Binary Classification Model

—

Apple vs Android

Presented by: Ziyang

As an investment analyst…

- Responsible to monitor Apple-related investment
- Company receives live Reddit post about Technology
- However, the system cannot identify where the post comes from in details
- It tends to group Apple and Android together
- Hinder monitor process

- ***Problem Statement***

  How can we effectively distinguish posts about Apple amongst data that consists of post from both Apple and Android?

Web Scrap

Exploratory Data Analysis

Pre-processing & Modeling

Challenge the model

Drawback & Conclusion

Web Scrap

Exploratory Data Analysis

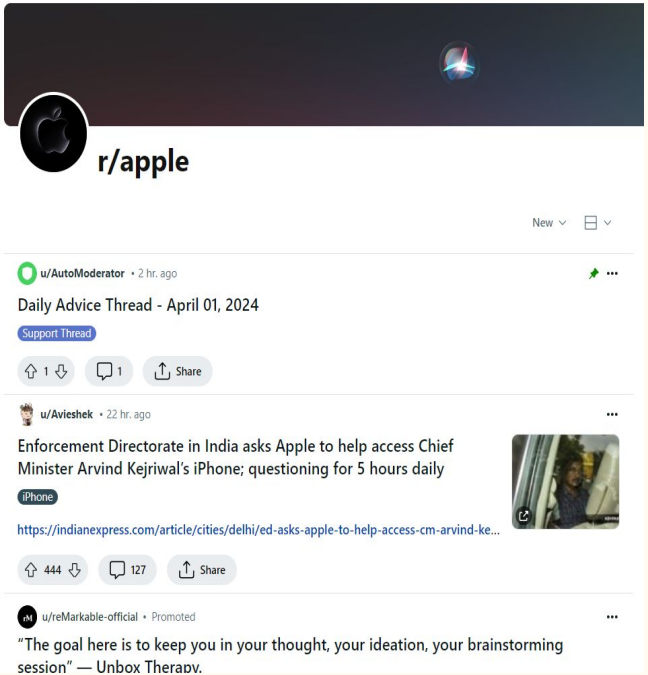
Pre-processing & Modeling

Challenge the model

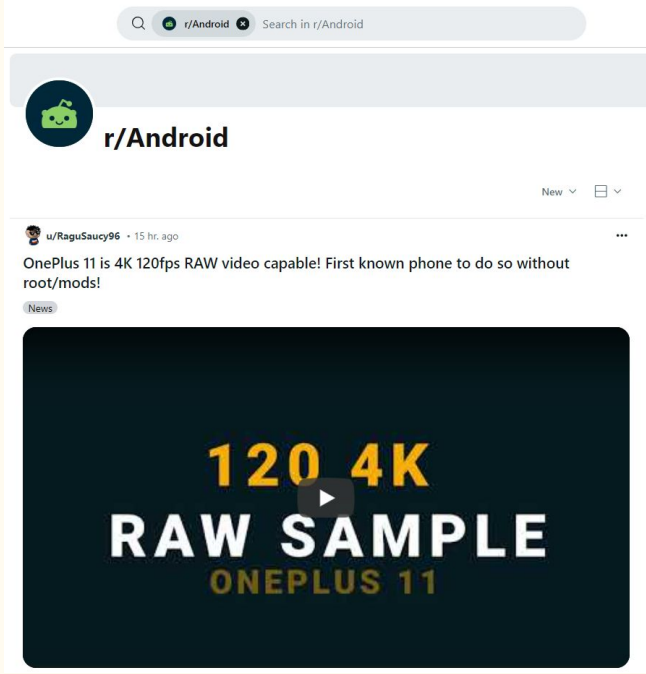
Drawback & Conclusion

