

short Note

The Data scientist 's Journey

โลกของ AI & Data Scientist Roadmap ฉบับย่อที่อ่านง่ายสำหรับทุกคน



อาชีพไหนก็ไม่ Sexy
เท่า Data Scientist



โดย ดร.นฤตม์ สุนทรานนท์
Data Scientist และไอติมบุฟเฟ่ Nerd



มีเพียง “ความรู้” เท่านั้นที่มนุษย์ใช้พลิก “โลก”
และเปลี่ยน “ชีวิต” เราจึงสร้างสรรค์ และส่งมอบ “ความรู้”
ในรูปแบบที่ดีกว่า เพื่อให้คนไทย “เรียนรู้” ได้ตลอดชีวิต

Only “Knowledge” can help human
change “The World” and “Their Lives”.
With this truth, it drives us to deliver
“Knowledge” for Thai being able to
“Learn” better everyday.



Short Note The Data Scientist's Journey

Writer	ดร.นฤตม์ สุนทรานนท์
Editor	ภีรพล คุษาเจริญ
Production Manager	วรพล ณธิกุล
Graphic Designer	ชวรินทร์ รัตนะ
Page Layout	สุรัสวดี วงศ์จันทร์สุข
Proofreader	พรรณรัตน์ ขุราชี
Publishing Coordinators	สุพัตรา อาจปฐ, สุรีย์รัตน์ จิ๋ว
Product Specialist	ศรัณย์ คมขำ

เครื่องหมายการค้าอื่นๆ ที่อ้างถึงเป็นของบริษัทนั้นๆ สงวนลิขสิทธิ์ตามพระราชบัญญัติลิขสิทธิ์ พ.ศ. 2537 โดยบริษัท ไอดีซี พรีเมียร์ จำกัด ห้ามลอกเลียนไม่ว่า ส่วนใดส่วนหนึ่งของหนังสือเล่มนี้ ไม่ว่าในรูปแบบใดๆ นอกจากจะได้รับอนุญาตเป็นลายลักษณ์อักษรจากผู้จัดพิมพ์เท่านั้น

บริษัท ไอดีซี พรีเมียร์ จำกัด จัดตั้งขึ้นเพื่อเผยแพร่ความรู้ที่มีคุณภาพสู่ผู้อ่านชาวไทย เรายินดีรับงานเขียนของนักวิชาการและนักเขียนทุกท่าน ท่านผู้สนใจกรุณาติดต่อผ่านทางอีเมลที่ infopress@idcpremier.com หรือทางโทรศัพท์หมายเลข 0-2962-1081 (อัตรานาที 10 คู่สาย) โทรสาร 0-2962-1084

สร้างสรรค์โดย



พิมพ์ครั้งที่ 1 พฤษภาคม 2566

ข้อมูลทางบรรณานุกรม :

ดร.นฤตม์ สุนทรานนท์

Short Note The Data Scientist's Journey

นนทบุรี : ไอดีซีฯ, 2566

296 หน้า

1. วิศวกรรมซอฟต์แวร์

I ชื่อเรื่อง

005.1

Barcode 885-916-101-047-0

ราคา 450 บาท

พิมพ์และจัดจำหน่ายโดย



บริษัท ไอดีซี พรีเมียร์ จำกัด

200 หมู่ 4 ชั้น 19 ห้อง 1901

จัสตินอินเตอร์เนชั่นแนลทาวเวอร์

ถ.แจ้งวัฒนะ อ.ปากเกร็ด จ.นนทบุรี 11120

โทรศัพท์ 0-2962-1081 (อัตรานาที 10 คู่สาย)

โทรสาร 0-2962-1084

พิมพ์ที่ บริษัท ดี.เค.ปรี้นติ้ง จำกัด

441/56 หมู่ 2 ต.บางบ่อ อ.บางบ่อ

จ.สมุทรปราการ 10560

โทรศัพท์ 0-2115-9105

โทรสาร 0-2115-9044

ร้านค้าและตัวแทนจำหน่าย

โทรศัพท์ 0-2962-1081-3 ต่อ 112-114

โทรสาร 0-2962-1084

สมาชิกสัมพันธ์

โทรศัพท์ 0-2962-1081-3 ต่อ 121

โทรสาร 0-2962-1084

PREFACE



จากกระแสความนิยมในไม่กี่ปีก่อนหน้านี้ เกี่ยวกับตำแหน่งงาน Data Scientist ถึงขนาดนิตยสารชื่อดังระดับโลกอย่าง Harvard Business Review (HBR) ปี 2012 บอกว่า จะเป็นงานที่ Sexy ที่สุดแห่งศตวรรษที่ 21 มีสถาบันการศึกษาเปิดหลักสูตรที่เกี่ยวข้องกับ Data Science และมีหลักสูตรการเรียน Online ต่างๆ มากมาย เพื่อสร้าง Data Scientist ให้เพียงพอกับความต้องการในอุตสาหกรรมต่างๆ เช่น โทรคมนาคม การเงิน คำปลีก พลังงาน รวมไปถึง Start-up Companies อีกจำนวนมาก

ประกอบกับความเชื่อที่ว่า Data Scientist เป็นตำแหน่งงานที่สำคัญในอนาคต เนื่องจากบริษัทต่างๆ ก็อยากปรับตัวไปสู่ Data-Driven Organization และค่าตอบแทนของ Data Scientist ก็สูงในระดับต้นๆ หากเทียบกับตำแหน่งงานอื่นๆ ทำให้มีผู้สนใจสายงานนี้เป็นจำนวนมาก และบ่อยครั้งก็เกิดคำถามว่า เส้นทางสู่อาชีพ Data Scientist เป็นอย่างไร?

หนังสือเล่มนี้เป็นการสรุปเส้นทางไปสู่อาชีพ Data Scientist ที่ต้องมีพื้นฐานใน 3 เรื่องคือ คณิตศาสตร์, การเขียนโค้ด และความเข้าใจในธุรกิจ รวมถึงความเข้าใจการใช้ประโยชน์จาก Data ที่มีอยู่มาสร้างคุณค่าทางธุรกิจ ในหนังสือพูดถึงเทคโนโลยีอย่าง Machine Learning และ Deep Learning รวมถึง Generative AI ที่โด่งดังในปัจจุบันอย่าง ChatGPT มีการยกตัวอย่าง Business Cases ต่างๆ ในอุตสาหกรรมการเงินและคำปลีก เพื่อให้ผู้อ่านได้เห็นภาพมากขึ้น นอกจากนี้ยังพูดถึง Soft Skills ที่จำเป็นสำหรับ Data Scientist เช่น ทักษะการสื่อสาร การเรียนรู้ด้วยตนเอง และปิดท้ายด้วยการเตรียมตัวสัมภาษณ์งาน โดยในแต่ละ Note จะมีการแนะนำแหล่งเรียนรู้เพิ่มเติมทั้งหนังสือและคอร์สออนไลน์

