



1101031

Data Science Foundation I

Porn-anant lamkhajornchai

Week 2



Topic

- การวิเคราะห์ข้อมูล (Data Analytics)
- การจัดกลุ่มข้อมูล (Data Clustering)



**Data
Analytics**

VS

**Data
Analysis**

Data Analytics

- ❑ A discipline that **encompasses the complete management of data** – including collecting, cleaning, organizing, storing, governing, and analyzing data
- ❑ Data analytics is **an overarching discipline** (which includes data analysis as a necessary subcomponent)
- ❑ Data analytics tries to **predict the future**

Data Analysis

- ❑ The **process of examining** in close detail the components of a given **data set** – separating them out and studying the parts individually and their relationship between one another
- ❑ Data analysis **is a process**
- ❑ Data analysis **looks at the past**

ที่มา: <https://itchronicles.com/big-data/data-analytics-vs-data-analysis-whats-the-difference/>

Data Analytics and Data Analysis Comparison Table

	Data Analytics	Data Analysis
Form	Data analytics is ‘ general ’ form of analytics which is used in businesses to make decisions from data which are data-driven	Data analysis is a specialized form of data analytics used in businesses to analyze data and take some insights of it
Structure	Data analytics consist of data collection and inspect in general and it has one or more users	Data analysis consisted of defining a data, investigation, cleaning, transforming the data to give a meaningful outcome
Tools	There are many analytics tools in a market but mainly R, Tableau Public, Python, SAS, Apache Spark, Excel are used	Data analysis is a specialized form of data analytics used in businesses to analyze data and take some insights of it

ที่มา: <https://www.educba.com/data-analytics-vs-data-analysis/>

Data Analytics and Data Analysis Comparison Table

	Data Analytics	Data Analysis
Sequence	Data analytics life cycle consist of Business Case Evaluation, Data Identification, Data Acquisition & Filtering, Data Extraction, Data Validation & Cleansing, Data Aggregation & Representation, Data Analysis, Data Visualization, Utilization of Analysis Results	The sequence followed in data analysis are data gathering, data scrubbing, analysis of data and interpret the data precisely so that you can understand what your data want to say
Usage	Data Analytics can be used to find masked patterns, anonymous correlations, customer preferences, market trends and other necessary information that can help to make more notify decisions for business purpose	Data analysis can be used in various ways like one can perform analysis like descriptive analysis, exploratory analysis, inferential analysis, predictive analysis and take useful insights from the data

Data Analytics



- **A process** that relies on methods and techniques to **taking raw data, mining for insights** that are relevant to the business's primary goals, and **drilling down into this information** to transform metrics, facts, and **figures into initiatives for improvement**

(Durgevic, [www](#), 2020)

Data Analytics



- A process of **cleaning, transforming, and modeling data** to discover useful information for business decision-making
- The purpose of data analysis is to **extract useful information from data** and **taking the decision** based upon the data analysis

(Guru99, www, 2020)

Data Analytics



- The science of **analyzing raw data** in order to make conclusions about that information
- Many of the techniques and processes of data analytics have been **automated into mechanical processes and algorithms** that work over raw data for human consumption

(Frankenfield, www, 2020)

Data Analytics



- The techniques used to **analyze data to enhance productivity and business gain**
- Data is **extracted from various sources** and is **cleaned** and **categorized to analyze** various behavioral patterns
- The techniques and the tools used vary according to the organization or individual

(Kappagantula, www, 2020)

รูปแบบการวิเคราะห์ข้อมูล



รูปแบบการวิเคราะห์ข้อมูล



การวิเคราะห์ข้อมูลแบบพื้นฐาน (Descriptive analytics)

- การวิเคราะห์ข้อมูลเพื่อตอบคำถามว่า “เกิดอะไรขึ้นบ้างในช่วงเวลาที่ผ่านมา”
- การวิเคราะห์ข้อมูลเพื่อแสดงผลของรายการทางธุรกิจ เหตุการณ์ หรือกิจกรรมต่าง ๆ ที่ได้เกิดขึ้นแล้วหรือกำลังเกิดขึ้น ในลักษณะที่ง่ายต่อการทำความเข้าใจหรือต่อการตัดสินใจกระทำสิ่งใดสิ่งหนึ่ง
- ตัวอย่างเช่น รายงานการขาย

รายงานผลการดำเนินงาน

รูปแบบการวิเคราะห์ข้อมูล

2

การวิเคราะห์แบบเชิงวินิจฉัย (Diagnostic analytics)

- การอธิบายถึงสาเหตุหรือปัจจัย และความสัมพันธ์ของปัจจัยหรือตัวแปรต่าง ๆ ที่มีความสัมพันธ์ต่อกันของสิ่งที่เกิดขึ้น
- ตัวอย่างเช่น
 - ❖ ความสัมพันธ์ระหว่างยอดขายต่อกิจกรรมทางการตลาดแต่ละประเภท ซึ่งจะช่วยให้ตัดสินใจในการผลิตและการจำหน่ายสินค้าในแนวทางที่เหมาะสม
 - ❖ กำไรที่หน่วยงานได้จากการจำหน่ายสินค้ามาจากลูกค้าเพิ่มขึ้น หรือการลดต้นทุนการผลิต

รูปแบบการวิเคราะห์ข้อมูล



การวิเคราะห์แบบพยากรณ์ (Predictive analytics)

- [illegible]

การวิเคราะห์แบบพยากรณ์ “ว่าที่ผู้สมัครผู้ว่าฯ กทม.”



ที่มา <https://www.bangkokbiznews.com/news/977137>

รูปแบบการวิเคราะห์ข้อมูล



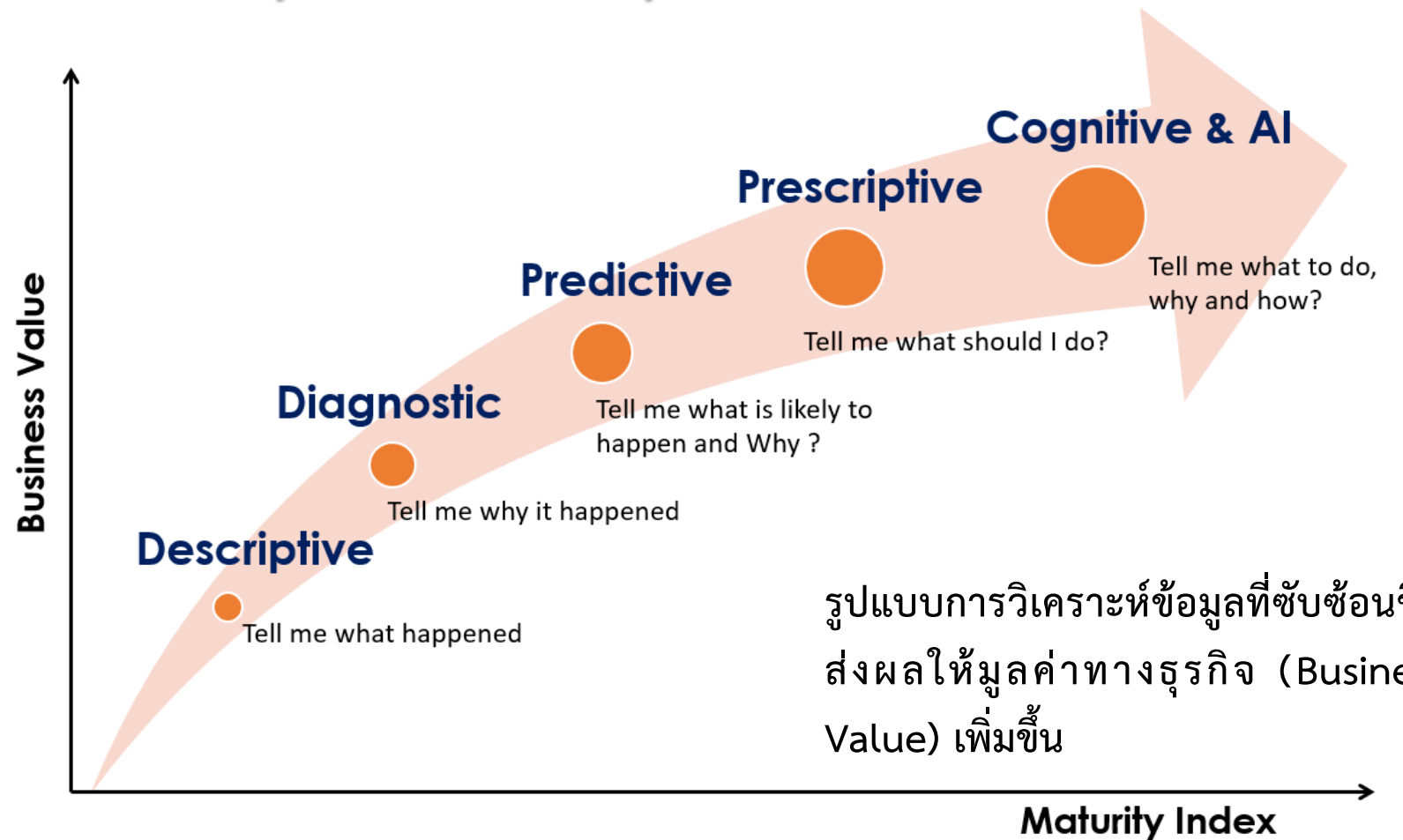
การวิเคราะห์แบบให้คำแนะนำ (Prescriptive analytics)

- การวิเคราะห์ข้อมูลที่มีความซับซ้อนที่สุด
- การพยากรณ์สิ่งต่าง ๆ ที่จะเกิดขึ้น พร้อมกับการระบุถึง
 - ❖ ข้อดีของเหตุการณ์ที่จะเกิดขึ้น
 - ❖ ข้อจำกัดของเหตุการณ์ที่จะเกิดขึ้น
 - ❖ ระยะเวลาของเหตุการณ์ที่จะเกิดขึ้น
 - ❖ การให้คำแนะนำ หรือทางเลือกต่าง ๆ และผลของแต่ละทางเลือก

รูปแบบการวิเคราะห์ข้อมูล



Analytics Maturity



รูปแบบการวิเคราะห์ข้อมูลที่ซับซ้อนขึ้น
ส่งผลให้มูลค่าทางธุรกิจ (Business Value) เพิ่มขึ้น

ตัวอย่างการวิเคราะห์ข้อมูล



ปี ค.ศ. 2012 ประธานาธิบดี บารัค ฮุสเซน โอบามา (Barack Hussein Obama) ใช้การวิเคราะห์ข้อมูล (Data analytics)

- วิเคราะห์ฐานเสี่ยงเลือกตั้งของตัวเอง และการปฏิสัมพันธ์ต่าง ๆ ของฐานเสี่ยงกับกิจกรรมการเลือกตั้ง
- มีการปรับกลยุทธ์การหาเสียงตามสภาพ / สถานการณ์ของพื้นที่แต่ละแห่ง ให้เหมาะสมกับกลุ่มเป้าหมาย และ ละเว้นการหาเสียงกับบางกลุ่ม
- วางแผนการหาเสียงให้มีการใช้ทรัพยากรบุคคล และ เวลาให้มีประสิทธิภาพมากที่สุด



ตัวอย่างการวิเคราะห์ข้อมูล



Uber ใช้การวิเคราะห์ข้อมูล (Data analytics)



- การตัดสินใจแบบอัตโนมัติเพื่อการปรับราคาค่าโดยสาร หากลูกค้ามีความต้องการใช้งานมาก ค่าโดยสารก็จะสูงขึ้นตามไปด้วย
- กรณีเหตุการณ์จับตัวประกันกลางกรุงชิตนีย์ ประเทศออสเตรเลียเมื่อปลายปี ค.ศ. 2014 อัตราค่าโดยสารของ Uber ในย่านดังกล่าวเพิ่มสูงขึ้น 4 เท่า เมื่อเทียบกับอัตราปกติ เนื่องจากลูกค้าทุกคนต้องการเดินทางกลับให้ถึงบ้านโดยเร็วที่สุด



องค์ประกอบสำคัญ ในการวิเคราะห์ข้อมูล

(Data Analysis Components)

องค์ประกอบสำคัญในการวิเคราะห์ข้อมูล



1

Roadmap and operating model

- การวิเคราะห์ข้อมูลขึ้นอยู่กับแบบจำลองกระบวนการ (Operating Model) ซึ่งมีการปรับเปลี่ยนไปตามวิสัยทัศน์ (Vision) และกลยุทธ์ (Strategy) ของหน่วยงาน
- หน่วยงานจะกำหนดโครงสร้าง กระบวนการ และสมรรถนะหลักต่าง ๆ ด้วยการจัดการข้อมูลที่มีประสิทธิภาพ ซึ่งผ่านกระบวนการวิเคราะห์ข้อมูลเชิงลึก

องค์ประกอบสำคัญในการวิเคราะห์ข้อมูล



Platform and data architecture

- แพลตฟอร์ม (Platform) ที่เหมาะสมในการวิเคราะห์ข้อมูลจะช่วยให้หน่วยงานสามารถจัดเก็บ ประมวลผล และวิเคราะห์ข้อมูลในระดับที่หน่วยงานต้องการได้อย่างมีประสิทธิภาพ



องค์ประกอบสำคัญในการวิเคราะห์ข้อมูล



3

Data security

- หน่วยงานควรตรวจสอบให้แน่ชัดว่าข้อมูลที่ตนเองจัดเก็บอยู่ นั้นมีการเปลี่ยนแปลง และใช้ประโยชน์ในลักษณะใดที่ไม่ส่งผลกระทบต่อความปลอดภัยของข้อมูลของหน่วยงาน

