

ภาพรวมของวิทยาการข้อมูล

ปัจจุบันเป็นยุคสมัยที่ข้อมูลมีปริมาณมหาศาล อันเนื่องมาจากการพัฒนาทางด้านวิทยาศาสตร์และเทคโนโลยี ในทุกๆ วินาทีมีข้อมูลจำนวนมากเกิดขึ้น และหลงเหลืออยู่ในระบบเครือข่ายคอมพิวเตอร์ อินเทอร์เน็ต และการสื่อสาร ข้อมูลที่ถูกจัดเก็บและรวบรวมไว้ถูกนำมาใช้ประโยชน์ในรูปแบบต่างๆ เช่น การอธิบายปรากฏการณ์ต่างๆ การทำนาย การตัดสินใจและการวางแผนทางธุรกิจ ในการใช้ประโยชน์จากข้อมูลนั้น ข้อมูลดิบมักจะถูกแปลงให้อยู่ในรูปสารสนเทศ ก่อนนำมาใช้ประกอบการวิเคราะห์ วิทยาการข้อมูล เป็นศาสตร์แขนงหนึ่งที่แสดงถึงแนวทางการใช้ประโยชน์จากเครื่องมือหรือวิธีการทางวิทยาศาสตร์ ในการสกัดสารสนเทศที่เป็นประโยชน์จากข้อมูลปริมาณมาก เพื่อให้เรามีความเข้าใจและเห็นถึงสิ่งที่ซ่อนอยู่ในข้อมูล รวมถึงการใช้ประโยชน์ในรูปแบบต่างๆ ในบทนี้จะแนะนำเกี่ยวกับวิทยาข้อมูล เพื่อให้ทราบถึงความหมาย สาขาวิชาอื่นๆ ที่เกี่ยวข้อง ระดับการวิเคราะห์และใช้ประโยชน์จากข้อมูล ตัวอย่างการใช้ประโยชน์จากข้อมูล และขั้นตอนการวิเคราะห์ข้อมูลโดยใช้เครื่องมือทางวิทยาการข้อมูล

1.1 ความหมายของวิทยาการข้อมูล

คำว่า **วิทยาการข้อมูล** มาจากคำในภาษาอังกฤษคำว่า *Data Science* ซึ่งเป็นการประสมคำระหว่างคำว่า *Data* หรือ **ข้อมูล** และคำว่า *Science* หรือ **วิทยาศาสตร์** ดังนั้น ก่อนจะกล่าวถึงความหมายของคำว่า **วิทยาการข้อมูล (Data Science)** จำเป็นที่จะต้องทราบความหมายของแม่คำทั้ง 2 คำนี้ก่อน

ข้อมูล (Data) คือ ข้อเท็จจริงเกี่ยวกับสิ่งต่างๆ เช่น คน สัตว์ พืช สิ่งของ สถานที่ และเหตุการณ์ เป็นต้น อาจเป็นค่าเชิงปริมาณหรือเชิงคุณภาพ ที่อยู่ในรูปของค่าตัวเลข ข้อความ รูปภาพ เสียง หรืออื่นๆ ข้อมูลที่ถูกเก็บรวบรวมและยังไม่ผ่านการประมวลผล เรียกว่า **ข้อมูลดิบ (Raw Data)** ส่วนข้อมูลที่ผ่านการประมวลผลหรือมีการจัดระเบียบแล้ว เรียกว่า **สารสนเทศ (Information)** ซึ่งสามารถนำไปใช้ประโยชน์ได้ต่อไป

วิทยาศาสตร์ (Science) เป็นกระบวนการค้นหาและจัดการความรู้อย่างเป็นระบบ มีระเบียบแบบแผน เพื่อให้ได้มาซึ่งองค์ความรู้ที่สามารถตรวจสอบได้

หากพิจารณาตามแม่คำแล้ว **วิทยาการข้อมูล (Data Science)** จึงหมายถึง การใช้กระบวนการทางวิทยาศาสตร์ที่มีระเบียบแบบแผน และสามารถตรวจสอบได้ ในการ สกัด องค์ความรู้ และ ความเข้าใจอย่างลึกซึ้ง จากข้อมูลที่มีปริมาณมาก เพื่อนำไปใช้ประโยชน์ต่อไป ราชบัณฑิตสภาได้ให้ความหมายของคำว่า **วิทยาการข้อมูล (Data Science)** ในพจนานุกรมศัพท์คอมพิวเตอร์และเทคโนโลยีสารสนเทศ ไว้ว่า “สหวิทยาการ ซึ่ง ประกอบด้วย วิทยาการคอมพิวเตอร์ คณิตศาสตร์ สถิติ และสาขาที่เกี่ยวข้อง ซึ่งใช้ในการวิเคราะห์ และทำความเข้าใจข้อมูลเพื่อนำไปใช้ประโยชน์ต่อไป”

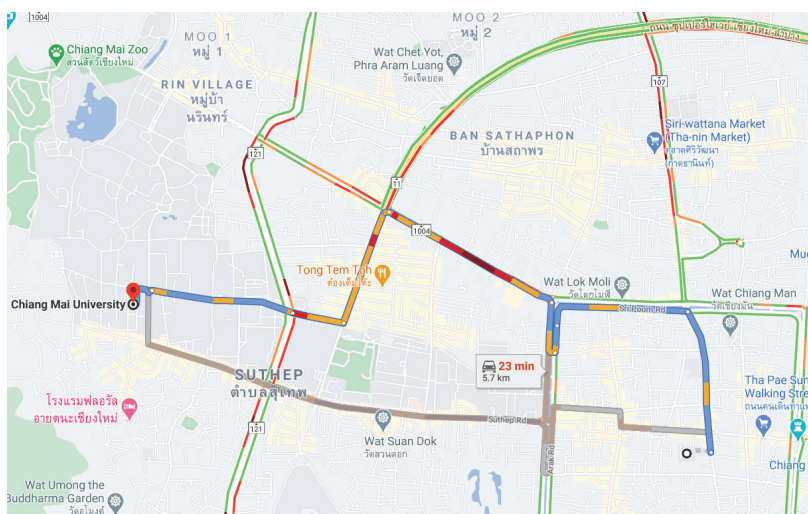
ความรู้ และ ความเข้าใจที่สกัดได้จากข้อมูลที่มีปริมาณมากนี้ จะเป็นประโยชน์อย่างยิ่งต่อการอธิบายปรากฏการณ์ต่างๆ ที่ซ่อนอยู่ในข้อมูล การทำนายหรือคาดการณ์สิ่งที่จะเกิดขึ้นเมื่อทราบข้อมูลบางส่วนหรือข้อมูลอดีต รวมไปถึง การใช้ประโยชน์เพื่อ ประกอบ การตัดสินใจ หรือ การวางแผนธุรกิจต่างๆ

1.2 ตัวอย่างการใช้ประโยชน์จากข้อมูล

Google Map

Google Map¹ เป็นบริการแผนที่ของกูเกิล (Google) ที่ผู้ใช้งานใช้ถนน รู้จักเป็นอย่างดี ความสามารถหนึ่งที่ของ Google Map คือการตรวจจับสภาพการจราจรบนท้องถนนที่จะแสดงแถบสีเขียว เหลือง หรือแดง บนถนนในแผนที่ เพื่อบ่งบอกสภาพการจราจร และการแนะนำเส้นทางที่คาดว่าจะใช้ระยะเวลาในการเดินทางจากต้นทางไปยังปลายทางน้อยที่สุด ซึ่งแสดงตัวอย่างดังภาพ 1.1 ความสามารถนี้เป็นตัวอย่างหนึ่งของการใช้ประโยชน์จากข้อมูลในการสร้างระบบผู้ช่วยแนะนำ (Recommendation System) โดยนำข้อมูล 2 ประเภท

1: Google Map
www.google.com/maps



ภาพ 1.1: แผนที่จาก Google Map แสดงสภาพการจราจรบนท้องถนนและเส้นทางแนะนำการเดินทางจากวัดเจดีย์หลวง จังหวัดเชียงใหม่ไปยังมหาวิทยาลัยเชียงใหม่ (ณ วันที่ 23 พฤศจิกายน 2563 เวลา 12:45น.)

