

# วิเคราะห์เพจ Facebook

## Physics Blueprint พลิกส์&ความนับด้วยศาสตร์ by พี่ต้ว

กำหนดความปังของยอด engagement ในแต่ละโพสต์



Presented by  
JOBJAB

# Data Collection

**Facebook Scrape -> ข้อมูลที่เลือกมาใช้มีทั้งหมด 7 features 341 posts  
( Time frame 24/02/2021 - 24/03/2022 )**

	date	time	type	post_text	comments	shares	reaction_count	
0	2022-03-25	16:01:00	image		NaN	1	0	638
1	2022-03-22	10:00:05	imageandcaption	เป็นสมการที่ทำแทนกันไม่ได้จริงๆ...แล้ว	14	180	1166	
2	2022-03-21	15:43:10	status		NaN	0	0	18
3	2022-03-19	11:00:01	imageandcaption	💥 พลาดไม่ได้ พื้นที่ความรู้คู่กับความสนุกไป❤️นี่...	0	9	44	
4	2022-03-19	05:47:10	status	พิสิกส์สามัญ จาก สสวท. กข้อสอบคุณภาพครับ	1	68	395	
...	...	...	...	...	...	...	...	
336	2021-03-02	15:19:07	status	🔴 มีเวลา空oy แต่ต้องใช้คะแนนเยอะໄก ❤️ พีตัวคัดม...	0	26	131	
337	2021-03-02	12:24:35	status		NaN	0	30	212
338	2021-03-01	13:18:31	video	⭐ ตั้งเป้าพิชิตพิสิกส์สามัญ 60+!! 🎉 แต่ยังเก็บ...	0	31	109	
339	2021-02-27	12:16:34	status	🔥 ใครอยากรเข้าวิศวะจุฬาต้องดู!! 🌞 ❤️ คะแนนขั้นต่ำว...	5	70	170	
340	2021-02-25	12:02:00	video	🔔 พธน์มี live สอนฟรีจากพีตัวรู้ยัง?! 🎙 1...	175	97	122	

341 rows × 7 columns

# Data Transformation

numeric - datetime

	date	time
0	2022-03-25	16:01:00
1	2022-03-22	10:00:05
2	2022-03-21	15:43:10
3	2022-03-19	11:00:01
4	2022-03-19	05:47:10
...	...	...
336	2021-03-02	15:19:07
337	2021-03-02	12:24:35
338	2021-03-01	13:18:31
339	2021-02-27	12:16:34
340	2021-02-25	12:02:00

month, day, hour

	post_hour	post_month	dow
0	16	3	4
1	10	3	1
2	15	3	0
3	11	3	5
4	5	3	5
...	...	...	...
336	15	3	1
337	12	3	1
338	13	3	0
339	12	2	5
340	12	2	3

341 rows × 5 columns

category

	post_hour	post_month	dow
0	Evening		1 Weekday
1	Morning		1 Weekday
2	Afternoon		1 Weekday
3	Morning		1 Weekend
4	Late_Night		1 Weekend
...	...	...	...
336	Afternoon		1 Weekday
337	Afternoon		1 Weekday
338	Afternoon		1 Weekday
339	Afternoon		1 Weekend
340	Afternoon		1 Weekday

341 rows × 3 columns

# Data Transformation

string - text

count\_text : numeric,  
wordcloud : category

	post_text
0	Nan
1	เป็นสมการที่ทำแทนกันไม่ได้จริงๆ... และ
2	Nan
3	💥พลาดไม่ได้ พื้นที่ความรู้คู่กับความสนุกໄກ❤️นี่...
4	ฟิสิกส์สามัญ จาก สสวท.ໄກข้อสอบคุณภาพครัวบ
...	...
336	sos มีเวลาอ้อย แต่ต้องใช้คะแนนเยอะໄก❤️ พีตัวคัดม...
337	Nan
338	⭐️ตั้งเป้าพิชิตฟิสิกส์สามัญ 60+!!ໄກ👉 แต่ยังเก็บ...
339	🔥ใครอยากรเข้าวิศวะจุฬาต้องอุป!!ໄก❤️ คะแนนขึ้นต่อว...
340	🔔พุธนี้มี live สอนฟรีจากพีตัวนรรย়ং?!ໄກ⌚ 1...

341 rows × 1 columns



	count_text	have_Tha	have_samak	have_Peang	have_teaknik
0	0	0	0	0	0
1	34	0	0	0	0
2	0	0	0	0	0
3	170	0	0	0	0
4	36	0	0	0	0
...	...	...	...	...	...
336	1484	1	1	1	1
337	0	0	0	0	0
338	1665	1	1	1	1
339	354	0	1	0	0
340	202	0	0	0	0

