

Project 3: Web APIs and NLP

by Cheryl Leong

Outline

Introduction

Web Scraping and Data Cleaning

EDA

Hyperparameter Tuning and Modelling

Summary and Improvements

Introduction

r/iphone and r/Google Pixel

r/iphone

vs

r/GooglePixel



Posted by u/SuitingUncle620 **Moderator** 53 minutes ago

Do you think that iPhones or Smartphones in general have reached a point of stagnation?

I've been thinking about this for a while now. I genuinely can't remember the last time I felt genuinely hyped about a new iPhone being released/announced. I remember prior to 2019 being constantly in the loop with regards to recent leaks & rumours about the upcoming iPhone lineup. These days I'm lucky to even *watch* the event live.

I just don't feel the same hype that I did back during (in my mind) the peak of smartphone innovations - which was when the iPhone X got unveiled. I remember the buzz surrounding the iPhone X's announcement, literally everyone was talking about it - how huge of a redesign it really was. Yet, these days no one seems to talk about new releases nearly as much anymore.

I personally own an iPhone XS since launch. I've been **wanting** to find a reason to upgrade but I simply can't justify it. The upgrades year over year have been so minimal

10 Comments Share Save ...



Posted by u/JimmyTran081195 4 hours ago

Discussion Do you think status bar on iPhone 14 Pros will be changed following screen design?

I hope that would be true, like that:

Left side: time, GPS, wifi (maybe include VPN if necessary) and alarm.

Right side: signal intensity, 5G and the pre8/8 Plus version of battery percentage.

7 Comments Share Save ...



Posted by u/TheGamingGoogolplex 3 hours ago

Discussion How does MagSafe not interfere with the display?

I was sat at my MacBook, thinking about the new iOS/macOS feature in which you can use the iPhone as a webcam. I was thinking Apple is potentially going to introduce a MagSafe connection to the back of newer MacBooks (whatever ones ship with Ventura) since that would make a hell of a lot of sense - at the moment they're asking people to buy a fairly unattractive clip. I then proceeded to think - no, they wouldn't do that because magnets would mess with the display. But.... It's in the iPhone? how do they avoid the problem of Magnetic interference with the display? is that a thing of the past? I



Posted by u/Chadical5000 23 hours ago

341 Moving from Pixel to iPhone - Not going well

I've been using Google Nexus/Pixel devices since the Nexus S. My most recent phone has been the Pixel 4a which I've loved and is the perfect size, but it's starting to get old. Unfortunately, Google doesn't seem to want to make smaller phones anymore. And once you've been used to constant updates from Google for the past 10+ years, moving over to a non-Pixel device with a custom skin (even if minimal), just seems like a bad move.

So I tried out a used iPhone 12 mini. I love the hardware and the small size of the device is great. CarPlay is absolutely superior to Android Auto. BUT, and this is a huge but, IOS feels soo limited. I can't believe how something so simple like customizing your home screen on an Android device is so ridiculously complicated and hacky to get the same look on IOS. Just try to put a digital clock on your homescreen with a transparent background on IOS and you will see what I mean. Also I'm not convinced that IOS looks

265 Comments Share Save ...



Posted by u/angrycatmeowmeow **Pixel 6 Pro** 10 hours ago

28 How are you liking your 6a?

A while back I bought my mom a 4XL for \$350. With the inflated trade in values, I figured I could upgrade her for pretty cheap to a 6a. The package is already looking pretty lost in the UPS system, still showing a delivery between 12-3PM today when it's almost 10PM here and it hasn't left Illinois. FedEx lost both my P6Ps when I bought them in November, so not getting the warm and fuzzies here. Especially since the lost P6Ps took weeks to resolve with Google support. Honestly not sure why I subjected myself to this again. Apparently I'm a masochist.

Anyway, just in case I do get the phone, how's yours working out and was your trade in smooth?

