

short Note



# The Data Scientist 's Journey

โลกของ AI & Data Scientist Roadmap ฉบับย่อที่อ่านง่ายสำหรับทุกคน



อาชีพไหนก็ไม่ Sexy  
เท่า Data Scientist



โดย ดร.นฤตม์ สุนทรานนท์  
Data Scientist และไอทีนินเฝ้า 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.นรุตม์ สุนทรานนท์
Editor	กิริพล คุษาเจริญ
Production Manager	วรพล ณิชกุล
Graphic Designer	ชวรินทร์ รัตนะ
Page Layout	สุรัสวดี วงศ์จันทร์สุข
Proofreader	พรรณรัตน์ ชุราณี
Publishing Coordinators	สุพัตรา อาจปรุ, 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](mailto:infopress@idcpremier.com) หรือทางโทรศัพท์หมายเลข 0-2962-1081 (อัตโนมัติ 10 คู่สาย) โทรสาร 0-2962-1084

## สร้างสรรค์โดย



พิมพ์ครั้งที่ 1 พฤษภาคม 2566

ข้อมูลทางบรรณานุกรม :

ดร.นรุตม์ สุนทรานนท์

Short Note The Data Scientist's Journey

นนทบุรี : ไอดีซีฯ, 2566

296 หน้า

1. วิศวกรรมซอฟต์แวร์

ชื่อเรื่อง

005.1

ISBN 978-616-487-438-1

ราคา 450 บาท

## พิมพ์และจัดจำหน่ายโดย



บริษัท ไอดีซี พรีเมียร์ จำกัด

200 หมู่ 4 ชั้น 19 ห้อง 1901

จัสมินอินเตอร์เนชั่นแนลทาวเวอร์

ถ.แจ้งวัฒนะ อ.ปากเกร็ด จ.นนทบุรี 11120

โทรศัพท์ 0-2962-1081 (อัตโนมัติ 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 ค้าปลีก 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 ในอุตสาหกรรมการเงินและค้าปลีก เพื่อให้ผู้อ่านได้เห็นภาพมากขึ้น นอกจากนี้ยังพูดถึง Soft Skills ที่จำเป็นสำหรับ Data Scientist เช่น ทักษะการสื่อสาร การเรียนรู้ด้วยตนเอง และปิดท้ายด้วยการเตรียมตัวสัมภาษณ์งาน โดยในแต่ละ Note จะมีการแนะนำแหล่งเรียนรู้เพิ่มเติมทั้งหนังสือและคอร์สออนไลน์

หวังเป็นอย่างยิ่งว่าหนังสือเล่มนี้จะเป็นไกด์ไลน์ให้กับทุกๆ คนที่อยากพัฒนาตนเองไปสู่ตำแหน่ง Data Scientist ได้มีแนวทางไปสู่จุดหมายได้ตามความตั้งใจ หรือผู้ที่อยู่ในสายงานนี้แล้วได้ขุดเวลาความรู้และทักษะต่างๆ ให้เฉียบคมมากยิ่งขึ้น หรือแม้แต่คนทั่วไปที่อยากเข้าใจโลกของ Data Science หนังสือเล่มนี้ก็ตอบใจทยัเช่นเดียวกัน

**ดร.นฤตม์ 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
- ✧ ผู้ร่วมก่อตั้งบริษัท 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 นอกจากนี้จะเข้าใจโลกของ Data Science และ Data Scientist Roadmap แล้ว ยังจะก้าวทันเทรนด์ AI ล่าสุดอย่าง Generative AI อีกด้วย

