Sentiment Analysis of Patient Feedback with Machine Learning

DS532 TEXT ANALYTICS
Srinakharinwirot University

AGENDA

About the Project 01

Our Workflows 02

Web Scraping 03

Text Pre-processing 04

05 Model Evaluation

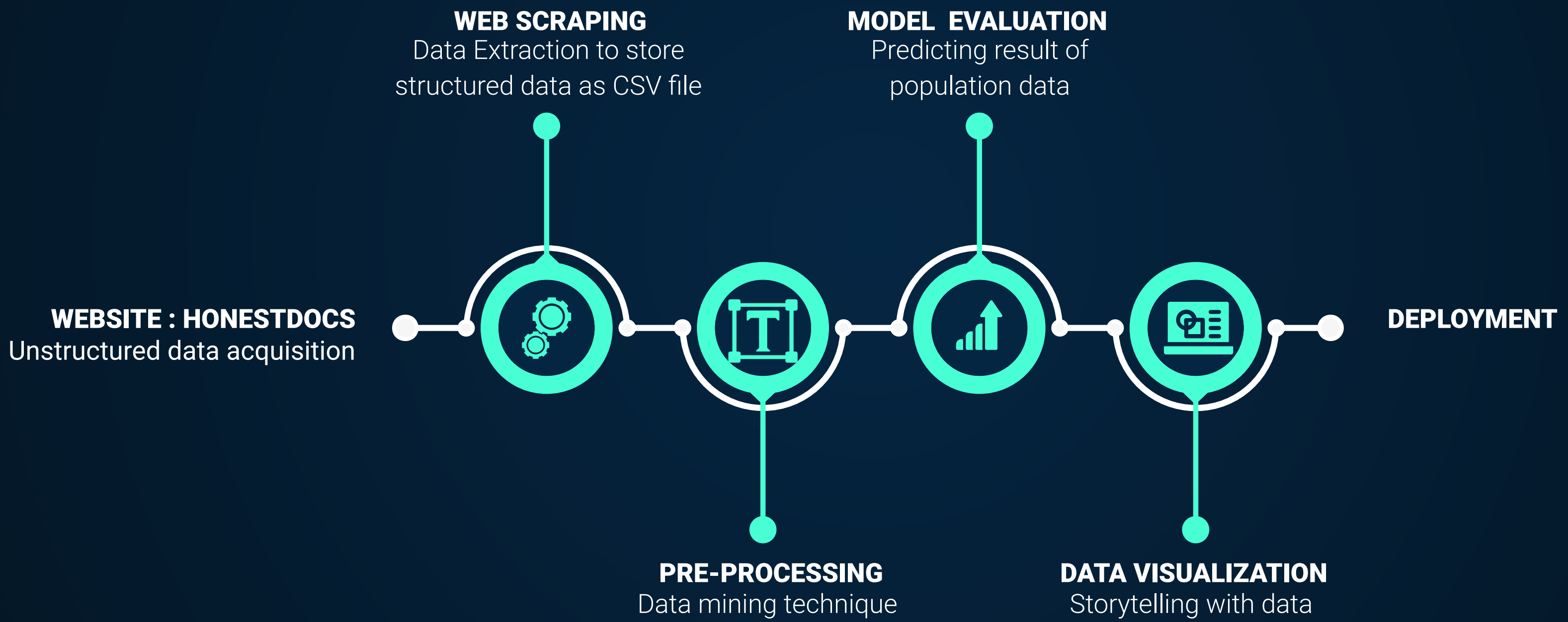
06 Data Visualization

07 Our Team

ABOUT THE PROJECT

Data scientist who was given a problem to analyze customer experiences of hospital services. Traditional survey has several limitations such as reluctance feedback. Moreover, it takes time to get data and the hospital cannot make decision on time to handle customer experiences.

OUR WORKFLOWS



WEB SCRAPING : Gathering of data

Patient Feedback are acquired by scraping from the website www.honestdocs.com.
We select the following three hospitals to make sentiment analysis.

COMMENTS

?

Siriraj Piyamaharajkarun Hospital

โรงพยาบาลศิริราช ปิยมหาราชการุณย์

https://www.honestdocs.co/hospitals/siriraj-piyamaharajkarun-hospital

COMMENTS

?

