Twitter Sentiment Analysis

Sean Hart September 15, 2022

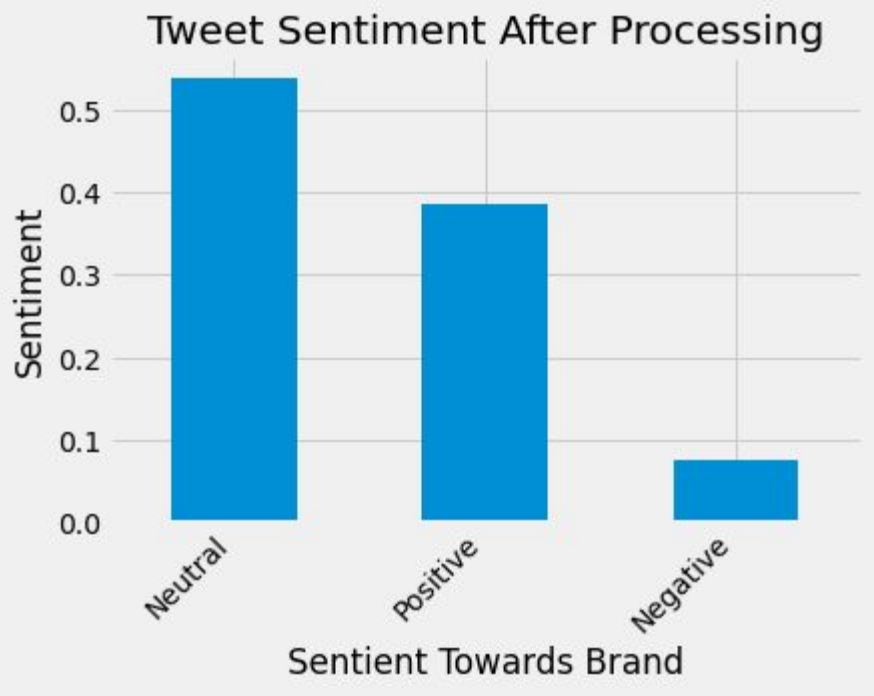
01

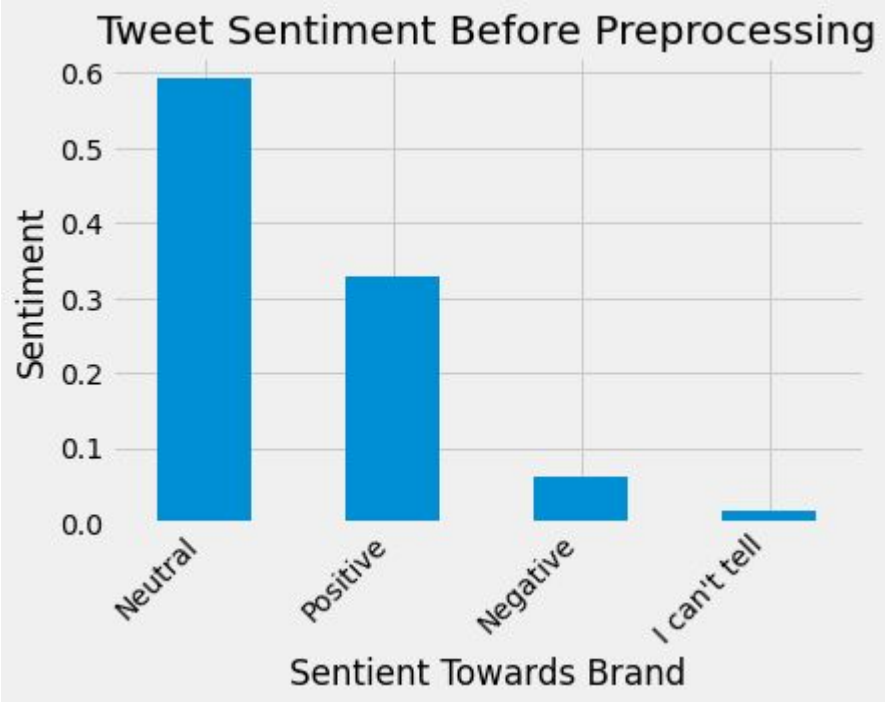
The Dataset

Initial Dataset

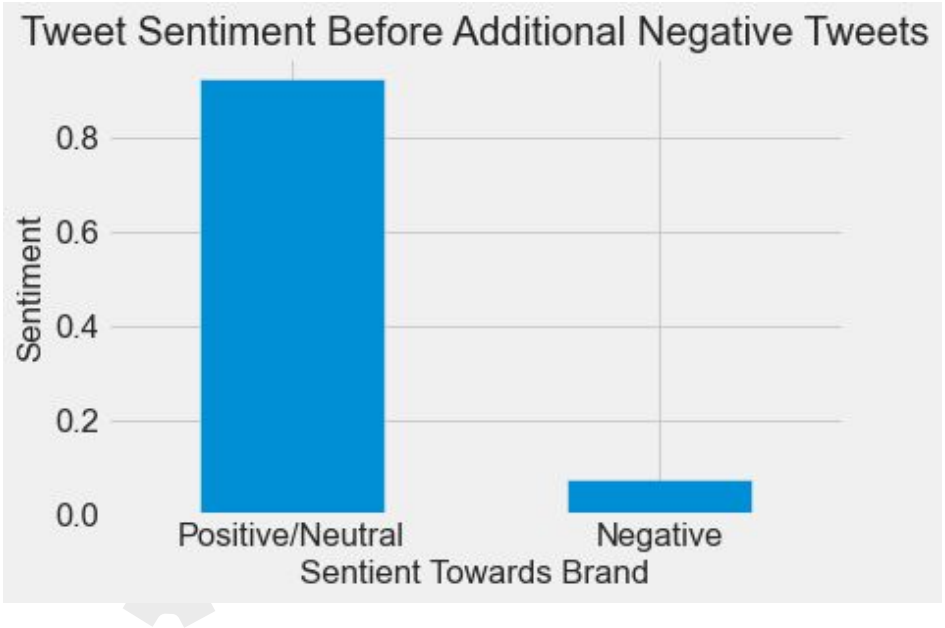
- 9902 Tweets from 2011 South by Southwest Conference
- Dataset included Tweet, Sentiment, and Object of Sentiment
- The dataset was highly imbalanced – very few negative sentiment examples.

