



Practical

DATA SCIENCE¹

with RapidMiner Studio

เล่ม



RAPIDMINER

No-Code/Low-Code Data Science Platforms

อธิบายการสร้าง Basic Machine Learning Models
ตั้งแต่ต้นจนจบ เหมาะสำหรับผู้เชี่ยวชาญสาขา
ทางธุรกิจที่อยากเป็น Citizen Data Scientist



WORKSHOP DATASET:
SERAZU.COM

ผู้แต่ง ดร.เอกสิทธิ์ พิชรวงศ์กัฏกา
บรรณาธิการ ทิรพล ภาษเจริญ

Practical Data Science with RapidMiner Studio เล่ม 1

Writer	ดร.เอกสิทธิ์ พัทธวงศ์ศักดิ์ดา
Editor	ภิรพล คชาเจริญ
Production Manager	วรพล ณธิกุล
Graphic Designer	ชวพันธ์ รัตนะ
Page Layout	สุรัสวดี วงศ์จันทร์สุข
Proofreader	สุนทรี บรรลือศักดิ์
Publishing Coordinators	สุพัตรา อาจปรู, สุรีย์รัตน์ จิ๋ว
Product Specialist	ศรัณย์ ชาทิสูทธิผล

RapidMiner Studio เป็นเครื่องหมายการค้าของบริษัท RapidMiner, Inc. และเครื่องหมายการค้าอื่นๆ ที่อ้างถึงเป็นของบริษัทนั้นๆ

สงวนลิขสิทธิ์ตามพระราชบัญญัติลิขสิทธิ์ พ.ศ. 2537 โดยบริษัท "ไอดีซี พรีเมียร์ จำกัด ห้ามลอกเลียนไม่ว่าส่วนใดส่วนหนึ่งของหนังสือเล่มนี้ ไม่ว่าในรูปแบบใดๆ นอกจากจะได้รับอนุญาตเป็นลายลักษณ์อักษรจากผู้จัดพิมพ์เท่านั้น

สร้างสรรค์โดย

จัดพิมพ์และจัดจำหน่ายโดย



ข้อมูลทางบรรณานุกรม
ดร.เอกสิทธิ์ พัทธวงศ์ศักดิ์ดา

Practical Data Science with RapidMiner Studio เล่ม 1
นนทบุรี : ไอดีซีฯ, 2565
264 หน้า

1. โปรแกรมสำหรับระบบการประมวลผลข้อมูลเฉพาะชนิด
I ชื่อเรื่อง

005.37

Barcode 885-916-101-074-6

พิมพ์ครั้งที่ 1 ตุลาคม 2565

ราคา 385 บาท

บริษัท ไอดีซี พรีเมียร์ จำกัด

200 หมู่ 4 ชั้น 19 ห้อง 1901

อาคารจัสตินอินเตอร์เนชั่นแนลทาวเวอร์

ถ.แจ้งวัฒนะ อ.ปากเกร็ด จ.นนทบุรี 11120

โทรศัพท์ 0-2962-1081 (อัตโนมัติ 10 คู่สาย)

โทรสาร 0-2962-1084

สมาชิกสัมพันธ์

โทรศัพท์ 0-2962-1081-3 ต่อ 121

โทรสาร 0-2962-1084

ร้านค้าและตัวแทนจำหน่าย

โทรศัพท์ 0-2962-1081-3 ต่อ 112-114

โทรสาร 0-2962-1084

พิมพ์ที่ บริษัท พงษ์วรินการพิมพ์ จำกัด

299-299/1 หมู่ 10 ถ.สุขุมวิท 107 ต.สำโรงเหนือ

อ.เมืองสมุทรปราการ จ.สมุทรปราการ 10270

โทรศัพท์ 0-2399-4525-31 โทรสาร 0-2399-4524



DATA SCIENCE PLATFORMS

ทางเลือกใหม่สำหรับ DATA SCIENTIST

องค์กรทุกขนาดในทุกอุตสาหกรรม ต่างมุ่งมั่นที่จะขับเคลื่อนด้วยข้อมูลมากขึ้น แต่ด้วยเกิดภาวะการขาดแคลน Data Scientists ที่จะมาตอบสนองได้ทันต่อความต้องการ ธุรกิจจึงสนใจแนวคิดใหม่ ที่จะมาเพิ่มขีดความสามารถให้กับคนที่มีอาชีพเป็นนักวิเคราะห์ข้อมูล ผู้เชี่ยวชาญด้านข้อมูล ผู้ที่มีประสบการณ์ในธุรกิจสาขาต่างๆ มาเรียนรู้แอปพลิเคชันซอฟต์แวร์บน Data Science and Machine Learning (DSML) Platforms โดยเรียกบุคลากรกลุ่มใหม่นี้ว่า “Citizen Data Scientists” และกลุ่ม Data Scientists ซึ่งเป็นผู้เชี่ยวชาญทางเทคนิคก็สามารถใช้ซอฟต์แวร์ DSML Platforms เพื่อ Develop & Deploy โมเดลไปรับใช้ในองค์กรได้เช่นเดียวกัน



