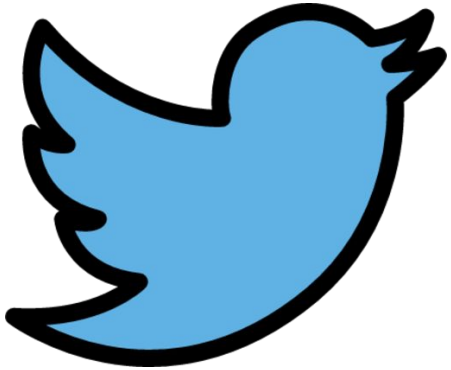
Twitter Sentiment Analysis

Overview



To do: Build an NLP model that can rate the sentiment of a Tweet based on its content.

Data: Twitter sentiment about Apple and Google products. The dataset comes from CrowdFlower via data.world.

Supervised: Human raters rated the sentiment in over 9,000 tweets.

The Data

The dataset has three features

- tweet_text → tweet
- emotion_in_tweet_is_directed_at → product
- is_there_an_emotion_directed_at_a_brand_or_product → sentiment

‘Tweet’ makes up 9,093 tweets

‘Product’ has 10 unique values

- iPhone, iPad or iPhone App, iPad, Google, Android, Apple, Android App, Other Google product or service, Other Apple product or service, and NaN

‘Sentiment’ has four unique values

- Negative emotion, Positive emotion, No emotion toward brand or product, and I can’t tell

Data Preprocessing

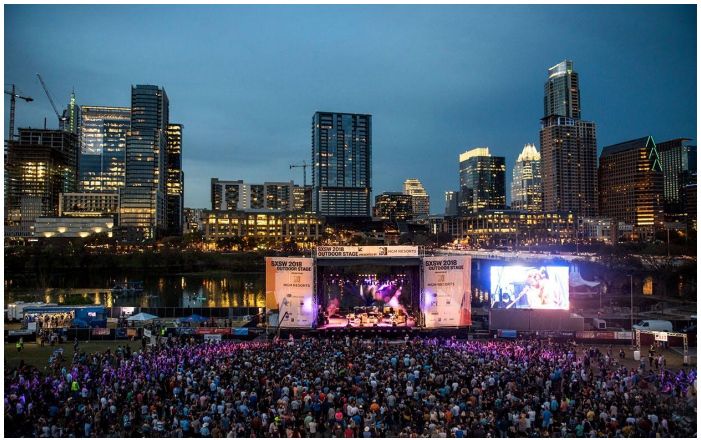
Remove usernames, hashtags, and hyperlinks from the text

.@wesley83 I have a 3G iPhone. After 3 hrs tweeting at #RISE_Austin, it was dead! I need to upgrade. Plugin stations at #SXSW.	. I have a 3G iPhone. After 3 hrs tweeting at , it was dead! I need to upgrade. Plugin stations at .
Ipad everywhere. #SXSW {link}	Ipad everywhere.

There were some missing values.

- One in the tweet column → dropped
- 5,802 in the product column
  - Changed to → “no answer”