Sirindhorn Hospital

โรงพยาบาลสิรินธร

https://www.honestdocs.co/hospitals/police-general-hospital

COMMENTS

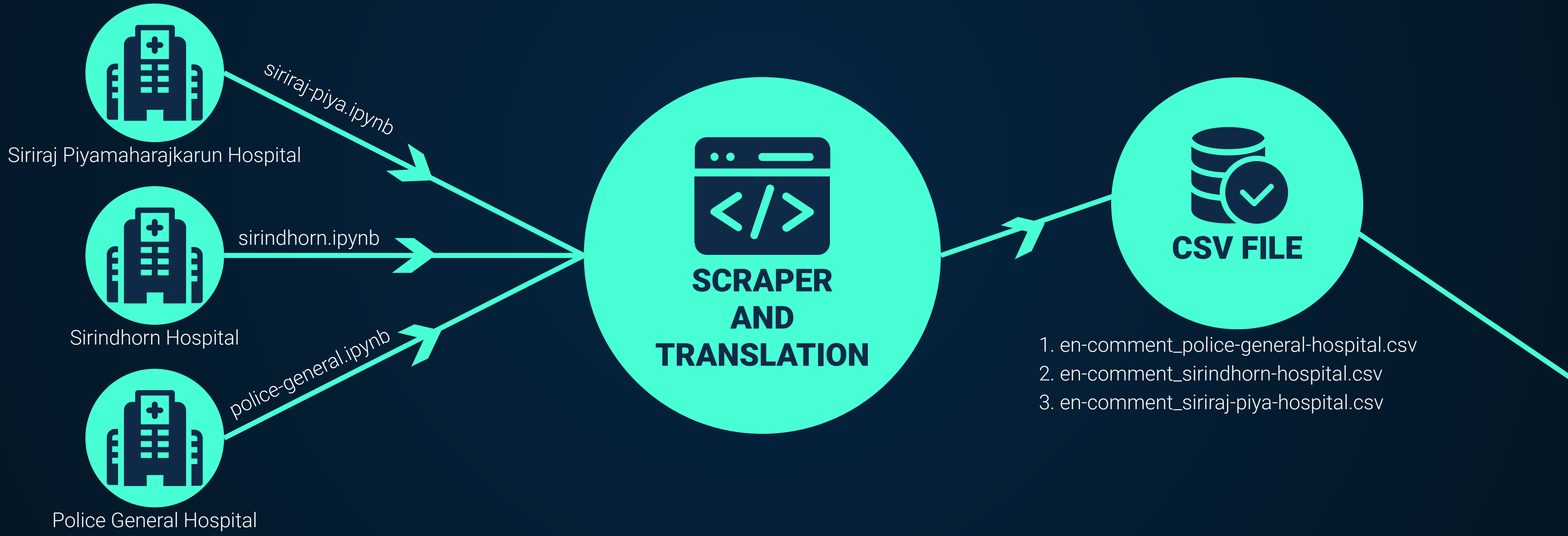
?

Police General Hospital

โรงพยาบาลตำรวจ

https://www.honestdocs.co/hospitals/sirindhorn-hospital

WEB SCRAPING : Process



WEB SCRAPING : Challenge

RUN SEPARATE FILES

It cannot run all at once due to limitations of library googletrans

REMOVE EMOJIS

Remove emojis from patient comments

REMOVE TEXT TOO LONG

To avoid an unexpected problem.

```python
#check emoji
import re
def remove_emojis(data):
    emoj = re.compile("["
        u"\U0001F600-\U0001F64F"  # emoticons
        u"\U0001F300-\U0001F5FF"  # symbols & pictographs
        u"\U0001F680-\U0001F6FF"  # transport & map symbols
        u"\U0001F1E0-\U0001F1FF"  # flags (iOS)
        u"\U00002500-\U00002BEF"  # chinese char
        u"\U00002702-\U000027B0"
        u"\U00002702-\U000027B0"
        u"\U000024C2-\U0001F251"
        u"\U0001f926-\U0001f937"
        u"\U00010000-\U0010ffff"
        u"\u2640-\u2642"
        u"\u2600-\u2B55"
        u"\u200d"
        u"\u23cf"
        u"\u23e9"
        u"\u231a"
        u"\ufe0f"  # dingbats
        u"\u3030"
                      "]+", re.UNICODE)
    return re.sub(emoj, '', data)
```

Remove Emojis

Siriraj Piyamaharajkarun Hospital

REMOVE EMOJIS

- Remove emojis from patient comments at index 394 or 365th comment

```python
'''
siripiya["en"] = siripiya.progress_apply(lambda x: th2en(x["comment"]), axis=1)
There is an unexpected issue at comment 395th  - index 394 - Found Emoji
395/396 [05:52<00:00,  1.11it/s]เคยไปหาแพทย์ผิวหนัง คุณหมอต่อพงษ์ (จำนามสกุลไม่ได้อะค่ะ) การบริการดีเยี่ยม👍🏻
คุณหมอให้คำปรึกษาดี ติดตามอาการ แต่ควรนัดล่วงหน้านะค่ะ จะได้ไม่ต้องรอนาน^^ ข้อดีของที่นี่ สถานที่⭐⭐⭐⭐⭐ ที่จอดรถมีเยอะค่ะ
ถ้ามาหาคุณหมอจอดฟรี ถ้าไม่ได้มาหาคุณหมอชม.100฿😅นะจ๊ะ ข้อเสีย รถติดมากจ้า🚗🚕🚙 ค่ารักษาที่นี่พอๆกับรพ.เอกชน
(เพราะรายได้ส่วนหนึ่งนำไปช่วยเหลือผู้ป่วยขาดแคลนจ้า)
so try to remove emoji and run above again
'''
print(siripiya.iloc[394,:].values[0])
siripiya.iloc[394,0] = remove_emojis(siripiya.iloc[394,0])
siripiya.iloc[394,:].values[0]
```

```
เคยไปหาแพทย์ผิวหนัง คุณหมอต่อพงษ์ (จำนามสกุลไม่ได้อะค่ะ) การบริการดีเยี่ยม👍🏻 คุณหมอให้คำปรึกษาดี ติดตามอาการ แต่ควรนัดล่วงหน้านะค่ะ จะได้ไม่ต้องรอนาน^^ ข้อดีของที่นี่ สถานที่⭐⭐⭐⭐⭐
'เคยไปหาแพทย์ผิวหนัง คุณหมอต่อพงษ์ (จำนามสกุลไม่ได้อะค่ะ) การบริการดีเยี่ยม คุณหมอให้คำปรึกษาดี ติดตามอาการ แต่ควรนัดล่วงหน้านะค่ะ จะได้ไม่ต้องรอนาน^^ ข้อดีของที่นี่ สถานที่ ที่จอดรถมีเยอะค่ะ
```

Remove text too long

Sirindhorn Hospital

REMOVE TEXT TOO LONG

Remove at index 14 to avoid
an unexpected problem.