มาใช้ประโยชน์ คือ 1) ข้อมูลอดีตเกี่ยวกับเวลาโดยเฉลี่ยที่ยานพาหนะผ่านส่วนหนึ่งของถนนในช่วงวันและเวลาต่างๆ และ 2) ข้อมูล แบบเรียลไทม์ ที่ส่งมาจากสมาร์ตโฟนผ่านแอปพลิเคชันต่างๆ หรืออุปกรณ์จีพีเอสซึ่งสามารถบอกความเร็วของรถยนต์ได้ เมื่อมีการใช้สมาร์ตโฟนโดยเฉพาะที่ติดตั้งระบบปฏิบัติการแอนดรอยด์ (Android) และมีการอนุญาตให้มีการเข้าถึงพิกัดจีพีเอสของเครื่องได้ สมาร์ตโฟนเครื่องนั้นจะมีการส่งข้อมูลแบบปกปิดตัวตนไปยังกูเกิล ทำให้กูเกิล ทราบความเร็วของรถยนต์ที่กำลังเคลื่อนที่ ข้อมูลที่ส่งออกมาจากสมาร์ตโฟนจำนวนมากมายจะถูกนำมาประมวลผลและนำเสนอบนแผนที่ด้วยแถบสีเพื่อบ่งบอกสภาพการจราจรบนท้องถนน ยิ่งไปกว่านั้น กูเกิลยังใช้ข้อมูลจากแอปพลิเคชัน Waze ที่ผู้ใช้แจ้งเตือนเหตุการณ์บนท้องถนนในการติดป้ายสัญลักษณ์แจ้งเตือนในแผนที่ด้วย [1]

ไอศกรีมรสใหม่จากข้อมูลสังคมออนไลน์

สังคมออนไลน์ (Social Media) นับว่าเป็นแหล่งข้อมูลขนาดใหญ่ ที่มีความหลากหลาย และมีพลวัตสูง ตัวอย่างการใช้ประโยชน์จากข้อมูลสังคมออนไลน์ ตัวอย่างหนึ่ง คือ การนำเอาความคิดเห็นของผู้ใช้งานสังคมออนไลน์ อย่างเช่น ทวิตเตอร์ มาเป็นใช้ในการนำเสนอไอศกรีมรสใหม่ของบริษัท Ben & Jerry's โดยทีมการตลาดของบริษัทพบว่าในช่วงที่มีพายุหิมะกลับมียอดขายไอศกรีมเพิ่มสูงขึ้น ซึ่งขัดแย้งกับความเชื่อกติที่ไอศกรีมควรขายดีในฤดูร้อน ทีมการตลาดจึงทำการหาข้อมูล เพื่อหาเหตุผลเบื้องหลังของปรากฏการณ์นี้ โดยทำการสำรวจจากข้อมูลในสังคมออนไลน์ จนพบว่าในช่วงพายุกำลังโหมกระหน่ำกลับมีผู้คนกลุ่มหนึ่งมีความสุขกับการทานไอศกรีมพร้อมๆ กับดูภาพยนตร์และรายการทีวีจาก Netflix เนื่องจากไม่สามารถออกจากที่พักได้ จากการค้นพบนี้จึงเป็นที่มาของกลยุทธ์การตลาดใหม่ โดยเมื่อมีช่วงเวลาที่ถูกคาดการณ์ว่าจะมีพายุฝนหรือพายุหิมะ บริษัทจะทำการโปรโมตให้คนออกไปหาซื้อไอศกรีมมาอุ่นเก็บไว้ และจากการที่ผู้คนมีความสุขกับการดูภาพยนตร์จาก Netflix พร้อมๆ กับรับประทานไอศกรีม จึงเป็นที่มาของไอศกรีมรสใหม่ นั่นคือ Netflix & Chill'd ของ Ben & Jerry [2]

NVIDIA Clara

NVIDIA Clara² เป็นอีกตัวอย่างหนึ่งของการใช้ประโยชน์จากข้อมูล โดยการนำข้อมูลทางการแพทย์ ที่รวบรวมไว้มาใช้ในการสร้างเทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence: AI) สำหรับเป็นผู้ช่วยทางการแพทย์ NVIDIA Clara เป็นโครงร่างซอฟต์แวร์ประยุกต์ด้านการแพทย์ สำหรับสร้างปัญญาประดิษฐ์สำหรับภาพถ่ายทางการแพทย์ (Medical Imaging) และการศึกษาจีโนม (Genomics) รวมไปถึงการพัฒนาแอปพลิเคชันและการนำไปใช้งานจริงในระบบโรงพยาบาล ในการพัฒนาปัญญาประดิษฐ์ในแนวทางของการเรียนรู้

[1] Bath Brindle (Feb. 2020). *How Does Google Maps Predict Traffic?* <https://electronics.howstuffworks.com/how-does-google-maps-predict-traffic.htm>. Accessed on 3 August 2020



ภาพ 1.2: ไอศกรีม รส Netflix & Chill'd ของบริษัท Ben & Jerry (ที่มา: www.benjerry.com/flavors/netflix-original-flavors)

[2] ณัฐพล ม่วงท่า (2563). *DATA-DRIVEN MARKETING การตลาดแบบฉลาดใช้ดาต้า*. กรุงเทพฯ: อมรินทร์ฮาวทู อมรินทร์ พรินต์ติ้ง แอนด์ พับลิชชิ่ง