South by Southwest(SXSW), Austin, Texas →



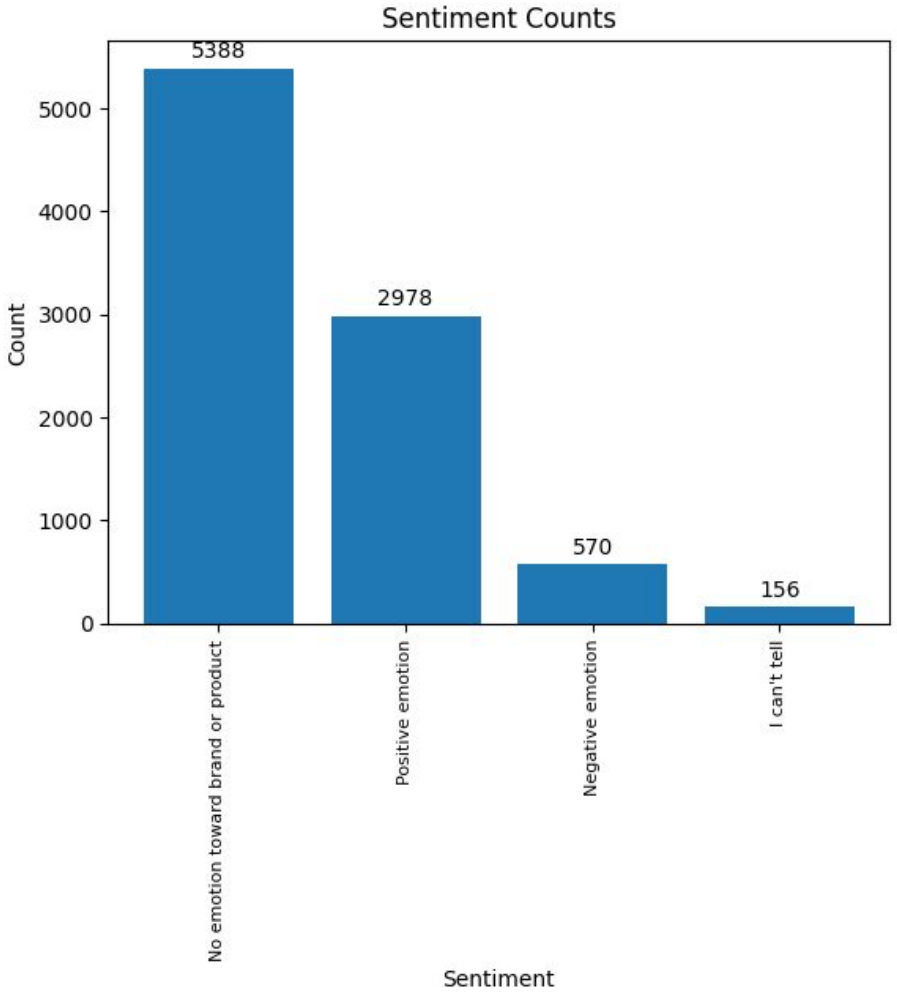
The Great Imbalance

Ternary

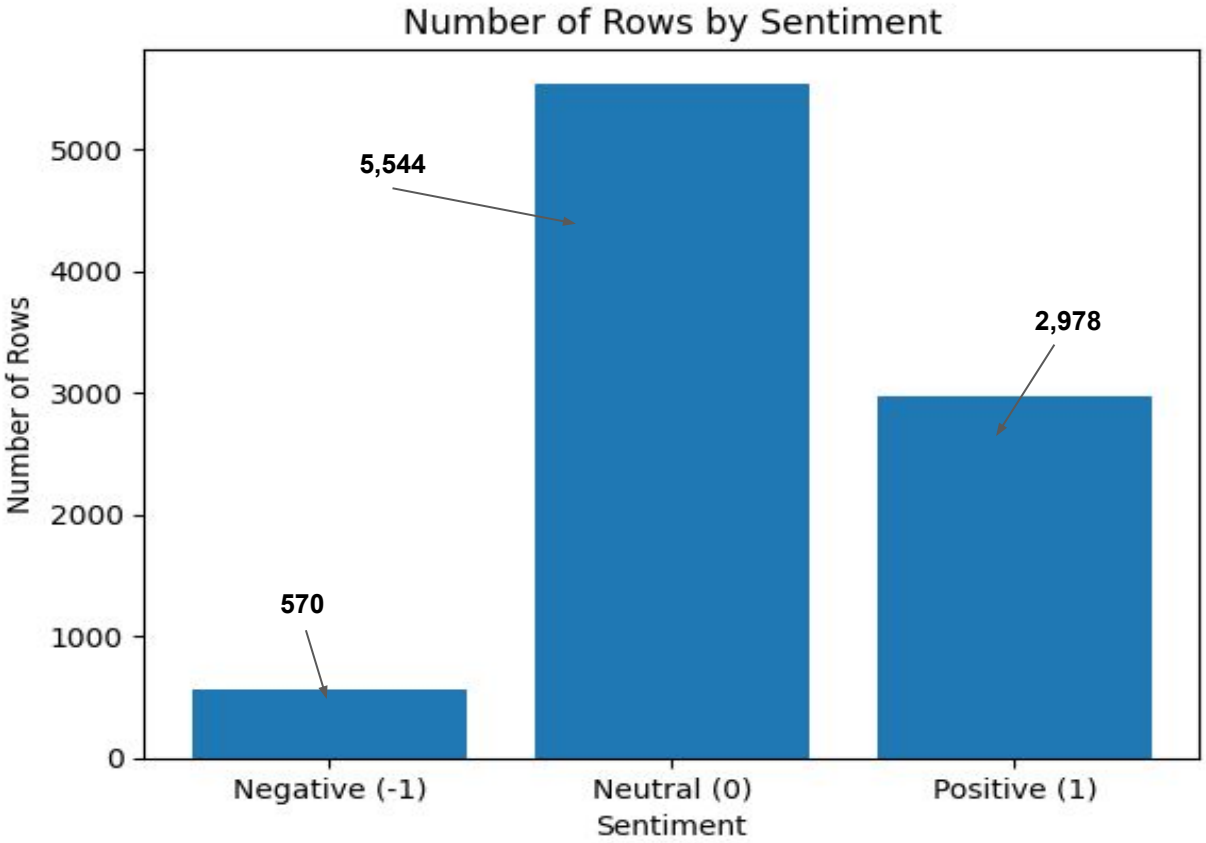
Neutral = “No emotion toward brand or product” + “I can’t tell”
Positive = “Positive emotion”
Negative = “Negative emotion”

Binary

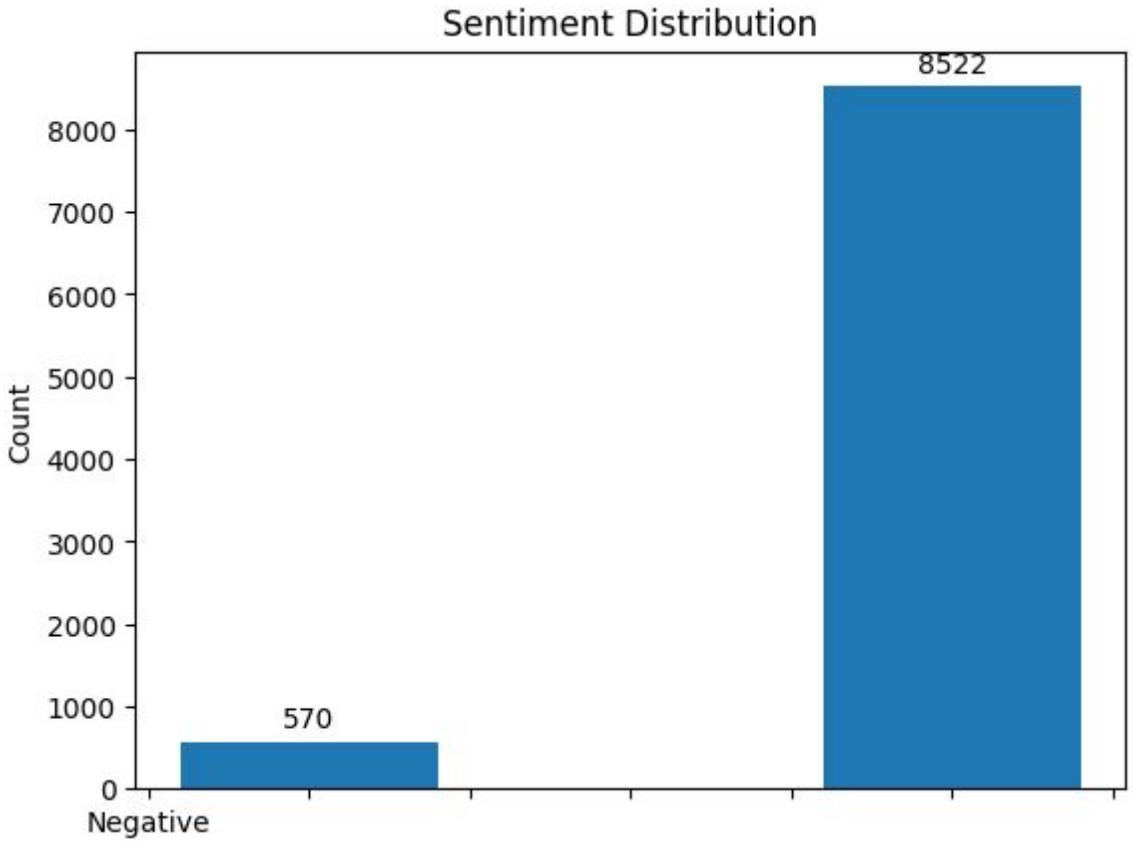
Not Negative = “No emotion toward brand or product” + “I can’t tell” + “Positive Emotion”
Negative = “Negative emotion”