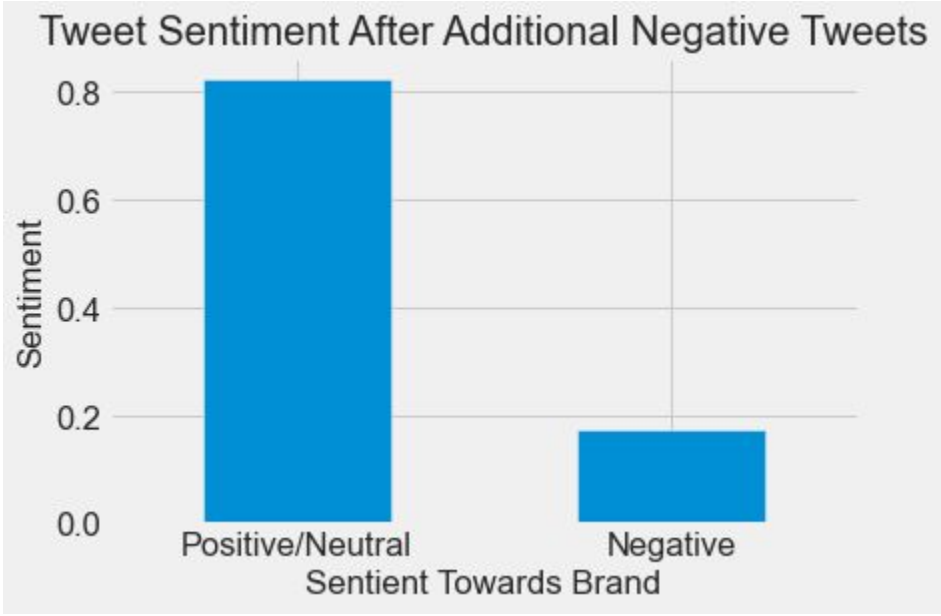
Sentiment Breakdown



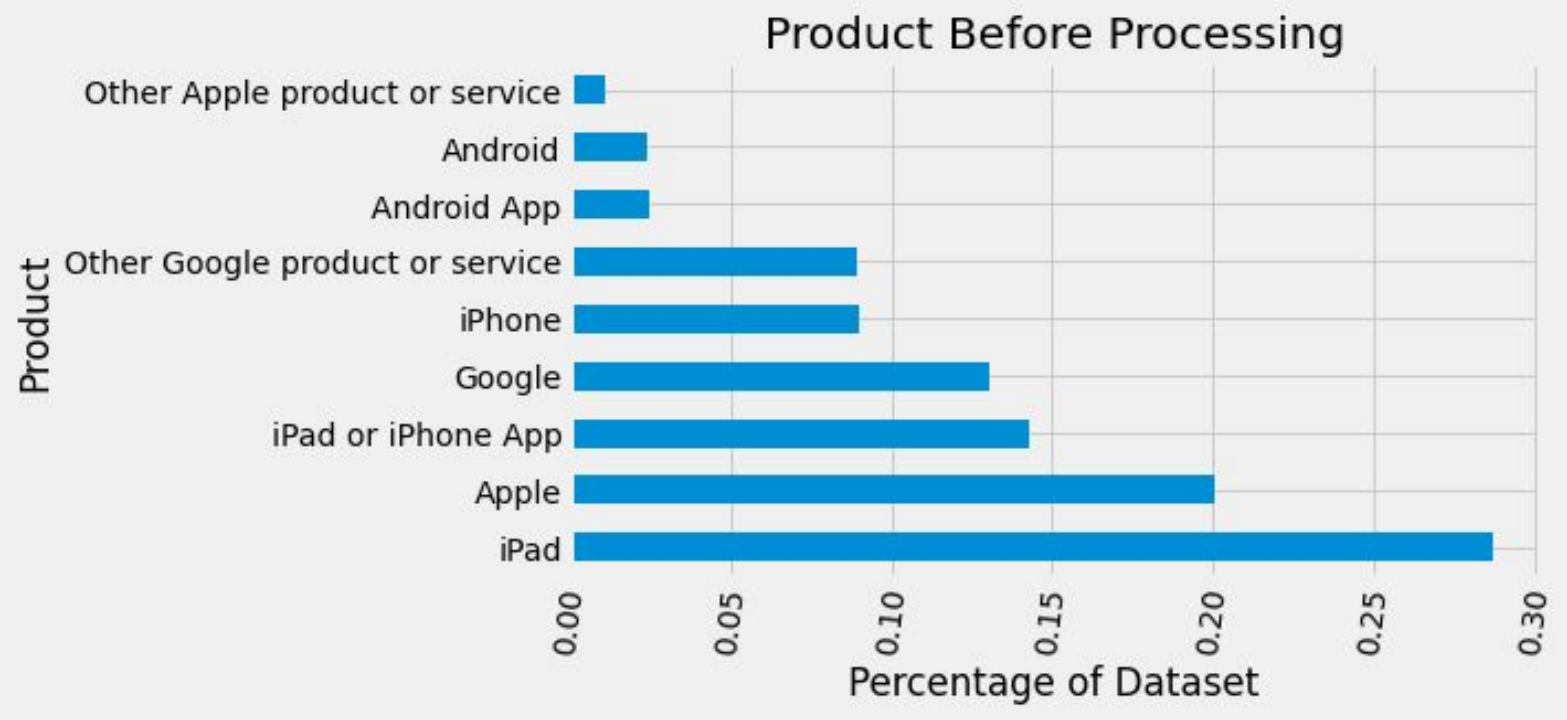


Sentiment - Binary Modeling

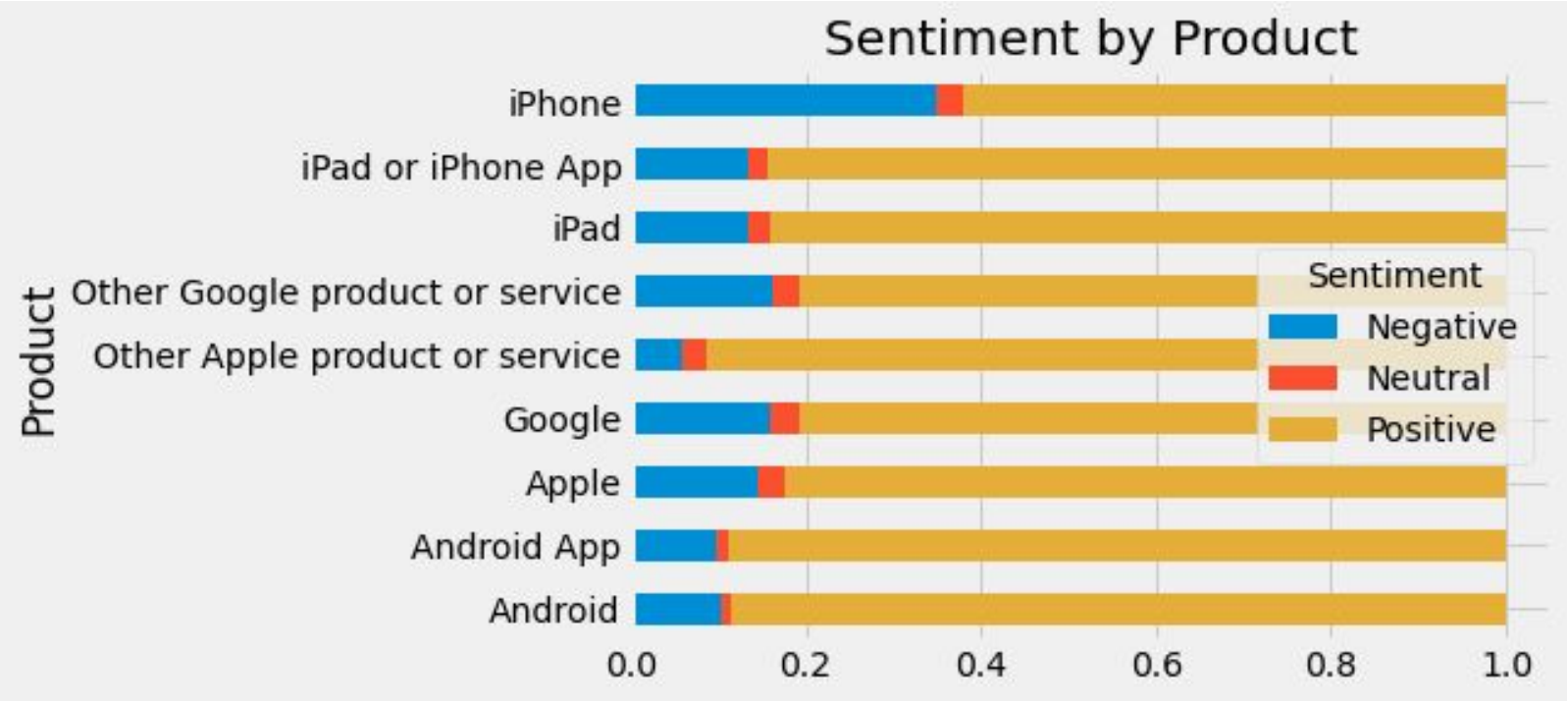




Product Breakdown

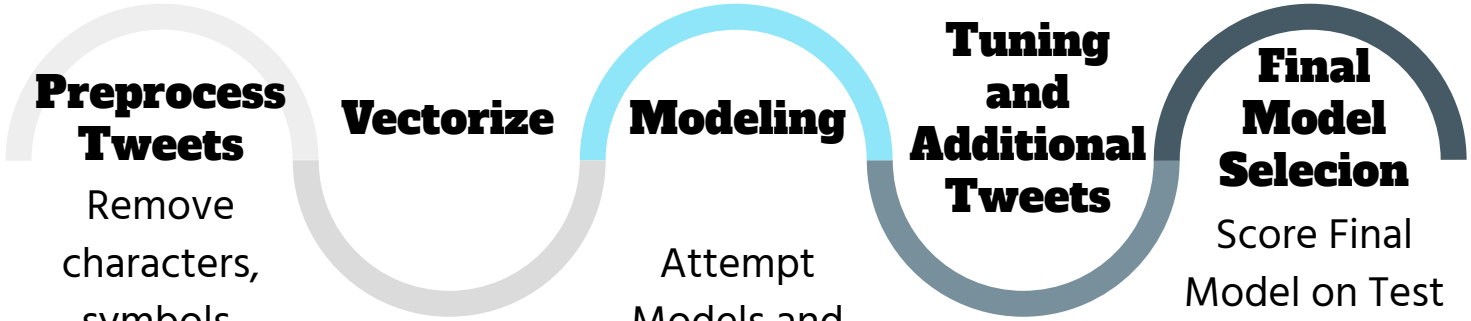


Product Breakdown



02 Project Metrics & Methods

