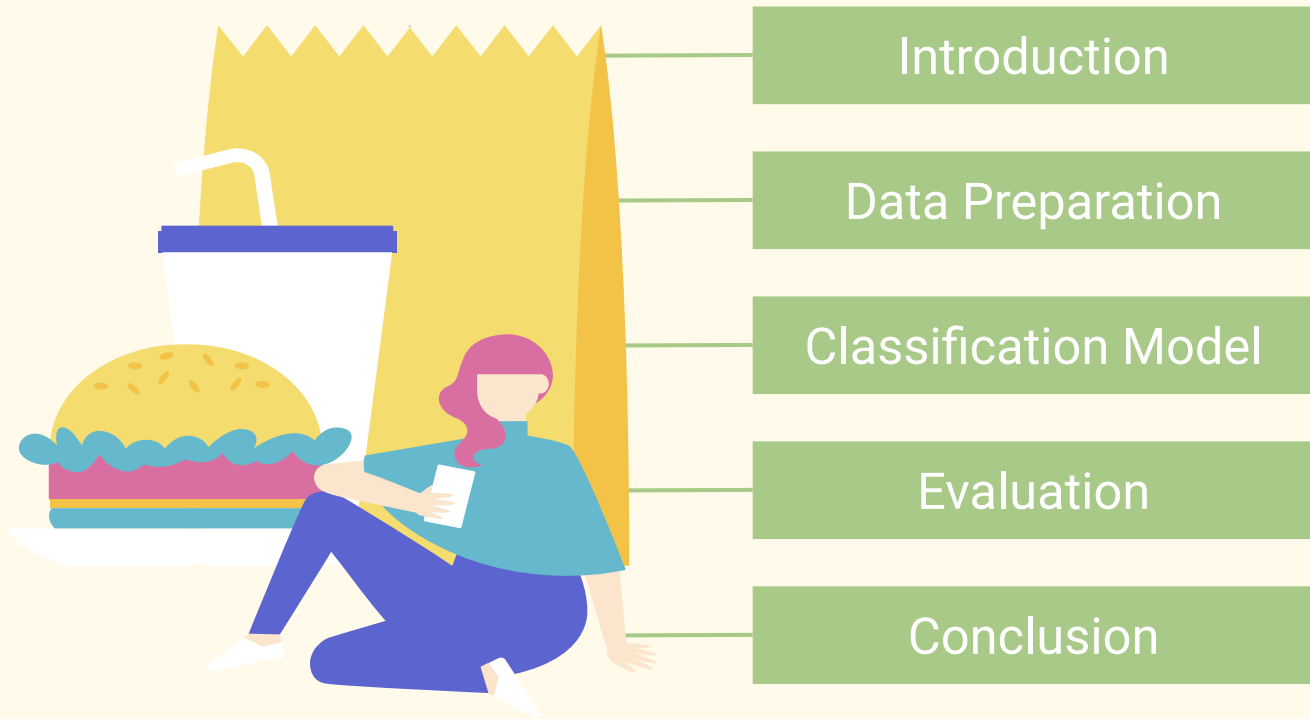


Fake Reviews Classification

Using Locality-sensitive
Hashing (LSH)



Agenda





Have you eaten?



7,500,000

searches/year on Wongnai

75%

of customers show that
review affects decision-making

53.6%

of people follows food influencers



Foodpanda เพื่อร้านค้าปรึกษาแก้ไขปัญาเบื้องต้นต่างๆ

...

4 เม.ย. · 🍕 · อยากปรึกษาอะ เจอแบบนี้แจ้งลบได้ไหมคะ ตอนแรกก็โกรธแต่ตอนนี้ก็กลายเป็นตลกไปแล้ว เพราะไม่มีอันไหนเป็นความจริงเลย เราเปลี่ยนน้ำมันทุกวัน และไม่ทอดค้าง ทอดซ้ำ ไก่เราก็ใหม่ทุกวันเพราะเราขายหมดทุกวัน ถ้าเพื่อนๆเจอแบบนี้จะทำยังไงกันคะ เราทำร้านมาด้วยความรัก ใส่ใจทุกกล่อง อันไหนคอมเม้นให้เราปรับปรุงเราพร้อม...