Ternary

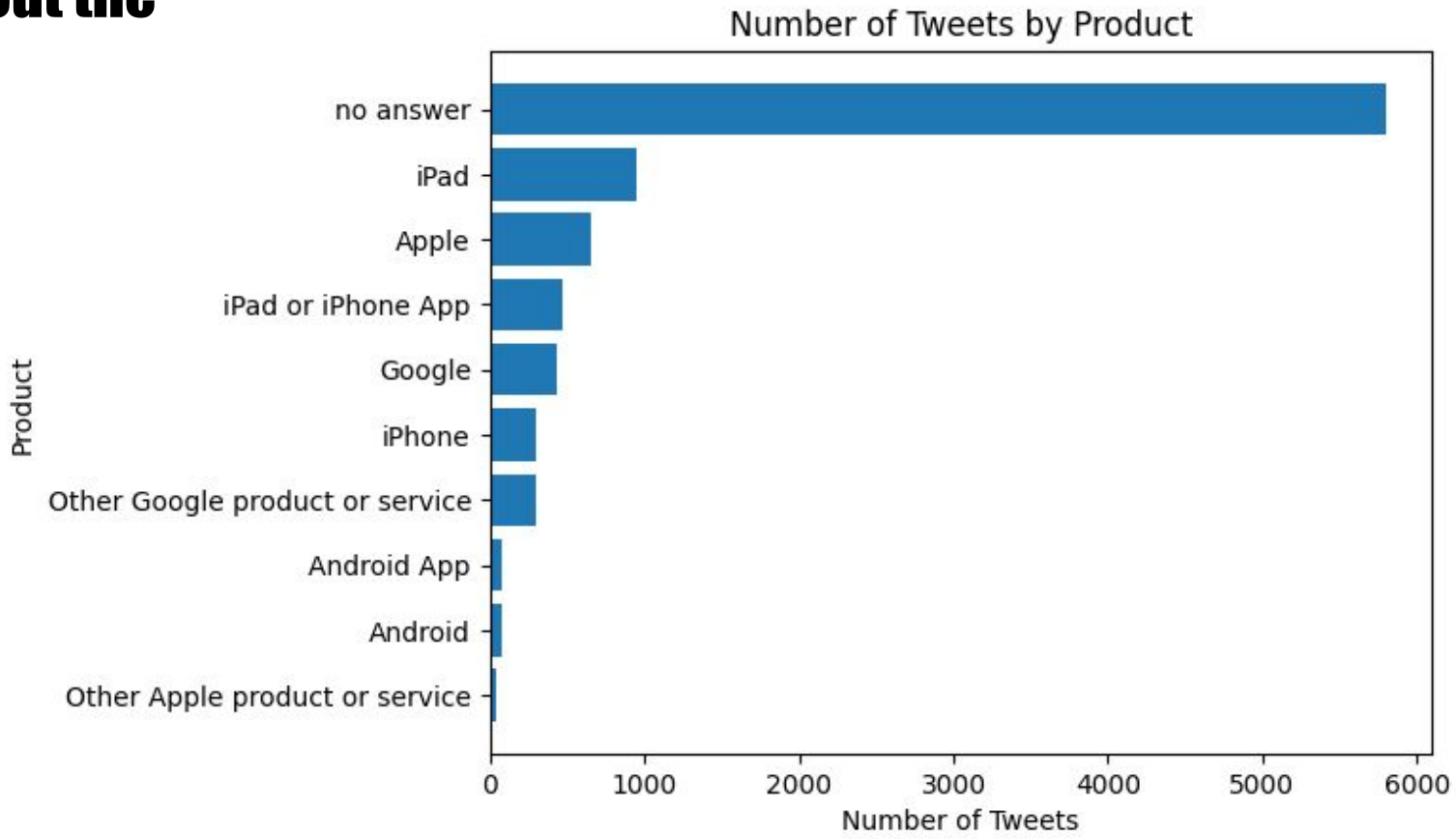

Number of Rows by Sentiment
Number of Rows
5000
4000
3000
2000
1000
0
5,544
2,978
570
Negative (-1)
Neutral (0)
Positive (1)
Sentiment