**“เรากำลังก้าวเข้าสู่ยุค ‘AI is Eating the World’ จงเลือกอาชีพที่โลกอนาคตต้องการ”**

**กัรวล คาเวริญ**

บรรณาธิการ

# CONTENTS

## PART 1 เส้นทางสู่อาชีพ DATA SCIENTIST

### ในโลกยุค AI .....10

#### NOTE 01 AI + BIG DATA + DATA SCIENCE

#### ขับเคลื่อนสู่โลกใหม่ ..... 11

#### NOTE 02 เข้าใจงานและแนวทางการเรียนรู้

#### เพื่อเป็น DATA SCIENTIST ..... 47

## PART 2 MATH & STATISTICS FOR DATA

### SCIENCE .....74

#### NOTE 03 พื้นฐานคณิตศาสตร์สำหรับ DATA SCIENCE.... 75

#### NOTE 04 พื้นฐานสถิติสำหรับ DATA SCIENCE ..... 101

## PART 3 PROGRAMMING LANGUAGES FOR

### DATA SCIENCE .....138

#### NOTE 05 ภาษา SQL & PYTHON สำหรับ

#### DATA SCIENCE ..... 139

## PART 4 การสร้าง AI ด้วย MACHINE LEARNING

### & DEEP LEARNING .....174

#### 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



# DATA SCIENTIST ROADMAP

A Quick Guide to Becoming a Data Scientist



**NOTE 01**  
Overview AI,  
Big Data,  
Data Science

**NOTE 02**  
Approaches to  
Learning Data  
Science

**NOTE 03**  
Mathematics  
for Data Science

**NOTE 04**  
Statistics for  
Data Science

**NOTE 05**  
Python and SQL  
for Data Science

**NOTE 06**  
Machine  
Learning

**NOTE 07**  
Deep  
Learning

**NOTE 10**  
Self-Learning  
Skills

**NOTE 11**  
Data Science  
Interview  
Preparation

**NOTE 08**  
Business  
Cases

**NOTE 09**  
Communication  
& Storytelling  
Skills

# PART 1

## เส้นทางสู่อาชีพ DATA SCIENTIST ในโลกยุค AI





**NOTE  
01**

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



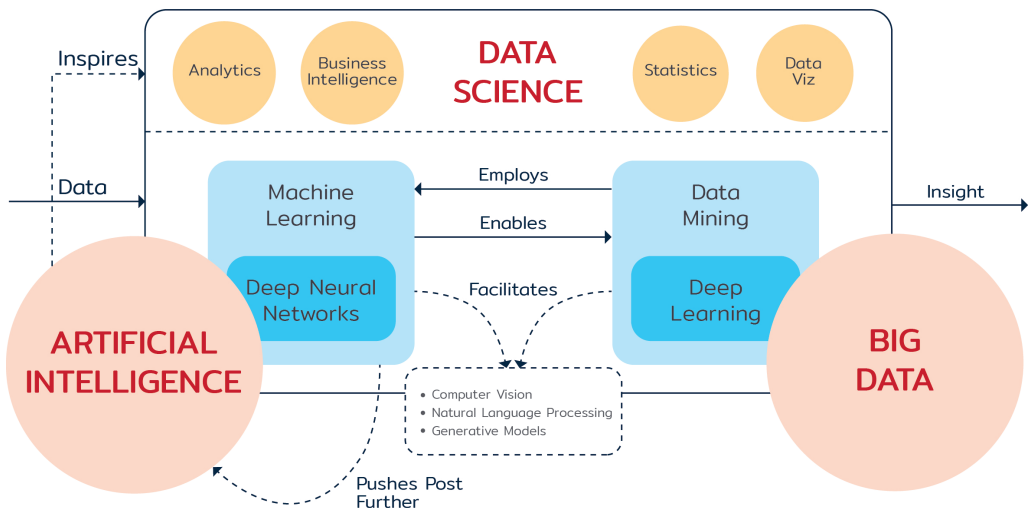
**“THE WORLD IS  
ONE BIG DATA PROBLEM.”**



**Andrew McAfee**

Co-director of the MIT initiative on the Digital Economy

- ★ **Data Science** วิทยาการด้านข้อมูลที่เกี่ยวข้องกับการใช้วิธีการทางสถิติและการคำนวณ เพื่อสกัดข้อมูลเชิงลึกและความรู้ออกมาจาก Big Data ประกอบด้วย เครื่องมือ เทคนิค และเทคโนโลยีทั้งหมดที่ช่วยให้เราสามารถจัดการกับข้อมูล และใช้ประโยชน์ในด้านต่างๆ
- ★ **AI** เป็นการสร้าง Machine / Computer อัจฉริยะที่เลียนแบบพฤติกรรมหรือการทำงานของมนุษย์ สร้างกระบวนการเรียนรู้เพื่อให้สามารถทำงานแทนมนุษย์ได้ มีเทคนิคต่างๆ ที่ใช้สร้าง AI เช่น Machine Learning, Deep Learning เพื่อสร้างความสามารถต่างๆ ที่ใช้ประโยชน์ได้ในโลกจริง เช่น การใช้เหตุผล การตัดสินใจ พยากรณ์ จดจำ 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

# การ Train Model และการนำไปใช้งาน

★ มีความแตกต่างระหว่าง

1. การทำ Data Science เพื่อหา Patterns และสร้าง Models (“Train / Build Model”)
2. การนำ Models ที่พัฒนาแล้วไปใช้งาน หรือ “Deployment” (ประมาณความน่าจะเป็นของข้อมูลใหม่จาก Model)

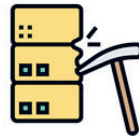
## 1. Train / Build Model

ข้อมูลในอดีต

Age	Gender	Income	Class
25	Male	50K	Yes
22	Male	20K	No
...			
50	Female	200K	Yes

ข้อมูล Train ที่ทราบ Class

Data Science



Model



## 2. Deployment

ข้อมูลใหม่

Age	Gender	Income	Class
30	Female	60K	?

ข้อมูลใหม่ที่ไม่ทราบ Class

Model



Class = Yes  
ความน่าจะเป็น 0.9

Deployment

การ Train Model (ภาพบน) และการนำ Model ไปใช้งาน (ภาพล่าง)

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

อธิบายรายละเอียดแต่ละขั้นตอนดังนี้

## 1. Business Understanding : การทำความเข้าใจ “ธุรกิจ”



**อะไรคือสิ่งที่ธุรกิจต้องการ?**

โครงการที่ดีควรเริ่มจากการทำความเข้าใจลูกค้าอย่างลึกซึ้ง การทำ Data Mining ก็เช่นกัน เห็นได้จากกระบวนการ CRISP-DM

การทำความเข้าใจธุรกิจ

1. กำหนดวัตถุประสงค์ทางธุรกิจ
2. ประเมินสถานการณ์
3. กำหนดเป้าหมายการทำ Data Mining
4. การจัดทำแผนโครงการ

- ★ สิ่งแรกจะต้องเข้าใจปัญหาที่จะแก้ไข ซึ่งอาจดูเหมือนจะชัดเจน แต่ในความเป็นจริงแล้ว การเปลี่ยนจาก “ปัญหาธุรกิจ” เป็น “ปัญหา Data Science” อาจมีความคลุมเครือ
- ★ การทำความเข้าใจธุรกิจเป็นงานที่ต้องใช้มุมมองของนักวิเคราะห์ธุรกิจ เพื่อเปลี่ยน “ปัญหาธุรกิจ” เป็น “ปัญหา Data Science (หรือ Data Mining)” ซึ่งอาจมีเพียง 1 ปัญหาหรือมากกว่า
- ★ กิจกรรมพื้นฐานการบริหารโครงการโดยส่วนใหญ่ตามหลักสากลมีดังนี้
  1. **กำหนดวัตถุประสงค์ทางธุรกิจ** : พยายามเข้าใจอย่างถึถ้วนจากมุมมองทางธุรกิจว่า “ลูกค้าต้องการอะไร” แล้วกำหนด “Business Success Criteria”
  2. **ประเมินสถานการณ์** : ตรวจสอบความพร้อมของทรัพยากร ข้อกำหนดของโครงการ ประเมินความเสี่ยงและแผนรองรับ และจัดทำ Cost-Benefit Analysis เพื่อประเมินความคุ้มค่า โดยเปรียบเทียบระหว่างประโยชน์ที่ได้รับและต้นทุนของโครงการ

# แนวทางการเรียนรู้เพื่อเป็น Data Scientist

Math และ Statistics		T		Note 03 และ 04
ภาษา SQL และ Python		T		Note 05
Machine / Deep Learning		T		Note 06 และ 07
Business Cases		T	S	Note 08
ทักษะด้านการสื่อสาร			S	Note 09
การเรียนรู้ด้วยตนเอง			S	Note 10
การสัมภาษณ์งาน		T	S	Note 11

แบ่งทักษะออกเป็น 2 ประเภท คือ Technical (Hard) และ Soft Skills

**T** Technical (Hard) Skills      **S** Soft Skills

## เส้นทางสู่การเป็น Data Scientist ในหนังสือเล่มนี้

- ★ Part 2 เรียนรู้เรื่อง Math & Statistics ซึ่งเป็นพื้นฐานสำคัญนำไปสู่เรื่องอื่นๆ ต่อไป
  - ★ Note 03 คณิตศาสตร์
  - ★ Note 04 สถิติ