6

ควา



ตนตางจ

3 ก.พ. · 📷

Foodpanda เพื่อร้านค้าปรึกษาแก้ไขปัญาเบื้องต้น ...

ความคิดเห็นที่ตรงกัน 3 ข้อความ



โอย!!!! เจอเหมือนกันคะ เมื่อเข้านี้เลย บอกว่าแพง ไม่คุ้มกับราคามั้ง ไร้มั้ง นุ่น นี่เมื่อก่อนเครียดมากคะ เราเป็นแม่ค้า พยายามทำให้ดีที่สุด แต่คนนะคะ ร้อยพ่อพันอย่าง บางรีวิวเพื่อหับถกมเลย บางรีวิวก้อแนะนำให้แก้ไข ตอน... ดูเพิ่มเติม

ถูกใจ · ตอบกลับ · 5 สัปดาห์

มีใครยกเลิกสัญญา เพราะเจอรีวิวลูกค้า
แปลกๆไหมคะ😂
ใจมันทนไม่ไหว
ใครมีลัษณ์ยกเลิกบ้างคะ
แพนด้าเอาแต่ลูกค้าเลย



ดาว 1

Foodpanda เพื่อร้านค้าปรึกษาแก้ไขปัญหาเบื้องต้น

...

5 ธ.ค. 2020 · 🍻 · เราสามารถขึ้นแบล็คลิส ลูกค้าอย่างไร ได้บ้างคะ วันนี้เจอรีวิว.. ขำวกล่องเดียว ไม่พอใจว่าปริมาณน้อย ทั้งๆที่เรามั่นใจว่าปรกติ ให้ 1 ดาว รีวิวตาดชนิดที่ว่า.. พวกขายสร้างตึก, ทูเรต ใครอยากกินร้านนี้ก็เชิญ (คือหยาบต๊ะ) ร้านโทรไปขอโทษพร้อมทำใหม่ให้ เค้าก็โอเค เหมือนสนุก และก็ไมรีวิวเพิ่มให้ เศร้าใจมากคะ บอกตรงๆ ทำมาดีตลอด เจอลูกค้าแบบนี้ รีวิวก็ลบไม่ได้ เหนื่อยจัง **หากเค้าสั่งมาอีก เรากดยกเลิก...



28

ความคิดเห็น 26 รายการ

ความคิดเห็นที่ตรงกัน 5 ข้อความ



อย่าไปอ่านคะบันทอนจิตใจ...เก็บพลังไว้ทำอย่างอื่น...เราทำดีที่สุดก้อพอ....ขายกลางวันอาจจะเจอคู่แข่ง....ขายกลางคืนอาจจะเจอคนเมา....รีวิว...เคยโทรไปถามลูกค้าที่รีวิวว่าทำไมไม่แดกอาการเป็นใจคะได้คำตอบว่า...ผมเมาคับ 🙄🙄

ถูกใจ · ตอบกลับ · 22 สัปดาห์ · แก้ไขแล้ว



7



ดนดาว 1

da Foodpanda เพื่อร้านค้าปรึกษาแก้ไขปัญหาเบื้องต้น

...

25 พ.ย. 2020 · 🍻 · โดนจนได้รีวิวดราปาบ จากลูกค้า ให้ 1 ดาว ต่ำหนี่เรว่าว สั่ง ราดหน้าทะเล ได้ราดหมู เห้ยริบดูเมนูตัวเอง ก็ได้ดูทาน ออกมาตังๆ ร้านกูไม่ได้ขายราดหน้าทะเลนี้หว่า ราคา ก็ไม่ใช่ละ ร้านผม ขาย 69 ที่ลูกค้าบอก 75 นะ เล่นซะเสีย ความรู้สึก กันทีเดียว เศร้าใจ #รีวิวดราปาบร้านค้า · รูปภาพจากโพสต์ของ Somporn Pitakwongchinda



