

Framework for Analyzing

Party Voting Patterns among Twitter Users

Based On Sentiment Prediction Results
Using BERT Architecture

Kantima Techaphonprasit 6410422027

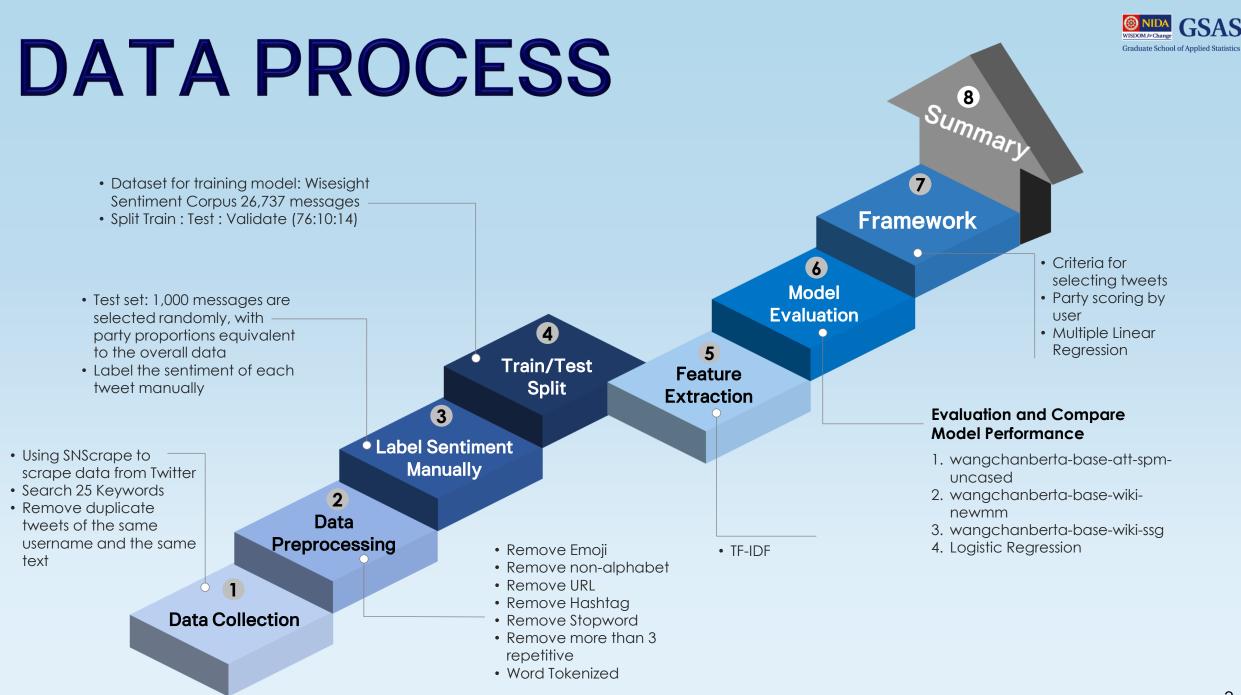
Independent Study | 21st July 2023 Aj. Ohm Sornil





OBJECTIVE

- To examine the relationship between <u>opinions</u>
 <u>expressed on the Twitter</u> platform and the <u>actual</u>
 <u>election results</u> using BERT architecture to classify the
 sentiment of tweets
- To investigate the political parties that users <u>favor or criticize</u> on the Twitter platform
- To examine the <u>correlation between various data factors</u>
 on <u>Twitter</u> (no. of tweets, no. of likes, no. of retweets,
 and sentiment of tweets) and election outcome ranking,
 with a focus on predicting early election results. The
 study seeks to identify key predictors from Twitter data
 to provide valuable insights for political analysis



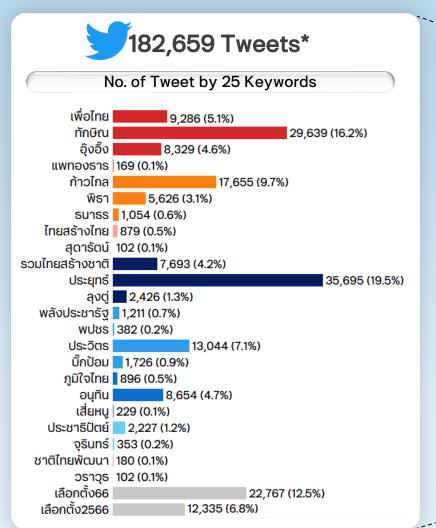
Data Collection

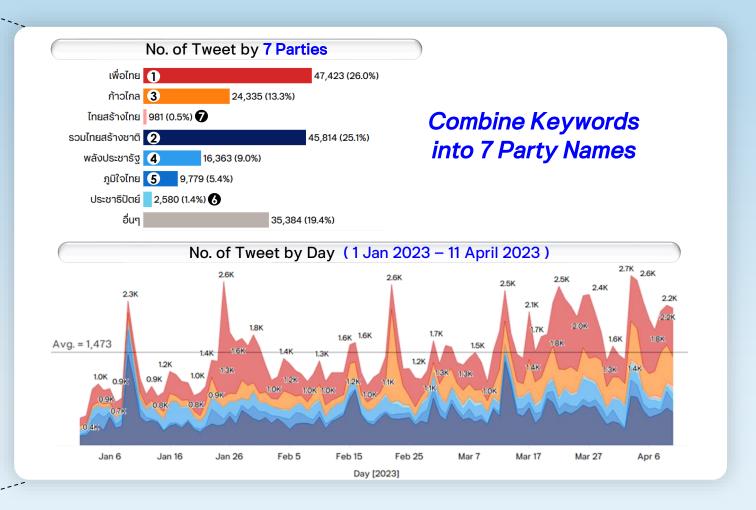
Using Snscrape To Scrape Tweet Data From Twitter



- Period: 1 Jan 2023 10 April 2023 (100 days) (Election Date: 14 May 2023)
- Search 25 Keywords
- 6 Features

(Datetime, Username, Text, Tweet ID, No. of Like, No. of Retweet)





Cleansing Data



- Remove Emoji
- Remove non-alphabet
- Remove URL
- Remove Hashtag
- Remove Stopword
- Remove more than 3 repetitive characters such as ดีมากกก to ดีมาก
- Word Tokenized



Example of Cleaned Texts

Text

"อนุทิน" ปลื้ม ชุมชน กทม.ต้อนรับดี มั่นใจ ตอกเสาเข็มคว้า ส.ส.เมืองหลวงได้แน่ 🎓 https://t.co/NvvL2VGeRU

#พรรคภูมิใจไทย #ภูมิใจไทย #พูดแล้วทำ https://t.co/ssKjReuqB6





🔞 🚱 🚱 ได้ๆ ชอบอ่ะ ยังไงก็ #เพื่อไทย

#เลือกตั้ง66 #รวมไทยสร้างชาติ #ลูงต่อยู่ต่อ บรรยาการมีแต่ความรัก

Text cleaned

อนทินปลื้มชมชนกทมต้อนรับดีมั่นใจตอก เสาเข็มคว้าสสเมืองหลวงได้แน่พรรคภมิ ใจไทยภูมิใจไทยพูดแล้วทำ

ได้ๆชอบอ่ะยังไงก็เพื่อไทย

เลือกตั้งรวมไทยสร้างชาติลงต่อย่ต่อ บรรยาการมีแต่ความรัก

Types of Word Tokens

- newmm: a dictionary-based Thai word tokenizer based on maximal matching from PyThaiNLP.
- spm: a <u>subword-level</u> token from **SentencePiece** library.
- sefr: a ML-based Thai word tokenizer based on Stacked Ensemble Filter and Refine (SEFR) [Limkonchotiwat et al., 2020] based on probabilities from CNN-based deepcut and SEFR tokenizer is loaded with engine="best".

Assign Sentiment Value In A Random Test Set Manually To Evaluate Model Performance



- Test set: 1,000 tweets are selected randomly,
- with party proportions equivalent to the overall data.
- Sentiment values are assigned by human annotators

Example Of Sentiment By Human

Text_cleaned	party	human_senti
วงการดาราต้องยอมแม่หน่อยสุดปังสุดารัตน์พรรคไทย	ไทยสร้างไทย	pos
ไทยสร้างไทยเบอร์เลือกตั้ง	ไทยสร้างไทย	neu
เคลียร์ชัดนโยบายหาเสียงเลือกตั้งเพิ่มเงินในกระ	ไทยสร้างไทย	neu

Label Sentiment



4-class text classification

	Negative	Neutral	Positive	Question
% of Total	35%	49%	14%	3%
No. of Message	348	487	137	28

Wisesight sentiment corpus is used to train models and to measure model performance







Split Train : Test : Validate (76:10:14)

Wisesight Sentiment

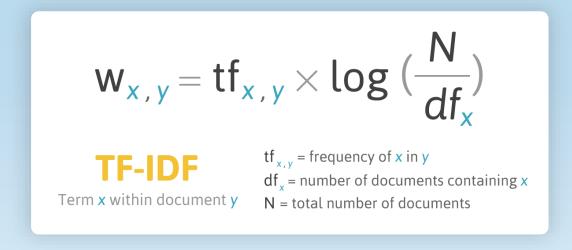
4-class text classification task based on social media posts and tweets.

