SAMSUNG VS APPLE

Project 3 : Web APIs & NLP

CONTENTS

1. Problem Statement

Problem

Marketing student who want to develop NLP for analyzing the competitive dynamics between Samsung and Apple to understand market trends and the factors influencing consumer preferences in the global electronics industry.

Background

The rivalry between Samsung and Apple is a longstanding and prominent competition in the consumer electronics industry.

Objectives

To develop NLP model that can automatically classify post into Samsung and Apple categories and see if can develop into another function that help gathering more data like classify customer sentiments for sentiment analysis or identifying emerging trends and topics in the discussions related to Samsung and Apple for market trend identification that would help with marketing research in the future.

2. Data Collection

Due to company policy, I can’t scraping data by myself.
My data was collected by Nozomi san.

Thank you. You’re the best !!!

2. Data Collection

	vote	title	text	date
0	2	Voice echo galaxy s2	Redditors,\n\n\nI'm using my sgs2 for a month ...	2011-07-28
1	5	Got a new Samsung TV do I have to use the Sams...	The TV is the un46d6000. I found generic wirel...	2011-11-12
2	5	The more I read about the Nexus Prime, the les...	Every day a new reason appears to make the Pri...	2011-11-22
3	2	Reddit, please help me choose (options and pre...	I am from India, and my budget is about 400 US...	2012-01-03
4	2	Samsung Galaxy R Battery?	This\n shows a comparison between Galaxy S2 an...	2012-01-08
...	...	...	...	...
10435	7	Should I be worried about losing the waterproo...	I accidentally placed my S10 on the charger fo...	2020-07-11
10436	2	Screen protector for A71	Can you please suggest some good screen protec...	2020-07-12
10437	1	How secure is Samsung nowadays compared to iOS?	As a br	
10438	1	Turn Subtitles permanently off on USB	Hey guy	
10439	3	Samsung Galaxy J5 problem	Hi everyone	

10440 rows × 4 columns

Samsung

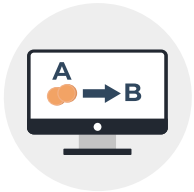
Apple

	vote	title	text	date
0	0	Apple Notes: When I download or share a note I...	As the title states, I create notes with a dar...	2023-10-24
1	8	Daily Advice Thread - July 05, 2023	Welcome to the Daily Advice Thread for /r/Appl...	2023-07-05
2	340	Mysterious air tag in rental van, then caught ...	As title says, we moved using a rental truck. ...	2023-06-25
3	170	Former Samsung Galaxy users, what was the stra...	I don't know if this has been posted already b...	2023-06-23
4	129	Apple Books irritation	So I finally understand why everyone's complai...	2023-06-18
...	...	...	...	...
19026	3	What is a good and free mind-mapping applicati...	I have to write an essay and mindmAps help.	2010-12-01
19027	5	DAE have Mac book pro track pad problems, like...	Not sure if this is the correct subreddit, but...	2010-12-01
19028	8	Bi-monthly "what apps do you use" ipad Posting	News: Washington Post, NYT, WSJ, Slate, the Ec...	2010-12-01
19029	0	Why does OSX fuck up text files all the time?	Is there a way to make TextEdit not take every...	2010-12-01
19030	4	I love AppleCare. (Update to SuperDrive issue ...	(\nHere is a link to the original post\n)\n\n\...	2010-12-01

19031 rows × 4 columns

3. Data Cleaning & Preprocessing

1



Check duplicated data and null value : None

2

labeling data by topic ‘Samsung’ & ‘Apple’

3

Concat 2 dataframe , reset index ,create new column for year and month of the post and drop column ‘date’

apple	19031
samsung	10440

apple	11419
samsung	10440

4

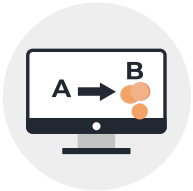
Check data balance & random drop data in apple 40%

5

Text Cleaning

- Remove HTML tags
- Remove non-alphanumeric characters
- Remove extra white spaces
- Make all text lowercase