Business Impacts



Customer



Restaurant



Business Platform

Data Collection

- Web Scraping from [wongnai](#)
- Python
 - Selenium & Chrome Webdriver



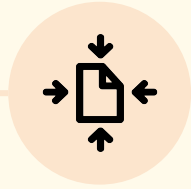
Data Collection

Column1	restaurant_name	Reviewer	comments	score	date	Follower	Review	Photos	Length	
0	0	Ai ching ข้าวมันไก่ (อ้ายจิง) ถนนดินสอ	Veerayuth K.	ร้านข้าวมันไก่เพิ่งเปิดใหม่ไม่นาน ตรงข้ามโรงเร...	5	เมื่อ 4 อาทิตย์ที่แล้ว	79	92	522	1423
1	1	Ai ching ข้าวมันไก่ (อ้ายจิง) ถนนดินสอ	Chit Chin	สืบทอดตำรับข้าวมันไก่ไหหลำ ร้านแรกอยู่ตลาดสี่ม...	5	เมื่อ เดือนที่แล้ว	2	45	172	226
2	2	Ai ching ข้าวมันไก่ (อ้ายจิง) ถนนดินสอ	Waan	กินคนเดียวแบบธรรมดาที่อ้อมแล้ว เนื้อไก่แบบชิ้นไ...	5	เมื่อ 4 อาทิตย์ที่แล้ว	1	11	33	871
3	3	Ai ching ข้าวมันไก่ (อ้ายจิง) ถนนดินสอ	Aprilly Pin	อร่อยสมคำล่ำลือมากๆ ข้าวเป็นเม็ดไม่แฉะ ไ...	5	เมื่อ อาทิตย์ที่แล้ว	6	36	64	216
4	4	Ai ching ข้าวมันไก่ (อ้ายจิง) ถนนดินสอ	Jean Jartupon Jean	มีโอกาสดิไปทานข้าวมันไก่ ร้าน Ai Ching (อ้าย...	5	เมื่อ เดือนที่แล้ว	0	1	5	811

Data Preparation

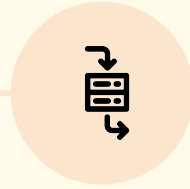


Exploratory

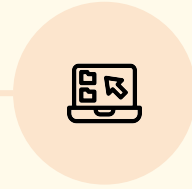


Descriptive

- Mean
- Q1
- Q3



Data
Cleaning



Labeling

Data Preparation

Column1		restaurant_name	Reviewer	comments	score	date	Follower	Review	Photos	Length
0	0	Ai ching ข้าวมันไก่ (ฮ้ายจิง) ถนนดินสอ	Veerayuth K.	ร้านข้าวมันไก่เพิ่งเปิดใหม่ไม่นาน ตรงข้าม โรงเร...	5	เมื่อ 4 อาทิตย์ที่แล้ว	79	92	522	1423
1	1	Ai ching ข้าวมันไก่ (ฮ้ายจิง) ถนนดินสอ	Chit Chin	สืบทอดตำรับข้าวมันไก่โฮหล่า ร้านแรกอยู่ตลาดสี่...	5	เมื่อ เดือนที่แล้ว	2	45	172	226
2	2	Ai ching ข้าวมันไก่ (ฮ้ายจิง) ถนนดินสอ	Waan	กินคนเดียวแบบธรรมดาที่อื่นแล้ว เนื้อไก่แบบชิ้น 1...	5	เมื่อ 4 อาทิตย์ที่แล้ว	1	11	33	871
3	3	Ai ching ข้าวมันไก่ (ฮ้ายจิง) ถนนดินสอ	Aprilly Pin	อร่อยสมคำล่ำลือมากฯฯ ข้าวเป็นเม็ดไม่แฉะ ไ...	5	เมื่อ อาทิตย์ที่แล้ว	6	36	64	216
4	4	Ai ching ข้าวมันไก่ (ฮ้ายจิง) ถนนดินสอ	Jean Jartupon Jean	มีโอกาสได้ไปทานข้าวมันไก่ ร้าน Ai Ching (ฮ้าย...	5	เมื่อ เดือนที่แล้ว	0	1	5	811