# PART 2

## MATH & STATISTICS FOR DATA SCIENCE



- ✦ ตัวอย่างการนำเซตไปใช้แก้ปัญหาในชีวิตจริง เช่น ในการวิเคราะห์ข้อมูล (Data Analysis) เราต้องการทราบว่า ลูกค้าที่มีอายุน้อยกว่า 25 ปี ที่อาศัยอยู่ในกรุงเทพฯ มีการตอบสนองต่อ Campaign ที่ผ่านมามากหรือน้อยเพียงใด? เราสามารถเลือกใช้ตัวดำเนินการแบบ Intersect ได้ (ลูกค้าที่มีอายุน้อยกว่า 25 ปี  $\cap$  ลูกค้าที่อยู่ในกรุงเทพฯ)
- ✦ ในทางปฏิบัติ การใช้ภาษา SQL ในการ Query ข้อมูล (รายละเอียดใน Note 05) ก็ต้องใช้ความรู้เรื่องเซต เช่น Intersect, Union, Subtraction ฯลฯ เพื่อกรองข้อมูลเฉพาะที่ต้องการจากฐานข้อมูล โดยการเชื่อมโยงข้อมูลระหว่าง 2 ตาราง

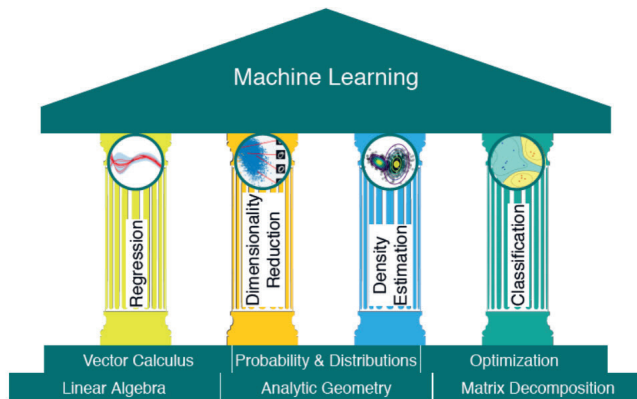
## พีชคณิตเชิงเส้น (LINEAR ALGEBRA) และการนำไปใช้แก้ปัญหาในชีวิตจริง

ทำไม “พีชคณิตเชิงเส้น” หรือ “Linear Algebra”  
จึงมีความสำคัญต่อ Machine Learning



Marc Deisenroth

DeepMind Chair of Machine Learning and Artificial Intelligence  
University College London



### Mathematical Foundations เป็นรากฐานและสี่เสาหลักของ Machine Learning

เครดิตภาพ : จากหนังสือ Mathematics for Machine Learning

โดย Marc Peter Deisenroth, A. Aldo Faisal และ Cheng Soon Ong

<https://deisenroth.cc>

## 🌟 ตัวอย่าง Python Source Code : Linear Regression

### # เรียกใช้ Numpy, Pandas, Matplotlib และ Scikit-Learn libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

### # Load ข้อมูล House Price ประกอบไปด้วยพื้นที่ (ตร.ว.) และราคา x 1,000 (บาท)

```
df = pd.read_csv('house_price.csv')
```

### # เก็บค่าพื้นที่ (input feature) ไว้ใน X และ ราคา (target) ไว้ใน y

```
X = df.Area
y = df.Price
```

### # ปรับรูปแบบให้เหมาะสมก่อน Train Model

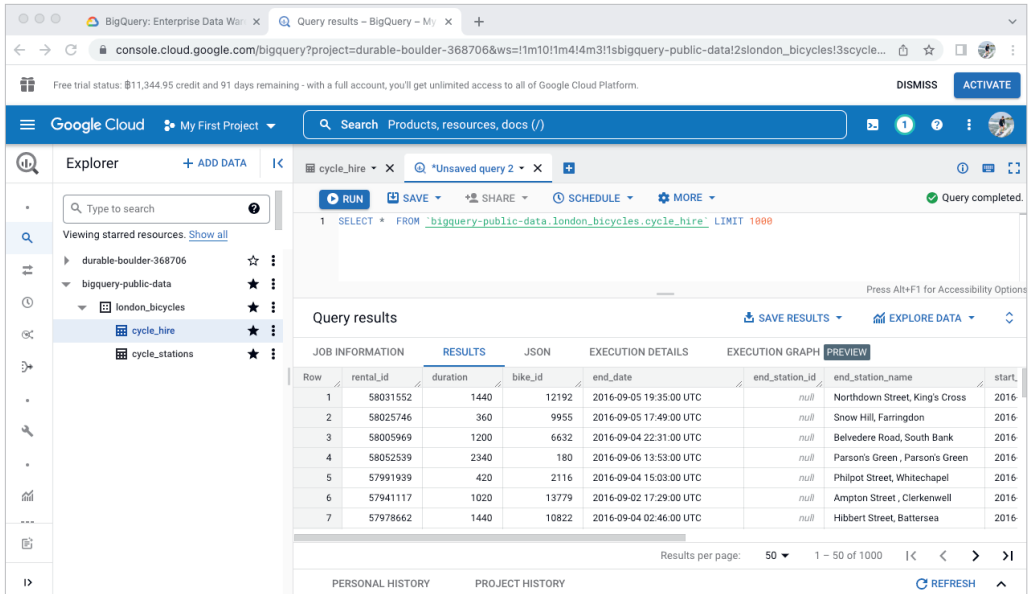
```
X = X.values.reshape(-1,1)
y = y.values.reshape(-1,1)
```

### # ทำการแสดงผลข้อมูล เพื่อทำความเข้าใจในเบื้องต้น

```
plt.scatter(X, y, s = 40, color='green', marker='D', alpha=.3,
label = 'Data')
plt.xlabel('Area (Sq.wa.)')
plt.ylabel('Price x 1,000')
plt.title('Area (Sq.wa) and Price x 1,000 Baht')
plt.show()
```