องค์ประกอบสำคัญในการวิเคราะห์ข้อมูล



Data governance and standards

- ธรรมาภิบาลข้อมูล (Data governance) ถือเป็นสิ่งสำคัญในกระบวนการวิเคราะห์ข้อมูล
- ธรรมาภิบาลข้อมูลครอบคลุมในประเด็นด้านการจัดการ การตรวจสอบกระบวนการรวบรวม จัดเก็บ ประมวลผล และการใช้ข้อมูลภายในหน่วยงาน
- หน่วยงานควรตระหนักและให้ความสำคัญในการดำเนินการที่เกี่ยวข้องกับธรรมาภิบาลข้อมูล และควรที่จะดำเนินการอย่างต่อเนื่อง

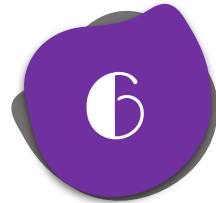
องค์ประกอบสำคัญในการวิเคราะห์ข้อมูล



Software and tooling

- ในการวิเคราะห์ข้อมูลของหน่วยงาน หน่วยงานควร
 - ❖ คัดเลือก กำหนด และใช้โปรแกรมและเครื่องมือต่าง ๆ ที่สนับสนุนการทำงาน
 - ❖ เลือกใช้โปรแกรมและเครื่องมือให้มีความเหมาะสมกับข้อมูลและสถาปัตยกรรมข้อมูลของหน่วยงาน

องค์ประกอบสำคัญในการวิเคราะห์ข้อมูล



Legacy migration

- หน่วยงานควรมีการวางแผน และเตรียมความพร้อมในการปรับเปลี่ยนระบบงาน (ระบบเดิมไปสู่ระบบใหม่) เมื่อต้องดำเนินงานด้านการวิเคราะห์ข้อมูล
 - ❖ ด้านบุคลากร
 - ❖ ด้านครุภัณฑ์ (วัสดุ-อุปกรณ์)
 - ❖ ด้านเทคโนโลยี

องค์ประกอบสำคัญในการวิเคราะห์ข้อมูล



Data acquisition

- หน่วยงานต้องทราบว่าข้อมูลสำคัญที่จะนำมาใช้ในการดำเนินธุรกิจ
 - ❖ ข้อมูลที่ต้องการอยู่ที่ใด
 - ❖ ข้อมูลที่ต้องการอยู่ในรูปแบบใด
 - ❖ หน่วยงานจะรวบรวมข้อมูลที่ต้องการได้อย่างไร

องค์ประกอบสำคัญในการวิเคราะห์ข้อมูล



Skills and roles

- บุคลากรในหน่วยงานจำที่รับผิดชอบด้านการวิเคราะห์ข้อมูล จำเป็นต้อง
 - ❖ มีทักษะ ความรู้ และความสามารถในการวิเคราะห์ข้อมูล
 - ❖ มีความเข้าใจในบทบาทหน้าที่ของตนเอง

องค์ประกอบสำคัญในการวิเคราะห์ข้อมูล



11

Real-time analytics

- หน่วยงานต้องการข้อมูลที่ผ่านการวิเคราะห์ข้อมูลแบบทันที เพื่อใช้ประโยชน์ในการตัดสินใจ และสามารถดำเนินธุรกิจที่ทันเวลา

12

Advanced analytics

- การวิเคราะห์ข้อมูลที่ดีควรที่จะนำเสนอข้อมูลที่คาดการณ์พยากรณ์ และให้คำแนะนำเพื่อนำไปใช้ประโยชน์ต่อไปได้

ประโยชน์ของการวิเคราะห์ข้อมูล



ประโยชน์เชิงธุรกิจ

- นำเสนอข้อมูลสินค้าและผลิตภัณฑ์ให้ตรงกับกลุ่มเป้าหมาย
- ช่วยตัดสินใจในการวางแผนธุรกิจ
- กำหนดนโยบายหน่วยงาน
- กำหนดกิจกรรมส่งเสริมการขาย
- วางแผนการดำเนินงานธุรกิจในอนาคต

ประโยชน์ของการวิเคราะห์ข้อมูล



2

ประโยชน์ด้านอื่น ๆ

- การคาดการณ์และการพยากรณ์เหตุการณ์ต่าง ๆ ที่คาดว่าจะเกิดขึ้นในอนาคต
- การกำหนดนโยบายและยุทธศาสตร์ของหน่วยงาน
- การวางแผนการบริหารของหน่วยงาน

ปัญหาที่พบในการวิเคราะห์ข้อมูล



ข้อมูลที่นำมาวิเคราะห์ที่ไม่มีคุณภาพ

- การขาดการระบุถึงแหล่งข้อมูลหลักและนิยามของการป้อนข้อมูลที่ชัดเจน
 - ❖ ผู้ใช้งานมีความเข้าใจข้อมูลที่แตกต่างกัน
 - ❖ พนักงานแต่ละฝ่ายให้ความสนใจเฉพาะกับหน่วยงานของตนเองเท่านั้น โดยไม่สนใจข้อมูลของฝ่ายอื่นที่เกี่ยวข้อง (กรณีที่หน่วยงานขนาดใหญ่ที่มีหน่วยงานย่อยจำนวนมาก)

ปัญหาที่พบในการวิเคราะห์ข้อมูล



ข้อมูลที่นำมาวิเคราะห์ที่ไม่มีคุณภาพ

- ขาดการกำหนดผู้ที่มีหน้าที่รับผิดชอบเกี่ยวกับการบริหารจัดการข้อมูล
- กำหนดผู้รับผิดชอบที่รับผิดชอบเกี่ยวกับการบริหารจัดการข้อมูลไม่เหมาะสม
- ความสม่ำเสมอในการบริหารจัดการข้อมูลของหน่วยงาน (ในระบบงานต่าง ๆ)
- ขาดการควบคุมการบริหารจัดการข้อมูลที่มีประสิทธิภาพ

ปัญหาที่พบในการวิเคราะห์ข้อมูล

2

ปัญหาด้านอื่น ๆ

- หน่วยงานมีการปรับเปลี่ยนให้บุคลากรที่ไม่มีทักษะ ความรู้ ความสามารถ และความเชี่ยวชาญมาทำงานด้านการจัดการและการวิเคราะห์ข้อมูล
- หน่วยงาน (ขนาดเล็ก) ขาดบุคลากรที่มีความรู้ ความสามารถในการจัดการและวิเคราะห์ข้อมูลโดยตรง



ปัญหาที่พบในการวิเคราะห์ข้อมูล

2

ปัญหาด้านอื่น ๆ

- หน่วยงานขาดการกำหนดวัตถุประสงค์และเป้าหมายในการวิเคราะห์ข้อมูล รวมถึงการนำข้อมูลที่ผ่านกระบวนการวิเคราะห์ข้อมูลไปใช้ประโยชน์
- หน่วยงานและบุคลากรในหน่วยงานขาดความเชื่อมั่นในข้อมูลที่ผ่านการวิเคราะห์ข้อมูล หรือเชื่อมั่นว่าสิ่งที่ตนเองคิด “ดีที่สุด” หรือ “ถูกต้องที่สุด”



ข้อควรพิจารณาใน
การวิเคราะห์ข้อมูล
ให้ประสบความสำเร็จ



ข้อควรพิจารณาในการวิเคราะห์ข้อมูล



เริ่มจาก “ประเด็นทางธุรกิจ”

ประเด็นที่หน่วยงานต้องพิจารณา ประกอบด้วย

- กลุ่มเป้าหมาย ลูกค้า หรือผู้รับบริการของหน่วยงานคือใคร
- การใช้งบประมาณที่มีจำกัดให้เกิดความคุ้มค่า และตรงกับกลุ่มเป้าหมายหลักของหน่วยงาน
- ต้นทุนในการผลิตสินค้าและบริการ (ความคุ้มค่า และมีประสิทธิภาพสูงสุด)
- การแก้ไขปัญหาเฉพาะหน้าอย่างทันท่วงที

ข้อควรพิจารณาในการวิเคราะห์ข้อมูล



ข้อมูลที่นำมาวิเคราะห์ไม่จำเป็นต้องเป็น Big Data

- การวิเคราะห์ข้อมูลไม่จำเป็นต้องใช้ข้อมูลขนาดใหญ่ (Big Data)
- การวิเคราะห์ข้อมูลไม่จำเป็นต้องใช้กับเครื่องมือเฉพาะด้าน อาจใช้โปรแกรมพื้นฐาน เช่น Microsoft Excel ก็ได้
- การวิเคราะห์ข้อมูลที่ประสบความสำเร็จ คือ การวิเคราะห์ การสังเคราะห์ การตีความ และการสรุปข้อมูลที่สอดคล้องกับวัตถุประสงค์ หรือนำไปใช้ประโยชน์

ข้อควรพิจารณาในการวิเคราะห์ข้อมูล



ควรมีการจัดการข้อมูลอย่างเหมาะสม

- หน่วยงานควรมีการจัดการข้อมูลให้อยู่รวมกัน
- ข้อมูลของหน่วยงานไม่ควรถูกจัดเก็บข้อมูลไว้ในที่ต่าง ๆ ที่แตกต่างกัน คนละแฟ้มข้อมูล หรือคนละระบบ
- การที่หน่วยงานจัดเก็บข้อมูลไว้ในสถานที่ที่แตกต่างกัน อาจส่งผลต่อการวิเคราะห์ การควบคุม การดูแล และการใช้งานข้อมูล

ข้อควรพิจารณาในการวิเคราะห์ข้อมูล



หน่วยงานต้องดำเนินการอย่างค่อยเป็นค่อยไป

- การวิเคราะห์ข้อมูลของหน่วยงาน ควรดำเนินการแบบค่อยเป็นค่อยไป ไม่ควรรีบร้อน
- การวิเคราะห์ข้อมูลจำเป็นต้องอาศัยความเชี่ยวชาญ และความชำนาญของผู้ปฏิบัติงาน

ข้อควรพิจารณาในการวิเคราะห์ข้อมูล



การวิเคราะห์ข้อมูลจำเป็นต้องใช้เวลา

- การวิเคราะห์ข้อมูลของหน่วยงานจำเป็นต้องอาศัยเวลาเนื่องจากพฤติกรรมของผู้บริโภคมีการเปลี่ยนแปลงอยู่ตลอดเวลา

ข้อควรพิจารณาในการวิเคราะห์ข้อมูล



การสร้างความเข้าใจเกี่ยวกับการวิเคราะห์ข้อมูล

หน่วยงานควรสร้างความเข้าใจเกี่ยวกับการวิเคราะห์ข้อมูลแก่ผู้เกี่ยวข้องในทุกภาคส่วน

- ผู้บริหารระดับสูงของหน่วยงาน (CEO หรือ CMO)
 - ❖ ควรมีความรู้และความเข้าใจเกี่ยวกับการวิเคราะห์ข้อมูล
 - ❖ ควรทราบถึงกระบวนการในการวิเคราะห์ข้อมูล
 - ❖ ควรทราบถึงศัพท์เทคนิคต่าง ๆ ที่ใช้ในการวิเคราะห์ข้อมูล
- บุคลากรที่เกี่ยวข้องในทุกภาคส่วนของหน่วยงาน ควรมีความรู้ความเข้าใจในการวิเคราะห์ข้อมูล

ข้อควรพิจารณาในการวิเคราะห์ข้อมูล



7

โครงสร้างหน่วยงานควรเอื้อต่อการทำงาน

- ทีมงานวิเคราะห์ข้อมูลจำเป็นต้องใช้บุคลากรจากหลายภาคส่วน / งานในหน่วยงานร่วมกัน

8

หน่วยงานต้องกำหนดวัฒนธรรมองค์กรเดียวกัน

- หน่วยงานควรปลูกฝังหรือกำหนดให้บุคลากรทุกคนมีความรู้เข้าใจเกี่ยวกับ “การวิเคราะห์ข้อมูล”

ข้อควรพิจารณาในการวิเคราะห์ข้อมูล



บุคลากรควรมีความรู้ความเข้าใจเกี่ยวกับการวิเคราะห์ข้อมูล

บุคลากรที่ทำหน้าที่ในการวิเคราะห์ข้อมูลควรมีความรู้ด้านการวิเคราะห์ข้อมูลอย่างลึกซึ้ง ทั้งด้าน

- กระบวนการ / ขั้นตอนการวิเคราะห์ข้อมูล
- วิธีการวิเคราะห์ข้อมูล
- การนำเสนอข้อมูล
- การสื่อสารกับบุคคลอื่น ๆ



แบบจำลองที่ใช้ใน
การวิเคราะห์ข้อมูล
แบบพหุภาคีมีอะไรบ้าง?



