

ปี 2563



พฤติกรรมการใช้อินเทอร์เน็ตใน ประเทศไทย

Data Mining

Presented by ได้นมดและจะดันรอ

THE TEAM



1. นายปิยพัทธ์ ปานะถึก 643020507-4
2. นางสาวพัฒน์นิศา ทองบ่อ 643020508-2
3. นางสาวนิมชนก วงศ์สายเชื้อ 643020510-5
4. นายภูริศ เครือชาไร่ 643020514-7
5. นางสาววิภาดา น่วงสูงเนิน 643020520-2
6. นางสาวสิรภัทร ไชยมาตย์ 643020523-6
7. นางสาวนุททัยชนก สรวงชัยภูมิ 643020525-2
8. นายอาภาวุธ จรุงรักษ์ 643020528-6
9. นายธนพร ก้านกิ่ง 643021264-9
10. นางสาวจินดาพร โพธิ์ภูมิ 643021262-3

ที่มาและ ความสำคัญ

ปัจจุบันเข้าสู่สังคมออนไลน์ ผู้คนได้รับข่าวสารอย่างรวดเร็ว ทันเหตุการณ์ ในทางเดียวกันก็ยังประสบกับปัญหาข่าวปลอมที่ก่อให้เกิดความเข้าใจผิดกันในสังคม สำนักงานพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ (สพธอ.) ได้สำรวจพฤติกรรมการใช้อินเทอร์เน็ตเป็นประจำทุกปีตั้งแต่ปี 2556 โดยมีวัตถุประสงค์หลัก คือ เพื่อรวบรวมข้อมูลที่แสดงลักษณะและแนวโน้มพฤติกรรมการใช้อินเทอร์เน็ตของคนไทยที่เปลี่ยนแปลงอย่างต่อเนื่อง มีข้อมูลที่สำคัญและจำเป็นต่อการพัฒนาธุรกรรมทางอิเล็กทรอนิกส์ไว้ให้บริการแก่ผู้ใช้อ้างอิงข้อมูลไม่ว่าจะเป็นภาครัฐ ภาคเอกชน และผู้สนใจ

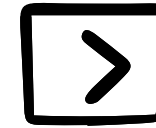
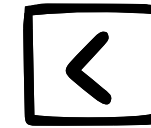
การทำ Data Mining ช่วยในการเข้าใจถึงพฤติกรรมการใช้สื่อที่ส่งผลทำให้เกิดการได้รับข่าวปลอม เพื่อใช้ในการหาแนวทางการแก้ไขปัญข่าวปลอมต่อไป

วัตถุประสงค์

REASON

เพื่อสร้างโมเดลทำนายการรับรู้ข่าวปลอม
จากข้อมูลสถิติการสำรวจพฤติกรรมผู้ใช้งาน
อินเทอร์เน็ตในประเทศไทย

ข้อมูลสถิติการสำรวจพฤติกรรมผู้ใช้งาน
อินเทอร์เน็ตในประเทศไทย



fake_news	online_acitivity_1	online_acitivity_2	online_acitivity_3	online_acitivity_4	online_acitivity_5	online_acitivity_6
2	1	0	1	1	1	1
2	1	0	0	1	0	0
2	1	1	1	1	1	1
2	1	0	1	1	1	0
2	0	1	0	1	0	1
2	1	0	0	0	1	0
2	1	0	1	1	1	1
2	1	0	0	0	0	0
1	1	1	0	1	1	0
2	1	1	0	1	1	1
2	1	0	0	1	1	1
2	1	0	0	1	1	0
2	1	0	0	0	1	0
2	1	0	0	1	1	0
2	1	0	1	1	1	1



ตัวอย่าง
DATA SETS





ตัวแปร



●●● ตัวแปร Y : TARGET
VARIABLE

fake_news

ตัวแปร X : FEATURE VARIABLES

●●● online_acitivity_1 online_acitivity_7
online_acitivity_2 online_acitivity_8
online_acitivity_3 online_acitivity_9
online_acitivity_4 online_acitivity_10
online_acitivity_5 online_acitivity_11
online_acitivity_6 online_acitivity_12

●●● online_acitivity_13
online_acitivity_14
online_acitivity_15
online_acitivity_16
online_acitivity_17
online_acitivity_18



