Project 3:
Classification of Reddit Posts

Team Members:
Hong Aik
Mitchelle
Shu Yi
Yong Gui
Wee Hong

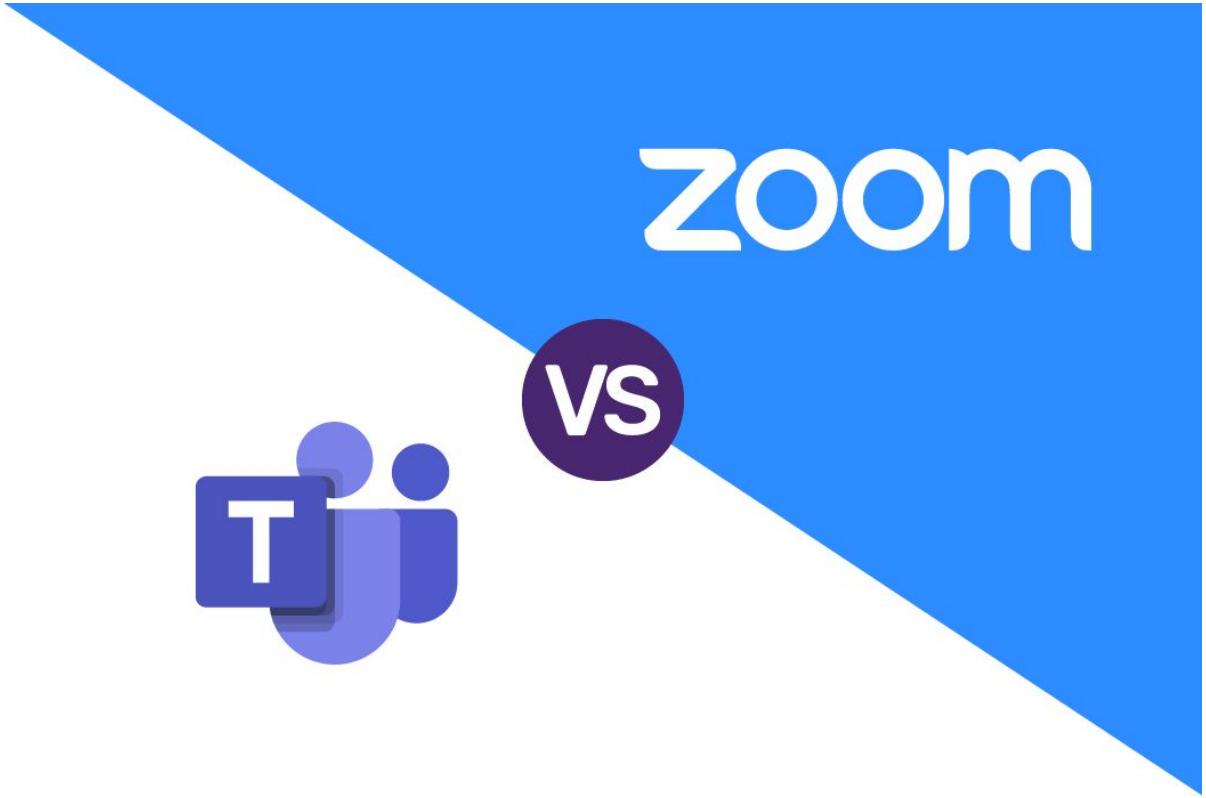
zoom


zoom

Introduction


zoom
zoom
zoom
zoom
Mitchelle Chua
Shu Yi
Wee Hong
Hong Aik

Problem Statement



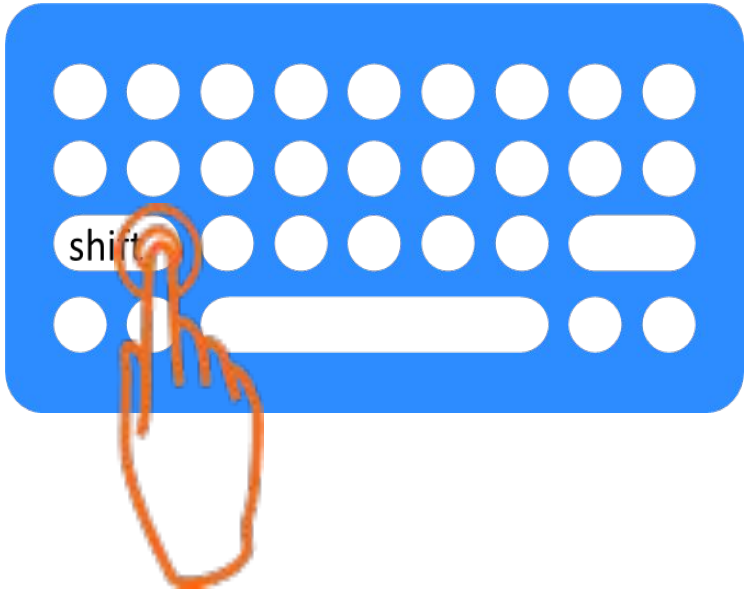
Data Collection & Cleaning

Data Collection