341 rows × 14 columns

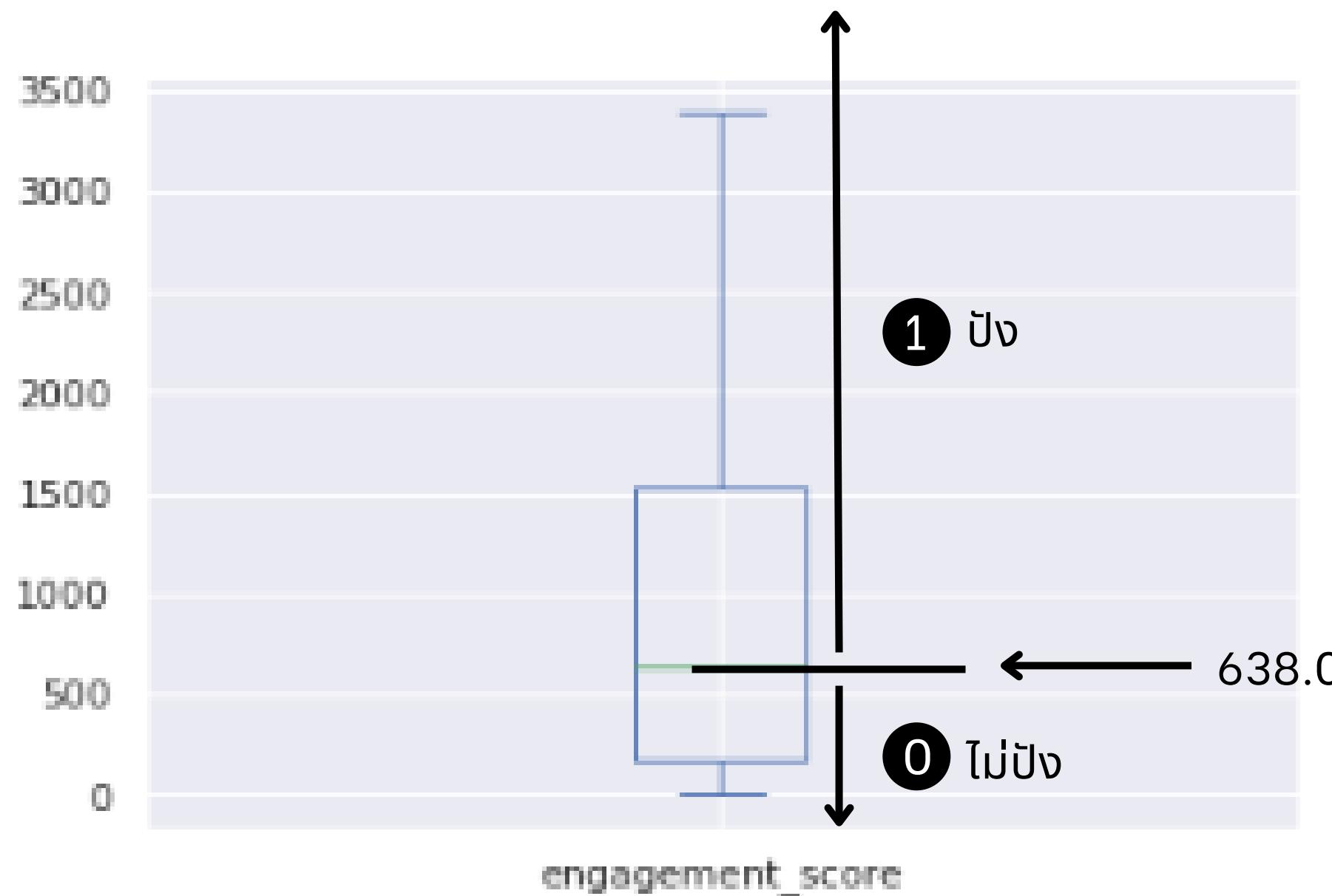
# Data Transformation

ตัวแปร Y

**Engagement score = reaction\_count + 5\*comments + 10\*shares**

	date	time	type	post_text	comments	shares	reaction_count	engagement_score
0	2022-03-25	16:01:00	image	NaN	1	0	638	643
1	2022-03-22	10:00:05	imageandcaption	เป็นสมการที่ทำแทนกันไม่ได้จริงๆ และ...	14	180	1166	3036
2	2022-03-21	15:43:10	status	NaN	0	0	18	18
3	2022-03-19	11:00:01	imageandcaption	💥 พลาดไม่ได้ พื้นที่ความรู้คุ้นความสนุกใน ❤️ นี่...	0	9	44	134
4	2022-03-19	05:47:10	status	พิสิกส์สามัญ จาก สวท. กขสอบคุณภาพครับ	1	68	395	1080
...	...	...	...	...	...	...	...	...
336	2021-03-02	15:19:07	status	🔴 มีเวลา空อย แต่ต้องใช้คะแนนเยอะ!! ก ❤️ พี่ด้วยคัม...	0	26	131	391
337	2021-03-02	12:24:35	status	NaN	0	30	212	512
338	2021-03-01	13:18:31	video	⭐ ตั้งเป้าพิชิตพิสิกส์สามัญ 60+!! ก 🌱 แต่ยังเก็บ...	0	31	109	419
339	2021-02-27	12:16:34	status	🔥 ใครอยากเข้าวิชาจุฬาต้องดู!! ก ❤️ คะแนนขั้นต่ำริ...	5	70	170	895
340	2021-02-25	12:02:00	video	🔔 พุธนี้มี live สอนฟรีจากพี่ตัวนั้นรู้ยัง!! ก 1...	175	97	122	1967

# Data Transformation



แบ่งเป็น 2 class ปัง/ไม่ปัง ที่ Median ด้วยค่าที่ได้  
จาก boxplot ไม่รวม outlier

# Data Transformation

engagement\_score



engagement\_score\_class

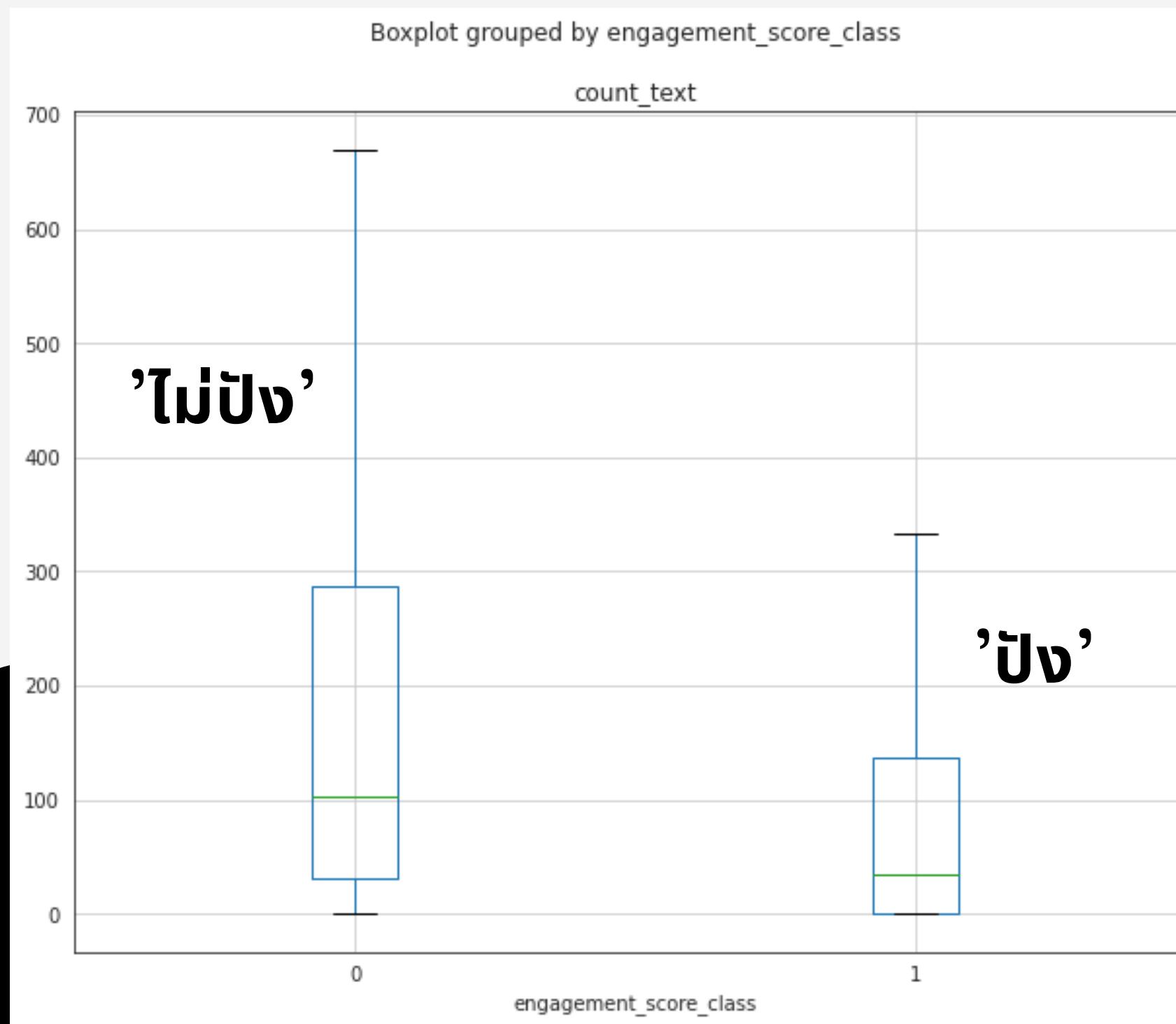
engagement_score_class	
0	1
1	1
2	0
3	0
4	1
...	...
336	0
337	0
338	0
339	
340	

341 rows × 1 columns

engagement\_score\_class  
0 171  
1 170  
dtype: int64



# Data Analysis



## imageandcaption

-> มีสัดส่วนโพสต์ไม้ปังมากกว่า



## image

-> มีสัดส่วนโพสต์ปังมากกว่า

# โพสต์ที่มี 'จำนวนคำน้อย' แล้วปัง เป็นอย่างไร?

- ข้อมูลส่วนใหญ่อยู่ในรูปภาพ และมีเพียงแค่ปั้นสัน ๆ หรือไม่มีแค่ปั้บเลย

Physics Blueprint ฟิสิกส์&ความคิดวิเคราะห์ by พี่ด้ว  
24 มีนาคม 2021 ·

100 แนวใจร้ายว่ารู้ครบทุกสูตรแล้ว  
ถึงจะเปลี่ยนแนวก็ไม่ต้องกลัว  
พี่ด้วสรุปสูตรที่ต้องรู้ มาให้น้องๆแล้วค้าบ ❤️

#วิชาสามัญ #ฟิสิกส์วิชาสามัญ #ฟิสิกส์พี่ด้ว  
#dek64 #เด็กชั่ว



แจกสูตร

Physics Blueprint ฟิสิกส์&ความคิดวิเคราะห์ by พี่ด้ว  
13 กุมภาพันธ์ ·

สมการ...กี่พมอยากให้คุณ  
ในว่าเล่นไกปีนี้

$$\log_{\text{I}} \text{ ❤️ } = \text{ U }$$

เพราะหัวใจพม log ไว้ที่ U

PHYSICS  
BLUEPRINT

ความคิดเห็น 12 รายการ แซร์ 222 ครั้ง

แจกเฉลย

รวมสกิต  
คะแบบ

มีม

# โพสต์ที่มี 'จำนวนคามาก' แล้วไม่ปัง เป็นอย่างไร?

- มีรูปภาพ พร้อมข้อมูลในแคปชั่นที่มีรายละเอียดเยอะ ครบถ้วน

Physics Blueprint พลิกส์&ความสนั่นไหวศาสตร์ by พี่ตัว  
11 พฤศจิกายน 2021 ·

11.11 ต้องห้ามพลาด 🚫  
ลงทะเบียนลุ้นของรางวัลวันสุดท้าย !!!  
เพียงแค่ลงทะเบียนใหม่!  
#Recheck ความพร้อมก่อนใคร  
เตรียม-ตัว-ติด มหา'ลัยในฝัน  
มีสิทธิ์ลุ้นรับของรางวัล iPad พร้อม Apple Pencil  
คอร์สเรียนฟรี คุปองส่วนลด 1,000 บาท  
และของรางวัลมูลค่ารวมกว่า 50,000 บาท  
สำหรับผู้ที่ลงทะเบียนเพื่อมาติวพร้อมกัน  
ช่วงเพื่อน 1 คน ได้รับสิทธิ์ลุ้นรับเพิ่ม 1 สิทธิ์  
ย้ำอีกครั้ง หมวดลงทะเบียนเพื่อรับของรางวัล  
วันที่ 11 พ.ย. 2564 เวลา 23.59 น.  
ประกาศผลรางวัลวันที่ 12 พ.ย. 64 เวลา 18.00 น. เป็นต้นไป  
เฉพาะน้องที่รับชมไลฟ์สดเท่านั้น !!!  
ทาง Youtube : SmartMathPro