6

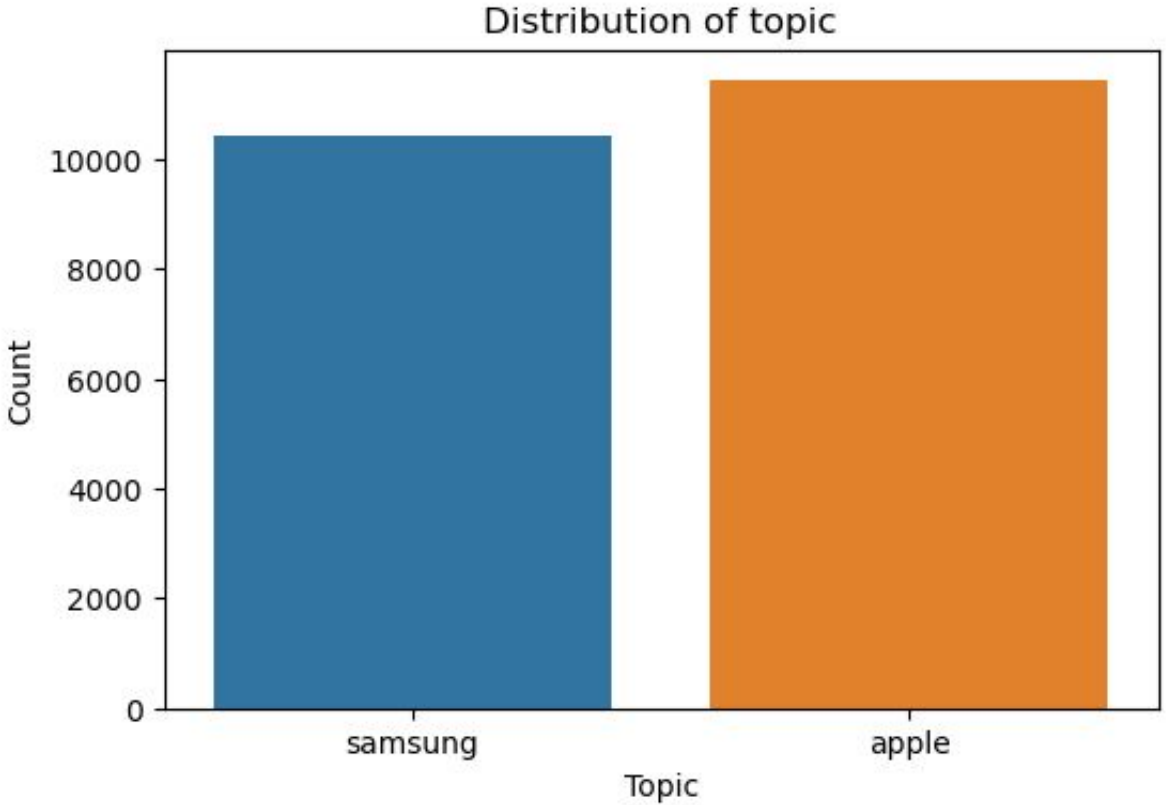


Create new column for post length & word count

Lemmatizing & Stop Word Removal

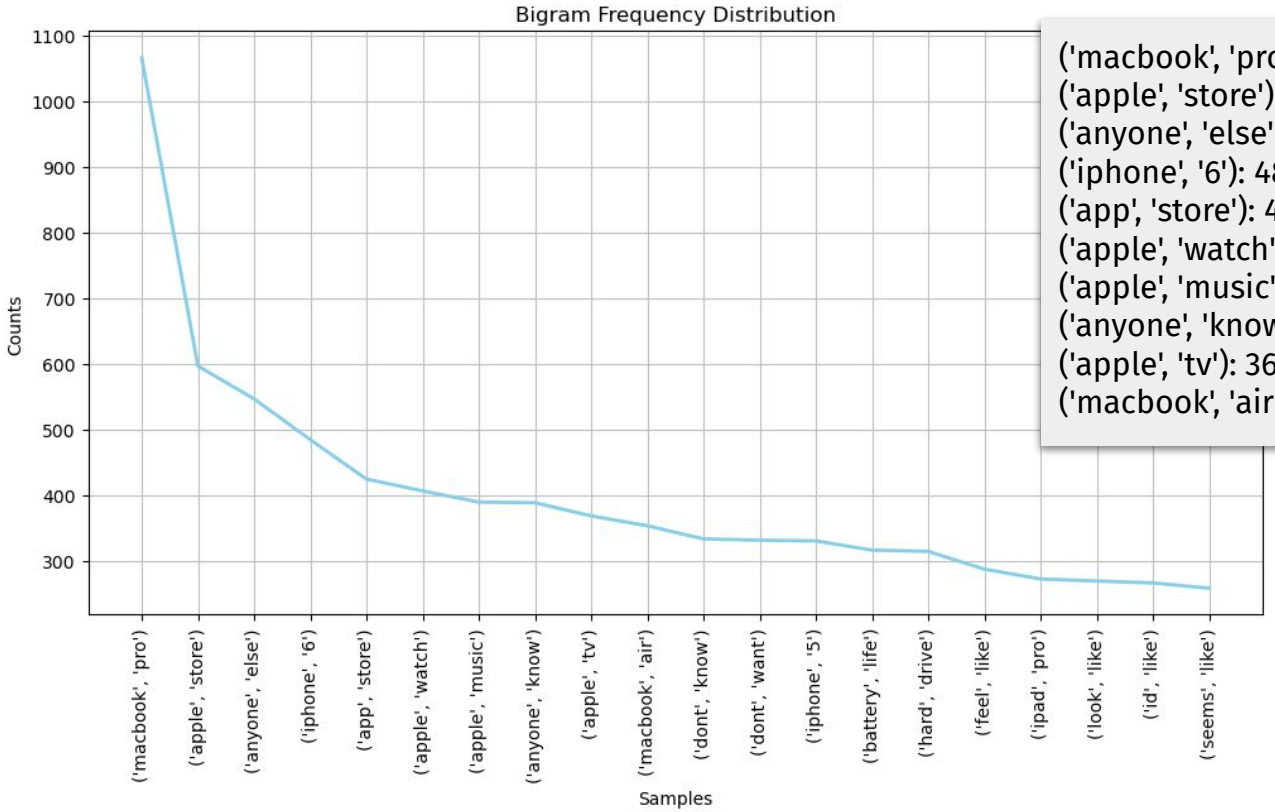
	vote	title	text	topic	year	month	post_length	word_count	tokenized	lemmatized	stemmed	lem_no_stopwords
0	2	Voice echo galaxy s2	redditors im using my sgs2 for a month now wor...	samsung	2011	7	404	74	redditors im using my sgs2 for a month now wor...	redditors im using my sgs2 for a month now wor...	redditor im use my sgs2 for a month now work g...	redditors using sgs2 month work great voicecal...
1	5	Got a new Samsung TV do I have to use the Sams...	the tv is the un46d6000 i found generic wirele...	samsung	2011	11	131	27	the tv is the un46d6000 i found generic wirele...	the tv is the un46d6000 i found generic wirele...	the tv is the un46d6000 i found gener wireless...	tv un46d6000 found generic wireless usb adapte...
2	5	The more I read about the Nexus Prime, the les...	every day a new reason appears to make the pri...	samsung	2011	11	601	111	every day a new reason appears to make the pri...	every day a new reason appears to make the pri...	everi day a new reason appear to make the prim...	every day new reason appears make prime look l...
3	2	Reddit, please help me choose (options and pre...	i am from india and my budget is about 400 usd...	samsung	2012	1	767	142	i am from india and my budget is about 400 usd...	i am from india and my budget is about 400 usd...	i am from india and my budget is about 400 usd...	india budget 400 usd 20000 indian rupee shorti...
4	2	Samsung Galaxy R Battery?	this shows a comparison between galaxy s2 and ...	samsung	2012	1	634	133	this shows a comparison between galaxy s2 and ...	this show a comparison between galaxy s2 and g...	thi show a comparison between galaxi s2 and ga...	show comparison galaxy s2 galaxy r battery spe...