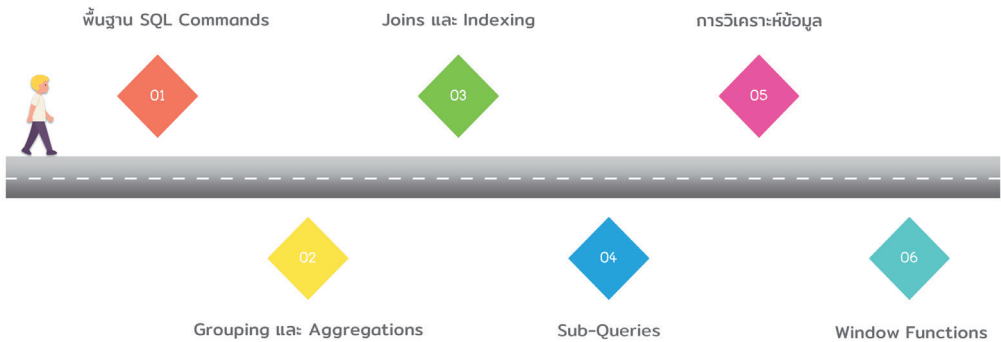
### # ทำการแปลงข้อมูลให้อยู่ใน Format ที่เหมาะสม

```
X = np.array(X).reshape(-1,1)
y = np.array(y).reshape(-1,1)
```



**ใน London Bicycle Hires จะมี 2 ตาราง แสดงการ Query ทุก Columns จำนวน 1,000 Rows ของตาราง "cycle\_hire"**

- ★ หรือกรณีผู้อ่านที่มีพื้นฐาน SQL แล้ว อาจทดลองกับ Dataset ที่มีความซับซ้อนมากขึ้น
- ★ แนวทางการเรียนรู้ SQL ของ Data Scientist อาจเริ่มจากพื้นฐาน SQL Commands, การทำ Grouping และ Aggregations, การ Join และ Indexing, การเขียน Sub-Queries, การวิเคราะห์ข้อมูล, การเขียน Window Functions



**แนวทางการเรียนรู้ SQL**

## AI แบ่งออกเป็น 3 ระดับตามความสามารถ



### Narrow AI

AI ที่ช่วยงานเฉพาะด้าน  
อย่างมีประสิทธิภาพ




### General AI

AI นำความรู้ที่เรียนรู้  
จากโดเมนหนึ่งไปใช้ในโดเมนอื่น



### Super AI

AI ที่ฉลาด  
กว่ามนุษย์หลายเท่า

 <p><b>A.I.</b></p> <p><i>"The theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages"</i></p> <p><b>OED</b></p>	AI Stages	<b>Artificial Narrow Intelligence (ANI)</b> Execute specific focused tasks, without ability to self-expand functionality	<b>Artificial General Intelligence (AGI)</b> Perform broad tasks, reason, and improve capabilities comparable to humans	<b>Artificial Super Intelligence (ASI)</b> Demonstrate intelligence beyond human capabilities
	Timing	Today	About 2040?	Soon after AGI
	Implications	Outperform humans in specific repetitive functions, such as driving, medical diagnosis and financial advice	Compete with humans across all endeavors, such as earning university degrees and convincing humans that it is human	Outperform humans, helping to achieve societal objectives or threatening human race
		<b>Jobs enhanced</b>	<b>Jobs at risk</b>	<b>Humanity at risk</b>

### 3 Stages ของ AI



#### Applied AI vs Narrow AI

Applied AI หมายถึง AI ที่มีความสามารถบางอย่างที่ทำได้ดีหรือดีกว่ามนุษย์ และนำความสามารถนั้นไปประยุกต์ใช้งานจริง เช่น การขับขี้อัตโนมัติ (Autonomous Driving), การจดจำใบหน้า (Face Recognition), การซื้อขายหุ้น (Stock Trading) ฯลฯ แต่มักถูกจัดอยู่ในกลุ่มเดียวกับ Narrow AI เนื่องจากการทำงานเฉพาะด้านเท่านั้น



2006

“Data is the New Oil”

Clive Humby ผู้ร่วมก่อตั้งบริษัท  
Dunnhumby

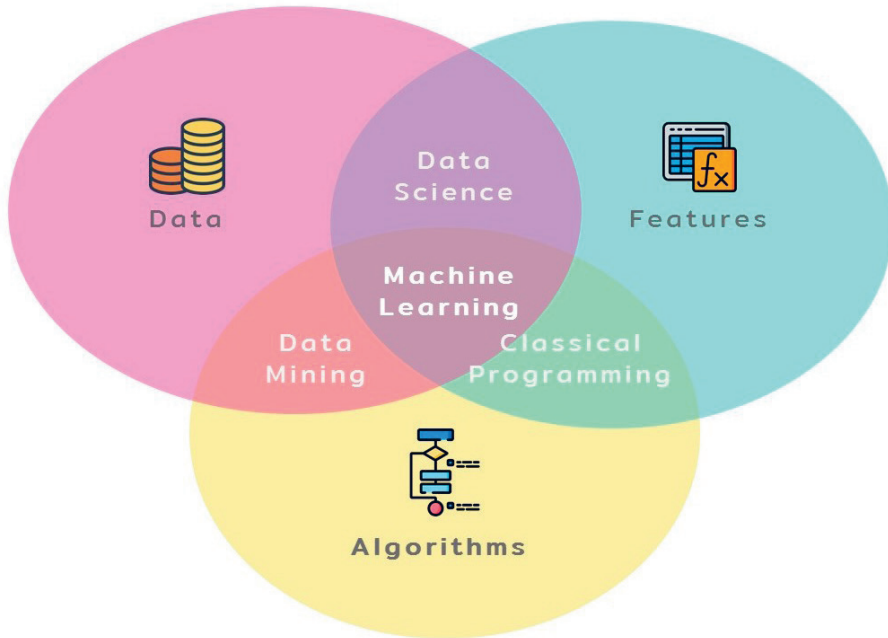


2016

“AI is the new Electricity”