สมัครติวและรับเอกสารติว  
ที่ <https://bit.ly/3pc18DN>  
หรือสแกน QR code ที่รูปได้เลย  
แล้วเจอกันน้าา

#recheck #เตรียมตัวติด #dek65 #dek66 #dek67 #เด็กชั่ว #ติวฟรี #สอบติด #live #เด็กสอบใหม่ #smartmathpro #พี่ปั้นsmmp



ขายคอร์ส



Physics Blueprint พลิกส์&ความสนั่นไหวศาสตร์ by พี่ตัว  
6 พฤศจิกายน 2021 ·

แคปหน้าจอไว้ตัววันไหนก็ไม่พลาด !!!

พบกับมหกรรมติวฟรี #Recheck ความพร้อมก่อนใคร

เตรียม-ตัว-ติด มหา'ลัยในฝัน

วันที่ 12 - 14 พฤศจิกายน 2564 นี้เท่านั้น

Live สด ทาง Youtube : SmartMathPro

11 พ.ย. 64

หมวดลงทะเบียนเพื่อสุ่มรับของรางวัล

12 พ.ย. 64

Talk Show

ร่วมพูดคุย #Recheck ความพร้อมก่อนสอบกับติวเตอร์ทั้ง 5

และพี่เนส SMP NEWS

ประกาศรางวัลจากกิจกรรมยิ่งช่วยยิ่งได้ลุ้น iPad และคอร์สเรียนฟรี

ที่สำคัญมีของรางวัลแจกสำหรับน้องๆ ในไลฟ์สดเท่านั้น!!

ตั้งแต่เวลา 18:00 น. เป็นต้นไป

13 พ.ย. 64

วันที่ 1 เตรียม ตัว ติด #Recheck

คณิตศาสตร์ โดย พี่ปั้น SmartMathPro

เวลา 14:00 -16:00 น.

ภาษาอังกฤษ โดย พี่เก่ง GAT ENG COOL COOL

เวลา 16:30 -18:30 น.

14 พ.ย. 64

วันที่ 2 เตรียม ตัว ติด #Recheck

เคมี โดย พี่กัปดัน Chemistry K

เวลา 13:00 -15:00 น.

ชีวะ โดย พี่หนุน BPN ONLINE+

เวลา 15:15 -17:15 น.

ฟิสิกส์ โดย พี่ตัว Physics Blueprint

เวลา 17:30 -19:30 น.

อย่าลืมช่วยเพื่อนมาติวพร้อมกัน

ช่วงเพื่อน 1 คน ได้รับสิทธิ์ลุ้นรับของรางวัลเพิ่ม 1 สิทธิ์ ยิ่งช่วยมาก ยิ่งมีสิทธิ์มาก

.

สมัครติวและรับเอกสารติว

ที่ : <https://bit.ly/3pc18DN>

หรือสแกน QR code ที่รูปได้เลย

แล้วเจอกันน้าาาาา

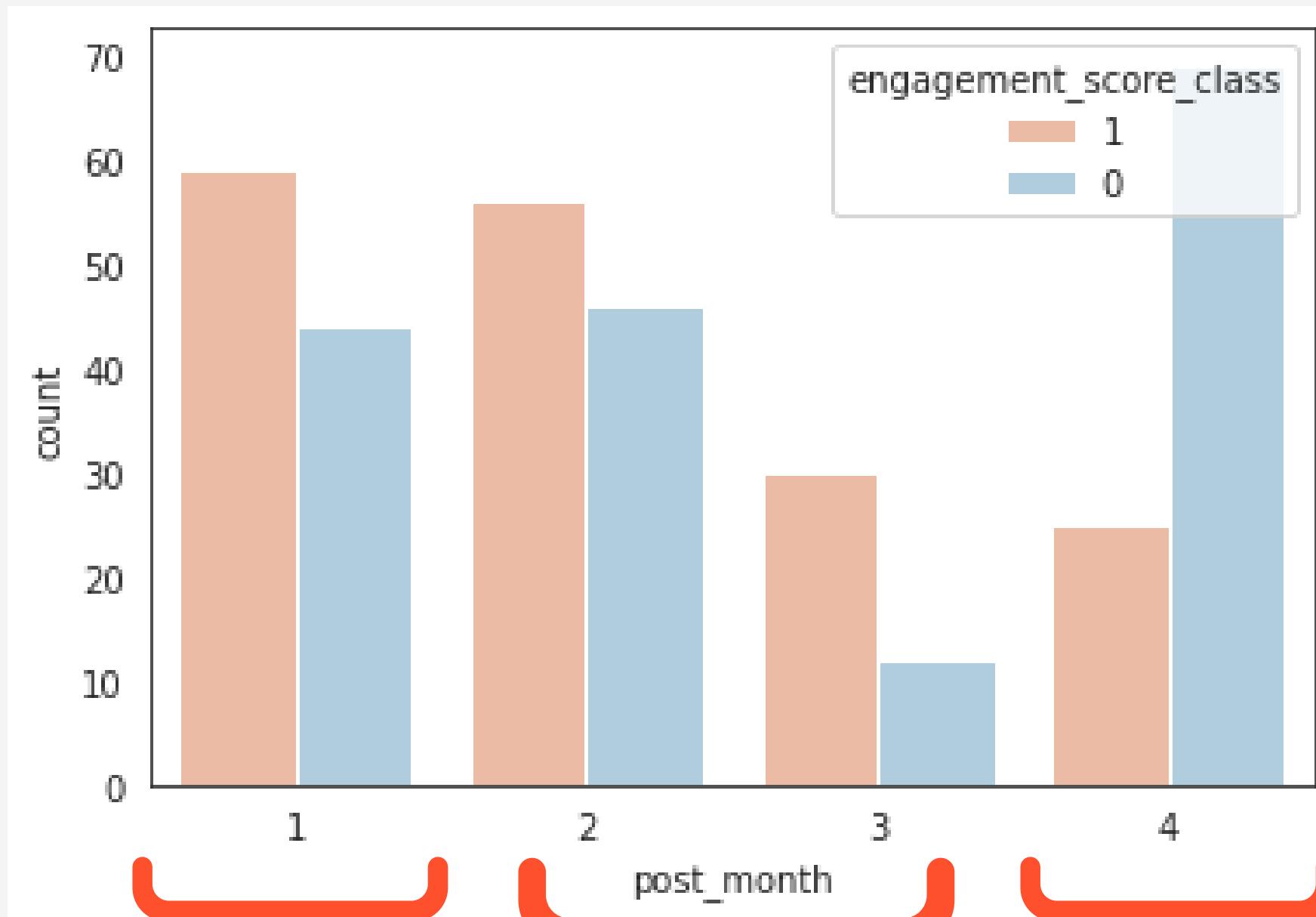
.