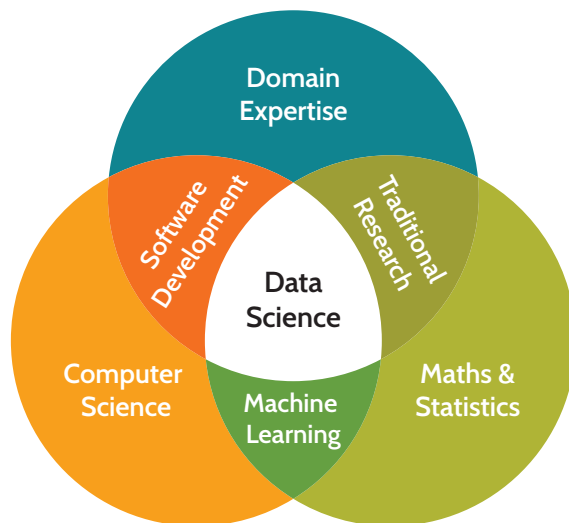
2: NVIDIA Clara

<http://developer.nvidia.com/clara>

ของเครื่อง (Machine Learning) ข้อมูลนับเป็นสิ่งสำคัญสำหรับการเรียนรู้ของตัวแบบปัญญาประดิษฐ์ (AI Model) เช่นเดียวกับการเรียนรู้ของมนุษย์ หากตัวแบบปัญญาประดิษฐ์ถูกสอนด้วยข้อมูลตัวอย่างที่มีปริมาณมาก มีความหลากหลาย และสามารถเป็นตัวแทนของข้อมูลทั้งหมดที่เป็นไปได้ ตัวแบบนั้นก็มีแนวโน้มที่จะมีความฉลาดและประสิทธิภาพสูง ทั้งนี้ขึ้นอยู่กับความซับซ้อนของตัวแบบด้วย นอกจากนี้เครื่องมือสำหรับการสร้างปัญญาประดิษฐ์แล้ว NVIDIA Clara ยังได้สร้างตัวแบบปัญญาประดิษฐ์ต้นแบบจากข้อมูลจำนวนมาก และสามารถปรับใช้ตัวแบบปัญญาประดิษฐ์ดังกล่าวโดยการสอนเพิ่มเติมจากข้อมูลที่มีความจำเพาะยิ่งขึ้นได้ ทำให้ลดระยะเวลา และค่าใช้จ่ายในการสร้างตัวแบบปัญญาประดิษฐ์ขึ้นใหม่ หลายองค์กรทางการแพทย์ทั้งในต่างประเทศ และในประเทศไทย เริ่มมีการนำเทคโนโลยีนี้มาพัฒนาต่อยอดและใช้งานบ้างแล้ว โดยเฉพาะทางด้านการวิจัย

1.3 สาขาวิชาที่เกี่ยวข้องกับวิทยาการข้อมูล

วิทยาการข้อมูล เป็นสหสาขาวิชาที่ผสมองค์ความรู้ทางด้าน คณิตศาสตร์และสถิติศาสตร์ (Mathematics and Statistics) วิทยาการคอมพิวเตอร์ (Computer Science) และความรู้เฉพาะด้านของสาขาที่เกี่ยวข้อง (Domain Expertise) เข้าไว้ด้วยกัน ซึ่งสามารถอธิบายความสัมพันธ์ระหว่างองค์ความรู้ทั้ง 3 ด้านนี้ได้ด้วยแผนภาพในภาพ 1.3



ภาพ 1.3: แผนภาพเวนนแสดงองค์ความรู้พื้นฐานของวิทยาการข้อมูล

ในการวิเคราะห์ข้อมูลที่มีจำนวนมหาศาลนั้น เราไม่สามารถใช้เพียงแรงงานคนในการทำงานได้ คอมพิวเตอร์จึงถูกนำมาใช้เป็นเครื่องมือเพื่อช่วยอำนวยความสะดวก ดังนั้น ความรู้ทางด้าน *วิทยาการคอมพิวเตอร์ (Computer Science)* จึงเข้ามามีบทบาทต่อสาขาวิทยาการข้อมูล หลายองค์กรได้นำระบบคอมพิวเตอร์มาช่วยในการรวบรวมข้อมูล ทั้งจากแหล่งข้อมูลปฐมภูมิ

และทฤษฎี มี การจัดเก็บข้อมูลต่างๆ ที่รวบรวมได้ในรูปแบบแฟ้มข้อมูล อิเล็กทรอนิกส์ (Electronic File) หรือฐานข้อมูล (Database) เมื่อข้อมูลถูกจัดเก็บลงในอุปกรณ์จัดเก็บข้อมูล การนำข้อมูลเหล่านั้นออกมาวิเคราะห์จึงอาศัยความรู้ทางวิทยาการคอมพิวเตอร์ในการพัฒนาเครื่องมือสำหรับเข้าถึง อ่าน และเขียนข้อมูล อีกทั้งปัจจุบันมีการพัฒนาซอฟต์แวร์ (Software) และภาษาโปรแกรม (Programming Language) จำนวนมากที่ประกอบด้วยเครื่องมือที่หลากหลายสำหรับอำนวยความสะดวกในการวิเคราะห์ข้อมูลที่มีพื้นฐานบนการคำนวณที่ซับซ้อน ยิ่งไปกว่านั้นความรู้ทางด้านวิทยาการคอมพิวเตอร์ยังถูกนำมาใช้ในการสร้างแผนภาพเพื่อนำเสนอข้อมูลหรือสารสนเทศที่ได้จากการวิเคราะห์ข้อมูล ซึ่งทำให้ง่ายต่อการสื่อสารกับบุคคลทั่วไป

แม้ว่าเครื่องมือในการวิเคราะห์ข้อมูลที่ถูกพัฒนาขึ้นจำนวนมากอยู่ในรูปของซอฟต์แวร์และภาษาโปรแกรม คอมพิวเตอร์ แต่เครื่องมือเหล่านั้นล้วนมีพื้นฐานบนความรู้ทางด้าน **คณิตศาสตร์และสถิติศาสตร์ (Mathematics and Statistics)** ทั้งสิ้น การวิเคราะห์ข้อมูลด้วยศาสตร์ทางด้านวิทยาการข้อมูลมีแนวคิดพื้นฐานอยู่บนปัญหาทางการคำนวณ ปัญหาธุรกิจ (Business Problem) จะต้องถูกแปลงให้อยู่ในรูปของปัญหาที่สามารถคำนวณได้ (Computable Problem) แล้วจึงนำเครื่องมือหรือวิธีการทางด้านคณิตศาสตร์ และสถิติมาใช้ในการแก้ปัญหาเหล่านั้น วิธีการคณิตศาสตร์ และสถิติ ถูกนำมาใช้เป็นเครื่องมือในแทบทุกขั้นตอนของการวิเคราะห์ข้อมูล ไม่ว่าจะเป็นการเตรียมและทำความสะอาดข้อมูล การอธิบายหรือพรรณาข้อมูล การวิเคราะห์ความสัมพันธ์ในข้อมูล การสร้างตัวแบบ (Model) เพื่อการทำนายปรากฏการณ์จากข้อมูล จนกระทั่ง การนำผลการวิเคราะห์ข้อมูลไปใช้ประกอบการวางแผนและตัดสินใจ ความรู้ความเข้าใจในคณิตศาสตร์ และสถิติที่เป็นพื้นฐานของวิธีการวิเคราะห์ข้อมูล นับว่ามีความสำคัญอย่างยิ่งที่จะทำให้การวิเคราะห์ข้อมูล สามารถเลือกใช้เครื่องมือได้อย่างเหมาะสม