	Negative	Neutral	Positive	Question
No. of Message	6,823	14,561	4,778	575
% of Total	26%		54%	18% 2%



TF-IDF: Term Frequency-Inverse Document Frequency

- It is a numerical representation of the importance of a term (word) in a collection of documents.
- TF-IDF is commonly used in natural language processing and information retrieval to represent and rank the significance of words in a corpus.





- TF-IDF(t, d) = Term Frequency (TF) * Inverse Document Frequency (IDF)
 Where:
- Term Frequency (TF) is the number of times a term 't' appears in a specific document 'd'.
- TF(t, d) = (Number of occurrences of term t in document d) / (Total number of terms in document d)
- Inverse Document Frequency (IDF) is a measure of how unique or important a term is across the
 entire collection of documents. It helps to down-weight terms that appear frequently in many
 documents, as they tend to be less informative.
- IDF(t) = log_e(Total number of documents / Number of documents containing term t)



Evaluation and Compare Model Performance



wangchanberta-baseatt-spm-uncased

- Largest WangchanBERTa trained on 78.5GB of Assorted Thai Texts with subword tokenizer SentencePiece

3 Wangchan **BERTa** WangchanBERTa sefr spm

wangchanbertabase-wiki-sefr

- WangchanBERTa trained on Thai Wikipedia Dump with word-level tokenizer SEFR

Model **Evaluation**

wangchanbertabase-wiki-newmm

- WangchanBERTa trained on Thai Wikipedia Dump with PyThaiNLP's word-level tokenizer newmm



4 Logistic Regression

Logistic Regression

- Logistic Regression trained on Wisesight Sentiment corpus with with PyThaiNLP's word-level tokenizer



Optimal Model Choice: wangchanberta-base-att-spm-uncased Outperforms in Both Datasets



Dataset

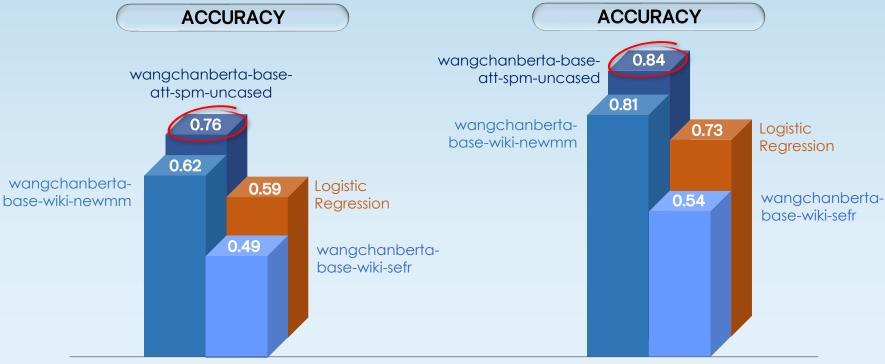
No	Model
1	wangchanberta-base-att-spm-uncased
2	wangchanberta-base-wiki-newmm
3	wangchanberta-base-wiki-sefr
4	Logistic Regression

Tweet Test Set (1,000 messages)											
Accuracy Precision Recall F1-Score											
0.76	0.76	0.76	0.74								
0.62	0.65	0.62	0.59								
0.49	0.24	0.49	0.32								
0.59	0.57	0.59	0.56								

Wisesight Validate Set (3,610 messages)										
Accuracy	Precision	Recall	F1-Score							
0.84	0.84	0.84	0.84							
0.81	0.81	0.81	0.8							
0.54	0.38	0.54	0.38							
0.73	0.73	0.73	0.71							

Model Evaluation





Incorporate Twitter data into the model for sentiment classification.



wangchanberta-baseatt-spm-uncased

Criteria For Selecting Tweets In Analysis Scope



182,659 Tweets

Remove **Duplicate Tweets**

> Related 7 Parties

One-party Mention

Exclude Question **Tweets**

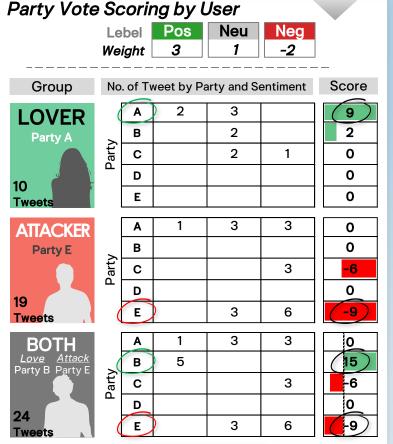
Users who tweeted < 300

Remove unrelated tweets

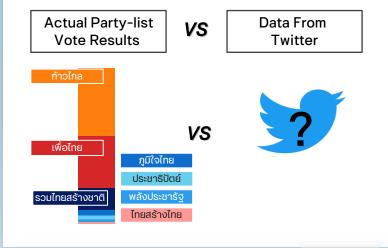


Analyze Tweets in 2 Levels

1) User Level Create User Table #Party #Party #Party Senti Like 22,979 Party Vote Scoring by User Pos Neu Neg



Analyze and contrast the proportion of votes for the 7 parties in the actual party-list vote results with the data extracted from Twitter.



Multiple Linear Regression

 To find the best-fitting linear equation that represents the relationship between the Vote Ranking and these independent variables.





Party Voting Patterns among Twitter Users Analysis

Incorporate Twitter data into the model for sentiment classification.



wangchanberta-baseatt-spm-uncased

Criteria For Selecting Tweets In Analysis Scope

182,659 Tweets

Remove **Duplicate Tweets**

> Related 7 Parties

One-party Mention

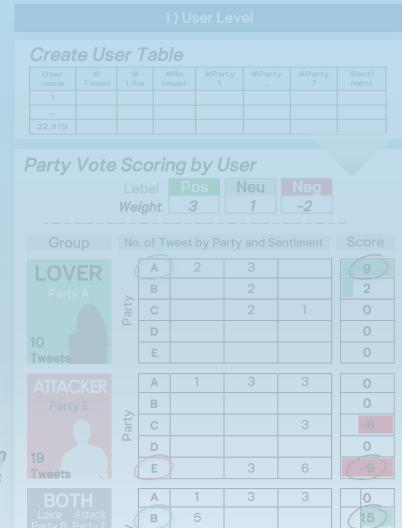
Exclude Question **Tweets**

Users who tweeted < 300

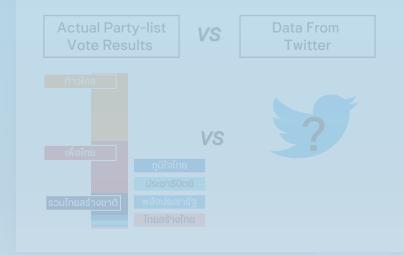
Remove unrelated tweets



Analyze Tweets in 2 Levels



Analyze and contrast the proportion of votes for the 7 parties in the actual party-list vote results with the data extracted from Twitter.



Multiple Linear Regression

 To find the best-fitting linear equation that Ranking and these independent variables.