นิยามตัวแปร

- ● ● fake_news = ทำนเคยพบเห็นข่าวปลอม (Fake News) บนโลกออนไลน์หรือไม
- online_acitivity_1 = ใช้ Social Media เช่น Facebook, Twitter, Instagram
- online_acitivity_2 = ใช้แอปพลิเคชันถ่ายทอดสด เช่น Facebook Live, Instagram Live , YouTube live
- online_acitivity_3 = เล่นเกมออนไลน์
- online_acitivity_4 = รับ-ส่งอีเมล
- online_acitivity_5 = ค้นหาข้อมูล (Search Engine) เช่น ค้นหาข้อมูลใน Google/Bing
- online_acitivity_6 = เรียนออนไลน์ (e-Learning)
- online_acitivity_7 = นทางาน/สมัครงานทางออนไลน์
- online_acitivity_8 = ซื้อขายสินทรัพย์เพื่อการลงทุน เช่น หุ้น กองทุนรวม
- online_acitivity_9 = ติดต้อสื่อสารออนไลน์ ทั้งการโทรศัพท์ และพูดคุย (Chat)

นิยามตัวแปร



- online_acitivity_10 = ขายสินค้าและบริการออนไลน์
- online_acitivity_11 = ดาวน์โหลดซอฟต์แวร์/เพลง/ละคร/ภาพยนตร์/เกม/ไอเทมในเกม
- online_acitivity_12 = ซื้อสินค้าและบริการออนไลน์
- online_acitivity_13 = ดูโทรทัศน์/ดูคลิป/ดูหนัง/ฟังเพลงออนไลน์
- online_acitivity_14 = อ่านข่าว/บทความ/หนังสืออิเล็กทรอนิกส์ (e-Book)
- online_acitivity_15 = ทำธุรกรรมทางการเงินออนไลน์
- online_acitivity_16 = ใช้งานบริการภาครัฐผ่านระบบออนไลน์ เช่น ชำระภาษีออนไลน์
- online_acitivity_17 = ทำงานผ่านระบบออนไลน์ ประชุมออนไลน์
- online_acitivity_18 = ซื้อประกันออนไลน์ เช่น ประกันCOVID19 ประกันสุขภาพ ประกันอุบัติเหตุ ฯลฯ

STEP 1

จัดการ Missing Data

- หาค่า missing value จากข้อมูลทั้งหมด

```
[ ] null_values = df.isnull().sum()
```

```
print(null_values)
```

```
fake_news      0
online_acitivity_1  0
online_acitivity_2  0
online_acitivity_3  0
online_acitivity_4  0
online_acitivity_5  0
online_acitivity_6  0
online_acitivity_7  0
online_acitivity_8  0
online_acitivity_9  0
online_acitivity_10  0
online_acitivity_11  0
online_acitivity_12  0
online_acitivity_13  0
online_acitivity_14  0
online_acitivity_15  0
online_acitivity_16  0
online_acitivity_17  0
online_acitivity_18  0
dtype: int64
```

STEP 2

Conditional Filtering

ตรวจสอบค่า Null หรือค่าผิดปกติ ใน DataFrame

- ตรวจสอบค่าใน column fake_news ว่าเป็นค่า 1 และ 2 หรือไม่

```
# Check fake_news column
valid_fake_news = all(df['fake_news'].isin([1, 2]))
valid_fake_news
```

```
True
```

- ตรวจสอบค่าใน column online_acitivity ทั้งหมด ว่าเป็นค่า 0 และ 1 หรือไม่

```
[ ] # Select columns from the 2nd to last
activity_columns = df.iloc[:, 1:]

# Check if all values are either 0 or 1
valid_values = (activity_columns.isin([0, 1])).all().all()

valid_values
```

```
True
```