27 Comments Share Save ...

Web Scraping and Data Cleaning

r/iphone and r/Google Pixel

Web Scraping

Web Scraping was done using Pushshift API:

<https://api.pushshift.io/reddit/search/submission>

Function was created to scrape 10,000 rows of data from each subreddit and stored in separate .csv files

However, upon exploring the datasets, there were many columns and rows that contained duplicate and redundant information and they were dropped, resulting in about 5,500 rows of useful data remaining for each subreddit

Data Cleaning

Additional functions were created to accomplish the following using regex:

- Lowercase all letters
- Remove numerical values, punctuations, non-ascii characters, words that contained 2 letters or less

Before:

	title	selftext	subreddit
0	Will the Pixel 5 be able to safely last anothe...	I change phones every 2-3 years and the Pixel ...	GooglePixel

After:

	title	selftext	subreddit
0	will the pixel able safely last another year...	change phones every years and the pixel curr...	GooglePixel

Data Processing

Both r/iphone and r/GooglePixel datasets were merged into one DataFrame and they were:

- Tokenized
- Lemmatized
- Introduced additional Stop Words that were common in both subreddits

```
add_words = ['phone', 'mobile', 'handphone', 'pro', 'would', 'like', 'get', 'app', 'screen']
```

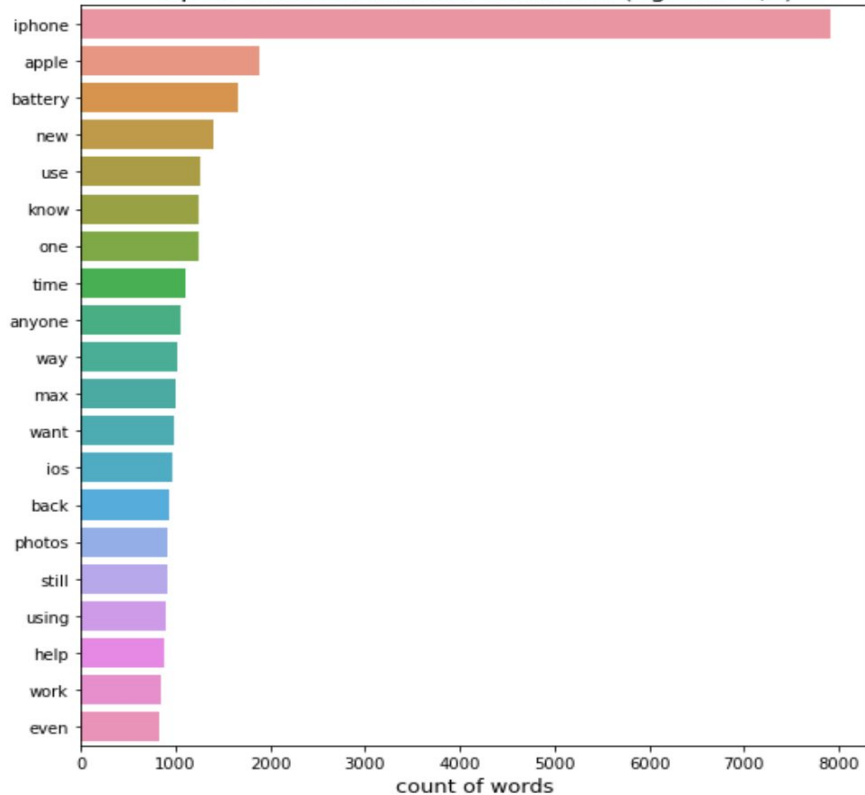
- Mapped subreddits to: iphone: 0 (50.7%) and GooglePixel: 1 (49.3%)

