2) ตรวจสอบคุณภาพของข้อมูลและความเหมาะสม
- 3) การบันทึกแหล่งที่มาของข้อมูลเพื่อการติดตามในภายหลัง

2.2.3 ผลลัพธ์: มีชุดข้อมูลที่จำเป็นพร้อมสำหรับการวิเคราะห์

2.3 การเตรียมข้อมูล (Data Preparation)

2.3.1 รายละเอียด

ขั้นตอนนี้เป็นการทำความสะอาดข้อมูลและเตรียมข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับการวิเคราะห์ ซึ่งอาจรวมถึงการจัดการกับค่าที่ขาดหาย ค่าผิดปกติ หรือการแปลงข้อมูลให้อยู่ในรูปแบบที่เครื่องมือสามารถนำไปใช้ได้

2.3.2 กิจกรรม

- 1) การทำความสะอาดข้อมูล เช่น การจัดการค่าที่ขาดหาย (Missing Values) และค่าผิดปกติ (Outliers)
- 2) การแปลงข้อมูล เช่น การจัดการกับตัวแปรหมวดหมู่ (Categorical Variables)
- 3) การรวมข้อมูลจากหลายแหล่งและการทำ Normalization

2.3.3 ผลลัพธ์: ข้อมูลที่ถูกทำความสะอาดและจัดเตรียมพร้อมสำหรับการวิเคราะห์

2.4 การวิเคราะห์ข้อมูล (Data Analysis)

2.4.1 รายละเอียด

นี่คือขั้นตอนสำคัญที่ใช้ในการวิเคราะห์ข้อมูลโดยใช้เครื่องมือและเทคนิคต่าง ๆ เช่น การวิเคราะห์เชิงสถิติ การวิเคราะห์ความสัมพันธ์ หรือการใช้ Machine Learning เพื่อทำนายหรือจำแนกประเภท

2.4.2 กิจกรรม:

- 1) การใช้การวิเคราะห์เชิงสถิติ เช่น การคำนวณค่าเฉลี่ย (Mean), ค่ามัธยฐาน (Median), และการวิเคราะห์ความสัมพันธ์ (Correlation)
- 2) การใช้ Machine Learning เช่น การจำแนกประเภท (Classification) หรือการจัดกลุ่ม (Clustering)
- 3) การทดสอบสมมติฐาน (Hypothesis Testing) และการวิเคราะห์แนวโน้ม (Trend Analysis)

2.4.3 ผลลัพธ์

ได้ข้อมูลเชิงลึก (Insights) หรือโมเดลที่สามารถใช้งานได้ เช่น โมเดลทำนายผลลัพธ์ หรือกฎความสัมพันธ์ในข้อมูล

2.5 การแสดงผลข้อมูล (Data Visualization)

2.5.1 รายละเอียด

หลังจากการวิเคราะห์ข้อมูลเรียบร้อยแล้ว ขั้นตอนนี้เกี่ยวข้องกับการนำเสนอผลการวิเคราะห์ให้กับผู้มีส่วนเกี่ยวข้องในรูปแบบที่เข้าใจง่าย การแสดงผลผ่านกราฟ แผนภูมิ หรือรายงาน จะช่วยให้การตีความข้อมูลทำได้ง่ายขึ้น

2.5.2 กิจกรรม

- 1) การสร้างกราฟ เช่น แผนภูมิแท่ง (Bar Chart), กราฟเส้น (Line Graph), และกราฟกระจาย (Scatter Plot)
- 2) การใช้เครื่องมือสร้างภาพข้อมูล เช่น Matplotlib, Seaborn หรือ Power BI
- 3) การสร้างแดชบอร์ดสำหรับการรายงานผลข้อมูลอย่างต่อเนื่อง

2.5.3 ผลลัพธ์

มีการนำเสนอข้อมูลในรูปแบบที่เข้าใจง่าย ทำให้การสื่อสารผลการวิเคราะห์มีประสิทธิภาพ