DATA PREPARATION

Conditional Filtering



- ตรวจสอบค่าที่ไม่ใช่ตัวเลขในข้อมูล

```
17] non_numeric_values = df.apply(lambda x: pd.to_numeric(x, errors='coerce')).isnull().sum()

print(non_numeric_values)
```

```
fake_news      0
online_acitivity_1  0
online_acitivity_2  0
online_acitivity_3  0
online_acitivity_4  0
online_acitivity_5  0
online_acitivity_6  0
online_acitivity_7  0
online_acitivity_8  0
online_acitivity_9  0
online_acitivity_10 0
online_acitivity_11 0
online_acitivity_12 0
online_acitivity_13 0
online_acitivity_14 0
online_acitivity_15 0
online_acitivity_16 0
online_acitivity_17 0
online_acitivity_18 1
dtype: int64
```

online_acitivity_18

S



9

- ลบ row ไม่ใช่ตัวเลข (เนื่องจากมีข้อมูลที่ผิดปกติไม่เกิน 5% ของข้อมูลทั้งหมด)

```
[41] non_numeric_rows = df.apply(lambda x: pd.to_numeric(x, errors='coerce')).isnull().any(axis=1)
      numeric_df = df[~non_numeric_rows]
      numeric_df
```

STEP 3

การแบ่งชุดข้อมูล:

- training set
- test set

```
# Features (X)
X = data.drop(columns=['fake_news']) # คอลัมน์กิจกรรมออนไลน์ทั้งหมด

# Target variable (y)
y = data['fake_news'] # เลือกคอลัมน์ 'fake_news' เป็นตัวแปรเป้าหมาย
```

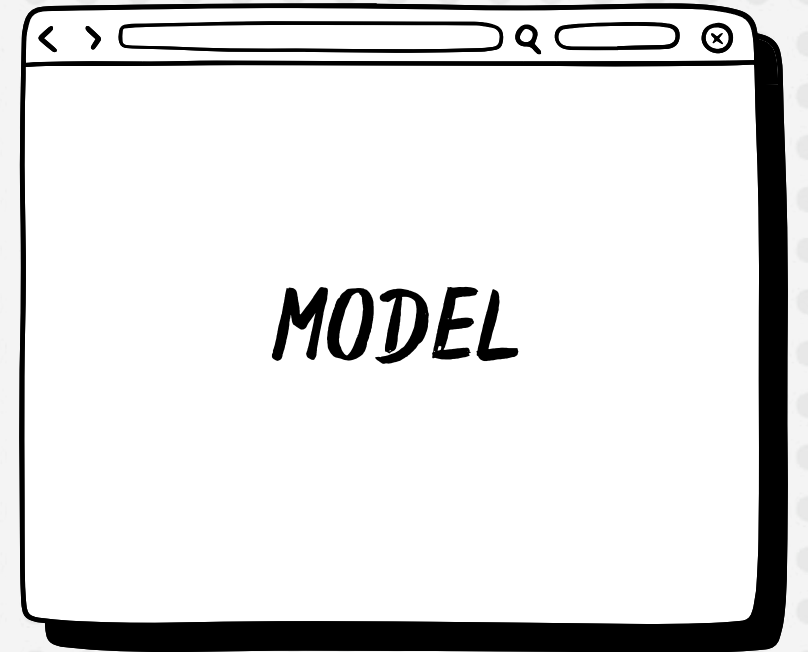
```
# แบ่งข้อมูลออกเป็นชุดในการ train และ test (train 80% test 20%)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# Print the shapes of the training and test sets
print("Training set shape (X):", X_train.shape)
print("Training set shape (y):", y_train.shape)
print("Test set shape (X):", X_test.shape)
print("Test set shape (y):", y_test.shape)
```

```
Training set shape (X): (17318, 18)
Training set shape (y): (17318,)
Test set shape (X): (4330, 18)
Test set shape (y): (4330,)
```

DATA PREPARATION

- Decision Trees: สามารถสร้างเงื่อนไขในการตัดสินใจเพื่อจำแนกว่าบุคคลที่เราสนใจอยู่ ว่าเคยพบเห็นข่าวปลอมหรือไม่ โดยใช้คุณลักษณะกิจกรรมทางออนไลน์ต่างๆ เป็นตัวชี้วัด
- K-Nearest Neighbors (KNN): โดยจำแนกว่าบุคคลใดๆ เคยพบเห็นข่าวปลอมบนโลกออนไลน์หรือไม่ โดยพิจารณาคุณลักษณะกิจกรรมทางออนไลน์ต่างๆ





ประโยชน์

ใช้เป็นเครื่องมือในการทำนายพฤติกรรมการใช้
อินเทอร์เน็ตของคนไทยในอนาคตได้
ซึ่งสามารถช่วยในการวางแผนและตัดสินใจใน
การดำเนินกลยุทธ์ต่างๆ
นำไปสู่การพัฒนานโยบายหรือการดำเนินการที่
เหมาะสมเพื่อส่งเสริมการใช้งานอินเทอร์เน็ตที่
ปลอดภัยและมีประสิทธิภาพ





THANK
YOU

SIZE THE MARKET

Apply the two ways to size the market - top down or bottom up.