EDA

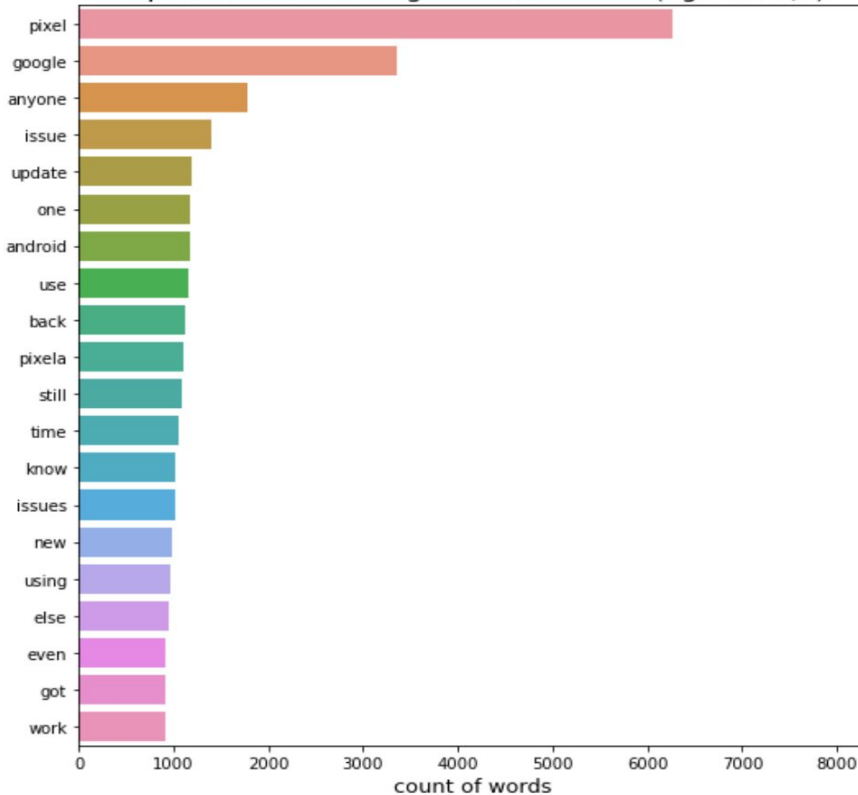
r/iphone and r/Google Pixel

Top 20 words (title and selftext combined)

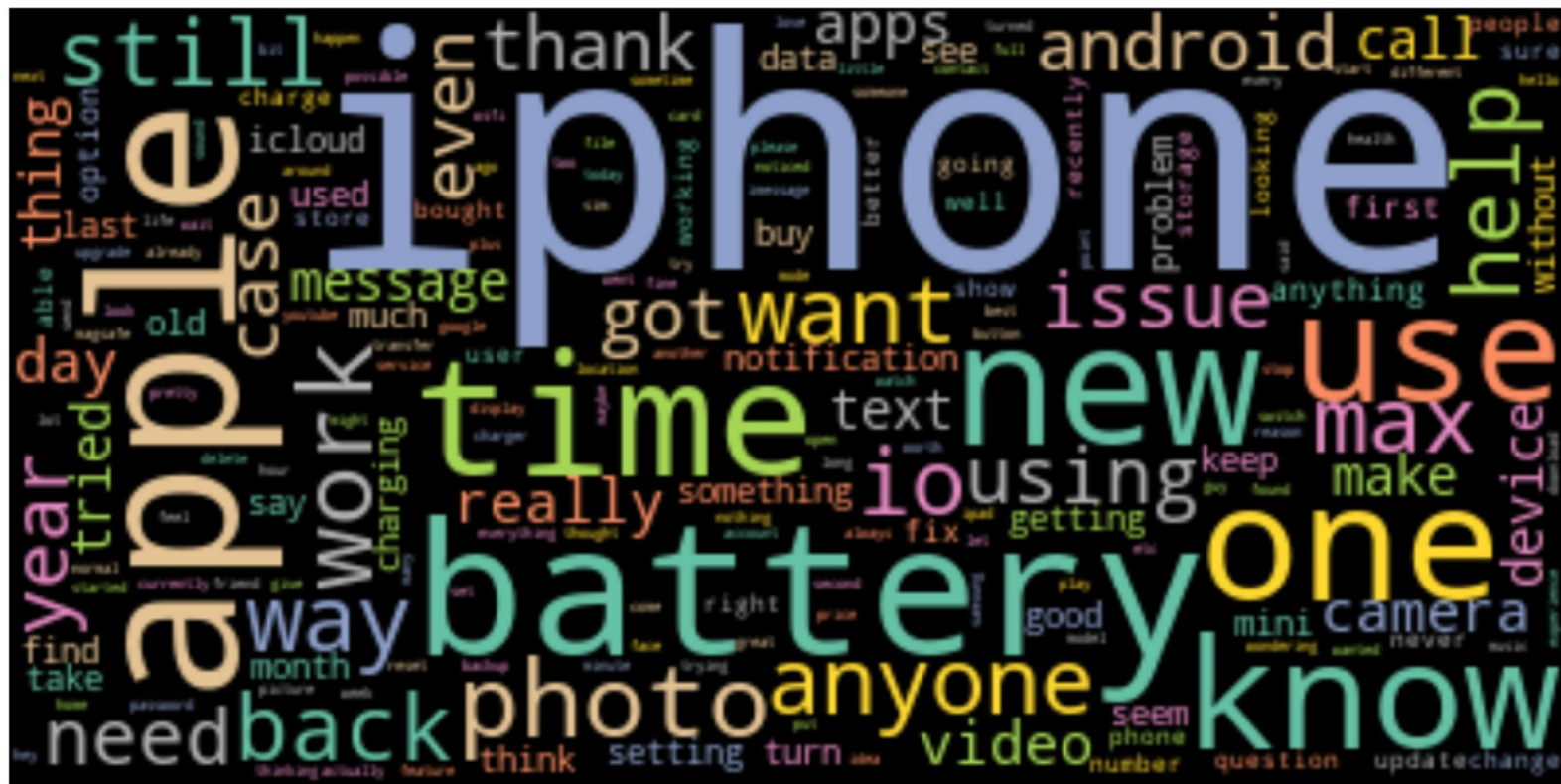
Top 20 words in r/iPhone combined (ngrams-1,1)