Binary

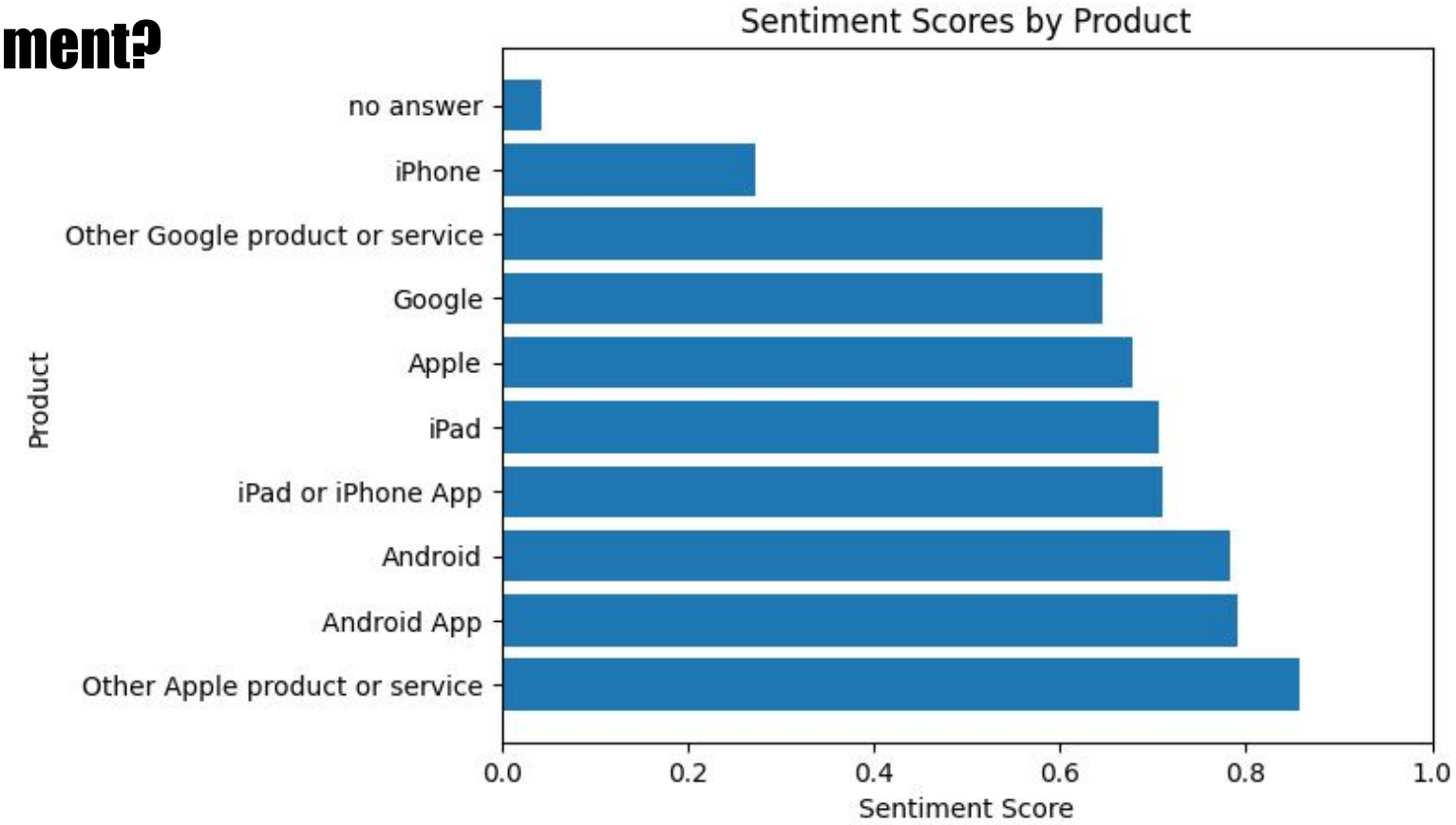

Sentiment Distribution
8522
570
Count
8000
7000
6000
5000
4000
3000
2000
1000
0
Negative

The Problem with Imbalanced Data

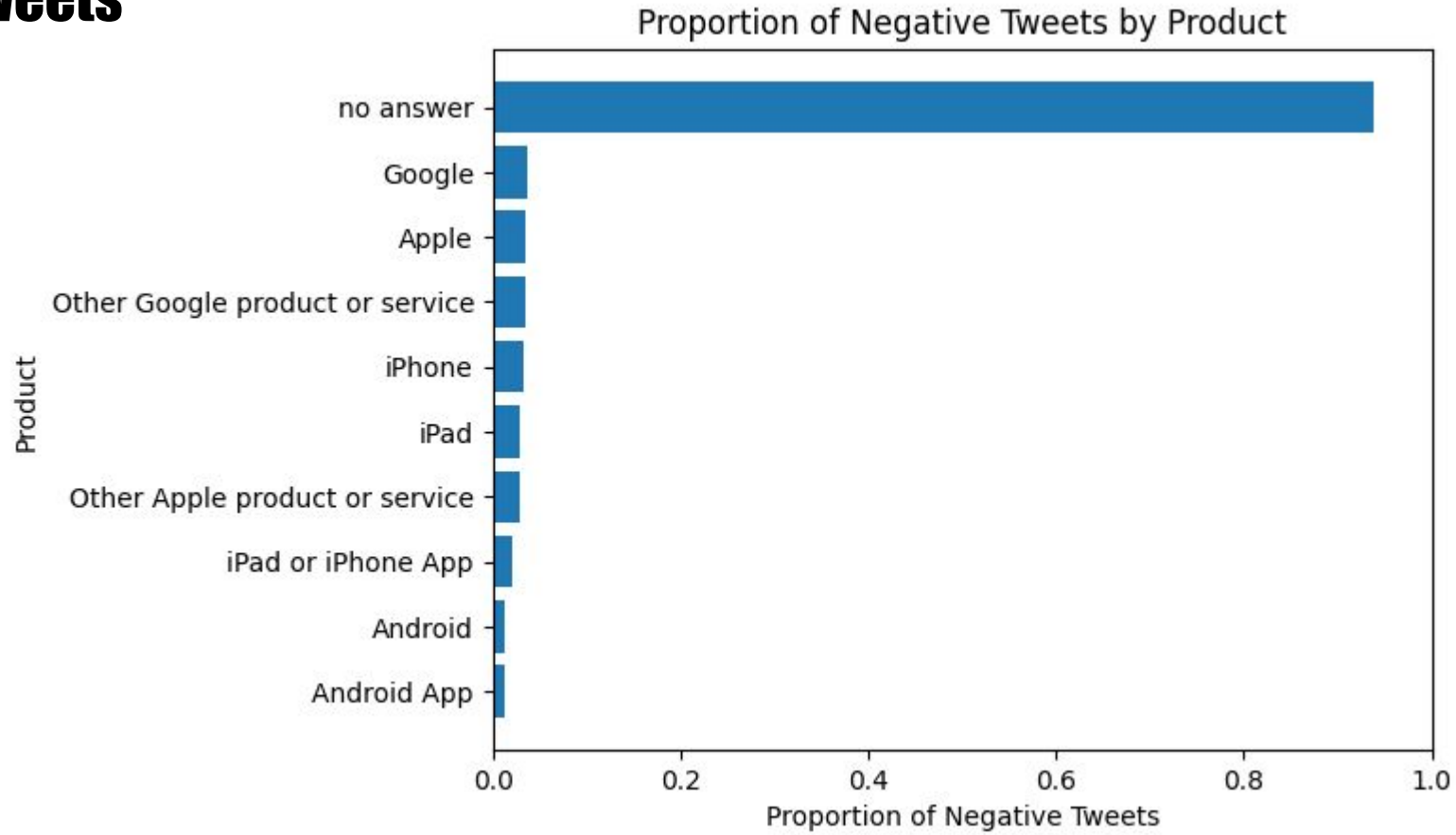
Which products were tweeted about the most?