Pushshift API (https://github.com/pushshift/api)

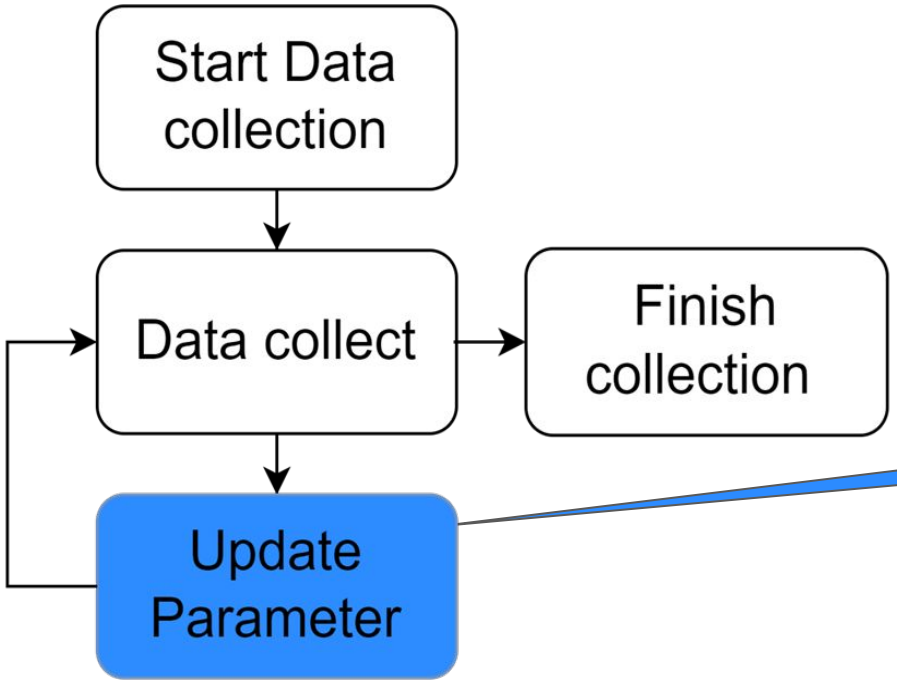
Parameters:

1. subreddit: “Zoom” / “MicrosoftTeams”
2. size: 100 (maximum)
3. after: *epoch value*
   a. first value: 1577836800
   b. Data and Time (GMT): 1 Jan 2020, 00:00



zoom

Data Collection



Pushshift API
(https://github.com/pushshift/api)

Parameters:

1. subreddit
2. size
3. after

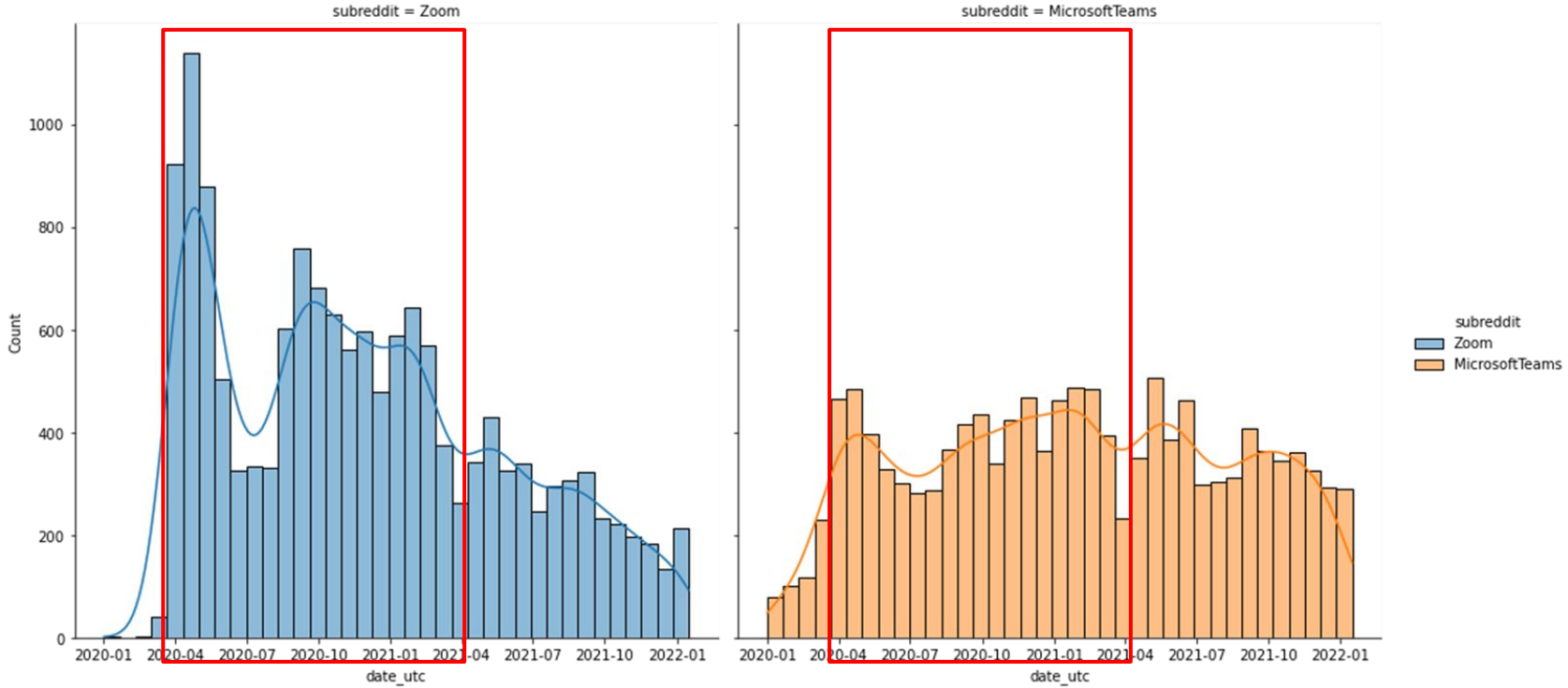
Cleaning

Cleaning of combined column of selftext and title:

- HTML Special entities (e.g. &amp)
- Hyperlinks
- Punctuation
- Whitespace
- Characters beyond Basic Multilingual Plane (BMP) of Unicode
- [removed]
- [deleted]

Time Period

Time Series graph for time frame selection



Preprocessing

Lemmatize

```
1 from nltk.stem import WordNetLemmatizer
2 lemmatizer = WordNetLemmatizer()
```

```
1 words=['feet','dogs','children','identify','this']
2 for word in words:
3     print(f"{word}: {lemmatizer.lemmatize(word)}")
```