หวังเป็นอย่างยิ่งว่าหนังสือเล่มนี้จะเป็นไกด์ไลน์ให้กับทุกๆ คนที่อยากพัฒนาตนเองไปสู่ตำแหน่ง Data Scientist ได้มีแนวทางไปสู่จุดหมายได้ตามความตั้งใจ หรือผู้ที่อยู่ในสายงานนี้แล้วได้ขัดเกลาความรู้และทักษะต่างๆ ให้เฉียบคมมากยิ่งขึ้น หรือแม้แต่คนทั่วไปที่อยากเข้าใจโลกของ Data Science หนังสือเล่มนี้ก็ตอบโจทย์เช่นเดียวกัน

ดร.บุรณิ สุนทรานนท์, Data Scientist และเจ้าของเพจ Nerd

Data Analytics | AI | Machine Learning | Deep Learning

ประวัติผู้เขียน

- ★ เจ้าของเพจ Nerd, Short Notes : Data Analytics | AI | Machine Learning | Deep Learning <https://www.facebook.com/nerd.data>
- ★ เคยมีประสบการณ์ในการนำทีม Data Science (Senior Data Science Manager) ที่ Central Digital (อุตสาหกรรมค้าปลีก)
- ★ เคยมีประสบการณ์ในการนำทีม Data Science (Lead Data Scientist) ที่ Krungsri Consumer (อุตสาหกรรมการเงิน)
- ★ เคยเป็นนักวิจัยด้าน Computer Vision สำหรับงานประยุกต์เรื่องเกษตรแม่นยำ (Precision Agriculture) ที่ GISTDA (Geo-Informatics and Space Technology Development Agency)
- ★ เป็นอาจารย์พิเศษและวิทยากรบรรยายเรื่อง Data Analytics | AI | Machine Learning | Deep Learning ในสถาบันการศึกษา รวมถึงหน่วยงานภาครัฐและเอกชนชั้นนำของประเทศ
- ★ ร่วมก่อตั้งบริษัท Dev Drone Mapper (DDM) ให้บริการด้าน Big Geo-Spatial Data Analytics (Drone / Satellite / Terrain Data)
- ★ จบปริญญาเอกสาขา Computer Science จาก French Mapping Agency (IGN) & University of Picardy Jules Verne สาธารณรัฐฝรั่งเศส

EDITOR'S NOTE



โลกของ Data Science มีกระแสเติบโตอย่างต่อเนื่องควบคู่กับการกำเนิดของอาชีพใหม่ๆ หนึ่งในนั้นก็คือ Data Scientist ที่รับบทบาทสำคัญในการพัฒนา AI Applications เพื่อมาตอบสนองธุรกิจที่ขับเคลื่อนด้วย AI เป็นอาชีพที่มุ่งเรียนวิทยาการทางด้าน Data Science ที่ใช้ในการสร้างโมเดล Intelligence ซึ่งเป็นหัวใจหลักของธุรกิจในอุตสาหกรรมต่างๆ รวมไปถึงบริษัท Start-up ยุคใหม่

เมื่อกระแสของ AI มีการเติบโตอย่างก้าวกระโดด อาชีพ Data Scientist จึงยังคงเป็นที่ต้องการในอนาคต และยังพัฒนาทักษะใหม่ๆ ได้ในอนาคต เช่น Prompt Engineering Skills ซึ่งเป็นรูปแบบข้อความ หรือ Prompt ที่ใช้สื่อสารกับ Generative AI (Gen AI) เพื่อสั่งให้มันสร้าง New Content ประเภทต่างๆ ที่สอดคล้องกับสิ่งที่เราต้องการ เช่น ให้ข้อมูลลูกค้า เขียนบทความ คิดแคมเปญโฆษณา วาดภาพ แต่งดนตรี ตัดต่อวิดีโอ ฯลฯ นิยมเรียกว่า AI สายสร้างสรรค์

หนังสือเล่มนี้เรียบเรียงโดย ดร.นฤตม์ สุนทรานนท์ (คุณใหม่) ซึ่งสามารถสรุปย่อจักรวาลของ Data Science มาเล่าให้เข้าใจง่ายๆ โดยไม่เน้นทฤษฎีหรือเนื้อหาหนักๆ เพื่อให้ทุกคนอ่านได้แม้ไม่ใช่คนสายเทคโนโลยี โดยนำเสนอ Data Scientist Roadmap เริ่มตั้งแต่อธิบายหลักการพื้นฐาน เทคนิคการพัฒนาระบบให้เรียนรู้จากข้อมูล ไปจนถึงการเตรียมตัวสัมภาษณ์งาน ครอบคลุมทั้ง Hard และ Soft Skills ในแต่ละ Note ยังแนะนำ Bootcamps, Short Courses, Podcasts และ Books / Blogs ที่คัดสรรมาจากประสบการณ์ตรงของผู้เขียน นอกจากจะเข้าใจโลกของ Data Science และ Data Scientist Roadmap แล้ว ยังจะก้าวทันเทรนด์ AI ล่าสุดอย่าง Generative AI อีกด้วย

“เรากำลังก้าวเข้าสู่ยุค ‘AI is Eating the World’ จงเลือกอาชีพที่โลกอนาคตต้องการ”