แบบจำลองการวิเคราะห์ข้อมูลแบบพยากรณ์



Classification Model

- แบบจำลองการจัดหมวดหมู่ (Classification Model) เป็นแบบจำลองการวิเคราะห์เชิงคาดการณ์พื้นฐานของการวิเคราะห์พยากรณ์ โดยมีการจัดกลุ่มข้อมูลตามรูปแบบเรียนรู้จากข้อมูลที่ปรากฏ
- แบบจำลองนี้ถือเป็นแบบจำลองที่ดีที่สุดในการวิเคราะห์ข้อมูล ที่ช่วยให้ผู้ปฏิบัติงานใช้เป็นคู่มือในการตัดสินใจในการดำเนินงานต่าง ๆ

แบบจำลองการวิเคราะห์ข้อมูลแบบพยากรณ์



Classification Model

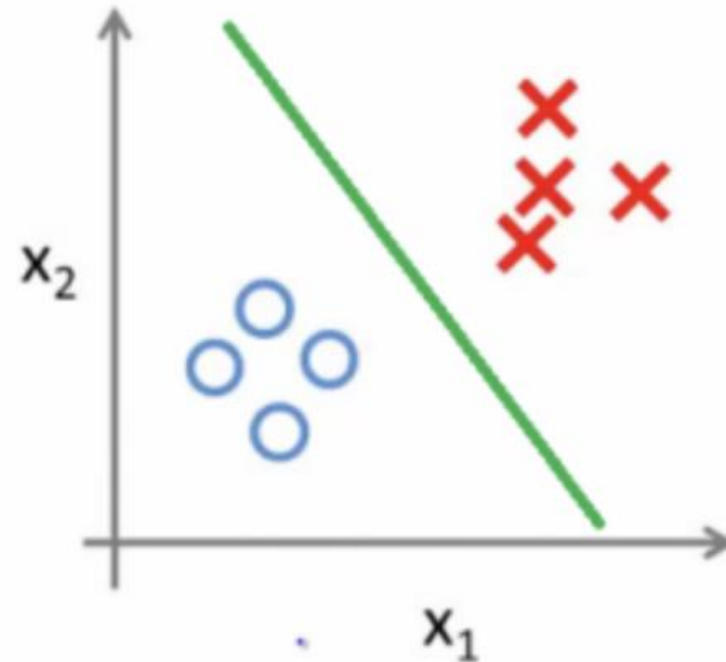
ตัวอย่างการใช้แบบจำลองการจัดหมวดหมู่ (Classification Model)

- ธุรกิจการค้าปลีก
“ลูกค้ากำลังจะเลิกกิจการหรือไม่”
- ธุรกิจการให้บริการเงินกู้
“เงินกู้จะได้รับการอนุมัติหรือไม่” หรือ
“ผู้สมัครใช้บริการเงินกู้ มีแนวโน้มที่จะผิดนัดชำระหนี้หรือไม่”
- ธุรกิจธนาคารออนไลน์
“ธุรกรรมที่ทำอยู่เป็นธุรกรรมที่หลอกลวงหรือไม่”

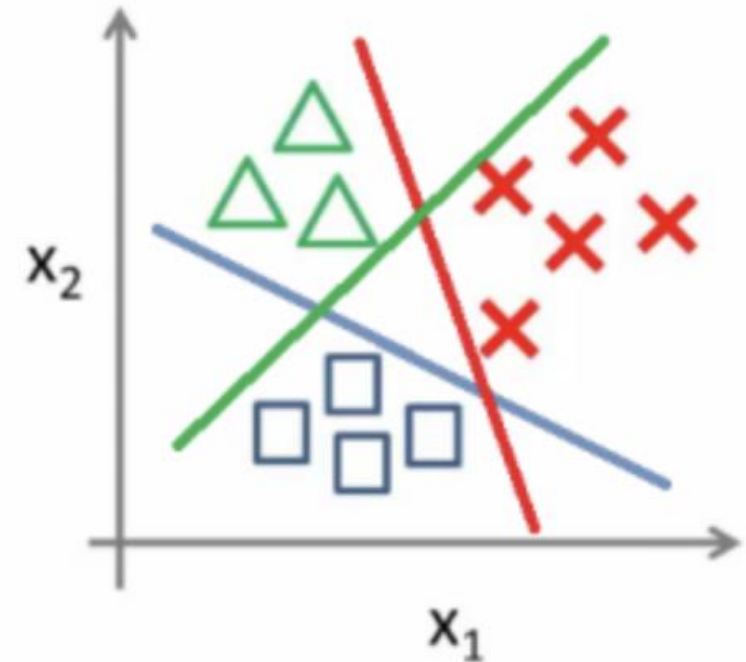
ตัวอย่างแบบจำลองการจัดหมวดหมู่ (Classification Model)



Binary classification:



Multi-class classification:



แบบจำลองการวิเคราะห์ข้อมูลแบบพยากรณ์



Clustering Model

- แบบจำลองการจัดกลุ่ม (Clustering Model) มุ่งเน้นจัดเรียงข้อมูลเป็นกลุ่ม โดยจำแนกข้อมูลตามคุณลักษณะที่คล้ายกัน/ต่างกันของข้อมูล
- การจำแนกแยกแยะข้อมูลด้วยแบบจำลองนี้จะทำให้หน่วยงานต่าง ๆ สามารถแบ่งหรือจำแนกผู้ใช้งาน ลูกค้า หรือประเด็นต่าง ๆ ตามคุณลักษณะที่กำหนดขึ้น หรือเงื่อนไขต่าง ๆ ที่หน่วยงานให้ความสนใจได้อย่างรวดเร็ว

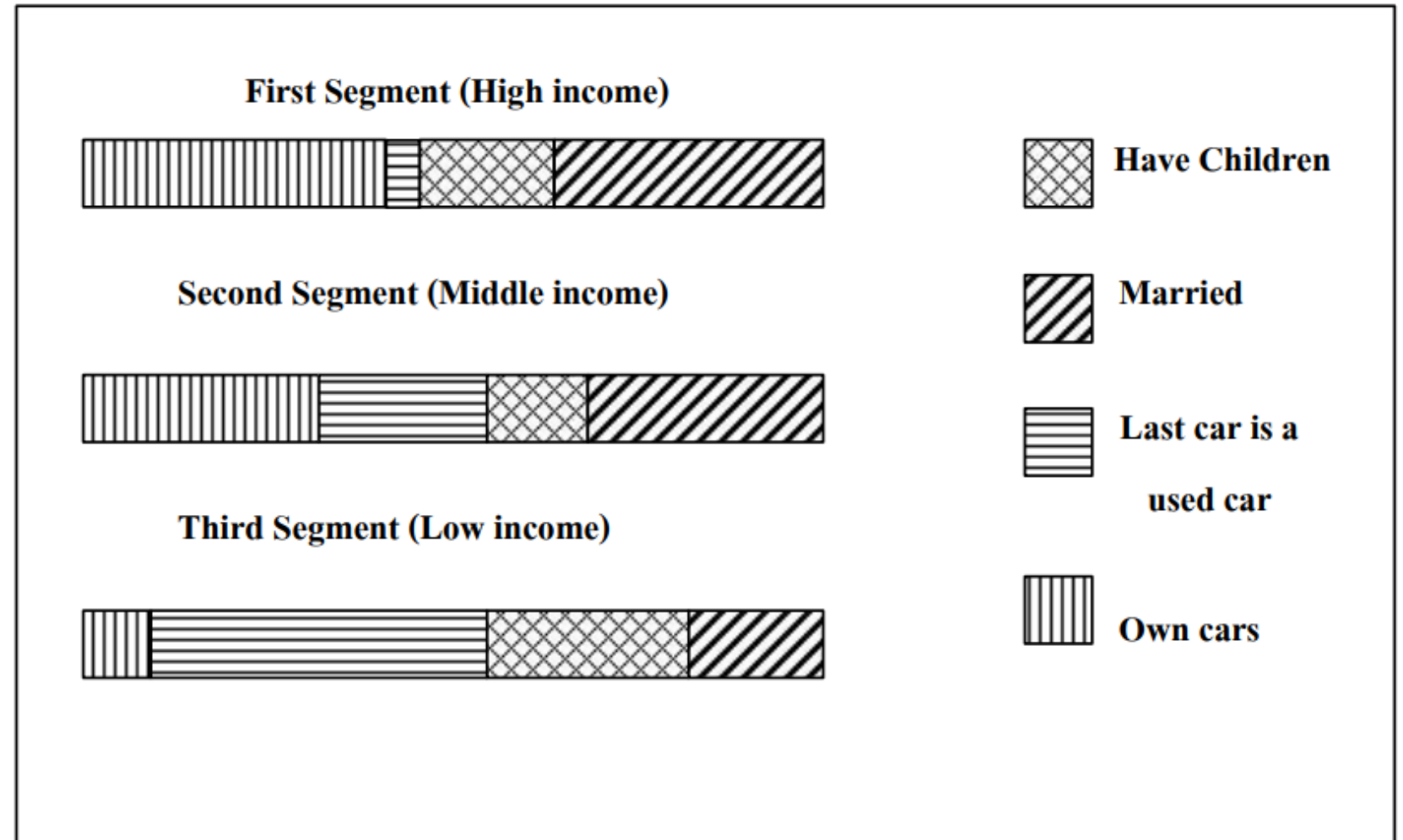
ตัวอย่างแบบจำลองการจัดกลุ่ม (Clustering Model)



บริษัทจำหน่ายรถยนต์ได้แยกกลุ่มลูกค้าออกเป็น 3 กลุ่ม คือ

- กลุ่มผู้มีรายได้สูง ($> \$80,000$)
- กลุ่มผู้มีรายได้ปานกลาง ($\$25,000$ to $\$80,000$)
- กลุ่มผู้มีรายได้ต่ำ (less than $\$25,000$)
- ลูกค้าในแต่ละกลุ่ม สามารถจำแนกเป็นกลุ่มย่อยได้ 4 กลุ่ม ดังนี้
 - ❖ Have Children
 - ❖ Married
 - ❖ Last car is a used car
 - ❖ Own cars

ตัวอย่างแบบจำลองการจัดกลุ่ม (Clustering Model)



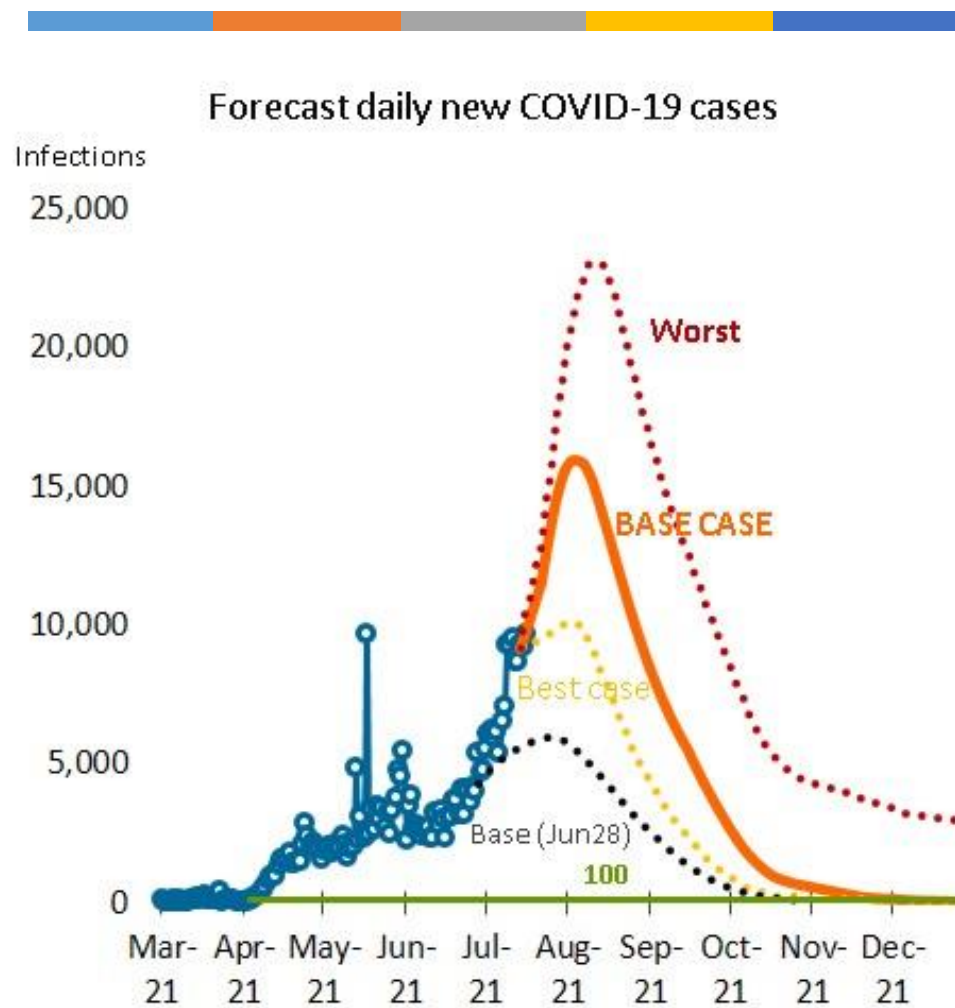
แบบจำลองการวิเคราะห์ข้อมูลแบบพยากรณ์



Forecast Model

- แบบจำลองพยากรณ์ (Forecast Model) เป็นแบบจำลองการพยากรณ์ที่เกี่ยวข้องกับ**การทำนายค่าข้อมูล (ข้อมูลเชิงปริมาณ)** โดยใช้การเรียนรู้จากข้อมูล (เดิม) ที่ปรากฏ
- แบบจำลองนี้อาจมีการใช้ค่าพารามิเตอร์ (Parameter) หลายตัวในการพยากรณ์ เช่น
 - ❖ การพยากรณ์อากาศ (Weather forecast)
 - ❖ การพยากรณ์จำนวนลูกค้าที่เข้ามาใช้บริการในแต่ละสัปดาห์
 - ❖ การพยากรณ์การระบาดของโรคต่าง ๆ

ตัวอย่างแบบจำลองพยากรณ์ (Forecast Model)