-6

-9

3

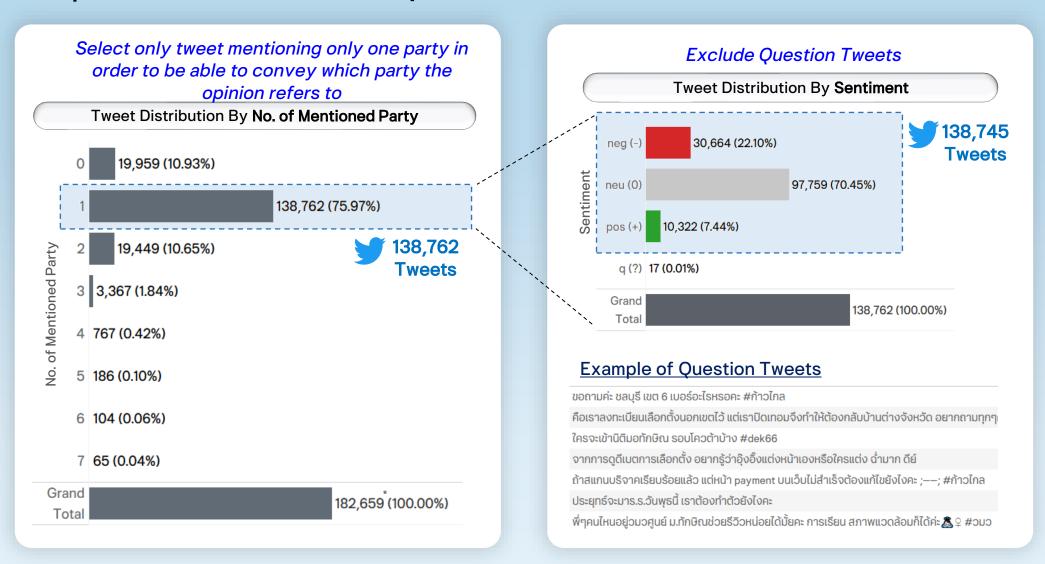
6

3

Criteria For Selecting Tweets In Analysis Scope



Select Only Tweet Mentioning Only One Party In Order To Be Able To Convey Which Party The Opinion Refers To And Exclude Question Tweets



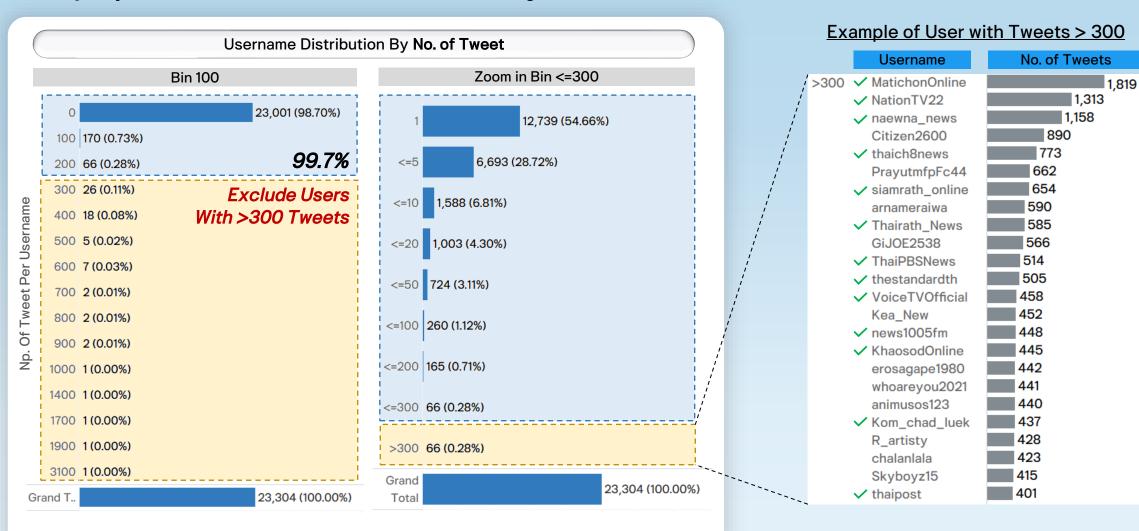
^{*}Already Removed duplicate tweets of the same username and the same text

Criteria For Selecting Tweets In Analysis Scope

WISDOM, for Change Graduate School of Applied Statistics

Focus On Tweets Posted By Users Who Tweeted Fewer Than 300 Tweets

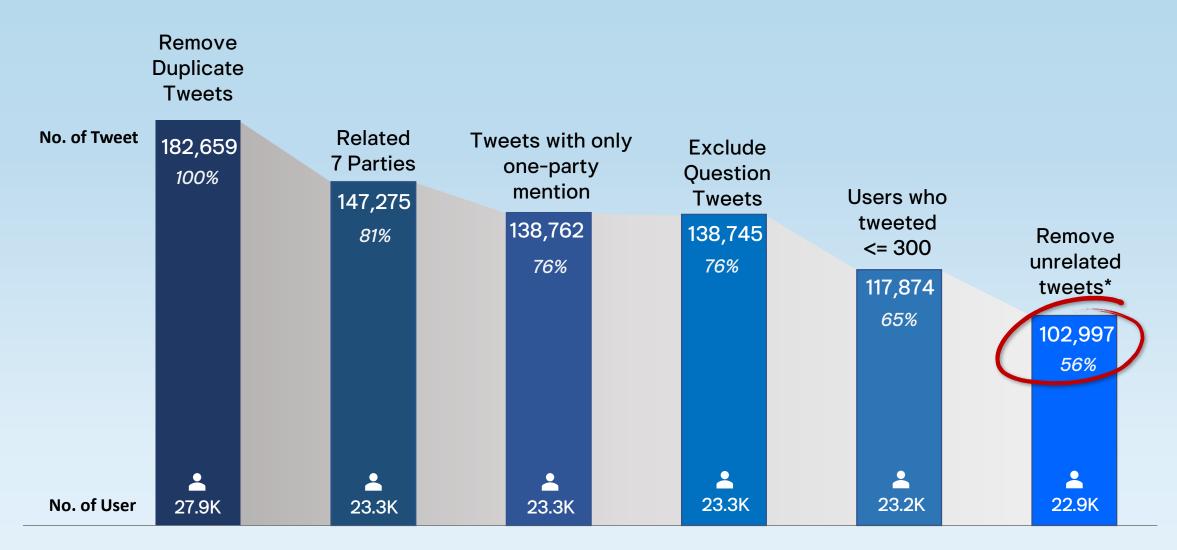
- 99.7% Of All Users Posted Less Than 300 Tweets
- The Majority Of Users With >300 Tweets Are News Organizations



Criteria For Selecting Tweets In Analysis Scope



103K Tweets Will Be Used To Analyze Party Voting Patterns Among Twitter Users



^{*}tweets contained these words "ซ้อปเลยโปรดีๆ|ส่วนลด|งานออนไลน์|พร้อมส่ง|ป้ายยา|ซ้อป|รับหิ้ว|รวมส่ง|โค้ดลด|แถมฟรี|ลดราคาเหลือ|สนใจทัก|ไม่มีขั้นต่ำ|ส่งฟรี|พร้อมโอน|ปล่อยกู้รายเดือน|ราคาชิ้นละ|ส่งต่อเสื้อผ้า|เสื้อเปิดไหล่|ดีลเดือด|ราคาถูก|ดู ดวง|พร้อมส่ง|Spotify|ถอนได้ไม่อั้น|ซื้อตู้เย็น|เสื้อกล้าม|มีงานมาแนะนำ|เกมมือถือ|เสื้อสงกราน|แอนจักรพงษ์|เที่ยวงาน|สอบถามจองคิว|ยำขนมจีน|ครีมบำรุงหน้า|มอทักษิณ,|Google Search Trend|นศ.ทักษิณ|รับสมัครงาน"

Incorporate Twitter data into the model for sentiment classification.



wangchanberta-baseatt-spm-uncased

Criteria For Selecting Tweets In Analysis Scope



182,659 Tweets

Users who tweeted < 300

Remove unrelated tweets



Analyze Tweets in 2 Levels

1) User Level

Create User Table

User name	# Tweet	# Like	#Re tweet	#Party 1	#Party 	#Party 7	Senti ment
1							
22,979							

Party Vote Scoring by User

Pos Neu Neg Weight 3

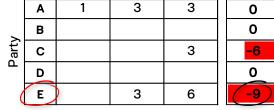
No. of Tweet by Party and Sentiment Group



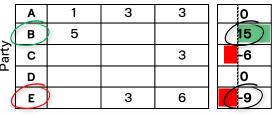
3 9 2 2 С 1 Ε

Score











Multiple Linear Regression

 To find the best-fitting linear equation that represents the relationship between the Vote Ranking and these independent variables.

Analyze and contrast the proportion of votes for the 7 parties

in the actual party-list vote results

with the data extracted from Twitter.