1.9 Billion

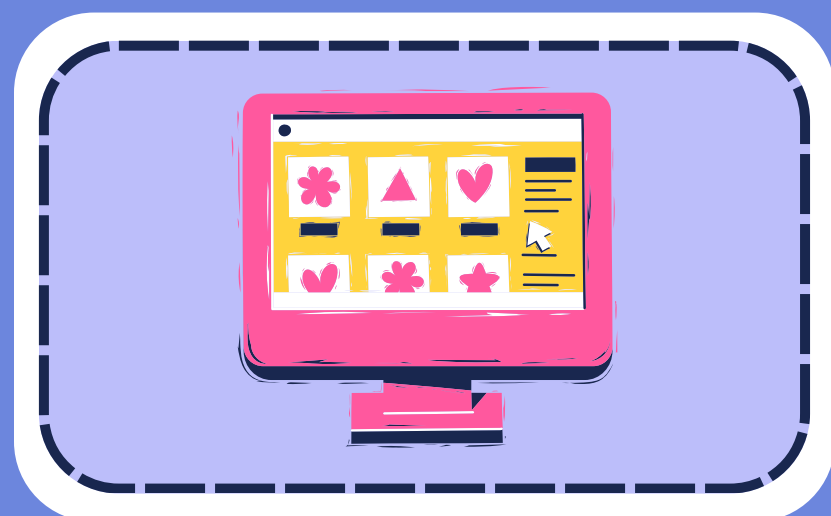
Total Available Market (TAM)

53 M

Serviceable Available Market (SAM)

10.6M

Serviceable Obtainable Market (SOM)