Top 20 words in r/GooglePixel combined (ngrams-1,1)



Word Cloud (r/iphone)



Word Cloud (r/GooglePixel)



Hyperparameter Tuning and Modelling

r/iphone and r/Google Pixel

Models

Multinomial Naive Bayes Classifier

Multinomial Naive Bayes with Count Vectorizer

Multinomial Naive Bayes and TF-IDF Vectorizer

Support Vector Machine Classifier

Support Vector Machines and CountVectorizer

Support Vector Machines with TF-IDF Vectorizer

Random Forest Classifier

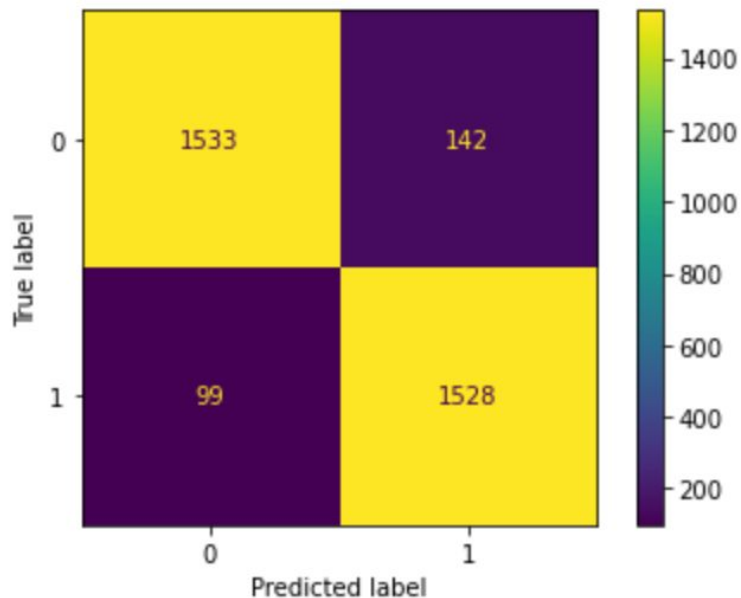
Random Forest and CountVectorizer

Random Forest and TF-IDF Vectorizer

Multinomial Naive Bayes Classifier

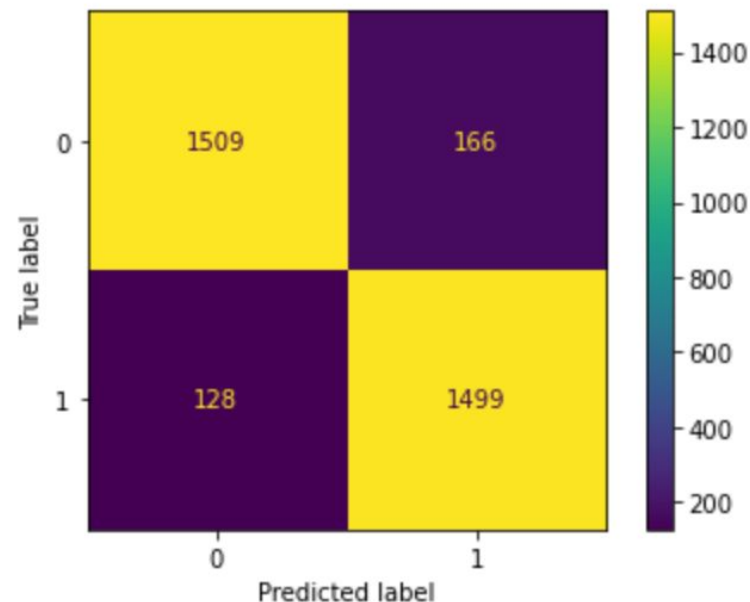
CountVectorizer with Multinomial Naive Bayes Model

Accuracy score of training set: 94.53%