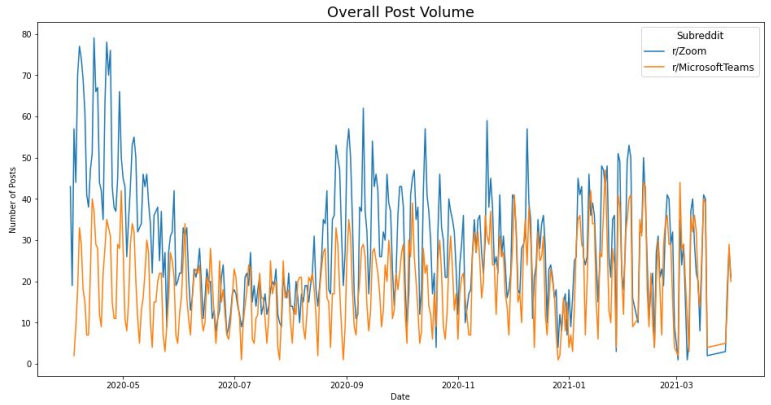
```
feet: foot
dogs: dog
children: child
identify: identify
this: this
```

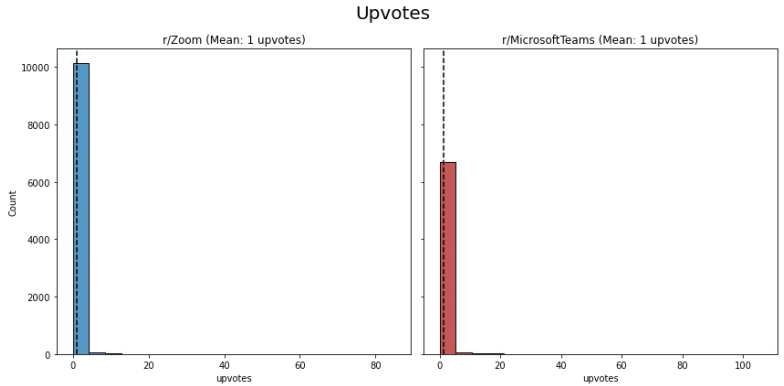
**Image from*
https://karkig.medium.com/understand-stemming-and-lemmatization-with-python-nltk-package-77973a727040