2.6 การนำผลลัพธ์ไปใช้ในการตัดสินใจและการดำเนินการ (Actionable Insights and Decision-Making)

2.6.1 รายละเอียด

ขั้นตอนสุดท้าย คือ การนำผลลัพธ์จากการวิเคราะห์ไปใช้ในการตัดสินใจเชิงกลยุทธ์หรือในการปรับปรุงกระบวนการทางธุรกิจ การใช้ข้อมูลเหล่านี้ช่วยให้องค์กรสามารถตัดสินใจได้อย่างมีประสิทธิภาพและรวดเร็ว

2.6.2 กิจกรรม

- 1) การตัดสินใจทางธุรกิจโดยอิงจากข้อมูลเชิงลึกที่ได้จากการวิเคราะห์
- 2) การวางแผนกลยุทธ์หรือแผนการดำเนินการเพื่อปรับปรุงกระบวนการทำงาน หรือ
- 3) การนำผลลัพธ์ไปทดสอบในการดำเนินงานจริง (Implementation)

2.6.3 ผลลัพธ์

การตัดสินใจและการดำเนินการที่อิงจากข้อมูล (Data-Driven Decision Making) ซึ่งช่วยเพิ่มประสิทธิภาพในการดำเนินงานหรือการตอบสนองต่อปัญหาทางธุรกิจ

ประโยชน์ของวัฏจักรการวิเคราะห์ข้อมูล

- 1) ความชัดเจนและการจัดระเบียบ: การใช้วัฏจักรช่วยให้การวิเคราะห์ข้อมูลเป็นไปอย่างเป็นระบบและจัดการได้ง่ายขึ้น

2) การปรับปรุงการตัดสินใจ: ทำให้สามารถใช้ข้อมูลในการสนับสนุนการตัดสินใจที่มีความถูกต้องและแม่นยำมากขึ้น

3) การทำงานเป็นทีม: วิศวกรนี้ช่วยให้ทีมงานที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล เช่น นักวิเคราะห์ข้อมูล นักวิทยาศาสตร์ข้อมูล และผู้จัดการโครงการ สามารถทำงานร่วมกันได้อย่างมีประสิทธิภาพ

4) การประเมินผลลัพธ์: ช่วยให้อาจวัดประสิทธิภาพของการวิเคราะห์ข้อมูลและปรับปรุงการทำงานในระยะยาวได้

3. การทำความเข้าใจประเภทของข้อมูล (Structured vs Unstructured Data)

การทำความเข้าใจประเภทของข้อมูล (Structured vs Unstructured Data) เป็นขั้นตอนสำคัญในการวิเคราะห์ข้อมูล เนื่องจากประเภทของข้อมูลที่แตกต่างกันต้องการเทคนิคและเครื่องมือที่แตกต่างกันในการจัดการและวิเคราะห์

3.1 ข้อมูลแบบมีโครงสร้าง (Structured Data)

ข้อมูลแบบมีโครงสร้าง (Structured Data) หมายถึงข้อมูลที่ถูกจัดระเบียบในรูปแบบที่มีโครงสร้างชัดเจน เช่น ตารางข้อมูลที่มีแถวและคอลัมน์ ข้อมูลประเภทนี้มักถูกจัดเก็บในฐานข้อมูลเชิงสัมพันธ์ (Relational Databases) ซึ่งง่ายต่อการค้นหาและจัดการโดยใช้ SQL (Structured Query Language) หรือเครื่องมือวิเคราะห์ข้อมูลทั่วไป

3.1.1 ลักษณะของข้อมูลแบบมีโครงสร้าง

- 1) รูปแบบตายตัว ข้อมูลจะมีรูปแบบที่กำหนดไว้ล่วงหน้า เช่น ค่าตัวเลข วันที่ หรือข้อความสั้น ๆ ที่ถูกจัดระเบียบในรูปแบบตาราง
- 2) จัดเก็บในฐานข้อมูลเชิงสัมพันธ์ ข้อมูลจะถูกจัดเก็บในแถว (Records) และคอลัมน์ (Fields) ในฐานข้อมูลเช่น MySQL, PostgreSQL, SQL Server
- 3) การเข้าถึงและการจัดการง่าย สามารถดึงข้อมูลและทำการคำนวณ วิเคราะห์ได้ง่ายด้วยคำสั่ง SQL หรือเครื่องมือวิเคราะห์ข้อมูล เช่น Python กับ Pandas หรือ Excel