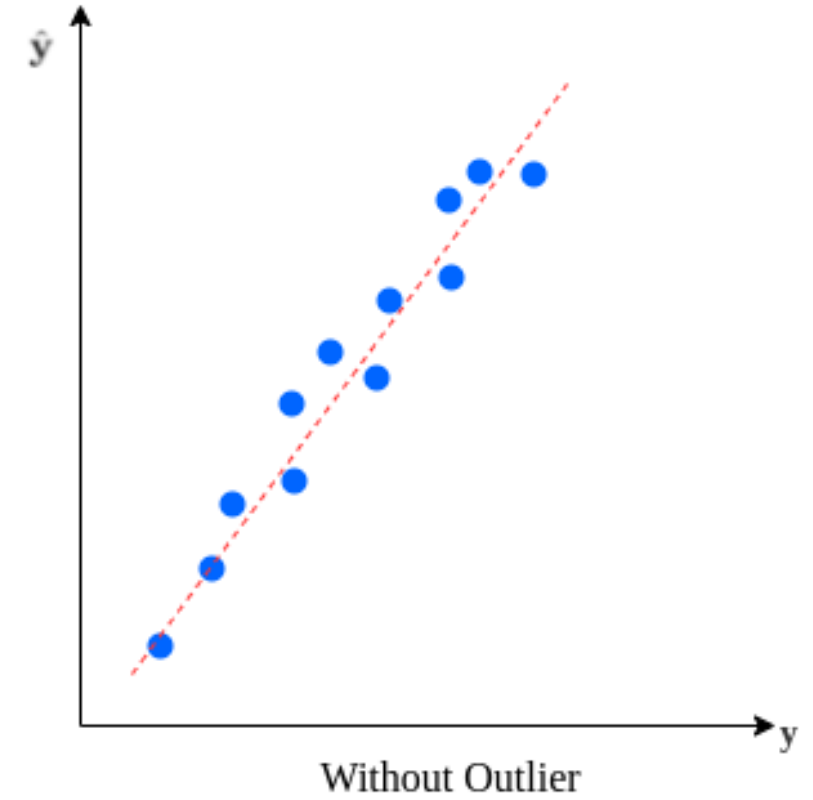
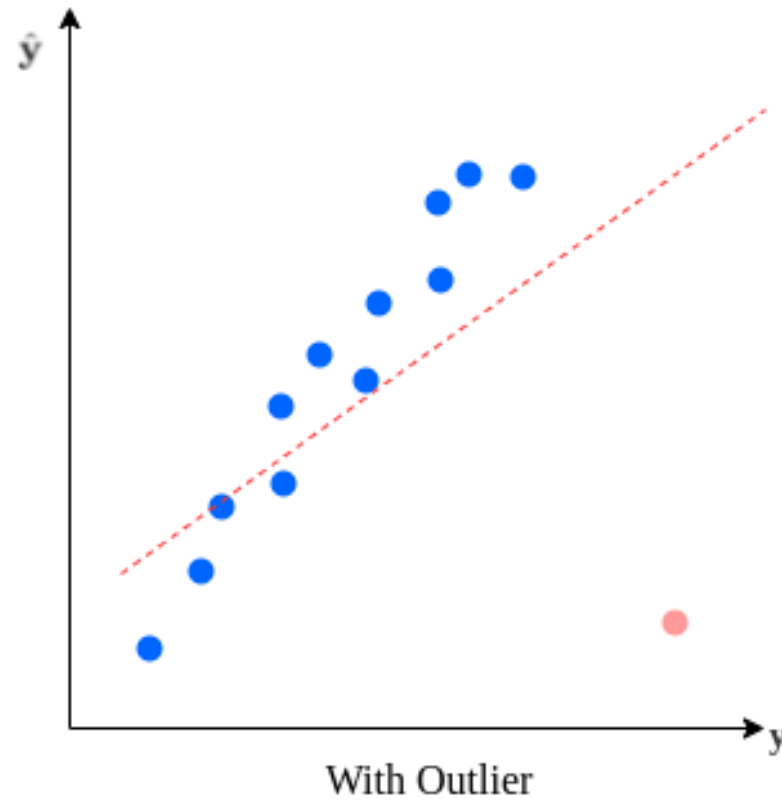
Note: forecast as of Jul 14, 2021
Source: WHO, Forecast by Krungsri Research

แบบจำลองการวิเคราะห์ข้อมูลแบบพยากรณ์

Outliers Model

- แบบจำลองค่าข้อมูลผิดปกติ (Outliers Model) เป็นแบบจำลองที่มุ่งเน้นการพิจารณาข้อมูลในชุดข้อมูล (Dataset) ที่มีความผิดปกติเป็นหลัก
- แบบจำลองนี้นิยมใช้ในธุรกิจการเงินและการธนาคาร และการค้าปลีก โดยวัตถุประสงค์เพื่อ
 - ❖ หาข้อมูลธุรกรรมที่มีความผิดปกติ
 - ❖ ลดปัญหาการฉ้อโกง
 - ❖ ลดข้อผิดพลาดที่อาจเกิดขึ้นกับสินค้า และ/หรือบริการต่าง ๆ

ตัวอย่างแบบจำลองค่าข้อมูลผิดปกติ (Outliers Model)



แบบจำลองการวิเคราะห์ข้อมูลแบบพยากรณ์

Time Series Model

- แบบจำลองอนุกรมเวลา (Time Series Model) มุ่งเน้นการใช้ลำดับจุดของข้อมูล (Data points) ที่มีการบันทึกไว้มาเป็นค่าพารามิเตอร์ในการวิเคราะห์ข้อมูลเป็นสำคัญ
- แบบจำลองนี้ช่วยให้ผู้วิเคราะห์เข้าใจชุดข้อมูลในช่วงเวลาใดช่วงเวลาหนึ่ง
- แบบจำลองนี้สามารถพยากรณ์ข้อมูลได้อย่างแม่นยำมากกว่าการใช้ค่าเฉลี่ยของข้อมูลต่าง ๆ

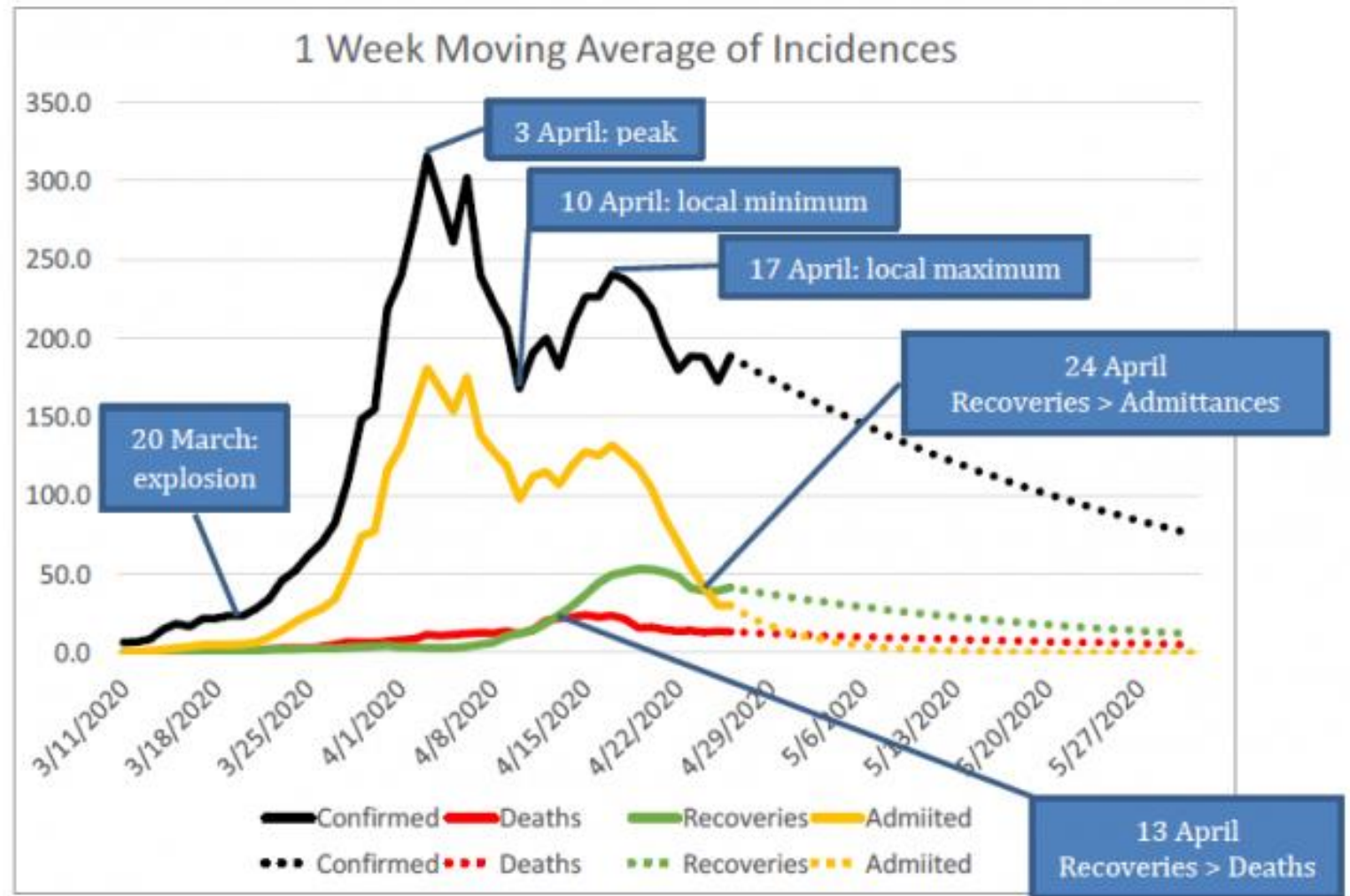
แบบจำลองการวิเคราะห์ข้อมูลแบบพยากรณ์

Time Series Model

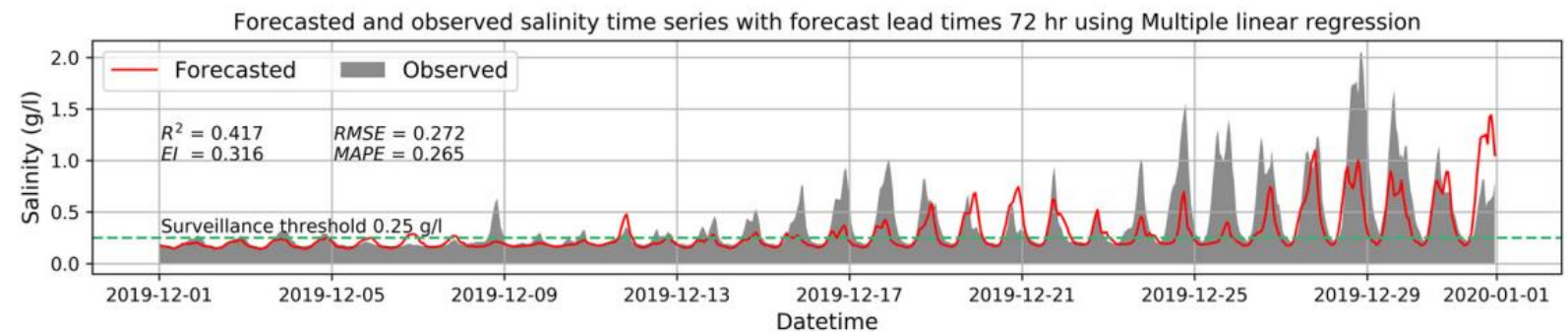
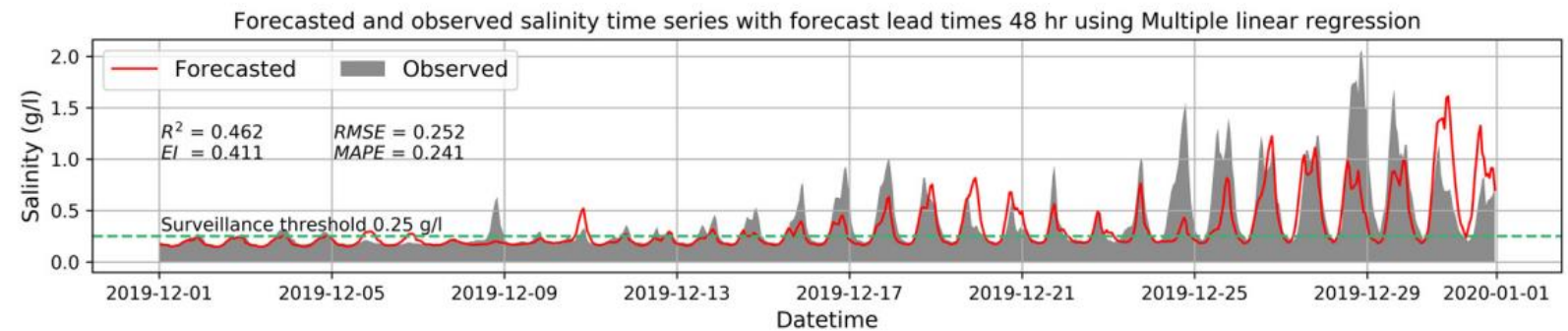
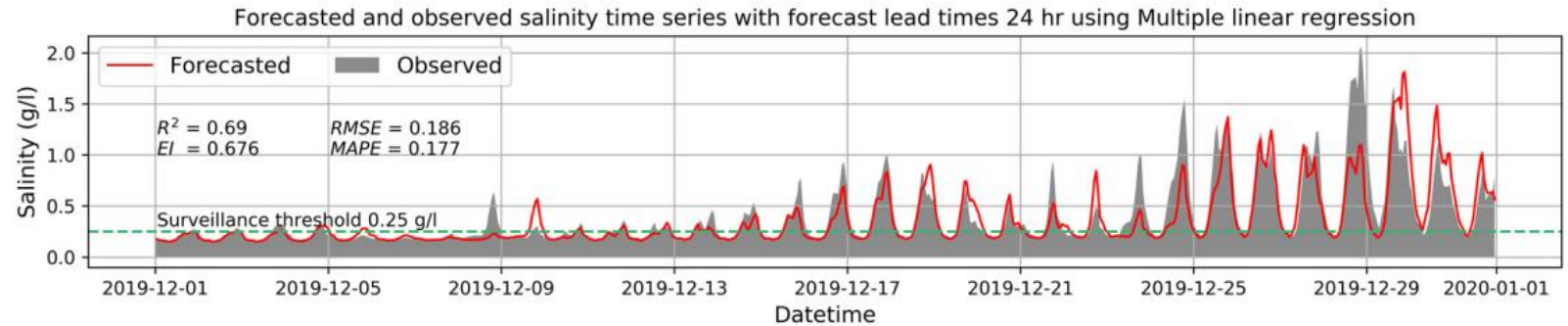
ตัวอย่างการใช้แบบจำลองอนุกรมเวลา