Project Steps



Project Metrics

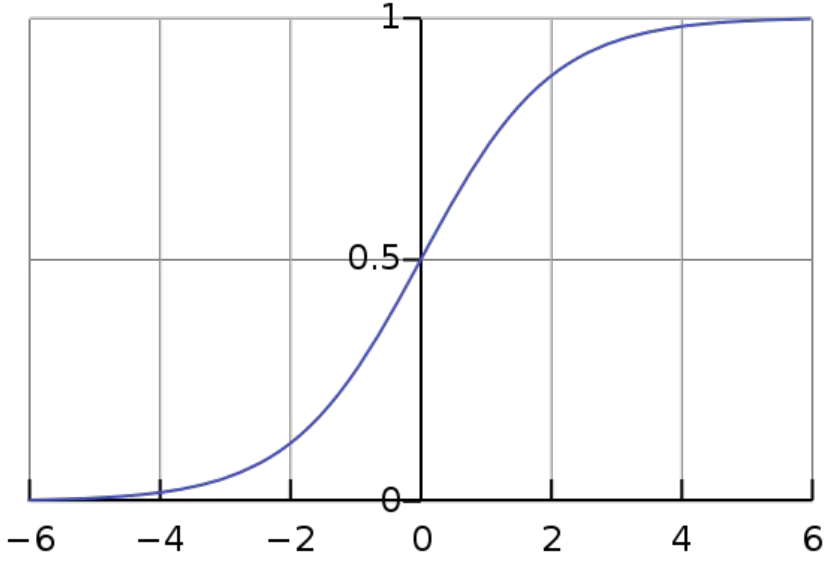
Recall - Casting the Net

- Recall – What percentage of Negative Tweets Were Captured
- Precision – What percentage of the time were the predicted actually Negative.
- Optimized with F-Beta 2 Scorer. Sought a Balance Between Recall and Precision.

FInal Model Selection

Tuned Logistic Regression Model

- Simple Model Nimble Model Using Bag-of-Words Vectors.
- Selected Models Using Cross-Fold-Validation and Grid Search
- Augmented the Training Data with an Additional 1219 Negative Tweets