Web Scrape

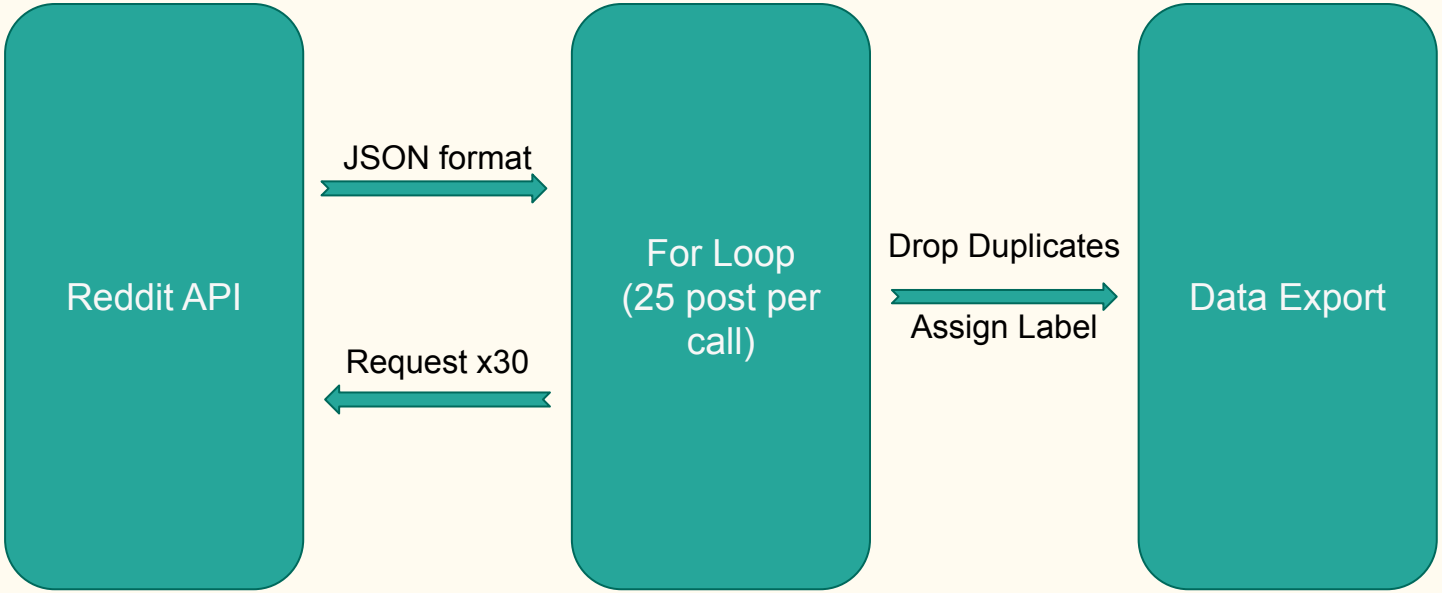
https://www.reddit.com/r/apple/new



https://www.reddit.com/r/Android/new



Web Scrape



*Only title of the post is scraped as they mostly summarise the content of the post, with less words count

Web Scrap

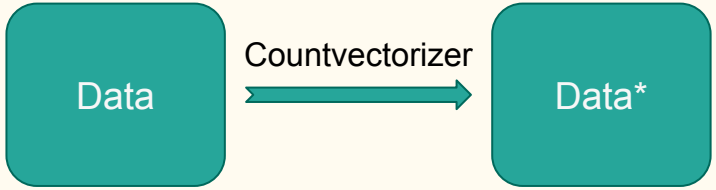
Exploratory Data Analysis

Pre-processing & Modeling

Challenge the model

Drawback & Conclusion

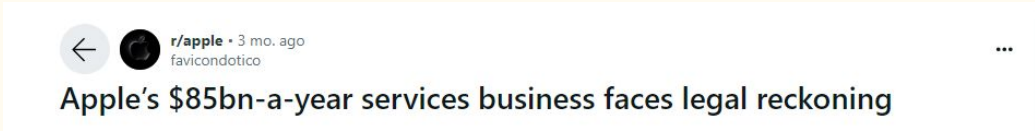
Exploratory Data Analysis




```
Index(['80', '8300', '85bn', '87', '8a', '8g3', '8gb', '8k', '8s', '90',
       '9000', '9300', '9400', '99', '9pro', '9to5google', '9to5mac', 'a17',
       'a25', 'a35', 'a55', 'a9', 'aaa', 'aawireless', 'ability', 'able',
       'absolutely', 'abused', 'abusive', 'accelerated'],
      dtype='object')
```