Before removing : 283 comments
After removing : 282 comments

```
[ ] #for check lenght with issue comment
    i = 0
    for x in sirin['comment']:
     if len(sirin['comment'][i]) > 1000:
      print(len(sirin['comment'][i]), "-> index : ", i)
     i=i+1
```

```
6534 -> index :  14
```

```
[ ] '''
    There is an unexpected issue at comment 15th  - index 14
    so try to remove one and run above again
    '''
    print(sirin.shape)
    sirin.drop(14, inplace=True)
    print(sirin.shape)
    sirin.head()
```

```
(283, 2)
(282, 2)
```

WEB SCRAPING : After Cleansing Data

After cleansing data, and implemented with Google Translate API using library googletrans along with export to csv files separately. There are 396, 282, and 262 comments for Siriraj Piyamaharajkarun Hospital, Sirindhorn Hospital, and Police General Hospital respectively.

COMMENTS

396

en-comment_police-general-hospital.csv

Siriraj Piyamaharajkarun Hospital

โรงพยาบาลศิริราช ปิยมหาราชการุณย์

COMMENTS

282

en-comment_sirindhorn-hospital.csv

Sirindhorn Hospital

โรงพยาบาลสิรินธร

COMMENTS

262

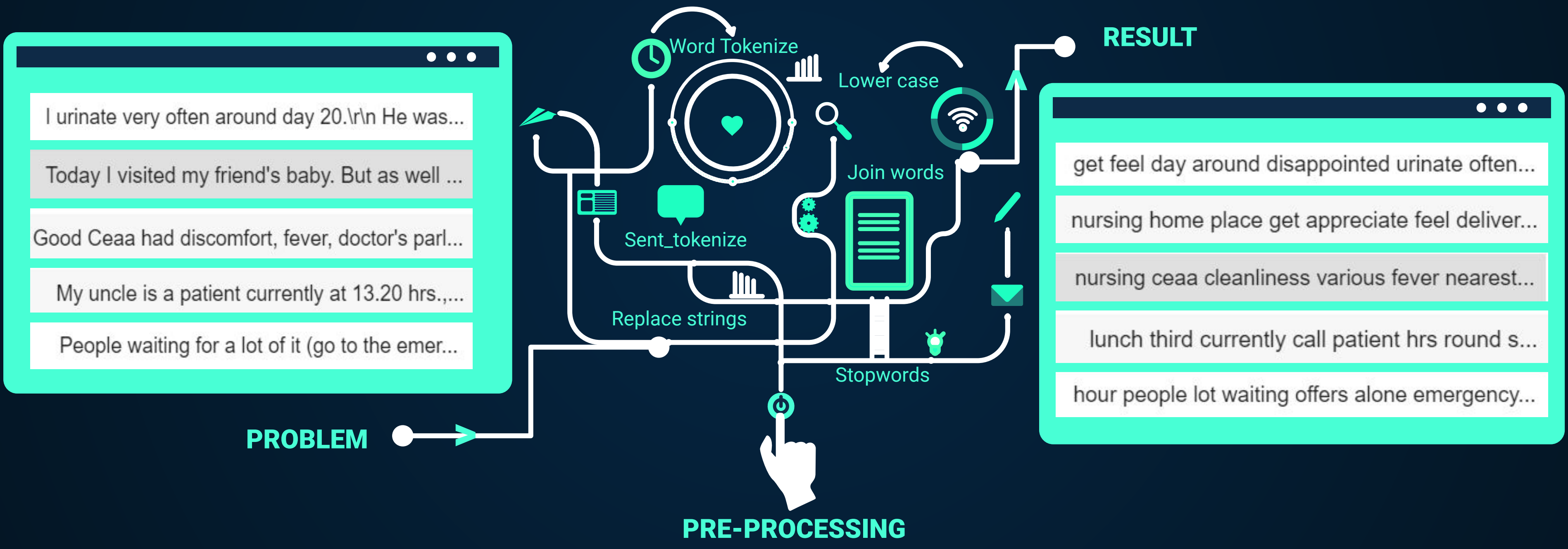
en-comment_siriraj-piya-hospital.csv

Police General Hospital

โรงพยาบาลตำรวจ

Updated April 2020

TEXT PREPROCESSING : Process



TEXT PREPROCESSING : Sentence-level sentiment

Example sentences

" nursing home place get appreciate feel delivered hope okay better doctors friends friend baby good like visited much well room clean visit rest... "

"hour people lot waiting offers alone emergency according speed wait good later doctor speak affordable one fine nurses keeping way although..."

"full back gynecology security obstetrics come details located wait doctors result good like chill yet much nurses never social promenade serving department flattering service near moving awaiting use conveniently"

"department key records call service patients book give outpatient medical fitting examination fast faster contacting center go queue"

"postpone call lot wanted burr learning feared told day search exploration doctors antenatal may doctor comes one unusable really old move department pregnancy numbers whose number net appointment stressed engaged help know sleep"