3.1.2 ตัวอย่างของข้อมูลแบบมีโครงสร้าง

- 1) ข้อมูลลูกค้า: ตารางที่บันทึกข้อมูลชื่อลูกค้า, อายุ, เบอร์โทรศัพท์, และอีเมล
- 2) ข้อมูลการขาย: ตารางที่เก็บข้อมูลยอดขายสินค้า ประกอบด้วยข้อมูลเช่น วันที่ซื้อ, จำนวนสินค้าที่ซื้อ, ราคาต่อหน่วย
- 3) ข้อมูลการเงิน: ตารางบันทึกยอดเงินเข้าออกจากบัญชีธนาคารพร้อมรายละเอียดการทำธุรกรรม

3.1.3 ข้อดีของข้อมูลแบบมีโครงสร้าง

- 1) การประมวลผลและการวิเคราะห์ง่าย เนื่องจากข้อมูลถูกจัดระเบียบแล้ว การค้นหาวิเคราะห์ และสรุปผลทำได้อย่างรวดเร็ว
- 2) สามารถจัดเก็บในระบบที่มีประสิทธิภาพสูง: ระบบฐานข้อมูลเชิงสัมพันธ์มีเครื่องมือและระบบที่รองรับการจัดเก็บและประมวลผลข้อมูลขนาดใหญ่

3.1.4 ข้อเสียของข้อมูลแบบมีโครงสร้าง

- 1) จำกัดรูปแบบของข้อมูล ไม่สามารถรองรับข้อมูลที่ไม่มีรูปแบบตายตัว เช่น ข้อความยาว รูปภาพ หรือวิดีโอ
- 2) ความยืดหยุ่นต่ำ การปรับเปลี่ยนโครงสร้างของข้อมูลในฐานข้อมูลอาจต้องใช้เวลา หรือการออกแบบโครงสร้างใหม่

3.2 ข้อมูลแบบไม่มีโครงสร้าง (Unstructured Data)

ข้อมูลแบบไม่มีโครงสร้าง (Unstructured Data) หมายถึง ข้อมูลที่ไม่มีการจัดระเบียบในรูปแบบที่ตายตัว และไม่สามารถจัดเก็บในรูปแบบตารางได้อย่างง่ายดาย ข้อมูลประเภทนี้ประกอบด้วยข้อความ รูปภาพ เสียง วิดีโอ หรือไฟล์ที่ไม่ได้จัดเก็บตามมาตรฐานฐานข้อมูลเชิงสัมพันธ์

3.2.1 ลักษณะของข้อมูลแบบไม่มีโครงสร้าง:

- 1) ไม่มีรูปแบบตายตัว: ข้อมูลไม่มีการจัดเรียงในรูปแบบที่ชัดเจนหรือเป็นระเบียบ เช่น เอกสารข้อความ (Text), ภาพถ่าย, วิดีโอ
- 2) ต้องการเครื่องมือเฉพาะในการวิเคราะห์: การจัดการและการวิเคราะห์ข้อมูลประเภทนี้ต้องใช้เทคนิคที่ซับซ้อน เช่น Natural Language Processing (NLP) สำหรับข้อความ หรือการวิเคราะห์ภาพสำหรับข้อมูลรูปภาพ
- 3) จัดเก็บในฐานข้อมูลแบบ NoSQL หรือระบบไฟล์: ข้อมูลประเภทนี้มักถูกจัดเก็บในฐานข้อมูลที่ไม่ใช่เชิงสัมพันธ์ (NoSQL Databases) เช่น MongoDB หรือในระบบการจัดเก็บไฟล์ทั่วไป เช่น Hadoop Distributed File System (HDFS)