- การวิเคราะห์จำนวนผู้ป่วยที่เข้ารับบริการในโรงพยาบาล
- การวิเคราะห์จำนวนนักท่องเที่ยวในช่วงเทศกาล/วันหยุด
- การวิเคราะห์ปริมาณการใช้น้ำประปาในเขตพื้นที่เทศบาล
- การคาดการณ์ความเค็มรุกตัวในแม่น้ำเจ้าพระยา
- การพยากรณ์ปริมาณน้ำไหลเข้าอ่างเก็บน้ำ

ตัวอย่างแบบจำลองอนุกรมเวลา (Time Series Model)



ตัวอย่างแบบจำลองอนุกรมเวลา (Time Series Model)

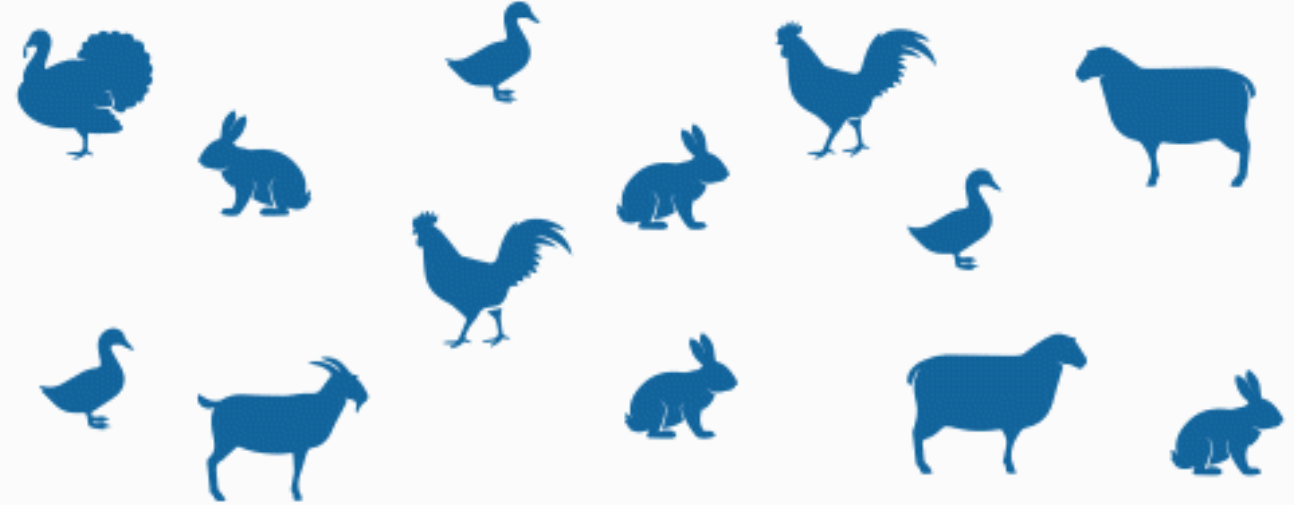




Classification

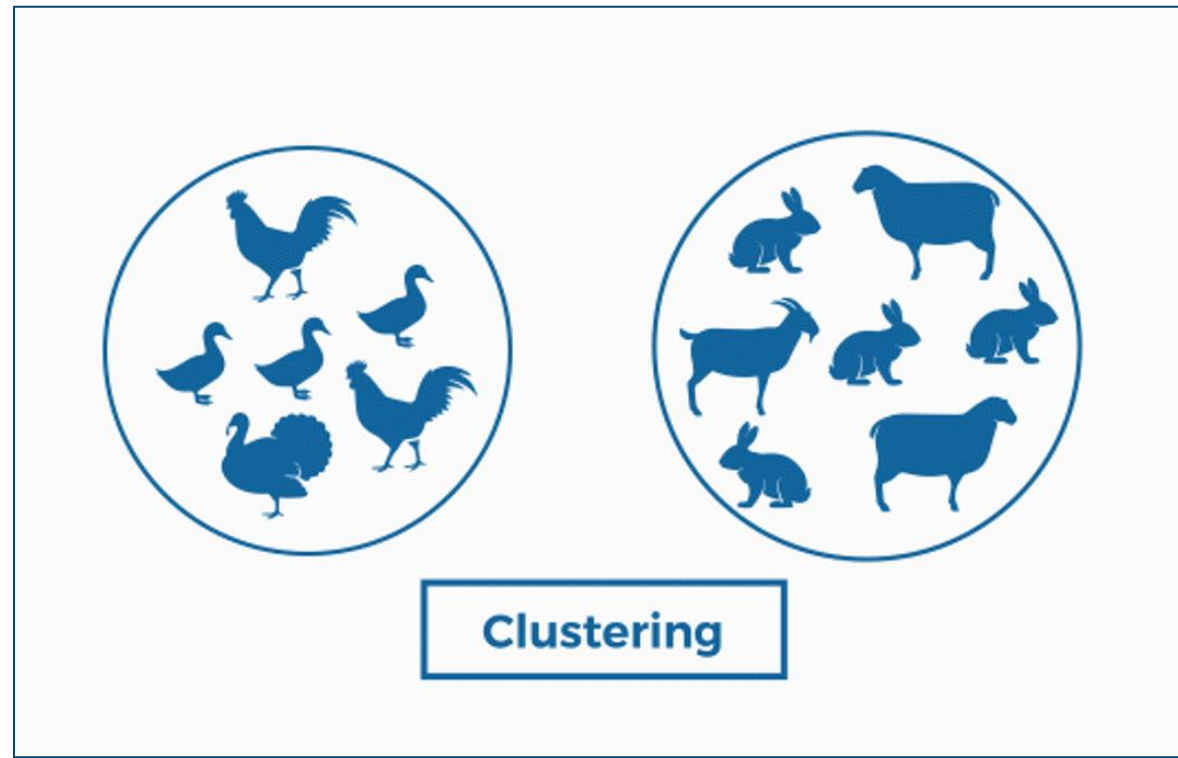
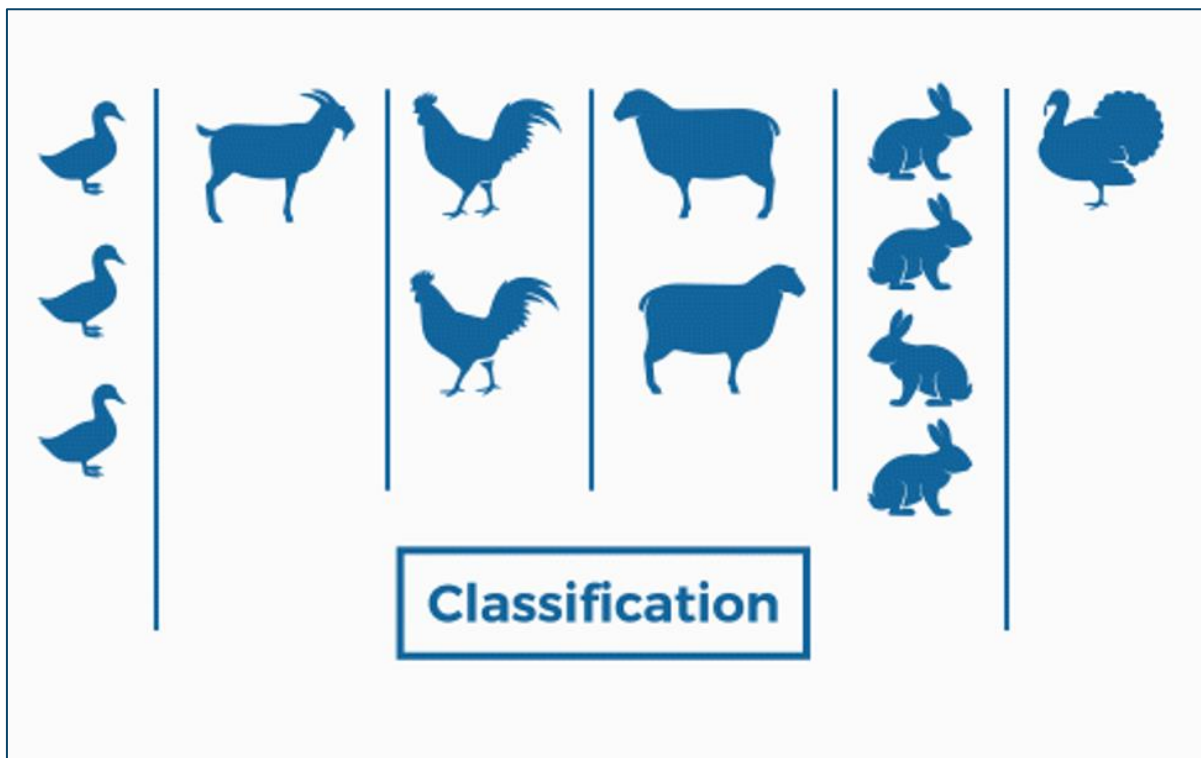
VS

Clustering



Classification

Clustering



Classification



- แบบจำลองการวิเคราะห์ข้อมูลประเภท Supervised Model
- แบบจำลองที่ต้องมีเป้าหมาย (Target) หรือตัวแปรที่ใช้วัดเป็นตัวกำหนดหรือต้นแบบ
- เป้าหมายของการจัดกลุ่มแบบ Classification คือ
 - ❖ การจำแนกข้อมูลเป็นกลุ่ม
 - ❖ การจำแนกข้อมูลที่สามารถจำแนกแยกแยะได้อย่างชัดเจน (Discrete) เช่น Yes/No เป็นต้น
- การประเมินผลลัพธ์ที่ได้จากการจัดกลุ่มแบบ Classification Model สามารถวัดค่าความแม่นยำ (Accuracy) ได้ เช่น การใช้ Confusion Matrix

Classification



- Classification เป็นกระบวนการสร้างโมเดลจัดการข้อมูลด้วยการจำแนกแยกแยะข้อมูลตามกลุ่มที่กำหนดไว้ เช่น
 - ❖ การจัดกลุ่มนักเรียนตามประวัติและผลการเรียน
 - ❖ การแบ่งประเภทลูกค้าตามรายได้
 - ❖ การแบ่งประเภทลูกค้าตามค่าใช้จ่ายในการซื้อสินค้าและบริการ
- การจัดกลุ่มแบบ Classification อาจไม่สามารถจำแนกคุณลักษณะของข้อมูลได้อย่างชัดเจน

Classification

วิธีการจำแนกข้อมูลด้วยวิธีการ Classification ได้แก่

1

Decision Tree

2

Naïve Bayes Method

3

K-nearest Neighbor (K-NN)

Classification

1

การวิเคราะห์ข้อมูลด้วยวิธี Decision Tree

- การนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบโครงสร้างต้นไม้
- รูปแบบการทำงาน
 - ❖ มีลักษณะแบบ Supervised Learning ซึ่งเป็นการเรียนรู้ของโมเดลแบบมีครูสอน
 - ❖ สามารถสร้างแบบจำลองการจัดหมวดหมู่ได้จากกลุ่มตัวอย่างข้อมูลที่กำหนดไว้ล่วงหน้า
 - ❖ สามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่

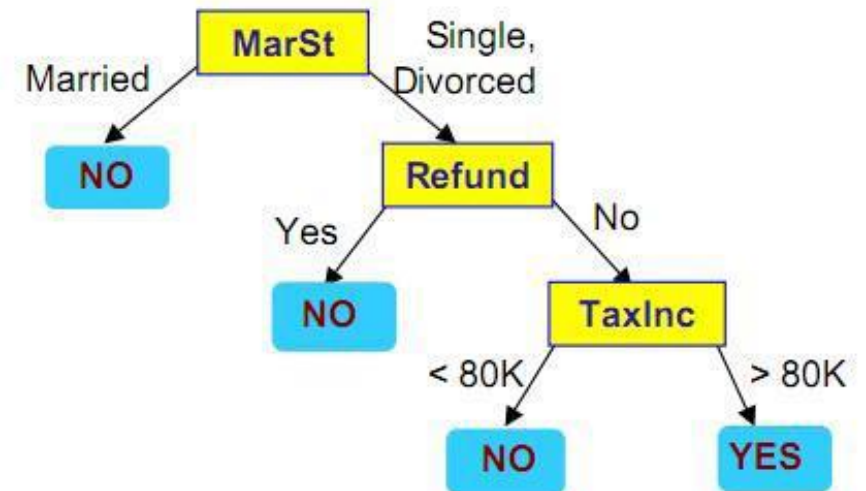
Classification

1

การวิเคราะห์ข้อมูลด้วยวิธี Decision Tree

➤ โครงสร้างของ Decision Tree ประกอบด้วย

- ❖ Root Node
- ❖ Child
- ❖ Leaf Node



การวิเคราะห์ข้อมูลด้วยวิธี Decision Tree



ตัวอย่างที่ 1

การพยากรณ์ว่าลูกค้าปัจจุบันที่เข้าสันทรัพย์อย่างใดอย่างหนึ่งไป
แล้วจะมีโอกาสตัดสินใจซื้อสันทรัพย์ชนิดนั้นไปเป็นของตนเองหรือไม่ ?