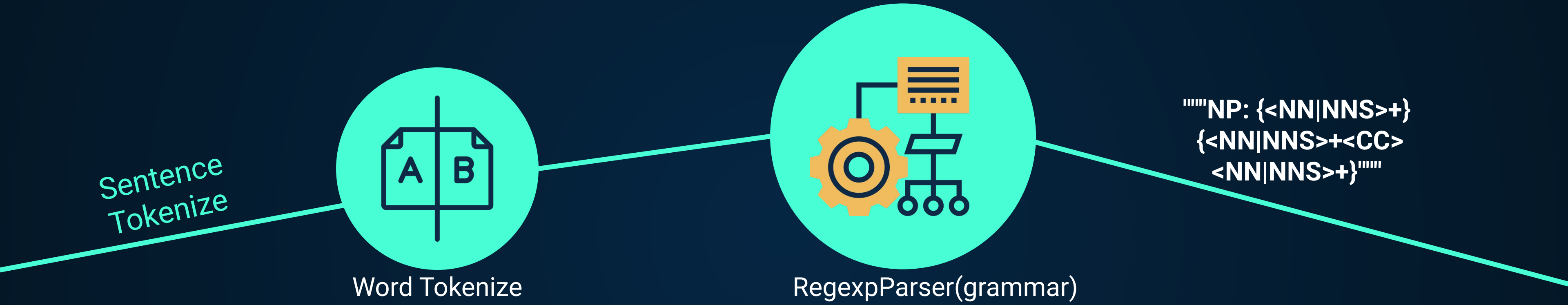
"minute say anc wait rarely nurses police hospital"

" ... wound services would expected medical heart asked study mention october nursing money people home three hostels treating worse coming work pressure wait brain return deposited returned gone lesions outcry mouth say shocked care skin disinfectant bad "

"bathroom full perhaps nursing even poor long time many meet car updated come emergency wait sunday saturday road queues spite areas..."

"lunch third currently call patient hrs round stunk uncle"

TEXT PREPROCESSING : Parsing




```
[Tree('S', [Tree('NP', [('nursing', 'NN'), ('home', 'NN'), ('place', 'NN')]), ('get', 'VBP'), ('appreciate', 'JJ'), Tree('NP', [('feel', 'NN')]), ('delivered', 'VBN')
 Tree('S', [('get', 'VB'), ('feel', 'JJ'), Tree('NP', [('day', 'NN')]), ('around', 'IN'), ('disappointed', 'JJ'), ('urinate', 'JJ'), ('often', 'RB'), Tree('NP', [('st
 Tree('S', [Tree('NP', [('nursing', 'NN'), ('ceaa', 'NN'), ('cleanliness', 'NN')]), ('various', 'JJ'), ('fever', 'RB'), ('nearest', 'JJS'), ('good', 'JJ'), ('followed
 Tree('S', [Tree('NP', [('months', 'NNS')]), ('assoc', 'VBP'), ('unusual', 'JJ'), Tree('NP', [('opens', 'NNS')]), ('see', 'VBP'), ('happening', 'VBG'), Tree('NP', [('
 Tree('S', [Tree('NP', [('nursing', 'NN'), ('ceaa', 'NN'), ('cleanliness', 'NN')]), ('various', 'JJ'), ('fever', 'RB'), ('nearest', 'JJS'), ('good', 'JJ'), ('followed
 Tree('S', [('high', 'JJ'), ('regularly', 'RB'), Tree('NP', [('state', 'NN')]), ('probably', 'RB'), ('private', 'JJ'), Tree('NP', [('siriraj', 'NN')]), ('congested',
 Tree('S', [('call', 'VB'), ('regularly', 'JJ'), Tree('NP', [('quality', 'NN'), ('schedule', 'NN')]), ('cleaned', 'VBD'), ('makes', 'VBZ'), Tree('NP', [('maintain', '
 Tree('S', [('high', 'JJ'), ('regularly', 'RB'), Tree('NP', [('state', 'NN'), ('time', 'NN')]), ('probably', 'RB'), ('private', 'JJ'), Tree('NP', [('siriraj', 'NN')])
 Tree('S', [Tree('NP', [('children', 'NNS')]), ('eloquence', 'JJ'), Tree('NP', [('place', 'NN')]), ('waiting', 'VBG'), ('explaining', 'VBG'), Tree('NP', [('paint', 'N
 Tree('S', [('well', 'RB'), Tree('NP', [('services', 'NNS')]), ('please', 'VBP'), ('various', 'JJ'), ('faster', 'RBR'), Tree('NP', [('treatment', 'NN'), ('heal', 'NN'
 Tree('S', [Tree('NP', [('bathroom', 'NN'), ('signs', 'NNS')]), ('call', 'VBP'), Tree('NP', [('time', 'NN'), ('feel', 'NN'), ('trolleys', 'NNS')]), ('nephrology', 'VB
 Tree('S', [('maintaining', 'VBG'), ('excellent', 'JJ'), Tree('NP', [('service', 'NN'), ('access', 'NN')])]),
 Tree('S', [Tree('NP', [('services', 'NNS'), ('levels', 'NNS')])]),
```

TEXT PREPROCESSING : Extract Noun Phrase

Parsing

Piyamaharajkarun Hospital

```
[ ] nps_siripya

[['nursing home place',
  'feel',
  'doctors friends',
  'room',
  'visit rest',
  'care today service',
  'relatives'],
 ['day',
  'staff doctors',
  'rooms',
  'level',
  'step design care',
  'patients',
  'services building'],
 ['nursing ceaa cleanliness',
  'medications discomfort doctor',
  'parlance knowledge wagga hygiene',
  'symptoms hospital',
  'patients',
```

Sirindhorn Hospital

```
[ ] nps_sirin

[['bathroom',
  'session lot home times',
  'emergency floor purchase staff tunes doctors',
  'doctor insurance cool',
  'nurses',
  'attention',
  'department restaurants',
  'tb',
  'morning snacks',
  'online'],
 ['hour people',
  'offers',
  'emergency',
  'speed wait',
  'doctor speak',
  'fine',
  'way',
  'department thank',
  'appointment morning treatment',
  'queue'],
 ['problem subjects',
  'answers',
  'doctor',
```