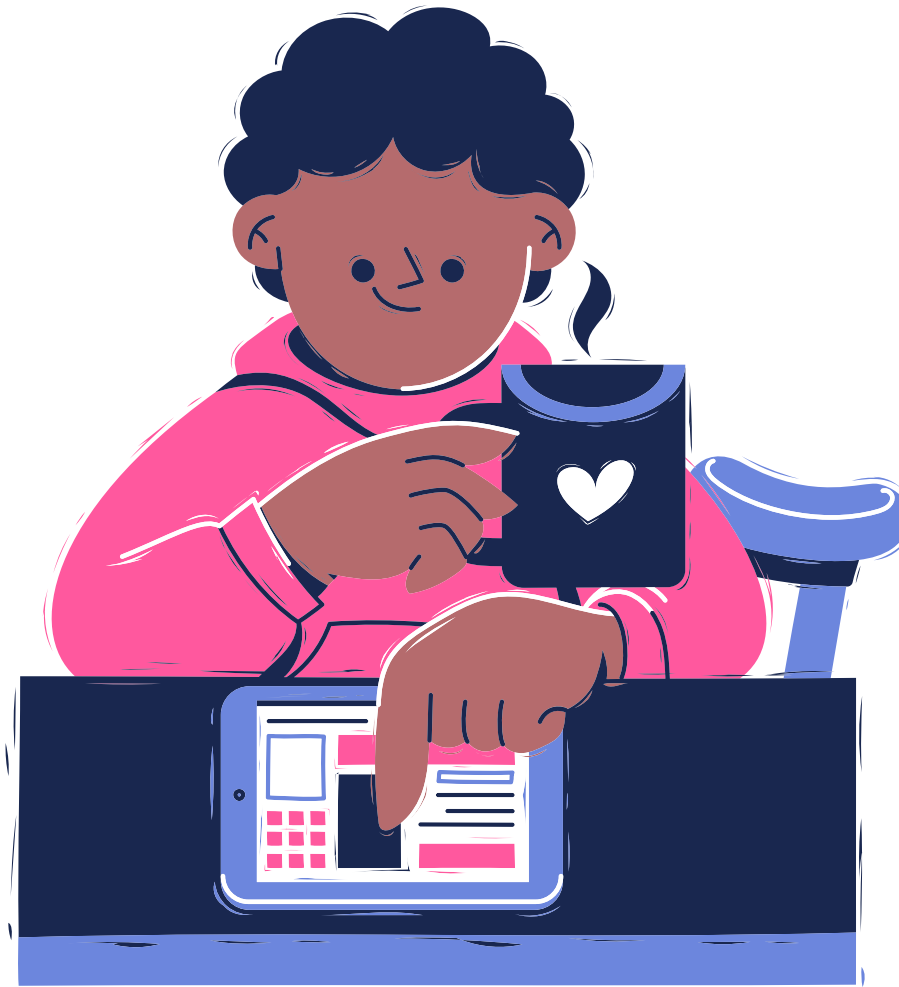


RESEARCH SKILLS

Basic skills to help start your research



Lesson Outline



- Overview of Research Skills
- Online Search Modifiers
- Search Tips
- The Credibility Test
- Skimming and Scanning



Think About...

your favorite gadget or invention.
Can you imagine life before this technology?





Today you will learn

- How to gather relevant information from multiple sources, particularly online ones
- Use effective search terms and assess the credibility and accuracy of each source

Search Term:

a word or combination of words or characters entered into a search engine in order to narrow down results



Basic Research Skills

With a topic on hand, here are the first few steps to kick off your research.

Use quality sources

Find a wide range of sources, such as books, magazines, and online articles. Make sure they are reliable.

Use online search modifiers

These are symbols and characters added to search terms for a more efficient online search.

Skim and scan

Skimming and scanning are reading methods of going through information quickly while noting down important details.



Online search modifiers

When conducting online research, combine use these modifiers with your search terms to get better and more specific results.

AND

It narrows down results by looking for **both** or **all** your search terms in each source.

OR or minus (-)

They tell the search engine to alternate or exclude search terms.

Quotation marks (" ")

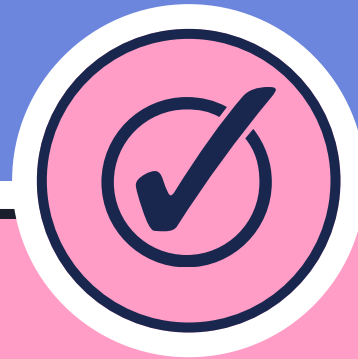
Enclosing search terms in quotation marks tells the search engine to look for the exact phrase.

THE CREDIBILITY TEST



Source Type

Examine the source thoroughly. Is it a book, a website article, a research paper, or a news report?



Authorship

Is the author an expert in the field about which they are writing?
Is the publisher reliable and well-known?



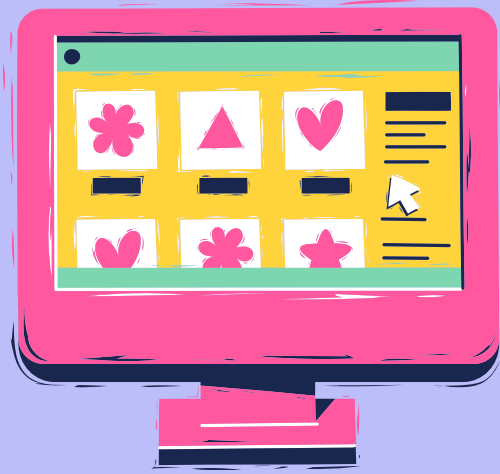
URL

Does the website end in ".edu," or ".org," indicating that it belongs to an educational institution, or a non-profit organization?



Up-to-date

Check the publication date. Is the information current, or has it recently been revised?



READING TECHNIQUES

When you are looking at many sources, use these techniques to help you read efficiently.

Skimming

Skim reading allows you to process a large amount of information quickly. Read the first sentence for each paragraph to get an overall understanding of what your reference is about.

Scanning

Scanning allows you to read and spot specific information you need. You can look for keywords related to your topic.

Share your thoughts

How can we know if our sources are credible and accurate?



TOPIC

**Source or
Website Link**

Summary of the essential
information

I know this source is credible because...

**Source or
Website Link**

Summary of the essential
information

I know this source is credible because...

**Source or
Website Link**

Summary of the essential
information

I know this source is credible because...