กัศวล คุษาเจริญ

บรรณาธิการ

CONTENTS

PART 1 เส้นทางสู่อาชีพ DATA SCIENTIST

ในโลกยุค AI10

NOTE 01 AI + BIG DATA + DATA SCIENCE

ขับเคลื่อนสู่โลกใหม่ 11

NOTE 02 เข้าใจงานและแนวทางการเรียนรู้

เพื่อเป็น DATA SCIENTIST 47

PART 2 MATH & STATISTICS FOR DATA

SCIENCE74

NOTE 03 พื้นฐานคณิตศาสตร์สำหรับ DATA SCIENCE.... 75

NOTE 04 พื้นฐานสถิติสำหรับ DATA SCIENCE 101

PART 3 PROGRAMMING LANGUAGES FOR

DATA SCIENCE138

NOTE 05 ภาษา SQL & PYTHON สำหรับ

DATA SCIENCE 139

PART 4 การสร้าง AI ด้วย MACHINE LEARNING

& DEEP LEARNING174

NOTE 06 MACHINE LEARNING การเรียนรู้ของเครื่อง... 175

NOTE 07 DEEP LEARNING การเรียนรู้เชิงลึก 213



PART 5 BUSINESS CASES 248

NOTE 08 การประยุกต์ใช้ DATA SCIENCE ในธุรกิจ 249

PART 6 SOFT SKILLS ทักษะสู่ความสำเร็จของ DATA SCIENTIST 260

NOTE 09 สื่อสารให้น่าสนใจ เล่าเรื่องให้ตรงประเด็น 261

NOTE 10 การเรียนรู้ด้วยตนเอง SELF-LEARNING 273

NOTE 11 วิธีเตรียมตัวสัมภาษณ์งาน Data Science 287



อาชีพไหนก็ไม่ Sexy เท่า Data Scientist

DATA SCIENTIST ROADMAP

A Quick Guide to Becoming a Data Scientist



PART 1

เส้นทางสู่อาชีพ DATA SCIENTIST ในโลกยุค AI



**NOTE
01**

**AI + BIG DATA
+ DATA SCIENCE
ขับเคลื่อนสู่โลกใหม่**



**"THE WORLD IS
ONE BIG DATA PROBLEM."**



Andrew McAfee

Co-director of the MIT initiative on the Digital Economy

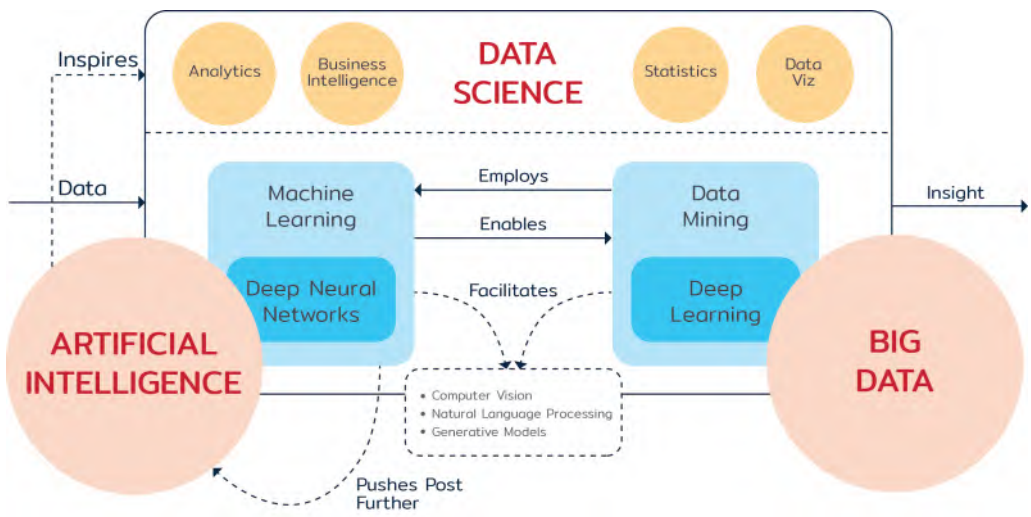
Data มาก, Problems ก็มาก

- ★ พวกเราอยู่ในยุคที่มี Data เป็นแกนหลัก ทุกสิ่งที่เกิดขึ้นในอดีตช่วยให้เราเข้าใจถึงประสิทธิภาพ การแก้ปัญหา การปรับปรุงกระบวนการ และความเข้าใจในตัวผู้บริโภค
- ★ แต่ปัญหาที่โลกกำลังเผชิญคือ Data มันใหญ่เกินไป และซับซ้อนเกินกว่าวิธีการประมวลผลข้อมูลแบบเดิมจะจัดการได้ นิยมเรียกว่า “Big Data”
- ★ การสกัดคุณค่าที่แท้จริงของ Data ออกมา จำเป็นต้องใช้วิทยาการ Data Science ซึ่งเป็นกระบวนการที่เกี่ยวข้องกับการใช้วิธีการทางสถิติและการคำนวณ
- ★ คุณค่าที่แท้จริงของ Data จำเป็นต้องผ่านการทำ Analytics เพื่อให้ได้มาซึ่ง Insights, Recommendations, Predictions หรือ Story ที่ช่วยผลักดันให้เกิดคุณค่าทางธุรกิจ
 - ★ Better Decisions สร้างการตัดสินใจที่ดีขึ้น
 - ★ New Efficiencies เพิ่มประสิทธิภาพการดำเนินการของธุรกิจ
 - ★ Process Automation สร้างกระบวนการให้เป็นอัตโนมัติ
 - ★ Better Experiences สร้างประสบการณ์ที่ดีขึ้นให้กับลูกค้า
- ★ ในปัจจุบัน Data เป็นหนึ่งในทรัพยากรที่ดีที่สุด ถูกนำมาใช้ผ่านเครื่องมือต่างๆ ในการพัฒนาองค์กร

3 สิ่งที่เปลี่ยนโลก AI + Big Data + Data Science

- ★ AI, Big Data, Data Science มีความสัมพันธ์กันและถูกใช้ร่วมกัน เพื่อนำไปสู่ความเข้าใจข้อมูล และสามารถนำมาใช้ประโยชน์ได้จริง
 - ★ **Big Data** ประกอบด้วยข้อมูลที่มีโครงสร้างและไม่มีโครงสร้างจำนวนมากมหาศาลจากแหล่งที่มาต่างๆ เช่น โซเชียลมีเดีย เซ็นเซอร์ อุปกรณ์ ฯลฯ เพื่อใช้เป็นวัตถุดิบสำหรับงาน Data Science เช่น Data Mining การสกัดข้อมูลเชิงลึกออกมา หรือ Machine Learning การพัฒนาโมเดลให้มีความสามารถเฉพาะ อย่างเช่น Computer Vision, Natural Language Processing, Generative AI เป็นต้น

- ★ **Data Science** วิทยาการด้านข้อมูลที่เกี่ยวข้องกับการใช้วิธีการทางสถิติและการคำนวณ เพื่อสกัดข้อมูลเชิงลึกและความรู้ออกมาจาก Big Data ประกอบด้วยเครื่องมือ เทคนิค และเทคโนโลยีทั้งหมดที่ช่วยให้เราสามารถจัดการกับข้อมูลและใช้ประโยชน์ในด้านต่างๆ
- ★ **AI** เป็นการสร้าง Machine/Computer อัจฉริยะที่เลียนแบบพฤติกรรมหรือการทำงานของมนุษย์ สร้างกระบวนการเรียนรู้เพื่อให้สามารถทำงานแทนมนุษย์ได้ มีเทคนิคต่างๆ ที่ใช้สร้าง AI เช่น Machine Learning, Deep Learning เพื่อสร้างความสามารถต่างๆ ที่ใช้ประโยชน์ได้ในโลกจริง เช่น การใช้เหตุผล การตัดสินใจ พยากรณ์ จัดจำ จัดกลุ่ม และสื่อสารกับมนุษย์



แผนภาพที่แสดงให้เห็นความสัมพันธ์ของ AI, Big Data และ Data Science

- ★ เมื่อวิทยาการ Data Science เติบโต จึงมีความต้องการบุคลากรที่มี AI Skills เช่น Coding AI, Data & Analytics Tools, Prompt Engineering และทักษะอื่นๆ ที่จำเป็นสำหรับการสกัดข้อมูลเชิงลึก สร้างโมเดลที่เรียนรู้จากข้อมูล และสร้าง AI Applications เช่น ChatGPT เป็น Generative Chatbot ที่ผู้ใช้สนทนากับ AI ได้ด้วยภาษามนุษย์ หรือ MidJourney เป็น Generative Art ที่สร้างภาพจากคีย์เวิร์ดที่เราป้อน



“DATA IS THE NEW OIL”

Clive Humby

World's Data Pioneers, Co-founder of Dunnhumby

เรากำลังอยู่ในยุคทองของ Data จึงถูกเปรียบเปรยว่า “ข้อมูลมีค่าดั่งน้ำมัน”

- ★ ในปี ค.ศ. 2006 Clive Robert Humby ผู้บัญญัติวลีอันโด่งดังว่า “Data is the new oil” ซึ่งเปรียบเหมือนกับการปฏิวัติอุตสาหกรรมการผลิตน้ำมัน ที่มีความท้าทายอย่างมากในกระบวนการกลั่นข้อมูลเพื่อประโยชน์ในการขับเคลื่อนนวัตกรรมทางธุรกิจ



Clive Humby คือใคร?

