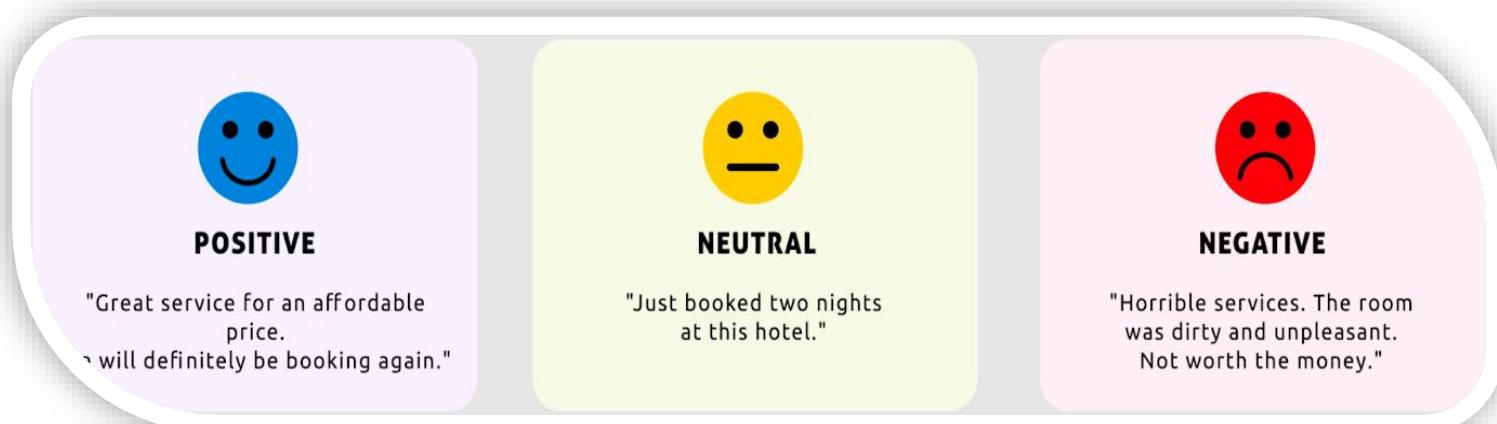


1. Sentiment Analysis on News Comment



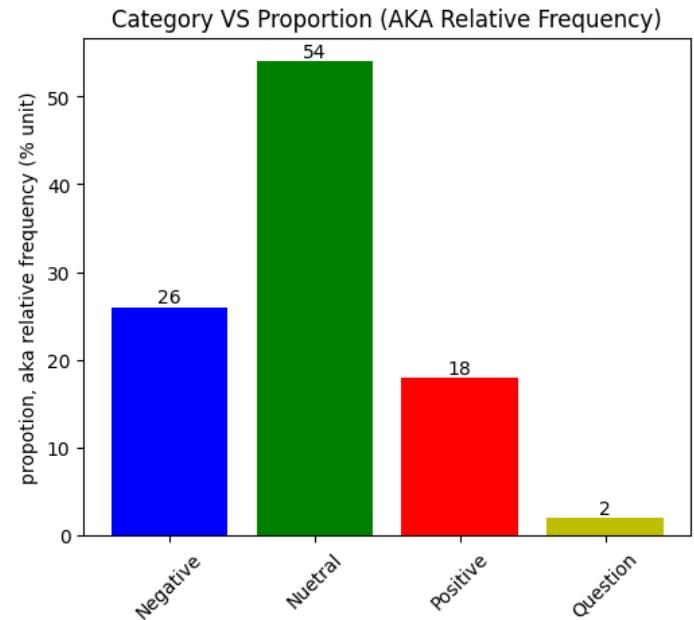
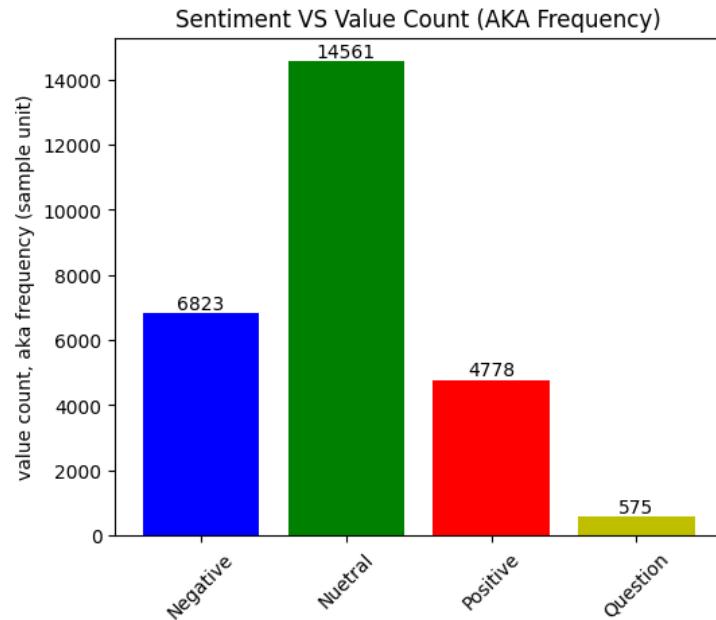
1.1 Dataset