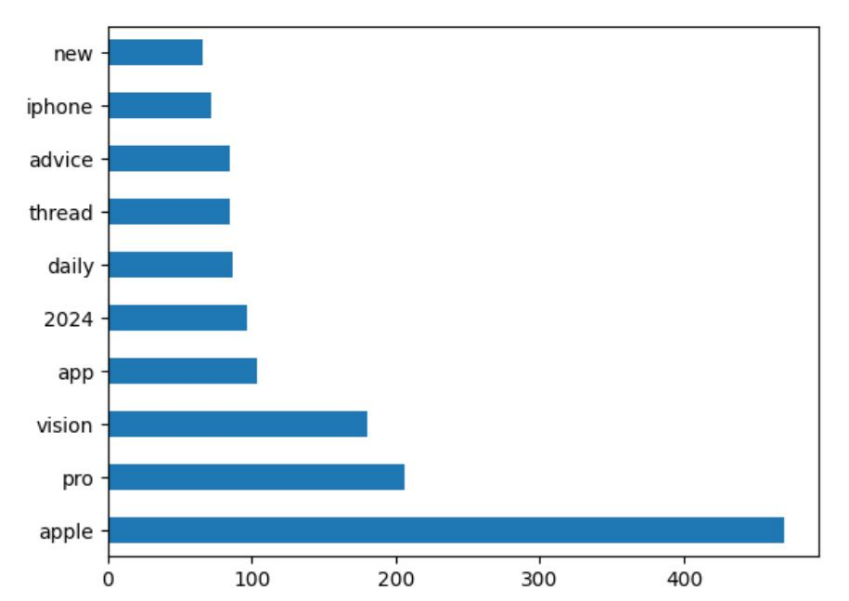
Countvectorizer:

- Convert a collection of text documents to a matrix of token counts
- English stop words are removed