Clive Humby มีประสบการณ์มากกว่า 40 ปี ในด้าน Customer Data และเป็นผู้บุกเบิก Consumer Centric Insights และ Data Analytics เป็นหนึ่งในผู้ก่อตั้งบริษัท Dunnhumby ร่วมกับ Edwina Dunn ที่ดำเนินธุรกิจระดับโลกด้านข้อมูลเชิงลึกในการปฏิบัติการใช้ข้อมูลการทำธุรกรรม เป็นผู้นำด้านการทำ Consumer Data Science มี Headquarter อยู่ที่ประเทศอังกฤษ (และยังมีสาขาอยู่ที่ประเทศไทยด้วย) อยู่เบื้องหลัง Clubcard ของ Tesco และ My Kroger Plus

★ Michael Palmer นักการตลาดอาวุโส, ANA Drives Growth ขยายความเพิ่มเติมว่า



“Data is just like crude. It’s valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc., to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value.”

- ★ “ข้อมูล” เปรียบเหมือนกับ “น้ำมันดิบ” มันมีค่า แต่ถ้ายังไม่ผ่านการกลั่นก็ใช้ประโยชน์อะไรไม่ได้ จะต้องผ่านกระบวนการกลั่นเพื่อแยกเป็นน้ำมัน ก๊าซ สารปิโตรเคมี ฯลฯ เพื่อนำไปผลิตเป็นน้ำมันเชื้อเพลิง และปิโตรเคมีภัณฑ์ประเภทต่างๆ ที่มีคุณค่าแตกต่างกัน มาขับเคลื่อนกิจกรรมทางธุรกิจที่ก่อเกิดผลกำไร
- ★ ข้อมูลดิบจะผ่านกระบวนการจัดการข้อมูลและการวิเคราะห์เพื่อหาคุณค่าในชุดข้อมูลต่างๆ จนค้นพบข้อมูลสำคัญเชิงลึกที่สามารถนำไปใช้ประโยชน์ และสร้างคุณค่าให้กับธุรกิจได้จริง



สิ่งที่มีมูลค่าในอนาคตไม่ใช่ น้ำมันอีกต่อไป แต่เป็นข้อมูล

เครดิต : นิตยสาร The Economist ในปี ค.ศ. 2017

เครดิตภาพ : <http://bitly.ws/rVVw>

- ★ คำว่า “Big Data” ไม่ใช่เรื่องไกลตัวเราอีกต่อไป การเกิดขึ้นของ Big Data ทำให้องค์กรต่างๆ มีความต้องการใช้ประโยชน์จากข้อมูล ทั้งจากภายในและภายนอกองค์กร การวิเคราะห์เพื่อทำความเข้าใจกับชุดข้อมูลต่างๆ ในเชิงลึก
- ★ ข้อมูลเชิงลึกถูกนำมาใช้กำหนดแนวทางในการทำธุรกิจที่มีประสิทธิภาพมากขึ้น ช่วยสนับสนุนการตัดสินใจในเรื่องต่างๆ ได้อย่างแม่นยำ
- ★ ใน Note นี้เราจะพูดถึงยุคแห่งการเติบโตของข้อมูล ความหมายของคำว่า Big Data การทำ Data Analytics ซึ่งเป็นการสร้างคุณค่าให้กับ Big Data ที่จำเป็นต้องอาศัยตำแหน่งงานต่างๆ ในทีม Data รวมไปถึง Data Scientist

ยุคแห่งการเติบโตของข้อมูล

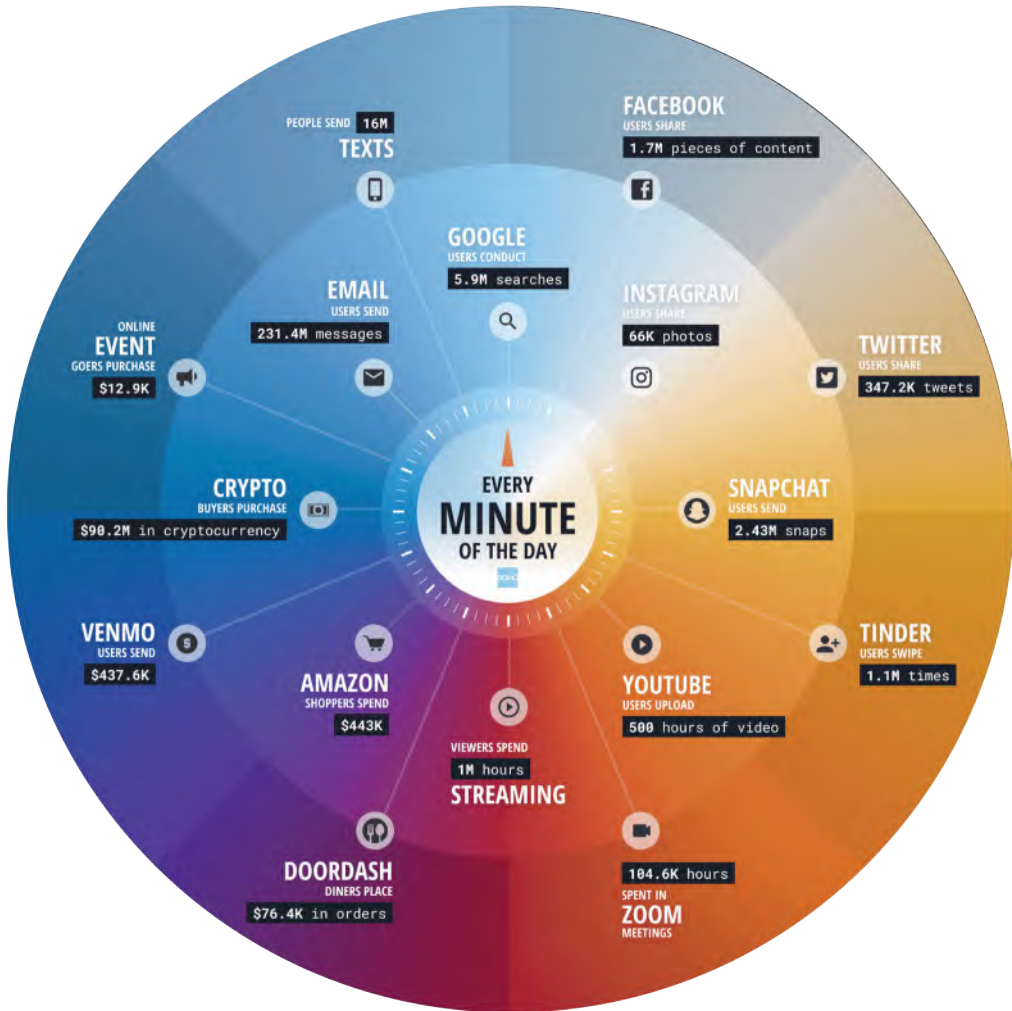
- ★ ในยุคปัจจุบัน พฤติกรรมของคนผูกพันกับโลกออนไลน์ นิยมใช้ Social Media Platforms ทำธุรกรรมต่างๆ เช่น มีการซื้อสินค้าผ่านระบบ E-Commerce มี Delivery Platforms บริการที่ช่วยให้เกิดการซื้อ-ขาย มีอุปกรณ์ไฟฟ้าในรูปแบบ Smart Home มี Smart Wearable ที่อำนวยความสะดวกให้ผู้สวมใส่ มีเทคโนโลยี IoT (Internet of Things) ควบคุมการทำงานผ่าน Mobile Applications ภายใต้ระบบเชื่อมโยงข้อมูลถึงกันบนเครือข่ายอินเทอร์เน็ตได้
- ★ การเข้าไปใช้งาน Platforms หรือ Applications เหล่านี้ เมื่อหลังจากจะมีการเก็บข้อมูลเพื่อบันทึกประวัติการใช้งานของผู้ใช้แต่ละรายไว้อย่างต่อเนื่อง เพื่อที่เจ้าของ Platforms จะได้นำข้อมูลเหล่านี้มาใช้วิเคราะห์หาข้อมูลเชิงลึก เพื่อประโยชน์ในการตัดสินใจคาดการณ์ หรือนำไปสร้างคุณค่าทางธุรกิจใหม่ๆ
- ★ ผลจากการที่มนุษย์เข้าไปใช้งาน Platforms ต่างๆ ที่เกิดขึ้นมากมาย
 - ★ Social Media : Facebook, Twitter, TikTok, Instagram
 - ★ Video Online Streaming : Netflix, YouTube
 - ★ E-commerce : Shopee, Lazada