4. Exploratory Data Analysis



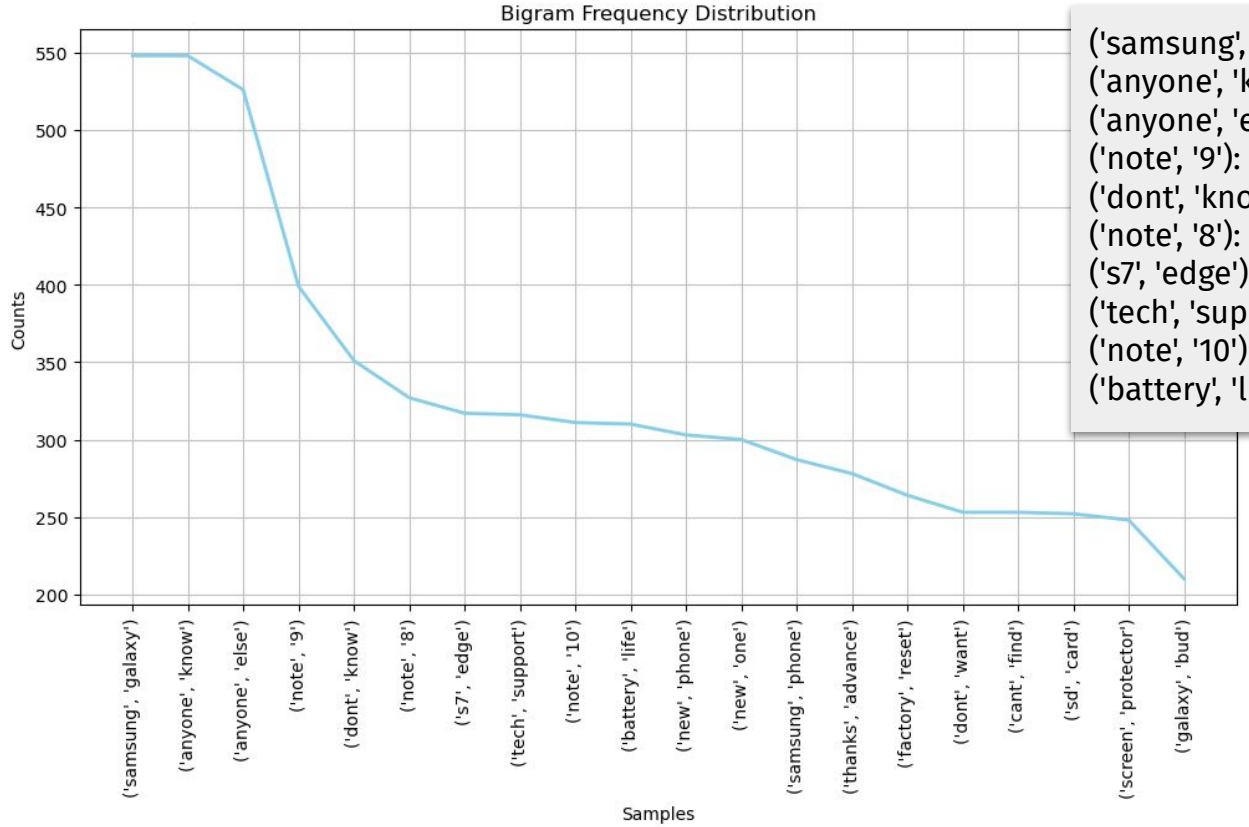
Apple : 0.522415
Samsung : 0.477585

4. Exploratory Data Analysis



('macbook', 'pro'): 1066
('apple', 'store'): 597
('anyone', 'else'): 547
('iphone', '6'): 485
('app', 'store'): 425
('apple', 'watch'): 407
('apple', 'music'): 390
('anyone', 'know'): 389
('apple', 'tv'): 369
('macbook', 'air'): 354