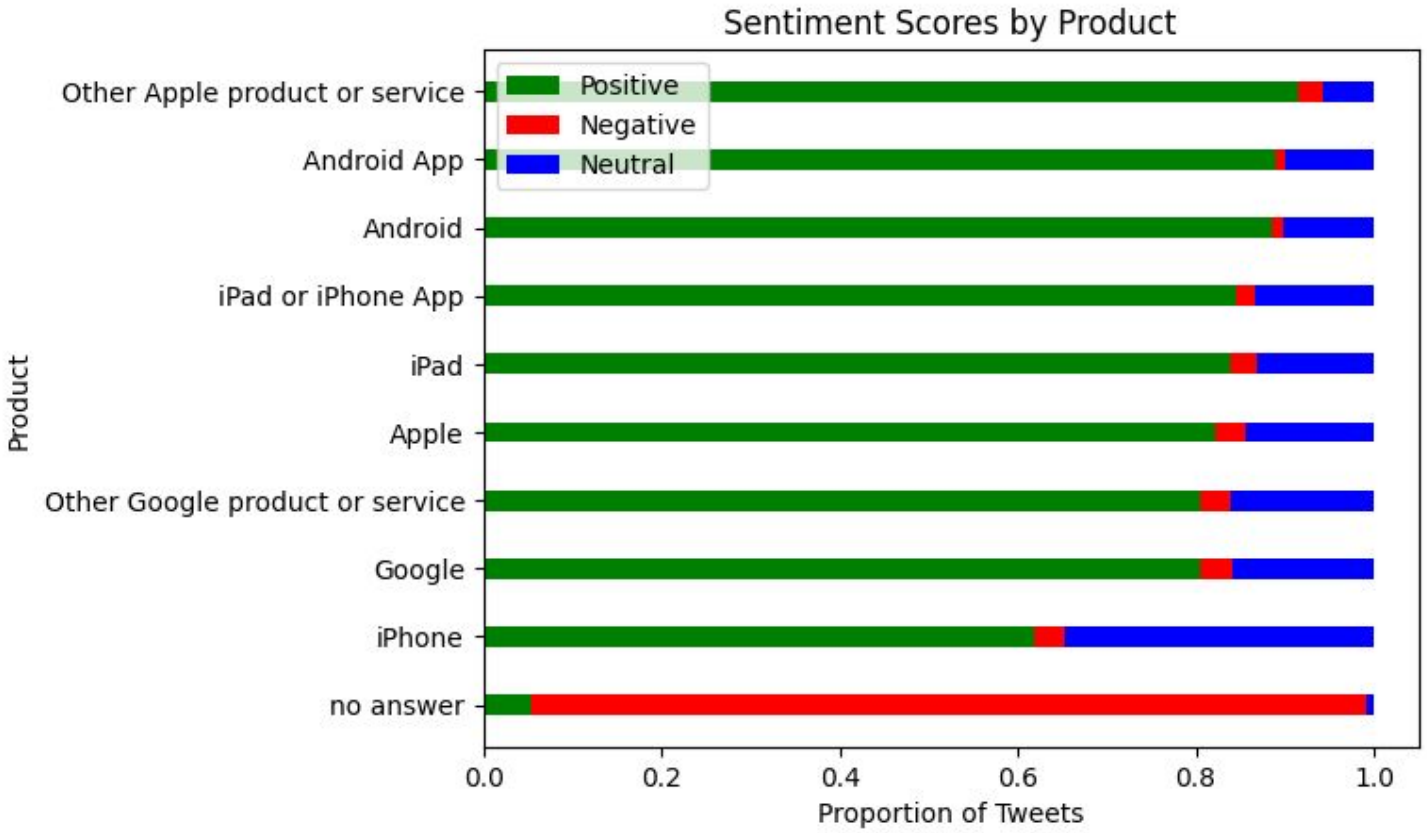
Which Product had the highest overall sentiment?



What are the negative tweets among the products?



How does the sentiment of tweets compare among the different products?



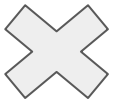
NLP Preprocessing

After data preprocessing	. I have a 3G iPhone. After 3 hrs tweeting at , it was dead! I need to upgrade. Plugin stations at .

Remove punctuation marks, special characters, convert all letters to lowercase	i have a 3g iphone after 3 hrs tweeting at it was dead i need to upgrade plugin stations at

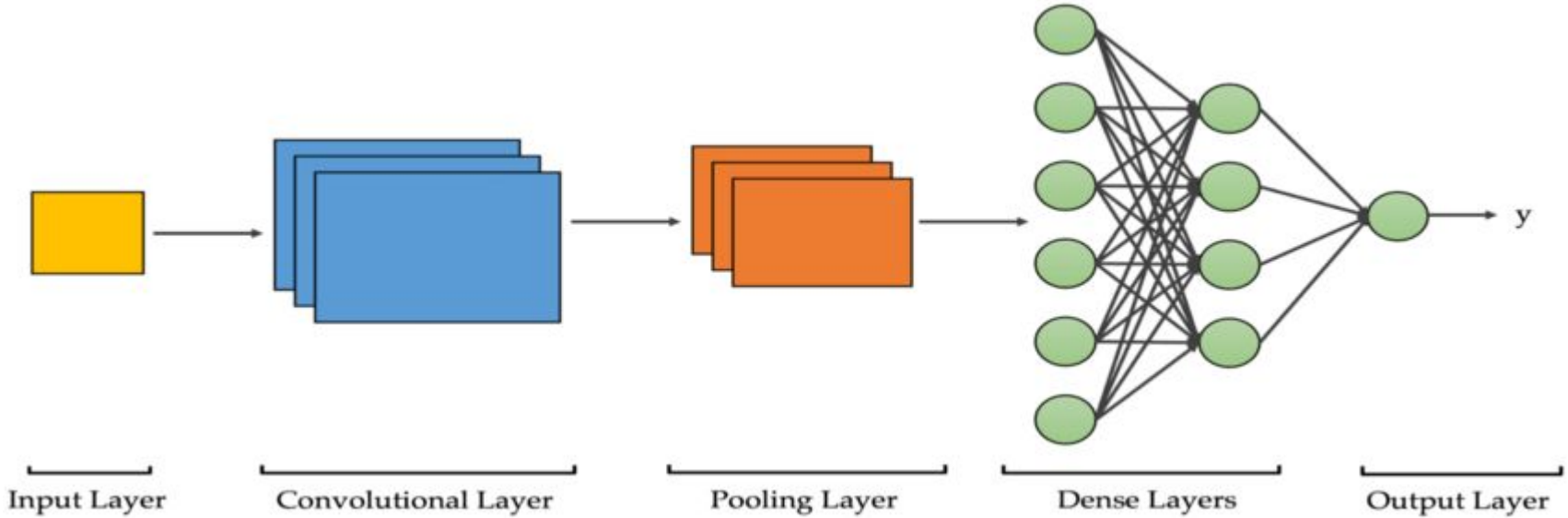
Tokenize the text into individual words	'i', 'have', 'a', '3g', 'iphone', 'after', '3', 'hrs', 'tweeting', 'at', 'it', 'was', 'dead', 'i', 'need', 'to', 'upgrade', 'plugin', 'stations', 'at'