- ★ อื่นๆ : อุปกรณ์ IoT (Internet of Things), IIoT หรือ Industrial IoT การเชื่อมต่อเครื่องจักรและระบบอุตสาหกรรมเข้ากับอินเทอร์เน็ต การเอาข้อมูลในกระบวนการผลิตมาใช้เพื่อที่จะปรับปรุงกระบวนการผลิตให้ดีขึ้น
- ✧ เนื่องจากความสามารถของเทคโนโลยีการประมวลผล การสื่อสาร และการจัดเก็บข้อมูลที่ดีขึ้น จึงนำไปสู่การเติบโตของข้อมูลอยู่ตลอดเวลา
- ✧ ในขณะที่เจ้าของ Platforms เองก็พยายามที่จะแข่งขันกัน เพื่อให้ผู้ใช้งานอยู่ใน Platforms ของตัวเองนานที่สุด เพราะรู้ว่าการที่ผู้ใช้งานมีส่วนร่วมกับ Platforms ของตัวเองมากจะเป็นเรื่องที่ดีในการสร้างความผูกพัน และยังเป็นผลพลอยได้ในการได้มาซึ่งข้อมูลอีกเป็นจำนวนมาก



ยุคแห่งการเติบโตของข้อมูล :
การใช้ชีวิตประจำวันของมนุษย์ส่งเสริมให้เกิดการสร้างข้อมูลอยู่ตลอดเวลา

- ★ เราอยู่ในยุคที่มีการสร้าง Data เป็นจำนวนมากในแต่ละวัน แผนภาพ EVERY MINUTE OF THE DAY ทำให้รู้ว่าในทุกๆ 1 นาที ผู้ใช้อินเทอร์เน็ตทั่วโลกทำอะไรบ้าง ในแต่ละปี ก็จะมี Platforms ใหม่ๆ ที่ได้รับความนิยมมากกว่าเข้ามาแทนที่ของเดิม



Data Never Sleeps 10.0 by Jace McLean, Director of Data Insights

ควาโอกาสจากข้อมูลที่แพร่หลาย

- ★ ข้อมูลที่มีอยู่มากมาย ทำให้ทุกๆ อุตสาหกรรมมุ่งไปที่การใช้ประโยชน์จากข้อมูลเพื่อ
ความได้เปรียบในการแข่งขัน (Competitive Advantage) ซึ่งเป็นความสามารถพิเศษ
ที่ยากจะเลียนแบบได้ หรืออาจต้องใช้เวลาอันยาวนานจึงจะตามทัน
- ★ ในอดีตบริษัทสามารถจ้างทีมนักสถิติ นักสร้างแบบจำลอง และนักวิเคราะห์ เพื่อ
สำรวจชุดข้อมูลด้วยตนเอง แต่ปัจจุบันปริมาณและความหลากหลายของข้อมูลนั้น
“มากเกินไป” จะวิเคราะห์ได้ด้วยตัวเอง
- ★ ขณะเดียวกันคอมพิวเตอร์เครือข่ายสื่อสารมีประสิทธิภาพสูงขึ้น และมีการเชื่อมโยง
ชุดข้อมูลเพื่อการวิเคราะห์ที่กว้าง (และลึก) กว่าเดิม
- ★ ปรากฏการณ์เหล่านี้ช่วยส่งเสริมให้มีการประยุกต์ใช้ Data Science / Data Mining
ในทางธุรกิจกันมากขึ้น



เมื่อโลกยุคใหม่นั้นมีข้อมูลเยอะเกินกว่าจะเข้าใจมันด้วยเครื่องมือเดิมๆ
ศาสตร์ Data Science & Tools ใหม่ๆ จึงถือกำเนิดขึ้นมาและมีบทบาทสำคัญ

- ★ สำหรับงานต่างๆ เช่น การตลาดที่ตรงกลุ่มเป้าหมาย การโฆษณาออนไลน์ และการทำ X-Selling (หรือ Cross-Selling) / Up-Selling
 - ★ X-Selling : การพยายามขายสินค้า Products อื่นๆ ให้ลูกค้า (ขายข้าม Products) เช่น ลูกค้าเงินฝากที่ธนาคารอาจจะเสนอให้ซื้อกองทุนหรือประกันชีวิต
 - ★ Up-Selling : การพยายามขายของ Products เดิมในปริมาณที่มากขึ้น
- ★ การปรับใช้ Data Science ในทางการตลาด มักใช้ในการบริหารจัดการความสัมพันธ์ลูกค้า หรือ CRM (Customer Relationship Management) เพื่อวิเคราะห์พฤติกรรมและเข้าใจลูกค้า นำไปสู่การสร้างแคมเปญมัดใจลูกค้า รักษาลูกค้าให้อยู่กับเรานานๆ
- ★ การปรับใช้ Data Science ในทางการเงิน มักใช้เพื่อสร้าง Credit Scoring เพื่อประเมินความเสี่ยงในการให้สินเชื่อลูกค้าแต่ละคน พิจารณาจากประวัติการใช้เครดิต และข้อมูลการเงินในอดีต โดยมี Machine Learning ทำงานช่วยพอร์ทอยู่เบื้องหลัง ในทาง Operations อาจใช้ในการตรวจจับการฉ้อโกง (Fraud) และการบริหารทรัพยากร
- ★ ผู้ค้าปลีกอย่าง Amazon และ Walmart มีการปรับใช้ Data Science ในภาคธุรกิจอย่างครอบคลุม ตั้งแต่การตลาด (Marketing) ไปจนถึงการจัดการห่วงโซ่อุปทาน (Supply Chain Management, SCM)