Comments

12,113 comments

Length

Reviewer

1,842 reviewers

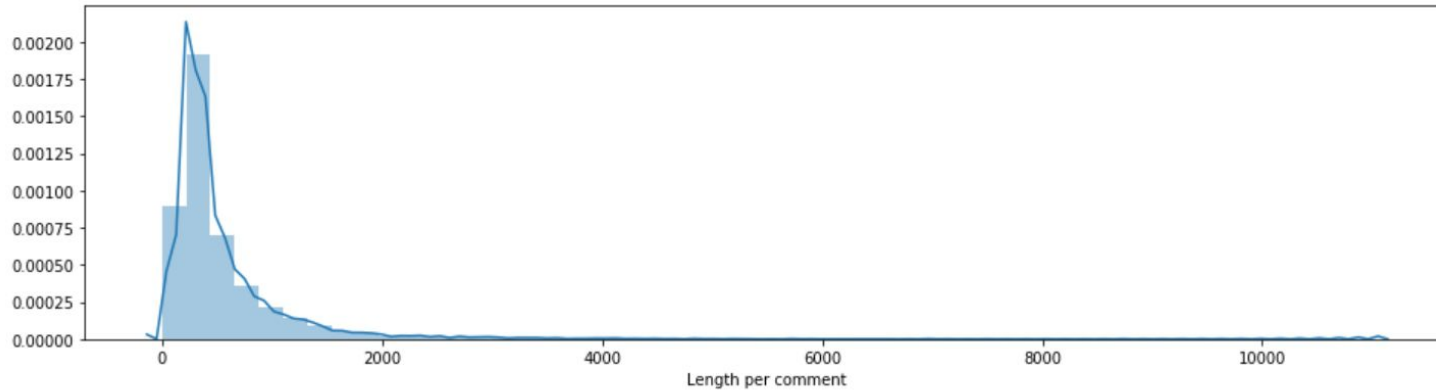
Review

Follower

Photos

Score

Data Preparation



Descriptive Statistics: Follower, Review, Photos, Length, score

Variable	N	N*	Mean	SE Mean	StDev	Variance	Minimum	Q1	Median	Q3
Follower	12859	0	2656.5	81.6	9254.1	85637898.9	0.0	5.0	22.0	265.0
Review	12859	0	456.09	7.19	814.89	664050.79	0.00	28.00	115.00	512.00
Photos	12859	0	4550	124	14049	197385958	0	98	509	3115
Length	12859	0	547.1	5.39	611.4	373813.0	0.00	238.0	369.0	623.0
score	12859	0	4.1054	0.00689	0.7809	0.6098	1.0000	4.0000	4.0000	5.0000

Variable	Maximum	Skewness	Kurtosis
Follower	75841.0	5.19	28.77
Review	8040.00	3.40	15.61
Photos	996544	29.53	1941.15
Length	11003.0	4.48	33.18
score	5.0000	-0.75	0.71