Police General Hospital

```
[ ] nps_police

[['security obstetrics',
  'details',
  'wait doctors',
  'chill',
  'nurses',
  'promenade',
  'department',
  'service',
  'use'],
 ['lot',
  'time',
  'appreciate broom',
  'everyone work',
  'knew doctors',
  'hospitals',
  'rooms',
  'doctor sweep',
  'cute',
  'debris',
```

Split DataFrame into positive and negative sentiment

Extract Noun Phrase

Positive = 1
Negative = 0

```
[114] df_siripya
```

	cids	sentences	sentiments	NP
0	0	nursing home place get appreciate feel deliver...	1	[nursing home place, feel, doctors friends, ro...
1	1	get feel day around disappointed urinate often...	1	[day, staff doctors, rooms, level, step design...
2	2	nursing ceaa cleanliness various fever nearest...	1	[nursing ceaa cleanliness, medications discomf...
3	3	months assoc unusual opens see happening day s...	0	[months, opens, day, fever, yes week, standard...
4	4	nursing ceaa cleanliness various fever nearest...	1	[nursing ceaa cleanliness, medications discomf...
...	...	...	...	...
324	390	price expensive helpful recommended smiling st...	1	[price, helpful, staff service hospital option...
325	391	easy located take big understand doctors nurse...	1	[understand doctors, siddique]
326	392	atmosphere voice lot problems selection time l...	1	[voice lot problems, time, security, restauran...
327	394	went treatment lot long place advantage find c...	1	[treatment lot, place advantage, car, wait, po...
328	395	price spacious acceptable graduate without exp...	0	[price, graduate, experience, doctor worry]

329 rows × 4 columns

```
[116] df_police
```

	cids	sentences	sentiments	NP
0	0	full back gynecology security obstetrics come ...	1	[security obstetrics, details, wait doctors, c...
1	1	actually bed went recognize lot decided exactl...	1	[lot, time, appreciate broom, everyone work, k...
2	2	lunch third currently call patient hrs round s...	0	[hrs round stunk uncle]
3	3	hour blocks actually second even need back lot...	1	[hour blocks, lot people system, get security,...
4	4	care sleeping room nurse son took	1	[care, room nurse son]
...	...	...	...	...
216	257	nurse doctor beaming generous	1	[doctor]
217	258	nursing lot accident input mab emergency speec...	0	[nursing lot accident input mab emergency spee...
218	259	service clear quick good maintaining explanation	1	[service, maintaining explanation]
219	260	well break wait division afraid quick long tim...	1	[division, time medics]
220	261	really official reception cheerful	1	[reception cheerful]

221 rows × 4 columns

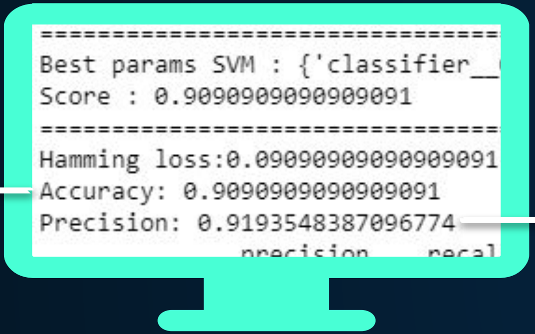
```
[115] df_sirin
```

	cids	sentences	sentiments	NP
0	0	bathroom nd went session lot home times okay i...	1	[bathroom, session lot home times, emergency f...
1	1	hour people lot waiting offers alone emergency...	1	[hour people, offers, emergency, speed wait, d...
2	2	problem subjects guidance early answers ration...	1	[problem subjects, answers, doctor, consult di...
3	3	problem subjects guidance early answers ration...	1	[problem subjects, answers, doctor, consult di...
4	4	home examination friendly doctors japan good p...	1	[home examination, doctors, press, doctor ever...
...	...	...	...	...
227	277	yap table offers generous dr	1	[yap table offers, dr]
228	278	successfully nursing need people lot france dr...	1	[need people, dream, encouragement see, change...
229	279	care nursing service staff clock night go c ti...	1	[care, service staff clock night, time, doctor...
230	280	detailed patients wait lot lost time medical e...	0	[patients, lot, time, examination]
231	281	sweatshops saying injection might exact state ...	0	[sweatshops, injection, state, duty, home, tim...