3.2.2 ตัวอย่างของข้อมูลแบบไม่มีโครงสร้าง:

- 1) ข้อความจากโซเชียลมีเดีย ข้อความหรือโพสต์จาก Twitter, Facebook ที่มีความยาวและเนื้อหาที่หลากหลาย
- 2) อีเมล ข้อความอีเมลที่มีข้อมูลทั้งรูปภาพ ข้อความ และไฟล์แนบ
- 3) รูปภาพและวิดีโอ ไฟล์สื่อมัลติมีเดียที่ใช้ในการวิเคราะห์ภาพหรือการรู้จำวัตถุ (Object Detection)
- 4) เสียง ไฟล์เสียงที่ใช้ในการวิเคราะห์เสียงพูด (Speech Recognition)

3.2.3 ข้อดีของข้อมูลแบบไม่มีโครงสร้าง:

1) รองรับข้อมูลได้หลากหลาย ข้อมูลที่ซับซ้อน เช่น รูปภาพ วิดีโอ และเสียง สามารถถูกวิเคราะห์เพื่อค้นหาข้อมูลเชิงลึกได้

2) แสดงข้อมูลเชิงลึกที่ซ่อนอยู่ ข้อมูลแบบไม่มีโครงสร้างมักมีรายละเอียดเชิงลึกที่สามารถค้นพบได้จากการวิเคราะห์ขั้นสูง เช่น การวิเคราะห์ข้อความเพื่อหาอารมณ์ (Sentiment Analysis)

3.2.4 ข้อเสียของข้อมูลแบบไม่มีโครงสร้าง

1) ความยากในการวิเคราะห์ ต้องใช้เทคนิคการประมวลผลข้อมูลขั้นสูงในการจัดการ เช่น การวิเคราะห์ข้อความ (Text Mining), การแยกภาพ (Image Recognition), หรือการแยกเสียง (Speech Recognition)

2) ต้องใช้ทรัพยากรสูง การจัดการและการวิเคราะห์ข้อมูลที่ซับซ้อนมักต้องการกำลังการประมวลผลและทรัพยากรที่สูง

3.3 ข้อมูลกึ่งโครงสร้าง (Semi-structured Data)

ข้อมูลกึ่งโครงสร้างคือข้อมูลที่อยู่ระหว่างข้อมูลแบบมีโครงสร้างและแบบไม่มีโครงสร้าง โดยข้อมูลประเภทนี้มีรูปแบบหรือโครงสร้างบางส่วน แต่ไม่เป็นระเบียบมากพอที่จะเก็บในตารางฐานข้อมูลเชิงสัมพันธ์

3.3.1 ลักษณะของข้อมูลกึ่งโครงสร้าง

1) มีองค์ประกอบของทั้งข้อมูลที่จัดเก็บอย่างมีระเบียบและข้อมูลที่ไม่มีโครงสร้าง

2) ตัวอย่างที่เห็นได้ชัดคือไฟล์ JSON และ XML ที่ข้อมูลมีการจัดเรียงในรูปแบบขั้น

(Hierarchical Structure)

3.3.2 ตัวอย่างของข้อมูลกึ่งโครงสร้าง

1) ไฟล์ JSON ข้อมูลที่เก็บในรูปแบบ JSON ซึ่งมีโครงสร้างคล้ายตารางแต่ไม่ได้เป็นตารางที่สมบูรณ์

2) อีเมล แม้ว่าอีเมลจะมีเนื้อหาที่ไม่เป็นโครงสร้าง แต่ก็มีข้อมูลโครงสร้างบางส่วน เช่น ที่อยู่ผู้ส่ง, วันที่, และหัวเรื่อง (Subject Line)