ยุคของการวิเคราะห์และใช้ประโยชน์จากข้อมูล

ในปัจจุบัน เราอยู่ในยุคที่เต็มไปด้วยข้อมูล (Data) ที่หลากหลาย และถูกสร้างขึ้นใหม่มากมาย ในแต่ละวัน อีกทั้ง มีข้อมูลบางประเภทที่ถูกสร้างขึ้นอย่างต่อเนื่องตลอดเวลา (Data Streaming) มีทั้งข้อมูลที่เกิดจากการดำเนินชีวิตประจำวันของพวกเราเอง และจากอุปกรณ์สมาร์ตดีไวส์ (Smart Devices) ต่างๆ หากจะกล่าวถึงแหล่งที่มาของข้อมูลที่สำคัญหลักๆ แล้วมีดังนี้

ข้อมูลจากการซื้อสินค้าออนไลน์ (Online Shopping Data)

ข้อมูลการช้อปปิ้งออนไลน์จะถูกรวบรวมเมื่อมีผู้คนออนไลน์ และเริ่มมีประสบการณ์กับการสั่งซื้อสินค้า ร้านค้าออนไลน์ก็จะเก็บรวบรวมข้อมูลเกี่ยวกับข้อมูลส่วนตัว ประวัติการซื้อ การใช้ยืมเฉลี่ย พฤติกรรมลูกค้า โดยเราสามารถจัดเก็บเป็นชุดข้อมูลการช้อปปิ้งออนไลน์ (Online Shopping Datasets) ของลูกค้าแต่ละรายได้

นอกจากนี้ ยังมีข้อมูลที่สำคัญที่มักถูกใช้ในการวิเคราะห์ เพื่อประเมินเว็บไซต์ในด้านต่างๆ เช่น อันดับเพจที่ถูกเข้าชมมากที่สุด อันดับสินค้าที่ถูกซื้อมากที่สุด ลิงก์ที่ลูกค้าชอบคลิก และข้อมูลการช้อปปิ้งออนไลน์ที่น่าสนใจอื่นๆ รวมถึงข้อมูลทางธุรกรรมที่ได้จากผู้ซื้อจริงๆ เช่น องค์ประกอบการใช้งานตะกร้าสินค้า (Shopping Cart), วิธีการชำระเงิน (Payment) และความสัมพันธ์ระหว่างแบรนด์กับลูกค้า (Brand Affinity) เป็นต้น

ข้อมูลจากการใช้งานโทรศัพท์มือถือ (Mobile Data)

ข้อมูลที่อยู่ในรูปแบบของเนื้อหาที่เผยแพร่บนอินเทอร์เน็ต ที่ส่งไปยังอุปกรณ์พกพาต่างๆ เช่น บทความ ข่าวสาร อีเมล วิดีโอสตรีมมิ่ง การประชุมออนไลน์ บริการบนระบบคลาวด์ การเล่นเกมโซเชียลบนแพลตฟอร์มต่างๆ เป็นต้น



ยิ่งเทคโนโลยีเครือข่ายมือถือมีความรวดเร็ว และมีเครือข่ายครอบคลุมพื้นที่มากเท่าไร ก็เอื้อต่อการสื่อสารข้อมูลมากขึ้นเท่านั้น ในชีวิตประจำวันของคนทั่วโลกทุกวันนี้ จำเป็นต้องพึ่งพาข้อมูลบนมือถือเพื่อทำกิจกรรมทางออนไลน์ต่างๆ ด้วยกันทั้งนั้น ไม่ว่าจะเป็นการศึกษา การค้นหาความรู้ การทำธุรกรรม การเลือกซื้อสินค้า การทำงาน รับส่งอีเมล การพักผ่อน การสื่อสาร ดูหนังฟังเพลง ฟังพอดแคสต์ หรือการขอคำปรึกษาทางการแพทย์ ฯลฯ โดยมี Apps ต่างๆ เป็นตัวกลาง

ด้วยข้อมูลต่างๆ เหล่านี้ถูกสร้างขึ้นตลอด 24 ชั่วโมงแต่ละวันโดยไม่มีวันหยุด จึงทำให้ปริมาณข้อมูลโมบายล์ที่ใช้เพิ่มมากขึ้นอย่างมหาศาลในทุกๆ ปี ฉะนั้น Data Strategy เพื่อเปลี่ยนข้อมูลให้เป็นโอกาสทางธุรกิจ จึงกลายเป็นกลยุทธ์กระแสหลักของการดำเนินธุรกิจในยุคใหม่

ข้อมูลจากอุปกรณ์ Internet of Things (IoT Sensors Data)

เป็นอุปกรณ์ประเภท Smart Devices ที่มีการเชื่อมต่อ และแลกเปลี่ยนข้อมูลกับอุปกรณ์อื่นๆ ผ่านสัญญาณอินเทอร์เน็ต โดยใช้เซนเซอร์ (Sensor) เพื่อตรวจจับหรือวัดค่าต่างๆ แล้วนำข้อมูลที่ได้จาก IoT Sensors ไปวิเคราะห์ อย่างเช่น