03

Results & Interpretation

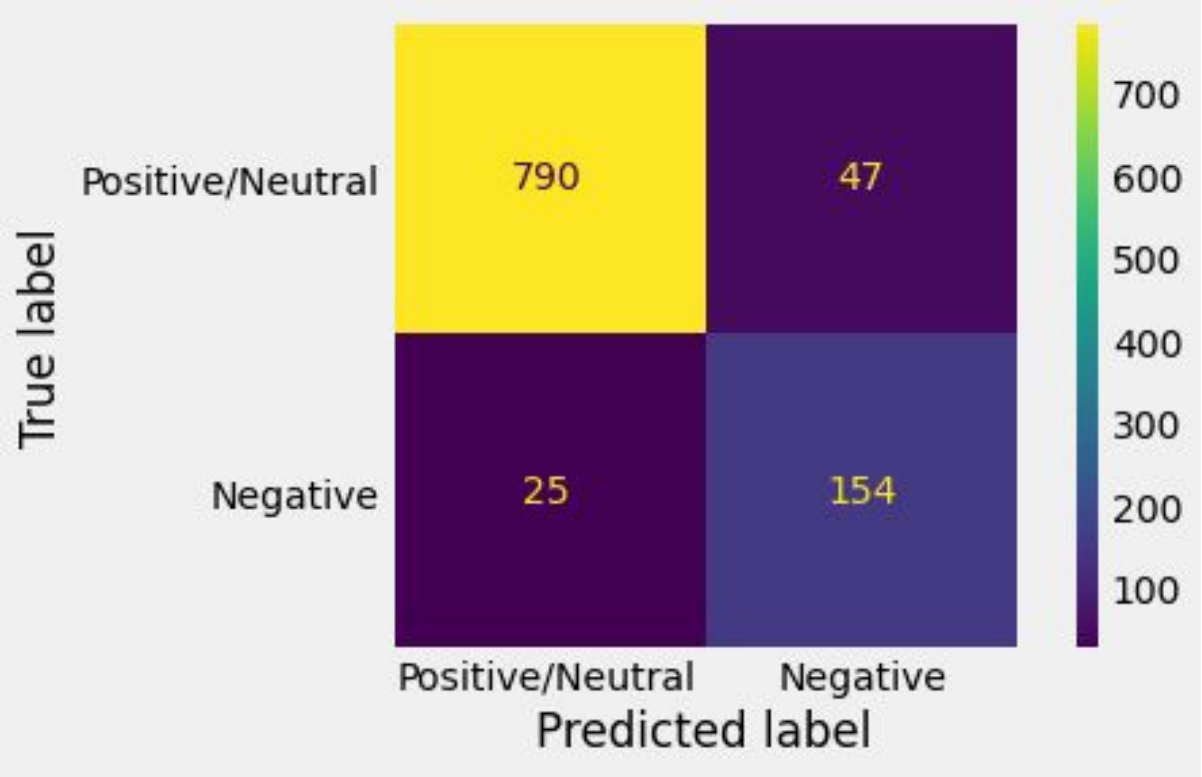
Final Model Results

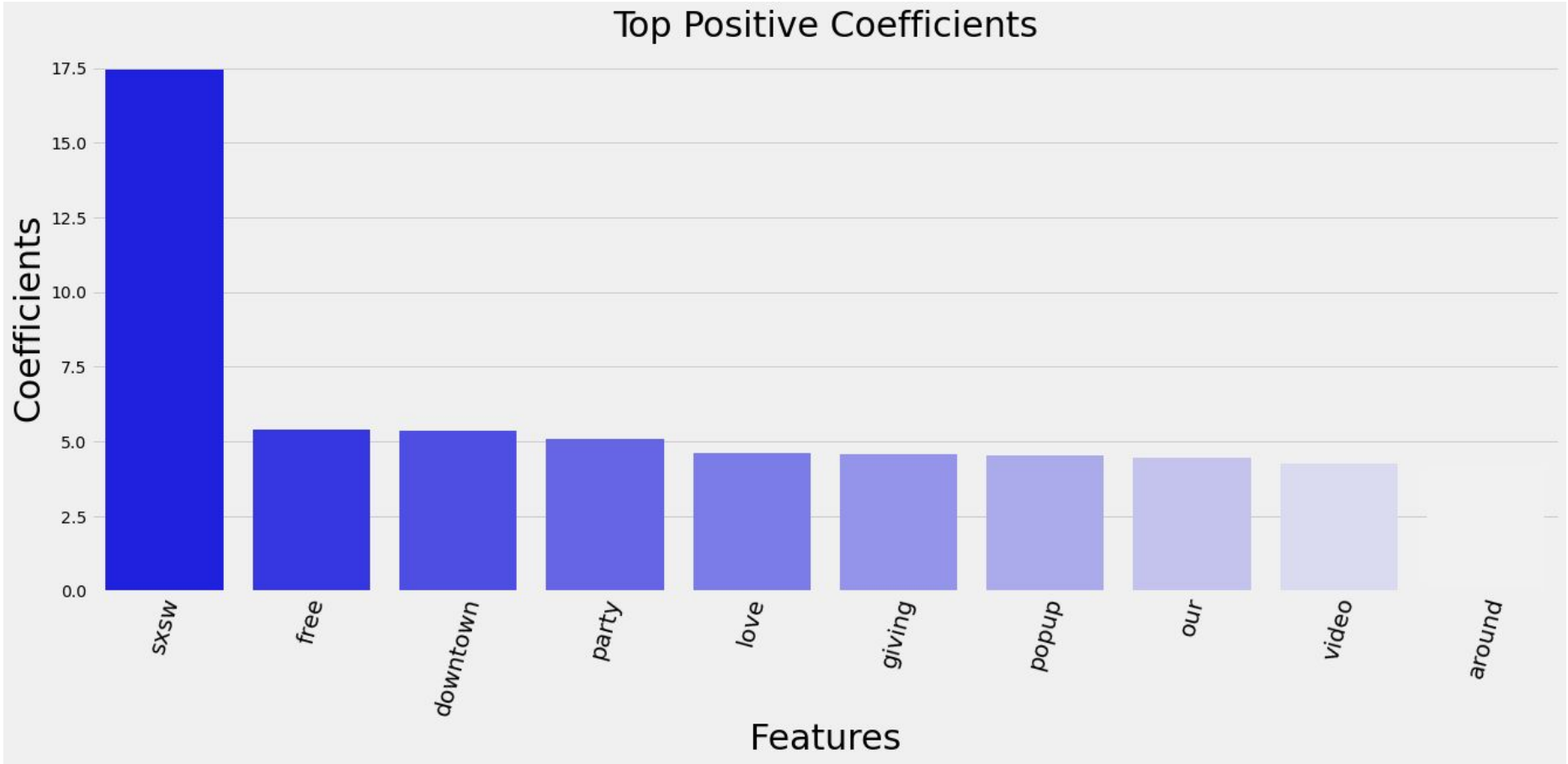
Recall Score - 86%

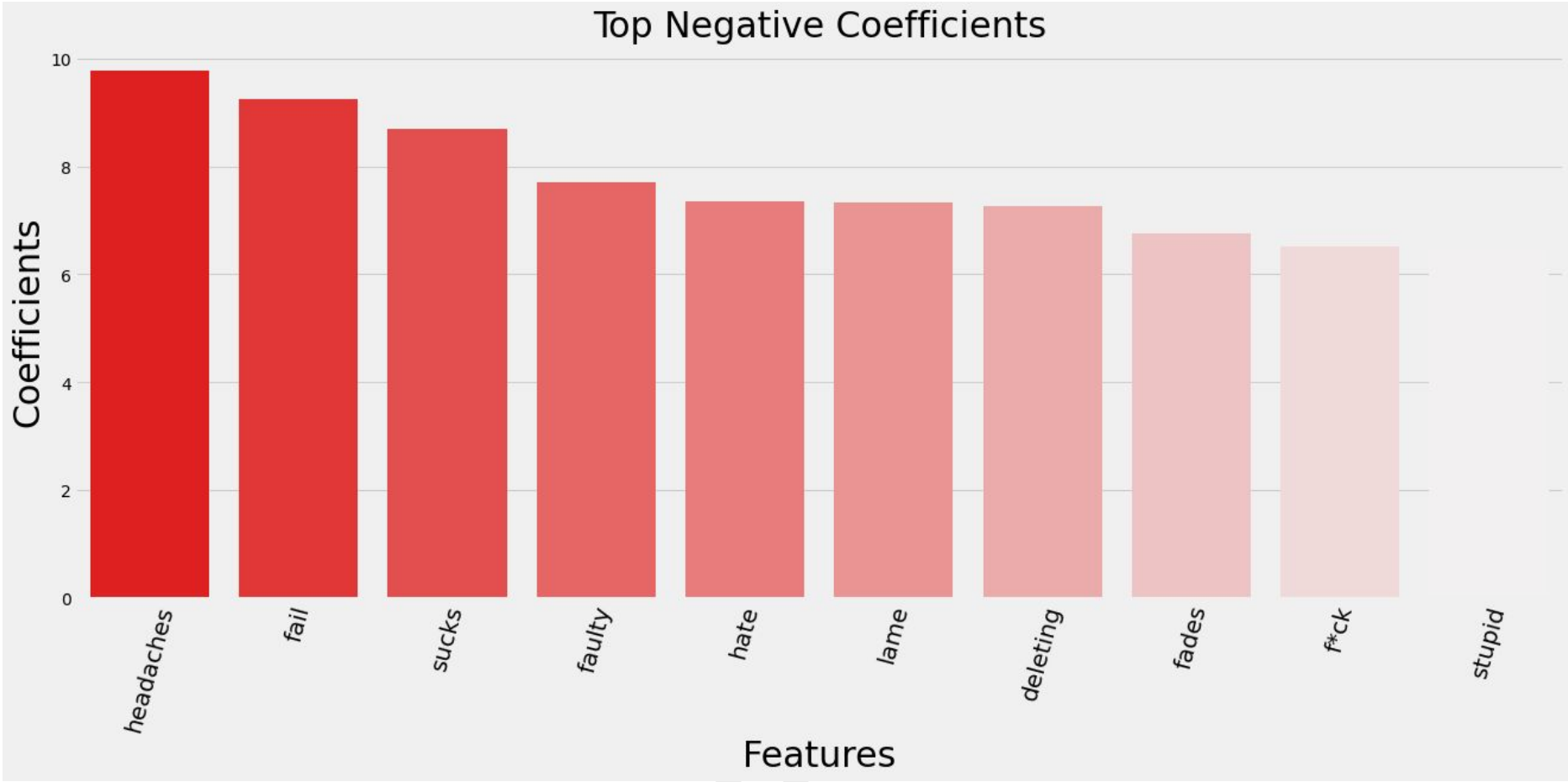
Precision - 77%

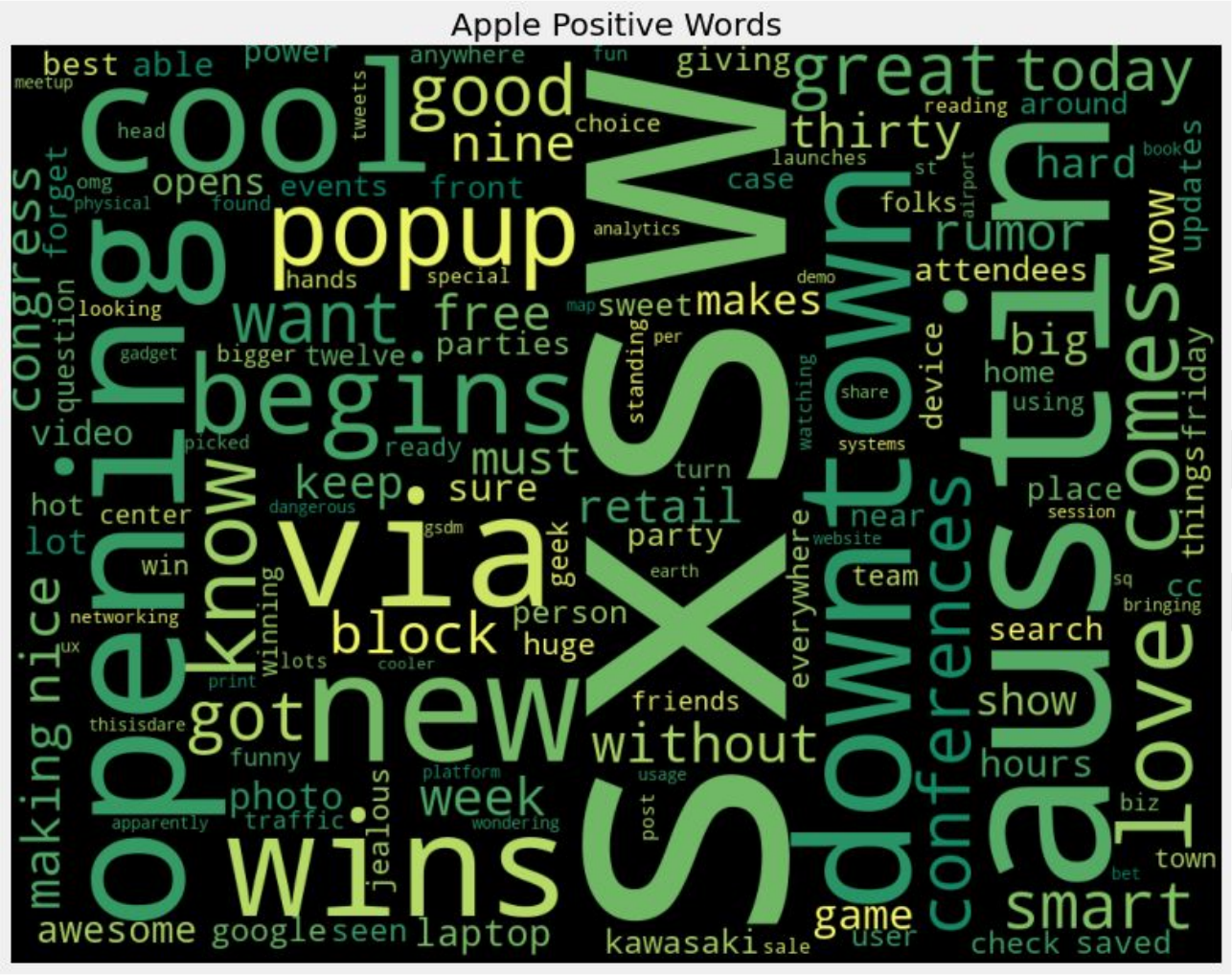
F1 Score - 81%

Accuracy - 93%

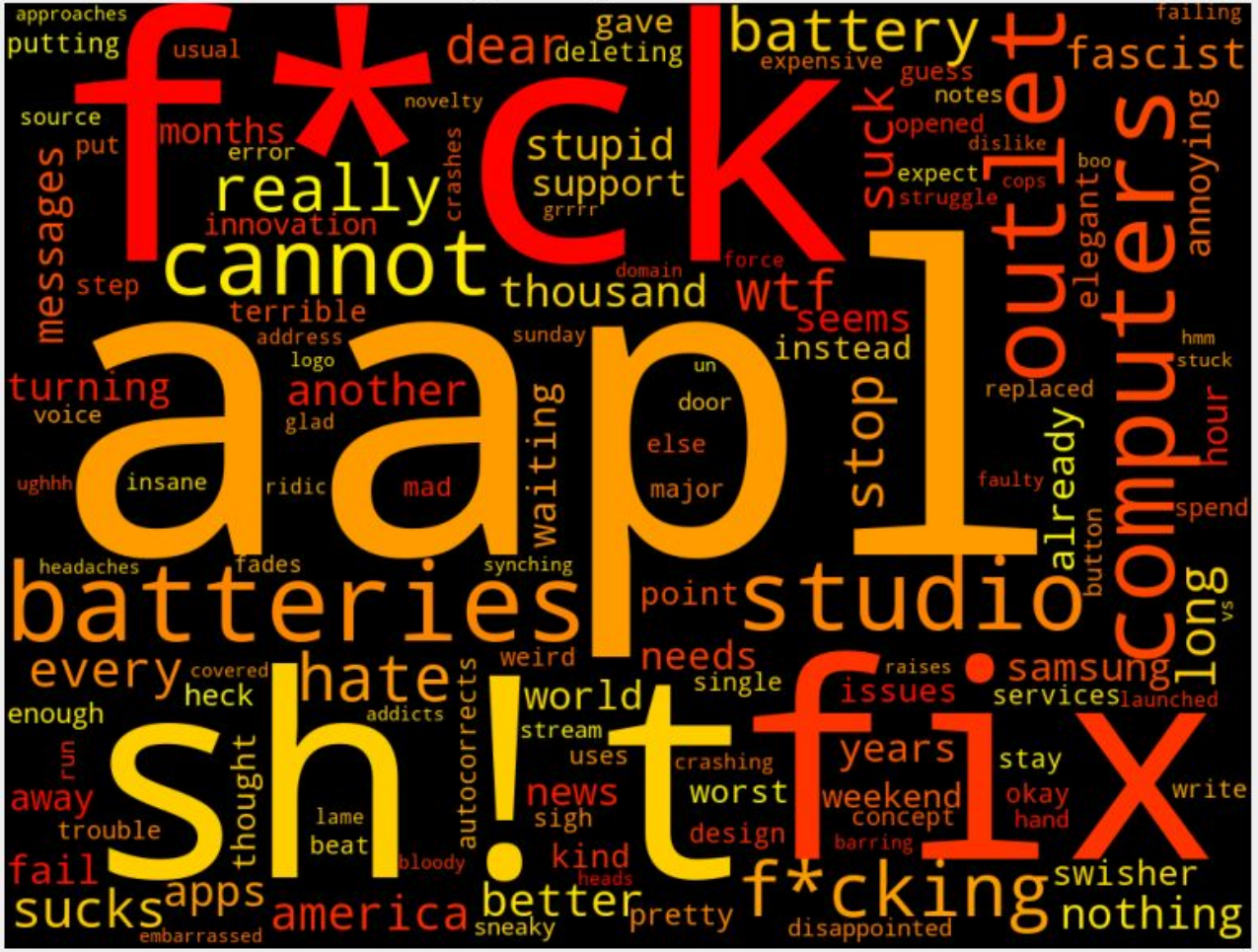



Top Positive Coefficients
Coefficients
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
sxsw
free
downtown
party
love
giving
popup
our
video
around
Features


Top Negative Coefficients
Coefficients