3.4 การวิเคราะห์ข้อมูลแบบ Structured และ Unstructured

3.4.1 การวิเคราะห์ข้อมูลแบบมีโครงสร้าง (Structured Data): ข้อมูลสามารถถูกวิเคราะห์ได้ง่ายโดยใช้เครื่องมือทั่วไป เช่น SQL หรือ Python กับไลบรารี Pandas ซึ่งสามารถจัดการกับข้อมูลที่จัดเก็บในรูปแบบตารางได้อย่างมีประสิทธิภาพ เช่น การวิเคราะห์ข้อมูลยอดขาย, การหาค่าเฉลี่ย หรือการคำนวณเชิงสถิติ

3.4.2 การวิเคราะห์ข้อมูลแบบไม่มีโครงสร้าง (Unstructured Data): ต้องใช้เครื่องมือเฉพาะในการวิเคราะห์ข้อมูลที่ซับซ้อน เช่น การวิเคราะห์ข้อความโดยใช้ Natural Language Processing (NLP) หรือการ

วิเคราะห์ภาพด้วยการใช้ Computer Vision ข้อมูลประเภทนี้อาจต้องใช้เวลาประมวลผลเบื้องต้นเพื่อให้เข้าใจง่ายขึ้นก่อนจะทำการวิเคราะห์ เช่น การแปลงข้อความเป็นโครงสร้างหรือการแยกส่วนประกอบของรูปภาพ

4. สาขาที่เกี่ยวข้องและการประยุกต์ใช้ในโลกรจริง

วิทยาศาสตร์ข้อมูล (Data Science) เป็นศาสตร์สหสาขา ซึ่งประกอบด้วยองค์ความรู้จากหลายด้านดังนี้:

สาขา	รายละเอียด
สถิติและคณิตศาสตร์ (Statistics & Mathematics)	ใช้ในการวิเคราะห์ข้อมูล สรุปผล และสร้างแบบจำลอง เช่น ค่ากลาง ความเบี่ยงเบน ความน่าจะเป็น
วิทยาการคอมพิวเตอร์ (Computer Science)	ใช้ในการเขียนโปรแกรม ประมวลผลข้อมูล และจัดการระบบฐานข้อมูล
การเรียนรู้ของเครื่อง (Machine Learning)	ใช้ในการสร้างโมเดลเพื่อคาดการณ์ พฤติกรรม หรือจำแนกข้อมูล
การแสดงผลข้อมูล (Data Visualization)	ใช้ในการแสดงผลข้อมูลให้เข้าใจง่าย เช่น กราฟ แผนภูมิ แดชบอร์ด
วิทยาการข้อมูลธุรกิจ (Business Intelligence)	เพื่อเชื่อมโยงข้อมูลกับการตัดสินใจทางธุรกิจ
ความรู้เฉพาะทาง (Domain Knowledge)	ขึ้นอยู่กับอุตสาหกรรม เช่น การเงิน การแพทย์ การศึกษา ฯลฯ

5. การประยุกต์ใช้ในโลกรจริง (Real-World Applications)

5.1 ธุรกิจและการตลาด

5.1.1 วิเคราะห์พฤติกรรมลูกค้า (Customer Segmentation)

5.1.2 พยากรณ์ยอดขาย (Sales Forecasting)

5.1.3 การตลาดเชิงพฤติกรรม (Targeted Marketing) ตัวอย่าง: Shopee วิเคราะห์การคลิกและการซื้อ เพื่อแนะนำสินค้า

5.2 การแพทย์และสาธารณสุข

5.2.1 วิเคราะห์ภาพทางการแพทย์ (เช่น X-ray, MRI)

5.2.2 คาดการณ์โรค (Disease Prediction)

5.2.3 ติดตามและควบคุมโรคระบาด (Epidemiological Modeling)

5.2.4 ตัวอย่าง: ระบบ IBM Watson ช่วยแพทย์วิเคราะห์ข้อมูลผู้ป่วยเพื่อแนะนำวิธีการรักษา

5.3 การเงินและธนาคาร

- 5.3.1 ตรวจจับการฉ้อโกง (Fraud Detection)
- 5.3.2 วิเคราะห์ความเสี่ยงด้านเครดิต (Credit Scoring)
- 5.3.3 การลงทุนอัตโนมัติ (Algorithmic Trading)
- 5.3.4 ตัวอย่าง: ธนาคารใช้ AI ประเมินความเสี่ยงลูกค้าก่อนปล่อยสินเชื่อ