Products Rating Prediction



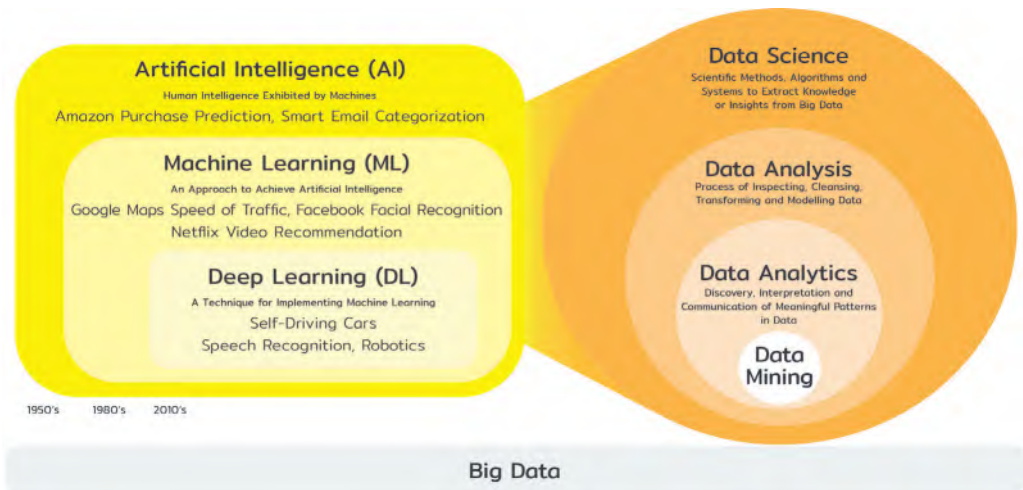
Demand Forecasting

**Amazon และ Walmart บริษัทเทคโนโลยีชั้นนำ
ที่มีการปรับใช้ Data Science ทุกภาคส่วนในองค์กร**



Credit Scoring หรือคะแนนเครดิต คือ ตัวชี้วัดความน่าจะเป็นในการคืนหนี้ของลูกค้าแต่ละราย อาจคำนวณจากปัจจัยต่างๆ โดยใช้วิธีการทางสถิติ (Statistical Methods) หรือการวิเคราะห์ขั้นสูง (Advanced Analytics) โดยใช้เทคนิคการสร้างโมเดลเพื่อคาดการณ์อนาคต

- ★ แต่ละบริษัทมีกลยุทธ์ด้าน Data Science ที่แตกต่างกัน บางบริษัทมองไปถึงการพัฒนาตัวเองไปเป็นบริษัท Data Science
- ★ การทำความเข้าใจปัญหาทางธุรกิจโดยใช้มุมมองจากข้อมูล เข้าใจหลักการและวิธีการดึงความรู้ที่เป็นประโยชน์จากข้อมูล มีแนวทางการคิดวิเคราะห์ข้อมูล และหลักการพื้นฐานที่ควรเข้าใจ นอกจากนี้ยังมีสิ่งที่ต้องใช้คือ สัญชาตญาณ ความคิดสร้างสรรค์ สามัญสำนึก และความรู้ในธุรกิจ
- ★ Data Science คือ หลักการพื้นฐานที่เป็นแนวทางในการสกัดความรู้จากข้อมูล ส่วน Data Mining คือ การใช้เทคโนโลยีเพื่อสกัดความรู้จากข้อมูล
- ★ คำว่า “Data Science” มักจะถูกนำไปใช้แพร่หลายมากกว่า “Data Mining”



Dr. Marcell Vollmer อธิบายความแตกต่างของ AI, ML, DL ร่วมกับ Data Science, Data Analysis & Analytics และ Data Mining สร้างภาพจำด้วยแผนภาพที่เรียบง่าย

เครดิตภาพ : <http://bitly.ws/rv2D>

Data Science Use Case 1 : Walmart สร้างโอกาสในวิกฤติเฮอริเคน Frances

คงจะไม่มีใครคาดเดาได้เลยว่า สินค้าอะไรใน Walmart จะขายดีก่อนเกิดพายุใหญ่
อะไรคือสิ่งที่ Walmart รู้เกี่ยวกับนิสัยของลูกค้า



กลยุทธ์เทคโนโลยีพลิกธุรกิจค้าปลีกด้วยแนวคิดที่ว่า
“หากเราเข้าใจการเปลี่ยนแปลงของสภาพอากาศ
และวางแผนสินค้าคงคลังอย่างเหมาะสม แล้วธุรกิจก็จะมีกำไร”

Walmart พัฒนาเทคโนโลยีคาดการณ์ยอดขายสินค้าล่วงหน้า
เพื่อสต็อกสินค้า Top Selling ในช่วงพายุเฮอริเคน

เครดิต : <http://bitly.ws/rV2G>



ผู้คนกำลังตัดสินใจตามสภาพอากาศมากขึ้น ส่วนหนึ่งเป็นผลมาจากการเพิ่มจำนวน
สมาร์ทโฟนและแอปพลิเคชัน เมื่อการพยากรณ์อากาศอยู่ใกล้แค่ปลายนิ้ว จึงทำให้ผู้คน
เช็คข้อมูลสภาพอากาศกันเพิ่มขึ้น นั้นสร้างผลกระทบโดยตรงต่อผู้ค้าปลีกอย่างมาก แต่
ก็ยังหา “โอกาส” ได้เช่นกัน สามารถศึกษารายละเอียดเพิ่มเติมได้ที่บทความ “What the
Weather Has in Stores” โดย IBM <http://bitly.ws/rVY6>