Perform lemmatization on the tokenized text	‘3g', 'iphone', '3', 'hr', 'tweeting', 'dead', 'need', 'upgrade', 'plugin', 'station'

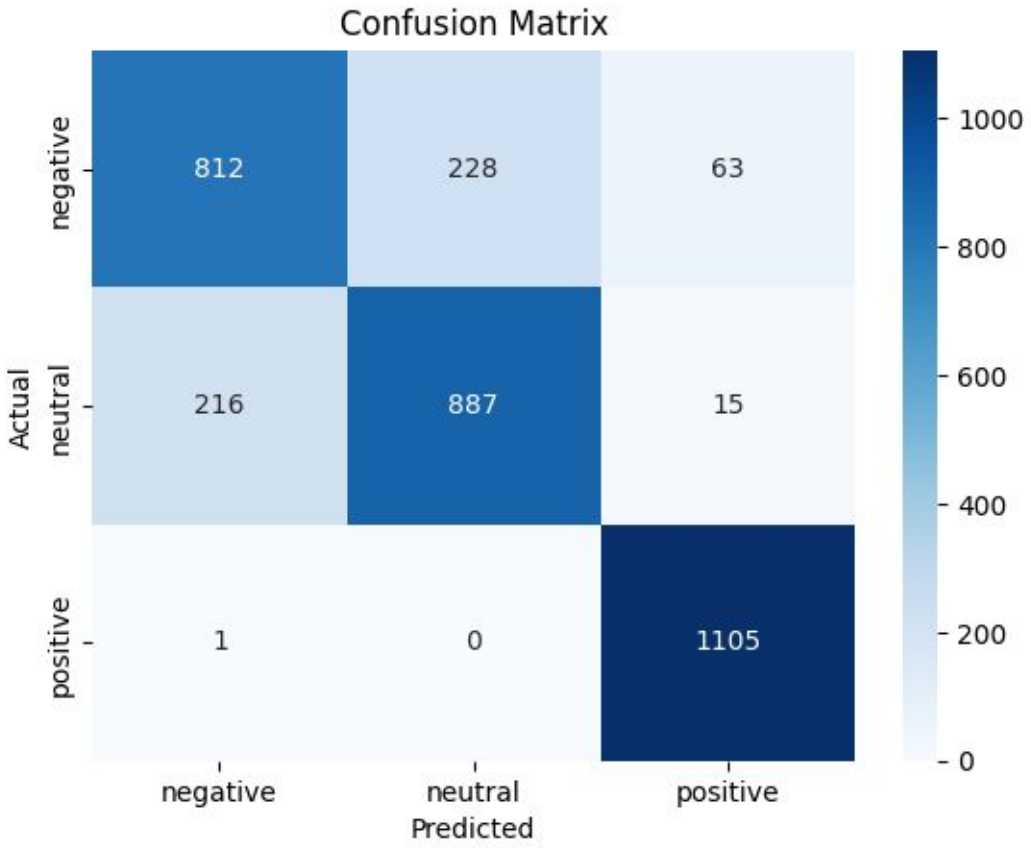


Perform stemming on the tokenized text	'3g', 'iphon', '3', 'hr', 'tweet', 'dead', 'need', 'upgrad', 'plugin', 'station'

Best Model classifying multiclass



Best Model classifying multiclass



Seven Layers

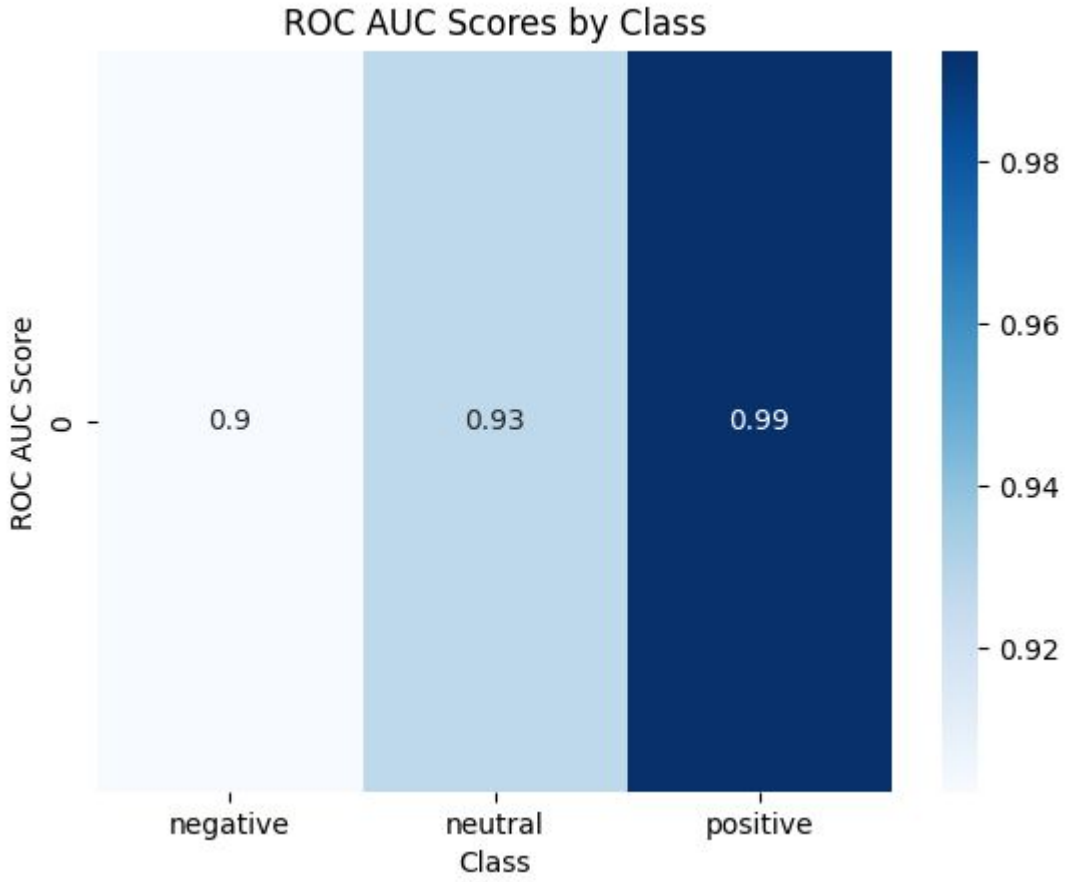
1. Embedding
2. Conv1D
3. Dropout
4. MaxPooling1D
5. Conv1D
6. Dropout
7. MaxPooling1D