VS



Party Voting Patterns among Twitter Users Analysis

Gather Data From Individual Twitter Users

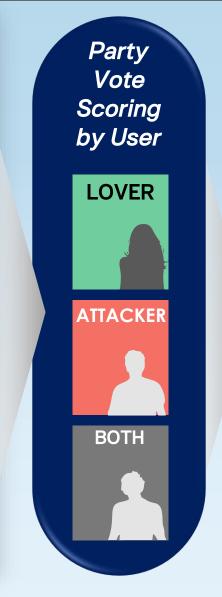
And Calculate Their Loved or Attacked **Towards Political Parties, Or Both.**







	#Tweet	%Sentiment	%Party	#Like	#Retwee
Username (user table			•		
Songpeep112	198	100.00%	100.00%	62	2
THESTATESTIMES	195	95.38%		136	34
agisna52456170	195	89.23%	43.08%	1	1
IIDrama_toxicII	194	56.70% 40.21%	71.65%	15,041	5,855
moveon0012	190	31.05% 62.11%	38.42% 51.05%	2,029	735
Kaoklai31	188	72.87%	87.77%	1,781	1,341
AgoH1987	188	94.68%	27.13% 30.85% <mark>30.32%</mark>	0	2
TeamOrangeParty	184	90.76%	96.74%	51,834	42,209
NoVemBerBitchy	182	76.92%	84.62%	23,257	6,186
PikaPhu_TH	179	36.31% 62.01%	79.33%	171	46
mindandwisdom	177	64.97%	44.07% 49.72%	6,885	2,891
chetawan_th	177	90.96%	100.00%	19,389	9,575
FuwCuYefpxj6mDP	176	43.75% 55.11%	100.00%	417	149
ucratis	175	34.86% 59.43%	33.14% 66.29%	325	197
b_B1B2B3	175	70.86%	29.14% 48.00%	17,485	21,962
haokhaws	172	33.72% 64.53%	72.67%	8,682	3,101
vfnBbPwD300dbKQ	171	55.56% 33.33%	84.80%	2,738	533
calya2007	171	91.23%	43.86%	0	0
ChookitTanawat	168	99.40%	100.00%	10	0
Kezar3659	166	49.40% 50.60%	58.43%	5,156	4,871
Niramon_ST	165	59.39% 35.76%	52.12% 36.36%	3,521	830
JiangWei_TH	164	43.29% 53.66%	55.49% 41.46%	4,173	1,420
RishadanPort	163	77.91%	49.69%	4,835	3,413
Banaranamaprang	163	51.53% 47.24%	84.05%	1,240	486
RungLOVE17	162	51.23% 48.77%	91.98%	862	280
boommari	161	68.94%	57.76%	1,970	925
Thairath_TV	157	96.82%	49.68%	4,168	1,187
Ch7HD	156	99.36%	30.13%	1,153	255
Unrulycat2511	155	79.35%	48.39%	66,769	153,085
shutup2557	155	69.03%	65.81%	19,033	26,228
Peter51787327	155	50.32%	65.16% 30.97%	1,979	563
pawutrxpawutr	155	67.10%	69.03%	1,922	1,165
IRh1almid0C3HkO	154	32.47% 48.70%	35.06% 53.25%	2,639	768



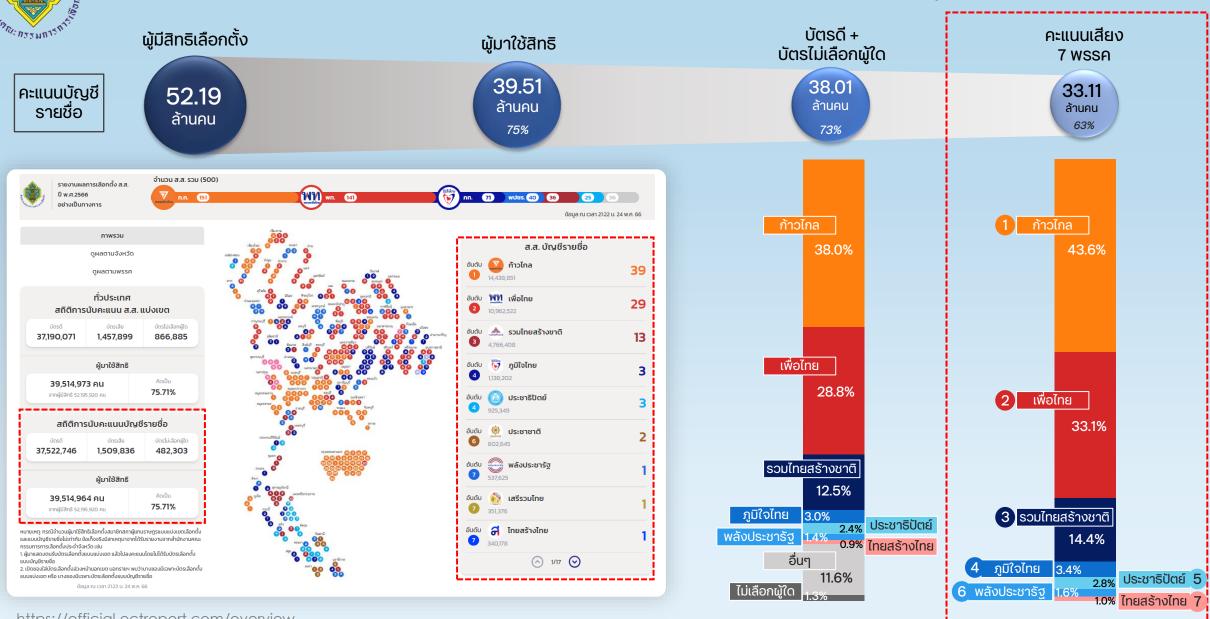
		Group			
Username	Group	LOVE Party	ATTACK Party		
Songpeep112	Lover	 พลังประชารัฐ 	Null		
THESTATESTIMES	Lover	 ภูมิใจไทย 	Null		
agisna52456170	Lover	 เพื่อไทย 	Null		
IIDrama_toxicII	Both	 รวมไทยสร้างชาติ 			
moveon0012	Both	 รวมไทยสร้างชาติ 	 ก้าวไกล 		
Kaoklai31	Lover	ก้าวไกล	Null		
AgoH1987	Lover	 ภูมิใจไทย 	Null		
TeamOrangeParty	Lover	 ก้าวไกล 	Null		
NoVemBerBitchy	Lover	 รวมไทยสร้างชาติ 	Null		
PikaPhu_TH	Both	เพื่อไทย	 รวมไทยสร้างชาติ 		
mindandwisdom	Both	 รวมไทยสร้างชาติ 	 เพื่อไทย 		
chetawan_th	Lover	• ก้าวไกล	Null		
FuwCuYefpxj6mDP	Lover	 รวมไทยสร้างชาติ 	Null		
ucratis	Both	ก้าวไกล	 เพื่อไทย 		
b_B1B2B3	Lover	ก้าวไกล	Null		
haokhaws	Both	เพื่อไทย	ก้าวไกล		
vfnBbPwD3OOdbKQ	Both	เพื่อไทย	 รวมไทยสร้างชาติ 		
calya2007	Lover	เพื่อไทย	Null		
ChookitTanawat	Lover	 พลังประชารัฐ 	Null		
Kezar3659	Both	ก้าวไกล	 รวมไทยสร้างชาติ 		
Niramon_ST	Both	 พลังประชารัฐ 	เพื่อไทย		
JiangWei_TH	Both	 รวมไทยสร้างชาติ 	ก้าวไกล		
RishadanPort	Lover	เพื่อไทย	Null		
Banaranamaprang	Both	 รวมไทยสร้างชาติ 	● เพื่อไทย		
RungLOVE17	Both		เพื่อไทย		
boommari	Both	เพื่อไทย	 รวมไทยสร้างชาติ 		
Thairath_TV	Lover	 รวมไทยสร้างชาติ 	Null		
Ch7HD	Lover	 รวมไทยสร้างชาติ 	Null		
Unrulycat2511	Both	 รวมไทยสร้างชาติ 	🖜 ประชาธิปัตย์		
shutup2557	Both	 ภูมิใจไทย 	 รวมไทยสร้างชาติ 		
Peter51787327	Both	 รวมไทยสร้างชาติ 	เพื่อไทย		
pawutrxpawutr	Both	เพื่อไทย	 รวมไทยสร้างชาติ 		
IRh1almid0C3HkO	Both	ก้าวไกล	 รวมไทยสร้างชาติ 		



รายงานผลการเลือกตั้ง ส.ส. ปี พ.ศ.2566 อย่างเป็นทางการ



โดยสนใจเฉพาะผลคะแนนบัญชีรายชื่อเพื่อใช้ในการเปรียบเทียบกับผลการโหวตของผู้ใช้ Twitter