ปัจจัยที่ใช้ในการวิเคราะห์ คือ

ระยะเวลาที่ลูกค้าได้เข้าสันทรัพย์

อายุของลูกค้า



การวิเคราะห์ข้อมูลด้วยวิธี Decision Tree



ตัวอย่างที่ 1

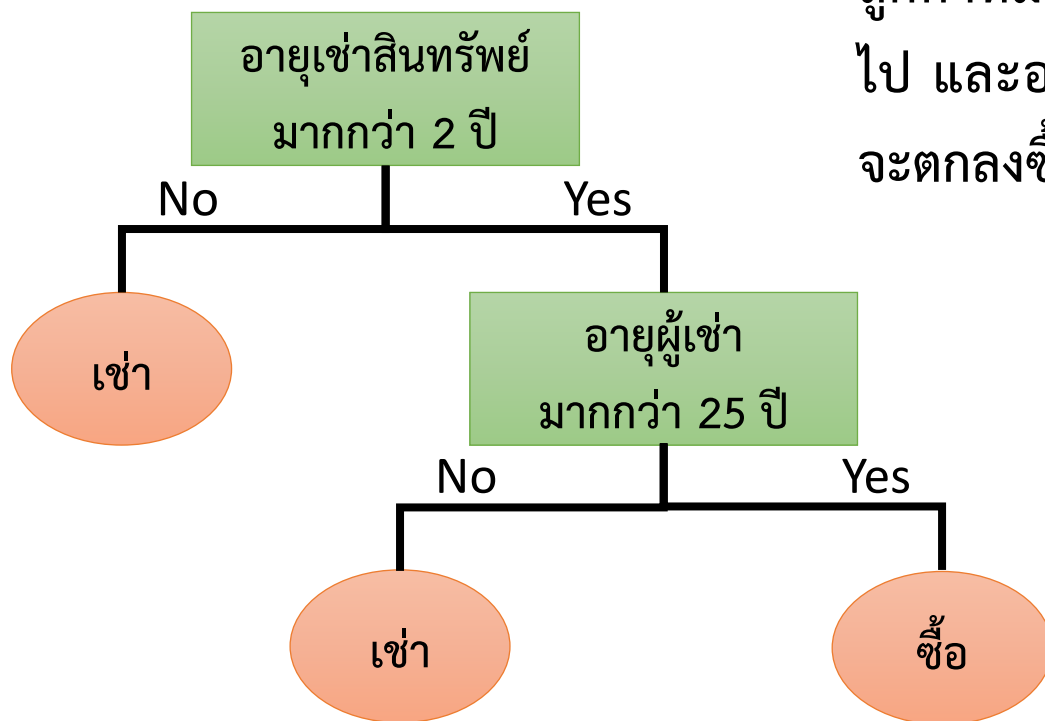


อายุเช่าสินทรัพย์	อายุผู้เช่า	การซื้อสินทรัพย์
2	25	Yes
3	22	Yes
1	28	No
5	30	Yes
4	27	No
2	21	No
3	23	No
...

การวิเคราะห์ข้อมูลด้วยวิธี Decision Tree

ตัวอย่างที่ 1

จากข้อมูลพบว่า
ลูกค้าที่มีอายุเช่าสินทรัพย์ตั้งแต่ 2 ปีขึ้นไป และอายุของผู้เช่าตั้งแต่ 25 ปีขึ้นไป จะตกลงซื้อสินทรัพย์เป็นของตนเอง



การวิเคราะห์ข้อมูลด้วยวิธี Decision Tree



ตัวอย่างที่ 2

การพยากรณ์การตัดสินใจในการเล่นกีฬา?

ปัจจัยที่ใช้ในการวิเคราะห์ คือ

ทัศนวิสัย (Outlook)

อุณหภูมิ (Temperature)

ความชื้นสัมพัทธ์ (Humidity)

ลมแรง (Windy)



การวิเคราะห์ข้อมูลด้วยวิธี Decision Tree

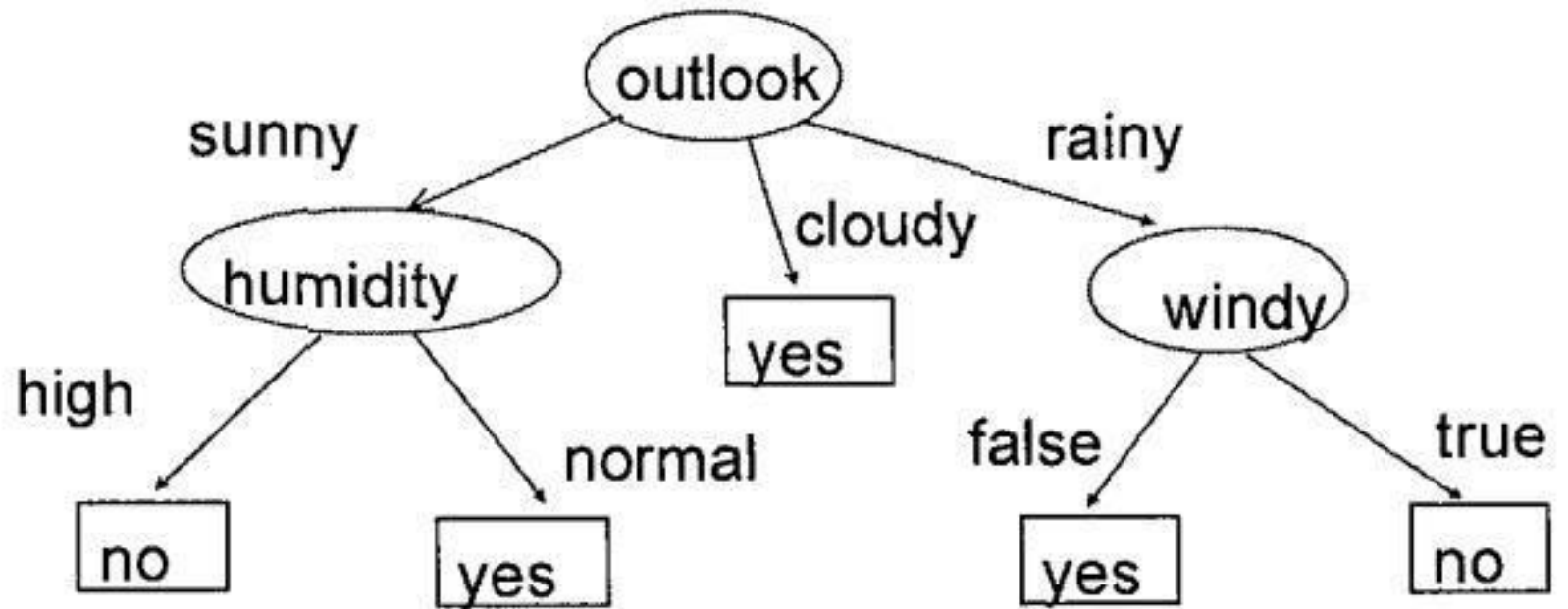
ตัวอย่างที่ 2

Outlook	Temperature	Humidity	Windy	Play(?)
sunny	hot	high	false	no
sunny	hot	high	true	no
cloudy	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
cloudy	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
cloudy	mild	high	true	yes
cloudy	hot	normal	false	yes
rainy	mild	high	true	no

การวิเคราะห์ข้อมูลด้วยวิธี Decision Tree

ตัวอย่างที่ 2

สรุปผลการพยากรณ์การตัดสินใจในการเล่นกีฬา



Classification

2

การวิเคราะห์ข้อมูลด้วยวิธี Naïve Bayes Method

- การวิเคราะห์ข้อมูลโดยหลักความน่าจะเป็น และใช้ทฤษฎีของ Bayes ในการแก้ปัญหา
- การวิเคราะห์ข้อมูลด้วยวิธี Naïve Bayesian Classification เป็นการหาความน่าจะเป็นของสิ่งที่ยังไม่เคยเกิดขึ้น ด้วยการคาดเดาจากสิ่งที่เคยเกิดขึ้นมาก่อน

การวิเคราะห์ข้อมูลด้วยวิธี Naïve Bayes Method



Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

การวิเคราะห์ข้อมูลด้วยวิธี Naïve Bayes Method



outlook	temperature	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

Classification



การวิเคราะห์ข้อมูลด้วยวิธี K-nearest Neighbor (K-NN)

- วิธีที่ง่ายและมีประสิทธิภาพ แต่การประมวลผลช้า
- ใช้หลักการเปรียบเทียบข้อมูลที่สนใจกับข้อมูลอื่นว่ามี ความคล้ายคลึงมากน้อยเพียงใด
- กรณีที่ข้อมูลที่กำลังวิเคราะห์อยู่ใกล้ข้อมูลใดมากที่สุด ระบบจะให้คำตอบเป็นเหมือนคำตอบของข้อมูลที่อยู่ใกล้ ที่สุดนั้น

Classification

3

การวิเคราะห์ข้อมูลด้วยวิธี K-nearest Neighbor (K-NN)

- การวิเคราะห์ข้อมูลด้วยวิธีการนี้ นิยมใช้ในการพยากรณ์ข้อมูลที่เป็นแบบประเภทนามบัญญัติ (Nominal) เช่น
 - ❖ เพศ (หญิง / ชาย)
 - ❖ ระดับการศึกษา (ต่ำกว่าประถม / ประถม / มัธยม / ปริญญาตรี / ปริญญาโท / ปริญญาเอก)
 - ❖ สถานภาพ (โสด / สมรส / หม้าย)

Classification



การวิเคราะห์ข้อมูลด้วยวิธี K-nearest Neighbor (K-NN)

- การวิเคราะห์ข้อมูลด้วยวิธีนี้ เป็นวิธีการที่ไม่ซับซ้อนและเข้าใจง่ายที่สุดที่ใช้ในการจำแนกประเภทข้อมูล
- มีลักษณะการทำงานแบบ Unsupervised learning ไม่ใช้ข้อมูลฝึกหัด (Training data) ในการสร้างแบบจำลอง แต่ใช้ข้อมูลมาเป็นตัวแบบจำลอง