Andrew Ng, Global Leader in AI

- ✪ ในปี ค.ศ. 2006 **Clive Robert Humby** หนึ่งในผู้บุกเบิก Data Analytics ผู้บัญญัติวลีอันโด่งดังไว้ว่า “Data is the new oil” ข้อมูลเปรียบเหมือนน้ำมัน มันมีค่า แต่ถ้าไม่แปรรูปก็นำไปใช้จริงไม่ได้ จึงนำน้ำมันดิบมาผ่านกระบวนการกลั่นเพื่อแปรรูปเป็นผลิตภัณฑ์ต่างๆ เช่น ก๊าซ น้ำมัน ปิโตรเคมี ฯลฯ ข้อมูลก็เหมือนกัน ต้องนำข้อมูลดิบไปผ่านกระบวนการจัดการข้อมูล และการวิเคราะห์ถึงคุณค่าในชุดข้อมูลต่างๆ จนค้นพบข้อมูลสำคัญเชิงลึกที่สามารถนำไปใช้ประโยชน์ได้จริง
- ✪ ในปี ค.ศ. 2016 **Andrew Ng** ผู้นำด้าน AI ระดับโลก ได้กล่าวเสริมว่า “AI is the new Electricity” ในที่นี้ Andrew เปรียบ AI เหมือนดังพลังงานไฟฟ้า มีข้อมูลอย่างเดียวไม่พอ ต้องมี AI เข้ามาใช้ในการประมวลผลเพื่อให้เกิดประโยชน์อย่างแท้จริง เปรียบได้กับพลังงานไฟฟ้าที่เราสามารถใช้ชาร์จโทรศัพท์ คอมพิวเตอร์ หรืออุปกรณ์ต่างๆ ที่อำนวยความสะดวกในการดำเนินชีวิตนั่นเอง หากยังจำสมการแบบง่ายๆ ที่กล่าวไว้ในตอนต้นว่า Big Data + Data Analytics = Value เช่นนี้ AI ก็เปรียบเสมือนเครื่องมือในการทำ Analytics เพื่อสร้างคุณค่านั้นเอง

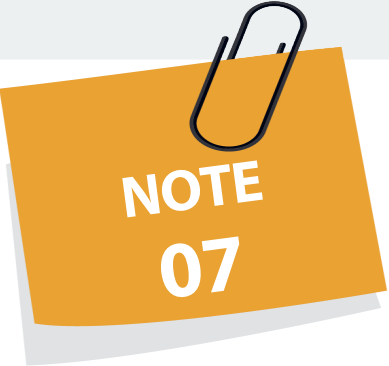


Venn Diagram สรุปองค์ประกอบของ ML

[https://vas3k.com/blog/machine\\_learning/](https://vas3k.com/blog/machine_learning/)


## ประเภทของ Classical Machine Learning

- ★ Machine Learning แบบดั้งเดิม (Classic) เป็นวิธีการที่มีรากฐานมาจากคณิตศาสตร์และสถิติ เช่น การค้นหารูปแบบจากตัวเลข การคำนวณระยะทางประเมินความใกล้เคียงของจุดข้อมูล การคำนวณทิศทางเวกเตอร์ ตัวอย่างกรณีศึกษาการนำไปใช้ประโยชน์ เช่น ระบบตรวจจับธุรกรรมที่ผิดปกติของ Bank หรือระบบแนะนำภาพยนตร์ของ Netflix
- ★ ธุรกิจเทคโนโลยีขนาดใหญ่นิยมใช้ Neural Network ด้วยเหตุผลว่า การเพิ่มความแม่นยำเพียง 1% อาจหมายถึงรายได้ที่เพิ่มขึ้น 1,000 ล้าน แต่อาจจะเป็นโครงการที่ใหญ่เกินไปสำหรับธุรกิจขนาดเล็ก จนมีเรื่องเล่าว่าเคยมีทีมหนึ่งใช้เวลา 1 ปี พัฒนาอัลกอริทึม Recommendation เพื่อใช้ในเว็บไซต์อีคอมเมิร์ซ (E-commerce) ก่อนจะพบว่าผู้เข้าชม 99% มาจาก Search Engine เวลาที่เสียไปจึงไร้ประโยชน์ เพราะผู้ใช้ส่วนใหญ่ไม่ได้เปิดหน้าเว็บไซต์หลัก