Party Voting Patterns among Twitter Users Analysis





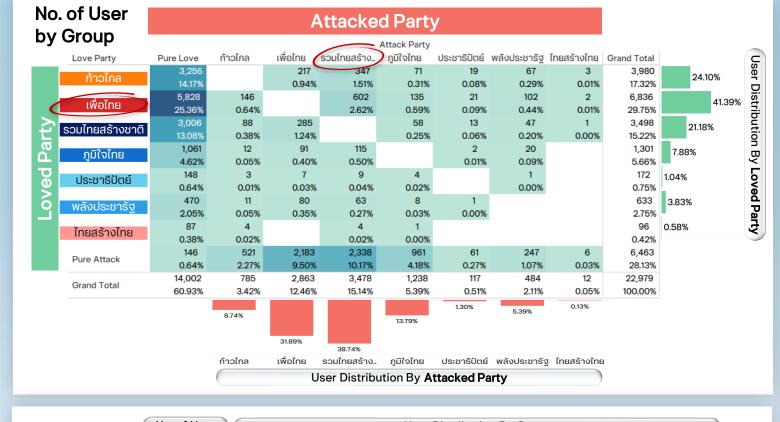
LOVER

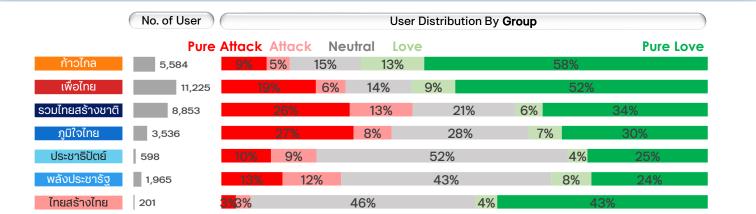
60%

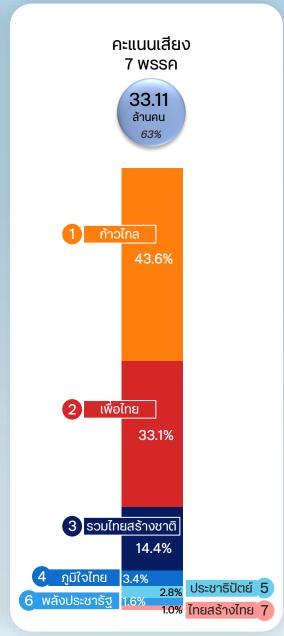
13.9K

ATTACKER28%
6.5K

12% 2.7K







Incorporate Twitter data into the model for sentiment classification.



wangchanberta-baseatt-spm-uncased

Criteria For Selecting Tweets In Analysis Scope



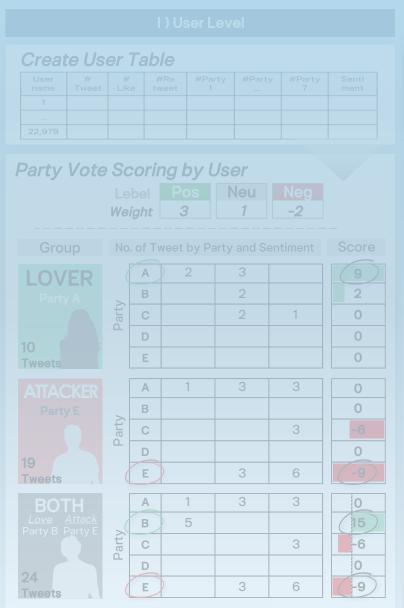
182,659 Tweets

Users who tweeted < 300

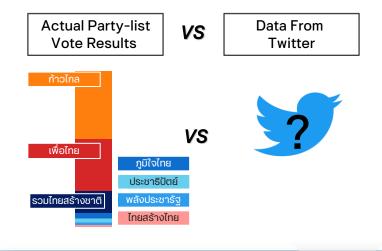
Remove unrelated tweets



Analyze Tweets in 2 Levels



Analyze and contrast the proportion of votes for the 7 parties in the actual party-list vote results with the data extracted from Twitter.



Multiple Linear Regression

• To find the best-fitting linear equation that represents the relationship between the Vote Ranking and these independent variables.

No. of Retweet No. of Tweet No. of Like Sentiment



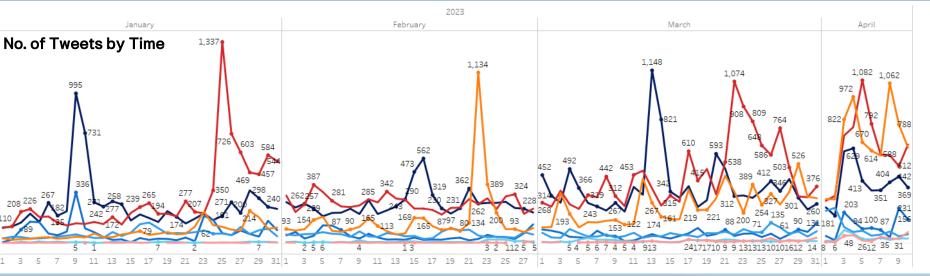
Party Voting Patterns among Twitter Users Analysis

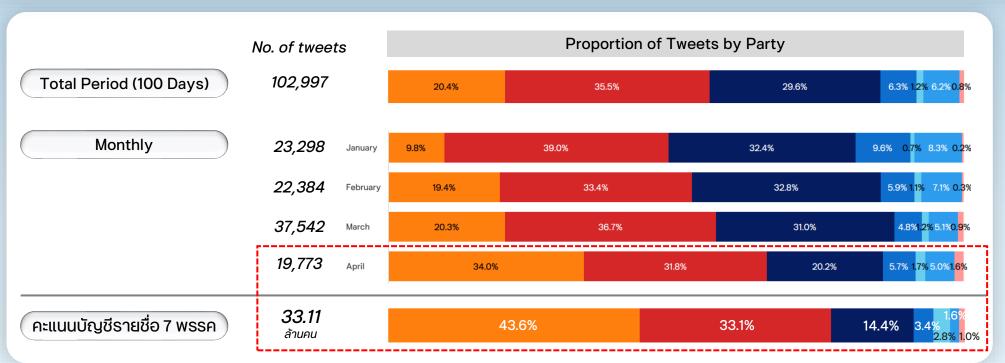
Compare The Proportion Of Votes For The 7 Parties



 The three political parties that are most trending on Twitter are เพื่อไทย, รวมไทย สร้างชาติ และก้าวไกล



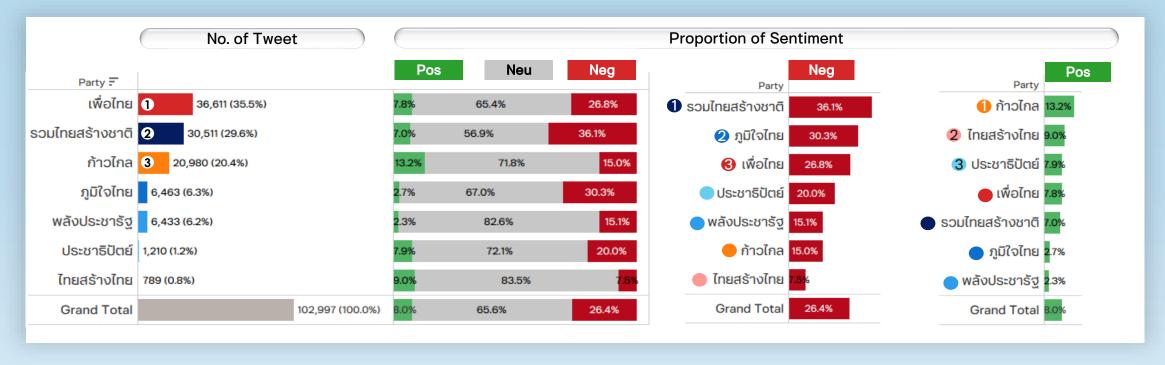




 As election day approaches, the proportion of the number of Tweets is expected to align more closely with the actual election results.

The Number Of Mentions In Tweets Doesn't Guarantee Corresponding Votes; Sentiment Analysis Is Also Essential



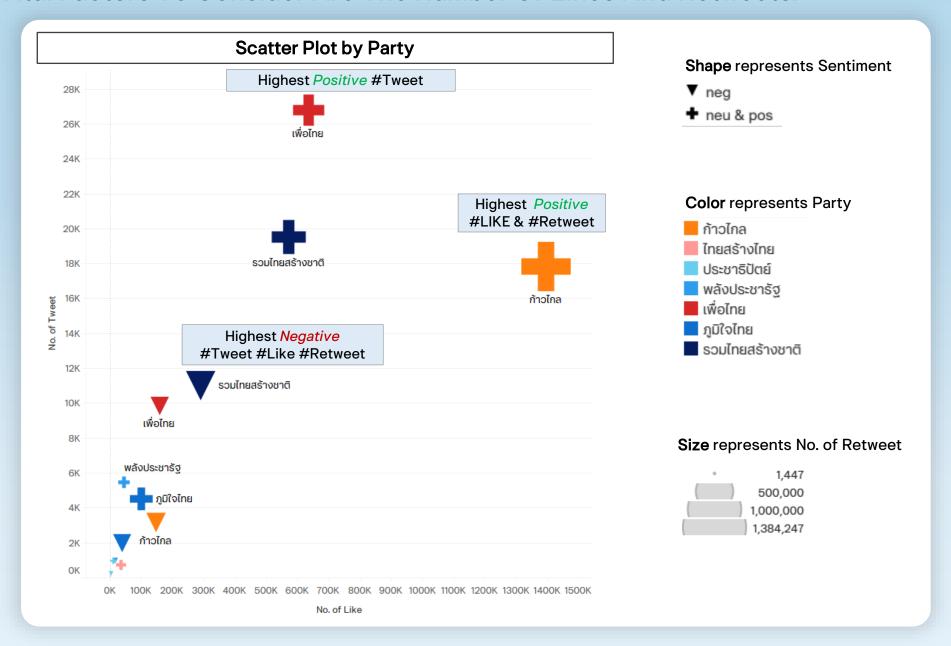


- เพื่อไทย receives the highest number of mentions among users in consideration of the number of Tweets, followed by รวมไทยสร้างชาติ and ก้าวไกล ,respectively
- Although a significant number of mentions in Tweets does not guarantee a corresponding number of votes, it is essential to consider the sentiment associated with the mentions.