องค์ความรู้ที่เป็นส่วนสำคัญที่สุดในการวิเคราะห์ข้อมูลด้วยศาสตร์ทางด้านวิทยาการข้อมูลนั้น คือ **ความรู้เฉพาะด้านของสาขาที่เกี่ยวข้อง (Domain Expertise)** ซึ่งเป็นความรู้เฉพาะด้านในสาขาวิชาที่เป็นที่มาของปัญหาทางธุรกิจ ความรู้เฉพาะด้านในปัญหาทางธุรกิจเป็นปัจจัยที่สำคัญต่อความสำเร็จของการดำเนินการวิเคราะห์ข้อมูลโดยตรง เริ่มต้นด้วยการวิเคราะห์ปัญหาจะต้องอาศัยผู้ที่มีความรู้ความเชี่ยวชาญในปัญหาทางธุรกิจช่วยใน การตั้งสมมติฐานและการกำหนดวัตถุประสงค์ของการวิเคราะห์ข้อมูล การวางแผนการเก็บรวบรวมข้อมูล ผู้เชี่ยวชาญเฉพาะด้านจะสามารถระบุได้ว่าข้อมูลใดมีความสำคัญหรือเกี่ยวข้องที่จะสามารถนำมาวิเคราะห์เพื่อตอบปัญหาทางธุรกิจได้ รวมไปถึงการเลือกใช้ข้อมูลหรือวิธีการวิเคราะห์ข้อมูลที่สอดคล้องกับความรู้เฉพาะด้าน จะทำให้ผลลัพธ์ที่ได้มีความน่าเชื่อถือและเป็นที่ยอมรับจากเจ้าของปัญหา นอกจากนี้ การนำผลการวิเคราะห์มาอภิปรายและสรุปผลก็ต้องอาศัยความรู้เฉพาะทางมา

อภิธานศัพท์

ตัวแบบ หรือ แบบจำลอง (Model)

หมายถึง สิ่งที่มนุษย์สร้างขึ้นเพื่อใช้แทนของจริง

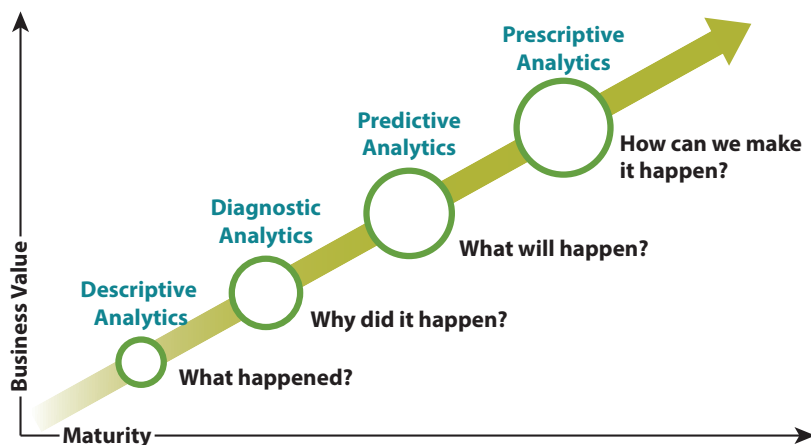
ตัวแบบเชิงคณิตศาสตร์และสถิติ

หมายถึง ตัวแบบที่ใช้ภาษาคณิตศาสตร์ในรูปแบบของสมการ อธิบายพฤติกรรมและความสัมพันธ์ระหว่างองค์ประกอบภายในระบบ

อธิบายและแสดงเหตุผลให้เป็นยอมรับ การวิเคราะห์ข้อมูลโดยปราศจากความ
รู้ความเข้าใจในบริบทของปัญหาทางธุรกิจนั้นอาจนำไปสู่ผลการวิเคราะห์ที่ไม่
สอดคล้องกับความเป็นจริง ย่อมทำให้ไม่เป็นที่ยอมรับและถูกนำไปใช้งานจริง

1.4 ระดับของการวิเคราะห์ข้อมูล

การ วิเคราะห์ ข้อมูล สามารถ แบ่ง ลำดับ ชั้น วุฒิ ภาวะ (Maturity) สำหรับ ใช้
ประเมินความก้าวหน้า ได้เป็น 4 ระดับ (แสดงดังภาพ 1.4) ดังนี้



ภาพ 1.4: ลำดับ ชั้น วุฒิ ภาวะ ของ การ
วิเคราะห์ข้อมูล

การวิเคราะห์เชิงพรรณนา (Descriptive Analytics)

ลำดับขั้นแรกของการวิเคราะห์ข้อมูล คือ การอธิบายหรือพรรณาข้อมูล
โดยปกติแล้วจะวิเคราะห์โดยการแจกแจงแต่ละ **ตัวแปร (Variable)** เครื่องมือ
ที่ช่วยในการอธิบายข้อมูลนี้ส่วนใหญ่เป็นการคำนวณค่าสถิติเชิงพรรณนา (De-
scriptive Statistics) เช่น ค่ากลางของข้อมูล ค่าความแปรปรวน และช่วงของ
ค่าข้อมูล เป็นต้น ปัญหาการวิเคราะห์ข้อมูลในระดับนี้มักเป็นเพียงคำถามที่
ต้องการทราบถึงสิ่งที่เกิดขึ้นจากข้อมูลเท่านั้น ตัวอย่างเช่น

- ▶ นักศึกษาที่ลงทะเบียนเรียนกระบวนวิชา 204123 ส่วนใหญ่มาจากคณะ
อะไร?
- ▶ ผู้เสียชีวิตจากเชื้อโรค COVID-19 มีอายุเฉลี่ยเท่าไร?

อภิธานศัพท์

ตัวแปร (Variable)

หมายถึง คุณลักษณะหรือคุณสมบัติหรือ
ปรากฏการณ์ของสิ่งต่างๆ ที่ผู้ศึกษาสนใจ
ตัวอย่าง แอปเปิลผลหนึ่งมีสีแดง รสชาติ
หวาน และมีน้ำหนัก 350 กรัม ชนิดผลไม้
สี รสชาติ และน้ำหนัก คือตัวแปรที่สนใจ
ส่วนแอปเปิลสีแดง รสหวาน และน้ำหนัก
350 กรัม เป็นค่าของตัวแปร

การวิเคราะห์เชิงวินิจฉัย (Diagnostics Analytics)

การวิเคราะห์ข้อมูลในลำดับขั้นที่สูงขึ้น คือ การวินิจฉัยความเป็นเหตุเป็นผล
หรือความสัมพันธ์กันระหว่างตัวแปร มีการนำเครื่องมือในการวิเคราะห์สหสัมพันธ์
(Correlation Analysis) หรือการวิเคราะห์ความสัมพันธ์ (Association Analy-
sis) มาใช้เพื่อหาเหตุผล ของสิ่งที่เกิดขึ้นจากข้อมูล ตัวอย่างเช่น

- ▶ เหตุใดนักศึกษาที่ลงทะเบียนเรียนกระบวนวิชา 204123 ส่วนมากมาจาก คณะวิศวกรรมศาสตร์?
- ▶ เหตุใดผู้เสียชีวิตจากเชื้อโรค COVID-19 มีอายุ 80 ปีขึ้นไป?
- ▶ เหตุใดผู้เข้ารับบริการที่สำนักทะเบียนและประมวลผล มีจำนวนมากที่สุด ช่วงเปิดภาคการศึกษา?