Dense Output layer/softmax activation

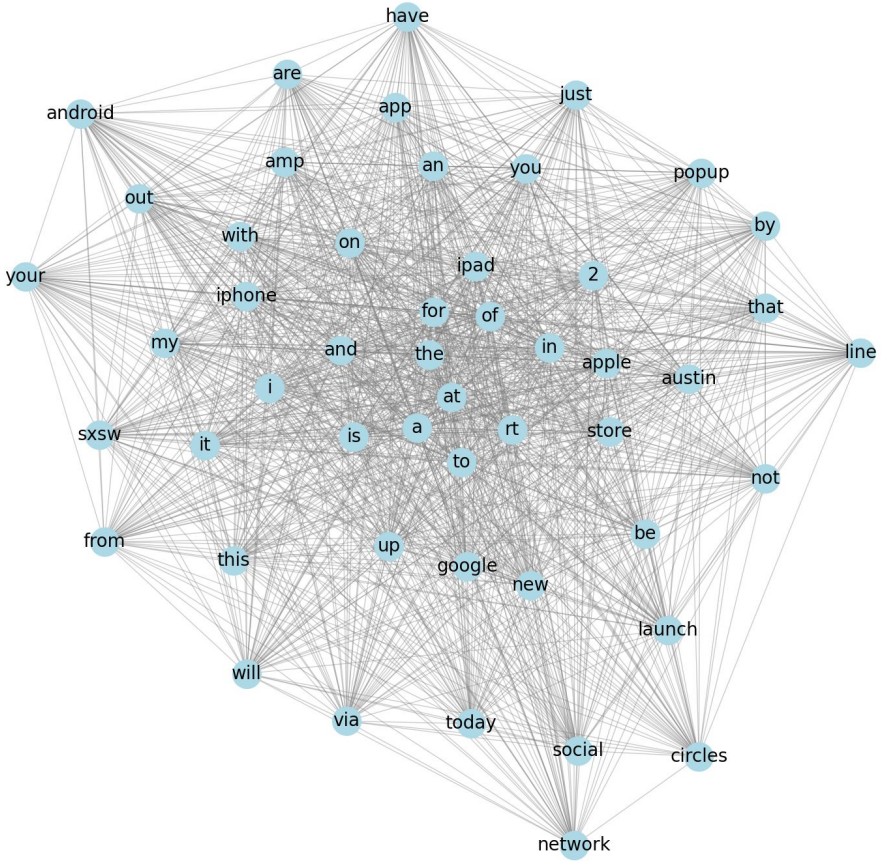
Training Accuracy: 95.4%

Test Accuracy: 84.2%

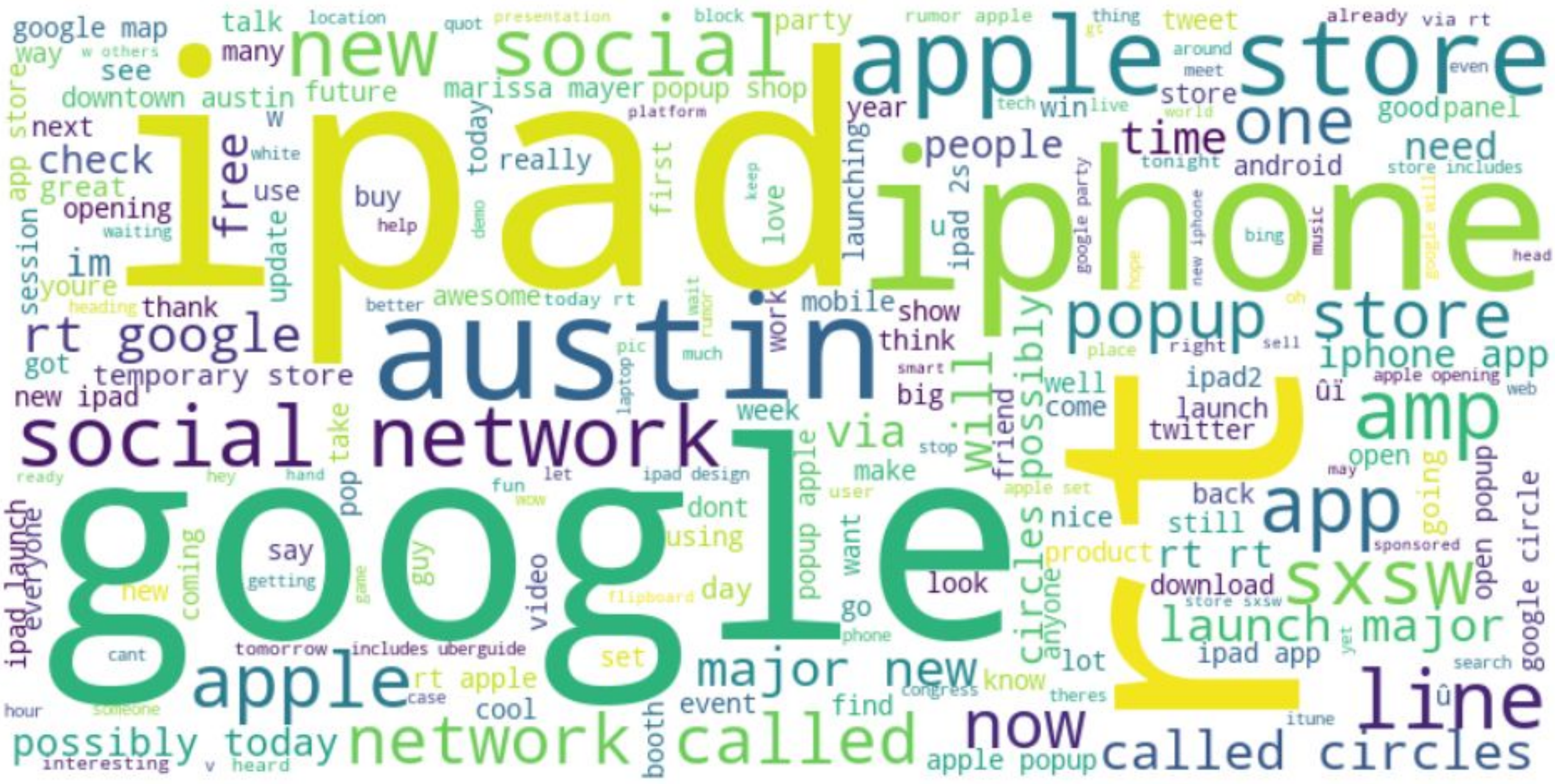
Strengths & Weaknesses



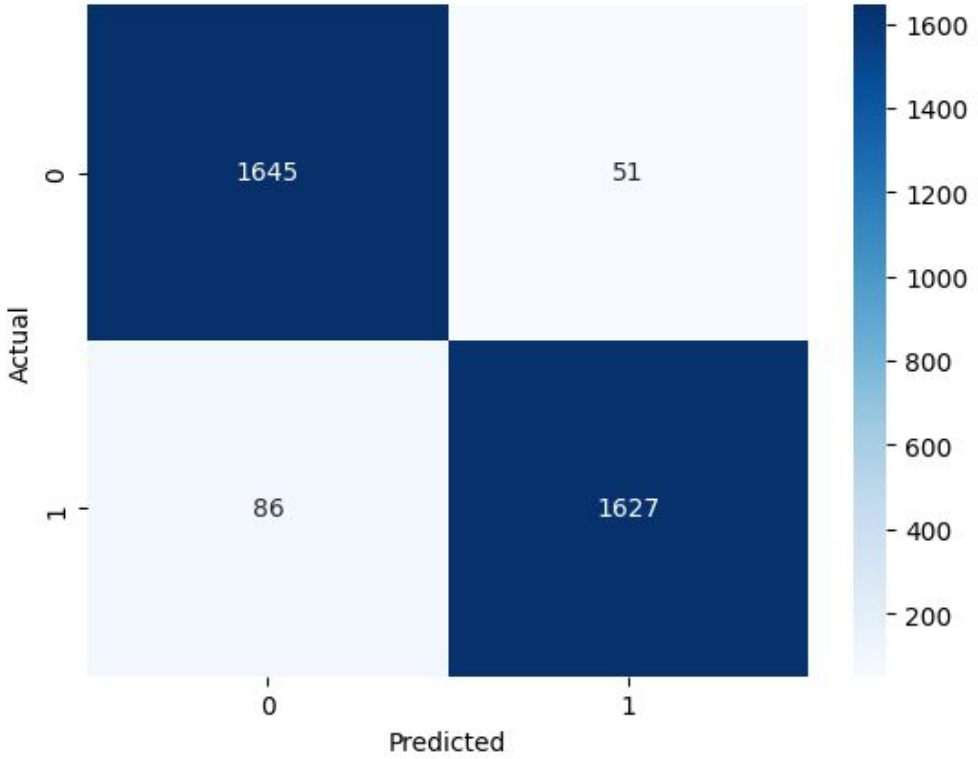
Network Graph



Word Cloud


ipad
google
iphone
austin
rt
apple store
new social
social network
network
called
apple
now
line
sxsw
app
amp
popup store
one
called circles
major new
launch major
iphone app
rt google
possibly today
temporary store
downtown austin
marissa mayer
popup shop
google map
people
time
need
android
apple popup
free
check
great
opening
im
youre
thank
got
new ipad
possibly
circles
will
via
launching
ipad 2s
love
first
today
really
buy
use
update
session
app store
next
see
way
many
future
talk
location
quot
presentation
block
party
rumor apple
thing
tweet
already
via rt
around
meet
even
store
win
good
panel
year
tonight
store includes
bing
music
head
mobile
show
think
big
week
make
dont
using
day
go
set
cool
event
find
know
lot
ipad app
download
rt rt
product
nice
still
back
twitter
launch
well
come
ipad2
open
going
open popup
google circle
friend
anyone
pop
take
video
coming
everyone
ipad launch
say
guy
new
cant
hour
tomorrow
includes uberguide
rt apple
booth
interesting
heard
awesome
better
today rt
work
much
pic
place
right
sell
apple opening
web
smart
stop
sponsored
search
itune
theres
congress
popup apple
want
look
ipad design
hand
hey
ready
let
fun
user
apple set
may
flipboard
getting
game
case
phone
help
white
waiting
demo
heading
keep
wait
rumor
laptop
google party
new iphone
hope
tech
live
world
platform
w
w others
google will
u

Best Model Classifying Binary



Count Vectorization
Multinomial NB
Training Accuracy: 97%
Test Accuracy: 96.4%

Misclassifications