Accuracy score of testing set: 92.7%



Multinomial Naive Bayes Model with TF-IDF Vectorizer

Accuracy score of training set: 95.47%
Accuracy score of testing set: 91.1%



Multinomial Naive Bayes Classifier Hyperparameter Tuning

Multinomial Naive Bayes with Count Vectorizer

```
{'cvec__max_df': 0.9,  
 'cvec__max_features': 5000,  
 'cvec__min_df': 8,  
 'cvec__ngram_range': (1, 2)}
```

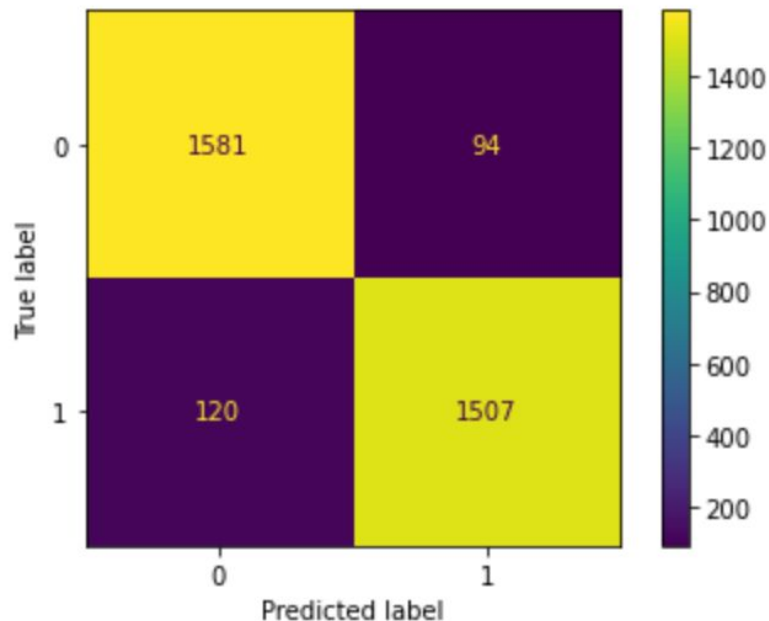
Multinomial Naive Bayes and TF-IDF Vectorizer

```
{'tvec__max_df': 0.9,  
 'tvec__max_features': 10000,  
 'tvec__min_df': 2,  
 'tvec__ngram_range': (1, 2)}
```


Support Vector Machines Classifier

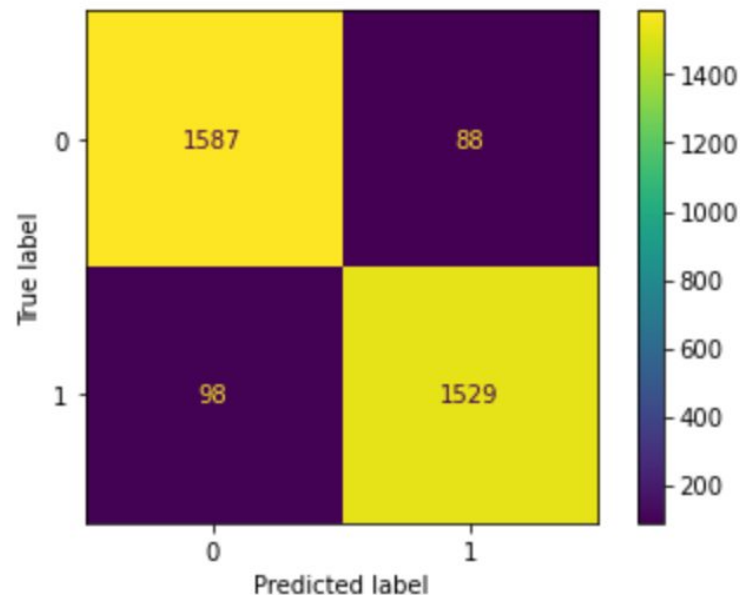
Support Vector Machine Model with CountVectorizer