#recheck #เตรียมตัวติด #dek65 #dek66 #dek67 #เด็กชั่ว #ติวฟรี #สอบติด #live #เด็กสอบใหม่

#smartmathpro #พี่ปั้นsmmp

ติวฟรี

# 'ไตรมาส' ส่งผลต่อความปังของโพสต์อย่างไร?



- สอ卜 GATPAT
- นักเรียนหาคอร์สโค้งสุดท้าย

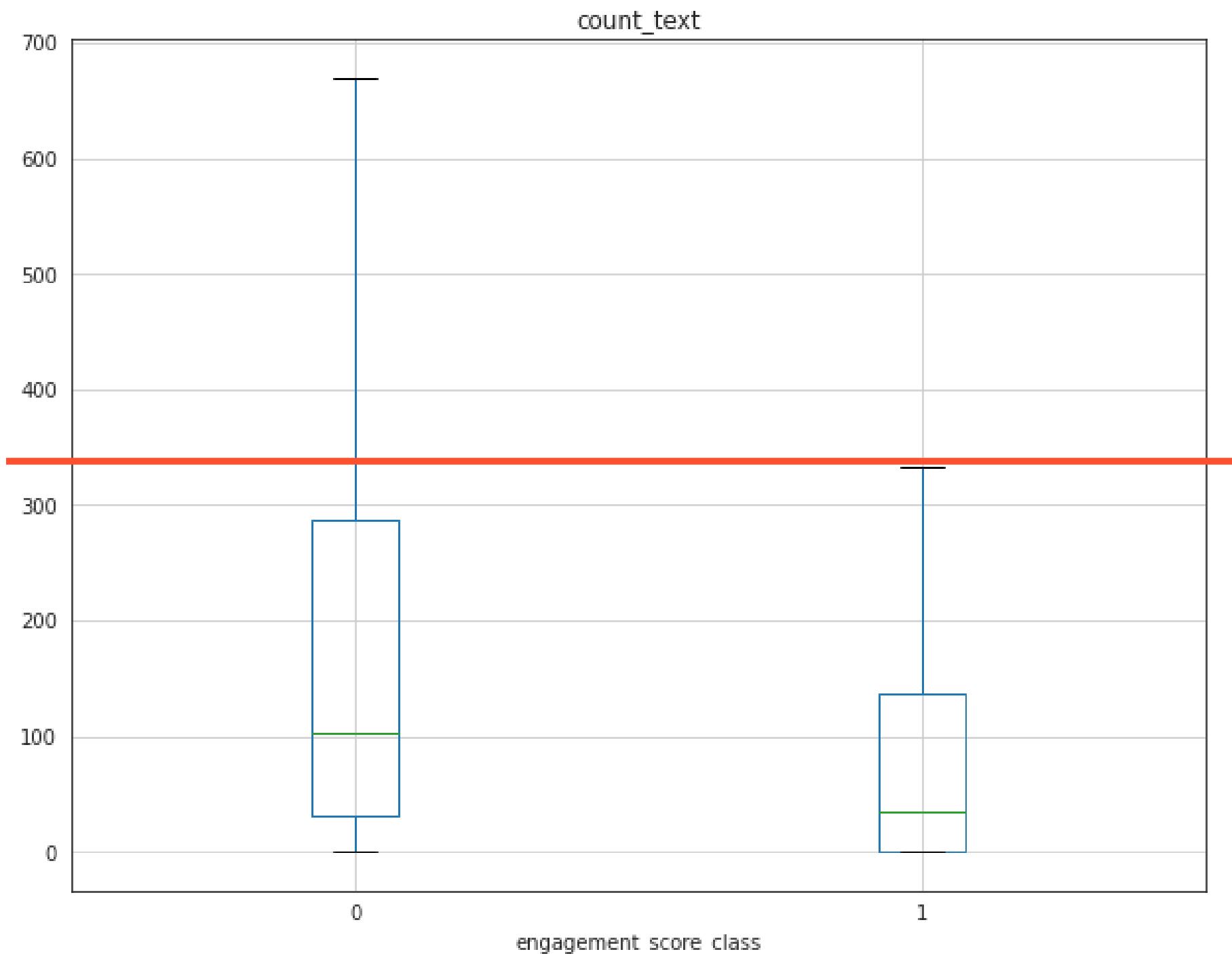
- นักเรียนหาคอร์สเรียน

- นักเรียนไม่ค่อยหาคอร์สเรียน

- ไตรมาส 1
  - ช่วงสอบ จะมีโพสต์ให้กำลังใจและชวนคุย
- ไตรมาส 2
  - โพสต์ขายคอร์สอยู่ระดับปานกลาง
- ไตรมาส 3
  - ไม่ค่อยมีโพสต์ขายคอร์ส
- ไตรมาส 4
  - โพสต์ขายคอร์สเยอะ

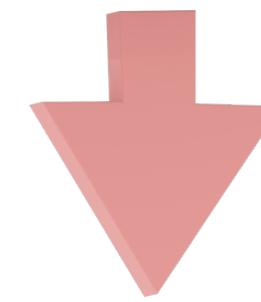
# Naive Classifier

Boxplot grouped by engagement\_score\_class



## Rule

- ถ้า  $\text{count\_text} \leq 350$ 
  - โพสต์ปัง (1)
- ถ้า  $\text{count\_text} > 350$ 
  - โพสต์ไม่ปัง (0)



Accuracy = 0.548

# Training Model

```
# Create training and test set  
y = df['engagement_score_class']  
X = df.drop(columns='engagement_score_class')  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=1)
```

แบ่งเป็น test data 25% และ train data 75% ของข้อมูลทั้งหมด

```
logreg = LogisticRegression()  
logreg.fit(X_train, y_train)  
  
cv_scores_lg = cross_val_score(logreg, X_train, y_train, cv=5)  
print("Average 5-Fold CV Score: {:.2f}".format(np.mean(cv_scores_lg)))
```

สร้าง Estimated Test Errors  
ด้วยวิธี K-fold Cross Validation โดยกำหนดค่า CV = 5