tweet	predicted label	actual label
my iphone was stolen and i got it back	negative	not negative
any ipad djs here need one for 3am tuesday douche bags need not apply	negative	not negative
this group next to me has 6 ppl the table everyone is using their phoneipad instead of talking to each other	negative	not negative
rt epic theres just one guy waiting in line for the ipad 2 in austin at sxsw	not negative	negative

Conclusion

Recommendations:

More data!

Next Steps:

Deploy the model to perform real-time sentiment analysis on new tweets

(Personal)Would like to see how it does on YouTube video comments

Thank You

Jennifer Casias

E-mail: casiasjc@gmail.com

Data Preprocessing

tweet

product

sentiment

	tweet_text	emotion_in_tweet_is_directed_at	is_there_an_emotion_directed_at_a_brand_or_product
0	.@wesley83 I have a 3G iPhone. After 3 hrs twe...	iPhone	Negative emotion
1	@jessedee Know about @fludapp ? Awesome iPad/i...	iPad or iPhone App	Positive emotion
2	@swonderlin Can not wait for #iPad 2 also. The...	iPad	Positive emotion
3	@sxsw I hope this year's festival isn't as cra...	iPad or iPhone App	Negative emotion
4	@sxtxstate great stuff on Fri #SXSW: Marissa M...	Google	Positive emotion
...	...	...	...
9088	Ipad everywhere. #SXSW {link}	iPad	Positive emotion
9089	Wave, buzz... RT @mention We interrupt your re...	NaN	No emotion toward brand or product
9090	Google's Zeiger, a physician never reported po...	NaN	No emotion toward brand or product
9091	Some Verizon iPhone customers complained their...	NaN	No emotion toward brand or product
9092	Ïį Ïà ü_ Ê Î Ò £ Á âå _ £ â_ ÛâRT @...	NaN	No emotion toward brand or product

What if...

I dropped those NaNs from ‘Product’ instead?

9,093 rows → 3,291 rows

Phew!

Still a HUGE imbalance!