10
8
6
4
2
0
headaches
fail
sucks
faulty
hate
lame
deleting
fades
f*ck
stupid
Features


Apple Positive Words
sxsw
austin
downtown
opening
cool
wins
via
new
good
great
today
popup
begins
want
free
know
comes
love
conferences
smart
hard
rumor
thirty
nine
opens
events
front
case
folks
attendees
makes
sweet
parties
twelve
bigger
video
ready
must
keep
sure
retail
party
block
person
huge
without
friends
hours
show
search
week
photo
traffic
jealous
awesome
google
seen
laptop
kawasaki
game
check
saved
user
sale
town
biz
bet
post
usage
platform
wondering
apparently
funny
thisisdare
got
making
nice
hot
lot
win
networking
ux
lots
cooler
print
geek
gsdm
dangerous
center
near
place
session
website
earth
everywhere
team
sq
cc
bringing
things
friday
big
home
using
share
device
watching
systems
standing
per
map
demo
special
hands
looking
question
gadget
congress
forget
omg
physical
found
head
meetup
best
able
power
anywhere
fun
giving
tweets
choice
launches
st
airport
book
updates
wow
reading
around
analytics
picked
turn
lots
Winning

Apple Negative Words



04

Conclusion & Future Work

Conclusions

- Attack the battery issue – major source of complaints.
- Throw some parties – people like free stuff, it can turn the tide of sentiment.
- People want to like brands at SXSW – it's a succes!

Future Work

- Analyze the use of emojis and GIFS in tweets.
- Gather location-based data in real-time to respond to negative and positive feedback on-site.
- Experiment with Sentiment forecasting - could any of these reactions be predicted by Twitter chatter in advance of the event?