	Q1	Q3 - MAXIMUM
FOLLOWER	5	265-75841
REVIEW	28	512-8040
PHOTOS	98	3115-996544
LENGTH	238	632-11003

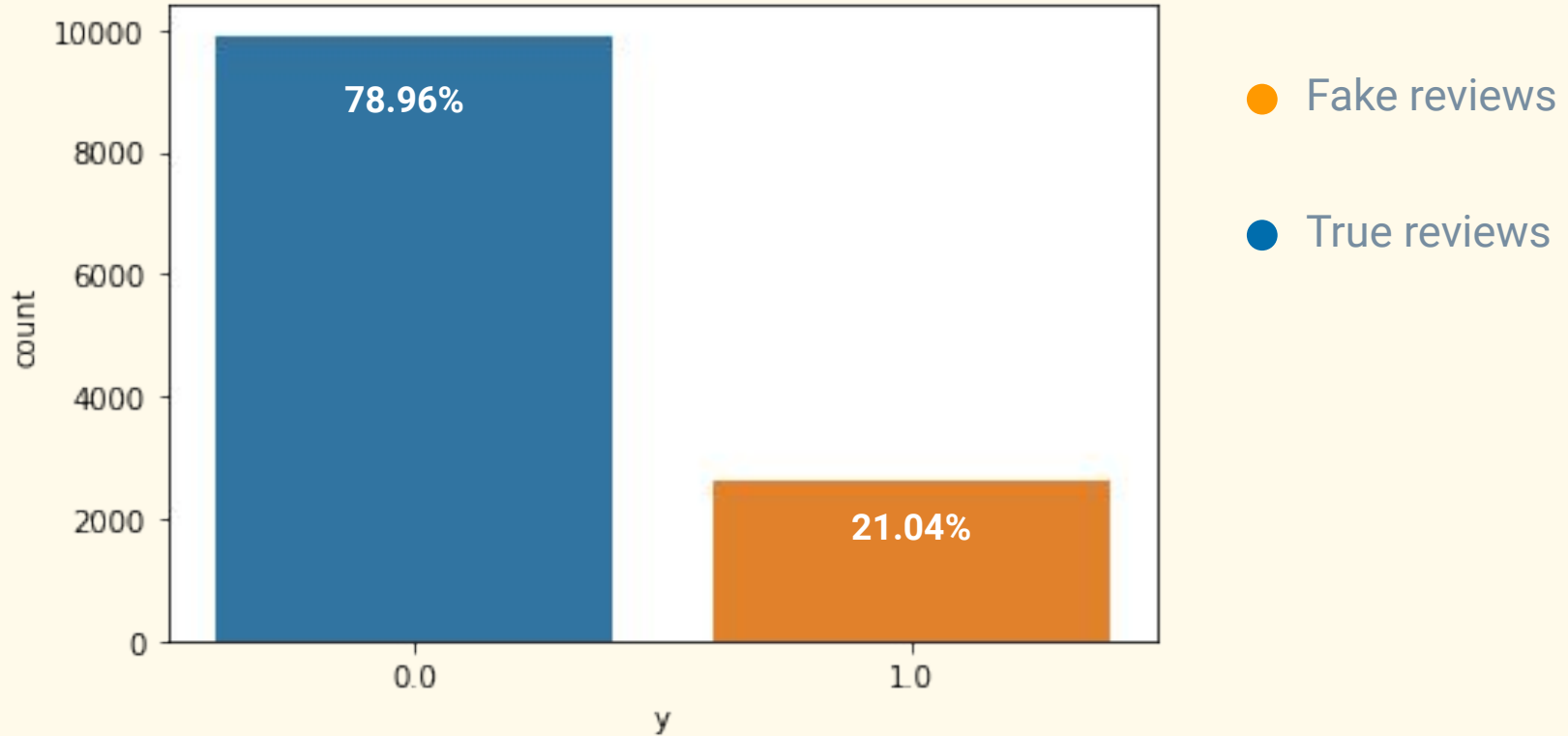
- Cleaning missing value
 - Microsoft Excel