Wisesight Sentiment Corpus:

- Social media messages in Thai language with sentiment label
- Sentiments include **negative**, **neutral**, **positive**, and **question**
- 26,737 messages
- <https://github.com/PyThaiNLP/wisesight-sentiment>

	Text	Sentiment
0		😊 Negative
1		😊 Negative
2		😊 Negative
3		😊 Negative
4	ร้า	Negative
...
26732	วิธีซื้อรถมือสองต้องดูยังไงตีครับตอนนั้นพี่จิมนะ...	Question
26733	- น้ำมันเครื่อง ทั้ง Fully / Semi เบนซิน - น้ำ...	Question
26734	Honda Civic Hatchback ที่มา กับสีแดง ราคาเท่าเด...	Question
26735	กำลังจะเริ่ม skincare 2 ตัวคือ stemfactor กัน ...	Question
26736	พีเค หนูอยากสอบถามเรื่องนึงค่ะ อาจจะไม่เกี่ยวก...	Question

26737 rows × 2 columns



1.2 Possible Solutions

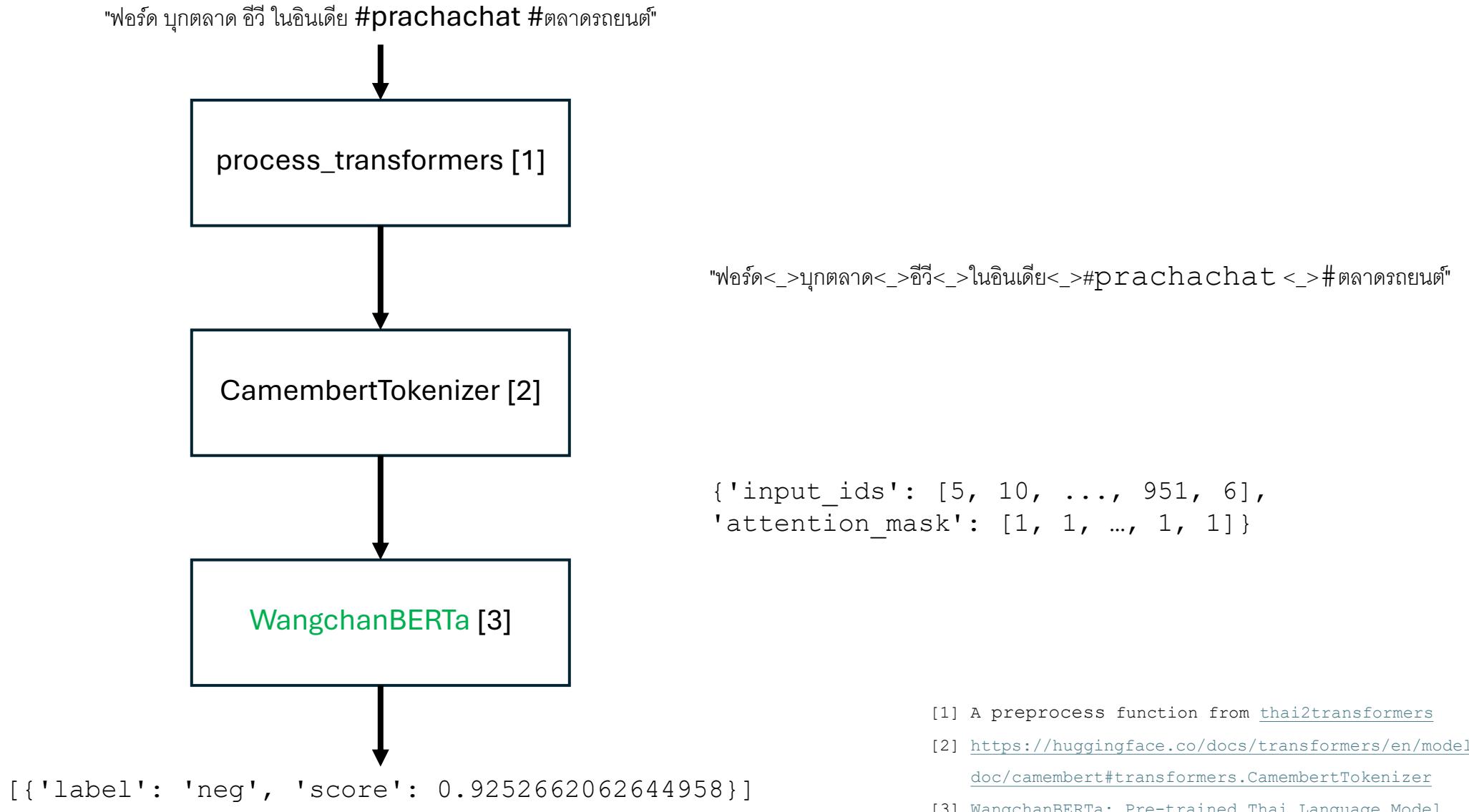
Solutions for sentiment analysis

	Feature + Model
Classical ML model	TF-IDF + Logistic Regression, SVM, or Naive Bayes etc.