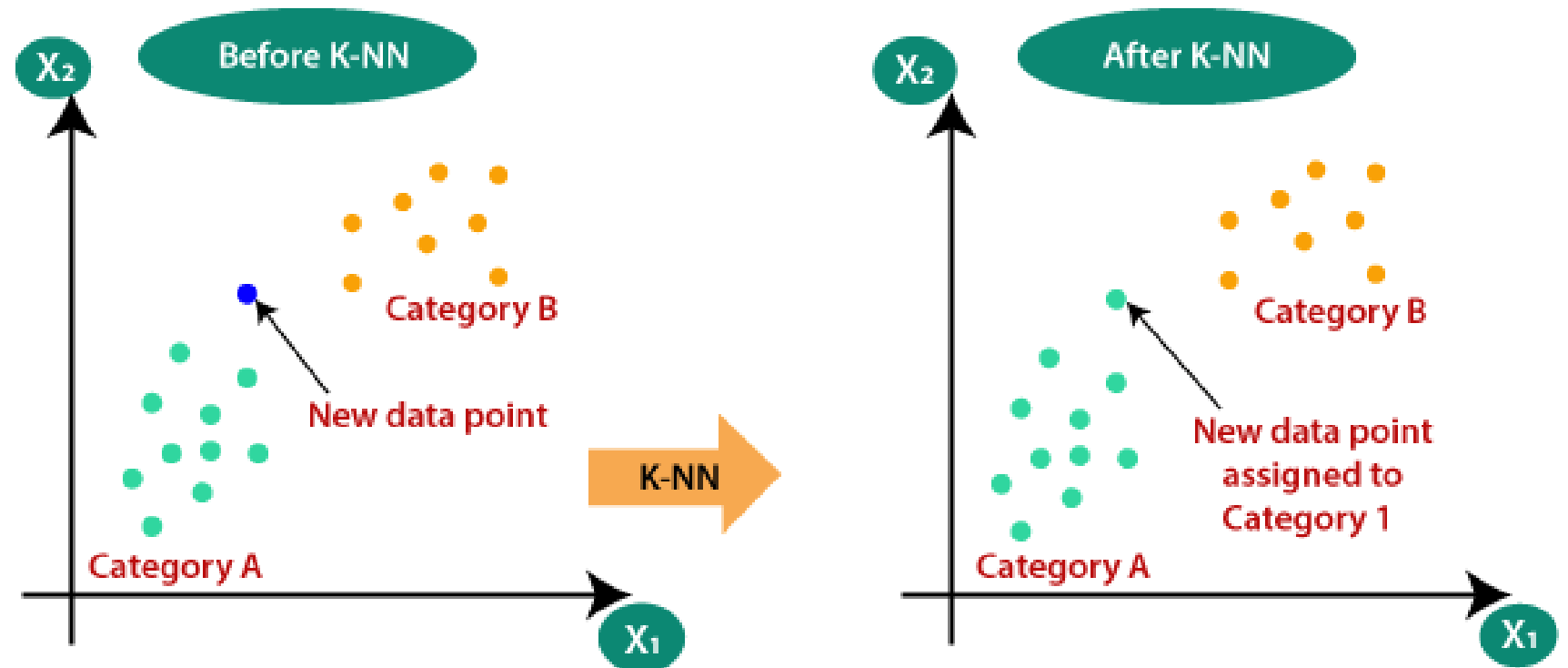
Classification



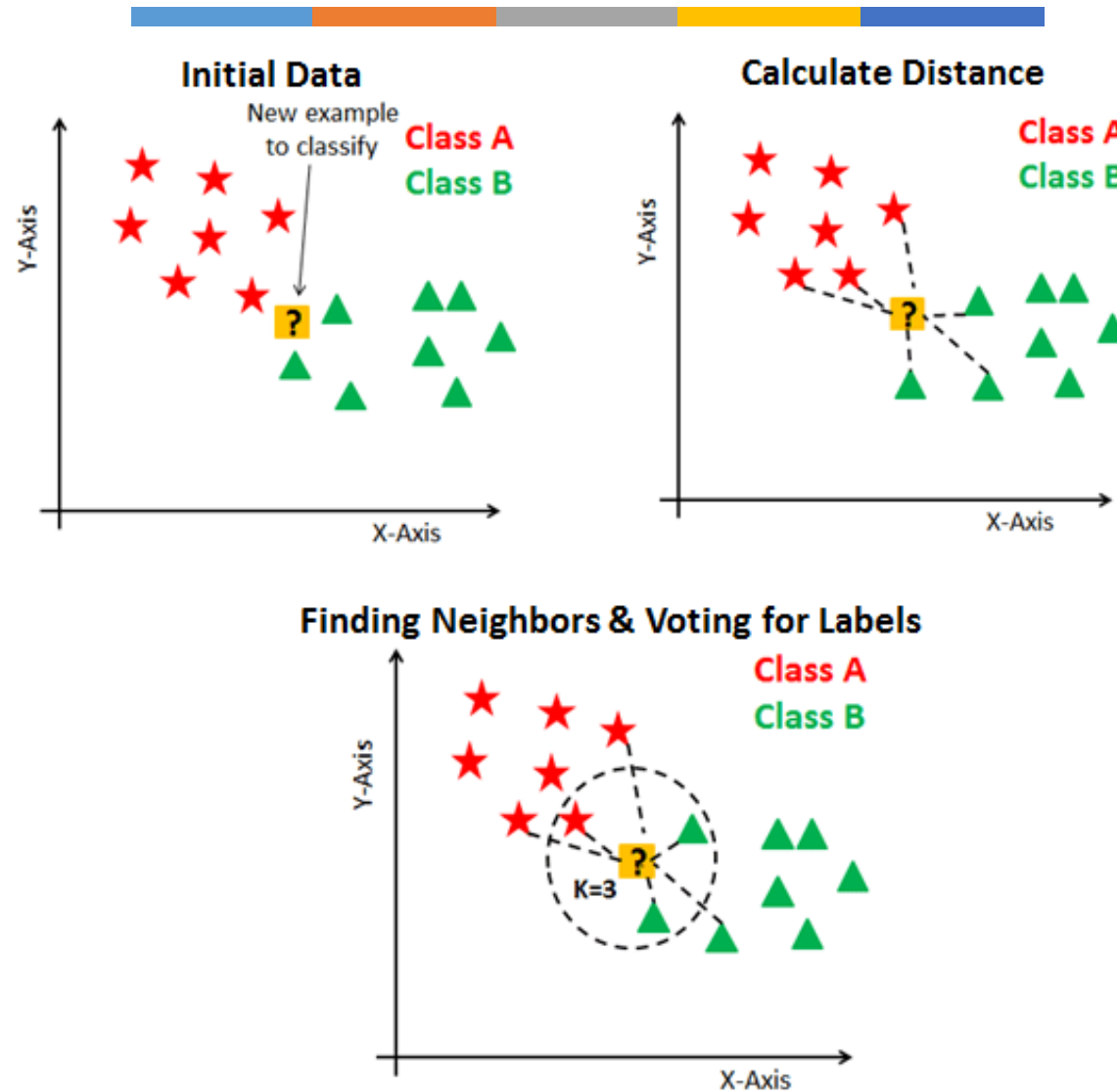
การวิเคราะห์ข้อมูลด้วยวิธี K-nearest Neighbor (K-NN)

- การวิเคราะห์ข้อมูลด้วยวิธีนี้พิจารณาคำตอบจาก
 - ❖ คำตอบของข้อมูลเรียนรู้ที่อยู่ใกล้ที่สุด K ตัวที่พบมากที่สุดเป็นคำตอบ
 - ❖ ให้ค่าน้ำหนักโดยการพิจารณาระยะห่างระหว่างข้อมูลที่สนใจกับข้อมูลที่อยู่ใกล้ที่สุด K ตัว

การวิเคราะห์ข้อมูลด้วยวิธี K-nearest Neighbor (K-NN)



การวิเคราะห์ข้อมูลด้วยวิธี K-nearest Neighbor (K-NN)



Clustering



- แบบจำลองการวิเคราะห์ข้อมูลประเภท Unsupervised Model
- แบบจำลองที่ไม่มีเป้าหมาย (Target) เป็นตัวกำหนดหรือต้นแบบ จึงไม่สามารถวัดผลเชิงค่าความแม่นยำ (Accuracy) ได้
- การจัดกลุ่มแบบ Clustering จะมุ่งเน้นการจำแนกข้อมูลด้วยการใช้คุณลักษณะ (Attribute) ในการจำแนกแยกแยะข้อมูล
- การประเมินผลลัพธ์ที่ได้จากการจัดกลุ่มแบบ Clustering Model นิยมใช้ประสบการณ์ของผู้ใช้ (User Experience) เป็นหลัก จากนั้นจึงจำแนกด้วยแบบจำลองการพยากรณ์ (Forecasting Model) ซึ่งมีความแม่นยำมากขึ้น

Clustering



Clustering เป็นการจัดกลุ่มข้อมูลที่มีความแตกต่างจากการจัดกลุ่มข้อมูลแบบ Classification

- จัดกลุ่มข้อมูลตามคุณลักษณะ (ความเหมือน/ความแตกต่าง)
- ไม่มีการกำหนดคลาสประเภทข้อมูลไว้ก่อน หรือไม่ทราบจำนวนกลุ่มล่วงหน้า
- เป็นการเรียนรู้แบบไม่มีผู้สอน (Unsupervised classification)

Clustering



Clustering เป็นกระบวนการจัดกลุ่มวัตถุตามคุณลักษณะหรือคุณสมบัติของวัตถุ

- การจัดกลุ่มวัตถุตามคุณลักษณะที่เหมือนกัน (Similarity)
- การจัดกลุ่มวัตถุตามคุณลักษณะที่แตกต่างกัน (Dissimilarity or Distance)

Clustering



วิธีการจำแนกข้อมูลด้วยวิธีการ Clustering ได้แก่



การวิเคราะห์กลุ่มแบบไม่เป็นขั้นตอน
(Nonhierarchical Clustering or K-means Clustering)



การวิเคราะห์กลุ่มแบบขั้นตอน
(Hierarchical Clustering)

Clustering



การวิเคราะห์ข้อมูลด้วยวิธี K-means Clustering

การวิเคราะห์กลุ่มแบบไม่เป็นขั้นตอน (Nonhierarchical Cluster Analysis) หรือ การแบ่งส่วน (Partitioning)

- โดยตัดแบ่ง (Partition) วัตถุออกเป็น K กลุ่ม
- แทนค่าแต่ละกลุ่มด้วยค่าเฉลี่ยของกลุ่ม ซึ่งใช้เป็นจุดศูนย์กลาง (Centroid) ของกลุ่มในการวัดระยะห่างของข้อมูลในกลุ่มเดียวกัน

Clustering



การวิเคราะห์ข้อมูลด้วยวิธี K-means Clustering

ข้อดีของการวิเคราะห์ข้อมูลด้วยวิธี K-means

- การหาค่าเฉลี่ยแบบ K-means สามารถคำนวณหาค่าได้เร็วกว่าการจัดกลุ่มแบบ Hierarchical
- ขั้นตอนการหาค่าเฉลี่ยแบบ K-means จะได้สมาชิกในกลุ่มหนาแน่นกว่าการจัดกลุ่มแบบ Hierarchical โดยเฉพาะถ้ากลุ่มเป็นวงกลม



Clustering



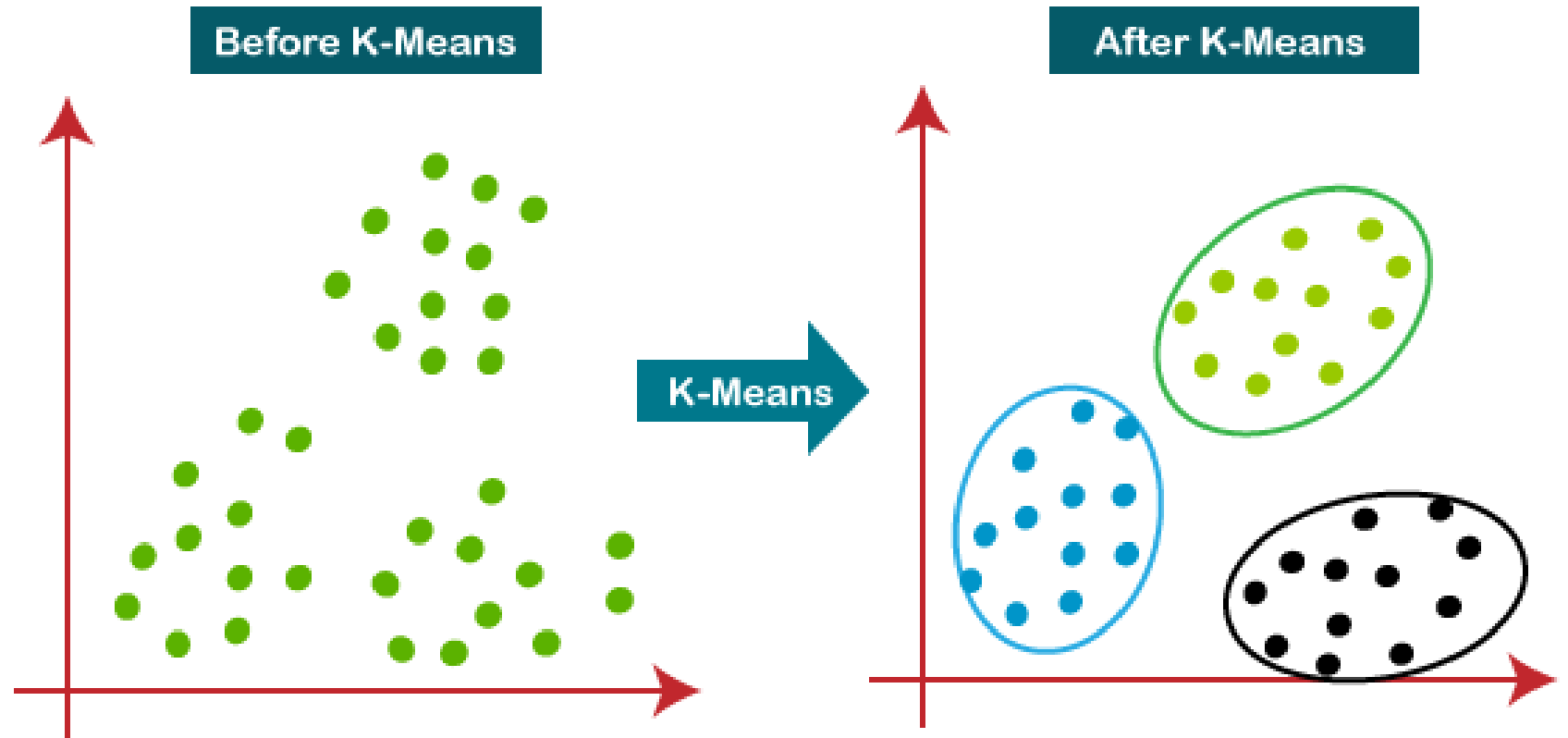
1

การวิเคราะห์ข้อมูลด้วยวิธี K-means Clustering

ข้อจำกัดของการวิเคราะห์ข้อมูลด้วยวิธี K-means

- การหาค่า K ที่เหมาะสมนั้นอาจคาดเดาได้ยาก
- การวิเคราะห์ข้อมูลอาจทำได้ไม่ดี หากกลุ่มข้อมูลไม่เป็นรูปร่างกลม
- การวิเคราะห์ข้อมูลอาจมีข้อจำกัดในเรื่องของขนาด ความหนาแน่น และรูปร่าง

การวิเคราะห์ข้อมูลด้วยวิธี K-means Clustering



การวิเคราะห์ข้อมูลด้วยวิธี K-means Clustering



Clustering

2

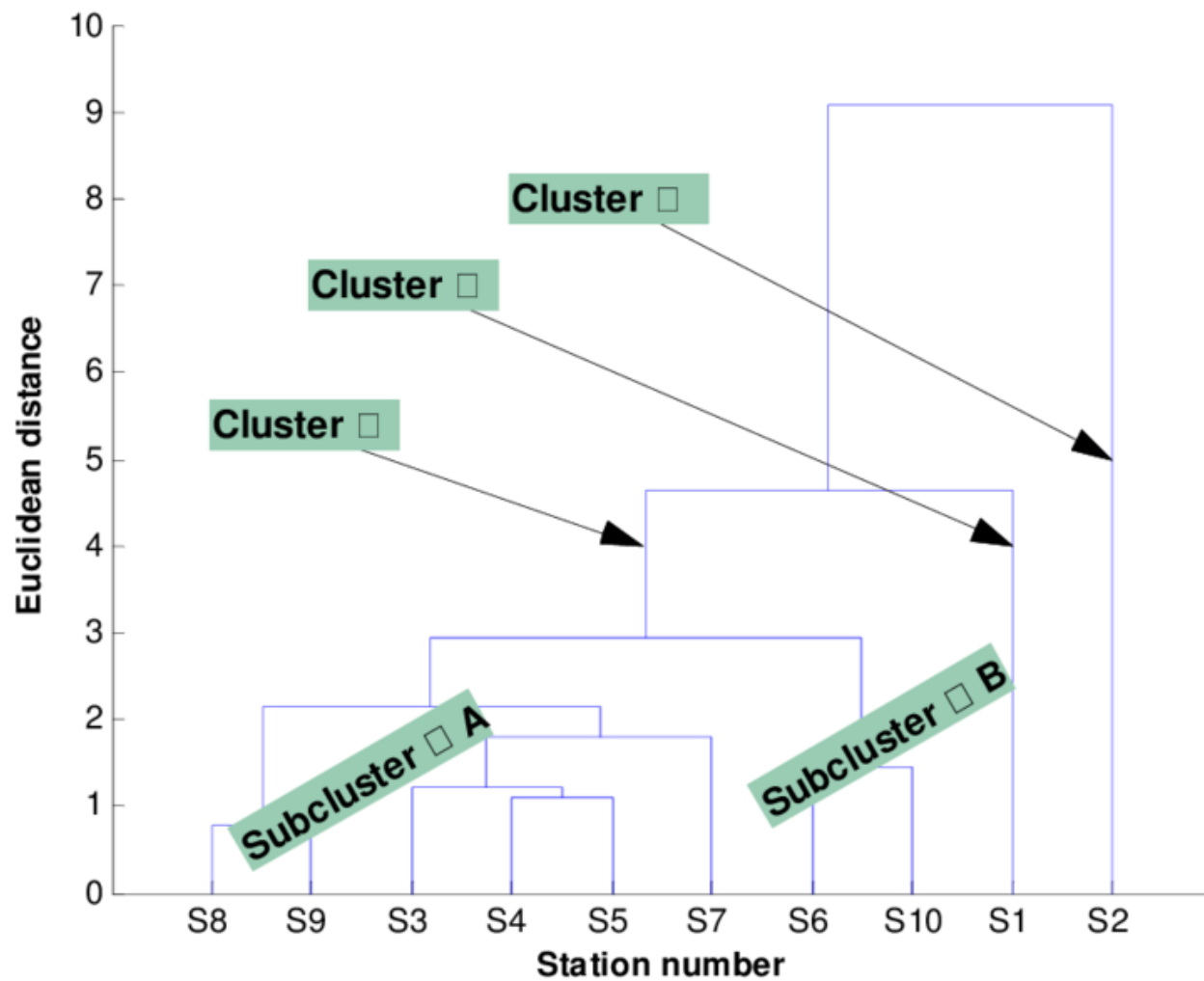
การวิเคราะห์ข้อมูลด้วยวิธี Hierarchical Clustering