Accuracy score of training set: 98.36%
Accuracy score of testing set: 93.52%



Support Vector Machine Model with TF-IDF Vectorizer

Accuracy score of training set: 100.0%
Accuracy score of testing set: 94.37%



Support Vector Machines Classifier Hyperparameter Tuning

Support Vector Machines and CountVectorizer

`{'C': 0.05, 'degree': 2, 'kernel': 'linear'}`

Support Vector Machines with TF-IDF Vectorizer

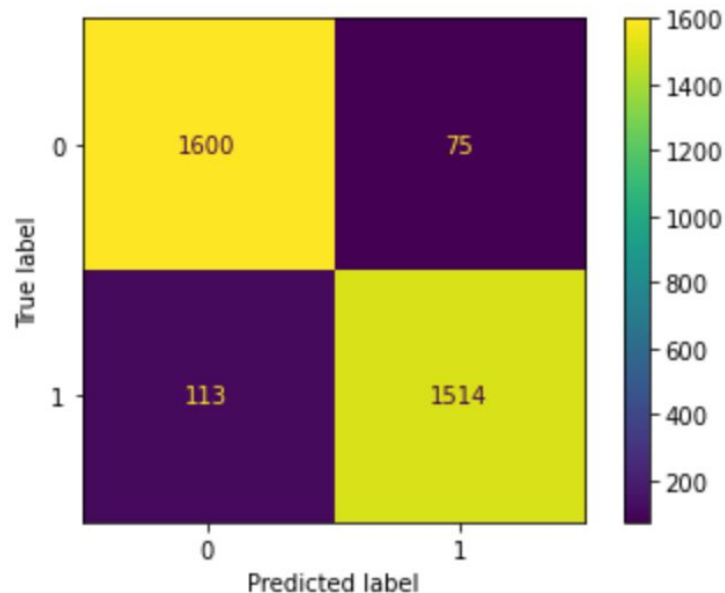
`{'C': 10, 'degree': 2, 'kernel': 'rbf'}`