การวิเคราะห์เชิงพยากรณ์ (Predictive Analytics)

การวิเคราะห์ข้อมูลในลำดับขั้นนี้มุ่งเน้นการทำนาย หรือพยากรณ์ค่าของ ตัวแปรที่ไม่ทราบค่า หรือค่าข้อมูลที่จะเกิดขึ้นในอนาคต เป็นการสร้างตัวแบบ ทางคณิตศาสตร์ หรือ สถิติ เพื่ออธิบายความสัมพันธ์ระหว่าง ตัวแปรที่ทราบ ค่า และตัวแปรที่ไม่ทราบค่า วิธีการที่นำมาใช้ในการวิเคราะห์เชิงพยากรณ์ สามารถแบ่งได้ 2 แนวทาง คือ 1) การทดลองเพื่อทดสอบสมมติฐาน โดยมีการตั้งสมมติฐานจากการสังเกต และทำการทดสอบ สมมติฐานนั้น โดยใช้วิธีการทางสถิติ และ 2) การใช้วิธีการเรียนรู้ของเครื่อง (Machine Learning) ในการสร้างตัวแบบสำหรับการทำนายค่าข้อมูล โดยใช้ชุดข้อมูลในอดีตที่ถูกเก็บ รวบรวมไว้เป็นข้อมูลสำหรับการเรียนรู้ เพื่อหาความสัมพันธ์ระหว่างตัวแปร [3] การวิเคราะห์ข้อมูลในลำดับขั้นนี้มักใช้เพื่อตอบคำถามทางธุรกิจที่ต้องการทราบ ถึงสิ่งที่จะเกิดขึ้น ตัวอย่างเช่น

- ▶ ราคาน้ำมันในสัปดาห์ถัดไปจะเพิ่มขึ้นหรือลดลงเท่าไร?
- ▶ พืชที่อยู่ในภาพถ่ายจัดอยู่ในสปีชีส์ใด?
- ▶ ร้านหนังสือจะสามารถขายคู่มือเตรียมสอบได้จำนวนกี่เล่ม เมื่อมีจำนวน นักเรียนที่สมัครสอบโควต้ามหาวิทยาลัยเชียงใหม่ 5,000 คน?

[3] Peter Prevos (2019). *Principles of Strategic Data Science: Creating Value from Data, Big and Small*. Birmingham, United Kingdom: Packt Publishing

การวิเคราะห์เชิงวางแผน (Prescriptive Analytics)

ลำดับขั้นของการวิเคราะห์ข้อมูลขั้นสูงสุดคือการวิเคราะห์เชิงวางแผน ซึ่งเป็นการวิเคราะห์ข้อมูล สำหรับช่วยในการวางแผน หรือการตัดสินใจ โดย คาดหวังผลลัพธ์ที่ดีที่สุดภายใต้เงื่อนไขที่กำหนด การวิเคราะห์ข้อมูลลำดับขั้นนี้ ปัญหาทางธุรกิจมักถูกแปลงให้อยู่ในรูปปัญหาการหาค่าที่เหมาะสมที่สุด (Optimization Problem) ก่อนจะประยุกต์ใช้เครื่องมือวิเคราะห์สำหรับการหาค่าเหมาะสม ที่สุดช่วยในการแก้ปัญหาอย่างอัตโนมัติ เครื่องมือวิเคราะห์ที่มีถูกนำมาใช้ช่วย ในการตัดสินใจอีกเครื่องมือหนึ่งคือการจำลองแบบ (Simulation) เนื่องจากการ ดำเนินการทดลองในบางสถานการณ์หรือบางสภาพแวดล้อมไม่สามารถทำได้ใน โลกของความเป็นจริง ดังนั้น จึงต้องจำลองสภาพแวดล้อมนั้นด้วยคอมพิวเตอร์ และสังเกตผลลัพธ์ที่ได้ก่อนนำไปสรุปเป็นแนวทางในการตัดสินใจ ตัวอย่าง ปัญหาทางธุรกิจที่ต้องอาศัยการวิเคราะห์เชิงวางแผน เช่น

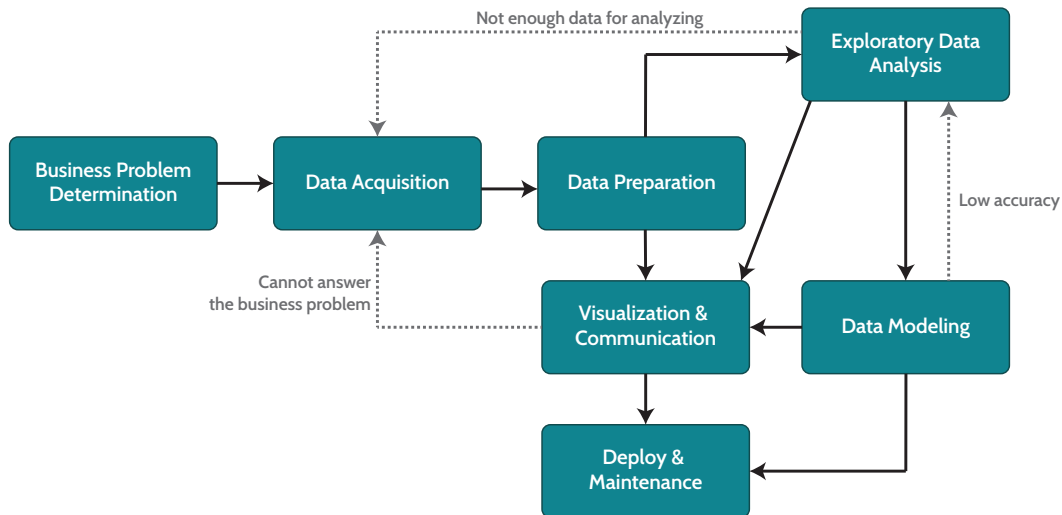
- ▶ เจ้าของฟาร์มเปิดควรสร้างโรงเรือนอย่างไร เพื่อให้สามารถรองรับจำนวนสัตว์เลี้ยงได้มากที่สุด?
- ▶ โรงงานผลิตแก้วควรผลิตสินค้าที่แตกต่างกัน 3 ชนิดจำนวนเท่าใด เพื่อให้ได้กำไรจากการขายสูงสุด?

1.5 ขั้นตอนทางวิทยาการข้อมูล

กระบวนการวิเคราะห์ข้อมูลในแนวทางของวิทยาการข้อมูล สามารถแบ่งออกได้ 7 ขั้นตอน แสดงลำดับการดำเนินงานได้ดังภาพ 1.5 โดยลำดับการดำเนินงานไปข้างหน้าแสดงด้วยลูกศรเส้นทึบ เริ่มต้นด้วยการกำหนดปัญหาธุรกิจ (Business Problem) จนกระทั่งการนำผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลไปใช้ประโยชน์ ในการดำเนินการวิเคราะห์ข้อมูลนั้นไม่จำเป็นต้องดำเนินงานให้ครบทุกขั้นตอน ทั้งนี้ขึ้นอยู่กับเป้าหมาย วัตถุประสงค์ และขอบเขตของปัญหาธุรกิจ อีกทั้งหากผลลัพธ์ที่ได้จากขั้นตอนใดขั้นตอนหนึ่งไม่เป็นไปตามความคาดหวังยังสามารถย้อนกลับไปแก้ไขในขั้นตอนก่อนหน้าได้ (แสดงการดำเนินการย้อนกลับด้วยลูกศรเส้นปะ พร้อมระบุปัญหาที่เกิดขึ้น) รายละเอียดการดำเนินงานในแต่ละขั้นตอนอธิบายได้ ดังนี้