การวิเคราะห์ข้อมูลที่นิยมใช้ในการจัดกลุ่ม Case หรือจัดกลุ่มตัวแปร โดยมีเงื่อนไขดังนี้

- วิธีนี้เหมาะสำหรับข้อมูลที่มีจำนวนตัวแปรไม่มาก
- วิธีนี้เหมาะสำหรับข้อมูลที่ไม่เกิน 200 ชุด (หากมีข้อมูลตั้งแต่ 200 ชุด ขึ้นไป ควรใช้วิธี K-Means Clustering)
- วิธีนี้ไม่จำเป็นต้องทราบจำนวนกลุ่มก่อนการวิเคราะห์
- วิธีนี้ไม่จำเป็นต้องทราบว่าข้อมูลมีตัวแปรใด หรือข้อมูลอยู่ในกลุ่มใด

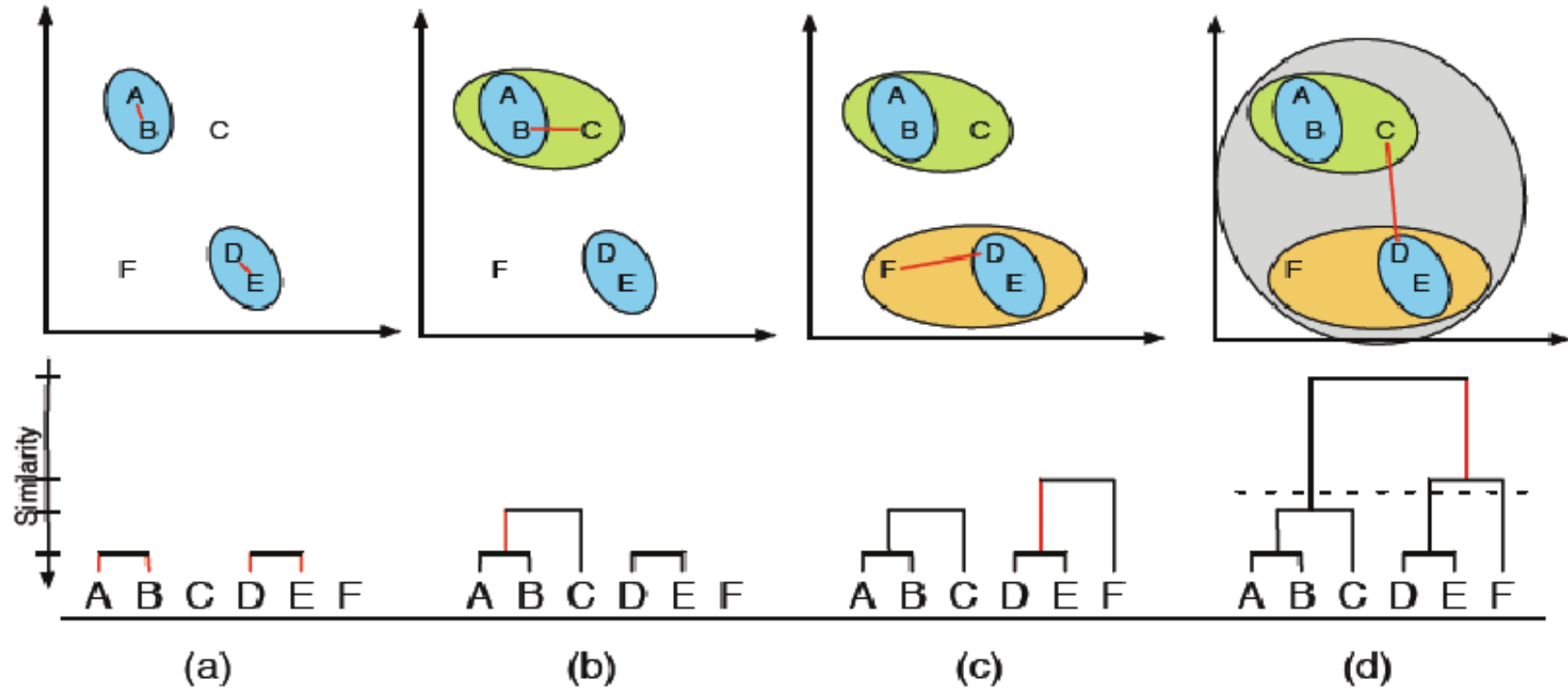


การวิเคราะห์ข้อมูลด้วยวิธี Hierarchical Clustering



การวิเคราะห์ข้อมูลด้วยวิธี Hierarchical Clustering

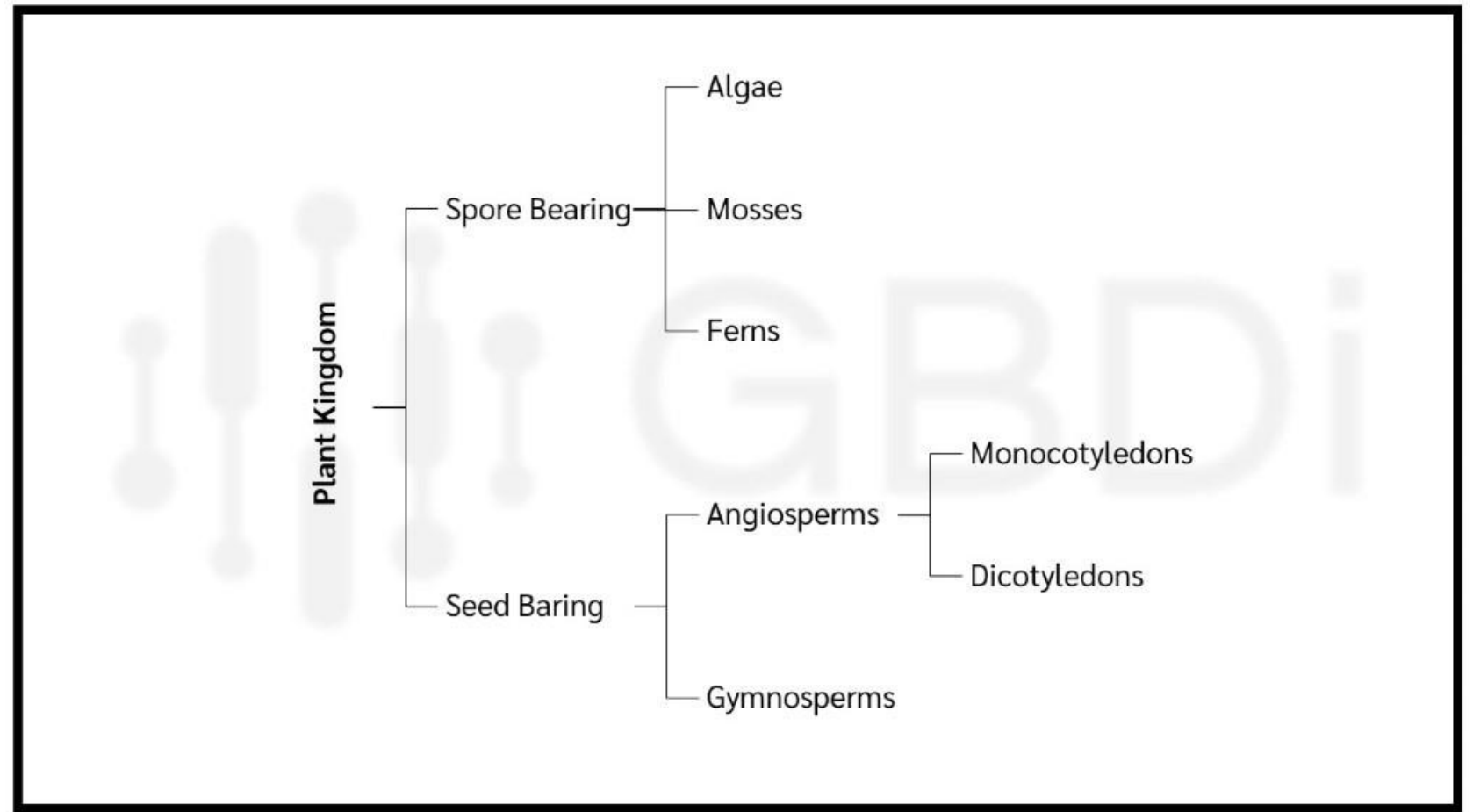
Example: Hierarchical Agglomerative Clustering



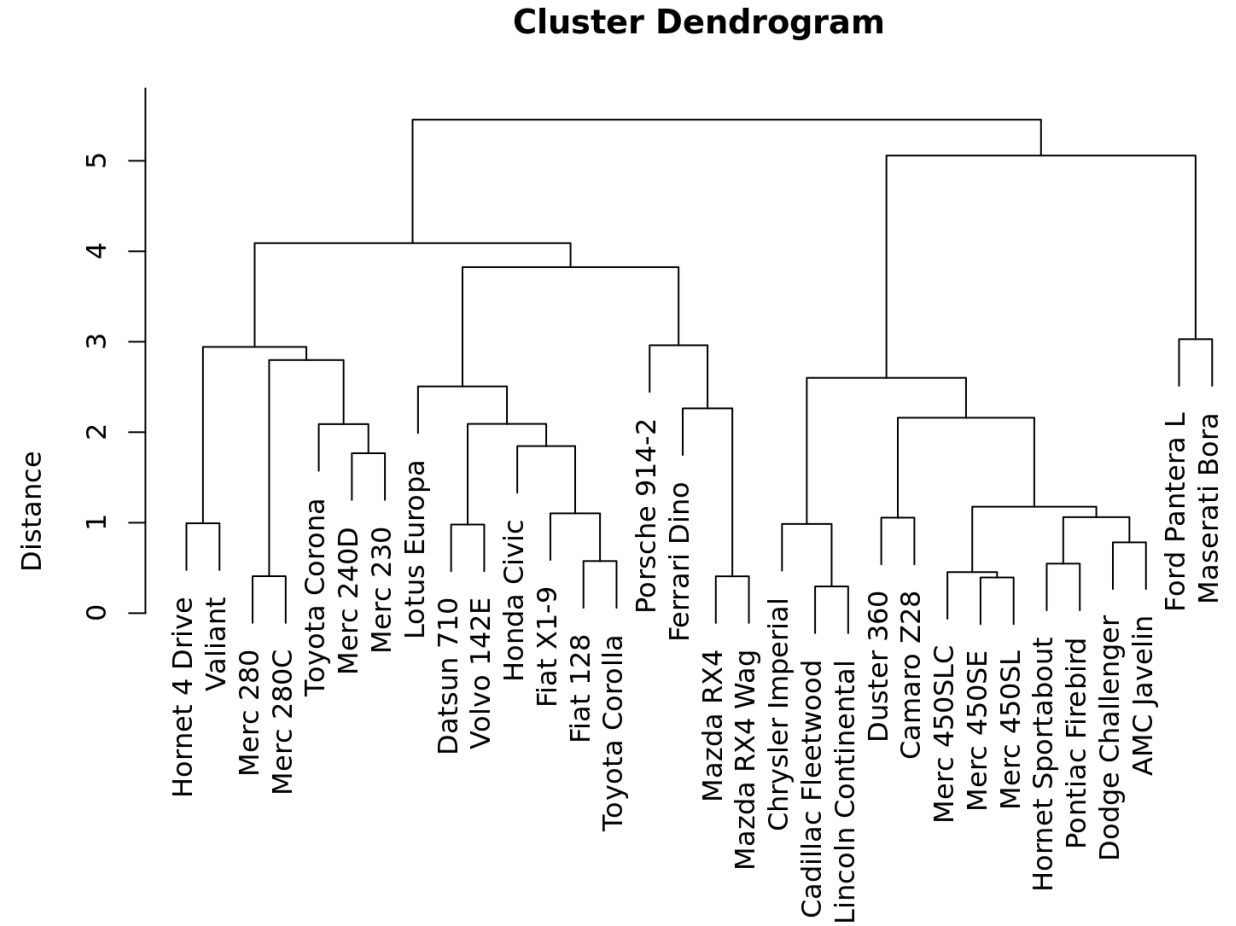
ที่มา

https://www.researchgate.net/figure/Example-of-hierarchical-clustering-clusters-are-consecutively-merged-with-the-most_fig3_273456906

การวิเคราะห์ข้อมูลด้วยวิธี Hierarchical Clustering



การวิเคราะห์ข้อมูลด้วยวิธี Hierarchical Clustering





Porn-anant Iamkhajornchai, Ph.D.

Institute of Digital Arts and Science, Suranaree University of Technology

e-mail: pornanant@sut.ac.th