NOTE  
07

## DEEP LEARNING การเรียนรู้เชิงลึก



**“Machine intelligence is the last invention  
that humanity will ever need to make.”**



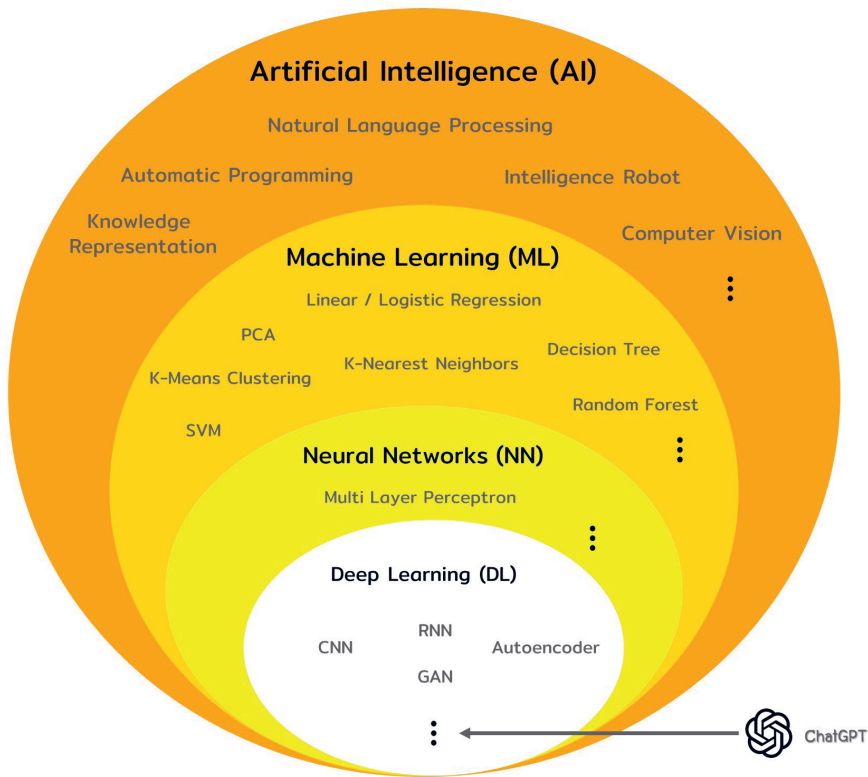
**Nick Bostrom**

Philosophy of Artificial Intelligence



# เกริ่นนำ

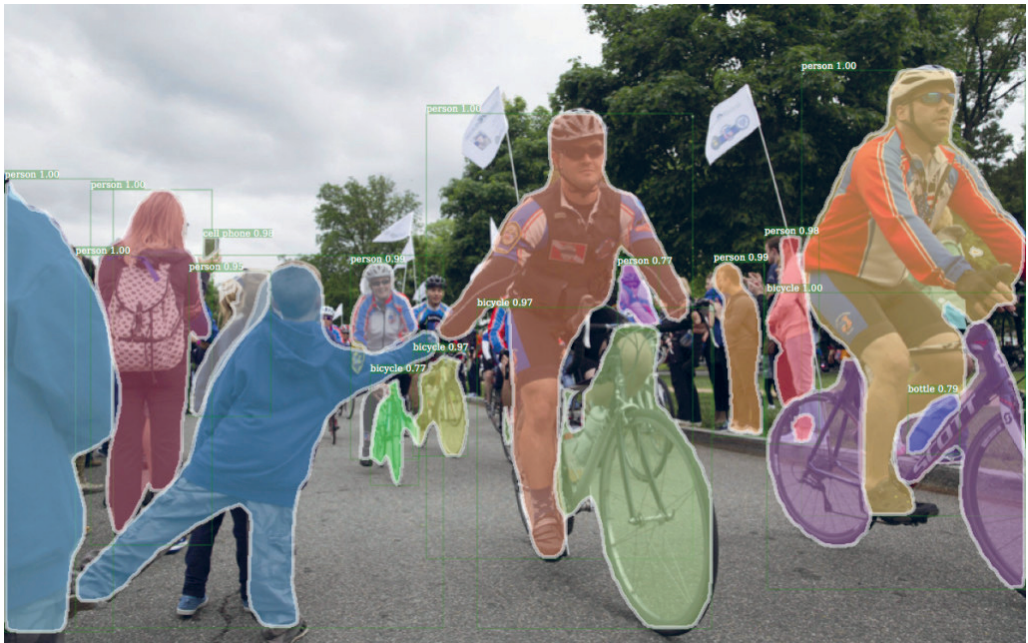
- ★ Note ที่แล้วได้พูดถึงประเภทของ Machine Learning (ML) และตัวอย่าง Algorithms ที่ได้รับความนิยม
- ★ สำหรับ Note นี้จะพูดถึง Machine Learning Algorithms ในกลุ่มของ Neural Networks ที่เรียกว่า Deep Learning (DL)
- ★ โดยจะยกตัวอย่าง Algorithms พื้นฐานที่ควรรู้ รวมถึงการประเมินประสิทธิภาพของโมเดล (Model Evaluation) ที่เหมาะจะนำไปใช้งานทั้ง ML และ DL Models
- ★ ท้ายสุดจะพูดถึง ChatGPT ซึ่งถูกพัฒนามานับพื้นฐานของ Deep Learning เป็น Natural Language Model ที่สามารถสร้างปฏิสัมพันธ์ และสนทนาในรูปแบบของข้อความกับมนุษย์ได้อย่างเป็นธรรมชาติ



ความสัมพันธ์ระหว่าง AI , ML และ DL

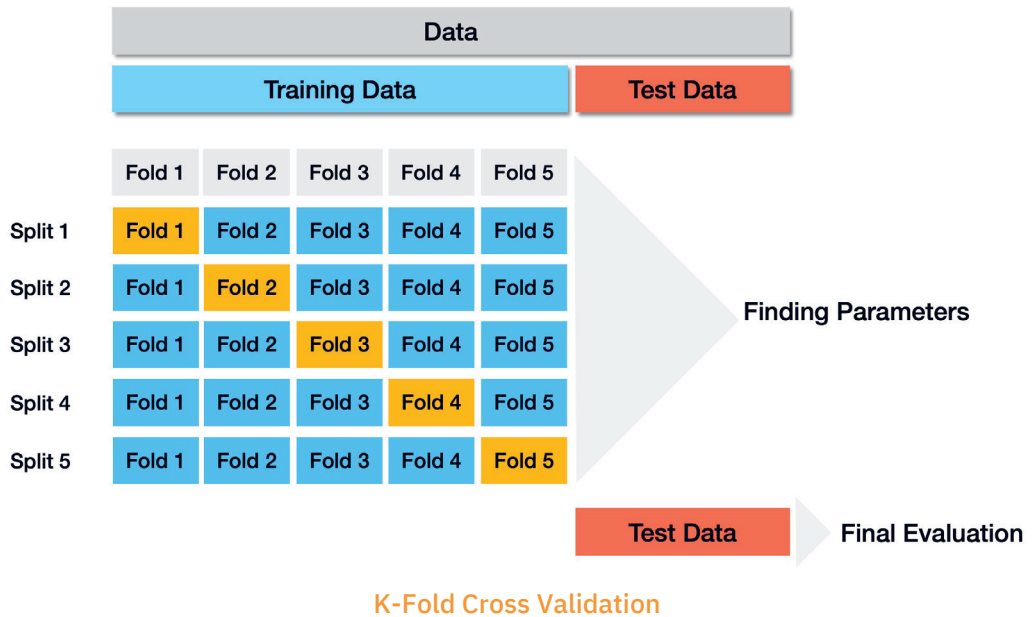
## Convolutional Neural Networks (CNN)

- ★ CNN ถูกใช้ในด้าน Computer Vision เช่น การค้นหาวัตถุบนรูปภาพ/วิดีโอ การจำแนกใบหน้า การถ่ายโอนสไตล์ การสร้าง/ปรับปรุงรูปภาพ การสร้าง Effects เช่น ภาพซ้ำ และการปรับปรุงคุณภาพของภาพ
- ★ ทุกวันนี้ CNN ถูกนำมาใช้กับข้อมูลที่เป็นรูปภาพและวิดีโอ แม้กระทั่งใน iPhone ของเรา
- ★ ในการทำงานกับรูปภาพ จำเป็นจะต้องดึง Features ออกมาเองด้วยมือเพื่อนำไปใช้ Train เช่น นี่คือหูและนี่คือหางของแมว
- ★ การกำหนด Features ด้วยมืออาจก่อให้เกิดปัญหา เช่น ถ้าเจอแมวสายพันธุ์หุ่บหรือแมวหันหลังให้กล้อง ก็จะมีปัญหาเพราะ NN จะไม่เห็นอะไรเลย