4. Exploratory Data Analysis

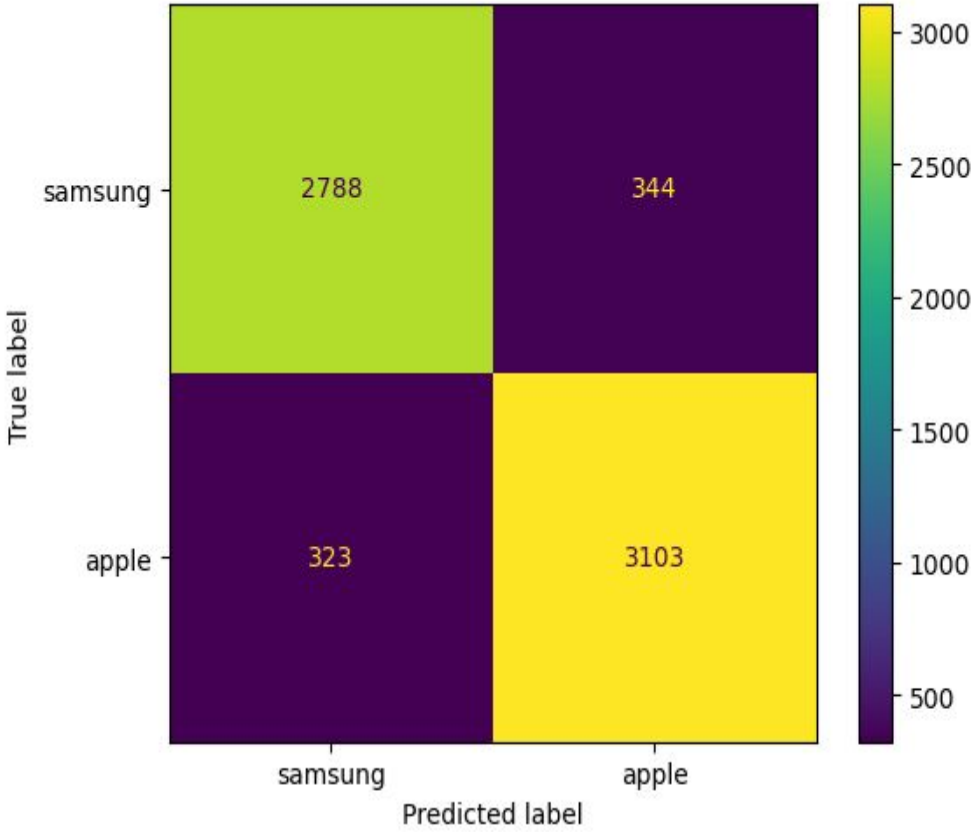


('samsung', 'galaxy'): 548
('anyone', 'know'): 548
('anyone', 'else'): 526
('note', '9'): 399
('dont', 'know'): 351
('note', '8'): 327
('s7', 'edge'): 317
('tech', 'support'): 316
('note', '10'): 311
('battery', 'life'): 310

5. Modeling & Evaluation

Logistic Regression

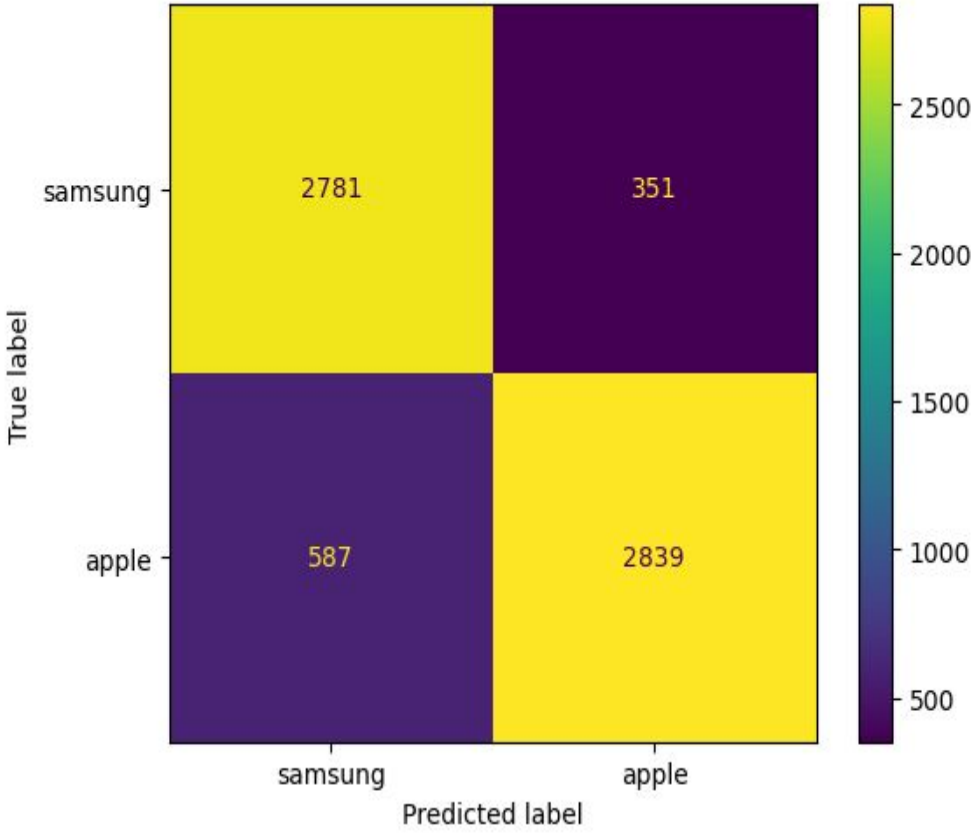
Train Accuracy	Test Accuracy	Precision	Recall	F1 Score
0.98313	0.90134	0.90020	0.90572	0.90295