- รวมไทยสร้างชาติ, despite ranking second in the number of tweets, has the highest proportion of negative tweets at 36%, followed by ภูมิใจไทย with 30% negative tweets.
- Despite being the third most tweeted party,
 ก้าวไกล has the highest percentage of positive comments at 13%, while ไทยสร้างไทย follows with 9% positive comments.

In Addition To The Number Of Tweets And Sentiment, Other Vital Factors To Consider Are The Number Of Likes And Retweets.





Compare The Proportion of.... Actual Party-list Vote Results vs No. of Tweets vs No. of Likes vs No. of Retweets



• The proportion of actual party-list vote results closely aligns with the proportion of positive and neutral likes.

		Actual Party-list Vote Results	All	Sentiment		Posi	tive + Neut	ral		Negative	
	Party	Actual	#Tweet	#Like	#Retweet	#Tweet	#Like	#Retweet	#Twe	t #Like	#Retweet
	ก้าวไกล	14,438,851	20,980	1,544,421	1,609,575	17,827	1,395,728	1,384,247	3,15	148,693	225,328
)er	เพื่อไทย	10,962,522	36,611	797,018	755,372	26,787	636,679	553,460	9,824	160,339	201,912
ᄪ	รวมไทยสร้างชาติ	4,766,408	30,511	865,664	1,170,978	19,501	573,706	659,274	11,010	291,958	511,704
Actual Number	ภูมิใจไทย	1,138,202	6,463	140,938	488,411	4,506	100,226	293,550	1,95	40,712	194,861
tua	ประชาธิปัตย์	925,349	1,210	14,309	34,299	968	10,330	21,175	242	2 3,979	13,124
Ac	พลังประชารัฐ	537,625	6,433	62,218	93,827	5,461	45,798	74,176	972	16,420	19,651
	ไทยสร้างไทย	340,178	789	37,659	58,568	730	36,487	57,121	59	1,172	1,447
	Total	33,109,135	102,997	3,462,227	4,211,030	75,780	2,798,954	3,043,003	27,21	7 663,273	1,168,027
	Party	Actual	#Tweet	#Like	#Retweet	#Tweet	#Like	#Retweet	#Twe	t #Like	#Retweet
	ก้าวไกล	43.6%	20.4%	44.6%	38.2%	23.5%	49.9%	45.5%	11.6	% 22.4%	19.3%
	เพื่อไทย	33.1%	35.5%	23.0%	17.9%	35.3%	22.7%	18.2%	36.1	% 24.2%	17.3%
Total	รวมไทยสร้างชาติ	14.4%	29.6%	25.0%	27.8%	25.7%	20.5%	21.7%	40.5	% 44.0%	43.8%
Į Į	ภูมิใจไทย	3.4%	6.3%	4.1%	11.6%	5.9%	3.6%	9.6%	7.2	% 6.1%	16.7%
% of	ประชาธิปัตย์	2.8%	1.2%	0.4%	0.8%	1.3%	0.4%	0.7%	0.9	% 0.6%	1.1%
	พลังประชารัฐ	1.6%	6.2%	1.8%	2.2%	7.2%	1.6%	2.4%	3.6	% 2.5%	1.7%
	ไทยสร้างไทย	1.0%	0.8%	1.1%	1.4%	1.0%	1.3%	1.9%	0.2	% 0.2%	0.1%
	Total	100%	100%	100%	100%	100%	100%	100%	100	% 100%	100%

Utilize Multiple Linear Regression To Establish The Relationship Between Vote Ranking And The Independent Variables

Independent Variables

WISDOM & Change

Graduate School of Applied Statistics

No. of Tweet

No. of Retweet

No. of Like

Sentiment

Multiple Linear Regression

- MLR is a statistical technique used to model the relationship between a dependent variable and two or more independent variables.
- In multiple linear regression, the goal is to find the best-fitting linear equation that represents the relationship between the dependent variable and the independent variables. The aim of multiple linear regression is to estimate the regression coefficients that minimize the sum of squared differences between the predicted values and the actual values of the dependent variable.
- This estimation is typically done using a method called ordinary least squares (OLS).

The equation takes the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon$$

Where:

- •Y represents the dependent variable or the variable being predicted.
- • X_1 , X_2 , ..., X_p represent the independent variables or predictors.
- • β_0 , β_1 , β_2 , ..., β_p are the regression coefficients or the parameters to be estimated. β_0 is the intercept term, and β_1 , β_2 , ..., β_p are the coefficients corresponding to the independent variables.
- • ϵ represents the error term or residual, which accounts for the variability not explained by the model.

Multiple Linear Regressions With The Smallest Mean Squared Error (MSE)



Are Equations With 3 Independent Variables (#Positive Retweet, #Negative Retweet, and #Neutral Retweet)

		1		Independent Variables													
	Dependent Variable		All Sentiment			Pos				Neu		Neg			Pos + Neutral		
	Party	Score Rank	#Tweet	#Like	#Retweet	#Tweet_ Pos	#Like_ Pos	#Retweet_ Pos	#Tweet_ Neu	#Like_ Neu	#Retweet_ Neu	#Tweet_ Neg	#Like_ Neg	#Retweet_ Neg	#Tweet_ Pos+Neu	#Like_ Pos+Neu	#Retweet_ Pos+Neu
ē	ก้าวไกล	1	20,980	1,544,421	1,609,575	2,767	255,837	242,641	15,060	1,139,891	1,141,606	3,153	148,693	225,328	17,827	1,395,728	1,384,247
quin	เพื่อไทย	2	36,611	797,018	755,372	2,845	82,358	49,519	23,942	554,321	503,941	9,824	160,339	201,912	26,787	636,679	553,460
5	รวมไทยสร้างชาติ	3	30,511	865,664	1,170,978	2,150	95,019	109,344	17,351	478,687	549,930	11,010	291,958	511,704	19,501	573,706	659,274
	ภูมิใจไทย	4	6,463	140,938	488,411	177	13,790	44,180	4,329	86,436	249,370	1,957	40,712	194,86	4,506	100,226	293,550
tua	ประชาธิปัตย์	5	1,210	14,309	34,299	95	550	81	873	9,780	21,094	242	3,979	13,124	968	10,330	21,175
AC	พลังประชารัฐ	6	6,433	62,218	93,827	148	244	187	5,313	45,554	73,989	972	16,420	19,651	5,461	45,798	74,176
	ไทยสร้างไทย	7	789	37,659	58,568	71	13,462	24,157	659	23,025	32,964	59	1,172	1,44	730	36,487	57,121
		<u>Total</u>	102,997	3,462,227	4,211,030	8,253	461,260	<u>470,109</u>	<u>67,527</u>	2,337,694	2,572,894	27,217	663,273	1,168,027	<u>75,780</u>	2,798,954	3,043,003
		=					•			•			•				
														,			
	ก้าวไกล	1	0.2037	0.4461	0.3822	0.3353	0.5546	0.5161	0.2230	0.4876	0.4437	0.1158	0.2242	0.1929	0.2352	0.4987	0.4549
tal	เพื่อไทย	2	0.3555	0.2302	0.1794	0.3447	0.1786	0.1053	0.3546	0.2371	0.1959	0.3610	0.2417	0.1729	0.3535	0.2275	0.1819
,0	รวมไทยสร้างชาติ	3	0.2962	0.2500	0.2781	0.2605	0.2060	0.2326	0.2569	0.2048	0.2137	0.4045	0.4402	0.4381	0.2573	0.2050	0.2167
<u>_</u>	ภูมิใจไทย	4	0.0627	0.0407	0.1160	0.0214	0.0299	0.0940	0.0641	0.0370	0.0969	0.0719	0.0614	0.1668	0.0595	0.0358	0.0965
0	ประชาธิปัตย์	5	0.0117	0.0041	0.0081	0.0115	0.0012	0.0002	0.0129	0.0042	0.0082	0.0089	0.0060	0.0112	0.0128	0.0037	0.0070
%	พลังประชารัฐ	6	0.0625	0.0180	0.0223	0.0179	0.0005	0.0004	0.0787	0.0195	0.0288	0.0357	0.0248	0.0168	0.0721	0.0164	0.0244
	ไทยสร้างไทย	7	0.0077	0.0109	0.0139	0.0086	0.0292	0.0514	0.0098	0.0098	0.0128	0.0022	0.0018	0.0012	0.0096	0.0130	0.0188
		<u>Total</u>	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

 Experiment with creating multiple linear regressions by importing different independent variables to estimate the regression coefficients that minimize the Mean Squared Error (MSE).