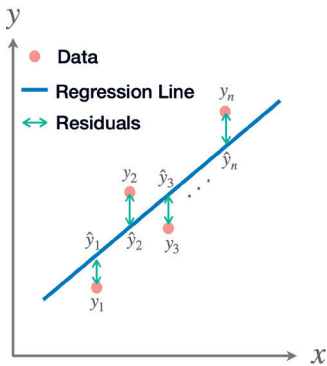


ตัวอย่างผลลัพธ์ของ Detectron ซึ่งเป็น Object Detection  
เครื่องมือที่พัฒนาโดย Meta (Facebook)

เครดิตภาพ : <https://ai.facebook.com/tools/detectron/>



★ หากมีข้อมูลจำนวนน้อย สามารถใช้วิธีการแบบ K-Fold Cross Validation ได้  
 ในภาพแสดงการแบ่งเป็น 5 Folds



$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

*MAE: Mean Absolute Error*

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

*MSE: Mean Squared Error*

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

*RMSE: Root Mean Squared Error*

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)}{\sum_{i=1}^N (y_i - \bar{y}_i)}$$

*R2 Score*

$\hat{y}$  : Predicted value of y  
 $\bar{y}$  : Average value of y

**Metrics สำหรับประเมินผล Regression Model**

- ✧ ในการประเมินประสิทธิภาพของ Model จะวัดที่ข้อมูล Test (Unseen Data)
- ✧ ตัวอย่าง Metrics ที่ใช้ในการวัดประสิทธิภาพของ Regression Model
  - ★ **MAE (Mean Absolute Error)** คำนวณได้โดยหาผลรวมของค่า Absolute (Error) เปรียบเทียบระหว่างค่าที่ทำนาย (Predicted Value) และค่าที่เกิดขึ้นจริง (Actual Value) แล้วหารด้วย N เพื่อเปลี่ยนเป็นค่าเฉลี่ย
  - ★ **MSE (Mean Square Error)** คำนวณได้โดยหาผลรวมของค่า Square (Error) เปรียบเทียบระหว่างค่าที่ทำนาย (Predicted Value) และค่าที่เกิดขึ้นจริง (Actual Value) แล้วหารด้วย N เพื่อเปลี่ยนเป็นค่าเฉลี่ย
  - ★ **RMSE (Root Mean Square Error)** ปรับปรุงจากค่า MSE โดยใส่ Square Root เพื่อให้แปลความหมายได้ดียิ่งขึ้น เป็น Metrics ที่นิยมใช้
  - ★ **R<sup>2</sup> Score** เป็นอีกหนึ่ง Metric ที่นิยมใช้ในการประเมินประสิทธิภาพ Regression Model มีค่าอยู่ระหว่าง 0 (แย่มาก) ถึง 1 (ดีที่สุด)

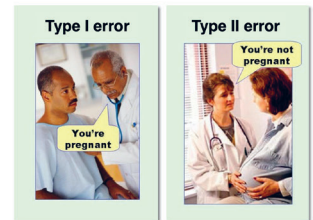
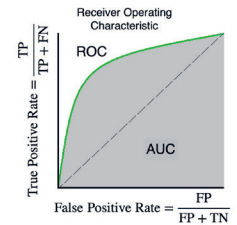
		Predicted Value	
		Positive	Negative
Actual Value	Positive	TP True Positive	FN False Negative <small>Type II error</small>
	Negative	FP False Positive <small>Type I error</small>	TN True Negative

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$


$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$




### Metrics สำหรับประเมินผล Classification Model

## Generative AI Chatbot : ChatGPT



- ★ ในตัวอย่าง พิมพ์ข้อความที่ Prompts ว่า “How to become a data scientist” AI ก็จะหาคำตอบมาให้เรา สรุปข้อมูลมาให้ด้วยภาษาที่เรียบเรียงมาอย่างดีโดยที่เราไม่ต้องหาข้อมูลเอง ผู้สนใจสามารถทดลองใช้งานได้ที่ <https://chat.openai.com/chat>



How to become a data scientist



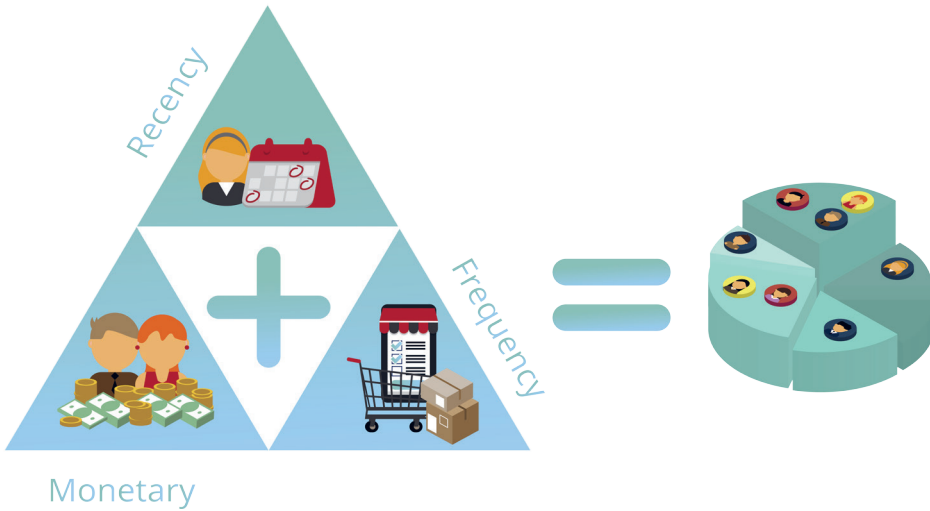
Becoming a data scientist typically requires a combination of education, experience, and skills in various fields such as statistics, mathematics, computer science, and domain knowledge. Here are some steps that can help you become a data scientist:

1. Education: A bachelor's or master's degree in a related field like computer science, statistics, mathematics, engineering, economics, or physics is usually required. You may also pursue a data science-specific degree or certification program.
2. Gain experience with programming languages: Data scientists must be proficient in programming languages such as Python, R, SQL, and others. Learning to use these languages will help you to manipulate and analyze data.
3. Acquire statistical knowledge: You should have a strong understanding of statistics and probability theory to be a good data scientist.
4. Develop domain knowledge: You should have domain knowledge in the industry that you are applying data science to. It will help you better understand the data that you work with.
5. Learn data visualization tools: Data visualization tools such as Tableau and PowerBI are necessary for communicating your findings to others.
6. Build a portfolio: Create a portfolio of projects that you have worked on. Showcasing your data science skills can help you stand out in the job market.
7. Network: Attend meetups, conferences, and connect with other data scientists. This can help you learn from others and stay up-to-date with the latest developments in the field.
8. Apply for internships: Internships are a great way to gain hands-on experience in the field and make valuable connections.
9. Apply for jobs: Apply for data scientist positions or related roles such as data analyst or machine learning engineer.