Random Forest Classifier

Random Forest Model with CountVectorizer

Accuracy score of training set: 96.18%

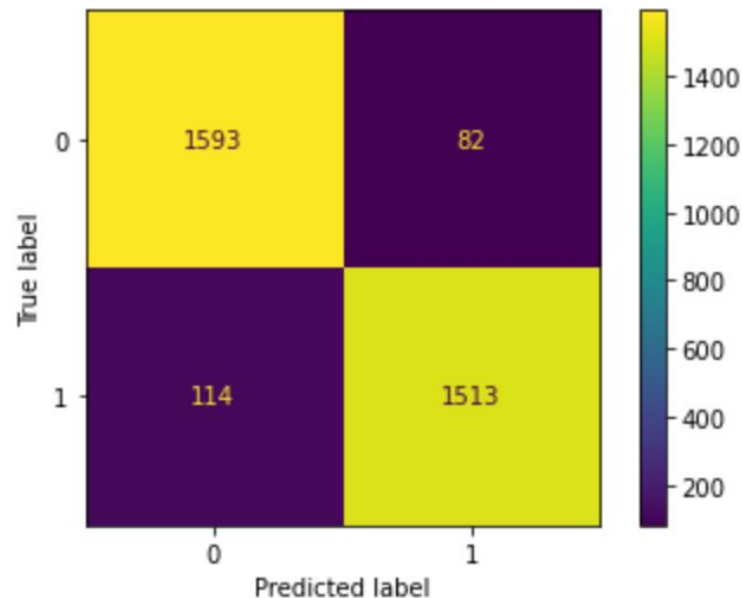
Accuracy score of testing set: 94.31%



Random Forest Model with TF-IDF Vectorizer

Accuracy score of training set: 97.07%

Accuracy score of testing set: 94.06%



Random Forest Classifier Hyperparameter Tuning

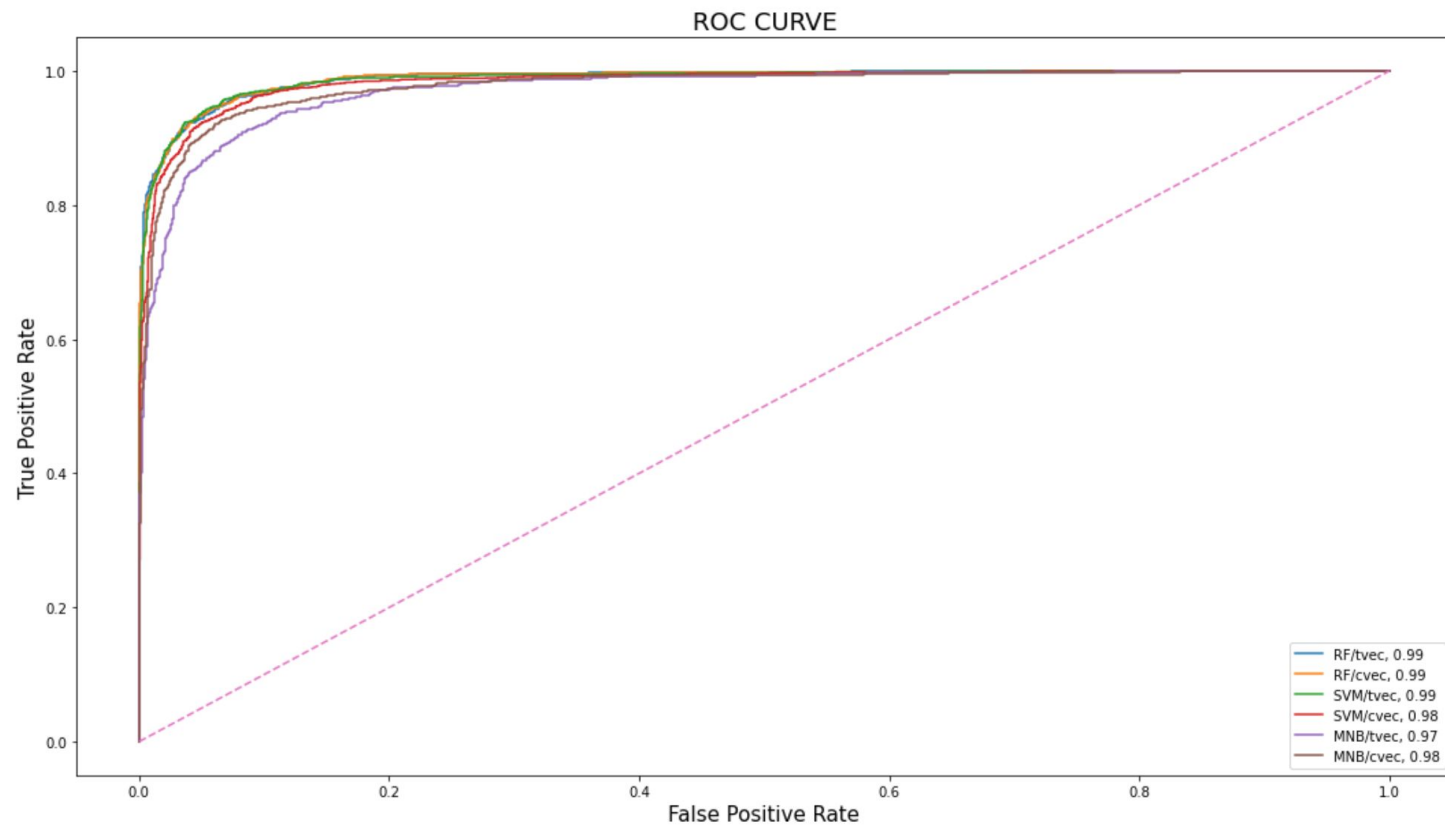
Random Forest and CountVectorizer

```
{'max_depth': None,  
 'min_samples_leaf': 5,  
 'min_samples_split': 20,  
 'n_estimators': 800}
```