232 rows × 4 columns

SCORE COMPARISON

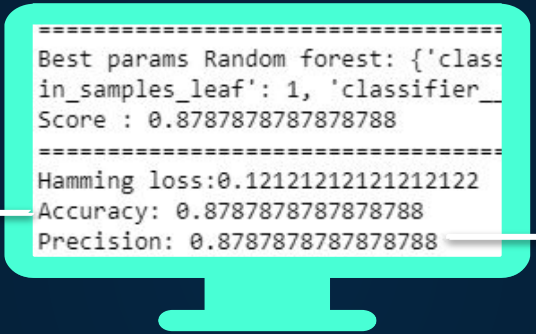
Siriraj Piyamaharajkarun Hospital



91% Accuracy

92% Precision

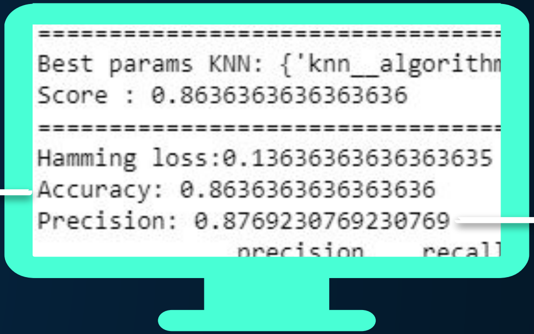
SVM



88% Accuracy

88% Precision

Random Forest



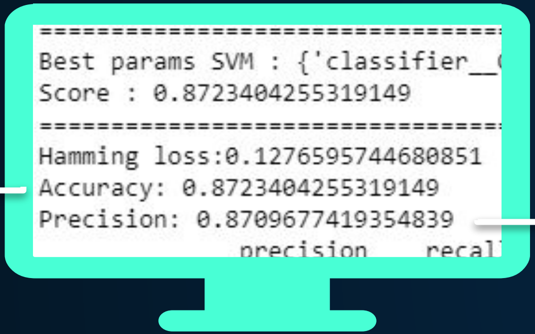
86% Accuracy

88% Precision

KNN

SCORE COMPARISON

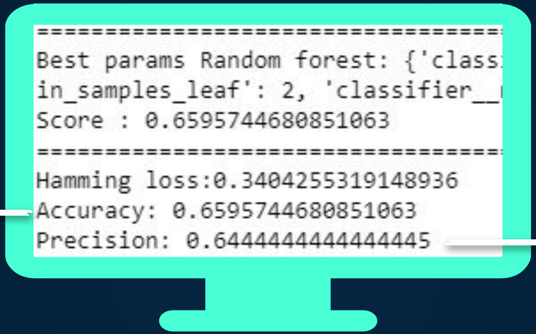
Sirindhorn Hospital


```
=====================================
Best params SVM : {'classifier__(
Score : 0.8723404255319149
=====================================
Hamming loss:0.1276595744680851
Accuracy: 0.8723404255319149
Precision: 0.8709677419354839
              precision    recall
```

87% Accuracy

87% Precision

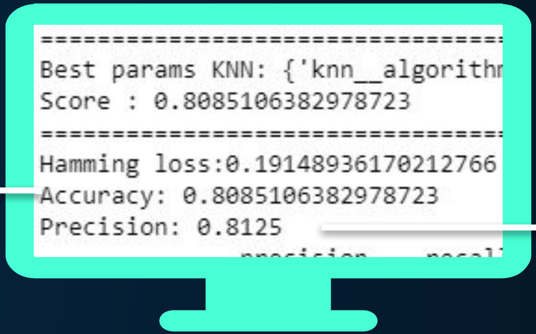
SVM


```
=====================================
Best params Random forest: {'class
in_samples_leaf': 2, 'classifier__
Score : 0.6595744680851063
=====================================
Hamming loss:0.3404255319148936
Accuracy: 0.6595744680851063
Precision: 0.6444444444444445
```

66% Accuracy

64% Precision

Random Forest


```
=====================================
Best params KNN: {'knn__algorithm
Score : 0.8085106382978723
=====================================
Hamming loss:0.19148936170212766
Accuracy: 0.8085106382978723
Precision: 0.8125
              precision    recall
```

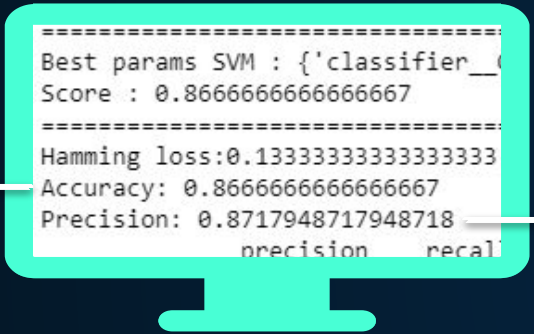
81% Accuracy

81% Precision

KNN

SCORE COMPARISON

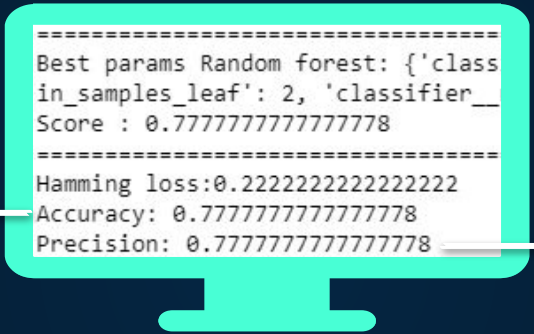
Police General Hospital



87% Accuracy

87% Precision

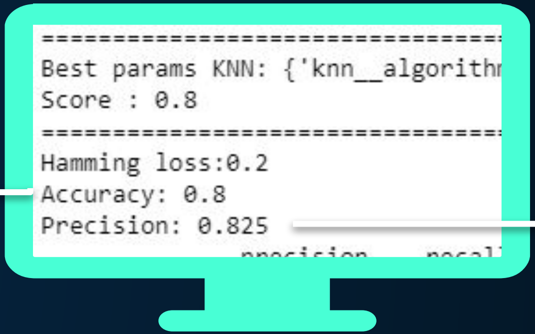
SVM



78% Accuracy

78% Precision

Random Forest



80% Accuracy

83% Precision

KNN

BEST MODEL

SVM

```
==================================
Best params SVM : {'classifier__C
Score : 0.9090909090909091
==================================
Hamming loss:0.09090909090909091
Accuracy: 0.9090909090909091
Precision: 0.9193548387096774
          precision    recall
```

91% Accuracy

92% Precision

Siriraj Piyamaharajkarun Hospital

```
==================================
Best params SVM : {'classifier__(
Score : 0.8723404255319149
==================================
Hamming loss:0.1276595744680851
Accuracy: 0.8723404255319149
Precision: 0.8709677419354839
          precision    recal
```

87% Accuracy

87% Precision

Sirindhorn Hospital

```
==================================
Best params SVM : {'classifier__(
Score : 0.8666666666666667
==================================
Hamming loss:0.1333333333333333
Accuracy: 0.8666666666666667
Precision: 0.8717948717948718
```