5.4 การศึกษา

- 5.4.1 วิเคราะห์ผลสัมฤทธิ์ของนักเรียน (Learning Analytics)
- 5.4.2 ระบบแนะนำรายวิชา (Course Recommendation)
- 5.4.3 พัฒนา Adaptive Learning System
- 5.4.4 ตัวอย่าง: แพลตฟอร์มเช่น Coursera ใช้ข้อมูลเพื่อแนะนำหลักสูตรที่เหมาะสมกับผู้เรียน

5.5 อุตสาหกรรมและ IoT

- 5.5.1 วิเคราะห์ข้อมูลจากเซนเซอร์ (Sensor Data Analysis)
- 5.5.2 คาดการณ์การซ่อมบำรุงเครื่องจักร (Predictive Maintenance)
- 5.5.3 เพิ่มประสิทธิภาพการผลิต (Process Optimization)
- 5.5.4 ตัวอย่าง: โรงงานใช้ข้อมูลจาก IoT ตรวจสอบอุณหภูมิและความดันเพื่อลดของเสีย

5.6 สื่อสารมวลชนและโซเชียลมีเดีย

- 5.6.1 วิเคราะห์เนื้อหาและความรู้สึก (Sentiment Analysis)
- 5.6.2 ตรวจจับข่าวปลอม (Fake News Detection)
- 5.6.3 การจัดอันดับข่าว/โพสต์ที่ได้รับความนิยม
- 5.6.4 ตัวอย่าง: Facebook ใช้ข้อมูลจากผู้ใช้งานเพื่อจัดลำดับโพสต์ใน News Feed

6. ตัวอย่างกรณีศึกษาเบื้องต้น

กรณีศึกษาที่ 1 บริบทของปัญหา : ร้านกาแฟ "Coffee@You" เปิดให้บริการในเมืองแห่งหนึ่งมานาน 1 ปี แต่ยอดขายไม่สม่ำเสมอในแต่ละวัน เจ้าของร้านต้องการทราบว่า:

- ลูกค้าซื้อกาแฟเวลาใดมากที่สุด?
- เมนูใดขายดี?
- วันใดของสัปดาห์ยอดขายสูงสุด?

วันที่	เวลา	เมนู	จำนวนแก้ว	ราคา (บาท)
--------	------	------	-----------	------------

01/06/2025	08:00	ลาเต้	2	120
01/06/2025	13:30	มอคค่า	1	60
02/06/2025	09:00	ลาเต้	3	180
03/06/2025	14:00	โกโก้	2	100
...	...	...	...	...

ขั้นตอนการวิเคราะห์ (แบบ Descriptive Analytics)

1. วิเคราะห์ยอดขายตามเวลา

พบว่าช่วง 08:00 – 10:00 น. มียอดขายสูงสุด

คำแนะนำ: จัดโปรโมชั่น “เช้ารับส่วนลด”

2. วิเคราะห์ยอดขายตามเมนู

ลาเต้ มียอดขายสูงสุด (35% ของยอดขายทั้งหมด)

รองลงมาคือ มอคค่า และ โกโก้

คำแนะนำ: โปรโมทลาเต้มากขึ้น หรือออกเมนูใหม่ใกล้เคียง

3. วิเคราะห์ยอดขายตามวันในสัปดาห์

ยอดขายมากสุดในวัน ศุกร์และเสาร์

คำแนะนำ: เสริมพนักงาน และจัดรายการส่งเสริมการขายในวันดังกล่าว

ผลลัพธ์ที่ได้

เจ้าของร้านเข้าใจพฤติกรรมลูกค้าดีขึ้น

สามารถกำหนดกลยุทธ์เวลาโปรโมชั่น / เมนูยอดนิยม

เพิ่มยอดขายเฉลี่ยได้ 15% ภายใน 1 เดือน

สิ่งที่เรียนรู้	รายละเอียด
การตั้งคำถามที่ดี	เริ่มจาก “อยากรู้อะไร?” ไม่ใช่แค่มีข้อมูล
ความสำคัญของข้อมูลสะอาด	ต้องจัดรูปแบบข้อมูลให้พร้อมใช้
การแปลผล	ข้อมูลที่ตีต้องถูก ตีความ ไม่ใช่แค่ “แสดง”
การนำไปใช้	ข้อมูลจะมีค่าก็ต่อเมื่อถูกนำไปตัดสินใจจริง