Mean Squared Error (MSE)

#Independent Variable			1	2	3		
Sentiment	-	Pos	Neg, Pos+Neu	Pos, Neu, Neg			
#Tweet	1.4834999	0.7800148	1.3033971	2.392686	1.2060886	0.991226016	0.733640916
#Like	0.7898197	1.2905683	0.8623104	1.9466772	0.9284993	0.768494314	0.500350176
#Retweet	0.78955	1.5932825	0.7701433	2.2920527	0.8908814	0.779889207	0.256295584



#Negative Retweet Has No Effect To Party Vote Score Ranking



MSE: 0.2563

Multiple Linear Regression

Objective: Are number of retweets in Twitter in each sentiment predictive of party vote score ranking?

Data collection:

	Score Rank	#Retweet_ Pos	#Retweet_ Neu	#Retweet_ Neg	Predict
ก้าวไกล	1	0.5161	0.4437	0.1929	1.05
เพื่อไทย	2	0.1053	0.1959	0.1729	1.92
รวมไทยสร้างชาติ	3	0.2326	0.2137	0.4381	2.82
ภูมิใจไทย	4	0.0940	0.0969	0.1668	4.51
ประชาธิปัตย์	5	0.0002	0.0082	0.0112	5.88
พลังประชารัฐ	6	0.0004	0.0288	0.0168	5.30
ไทยสร้างไทย	7	0.0514	0.0128	0.0012	6.52
	<u>Total</u>	1.0000	1.0000	1.0000	

Estimation of Coefficients (OLS):

Test Hypothesis:

Assumptions 1: Linear Model H_0 : $\beta i = 0$ H_1 : $\beta i \neq 0$ > rmodel <- aov(Party~ MLR\$V7+MLR\$V10+MLR\$V13, data = dataf)</pre> > summary(rmodel) Df Sum Sq Mean Sq F value Pr(>F) Alpha 1 16.847 16.847 28.171 0.0131 * <0.05 Reject Ho #RT Pos #RT Neu 1 9.057 9.057 15.145 0.0301 * < 0.05 Reject Ho >0.05 Do not reject Ho #RT Nea 1 0.302 0.302 0.505 0.5285 Residuals 3 1.794 0.598

- At the 0.05 level of significance, RT_Pos and RT_Neu has effect to party vote score ranking
- At the 0.05 level of significance, RT_Neg has no effect to party vote score ranking

X

Assumptions 2: Independency

H1: rho≠0 > durbinWatsonTest(model) lag Autocorrelation D-W Statistic p-value

lag Autocorrelation D-W Statistic p-value
1 0.04917327 1.770989 0.332
Alternative hypothesis: rho != 0 > Alpha 0.05
Do not reject Ho

• At the 0.05 level of significance, Random errors are independent.

Assumptions 3: Mean (error) = 0

 $\mu_{\varepsilon} = 0$

Follow immediately from LS method of estimation



H0: rho = 0



Score Rank

= 6.1211

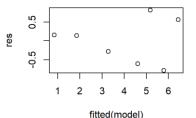
+14.6953 (#Retweet_Pos)

-27.7285 (#Retweet_Neu)

-1.8143 (#Retweet_Neg)

Assumptions 4: Homogeneity of Variance

Assumptions 6 : No outliers

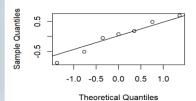


- Random errors are homogeneity of variance
- No outlier



Assumptions 5 : Normal Distribution

Normal Q-Q Plot



- Random errors follow line
- : Random errors are Normal Distribution





MLR With 2 Independent Variables Is Consistent With All Assumptions. (#Positive Retweet & #Neutral Retweet)



MSE: 0.2994

Multiple Linear Regression

Objective: Are number of retweets in Twitter in each sentiment predictive of party vote score ranking?

Data collection:

	Score Rank	#Retweet_ Pos	#Retweet_ Neu	Predict
ก้าวไกล	1	0.5161	0.4437	0.85
เพื่อไทย	2	0.1053	0.1959	1.86
รวมไทยสร้างชาติ	3	0.2326	0.2137	3.28
ภูมิใจไทย	4	0.0940	0.0969	4.61
ประชาธิปัตย์	5	0.0002	0.0082	5.79
พลังประชารัฐ	6	0.0004	0.0288	5.18
ไทยสร้างไทย	7	0.0514	0.0128	6.43
	<u>Total</u>	1.0000	1.0000	

Estimation of Coefficients (OLS):

F-statistic: 24.71 on 2 and 4 DF. p-value: 0.005605

> durbinWatsonTest(model)

Assumptions 2:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

Score Rank

= 6.0244

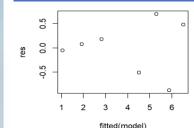
+15.32 (#Retweet_Pos)

-29.4906 (#Retweet_Neu)

Assumptions 4:

Homogeneity of Variance

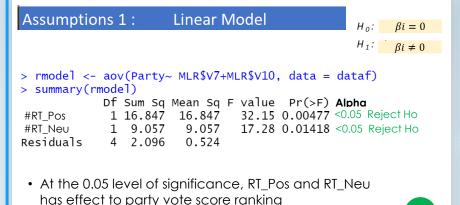
Assumptions 6 : No outliers



- : Random errors are homogeneity of variance
- No outlier



Test Hypothesis:



Assumptions 3: Mean (error) = 0

Random errors are independent.

0.2150905

• At the 0.05 level of significance,

Alternative hypothesis: rho != 0

 $\mu_{\underline{\varepsilon}} = 0$

Follow immediately from LS method of estimation

Independency

1.405625

lag Autocorrelation D-W Statistic p-value



H0: rho = 0

H1: rho ≠ 0

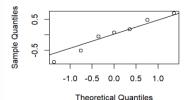
0.204

> Alpha 0.05

Do not reject Ho

Assumptions 5: Normal Distribution

Normal Q-Q Plot



- Random errors follow line
- Random errors are Normal Distribution









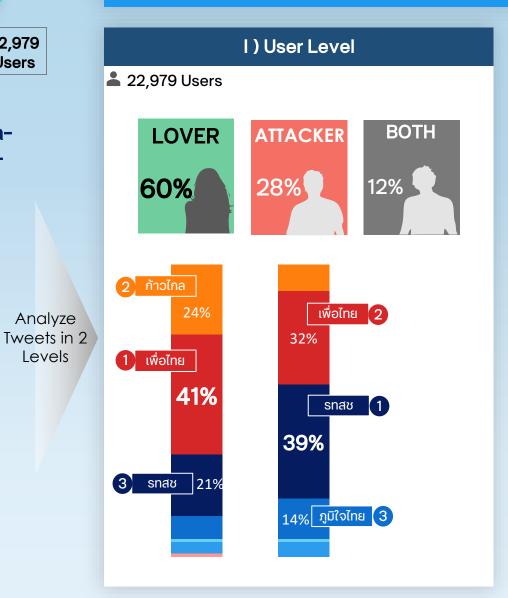




wangchanbertabase-att-spmuncased



is applied to classify sentiment



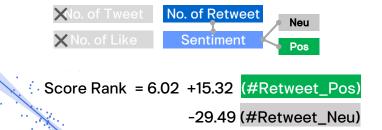
II) Tweet Level

Mere tweet count is insufficient for predicting the selection outcome; considering the number of likes and retweets, along with sentiment analysis, enhances prediction accuracy.

Party with Max Count					
Sentiment	All	All Pos Neu Neg			
#Tweet	เพื่อไทย	เพื่อไทย	เพื่อไทย	รทสช	
#Like	ก้าวไกล	ก้าวไกล	ก้าวไกล	รทสช	
#Retweet	ก้าวไกล	ก้าวไกล	ก้าวไกล	รทสช	

Multiple Linear Regression

The relationship between a Twitter user's number of positive and neutral retweets and a political party's vote ranking can be described by a linear correlation, as expressed in this equation.