# Training Model

Model

Logistic  
Regression

Linear  
Classification

Decision Tree

Random  
Forest

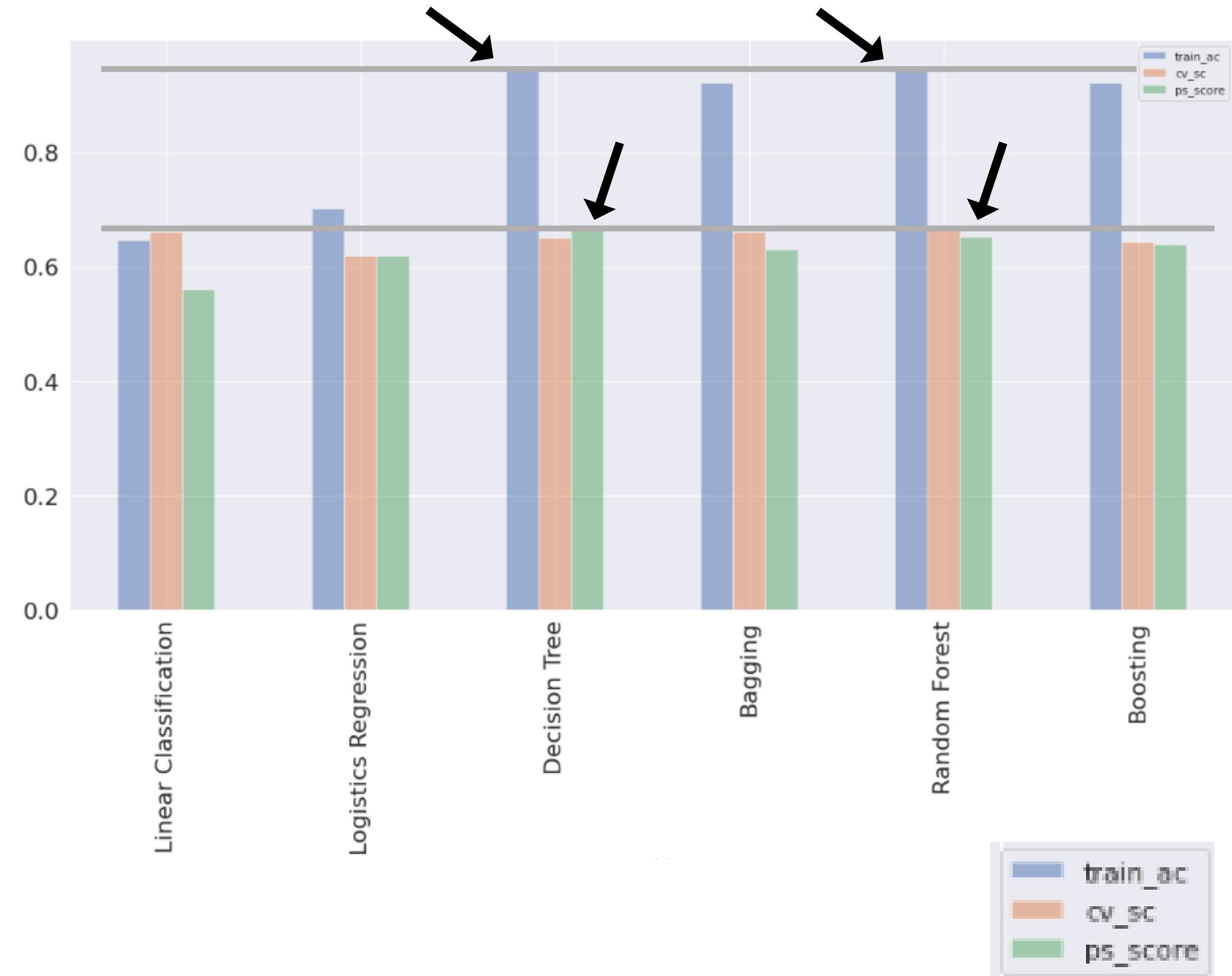
Bagging

Boosting tree

# Select Model

## Default Parameter

	Model	train_ac	cv_sc	ps_score
0	Linear Classification	0.647059	0.662745	0.561629
1	Logistics Regression	0.701961	0.619608	0.618912
2	Decision Tree	0.949020	0.650980	0.665937
3	Bagging	0.921569	0.662745	0.630853
4	Random Forest	0.949020	0.670588	0.653023
5	Boosting	0.921569	0.643137	0.640465



# Model Evaluation

	Model	train_ac	cv_sc	ps_score
0	Linear Classification	0.647059	0.662745	0.561629
1	Logistics Regression	0.701961	0.619608	0.618912
2	Decision Tree	0.949020	0.650980	0.665937
3	Bagging	0.921569	0.662745	0.630853
4	Random Forest	0.949020	0.670588	0.653023
5	Boosting	0.921569	0.643137	0.640465

Overfit



Decision tree

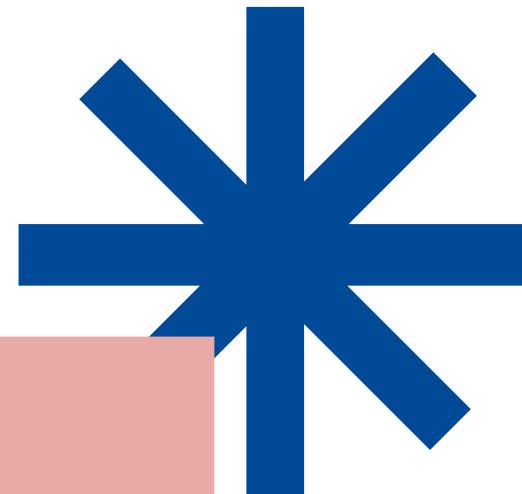
Train accuracy ≈ 0.949

Validation accuracy ≈ 0.651

Train accuracy ≈ 0.949

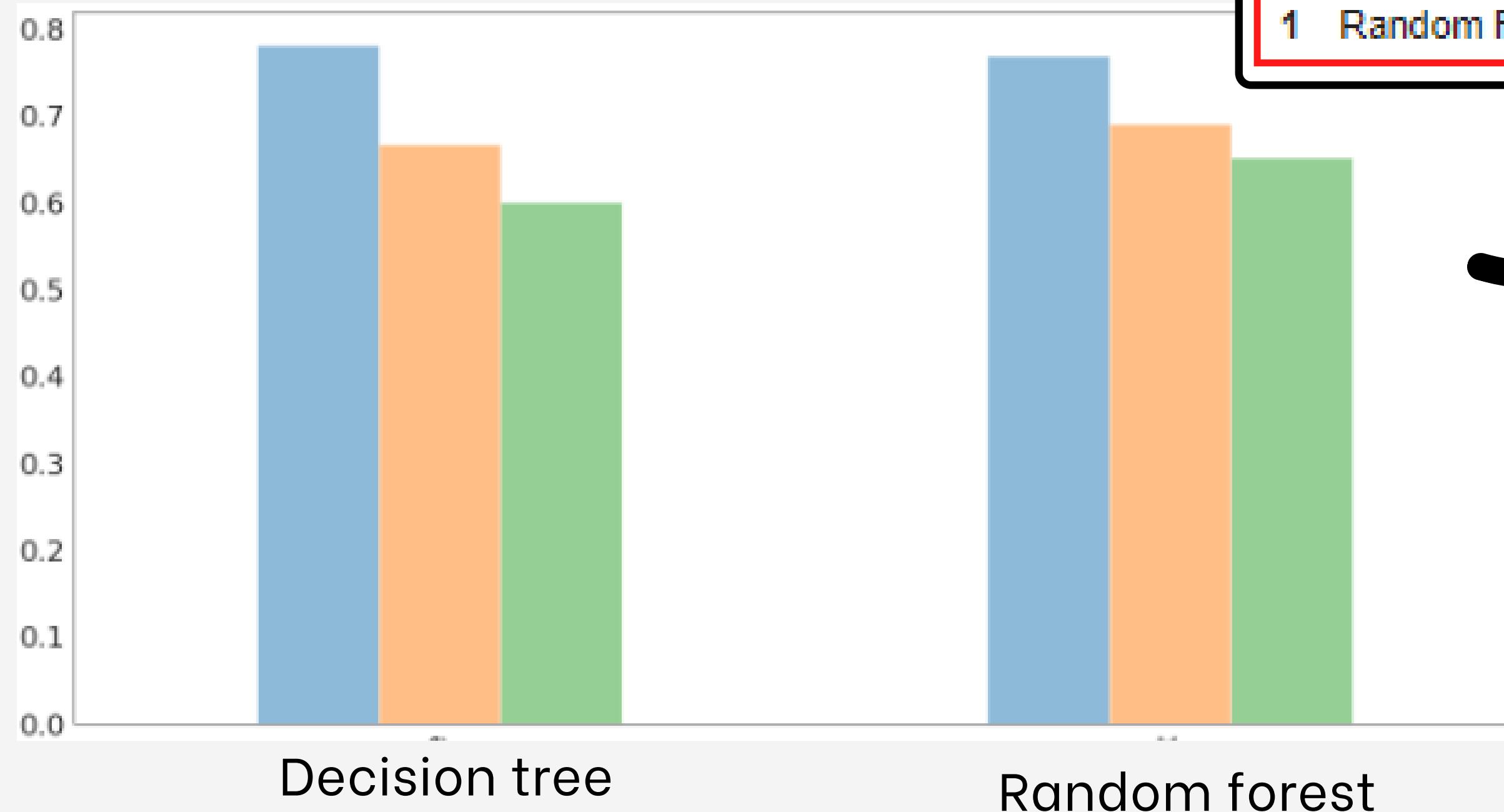
Validation accuracy ≈ 0.671

Random forest

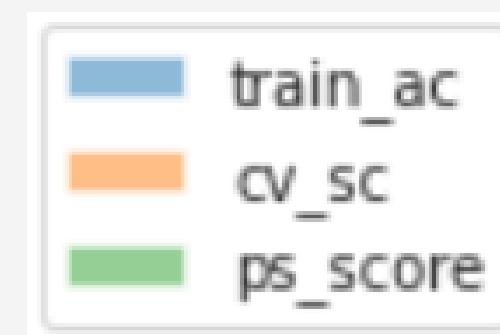
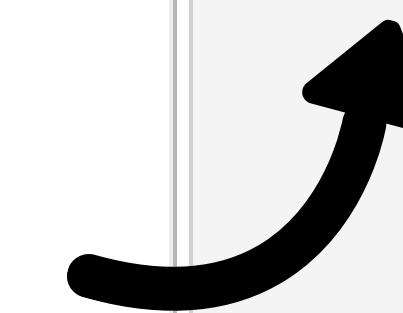


# Model Improvement

Feature Important + GridSearch



Model	train_ac	cv_sc	ps_score
Decision Tree	0.780392	0.666667	0.599333
Random Forest	0.768627	0.690196	0.651868