- **เซนเซอร์วัดอุณหภูมิ (Temperature Sensors)** ใช้ตรวจจับการเปลี่ยนแปลงของอุณหภูมิ มีประโยชน์ในการคอนโทรลอุณหภูมิในโรงงาน เช่น

ในโรงงานแปรรูปอาหารทะเลแช่แข็ง การรักษาคุณภาพเพื่อคงความสดใหม่ของวัตถุดิบ ก่อนถึงโรงงานแปรรูปเป็นเรื่องที่สำคัญมาก จะต้องควบคุมอุณหภูมิภายในห้องเย็นให้ได้ตามมาตรฐาน

หรือการวัดอุณหภูมิในดิน ซึ่งเป็นปัจจัยสำคัญที่ส่งผลต่อการเจริญเติบโตและการพัฒนาการของพืช เซนเซอร์ช่วยให้เราเห็นความผันผวนของอุณหภูมิได้ง่าย และสามารถแก้ไขได้อย่างรวดเร็วก่อนที่จะเกิดปัญหา เป็นต้น



- **เซนเซอร์วัดค่าทางชีวภาพ (Biomedical Sensors)** ถูกใช้ในอุตสาหกรรมทางการแพทย์ ในการเก็บข้อมูลเพื่อวิเคราะห์ปัญหาด้านสุขภาพ หากตรวจพบข้อมูลที่บ่งบอกถึงความผิดปกติของร่างกาย ก็จะได้เข้ารับการรักษาเสียตั้งแต่เนิ่นๆ หรือใช้เพื่อติดตามการรักษาผู้ป่วย เรามักพบ Biomedical Sensors ได้ในอุปกรณ์สวมใส่ไว้กับร่างกาย (Wearable) เช่น Smart Watch ที่นอกจากจะใช้ดูเวลาแทนนาฬิกาแล้ว ยังมีฟังก์ชันอื่นๆ อีกมากมาย เช่น ใช้วัดความดันโลหิต วัดอัตราการเต้นของหัวใจ เป็นต้น โดยมีการทำงานผ่าน Apps

นอกจากนี้ ยังมีเซนเซอร์อีกมากมาย เช่น เซนเซอร์วัดค่าทางเคมี (Chemical Sensors), เซนเซอร์วัดความชื้น (Humidity Sensors), เซนเซอร์วัดความดัน (Pressure Sensors) และเซนเซอร์วัดแสง (Light Sensor) ฯลฯ

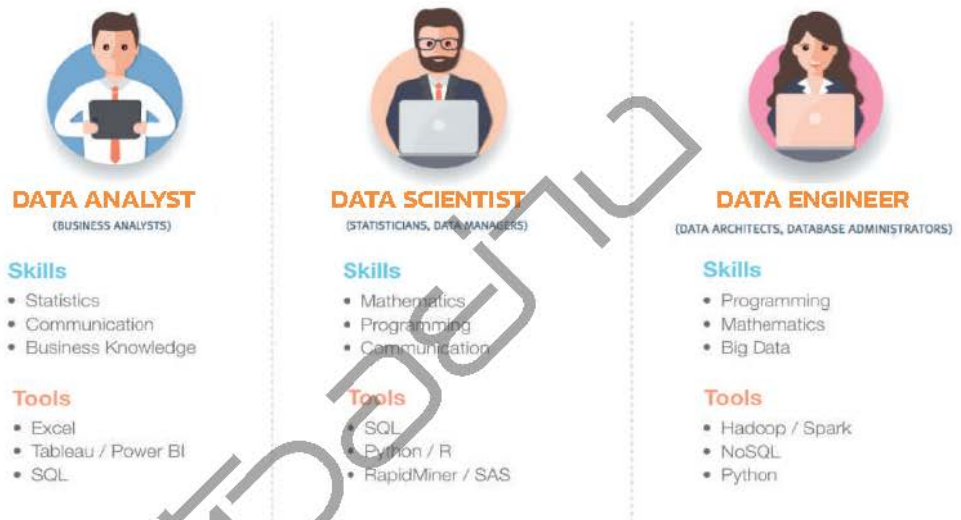
ด้วยการพัฒนาอย่างรวดเร็วของเทคโนโลยีทางสารสนเทศและคอมพิวเตอร์ ทำให้เราสามารถเก็บข้อมูลได้มากขึ้นด้วยค่าใช้จ่ายที่ถูกลง แต่ถ้าเราเก็บข้อมูลไว้มากมายแล้วไม่สามารถนำมาสร้างคุณค่าใดๆ ได้อย่างคุ้มค่า ก็คงไม่เกิดประโยชน์มากนัก ดังนั้น หลายๆ องค์กรจึงอยากนำข้อมูลเหล่านี้มาทำการวิเคราะห์อย่างจริงจัง แล้วหาวิธีใช้ประโยชน์จากข้อมูลเชิงลึกที่ได้รับ ซึ่งการวิเคราะห์ข้อมูลนี้อาจจะเป็นแบบง่ายๆ โดยใช้การหาค่าทางสถิติ อย่างเช่น