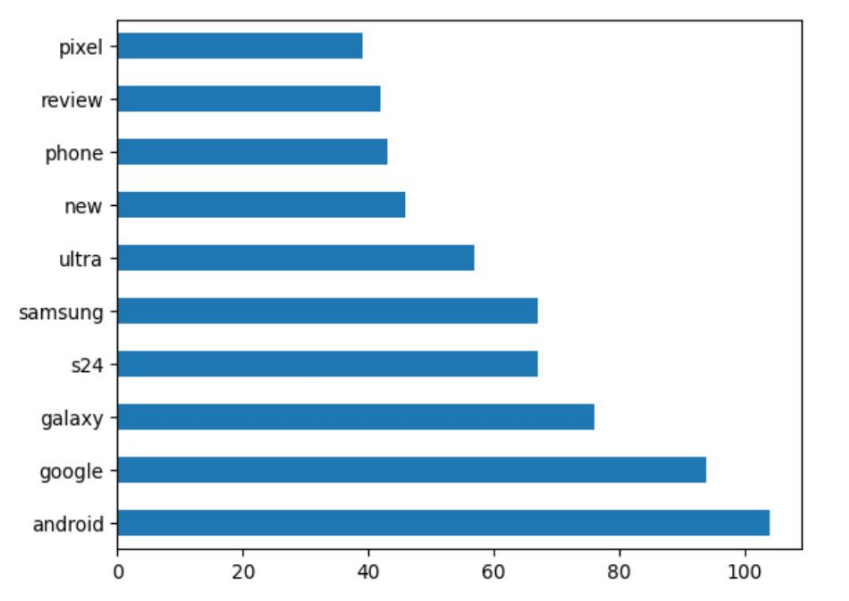


Exploratory Data Analysis

Apple

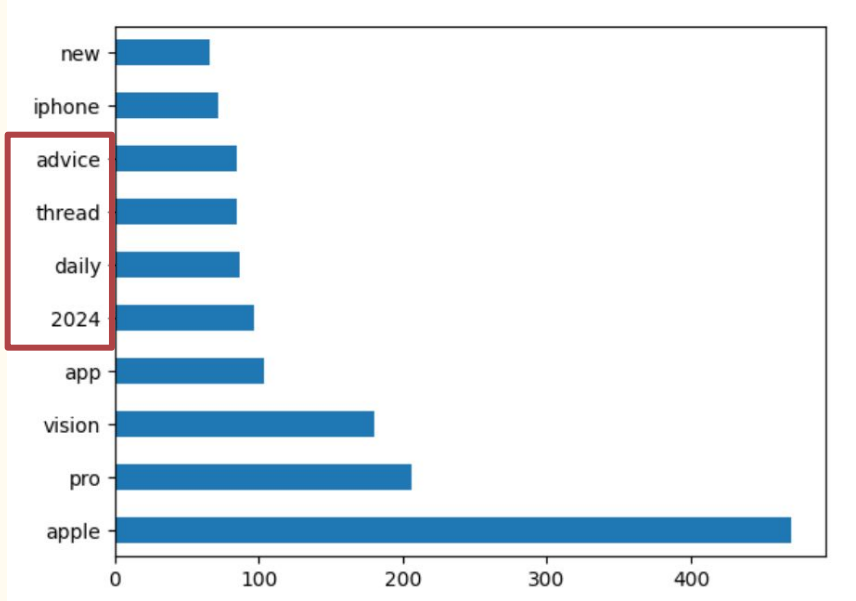


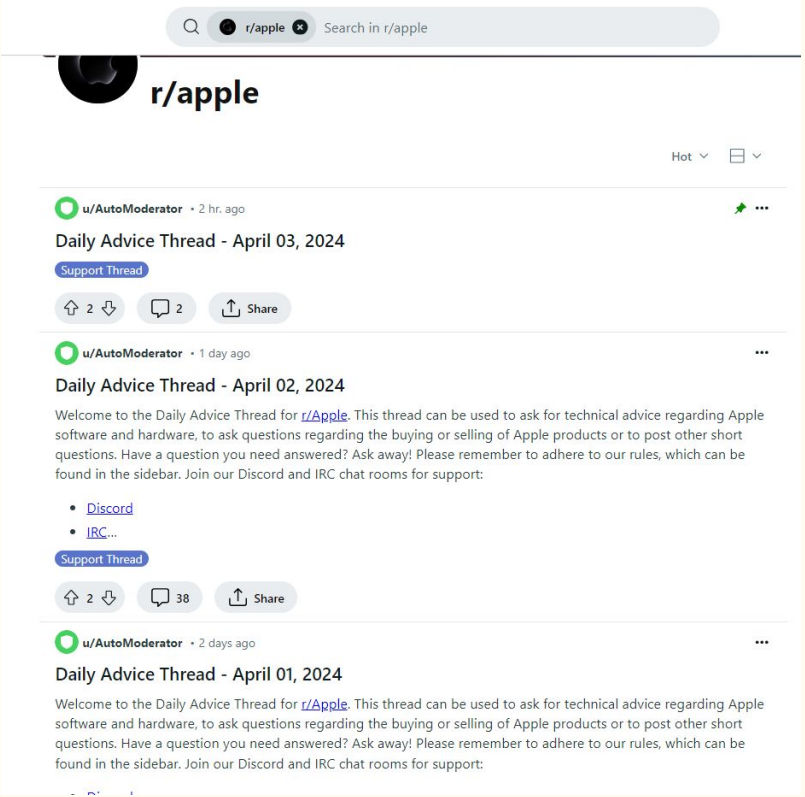
Android



Exploratory Data Analysis

Apple





Web Scrap

Exploratory Data Analysis

Pre-processing & Modeling

Challenge the model

Drawback & Conclusion

Pre-processing & Modeling

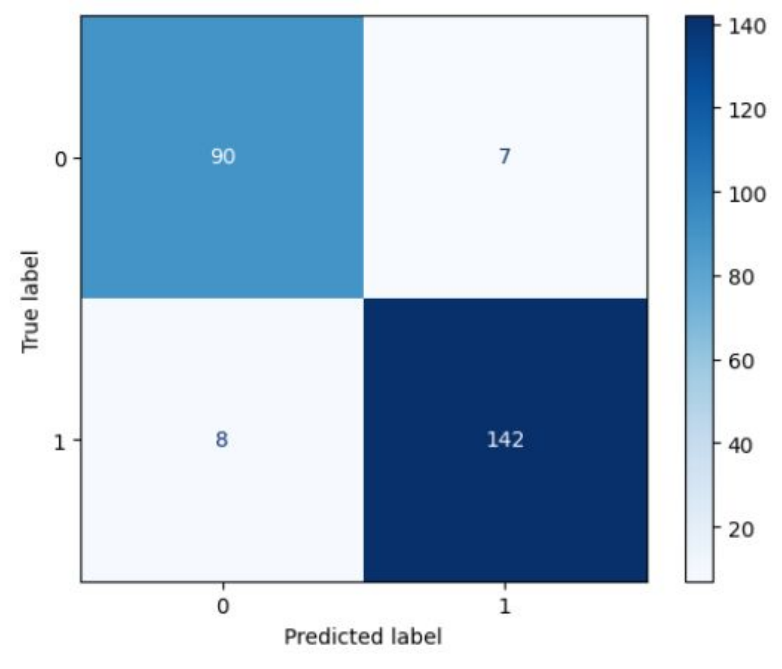
*Countvectorizer

GridSearch 5-fold Cross Validation:

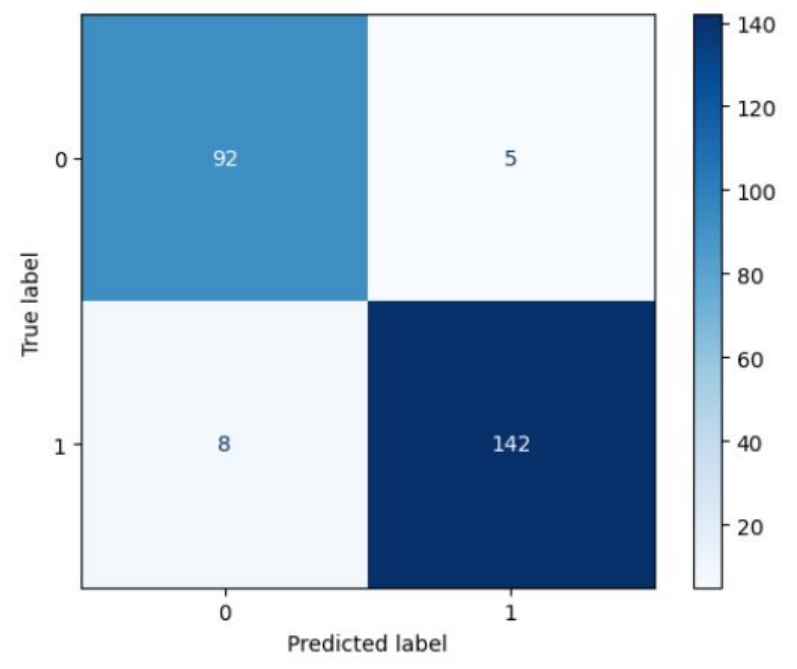
- On Countvectorizer parameters
- On Logistic Regression hyperparameters

Pre-processing & Modeling

Naive Bayes



Logistic Regression



Pre-processing & Modeling

Baseline model accuracy 50%, just random chance!

Naive Bayes:

- Recall: 95%
- Precision: 95%
- Accuracy: ***93.92%***

Logistic Regression:

- Recall: 95%
- Precision: 96%
- Accuracy: ***94.70%***

Web Scrap

Exploratory Data Analysis

Pre-processing & Modeling

Challenge the model

Drawback & Conclusion

r/honeymoonplanning

Create a post | Join

New

u/ehtsstarrw10 • 9 hr. ago

Caribbean Vegan

Has anyone been to the Caribbean islands as a vegan? Me and my soon to be wife want to go somewhere in the Caribbean and be able to eat as we are fully vegan. Any help or recommendations would be amazing !

