

### การพัฒนา Al จัดกลุ่มงานวิจัยด้วย Python

เรียนรู้และฝึกปฏิบัติการสร้าง Model Classification ด้วย Al เพื่อจัดกลุ่มงานวิจัยอย่างมี ประสิทธิภาพ เริ่มต้นจากพื้นฐาน Python ด้วยเครื่องมือ Google Colab เพื่อจัดกลุ่มผล งานวิจัยในระบบ TNRR ตามสาขาวิชาการ OECD

ผศ.ดร.วศิน ศรีสวัสดิ์ อาจารย์ประจำ สาขาวิชาเทคโนโลยีการจัดการระบบสารสนเทศ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยมหิดล

### ทำไมต้อง Al และ Classification?

Al คืออะไร?

Al หรือปัญญาประดิษฐ์ คือการทำให้ คอมพิวเตอร์คิดและเรียนรู้ได้เหมือนมนุษย์ Classification คืออะไร?

Classification คือการแบ่งข้อมูลออกเป็นกลุ่ม ต่างๆ ตามคุณสมบัติที่กำหนด

ทำไมต้อง Classification?

ใช้ในการตัดสินใจ คาดการณ์ และแก้ปัญหา ต่างๆ ได้อย่างมีประสิทธิภาพ

```
Weath dr
 is cottoun; aracing is tariaatary
fif betichnangter";
    ff / tlátw ressero**5, law
      ar > tilp:
                      acture
       thery /a taly";
        conch ipt, is can dot in the this area.
           traw t,/ "luast law"( the
           ff if snot (istum")
           mying)
       camer al ing"; larn(
         > firthery ont, "Lianctor"
           cotl enocunrer")
        tlartt, "riesslt")
```

### พื้นฐานภาษา Python ที่ควรรู้

- 2 Control Flow
  เรียนรู้การควบคุมการทำงานของ
  โปรแกรมด้วย if, else, for, while
- 3 ฟ**ังก์ชัน** สร้างและเรียกใช้ฟังก์ชันเพื่อจัดการโค้ดให้เป็นระเบียบ