- ยอดขายรวมในแต่ละสาขา
- จำนวนผู้เข้าชมเว็บไซต์เฉลี่ยแต่ละเดือน ในช่วงไตรมาสที่หนึ่งของปี 2022
- การวิเคราะห์ในลักษณะของการคาดการณ์ล่วงหน้า เช่น ในอีก 3 เดือนข้างหน้า ยอดขายโทรศัพท์มือถือ iPhone จะเป็นจำนวนกี่เครื่อง



อาชีพทางด้าน Data Science

เนื่องจาก Data Science ไม่ใช่งานที่สามารถทำได้สำเร็จได้ด้วยคนเพียงคนเดียว จำเป็นต้องมีทีมสนับสนุน จึงทำให้เกิดอาชีพใหม่ขึ้นมาอย่างน้อย 3 อาชีพ ซึ่งมีบทบาทที่แตกต่างกันครับ ใครทำหน้าที่อะไร และต้องมีทักษะความเชี่ยวชาญอะไรเป็นพิเศษ ผมสรุปไว้ให้ดังนี้



รูปที่ 1 - 1 แสดงตำแหน่งงานที่เกี่ยวข้องกับ Big Data

http://insidebigdata.com/wpcontent/uploads/2016/04/springboard_infographic.jpg

- **Data Analyst** หรือ **นักวิเคราะห์ข้อมูล** จะมีหน้าที่ในการวิเคราะห์ข้อมูลเชิงธุรกิจ และนำผลลัพธ์ที่ได้มาจัดทำรายงานสรุป โดยแสดงผลในรูปแบบต่างๆ ที่เข้าใจง่ายสำหรับผู้ใช้งาน (Business User) เช่น รูปแบบของชาร์ต (Chart) หรือแดชบอร์ด (Dashboard) เป็นต้น นอกจากนี้ อาจจะช่วยแปลโจทย์ทางด้านธุรกิจให้เป็นหัวข้อทางการวิเคราะห์ข้อมูล เพื่อใช้ประสานงานกับนักวิทยาศาสตร์ข้อมูล สำหรับนำไปวิเคราะห์ข้อมูลในเชิงลึกต่อไปได้อีกด้วย



ทักษะที่สำคัญของนักวิเคราะห์ข้อมูล คือ การใช้ซอฟต์แวร์ในการทำ Self-Service Analytics ได้เช่น Excel, Tableau, Power BI, Google Data Studio หรือเขียนโปรแกรมด้วยภาษา SQL ได้ เพื่อแสดงผลการวิเคราะห์เบื้องต้นให้เข้าใจได้ง่าย

- **Data Scientist** หรือ **นักวิทยาศาสตร์ข้อมูล** จะมีหน้าที่วิเคราะห์ข้อมูลเชิงลึกด้วยเทคนิคต่างๆ เช่น วิทยาศาสตร์ข้อมูล (Data Science), เหมืองข้อมูล (Data Mining), การเรียนรู้ของเครื่อง (Machine Learning) โดยอาจจะมีการพัฒนาโมเดล (Model) ทางคณิตศาสตร์ขึ้นมาใหม่ เพื่อให้ตอบโจทย์ทางธุรกิจของผู้ใช้ได้มากที่สุด

ทักษะที่สำคัญของนักวิทยาศาสตร์ข้อมูล คือ ความเข้าใจทางคณิตศาสตร์และสถิติ สามารถเขียนโปรแกรม หรือใช้งานซอฟต์แวร์เพื่อทำการวิเคราะห์ข้อมูลเชิงลึกได้ โดยภาษาโปรแกรมมิ่งที่ได้รับความนิยม เช่น การเขียนโปรแกรมด้วยภาษา R, Python และ SQL หรือซอฟต์แวร์ RapidMiner

- **Data Engineer** หรือ **วิศวกรข้อมูล** ตำแหน่งนี้จะไม่ค่อยได้ติดต่อกับผู้ใช้โดยตรงจะเน้นในส่วนของการจัดการดูแลระบบ เพื่อให้รองรับการทำงานของนักวิทยาศาสตร์ข้อมูลให้มีประสิทธิภาพทั้งในแง่ของความเร็วในการประมวลผล และการรองรับข้อมูลที่มีปริมาณมากได้

ทักษะที่สำคัญของวิศวกรข้อมูล คือ การใช้เทคโนโลยีในการจัดการข้อมูลต่างๆ เช่น SQL ในการจัดการฐานข้อมูล หรือการติดตั้งและจัดการระบบ Hadoop ภายในองค์กร