# Model Evaluation

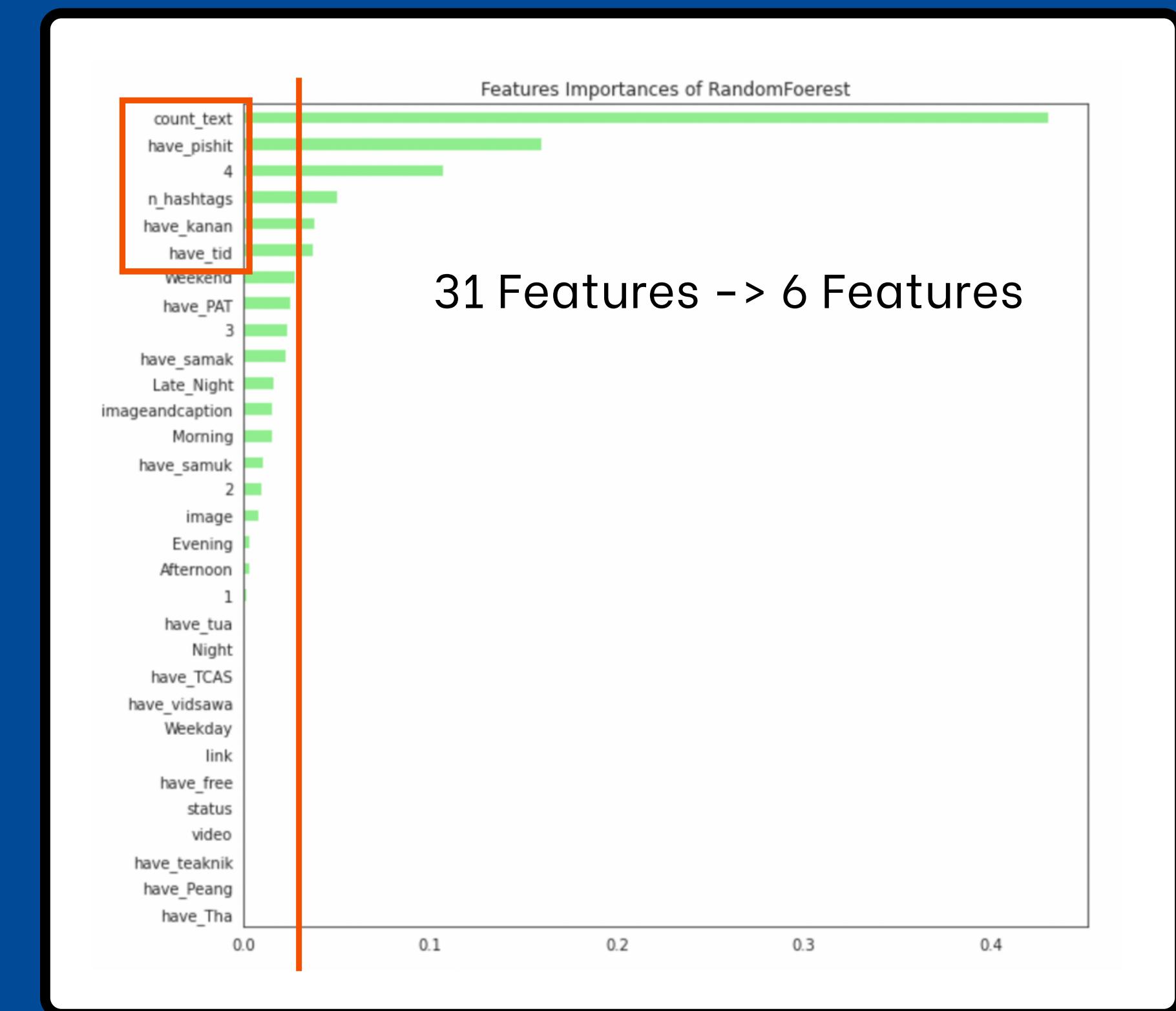
## Feature Importances

- 1.count\_text
- 2.have\_pishit
- 3.lmsuna 4
- 4.n\_hashtags
- 5.have\_kanan
- 6.have\_tid

## Parameters by Gridsearch

n\_estimators=9,max\_depth = 6, min\_samples\_leaf=4,random\_state=4, criterion='entropy',max\_features= 'log2'

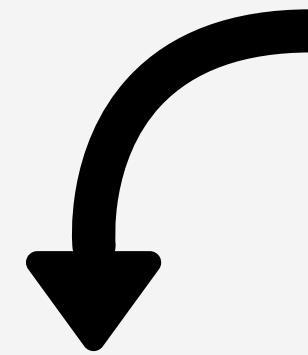
## Random Forest



# Prediction

## Confusion Matrix

Predict



Actual

		1	0
1	28	16	
0	14	28	

	precision	recall	f1-score	support
0	0.67	0.64	0.65	44
1	0.64	0.67	0.65	42
accuracy			0.65	86
macro avg	0.65	0.65	0.65	86
weighted avg	0.65	0.65	0.65	86