2 | 2 | Share

u/SpartyOn32 • 13 hr. ago

Late July 2024 Honeymoon

Hi everyone! Unfortunately we just found out that our Tahitian cruise was cancelled. We thought we had everything planned and finalized, but now we're back to the drawing board. Our budget is about $20k and we're open to a land trip in Tahiti, a replacement cruise on Paul Gauguin (although this is much pricier than our original Windstar cruise), or a pivot to Hawaii? Any thoughts or suggestions? Thanks!

0 | 1 | Share

u/IBKR_Official • Promoted

What interest rate does your broker pay? Interactive Brokers' clients earn up to 4.83% on uninvested USD cash balances.

place for planning honeymoons

While wedding planning is fun, it's also fun to talk about and make recommendations for the real fun which comes after the wedding. The honeymoon! All couples welcome.

9.1K	5	Top 7%
Members	Online	Rank by size

RULES

1. Be Civil
2. Spam
3. Advertising

BANNER/ICON CREDITS

Icon made by Icon Pond from www.flaticon.com.

Banner photo credits (not all pictures show on all devices). Mods are open to suggestions for new pictures as well. All

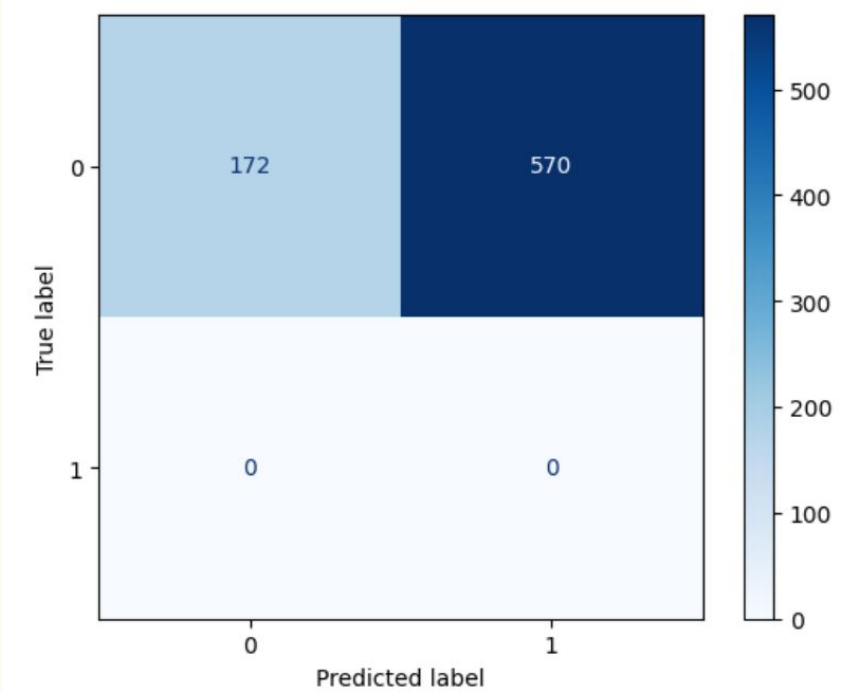