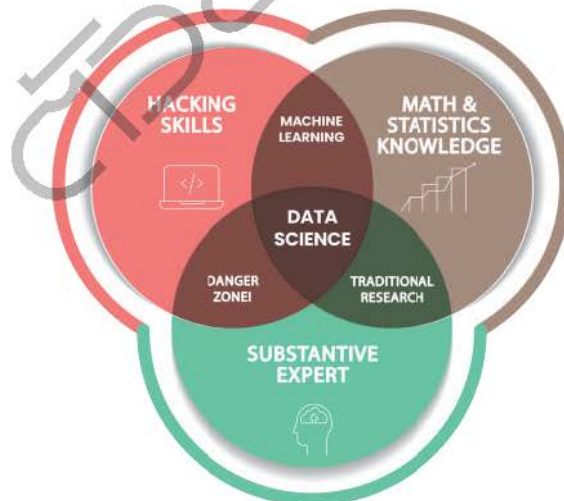


Data Scientist และ Citizen Data Scientist

การคาดการณ์อนาคตคือ การวิเคราะห์ข้อมูลในระดับที่ซับซ้อนที่เป็นส่วนหนึ่งของแนวทาง **วิทยาศาสตร์ข้อมูล** หรือ **Data Science** เป็นศาสตร์ที่กำลังเป็นกระแสนิยมในปัจจุบัน ธุรกิจและอุตสาหกรรมด้านต่างๆ ทั่วโลกต่างตื่นตัวกับการขับเคลื่อนด้วยเทคโนโลยีทางด้าน Data Science อย่างเห็นได้ชัด มองเห็นศักยภาพจนเป็นที่ประจักษ์แล้วว่า มันสร้างให้เกิดคุณค่าต่อธุรกิจได้อย่างแท้จริงจากตัวอย่างเคสมากมาย และเป็นกลยุทธ์สำคัญต่อความอยู่รอดในอนาคต

Data Scientist สาย Depth (ดั้งเดิม)

ด้วยเหตุผลข้างต้น จึงมีความต้องการบุคลากรทางด้าน **วิทยาศาสตร์ข้อมูล** หรือ **Data Science** เป็นจำนวนมาก แต่การพัฒนาบุคลากรทางด้าน Data Science ก็ไม่ได้ง่ายครับ เพราะเป็นเรื่องใหม่ที่ต้องอาศัยองค์ความรู้และทักษะหลายสาขาประกอบกัน โดยส่วนผสมขององค์ความรู้ทาง Data Science ที่ถูกอธิบายอย่างเป็นรูปธรรมเคยถูกกล่าวไว้โดย Drew Conway ดังแสดงในรูปที่ 1 – 2 เมื่อปี ค.ศ. 2010

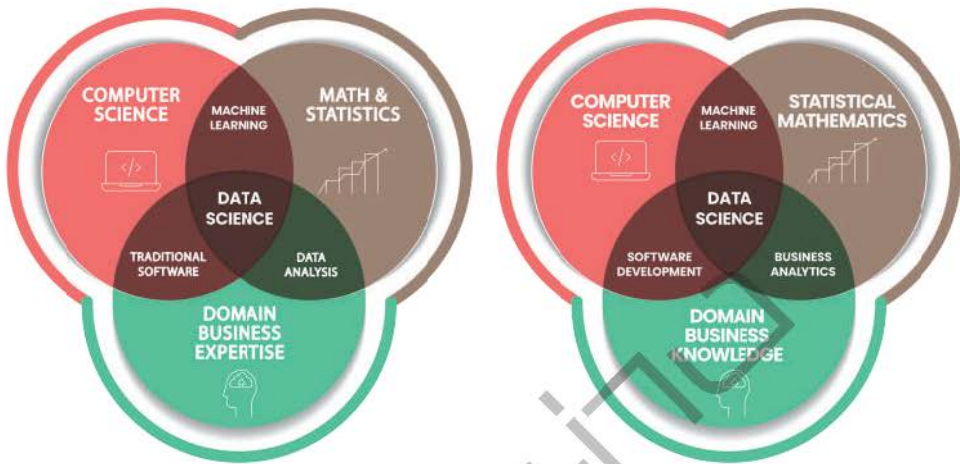


รูปที่ 1 – 2 แผนภาพ Data Science Venn Diagram โดย Drew Conway

<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>



ต่อมาได้เกิดแผนภาพเพื่อใช้อธิบาย Data Science ด้วยมุมมองใหม่ๆ ที่ดูเป็นทางการกว่าต้นฉบับอีกหลายแบบ แต่ความหมายก็คล้ายคลึงกัน ผมขออธิบายสรุปในภาพรวมไว้ดังนี้