14


Fake Reviews Criteria for Data Labeling



Fake Reviews Criteria for Data Labeling



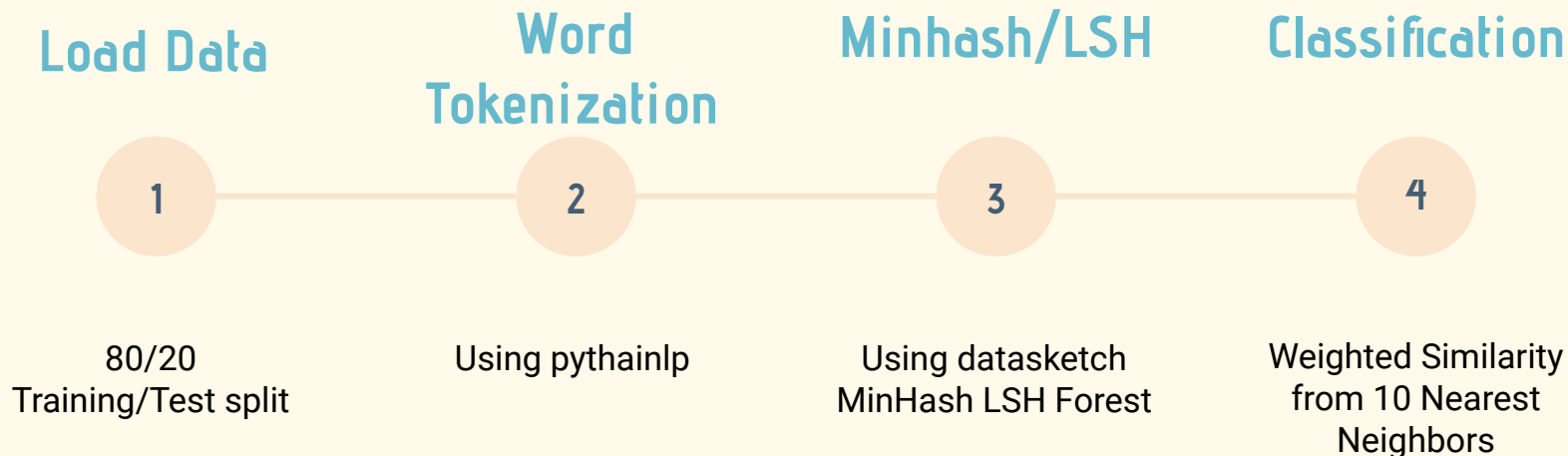
Data Labeling

	Column1	restaurant_name	Reviewer	comments	score	date	Follower	Review	Photos	Length	y
0	0	26BraisedBeef (26 ยี่ สับ หลก) ทำพระ จันทร์ - ...	Auditor Journey	 ตำนานบทใหม่จากก๊วยเตี๋ยวนายโศยทำ พระอาท...	4	เมื่อ อาทิตย์ที่ แล้ว	126	504	3115	1777	1
1	1	26BraisedBeef (26 ยี่ สับ หลก) ทำพระ จันทร์ - ...	สาวเหนือเซา กีน	มาร้านนี้เพราะได้ดูรีวิวในรายการขอส แต่เราไม่ ท...	4	เมื่อ 3 เดือนที่ แล้ว	41	110	1226	393	1
2	3	26BraisedBeef (26 ยี่ สับ หลก) ทำพระ จันทร์ - ...	Bom+	#MichelinThailand2020 #MichelinPlate2020\๓๒๖...	4	11 ก.ย. 2020	62230	1440	20882	4072	1
3	4	26BraisedBeef (26 ยี่ สับ หลก) ทำพระ จันทร์ - ...	Bigstargirl	ปกติชอบแวะไปทานร้านก๊วยเตี๋ยวนายโศยตรง พระอาท...	5	22 ต.ค. 2020	42	328	3767	2904	1
4	5	26BraisedBeef (26 ยี่ สับ หลก) ทำพระ จันทร์ - ...	Muummy	ร้านเนื้อหายากก๊วยเตี๋ยวนายโศยย้ายถนน พระ...	4	19 ต.ค. 2020	18	573	5432	483	0

```
df.describe()
```

	Column1	score	Follower	Review	Photos	Length	y
count	12113.000000	12113.000000	12113.000000	12113.000000	12113.000000	12113.000000	12113.000000
mean	151.566334	4.093536	2678.789895	459.215471	4511.765871	554.411376	0.263353
std	123.310362	0.787693	9287.242494	818.238836	11030.441844	625.923363	0.440471
min	0.000000	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000
25%	51.000000	4.000000	5.000000	28.000000	95.000000	236.000000	0.000000
50%	122.000000	4.000000	22.000000	114.000000	501.000000	362.000000	0.000000
75%	227.000000	5.000000	286.000000	516.000000	3126.000000	636.000000	1.000000
max	594.000000	5.000000	75841.000000	8040.000000	89493.000000	11003.000000	1.000000

Classification Model



ekzhu/datasketch

MinHash, LSH, LSH Forest, Weighted MinHash,
HyperLogLog, HyperLogLog++, LSH Ensemble

Train/Test Split

Full Data Set (100%)

Training Set (80%)