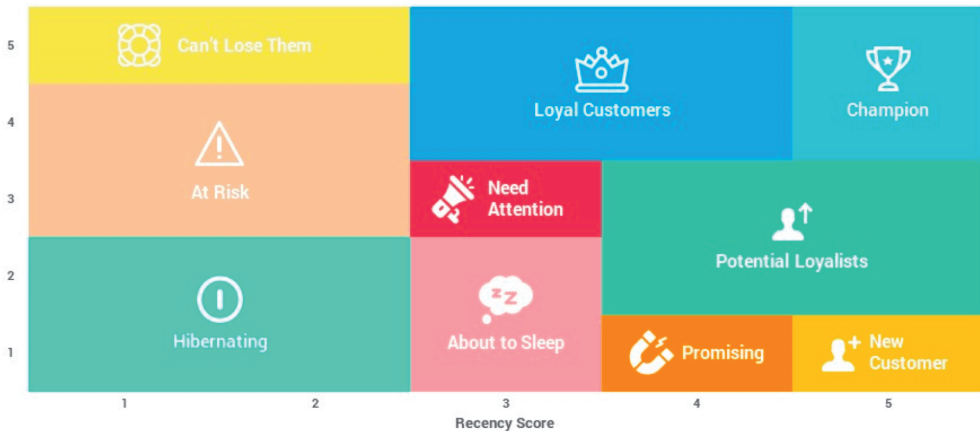
Overall, becoming a data scientist requires dedication and a willingness to continually learn and adapt to new technologies and methods.

**ตัวอย่างการใช้งาน ChatGPT และผลลัพธ์เมื่อใส่ข้อความว่า  
“How to become a data scientist” (จะเป็นนักวิทยาศาสตร์ข้อมูลได้อย่างไร)**



## RFM Analysis Model - RFM Summary Report!

RFM Analysis Model - RFM Summary Report!



RFM Analysis ใช้ในการแบ่งลูกค้าเพื่อสร้างแคมเปญ (Campaign)  
หรือรักษาสถานลูกค้า (Retention)

เครดิตภาพ : <https://www.annexcloud.com/blog/revolutionising-segmentation-individualisation-using-rfm-to-step-further/>



เครื่องมือที่ใช้แบ่งกลุ่มลูกค้าอาจแบ่งกลุ่มมาให้จำนวนมาก เราไม่จำเป็นต้องทำการตลาดกับทุกกลุ่มเป้าหมายให้เลือกทำการตลาดกับกลุ่มลูกค้าตามความเหมาะสม

เข้าใจผู้ฟัง



01

เตรียม ฝึกฝน



03

นำเสนองาน



05

02



ทำ Pitch Deck

04



ตรวจการตั้งค่า

06



มีส่วนร่วมกับผู้ฟัง

### แนวทางการนำเสนอ Data Science Project

- ★ **ขั้นที่ 1** ทำความเข้าใจผู้ฟัง และเป้าหมายของผู้ฟัง

  - ★ กำลังนำเสนอใคร ตำแหน่งอะไร ทำไมถึงเข้าร่วม
  - ★ ระบุคำถามที่ผู้ฟังต้องการคำตอบ
  - ★ เสี่ยงศัพท์ทางเทคนิค หากผู้ฟังไม่เข้าใจ
- ★ **ขั้นที่ 2** ทำ Pitch Deck (Presentation) ที่น่าสนใจ

  - ★ ทำสไลด์ที่เรียบง่าย มีหัวเรื่องที่ชัดเจน
  - ★ ใช้รูปภาพและกราฟ เพื่อนำเสนอข้อมูลเชิงลึกที่ค้นพบ
  - ★ ใช้ข้อความบนสไลด์ให้น้อยที่สุด (น้อยกว่า 15 คำ)
- ★ **ขั้นที่ 3** เตรียม ฝึกฝน และปรับปรุงการนำเสนอให้ดีขึ้น

  - ★ เตรียมสคริปต์คร่าวๆ สำหรับสิ่งที่จะพูด
  - ★ ส่งงานนำเสนอให้เพื่อน / เพื่อนร่วมงาน และขอคำติชม / ความคิดเห็น
  - ★ แก้ไขข้อผิดพลาด ปรับปรุงสไลด์ให้ชัดเจนและกระชับ

- ✦ การไปสู่อาชีพ Data Scientist อาจมีได้หลายเส้นทาง ในที่นี้จะแสดงตัวอย่างหนึ่งของ Data Scientist Roadmap หรือเส้นทางการเรียนรู้เพื่อก้าวสู่โลกแห่งอาชีพ Data Scientist

- ★ เรียนรู้พื้นฐาน SQL และ Python

- ★ พื้นฐานคณิตศาสตร์และสถิติ

Fundamental Level

- ★ เรียนรู้ SQL และ Python ในระดับกลางและสูง

- ★ การทำ Data Visualization

- ★ เรียนรู้ Machine Learning Algorithms

- ★ การทำ Data Manipulation

- ★ เรียนรู้ Model Deployment

Intermediate Level

- ★ เรียนรู้ Deep Learning Algorithms

Advanced Level

- ★ การทำ Projects และ Use Cases ต่างๆ (Data Science Portfolio)

Data Science Projects

- ★ การเตรียมตัวสัมภาษณ์งาน

- ★ ก้าวแรกสู่อาชีพ Data Scientist

First Step in Career World





# 18 BEST DATA SCIENCE PODCASTS

Data Skeptic	Data Science Ethics	Data Science at Home	Data Crunch	SuperData Science	Data Cast
Data Science Imposters	Women in Data Science	Linear Digressions	Data Futurology	Data Science in Production	Cyentia Podcast
Talk Python To Me	O'Reilly Radar - The O'Reilly Data Show Podcast	The AI Podcast	DATAHACK RADIO	Half Stack Data Science	This Week in Machine Learning & AI (TWIML & AI)

**datasciencedojo**  
data science for everyone

© Copyrights Reserved. <https://datasciencedojo.com>

## Podcast ยอดนิยมนในเรื่อง Data Science

เครดิตภาพ : <https://datasciencedojo.com/blog/best-data-science-podcasts/>