รูปที่ 1 – 3 แผนภาพ Data Science Venn Diagram แบบอื่นๆ

- **Computer Science** ความรู้ทางด้านวิทยาศาสตร์คอมพิวเตอร์ หรือไอที เช่น เข้าใจองค์ประกอบและระบบการทำงานของคอมพิวเตอร์ การสื่อสารข้อมูลบนเครือข่าย การเขียนโปรแกรมทาง Data Science โดยใช้ภาษาที่จำเป็น เช่น Python, R หรือ SQL เป็นต้น หรือการใช้ซอฟต์แวร์สำเร็จรูปประเภท No Code/Low Code
- **Math and Statistics** ความรู้ทางด้านคณิตศาสตร์และสถิตินั้น เป็นรากฐานที่สำคัญของการวิเคราะห์ข้อมูล เนื่องจากข้อมูลที่ถูกนำมาใช้ในกระบวนการวิเคราะห์จะเป็นข้อมูลทางสถิติไม่ว่าจะเป็นข้อมูลเชิงปริมาณ (Quantitative Data) และข้อมูลเชิงคุณภาพ (Qualitative Data) ไปจนถึงเทคนิคการพัฒนาโมเดลการเรียนรู้ของเครื่อง (Machine Learning)
- **Domain/Business Knowledge** การเข้าใจลักษณะธุรกิจอย่างลึกซึ้งเป็นเรื่องที่สำคัญมาก หากขาดความรู้ความเข้าใจในธุรกิจอย่างถ่องแท้ ก็ยากที่จะพัฒนาโปรเจกต์ทาง Data Science ให้ตอบโจทย์ หรือแก้ปัญหาให้กับธุรกิจได้ มีโอกาสล้มเหลวสูง ซึ่งข้อนี้ทำให้บริษัทหาคนที่เหมาะสมได้ยากมากทีเดียวครับ



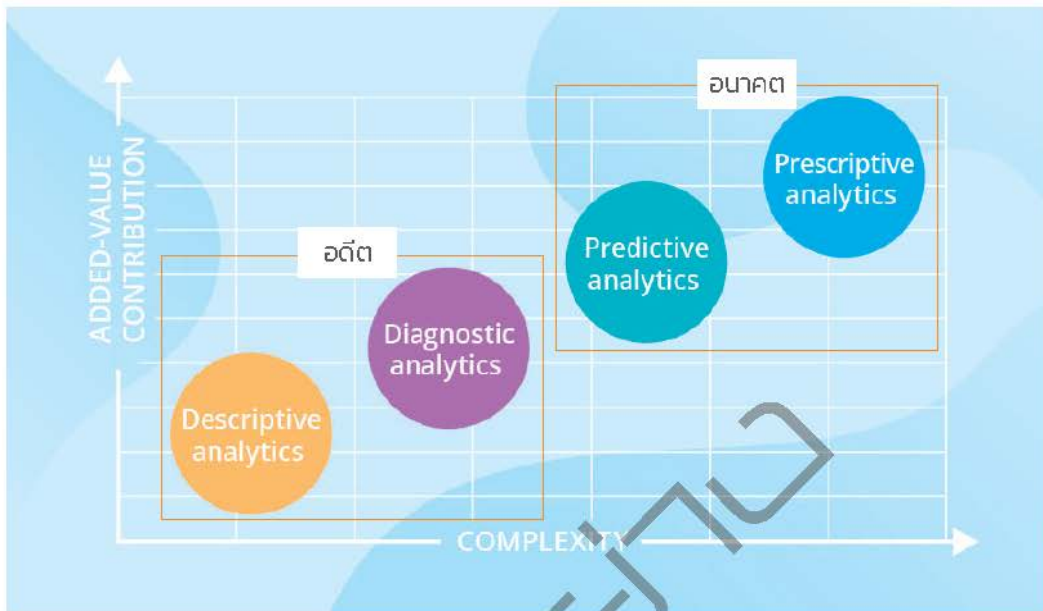
ด้วยอุปสรรคสำคัญในการสรรหาบุคลากร จึงเกิดแนวทางใหม่คือ การเลือกคนที่มีคุณสมบัติเหมาะสม โดยไม่จำกัดว่าจะเป็นบุคลากรภายในหรือภายนอกองค์กร หรือจะเป็นสายเทคนิคอล (Technical) หรือสายธุรกิจในฟังก์ชันงานต่างๆ (Nontechnical) ก็ตาม ให้มาเรียนรู้เรื่องการวิเคราะห์ข้อมูล และการใช้ซอฟต์แวร์แอปพลิเคชันทางด้าน Data Science and Machine Learning (DSML) Platforms เพิ่มเติม โดยเน้นผลลัพธ์ทางธุรกิจเป็นหลัก อาจไม่ต้องถึงขั้นมีสเกลสูงทางด้านการเขียนโปรแกรมก็ได้ เพราะสามารถใช้ซอฟต์แวร์สำเร็จรูปแทนได้ เราเรียก Data Scientist ประเภทนี้ว่า **“Citizen Data Scientist”**

Citizen Data Scientist สาย Simplicity (ทางเลือกใหม่)

Citizen Data Scientist คำคำนี้ถูกกล่าวขึ้นโดย Gartner ที่ต้องการสื่อถึงกลุ่มคนที่เหมาะจะผันตัวตัวเองมาทำงานด้านนี้ โดยธุรกิจมีส่วนช่วยเพิ่มศักยภาพด้วยเครื่องมือ และการสนับสนุนที่จำเป็นในการเปลี่ยนสายงานให้มีสเกลด้าน Data Science โดยกลุ่มคนที่กล่าวถึง เช่น นักวิเคราะห์ข้อมูลขั้นสูง (Data Analytics Professionals), ผู้เชี่ยวชาญด้านข้อมูล (Data Professional) หรือผู้เชี่ยวชาญด้านข้อมูลในสาขาต่างๆ (Business/Domain Experts) เช่น การขาย การตลาด การเงิน หรือทรัพยากรบุคคล ฯลฯ