กรณีศึกษาที่ 2 โรงเรียนต้องการลดการขาดเรียนของนักเรียน

บริบทของปัญหา โรงเรียนแห่งหนึ่งสังเกตว่า นักเรียนบางคนมีแนวโน้มขาดเรียนบ่อยในบางวัน และอยากรู้ว่าปัจจัยใดเกี่ยวข้องกับการขาดเรียน เพื่อหาทางลดอัตราการขาดเรียน

รหัสนักเรียน	วันที่	สถานะการมาเรียน	อากาศ (ฝนตก/ไม่ตก)	วิชาช่วงเช้า
60101	01/06/2025	มา	ไม่ตก	คณิต
60102	01/06/2025	ขาด	ฝนตก	วิทย์
60101	02/06/2025	ขาด	ฝนตก	อังกฤษ
...	...	...	...	...

การวิเคราะห์เชิงพรรณนา

วันที่ฝนตก: อัตราขาดเรียน สูงกว่าปกติ
นักเรียนบางคนขาดบ่อยในวันจันทร์และศุกร์
วิชาบางวิชา (เช่น วิทย์) มีอัตราขาดเรียนมากกว่าวิชาอื่น

แนวทางปรับปรุงจากข้อมูล

เตรียมรถรับส่งช่วงฝนตก
มีระบบเตือนนักเรียนและผู้ปกครองล่วงหน้า
ปรับกิจกรรมช่วงเช้าให้ดึงดูดใจมากขึ้น

กรณีศึกษา 3: คลินิกต้องการจัดคิวแพทย์ให้เหมาะสมกับช่วงเวลาที่คนไข้หนาแน่น

บริบทของปัญหา : คลินิกเอกชนมีคนไข้จำนวนมากในบางช่วงเวลา ทำให้ต้องรอนาน และเกิดความไม่พอใจ ผู้บริหารจึงอยากรู้ว่าเวลาใดคนไข้หนาแน่น เพื่อจัดแพทย์และพนักงานให้เพียงพอ

ข้อมูลสมมติ

วันที่	เวลา	จำนวนคนไข้	แพทย์ที่ประจำ
01/06/2025	09:00	10	Dr.A
01/06/2025	15:00	4	Dr.B
02/06/2025	10:00	12	Dr.A

...	...	...	...
-----	-----	-----	-----

การวิเคราะห์เชิงพรรณนา

พบว่าช่วงเช้า 09:00–11:00 คนไข้มากที่สุด

บ่ายมีคนน้อย

วันจันทร์และศุกร์เป็นวันที่มีคนไข้สูงสุด

ผลการนำไปใช้

เพิ่มแพทย์ประจำในช่วงเช้า

ลดจำนวนแพทย์ช่วงบ่ายเพื่อประหยัดต้นทุน

แจ้งเตือนคนไข้ให้จองคิวล่วงหน้าคนน้อย

กรณี	เครื่องมือที่ใช้	เป้าหมาย	ผลที่ได้
ร้านกาแฟ	ตาราง, กราฟ, ค่าเฉลี่ย	วิเคราะห์พฤติกรรมลูกค้า	เพิ่มยอดขาย
โรงเรียน	การแจกแจงความถี่	วิเคราะห์สาเหตุขาดเรียน	วางแผนลดการขาด
คลินิก	กราฟแท่ง, เวลา	วิเคราะห์ช่วงแออัด	จัดตารางแพทย์เหมาะสม