ขั้นตอนที่ 1 การกำหนดปัญหาทางธุรกิจ (Business Problem Determination)

การดำเนินการวิเคราะห์ข้อมูลเริ่มต้นด้วยการทำความเข้าใจและวิเคราะห์ปัญหาธุรกิจ เพื่อให้ได้มาซึ่งวัตถุประสงค์หรือเป้าหมายของปัญหาที่ชัดเจน พร้อมทั้ง ระบุขอบเขตของปัญหา มักจะเป็นการทำงานร่วมกันของเจ้าของปัญหาธุรกิจและนักวิทยาการข้อมูล ขั้นตอนแรกนี้มีความสำคัญอย่างยิ่งที่นำไปสู่ความสำเร็จของการวิเคราะห์ข้อมูล หากนักวิทยาการข้อมูลไม่สามารถทำความเข้าใจหรือเข้าใจไม่ถูกต้องเกี่ยวกับปัญหาธุรกิจ ย่อมทำให้การกำหนดเป้าหมายของการวิเคราะห์ข้อมูลผิดพลาด และได้ผลลัพธ์ไม่ตรงตามความต้องการของเจ้าของปัญหาธุรกิจ ความรู้เฉพาะด้านของปัญหาธุรกิจมีส่วนสำคัญอย่างยิ่งที่จะช่วยให้การทำความเข้าใจปัญหาทำได้ง่ายยิ่งขึ้น ดังนั้น เมื่อได้รับปัญหาธุรกิจ จึงควรต้องศึกษาความรู้เฉพาะด้านที่เกี่ยวข้องกับปัญหานั้นก่อน ซึ่งจะทำให้การสื่อสารระหว่างเจ้าของปัญหาและนักวิทยาการข้อมูลเป็นไปได้อย่างราบรื่น และมีความเข้าใจที่ตรงกัน การดำเนินงานในขั้นตอนนี้อาจจะใช้ระยะเวลามาก อย่างไรก็ตามหากสามารถกำหนดวัตถุประสงค์ เป้าหมาย และขอบเขตของปัญหาธุรกิจได้อย่างถูกต้อง และชัดเจนแล้ว โอกาสในการวิเคราะห์ข้อมูลให้ประสบความสำเร็จและถูกนำไปใช้ได้จริงย่อมมีสูงตามไปด้วย



ภาพ 1.5: กระบวนการวิเคราะห์ข้อมูลในแนวทางของวิทยาการข้อมูล ประกอบด้วย 7 ขั้นตอน (ลูกศรเส้นทึบแสดงลำดับการดำเนินการไปข้างหน้า และ ลูกศรเส้นปะแสดงการย้อนกลับการดำเนินงาน)

ขั้นตอนที่ 2 การได้ข้อมูล (Data Acquisition)

ขั้นตอนต่อมา คือ การได้ข้อมูลสำหรับนำมาวิเคราะห์เพื่อแก้ปัญหาธุรกิจ ไม่ใช่เพียงแค่การรวบรวมข้อมูลมาได้เท่านั้น ในขั้นตอนนี้ยังรวมถึงการออกแบบ และวางแผนการรวบรวมและจัดเก็บข้อมูลด้วย การเก็บรวบรวมข้อมูลให้ได้ อย่างครบถ้วนและเพียงพอต่อการนำไปวิเคราะห์เพื่อตอบโจทย์ปัญหาธุรกิจ นั้นเป็นสิ่งสำคัญ การได้ข้อมูลมานั้นย่อมมีค่าใช้จ่ายและต้องใช้เวลาในการ เก็บรวบรวมข้อมูลเสมอ หากทำการเก็บรวบรวมข้อมูลมากเกินไปจนความจำเป็นต่อการนำไปวิเคราะห์น้อยลงหมายถึงการลงทุนที่มีค่าใช้จ่ายสูงเมื่อเปรียบเทียบกับ ผลลัพธ์ที่ได้ แต่การเก็บรวบรวมข้อมูลที่น้อยจนไม่สามารถนำไปวิเคราะห์ได้ ย่อมหมายถึงการลงทุนที่สูงเปล่า ดังนั้นในขั้นตอนการได้ข้อมูลนี้จึงต้องอาศัย การวางแผนและการบริหารจัดการที่ดี

ขั้นตอนที่ 3 การเตรียมข้อมูล (Data Preparation)

ข้อมูลที่ได้มาจากขั้นตอนที่ 2 จะถูกนำมาจัดเตรียมให้พร้อมสำหรับการนำไปวิเคราะห์ในขั้นตอนนี้ ข้อมูลที่รวบรวมได้จะต้องถูกพิจารณาถึงคุณภาพ ของข้อมูล ข้อมูลที่มีคุณภาพ ย่อมส่งผลให้การวิเคราะห์ข้อมูลทำได้ อย่างมีประสิทธิภาพ และได้ผลลัพธ์ที่คาดหวังได้ว่าจะถูกต้องตามความเป็นจริง ใน ขั้นตอนนี้ข้อมูลจะถูกตรวจสอบและปรับปรุงคุณภาพของข้อมูลด้วยวิธีการ ต่างๆ จากนั้นข้อมูลจะถูกแปลงให้อยู่ในรูปแบบที่เหมาะสมสำหรับการวิเคราะห์ ข้อมูลจะถูกเปลี่ยนแปลงรูปแบบหรือโครงสร้างการจัดเก็บ รวมไปถึงการปรับ ค่าข้อมูลให้เป็นมาตรฐาน (Standardization) และการทำให้เป็นบรรทัดฐาน (Normalization) ในบางครั้งอาจมีการสกัดลักษณะเด่น (Feature Extraction) หรือการสร้างลักษณะเด่น (Feature Generation) จากข้อมูลเดิม ซึ่งทำให้ง่าย ต่อการทำความเข้าใจลักษณะและโครงสร้างของข้อมูล และนำไปสู่การแนวทาง

หรือวิธีการวิเคราะห์ข้อมูลที่เหมาะสมต่อไป

ขั้นตอนที่ 4 การวิเคราะห์ข้อมูลด้วยการสำรวจ (Exploratory Data Analysis)

เมื่อข้อมูลถูกตรวจสอบคุณภาพและถูกแปลงให้อยู่ในรูปแบบที่เหมาะสมแล้ว การวิเคราะห์ข้อมูลที่สมควรเริ่มต้นด้วยการสำรวจข้อมูล เพื่อทำความเข้าใจข้อมูลที่มีอยู่ในเบื้องต้น ในขั้นตอนนี้เป็นการวิเคราะห์ข้อมูลเชิงพรรณนาและการวิเคราะห์ข้อมูลเชิงวินิจฉัย ซึ่งเป็นการวิเคราะห์ข้อมูลขั้นเริ่มต้น ด้วยเครื่องมือทางสถิติเชิงพรรณนา การแสดงภาพข้อมูล (Data Visualization) และการวิเคราะห์ความสัมพันธ์ที่เกิดขึ้นระหว่างตัวแปรต่างๆ ในบางปัญหาระบุถึงอาจใช้เพียงการวิเคราะห์ข้อมูลในขั้นตอนนี้เท่านั้นและสามารถนำผลลัพธ์จากขั้นตอนนี้มาเสนอต่อเจ้าของปัญหาได้ อย่างไรก็ตามหากปัญหาระบุถึงต้องอาศัยการวิเคราะห์เชิงพยากรณ์หรือการวิเคราะห์เชิงวางแผนไข ในการแก้ปัญหา การวิเคราะห์ในขั้นตอนนี้จะช่วยให้สามารถเลือกตัวแปรหรือคุณลักษณะที่มีศักยภาพและเหมาะสมสำหรับใช้ในการพยากรณ์หรือหาค่าที่เหมาะสมได้ เรียกวิธีการนี้ว่า การเลือกลักษณะเด่น (Feature Selection)