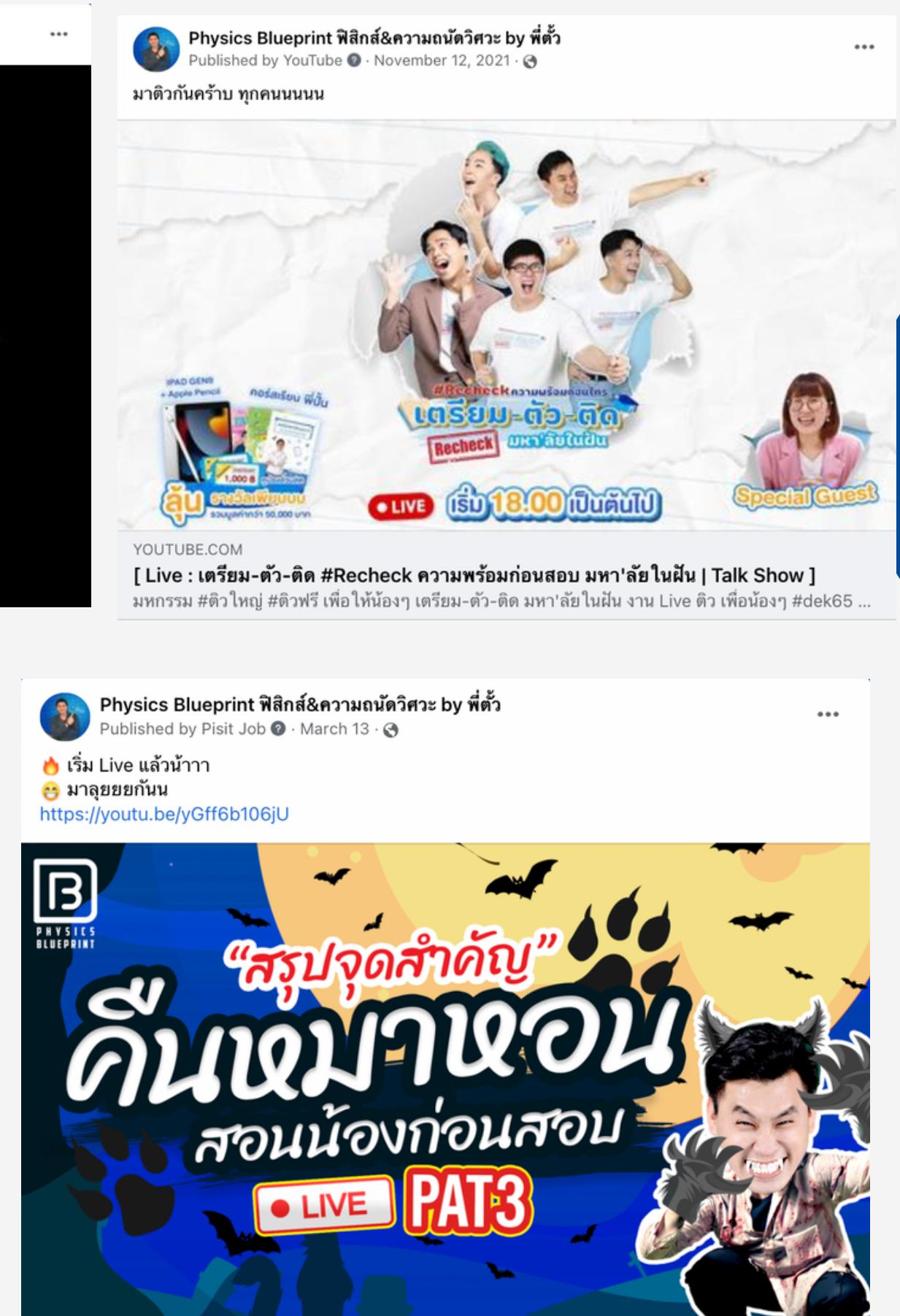
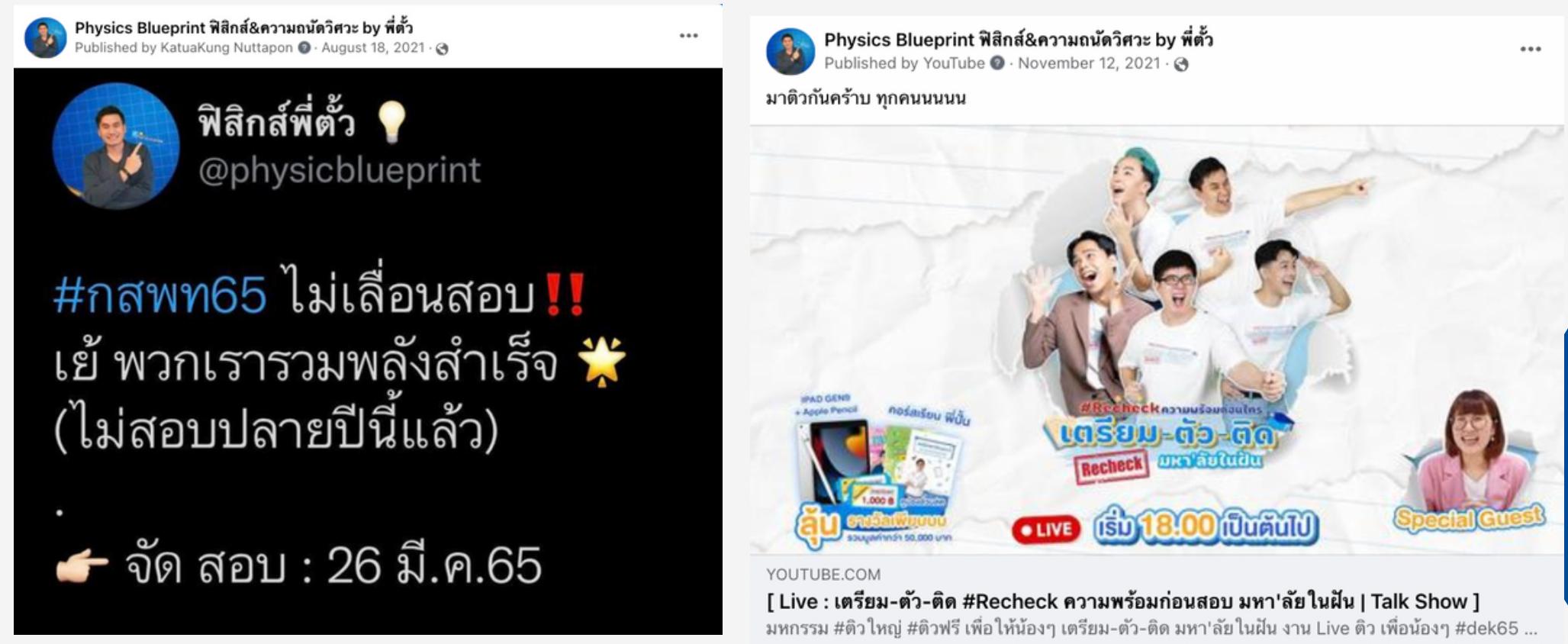
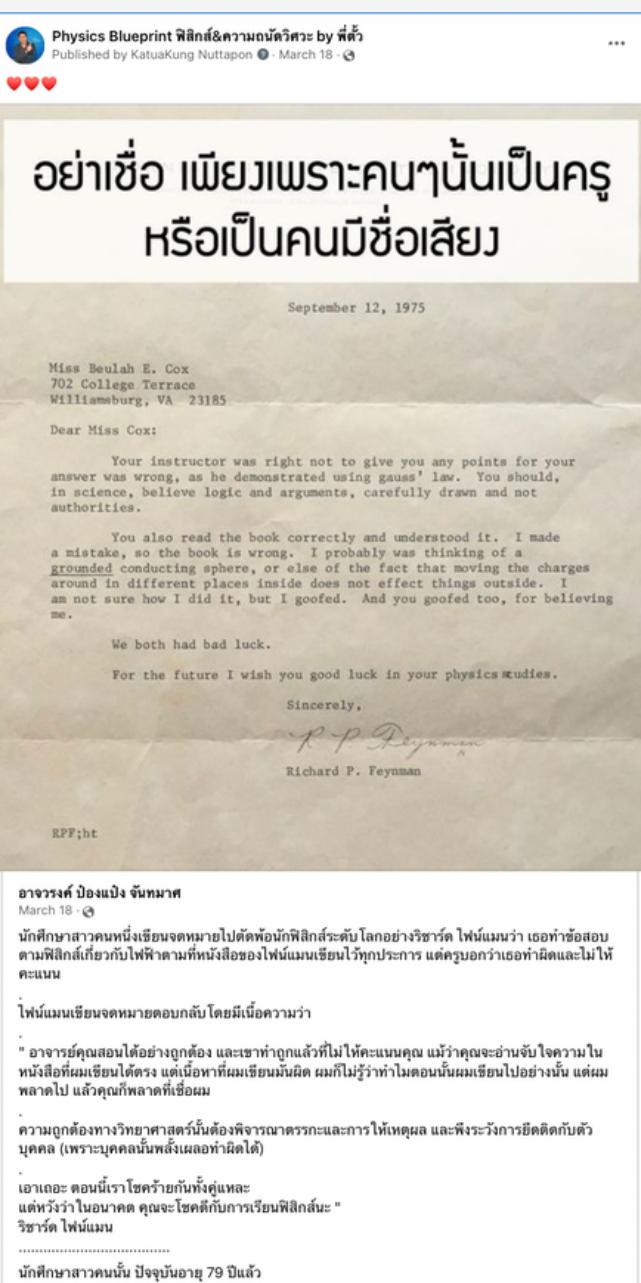
# ประเภทของโพสท์ที่ถ่ายผิด

## โพสท์เดาเป็น 1 แต่จริงๆ เป็น 0



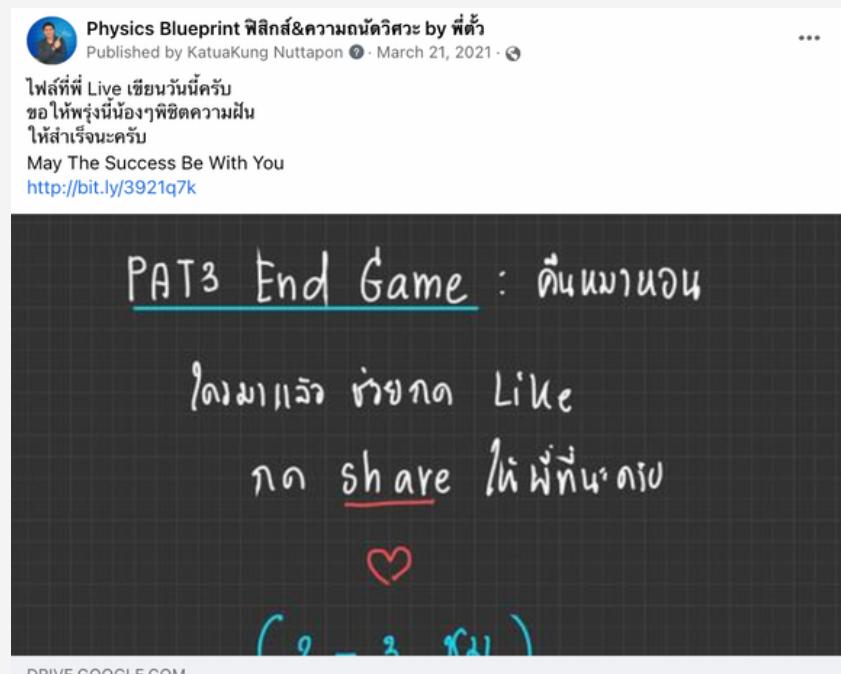
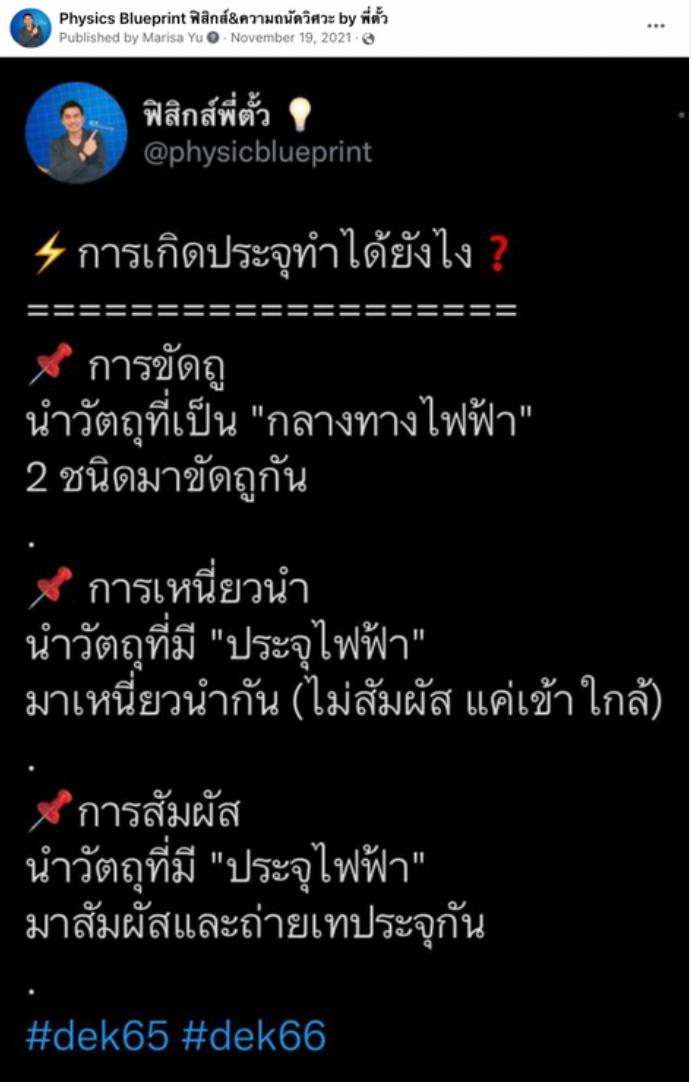
- count\_text น้อย และเป็นรูปภาพ แต่ไม่เดลไม่สามารถจับ Content ในรูปภาพได้