Test Set (20%)

```
df_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 9690 entries, 10957 to 6061
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Column1          9690 non-null   int64
1   restaurant_name  9690 non-null   object
2   Reviewer         9690 non-null   object
3   comments         9690 non-null   object
4   score            9690 non-null   int64
5   date             9690 non-null   object
6   Follower         9690 non-null   int64
7   Review           9690 non-null   int64
8   Photos           9690 non-null   int64
9   Length           9690 non-null   int64
10  y                9690 non-null   int64
dtypes: int64(7), object(4)
memory usage: 908.4+ KB
```

```
df_test.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2423 entries, 4052 to 4043
Data columns (total 11 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Column1          2423 non-null   int64
1   restaurant_name  2423 non-null   object
2   Reviewer         2423 non-null   object
3   comments         2423 non-null   object
4   score            2423 non-null   int64
5   date             2423 non-null   object
6   Follower         2423 non-null   int64
7   Review           2423 non-null   int64
8   Photos           2423 non-null   int64
9   Length           2423 non-null   int64
10  y                2423 non-null   int64
dtypes: int64(7), object(4)
memory usage: 227.2+ KB
```

Word Tokenization



Input

'สี่อ้ว'ยังอรร้อยเหมือนเดิม



Thai Word Tokenize
With normalize



Output

['สี่อ้ว', 'ยัง', 'อรร้อย', 'เหมือนเดิม']

Model Evaluation

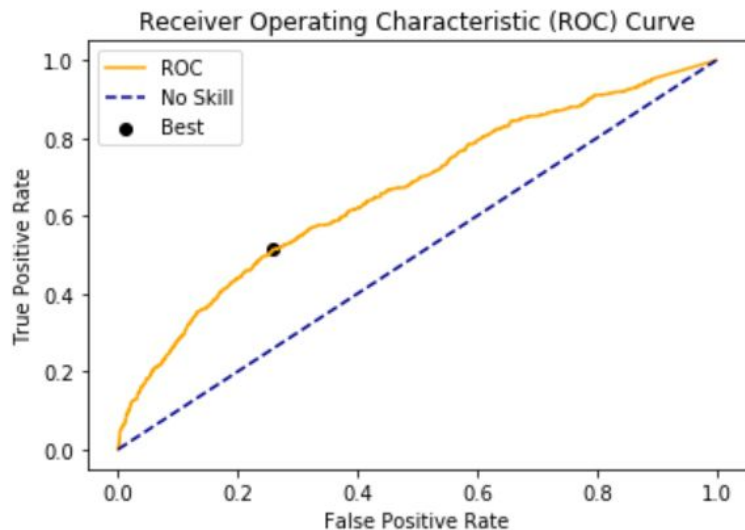
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