87% Accuracy

87% Precision

Police General Hospital

VISUALIZATION - Word Cloud

Siriraj Piyamaharajkarun Hospital

POSITIVE WORDS

NEGATIVE WORDS

VISUALIZATION - Word Cloud

Sirindhorn Hospital

POSITIVE WORDS

NEGATIVE WORDS

VISUALIZATION - Word Cloud

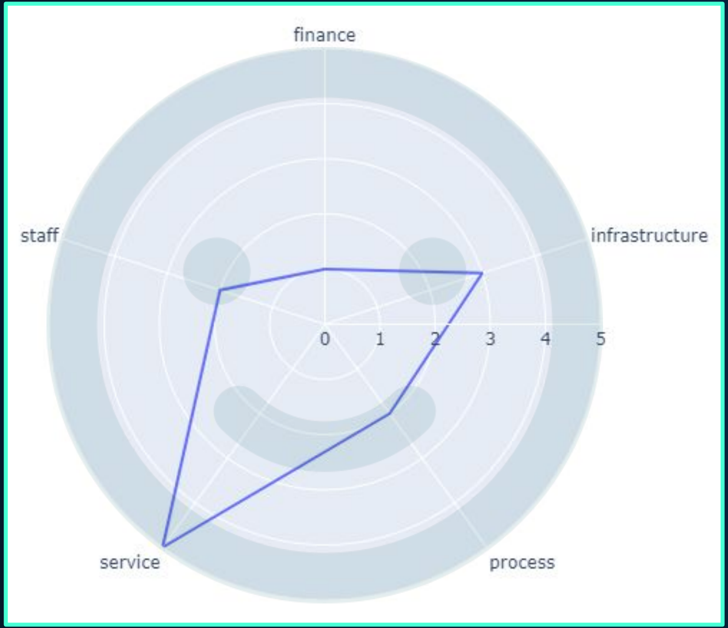
Police General Hospital

POSITIVE WORDS

NEGATIVE WORDS

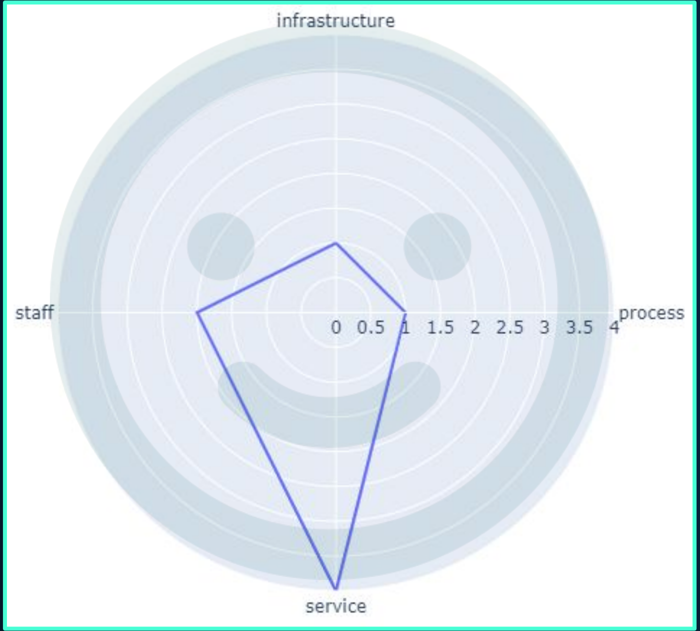
VISUALIZATION - Radar Chart

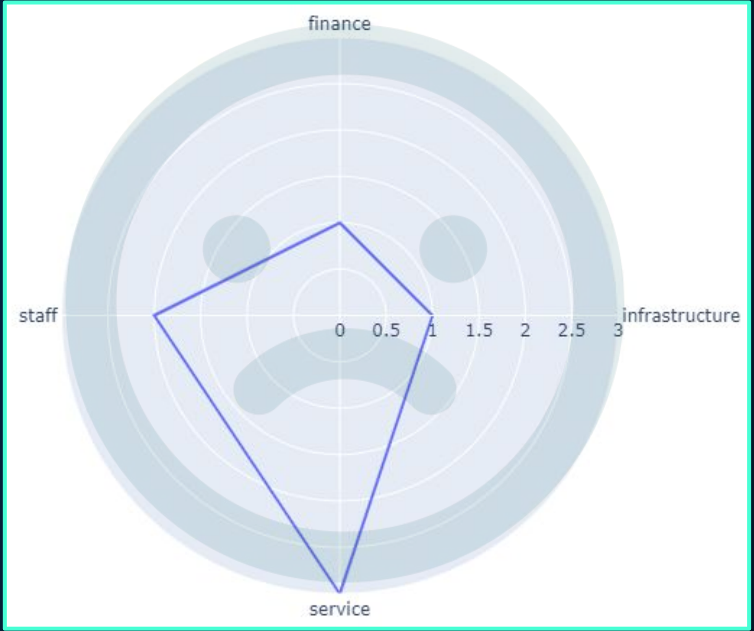
Siriraj Piyamaharajkarun Hospital



VISUALIZATION - Radar Chart

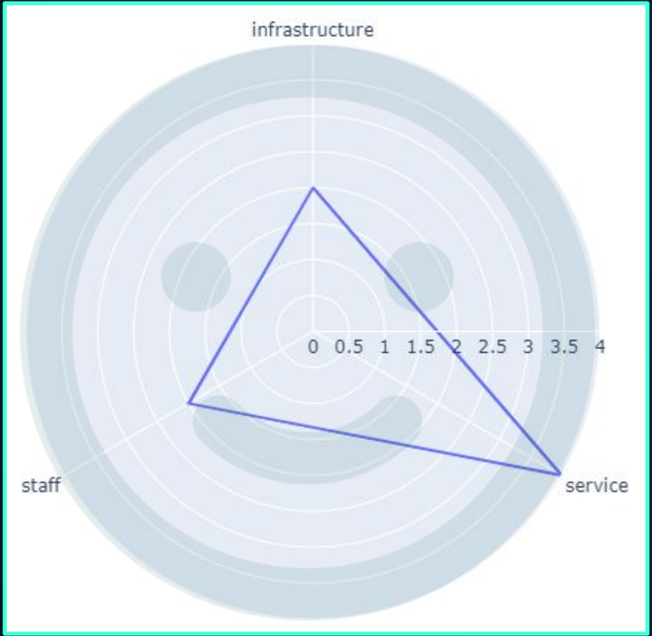