Random Forest and TF-IDF Vectorizer

```
{'max_depth': None,  
 'min_samples_leaf': 5,  
 'min_samples_split': 30,  
 'n_estimators': 800}
```

ROC Curve




Summary and Improvements

r/iphone and r/Google Pixel

Summary


Estimator	Vectorizer	Train Accuracy	Test Accuracy	Generalization	Sensitivity	Specificity	AUC
Multinomial Naive Bayes	Count	94.53%	92.7%	1.83%	0.94	0.91	0.98
Multinomial Naive Bayes	Tfidf	95.47%	91.1%	4.37%	0.92	0.9	0.97
Support Vector Machine	Count	98.36%	93.52%	4.84%	0.93	0.94	0.98
Support Vector Machine	Tfidf	100%	94.37%	5.63%	0.94	0.95	0.99
Random Forest	Count	96.18%	94.31%	1.87%	0.93	0.96	0.99
Random Forest	Tfidf	97.07%	94.06%	3.01%	0.93	0.95	0.99

Misclassified rows: r/iphone vs r/GooglePixel


 r/iphone · Posted by u/V3N0M3 1 day ago

Discussion iPhone SE 2022: A phone that has something no other iPhone has but is a phone hated by reviewers.


97 upvotes · 109 comments · 0 awards

 r/iphone · Posted by u/anjibotan 9 days ago

iPhone SE versus Pixel 6A: what Google got right and Apple got wrong
theverge.com/232886...




194 upvotes · 188 comments · 0 awards

 r/iphone · Posted by u/Oceansun_2004 15 days ago


Fold 3 to i phone

92 upvotes · 57 comments · 0 awards

 r/iphone · Posted by u/Geminiun 19 days ago


Discussion I switched to an iPhone 13 for the first time after using an Android for the last 12 years, these are my thoughts after 72 hours!

1.2k upvotes · 370 comments · 1 award

 r/iphone · Posted by u/thanirs 21 days ago


Discussion Things you would miss if you came from a Pixel

39 upvotes · 77 comments · 0 awards

 r/iphone · Posted by u/asexualloverboy 26 days ago


Discussion Iphone is just too good a phone when I match android to it

421 upvotes · 296 comments · 1 award

 r/iphone · Posted by u/MichaelRahmani 26 days ago


Discussion iphone glass doesn't feel as smooth as my pixel

0 upvotes · 18 comments · 0 awards

 r/GooglePixel · Posted by u/ArizonaRenegade 3 hours ago


General Trade-in questions. How strict is Google, when it comes to their trade-in standards? Will they accept an iCloud locked iPhone?

1 upvote · 2 comments · 0 awards

 r/GooglePixel · Posted by u/Affectionate_Neck355 17 hours ago


Pixel 6 Pro Just got Google Pixel 6 Pro. Are there solutions to the issues I'm having?

0 upvotes · 10 comments · 0 awards

 r/GooglePixel · Posted by u/Coconuttery Pixel 5 19 hours ago


A Google tracker product (like Tile/Airtag)

0 upvotes · 8 comments · 0 awards

 r/GooglePixel · Posted by u/GodlyArchitect 21 hours ago


Pixel 5 Bumper Case for Pixel 5?

0 upvotes · 11 comments · 0 awards

 r/GooglePixel · Posted by u/ThePazzoDrago 23 hours ago

Managing texts on Pixel and iPhone

1 upvote · 8 comments · 0 awards

 r/GooglePixel · Posted by u/Chadical5000 23 hours ago

Moving from Pixel to iPhone - Not going well

341 upvotes · 266 comments · 0 awards

Improvements

Increase the number of datasets used

Try using different Stop Words

Use algorithms such as Bagging Classifier

Explore Random Forest with CountVectorizer further by introducing more trees and depth

The End

r/iphone and r/Google Pixel