Challenge the model

Web Scrape

Assign Label

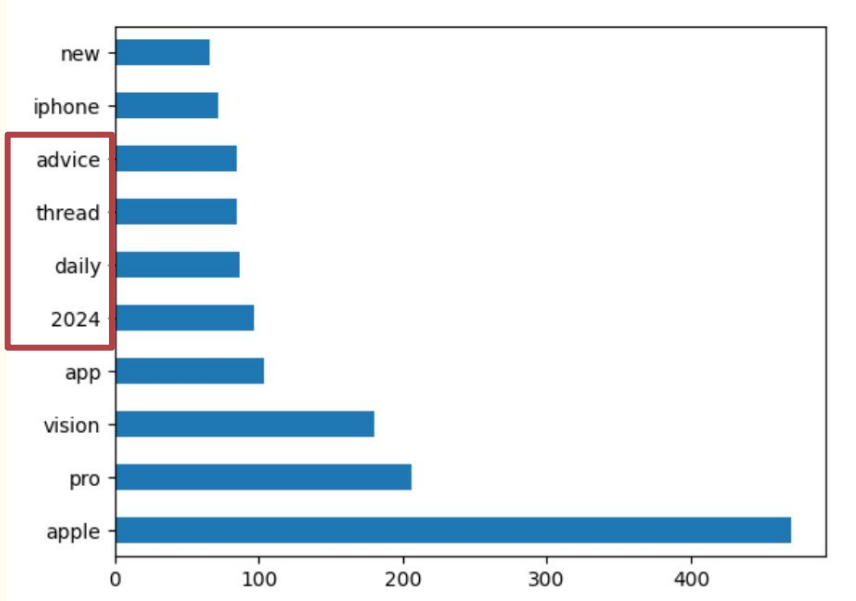
Transform

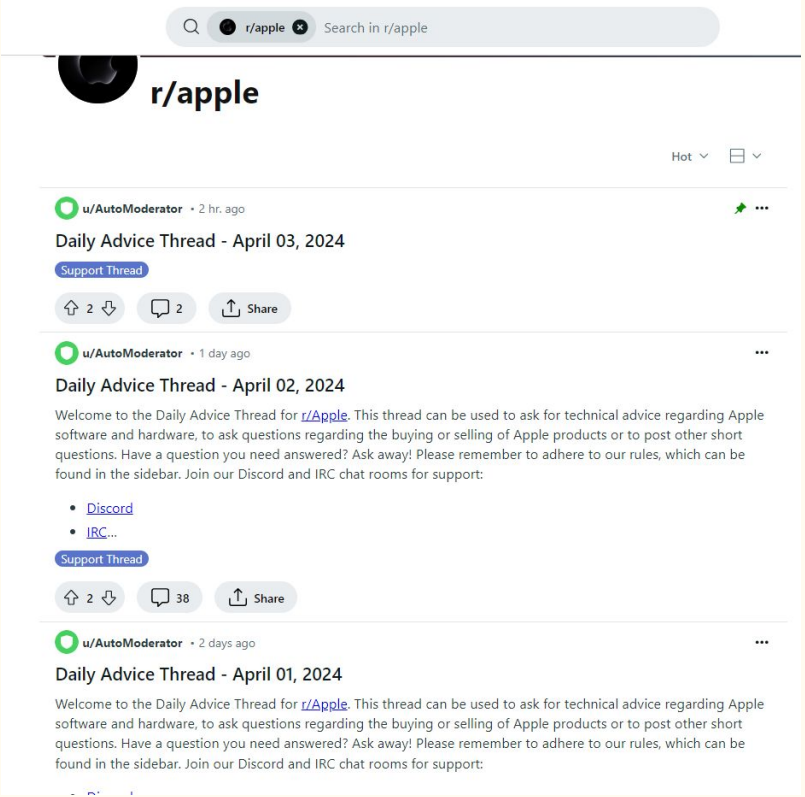
Predict



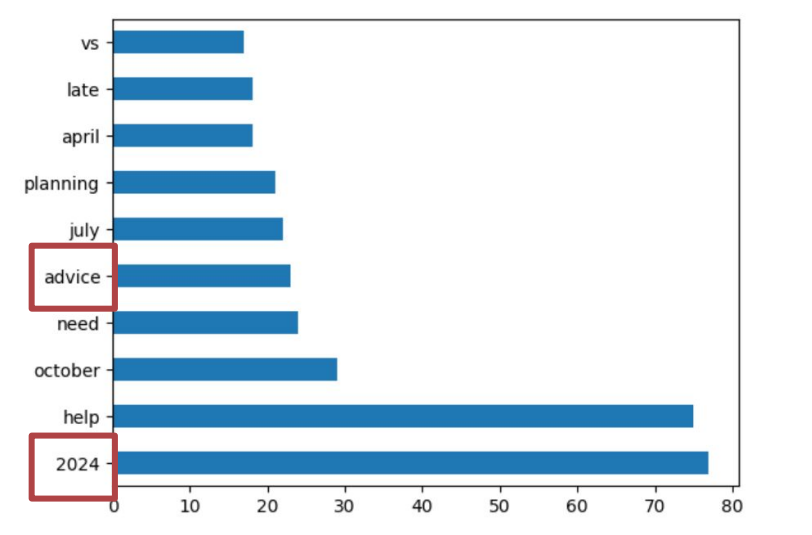
Accuracy: 23%

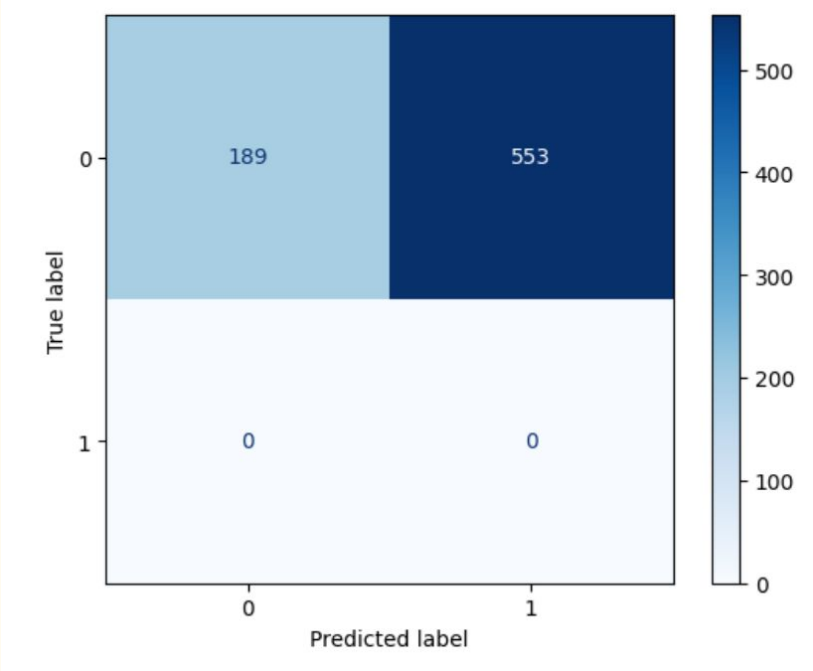
Apple


new
iphone
advice
thread
daily
2024
app
vision
pro
apple
0
100
200
300
400


r/apple
Search in r/apple
r/apple
Hot
u/AutoModerator • 2 hr. ago
Daily Advice Thread - April 03, 2024
Support Thread
2
2
Share
u/AutoModerator • 1 day ago
Daily Advice Thread - April 02, 2024
Welcome to the Daily Advice Thread for r/Apple. This thread can be used to ask for technical advice regarding Apple software and hardware, to ask questions regarding the buying or selling of Apple products or to post other short questions. Have a question you need answered? Ask away! Please remember to adhere to our rules, which can be found in the sidebar. Join our Discord and IRC chat rooms for support:
Discord
IRC...
Support Thread
2
38
Share
u/AutoModerator • 2 days ago
Daily Advice Thread - April 01, 2024
Welcome to the Daily Advice Thread for r/Apple. This thread can be used to ask for technical advice regarding Apple software and hardware, to ask questions regarding the buying or selling of Apple products or to post other short questions. Have a question you need answered? Ask away! Please remember to adhere to our rules, which can be found in the sidebar. Join our Discord and IRC chat rooms for support:

Challenge the model





Accuracy: 25%

Web Scrap

Exploratory Data Analysis

Pre-processing & Modeling

Challenge the model

Drawback & Conclusion

Drawbacks

- Unable to generalise
  - It is only trained on two subsets of a full set, lack holisticity
  - “Problem Statement” problem?
- Redundant information not optimised
  - E.g. “Daily Advice Thread”, or other numerical values
  - Data cleaning via Regex
- Too easy, or “naive”
  - Keywords appearance, e.g. “apple” and “android”
  - Removal of keywords could impact accuracy

Conclusion

- *Problem Statement*

  How can we effectively distinguish posts about Apple amongst data that consists of post from both Apple and Android?

★ Binary classification model - with 94% accuracy
★ Limited to Apple and Android only
★ Consider using LLM for generalisation purpose

THE END.