บุคลากรเหล่านี้ต่างก็มีความรู้เกี่ยวกับปัญหา หรือความท้าทายทางธุรกิจที่ต้องเผชิญอยู่แล้ว พอได้รู้จักการใช้ซอฟต์แวร์สำเร็จรูปที่มีความฉลาด (DSML Platforms) ก็สามารถนำข้อมูลที่มีอยู่แล้วมาวิเคราะห์ แล้วสร้างโมเดลเชิงคาดการณ์ถึงสิ่งที่อาจเกิดขึ้นในอนาคต (Predict) หรือแนะนำสิ่งที่ควรทำ (Prescribe) ตามระดับความซับซ้อนของการวิเคราะห์

หากเป็นเมื่อก่อนเราต้องพึ่งพานักวิทยาศาสตร์ข้อมูล นักสถิติ หรือนักคณิตศาสตร์เท่านั้น ทุกวันนี้ แนวคิด Citizen Data Scientist จึงกลายเป็นกระแสหลักจนอาจกล่าวได้ว่า “The Age of the Citizen Data Scientist”



► รูปที่ 1 - 4 การทำความเข้าใจการวิเคราะห์ข้อมูลในระดับที่ซับซ้อนและคุณค่าที่ได้รับ

เครดิตภาพ : www.scnsoft.com/blog/4-types-of-data-analytics

อยากให้เข้าใจว่า Citizen Data Scientist ไม่ได้จะเข้ามาแทน Data Scientist ดังเดิมแต่อย่างใด แต่คือส่วนที่เข้ามาเติมเต็มให้ทีม Data Science มีความยืดหยุ่นมากขึ้น เพราะการใช้แพลตฟอร์มหรือซอฟต์แวร์สำเร็จรูปนั้นทำได้ง่าย รวดเร็ว และประหยัด ที่สำคัญคือ หาบุคลากรที่เหมาะสมได้ง่ายกว่า ช่วยให้ธุรกิจสามารถปรับเปลี่ยนเพื่อเปลี่ยนแปลงด้วย Data Science ได้จริงมากขึ้นครับ

ด้วยแนวคิดดังกล่าว ผมจึงตั้งใจเขียนหนังสือเล่มนี้ขึ้นมา เพื่อให้โอกาสกับคนที่อยากทำงานด้าน Data Science ได้เข้าใจแนวคิดการวิเคราะห์ข้อมูลต่างๆ ทาง Machine Learning โดยเรียนรู้จากการลงมือปฏิบัติจริงโดยใช้ซอฟต์แวร์ที่ชื่อว่า **RapidMiner** ซึ่งถูกกล่าวถึงในรายงาน 2 ฉบับของ Gartner ว่า “RapidMiner is a trustworthy choice.”

อ้างอิง : <https://rapidminer.com/downloads/2022-gartner-market-guides/>



RapidMiner เครื่องมือสำหรับ Citizen Data Scientist

RapidMiner คือ **Data Science Platform** ที่มีส่วนผลักดันให้เกิดการมีส่วนร่วมในงาน Data Science มากขึ้น มีวิธีการใช้งานที่เหมาะสมสำหรับทุกคนทุกระดับทักษะ แพลตฟอร์มนี้มีขั้นตอนการดำเนินงานตามวงจรที่ช่วยให้งานสามารถดำเนินไปได้อย่างรวดเร็ว เริ่มตั้งแต่การเตรียมข้อมูล (Data Prep) ไปจนถึงการใช้งานโมเดล (Model Ops)



รูปที่ 1 – 5 RapidMiner : Data Science Lifecycle



RapidMiner อนาคตของทีมสนับสนุน

บทบาทในภาพรวมของผู้ที่มีส่วนเกี่ยวข้องหลักๆ มีดังนี้



- **Data Science Experts :** แม้ว่าบุคลากรกลุ่มนี้จะมีความพร้อมในการดำเนินโปรเจกต์ที่ซับซ้อนของงาน Data Science อยู่แล้ว แต่ RapidMiner เป็นเหมือนทางเลือกที่ช่วยให้งานบางตัวสำเร็จเร็วขึ้น เพื่อจะได้เอาเวลาที่มีค่าไปทุ่มให้กับงานที่มีมูลค่าสูงกว่า หรือแม้กระทั่งการให้ความช่วยเหลือกลุ่มคนที่เขียนโค้ดไม่ได้ที่มี ซึ่งเป็นประโยชน์ต่อธุรกิจอย่างแท้จริง