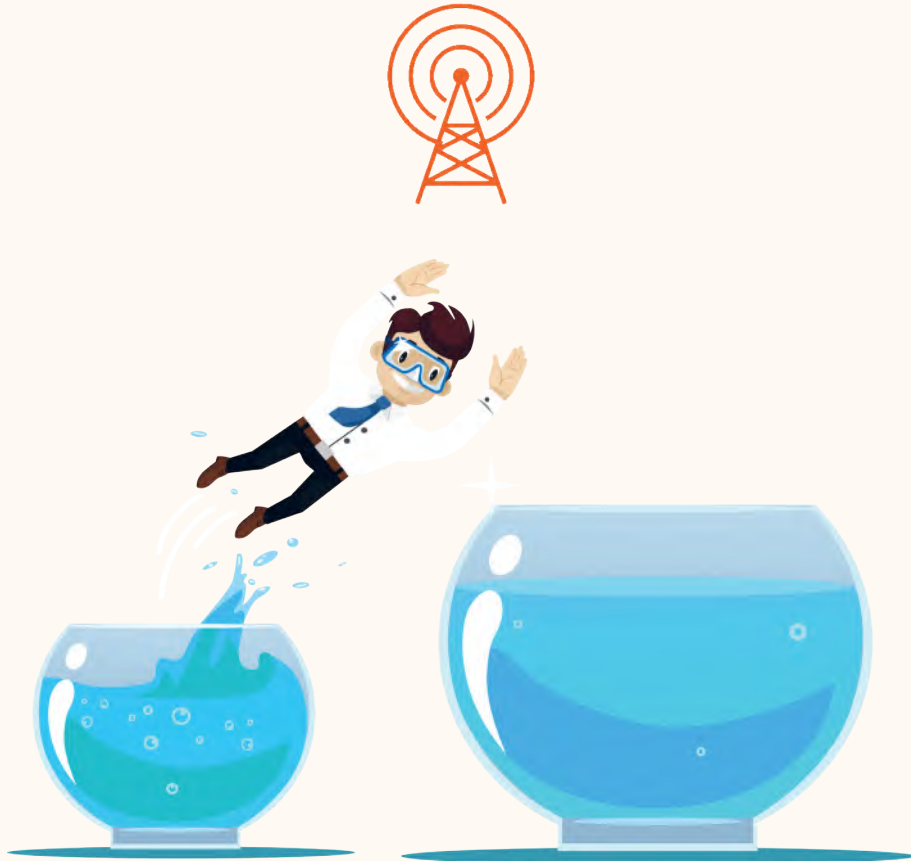
- ★ ในเหตุการณ์เฮอริเคน Frances บริษัทห้างค้าปลีกรายใหญ่ในสหรัฐอเมริกาอย่าง Walmart ได้ใช้ Predictive Technology ซึ่งเป็นเทคโนโลยีเพื่อการทำนายสินค้าที่น่าจะขายดี โดยดูจากข้อมูลประวัติย้อนหลังเมื่อเฮอริเคน Charley เข้าโจมตีเมื่อหลายสัปดาห์ก่อน (Walmart มีข้อมูลลูกค้าเป็นล้านๆ โบทันคลังข้อมูล)
- ★ การใช้ประโยชน์จากข้อมูล เช่น ทำนายว่าคนที่อยู่ในเส้นทางของเฮอริเคนน่าจะซื้อน้ำขวดเพิ่มขึ้น (อันนี้น่าจะชัดเจนอยู่แล้ว) หรือคาดการณ์ยอดขายที่เพิ่มขึ้นเพื่อให้แน่ใจว่า Walmart แต่ละสาขามีสินค้าอย่างเพียงพอ
- ★ นักวิเคราะห์ตรวจสอบข้อมูลประวัติยอดขายย้อนหลังของ Walmart เพื่อสำรวจว่าสินค้าใดมีความต้องการมากกว่าปกติ และได้ค้นพบข้อมูลที่ไม่เคยรู้มาก่อน เช่น ก่อนเกิดเฮอริเคน สินค้าขายดีคือ “Pop-Tarts รสสตรอว์เบอร์รี่” มียอดขายเพิ่มขึ้น 7 เท่า และสินค้าที่ขายดีที่สุดก่อนเกิดเฮอริเคนคือ “เบียร์”
- ★ Walmart ได้เรียนรู้ว่า Pop-Tarts รสสตรอว์เบอร์รี่เป็นหนึ่งในอาหารที่คนนิยมซื้อมากที่สุดทั้งก่อนและหลังเกิดพายุ เนื่องจากเป็นอาหารพร้อมทาน ทานได้ทุกมื้อ และเก็บรักษาได้นาน ไม่ต้องมีขั้นตอนปรุงสุก ไม่ต้องแช่เย็น
- ★ ด้วย Vision ของทีมบริหาร Walmart “มองวิกฤติเป็นโอกาส” เริ่มจากการย้อนดูข้อมูลประวัติการขายย้อนหลัง จนพบข้อมูลสำคัญที่นำไปสู่การสร้างโอกาสใหม่ ต่อยอดด้วยเทคโนโลยี Predictive นำมาใช้เพื่อพยากรณ์ยอดขายสินค้า Top Selling ทั้งก่อนและหลังพายุเฮอริเคน



ผู้บริหารจำเป็นต้องเข้าใจศักยภาพของ Data Science แม้ว่าจะไม่ได้ทำมันด้วยตัวเองก็ตาม การคิดวิเคราะห์ข้อมูลอย่างเป็นระบบจะช่วยให้การพิจารณาข้อเสนอโครงการ (Proposal) ทำได้อย่างเป็นระบบ และตัดสินใจว่ามีข้อบกพร่องหรือไม่?

Data Science Use Case 2 : การทำ Churn Prediction พูกลูกค้าก่อนย้ายค่าย

เราจะรู้ได้อย่างไรว่า ใครบ้างคือลูกค้าที่กำลังจะยกเลิกใช้บริการ (Churn)
Customer Churn Prediction in Telecommunication



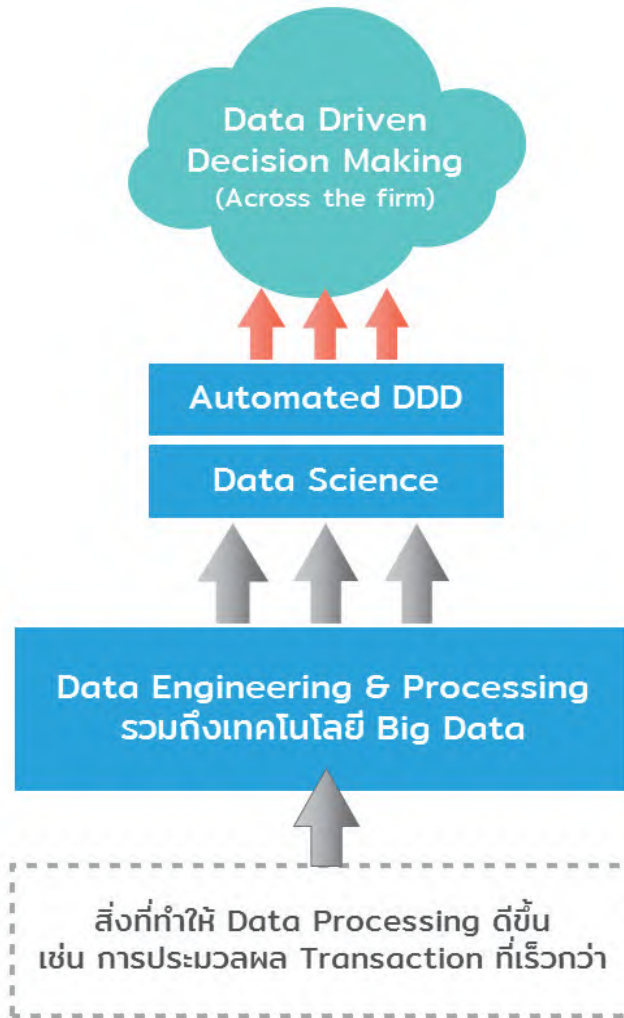
Churn Prediction เครื่องมือวิเคราะห์แนวโน้มการยกเลิกการใช้บริการของลูกค้า
เพื่อที่เราจะได้หาวิธีรับมือก่อนที่จะลูกค้าจะหนีไปหาคู่แข่งของคุณ