Sirindhorn Hospital

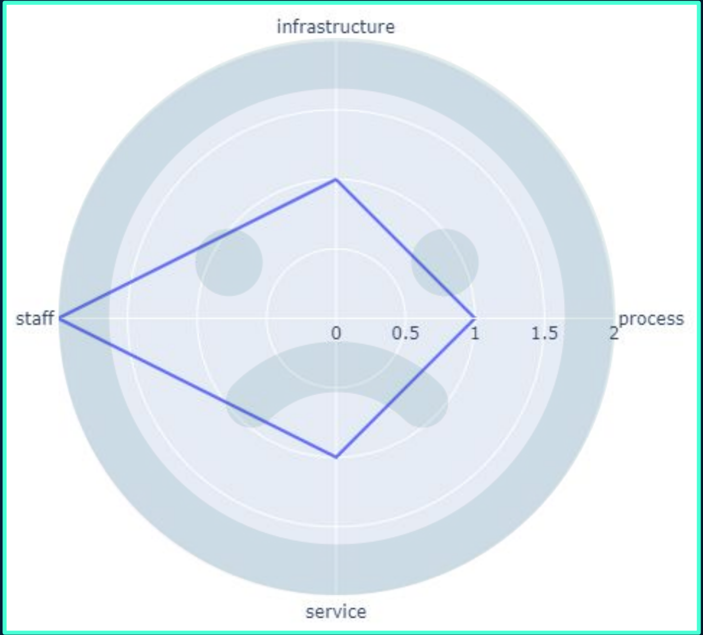




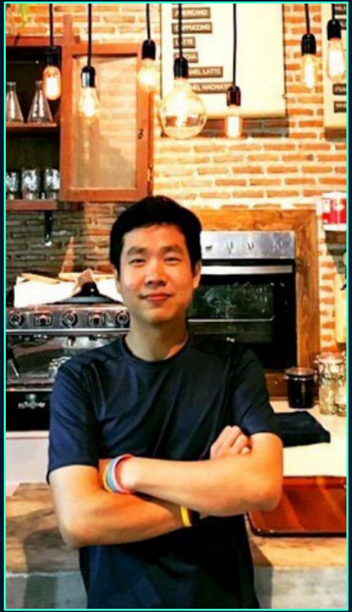
VISUALIZATION - Radar Chart

Police General Hospital





THE TEAM

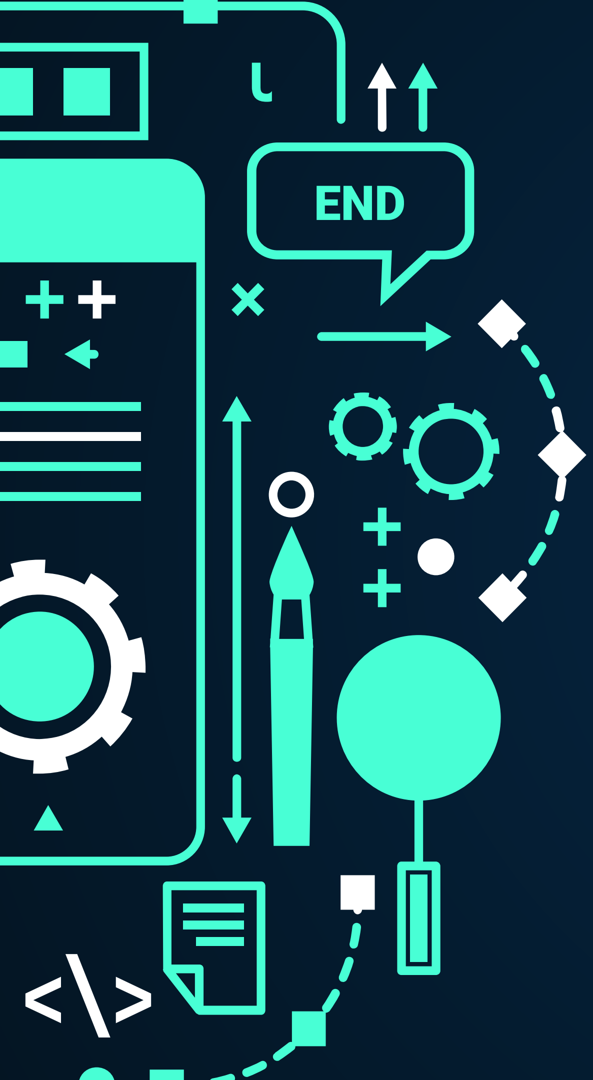


Teerawit Seekasamit
62199130233

Vijitra Thongthanachote
62199130377

Thiphawan Sawaengmee
62199130374

Piyawan Thongploy
62199130237



THANKS!

Dear Ph.D Ratchainant Thammasudjarit,

Good day to you, this message is to simply thank you !

You have educated us countless lessons in academic, as well as enlighten us through your teachings !