- **Information Technology (IT) :** เมื่อโมเดลถูกสร้างสำเร็จแล้ว ฝ่ายไอทีมีบทบาทสำคัญในการมอนิเตอร์ว่า โมเดลสามารถดำเนินการได้อย่างเหมาะสม และเชื่อมต่อกับระบบที่มีอยู่โดยไม่กระทบต่อความปลอดภัย เพื่อให้องค์กรได้รับมูลค่าทางธุรกิจเพิ่มขึ้นจากโมเดล RapidMiner มีมาตรฐานความปลอดภัยระดับสูง โดยแอดมินสามารถควบคุมการเข้าถึงข้อมูลและโครงการต่างๆ ได้อย่างเต็มที่ เพื่อให้เป็นไปตามนโยบายและข้อบังคับขององค์กร



- **Domain Experts :** บุคลากรเหล่านี้ต่างก็มีความรู้เกี่ยวกับปัญหาทางธุรกิจที่ต้องเผชิญอยู่แล้ว โดยจะต่อยอดด้วยการสร้างโมเดลเชิงคาดการณ์ (Predictive Models) ที่จะช่วยให้การตัดสินใจที่สำคัญมีประสิทธิภาพมากยิ่งขึ้น คนกลุ่มนี้มักเขียนโค้ดไม่ได้ จึงสามารถใช้ RapidMiner เป็นเครื่องมือสร้างการทำนายผลลัพธ์ในอนาคตได้



RapidMiner คืออะไร

ในยุคของการใช้ประโยชน์จากข้อมูล การพยากรณ์อนาคตกำลังเป็นที่นิยม ทำให้การวิเคราะห์เชิงพยากรณ์ ซึ่งเป็นการวิเคราะห์ระดับสูงกลายเป็นเครื่องมือที่สำคัญ RapidMiner คือ แพลตฟอร์ม Automated Data Science แบบ No Code/ Low Code สำหรับสร้างโมเดลเชิงคาดการณ์ เพื่อทำนายผลลัพธ์ในอนาคตได้แบบอัตโนมัติ ตอบโจทย์ทั้งบุคลากรสายเทคนิค และสายธุรกิจในฟังก์ชันงานต่างๆ

Best Practices for Reskill and Upskilling

Workshops : RapidMiner Turbo Prep และ Auto Model

- Marketing : การแบ่งกลุ่มลูกค้าด้วยวิธี RFM Segmentation
- Human Resources : การคาดการณ์การลาออกของพนักงาน
- Logistics : การแนะนำรถยนต์ที่เหมาะสมสำหรับการขนส่ง
- Energy : การพยากรณ์ปริมาณการใช้ไฟฟ้าเพื่อทำความร้อน
- Healthcare : การพยากรณ์ว่าใครจะเป็นโรคเบาหวานบ้าง

การพัฒนา Machine Learning UU RapidMiner Platform ถืออย่างไร

- AI for Everyone ออกแบบมาให้ง่ายสำหรับทุกคน
- เปลี่ยนงานที่ยุ่งยากมาเป็นซอฟต์แวร์ที่ใช้ง่าย รวดเร็ว และประหยัด
- ช่วยทำงานที่ไม่ซับซ้อน เพื่อให้โปรแกรมเมอร์ไปทำงานที่มีมูลค่าสูงกว่า
- คนสาย Tech และ Non-Tech ทำงานร่วมกันได้บนแพลตฟอร์มเดียวกัน
- ซัพพอร์ต Applications, Data Sources, Systems, Coding Languages, Machine Learning Libraries และเทคโนโลยี Cloud ที่หลากหลาย

ครอบคลุมเทคนิคการสร้างโมเดล • Decision Tree • Naive Bayes • Linear Regression • Logistic Regression

การสร้างโมเดล (Build Model) ในแต่ละเทคนิค (Algorithm) จะแบ่งออกเป็น 3 หัวข้อ



1 สรุปแนวคิดของการสร้างโมเดล

อธิบายแนวคิดแบบกระชับ พร้อมตัวอย่างง่ายๆ



2 ข้อดี - ข้อจำกัดของเทคนิค

ประโยชน์ที่จะได้รับ รวมถึงข้อจำกัดการใช้งาน



3 RapidMiner Auto Model

เวิร์กช็อปการสร้างโมเดลด้วย RapidMiner แบบ Step by Step

Programs and Fields of Study

✓ นักวิเคราะห์ข้อมูล

✓ ผู้เชี่ยวชาญด้านข้อมูล

✓ นักสถิติหรือนักคณิตศาสตร์

✓ ผู้เชี่ยวชาญด้านธุรกิจสาขาต่างๆ

✓ บุคลากรสาย IT/ICT

✓ นักวิทยาศาสตร์ข้อมูล



ดร.เอกสิทธิ์ พงษ์ทองकर्ณ
Data Science & RapidMiner Expert
กัศพล ภิชาเจริญ
บรรณาธิการ

สั่งหนังสือผ่านช่องทางออนไลน์ได้ที่



จัดจำหน่ายโดย IDC

Barcode 885-916-101-074-6



ราคา 385 บาท