- ★ Mega-Telco คือ หนึ่งในบริษัทสื่อสารที่ใหญ่ที่สุดในสหรัฐอเมริกา กำลังประสบปัญหาในการรักษาลูกค้าเดิม พบว่ามีลูกค้าโทรศัพท์มือถือ 20% ยกเลิกสัญญาเมื่อสัญญาสิ้นสุดลง และการหาลูกค้ารายใหม่นั้นยากขึ้น เนื่องจากตลาดโทรศัพท์มือถืออยู่ในช่วงอิ่มตัว
- ★ การย้ายค่ายของลูกค้าจากค่ายหนึ่งไปยังอีกค่ายหนึ่ง เรียกว่า “Churn” หรือ “Customer Churn” (อาจหมายถึง หยุดซื้อ หยุดใช้บริการ เปลี่ยนใจ แล้วแต่บริบท)
- ★ ค่ายหนึ่งต้องสร้างแรงจูงใจเพื่อโน้มน้าวลูกค้าใหม่ เช่น โปรยค่ายเบอร์เดิม, โปรยค่ายเสนอส่วนลด 50% ในขณะที่อีกค่ายหนึ่งต้องสูญเสียรายได้เมื่อลูกค้าย้ายค่าย
- ★ สังเกตว่า “การหาลูกค้าใหม่” นั้นมีต้นทุนที่สูงกว่า “การรักษาลูกค้าเดิม” มาก ดังนั้น จึงควรจัดสรรงบประมาณการตลาดเพื่อป้องกันการเกิด Customer Churn จึงมีแคมเปญที่ออกแบบมาเพื่อมัดใจลูกค้ากลุ่มนี้โดยเฉพาะ
- ★ ภารกิจของ Data Science คือ การนำข้อมูลมาวิเคราะห์และทำนายลูกค้าที่มีโอกาสที่จะย้ายค่ายด้วย Churn Prediction โดยใช้ Machine Learning เพื่อมอบข้อเสนอพิเศษมัดใจก่อนที่จะสูญเสียลูกค้ากลุ่มนี้ไปเมื่อสัญญาสิ้นสุดลง
- ★ Mega-Telco ควรเลือกลูกค้าที่รับข้อเสนอพิเศษอย่างไร? เพื่อไม่ให้งบประมาณในการมัดใจลูกค้ากลุ่ม Customer Churn บานปลาย
- ★ โจทย์แบบนี้มีความซับซ้อนมากขึ้น จะอธิบาย Stage ต่างๆ ของการทำ Analytics ในส่วนถัดไป



การนำ Data Science มาใช้เพื่อรักษาลูกค้าไว้ (Retention) เป็นที่แพร่หลายอย่างมาก โดยเฉพาะในธุรกิจโทรคมนาคมและการเงิน

การตัดสินใจที่ขับเคลื่อนด้วยข้อมูลในทุกระดับงาน (Data-Driven Decision-Making : DDD)



Data Science ในบริบทของกระบวนการที่เกี่ยวข้องกับ Data ขององค์กร

เครดิต : Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking โดย Foster Provost และ Tom Fawcett

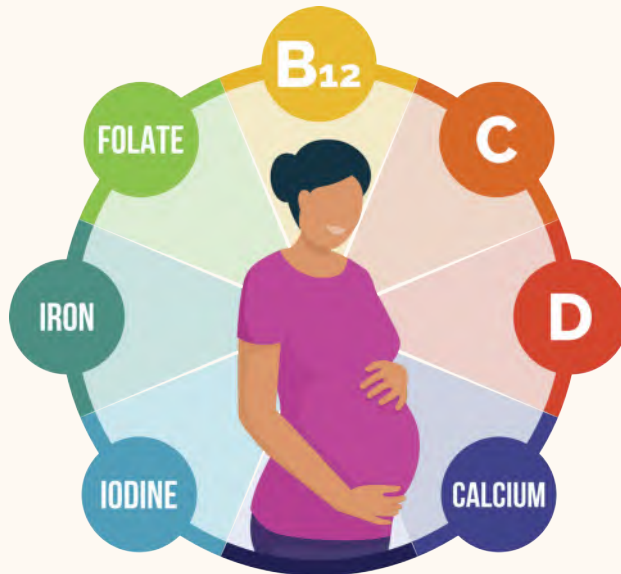
- ★ Data Science เกี่ยวข้องกับหลักการ / กระบวนการ / เทคนิค ในการทำความเข้าใจปรากฏการณ์ผ่านการวิเคราะห์ข้อมูล (แบบอัตโนมัติ) และเป้าหมายของ Data Science คือ ช่วยให้การตัดสินใจดีขึ้น (ในเรื่องที่ธุรกิจสนใจ)
- ★ Data-Driven Decision-Making (DDD) คือ การตัดสินใจบนพื้นฐานการวิเคราะห์ข้อมูล แทนที่จะใช้สัญชาตญาณอย่างเดียว เช่น นักการตลาดสามารถเลือกโฆษณาจากประสบการณ์ตัวเอง หรืออาจวิเคราะห์ข้อมูลการตอบสนองต่อโฆษณาที่ต่างกันของผู้บริโภค หรือสามารถใช้วิธีการทั้ง 2 ผสมผสานกัน
- ★ DDD ช่วยให้ประสิทธิภาพมากขึ้น ลดความสับสน ช่วยเพิ่มผลตอบแทนจากสินทรัพย์ ผลตอบแทนต่อผู้ถือหุ้น ประสิทธิภาพการใช้สินทรัพย์และมูลค่าตลาด

การประมวลผลข้อมูลและบิ๊กดาต้า (Data Processing & Big Data)

- ★ การประมวลผลข้อมูลไม่ถือเป็น Data Science แต่วิศวกรรม / การประมวลผลข้อมูลมีความสำคัญต่อการสนับสนุน Data Science
- ★ เทคโนโลยีการประมวลผลข้อมูลมีความสำคัญสำหรับธุรกิจที่มีข้อมูลจำนวนมาก แต่ไม่เกี่ยวกับการดึงความรู้ / การตัดสินใจ ที่ขับเคลื่อนด้วยข้อมูล (DDD) เช่น ประมวลผลธุรกรรม ประมวลผลระบบเว็บ การจัดการแคมเปญโฆษณาออนไลน์ อย่างมีประสิทธิภาพ
- ★ “Big Data” คือ ชุดข้อมูลที่มีขนาดใหญ่เกินกว่าที่จะใช้ระบบประมวลผลแบบดั้งเดิม จึงต้องใช้เทคโนโลยีการประมวลผลแบบใหม่ (เช่น Hadoop, HBase และ MongoDB)
- ★ จากการศึกษานักเศรษฐศาสตร์ Prasanna Tambe พบว่าเทคโนโลยี Big Data ช่วยเพิ่มผลิตภาพ หรือ Productivity และลดปัจจัยที่ทำให้เกิดความสับสนในองค์กรต่างๆ

Data Science Use Case 3 : “ลูกค้าคนไหนกำลังท้อง” รู้ก่อนเพื่อชิงความได้เปรียบ

ถ้าเราอยากรู้ว่าลูกค้าท้องหรือเปล่า แม้ว่าเธอจะไม่อยากบอกให้เราทราบ คุณทำได้ไหม?
Andrew Pole นักสถิติของธุรกิจค้าปลีก Target ถูกถามจากฝ่ายการตลาด



การวิเคราะห์เชิงคาดการณ์จากพฤติกรรมการจับจ่ายที่เปลี่ยนไป
และเป็นรายการที่เข้าข่ายหญิงตั้งครรภ์ แล้วนำมาสร้างโมเดลทำนายการตั้งครรภ์

เครดิต : How Companies Learn Your Secrets โดย Charles Duhigg,
New York Times, <http://bitly.ws/rW84>

**“We’re living through
a golden age of behavioral research”**

Eric Siegel

PhD, ผู้ก่อตั้ง Predictive Analytics World