 No tweet information is available around the time of the election. (11 April 2023 – 14 May 2023)

Future Work

- In this analysis, emojis were excluded from the text. However, it is important to acknowledge that emojis can convey emotions and should be taken into consideration for recognizing sentiments.
- · This framework should be backtested with other data sets.





Code_Colab

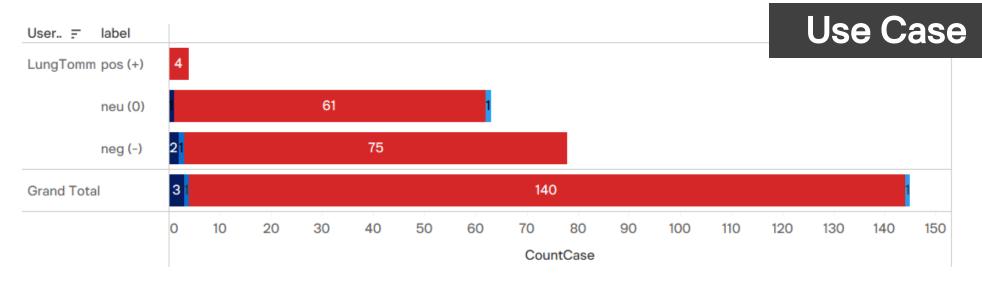
https://colab.research.google.com/drive/1caasFG6ha3rR9Wu3g6tsjU-ZhT2TmSMw?usp=sharing



Literature Review

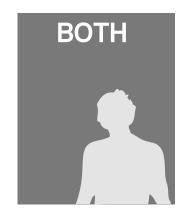
		3			Metric	Performane
2021	PREDICTING THE 2020 US PRESIDENTIAL ELECTION WITH TWITTER	VADER sentiment and structural data with aggregate polling, a time series analysis	2020 US Presidential election	Twitter A) 1.7 million tweet data set from Kaggle user B) scraping historical tweets from multiple prominent accounts. (57 accounts,3200 tweets) from presidential candidates, news organizations, political pundits, and politicians	MAE	This method performed worse than its baseline of polling predictions
2016	Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter	Dictionary Based, Naive Bayes and SVM	2016 India general state election	Twitter - 42,235 tweets (1 month during the campaigning period)	Accuracy	- SVM predicted a 78.4% chance that the BJP would win - As it turned out, BJP won 60 out of 126 constituencies in the 2016 general election
2022	Sentiment Analysis on Twitter for the Major German Parties during the 2021 German Federal Election	transformer-based models (BERT)	2021 German federal election	Twitter 58,000 tweets from 7 parties	Accuracy, macro F1, weighted f1 score	- BERT model achieved an accuracy of 93.3% and macro f1 score of 93.4%.
2022	Understanding Social Media Behavior in Philippines Presidential Election using Natural Language Processing	SVM (to classify sentiment) - LDA (to discover hidden topics from the given corpus) - K-means (to clusters topics)	2022 Philippines Presidential election	Facebook Posts from 300 users with 3,210 posts	Accuracy	SVM accuracy 73% -The topic models and cluster models were evaluated using a human annotator (50.1% which shows an average level of the quality of the topic models)
2022	A Semi-Supervised Approach to Sentiment Analysis of Tweets during the 2022 Philippine Presidential Election	Multinomial Naïve Bayes model	2022 Presidential Election in the Philippines	Twitter - 28 February to 8 May 2022 (2 months) (the campaign period) - 114,851 tweets	Accuracy	an accuracy rate of 84.83% higher than the researchers' previous study's accuracy rate.





Username	label	ATTACK Party	Text
LungTomm	pos (+)	เพื่อไทย	@HS2_QRK5_s9 พรรคเพื่อทักษิณส่งหมาลงควายสีแดงมันก็เลือก
LungTomm	pos (+)	เพื่อไทย	@forevertum ท่านทักษิณชอบคนแบบนี้มาก หลอกไปเผาบ้านเกิดหลอกมาแล้ว หลอกไปตายทำมาแล้ว สร้างให้เป็นไพร่ด้วยศำ
LungTomm	pos (+)	เพื่อไทย	@forevertum ท่านทักษิณชอบคนประเภทนี้มากครับ ท่านเคยพูดว่าคนที่ไม่ชอบอ่านไม่ชอบค้นคว้าเป็นคนที่หลอกใช้ง่ายครับ
LungTomm	pos (+)	เพื่อไทย	@hEtlowCXN399XXW @PheuThaiParty ท่านทักษิณชอบคนโง่เพราะปกครองง่าย https://t.co/MGqbYj1JBe

Lunglonum	+	/	-	
เพื่อไทย	4	61	75	-77
รวมไทยสร้างช	าติ	1	2	-3
ก้าวไกล				ĺ
พลังประชารัฐ				
ภูมิใจไทย			1	-2
ประชาธิปัตย์				
ไทยสร้างไทย				
			145	







Username	ATTACK Party		ATTACK Party Text			
JiangWei_TH	neu (0)	ก้าวไกล	®SiaJampathong ใช่ที่นี่มั้ยครับ ที่ว่า กดขี่ กิดกันการรวมตัวเป็นสหภาพแรงงาน หวังว่าจะได้รับเกียรติจากคุณเชีย มา			
JiangWei_TH	neu (0)	ก้าวไกล	@ongpadipat เป็นโอกาสอันดีที่ชาวพิษโลกจะได้เจอหน้า สส ของท่านแล้ว หลังจากท่านแทบไม่เคยเห็นหน้ามันเลย ตร			
JiangWei_TH	neu (0)	ก้าวไกล	@Wit_TheRipper @ChawalitKongpet เห็นว่ามีดีกรีด๊อกเตอร์ ก็เลยคาดหวัง แม่แต่ด๊อกเตอร์ ก็คงคาดหวังอะไรไม่ไ			
JiangWei_TH	neu (0)	ก้าวใกล	นั่นใง พิธาเอ็งชะตาขาดแล้ว เค้าเตรียมหัวหน้าใหม่ไว้แล้ว #ก้าวไกล https://t.co/b6pjAyS8lw			
JiangWei_TH	neu (0)	ก้าวไกล	แต่โปรดระวัง วันหนึ่ง เพื่อนคุณอาจได้ใช้บริการลูกสาวของคุณ วันหนึ่ง เพื่อนบ้านคุณ อาจได้ใช้บริการ เมียของคุณ			
JiangWei_TH	neu (0)	ก้าวไกล	@IIDrama_toxicII จะบรรยายยังใงหมด มันเยอะมากนะ A41รีม ยังไม่รู้จะพอมั้ย #ก้าวไกล			
JiangWei_TH	neu (0)	ก้าวไกล	@lawyerJammy ยืนยันก็ดีครับ จะได้เป็นหลักฐานอีกขึ้น ให้ กกต พิจารณา ก็ว่ากันไปตามหลักฐานและเจตนาครับ			

Username	label	LOVE Part	y		
JiangWei_TH	pos (+)	รวมไทยสร้างชาติ	8 ปิของลุงดู่ เข้ามาจัดการ รถมือสอง ที่ผุพัง ไม่มีใครเหลียวมอง จนกลายเป็น เฟอรารี่ ได้ในวันนี้ จนมีโจร ขโมย จ้องตาเป็นมัน ลุงดู่ เก่ง จริงๆครับ ผมนับถือ #รวมไทยสร้างชาติ #ลุงดู่ #รวมไทยสร้างชาติ ๒๒ #รทสช		
JiangWei_TH	pos (+)	รวมไทยสร้างชาติ	carmen167057816 #ทีมลุงคู่ ด้วยครับ #รทสช #รวมไทยสร้างชาติ		
JiangWei_TH	neu (0)	รวมไทยสร้างชาติ	⊉ApWtboy แหม่ โยงไปถึงประยุทธ์จนได้เนาะ		
JiangWei_TH	neu (0)	รวมไทยสร้างชาติ	พรุ่งนี้ #ลุงตู่ ลงไปภูเก็ต ไปตรวจศูนย์ประชุม คนภูเก็ต อย่าลืมมาดูคลิปนี้กันนะครับ #ทีมลุงตู่ #รทสช #รวมไทยสร้างชาติ		
JiangWei_TH	neg (-)	รวมไทยสร้างชาติ	@Bobthelove2 ถึงแม้หาก พล อ ประยุทธ์ จะประกาศเลิกเล่นการเมือง ผมก็ไม่กลับไปเลือกคุณแล้ว ผมเห็นคุณในฐานะผู้นำมาแล้ว เข็ดแล้วครับ #อภิสิทธิ์ #ปชป		



Use Case