ขั้นตอนที่ 5 การสร้างตัวแบบข้อมูล (Data Modeling)

ในการวิเคราะห์เชิงพยากรณ์หรือการวิเคราะห์เชิงวางแผนไข ข้อมูลที่ได้รวบรวมและจัดเตรียมไว้จะถูกใช้ในการสร้างตัวแบบทางคณิตศาสตร์หรือสถิติ ซึ่งจะถูกนำไปใช้ประโยชน์ในการทำนายหรือช่วยในการตัดสินใจเพื่อตอบปัญหาระบุถึง ในขั้นตอนนี้ข้อมูลจะถูกแบ่งเป็น 2 ชุด คือ 1) ชุดข้อมูลเรียนรู้ (Training Dataset) สำหรับใช้ในการเรียนรู้ของตัวแบบ และ 2) ชุดข้อมูลทดสอบ (Test Dataset) สำหรับใช้วัดประสิทธิภาพของตัวแบบ นักวิทยาการข้อมูลจะต้องเลือกใช้ตัวแบบที่เหมาะสมกับปัญหาระบุถึง ทำการสร้างตัวแบบด้วยชุดข้อมูลเรียนรู้ เมื่อได้ตัวแบบที่ดีแล้ว ตัวแบบจะต้องถูกทดสอบและประเมินประสิทธิภาพด้วยชุดข้อมูลทดสอบก่อนนำไปใช้งาน

ขั้นตอนที่ 6 การสร้างภาพข้อมูลและการนำเสนอผลการวิเคราะห์ (Visualization & Communication)

เมื่อทำการวิเคราะห์ข้อมูลแล้ว ผลลัพธ์จากการวิเคราะห์ข้อมูลจะถูกนำเสนอต่อเจ้าของปัญหาระบุถึง ข้อมูลและผลการวิเคราะห์ข้อมูล มักจะถูกนำเสนอโดยใช้ภาพเป็นสื่อกลาง เพื่อให้ง่ายต่อการสื่อสารและการทำความเข้าใจ เทคนิคการแสดงผลภาพข้อมูล จะถูกเลือกใช้อย่างเหมาะสม เพื่อสร้างการนำเสนอที่มีประสิทธิภาพ สามารถสื่อสารให้เกิดความเข้าใจอย่างถูกต้อง ตรงตามวัตถุประสงค์ของการนำเสนอข้อมูล

ขั้นตอนที่ 7 การนำไปใช้ประโยชน์และการบำรุงรักษา (Deploy & Maintenance)

ผลการวิเคราะห์ข้อมูลที่ได้รับการตรวจสอบและยอมรับจากเจ้าของธุรกิจแล้ว อาจจะถูกนำเสนอให้เป็นรับรู้และใช้งานจากบุคคลทั่วไปในรูปแบบรายงาน สื่อมัลติมีเดีย หรือแอปพลิเคชัน ผ่านทางช่องทางต่างๆ อีกทั้งยังต้องมีการบำรุงรักษาให้ข้อมูล หรือแอปพลิเคชันเหล่านั้น มีความถูกต้อง ทันสมัย และสามารถทำงานได้อย่างปกติอยู่เสมอ

ตัวอย่าง 1.1

ปัญหาธุรกิจ สวนสัตว์แห่งหนึ่งต้องการพัฒนาแอปพลิเคชันบนสมาร์ตโฟน สำหรับนักท่องเที่ยว ที่สามารถถ่ายรูปนกแต่ละชนิดในสวนสัตว์และระบุชนิดของนกในภาพถ่าย พร้อมคำอธิบายได้

จากปัญหาธุรกิจข้างต้น สามารถออกแบบ การดำเนินการวิเคราะห์ข้อมูลได้ 7 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 กำหนดปัญหาธุรกิจ เป้าหมายของปัญหาธุรกิจนี้ คือ แอปพลิเคชัน ที่สามารถรับภาพถ่ายนกและระบุชนิดของนกในภาพถ่าย จะเห็นได้ว่าเป็นการทำนายหรือพยากรณ์ชนิดของนกจากข้อมูลภาพถ่าย ดังนั้น ปัญหานี้จึงต้องใช้ *การวิเคราะห์เชิงพยากรณ์* ในการแก้ปัญหา อีกทั้งลักษณะของปัญหาเป็นปัญหาการรู้จำหรือปัญหาการจำแนก (Classification Problem) นั่นคือการจำแนกชนิดของนกจากภาพถ่ายนก

ขั้นตอนที่ 2 ออกแบบและรวบรวมข้อมูล จากการวิเคราะห์ปัญหาธุรกิจในขั้นตอนที่ 1 ปัญหาธุรกิจนี้เป็นปัญหาการจำแนก และต้องใช้การวิเคราะห์เชิงพยากรณ์ในการแก้ปัญหา ซึ่งจะต้องทำการสร้างตัวแบบสำหรับจำแนกข้อมูลรูปภาพ ในการสร้างตัวแบบสำหรับการจำแนกข้อมูล จะต้องเตรียมชุดข้อมูล 2 ชุด คือ ชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ ดังนั้น ในขั้นตอนการรวบรวมข้อมูลนี้จะต้องทำการเก็บภาพถ่ายนกแต่ละชนิด พร้อมป้ายระบุชนิดของนกในภาพ อาจได้มาจากการรวบรวมจากเว็บไซต์ต่างๆ หรือถ่ายภาพนกในสวนสัตว์

ขั้นตอนที่ 3 เตรียมข้อมูล เนื่องจากภาพถ่ายนกที่รวบรวมมาได้จากหลายแหล่งข้อมูล อาจมีคุณลักษณะของภาพแตกต่างกัน ในขั้นตอนนี้จะทำการเตรียมข้อมูล โดยปรับขนาดของภาพให้เท่ากัน ในกรณีที่รูปหนึ่งรูปมีนกหลายตัวจะทำการแยกภาพรูปนกให้ภาพหนึ่งภาพมีนกหนึ่งตัว อีกทั้งจะการจัดระเบียบไฟล์ข้อมูลโดยแยกภาพนกที่มีชนิดเดียวกันอยู่ในโฟลเดอร์ข้อมูลเดียวกัน ภาพนกต่างชนิดกันจัดแยกโฟลเดอร์ข้อมูลกัน และทำการแยกข้อมูลออกเป็น 2 ชุด คือชุดข้อมูลเรียนรู้ และชุดข้อมูลทดสอบ ต่อมา

จะต้องทำการสกัดลักษณะเด่น (Feature) จากภาพ สมมติว่า เราสามารถสกัดลักษณะเด่นได้ ดังนี้