5. Modeling & Evaluation

Naive Bayes

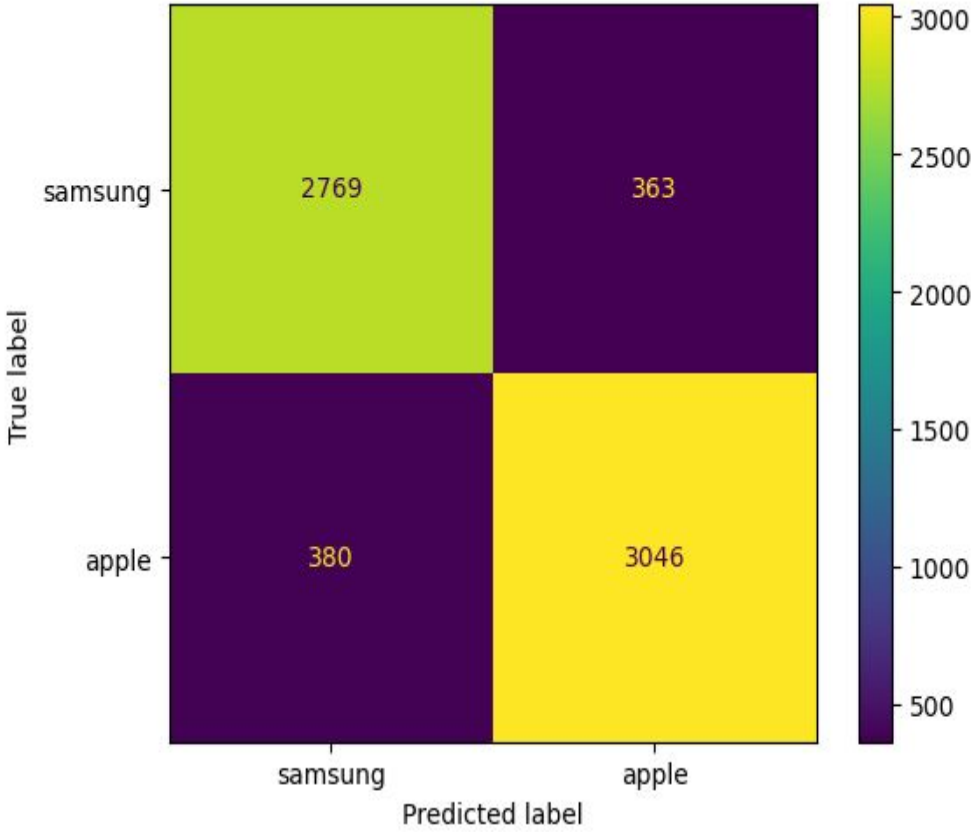
Train Accuracy	Test Accuracy	Precision	Recall	F1 Score
0.87000	0.85696	0.88996	0.82866	0.85822



5. Modeling & Evaluation

Random Forests

Train Accuracy	Test Accuracy	Precision	Recall	F1 Score
0.99405	0.88670	0.89351	0.88908	0.89129



Model Performance

Model	Vectorizer	Train Accuracy	Test Accuracy	Precision	Recall	F1 Score
Logistic Regression	CVEC	0.98313	0.90134	0.90020	0.90572	0.90295
Naive Bayes	CVEC	0.87000	0.85696	0.88996	0.82866	0.85822
Random Forests	CVEC	0.99405	0.88670	0.89351	0.88908	0.89129

6. Conclusions & Recommendations

Conclusions

NLP model have ability to classify post into the categories of Samsung and Apple with acceptable in a prediction model accuracy.So it can develop to classify more detail in future , for next step should be develop with similar data like categorize brand mentions for Samsung and Apple and classify customer sentiments expressed in reviews and social media comments.

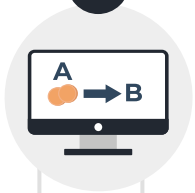
Recommendations

Exploring more additional features, such as competitor mentions to identify and extract mentions of competitors in the text. Understanding how Samsung and Apple are discussed in comparison to each other and to other competitors can be informative.

THANK YOU

3. Data Cleaning & Preprocessing

1



Check duplicated data and null value : None

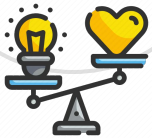
2

labeling data by topic ‘Samsung’ & ‘Apple’

3

Concat 2 dataframe , reset index ,create new column for year and month of the post and drop column ‘date’

4



Check data balance & random drop data in apple 40%

apple	19031
samsung	10440

apple	11419
samsung	10440

5

Text Cleaning

- Remove HTML tags
- Remove non-alphanumeric characters
- Remove extra white spaces
- Make all text lowercase

6

Create new column for post length & word count