- ไม่ได้เก็บข้อมูลของโพสต์ ตั้งทางก็มีการแชร์ต่อ

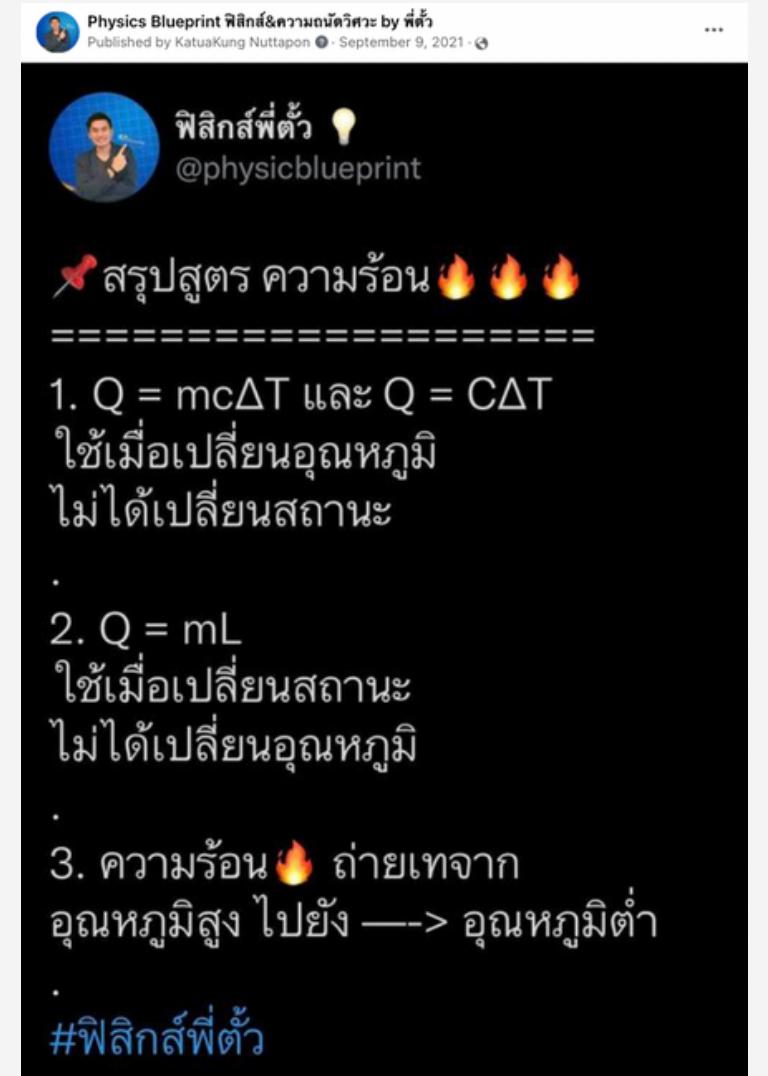
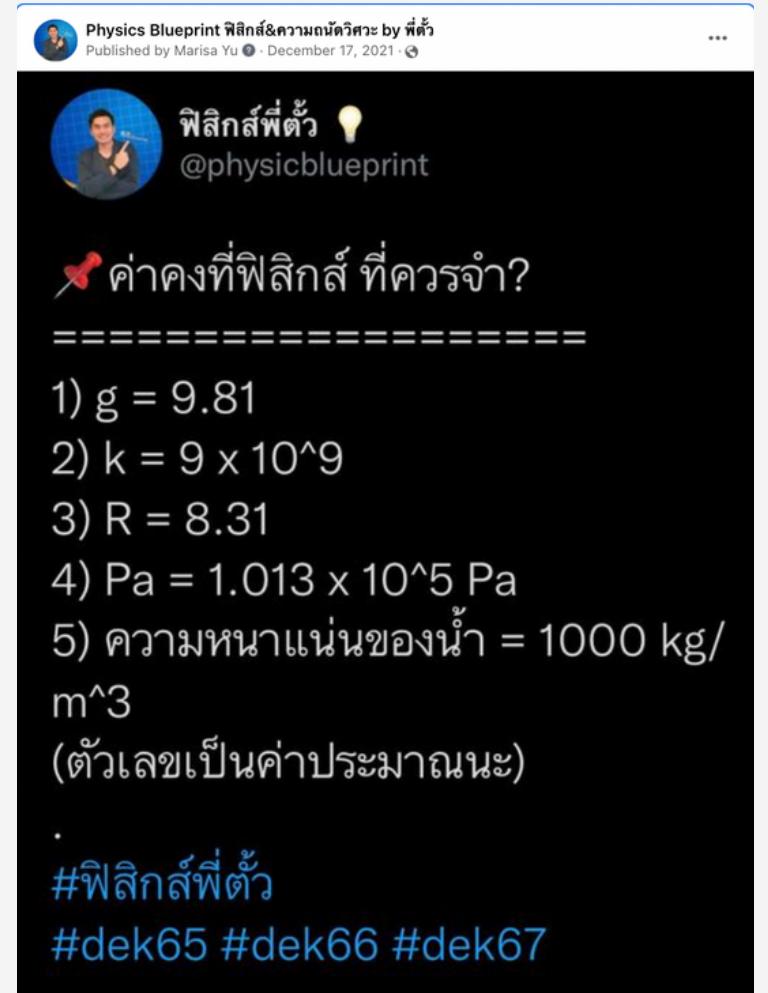


- ข้อมูลประเภท link มีน้อยมาก ไม่เดลไม่สามารถจับ pattern ได้

# ประเภทของโพสท์ที่ถูกผิด



# โพสท์เดาเป็น 0 แต่จริงๆ เป็น 1



- โพสต์อยู่ในไตรมาส 4 มีโพสต์ขายคอร์ส เยอะ เลยเดาเป็น 0
- แต่ Content มีประประโยชน์ต่อเด็ก ๆ ที่กำลังเดือด 55555

# Summary

Accuracy

Naive Model  
54.8%

→  
+10.2%

RandomForest Model  
65.0%

การนำไปใช้งาน

ทำอย่างไรให้โพสต์ปัง?

- จำนวนคำน้อย ๆ กระชับ ไม่ต้องมีรายละเอียดเยอะ อาจเป็นการสรุปข้อมูลในรูป หรือมี link ให้อ่านรายละเอียดต่อเอง
- เน้นสื่อสารด้วยรูปภาพ
- โพสต์ขายคอร์สให้กระจายทุกไตรมาส ไม่ต้องกระจุกใบบางไตรมาสมากเกินไป