Model Evaluation

No Skill: ROC AUC=0.500

LSH: ROC AUC=0.663

Threshold value is: 0.31868131868131866



Recall

0.514

Precision

0.409

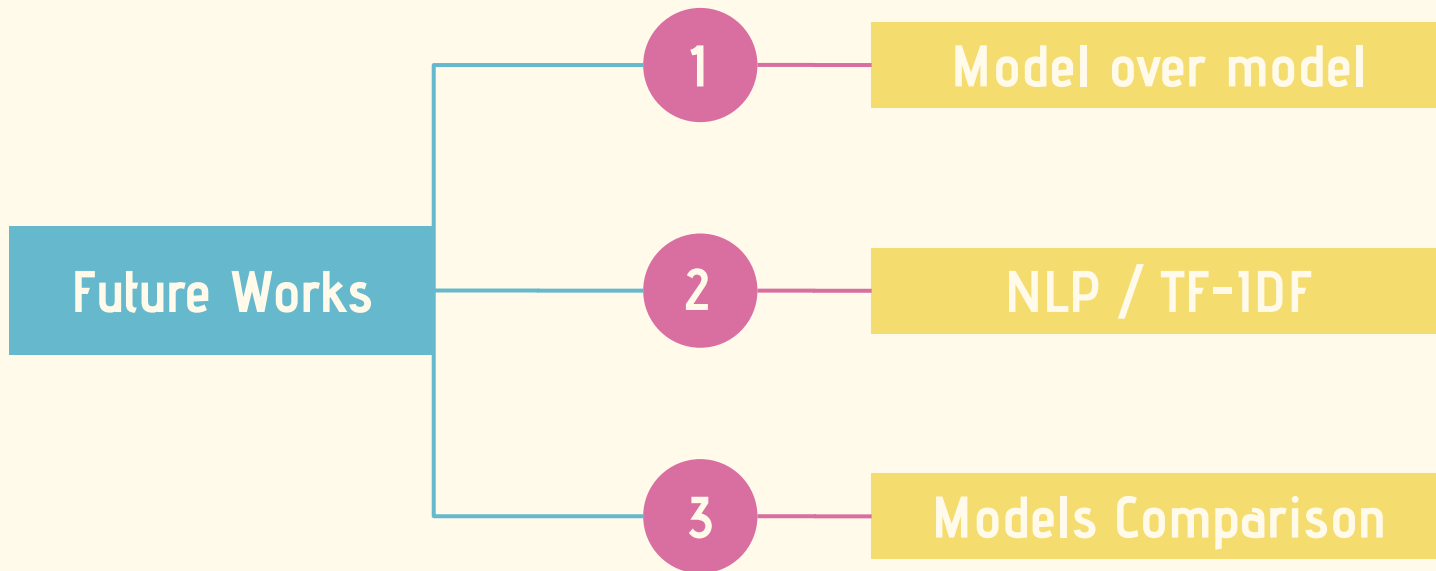
F1-Score

0.456

Accuracy

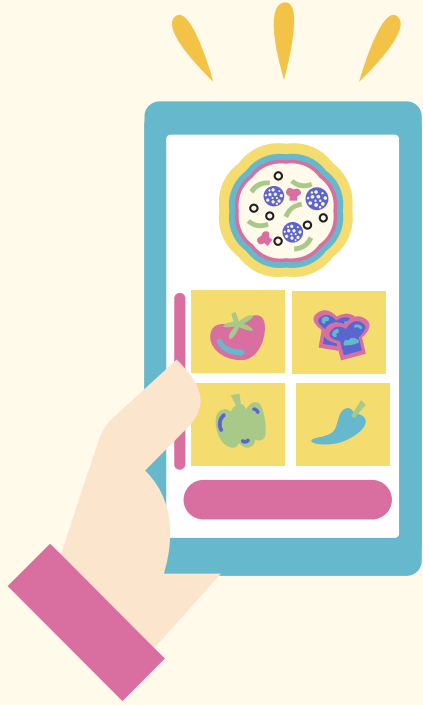
0.682

Future Works





1,800,000 Baht/day



225,000 / day

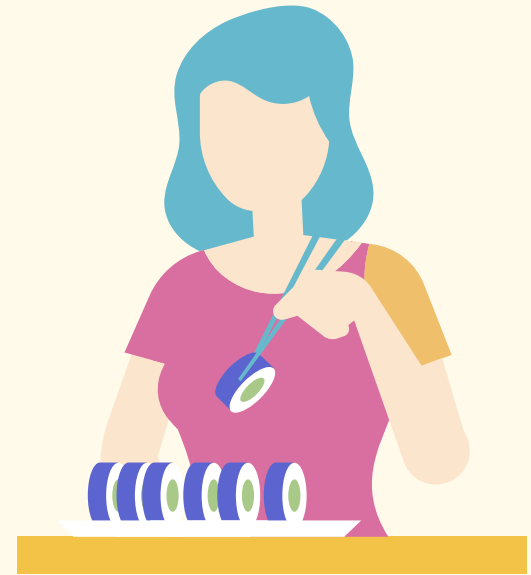
$\pm \frac{1}{2}$ rounds





430K

Restaurants



Advantages

Revenues

Confidences



Creditability

Customer
Satisfactions

True Sharing

Business Partners



Other Business Platform



Team Members



PEARPLOY THIPPRASERT

6310422030

RUETINAN HANPRASOPWAT

6310422033

RAKCHANOK THONGKUMPA

6310422039

ATHITTAYA SRIARAM

6310422045

WEERASAK KAROON

6310422046

NACHASA KHONGCHU

6310422050

Thank you