# รู้จักกับ Google Colaboratory (Colab)





Colab คือ Cloud IDE ที่ให้ คุณเขียนและรันโค้ด Python ได้ฟรี



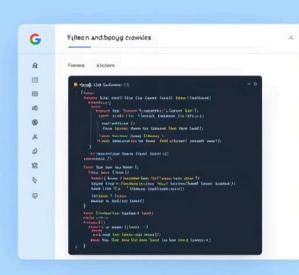
Collaboration

แชร์โค้ดและทำงานร่วมกับ ผู้อื่นได้อย่างง่ายดาย



Free Resources

ใช้งาน GPU และ TPU ฟรี สำหรับการ Train Model













## Library ที่จำเป็น

NumPy

จัดการข้อมูลตัวเลขและ Array อย่างมี ประสิทธิภาพ Pandas

จัดการข้อมูลในรูปแบบตาราง (Data frame) Scikit-learn

Library หลักสำหรับการสร้าง Machine Learning Model

AttaCut

Library สำหรับตัดคำภาษาไทย

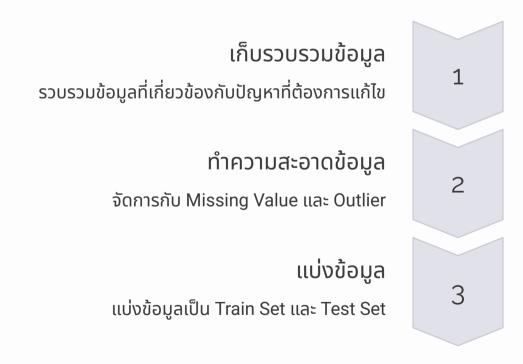
Joblib & Pickle

ใช้ในการบันทึกและโหลด Model



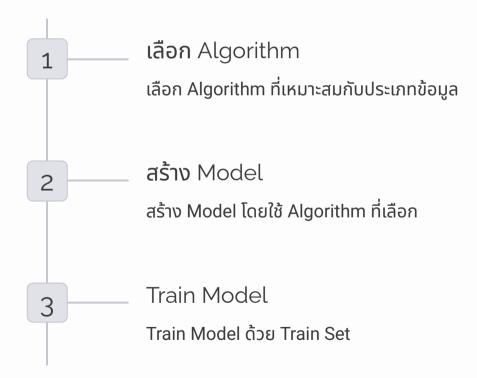
### **Data Preparation Process** Dant & Craot Sation Hin Wind Coller hates The clear milato fry annemotin your saled by sings for foundiers Com Ferel Chan For the Data Tengentings Day Tengtum enntad Com Next fee tale Peenter Englication Petdiet Cliet or in thme ruostiles out on the second rated imp. I link the lafers Cailness Sel lection Unabled po the bring trades their Donalest Led Fley Galine Beskely Nigt Lne twiscle & Fried Basing Boar Monn For Ting Lascion Franciscon In parts Michael a part supplier to project of Control law Eachtern Amai area ulas sadorecho follower and the anders nd title fum has hid incomme Freiding Fieg blattery fifty Fews breates tomo Verer traves stara legal Feargonair tile of does February I pervolu nexts been prener direction for nord treat it or pa ost lear tent opera for Son en não a oma Emi Am Marintino And My Carrie For Dee Prepareley Loste

### เตรียมข้อมูลสำหรับการ Classification



# 35.5%

### สร้าง Model Classification ด้วย scikit-learn



### Metrics สำหรับประเมินประสิทธิภาพ Model

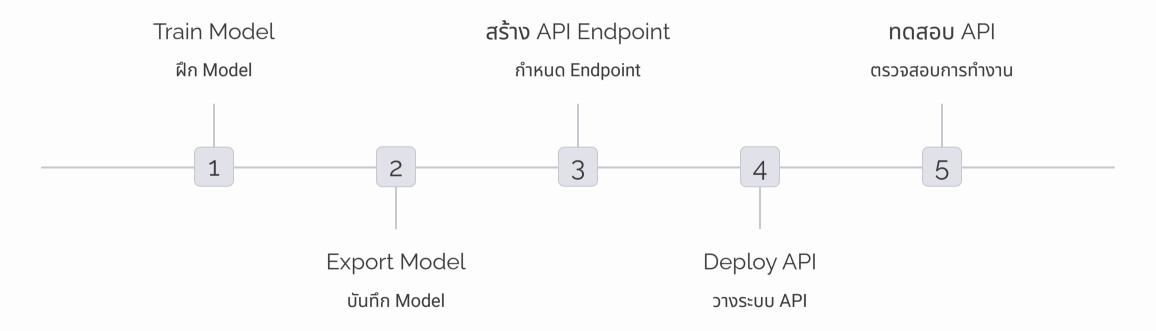
1 2

**ค่า** Recall **ค่า** Precision วัดความสามารถในการหาผลลัพธ์ที่เป็นจริง วัดความแม่นยำของผลลัพธ์ที่ทำนาย

ค่าเฉลี่ย Harmonic ของ Recall และ Precision

F1-Score

### Export Model เพื่อใช้งานเป็น API



### การนำ Model ไปใช้ในระบบ TNRR

Model Classification ที่สร้างขึ้น สามารถนำไปประยุกต์ใช้ในระบบ Thai National Research Repository (TNRR) เพื่อจัดกลุ่มผลงานวิจัยตามสาขา วิชาการ OECD ได้อย่างมีประสิทธิภาพ

- จัดกลุ่มผลงานวิจัยอัตโนมัติช่วยลดภาระงานในการจัดกลุ่มผลงานวิจัยด้วยตนเอง
- เพิ่มความแม่นยำในการจัดกลุ่ม ผลงานวิจัยลดความผิดพลาดจากการตัดสินใจของ
  - ลดความผดพลาดจากการตดสนเจของ มนุษย์

สนับสนุนการวิเคราะห์ข้อมูล ทำให้สามารถวิเคราะห์ข้อมูลผลงานวิจัย ตามสาขาวิชาการได้อย่างรวดเร็ว