- ▶ ขนาดตัว ปีก ขา และเท้า
- ▶ ลักษณะปาก หัว หาง ปีก และเท้า
- ▶ สีบริเวณหน้า ปีก หาง และหน้าท้อง
- ▶ จำนวนตา ขา ปาก และนิ้วเท้า

ข้อมูลของนกแต่ละตัว ประกอบด้วย ลักษณะเด่นที่สกัดได้จากภาพ และป้ายระบุชนิด สามารถจัดเก็บในรูปแบบตารางข้อมูลได้

ขั้นตอนที่ 4 วิเคราะห์ข้อมูลด้วยการสำรวจ เมื่อข้อมูลถูกแปลงให้อยู่ในรูปแบบที่เหมาะสม โดยประกอบด้วยลักษณะเด่น และป้ายระบุชนิดแล้ว ต่อมาจะทำการวิเคราะห์ลักษณะการกระจายของข้อมูลลักษณะเด่นแต่ละตัวแปร และความสัมพันธ์ระหว่างแต่ละตัวแปรและป้ายระบุชนิด พบว่าลักษณะเด่นที่ระบุ จำนวนตา ขา ปาก และนิ้วเท้า ไม่สามารถใช้ในการจำแนกชนิดของนกได้ ดังนั้น จึงจะไม่นำลักษณะเด่นเหล่านี้ไปใช้ในการสร้างตัวแบบสำหรับจำแนกชนิดของนก

ขั้นตอนที่ 5 สร้างตัวแบบข้อมูล ในขั้นตอนนี้จะทำการสร้างตัวแบบสำหรับการจำแนกชนิดของนก โดยอาศัยชุดข้อมูลเรียนรู้ในการสร้างตัวแบบ และทำการทดสอบประสิทธิภาพในการรู้จำของตัวแบบด้วยชุดข้อมูลทดสอบ อาจทำการศึกษาตัวแบบที่หลากหลาย และเปรียบเทียบประสิทธิภาพในการรู้จำ เพื่อให้ได้ตัวแบบที่สามารถจำแนกชนิดของนกจากรูปภาพได้ดีที่สุด

ขั้นตอนที่ 6 นำเสนอผลการวิเคราะห์ข้อมูล เมื่อได้ตัวแบบสำหรับการรู้จำชนิดนกจากภาพถ่าย ที่มีประสิทธิภาพเพียงพอต่อการนำไปใช้งานแล้ว ต่อมาจะต้องนำเสนอผลการวิเคราะห์ให้แก่เจ้าของปัญหาธุรกิจ ตั้งแต่กระบวนการที่ใช้ในการวิเคราะห์ข้อมูล ความเกี่ยวข้องระหว่างวิธีการวิเคราะห์ข้อมูล และความรู้จากผู้เชี่ยวชาญเรื่องนก ผลการสร้างตัวแบบและประสิทธิภาพในการรู้จำภาพนก

ขั้นตอนที่ 7 นำไปใช้ประโยชน์ เมื่อเจ้าของปัญหาธุรกิจยอมรับผลการวิเคราะห์ข้อมูลและตัวแบบที่นำเสนอแล้ว ในขั้นตอนสุดท้ายนี้ ตัวแบบที่สร้างขึ้นจะถูกส่งมอบให้นักพัฒนาโปรแกรมนำไปพัฒนาเป็นแอปพลิเคชันบนสมาร์ตโฟน และเผยแพร่ต่อไป

1.6 เอกสารอ้างอิงและแหล่งศึกษาเพิ่มเติม

1. Brindle, Bath (2020). *How Does Google Maps Predict Traffic?* <https://electronics.howstuffworks.com/how-does-google-maps-predict-traffic.htm>. Accessed on 3 August 2020
2. ม่วงท่า, ณัฐพล (2563). *DATA-DRIVEN MARKETING การตลาดแบบฉลาดใช้ดาต้า*. กรุงเทพฯ: อมรินทร์ฮาวทู อมรินทร์พริ้นติ้ง แอนด์พับลิชชิ่ง
3. Prevos, Peter (2019). *Principles of Strategic Data Science: Creating Value from Data, Big and Small*. Birmingham, United Kingdom: Packt Publishing
4. Simplilearn (2018). *Data Science In 5 Minutes | Data Science For Beginners | What Is Data Science?*. <https://www.youtube.com/watch?v=X3paOmcTjQ&t=22s>. Accessed on 15 July 2019

1.7 แบบฝึกหัดท้ายบท

1. จงยกตัวอย่างงานประยุกต์ที่นำข้อมูลมาใช้ประโยชน์ในการแก้ปัญหา หรือพัฒนาแอปพลิเคชันที่นักศึกษาใช้ในชีวิตประจำวัน
2. ตัวอย่างการใช้ประโยชน์จากข้อมูล กรณีศึกษา Google Map จัดเป็นการวิเคราะห์ข้อมูลในลำดับขั้นวุฒิภาวะ (Data Analysis Maturity) ใด จงอธิบายเหตุผล
3. ขั้นตอนทางวิทยาการข้อมูลใดบ้าง ที่ต้องอาศัยความรู้เฉพาะด้านของสาขาที่เกี่ยวข้อง (Domain Expertise) กับปัญหาธุรกิจ ในการดำเนินงาน จงอธิบาย
4. **ปัญหาธุรกิจ** กรมอุตุนิยมวิทยามีข้อมูลอุณหภูมิ ความกดอากาศ ปริมาณ ความชื้น และกระแสลม ณ บริเวณต่างๆ ของประเทศไทย เจ้าหน้าที่กรมคนหนึ่งต้องการใช้อยู่มุลดังกล่าวในการพยากรณ์ ความเป็นไปได้ที่ฝนจะตกในบริเวณต่างๆ
จากปัญหาธุรกิจข้างต้น จงตอบคำถามต่อไปนี้
 - a) การวิเคราะห์ข้อมูลเพื่อแก้ปัญหาธุรกิจนี้จัดอยู่ในลำดับขั้นวุฒิภาวะของการวิเคราะห์ข้อมูลลำดับขั้นใด
 - b) ในการดำเนินงานเพื่อแก้ปัญหาธุรกิจ ควรดำเนินการขั้นตอนทางวิทยาการข้อมูลขั้นตอนใดบ้าง จงอธิบายลำดับขั้นตอนการวิเคราะห์ข้อมูล
5. **ปัญหาธุรกิจ** ผู้บริหารมหาวิทยาลัยแห่งหนึ่งต้องการวิเคราะห์ภาวะการมีงานทำ และสายอาชีพของนักศึกษาที่สำเร็จการศึกษาแล้วของคณะต่างๆ ในมหาวิทยาลัย
จากปัญหาธุรกิจข้างต้น จงตอบคำถามต่อไปนี้
 - a) การวิเคราะห์ข้อมูลเพื่อแก้ปัญหาธุรกิจนี้จัดอยู่ในลำดับขั้นวุฒิภาวะของการวิเคราะห์ข้อมูลลำดับขั้นใด
 - b) ในการดำเนินงานเพื่อแก้ปัญหาธุรกิจ ควรดำเนินการขั้นตอนทางวิทยาการข้อมูลขั้นตอนใดบ้าง จงอธิบายลำดับขั้นตอนการวิเคราะห์ข้อมูล