Deep learning	Word embeddings + RNN or LSTM
Transformer based model	mBERT or WangchanBERTa

Model	Wisesight	Wongnai	Generated Reviews (Rating)	Prachathai
<i>Existing multilingual models</i>				
XLMR [Conneau et al., 2019]	73.57	62.57	64.91	68.18
mBERT [Devlin et al., 2018]	70.05	47.99	62.14	66.47
<i>Our baseline models</i>				
Navie Bays SVM	72.03	58.38	59.68	66.77
ULMFit (thai2fit)	70.95	61.79	64.33	66.21
<i>Our pretrained RoBERTa_{BASE} model</i>				
WangchanBERTa	76.19	63.05	64.66	69.78

ผลการทดสอบบนชุดข้อมูลสำหรับโจทย์ **sequence classification** บนชุดข้อมูลทดสอบ (**test set**) โดยเกณฑ์ในการวัดผลคือ **micro-averaged F1 score** ([WangchanBERTa: Pre-trained Thai Language Model](#))

1.3 Workflow for WangchanBERTa Inference



1.4 WangchanBERTa Evaluation

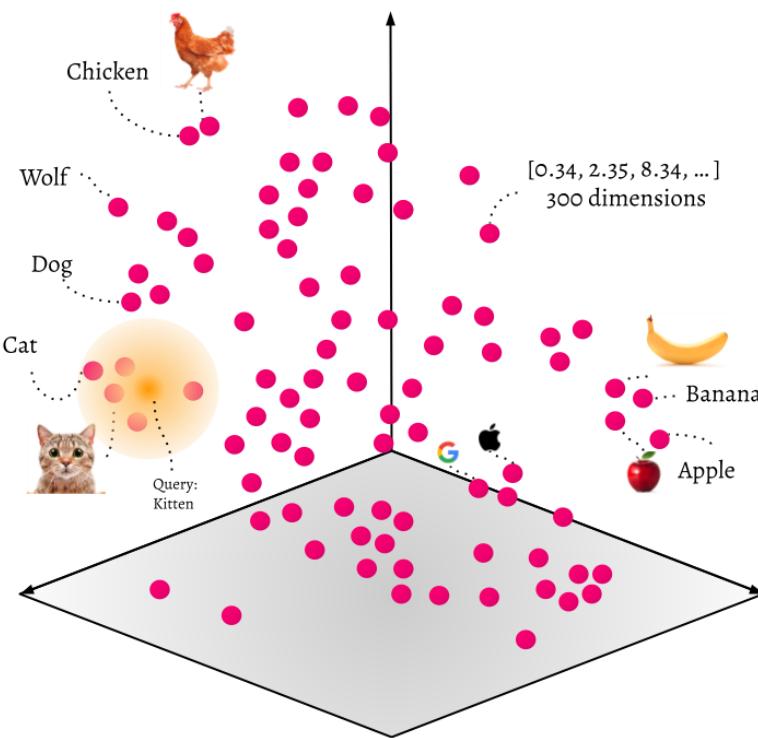
- 4 news posts from the Ch7HD page
- 139 comments
- All are labeled by ChatGPT
- Weighted F1 = ~54.32

Out[27]:

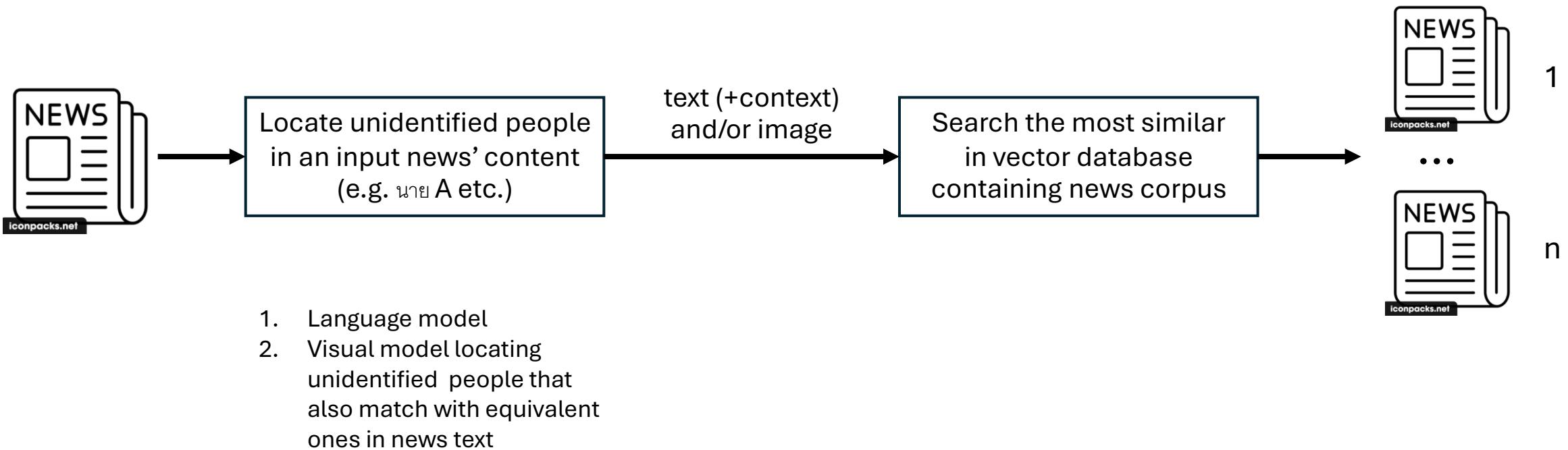
	News title	FB comment	Sentiment
0	ธอส.ปล่อยสินเชื่อคนอยากมีบ้าน ผ่อนนานสุด 40 ปี...	ผ่อนบ้านดอกเบี้ยแพงมาก..	neg
1	ธอส.ปล่อยสินเชื่อคนอยากมีบ้าน ผ่อนนานสุด 40 ปี...	เดินเรื่องยากมั้ย	q
2	ธอส.ปล่อยสินเชื่อคนอยากมีบ้าน ผ่อนนานสุด 40 ปี...	ป่วยหัวครับผ่อนบ้านนานมา12ปีกับน.อ.ส.เหลือเงินต...	neg
3	ธอส.ปล่อยสินเชื่อคนอยากมีบ้าน ผ่อนนานสุด 40 ปี...	เริ่มผ่อนตั้งปี2550	neu
4	ธอส.ปล่อยสินเชื่อคนอยากมีบ้าน ผ่อนนานสุด 40 ปี...	ใบงวดสุดท้ายที่จ่ายให้กับน.อ.ส.สาขาเก้าแสน จ.ส...	neu
...
134	'นึกถ่าย' เพย รู้สึกเจ็บปวด หลัง ตร.ถูกโ久มตี ...	รับไม่ได้อะไรค่ะ รู้ๆกันอยู่	neg
135	'นึกถ่าย' เพย รู้สึกเจ็บปวด หลัง ตร.ถูกโ久มตี ...	ถ้าทำดีแล้วทำต่อไปครับถ้าทำไม่ดีว่ากันด้วยหลัก...	pos
136	'นึกถ่าย' เพย รู้สึกเจ็บปวด หลัง ตร.ถูกโ久มตี ...	ต่าราชตีก็มีแต่น้อยกว่าต่าราช Lewa	neg
137	'นึกถ่าย' เพย รู้สึกเจ็บปวด หลัง ตร.ถูกโ久มตี ...	มันจริง	neu
138	'นึกถ่าย' เพย รู้สึกเจ็บปวด หลัง ตร.ถูกโ久มตี ...	ความเจ็บปวดมันมากกว่าความอับอายหรือครับ ทำน้ำ...	neu

139 rows × 3 columns

2. Searching Unidentified People in News Corpus



2.1 System Overview



Thanks for watching