EDA

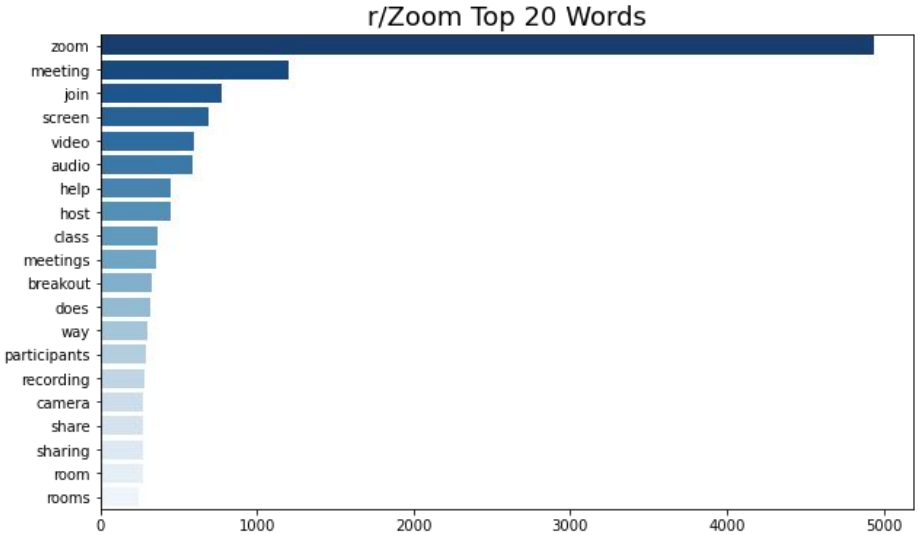
EDA Summary

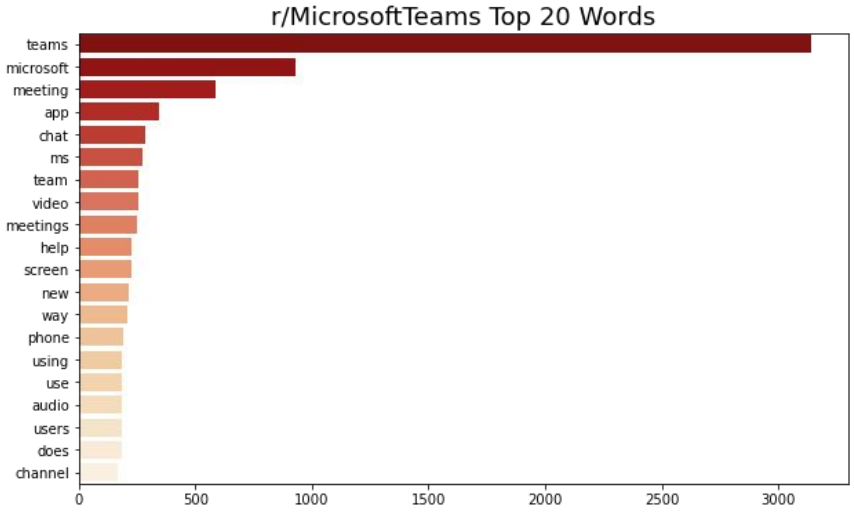
- Analysis of extracted data from r/Zoom & r/MicrosoftTeams posting
  - Analyzing the post volume
  - Analyzing reddit specific data like upvote, words length in selftext and others
- Analyzing text in both title and selftext column
- Perform sentiment analysis using VADER
- Research on words with Scattertext



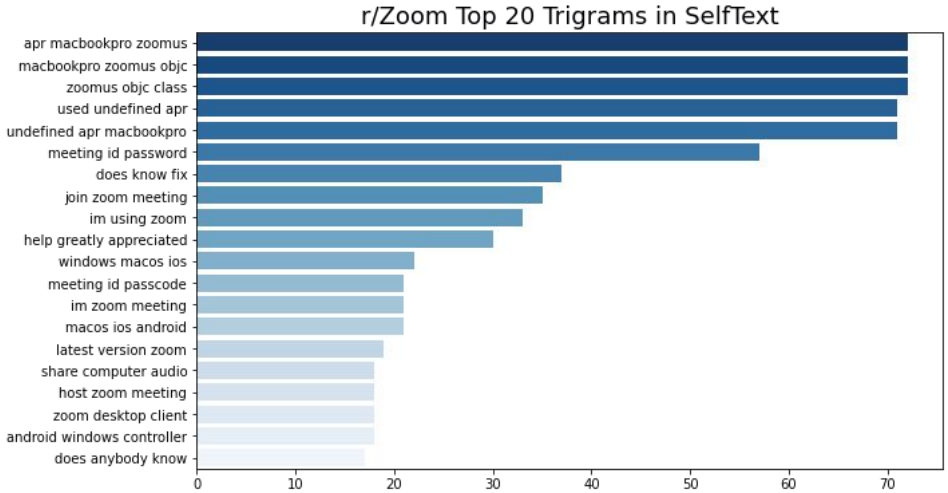


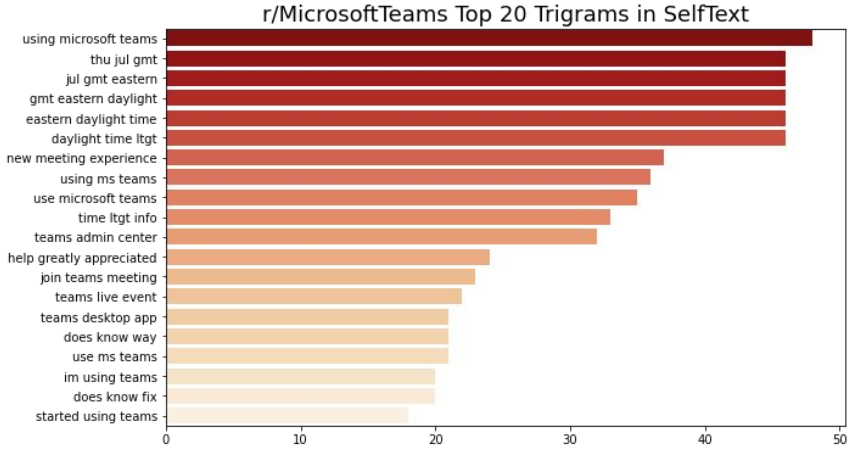
Analyzing Words in Title



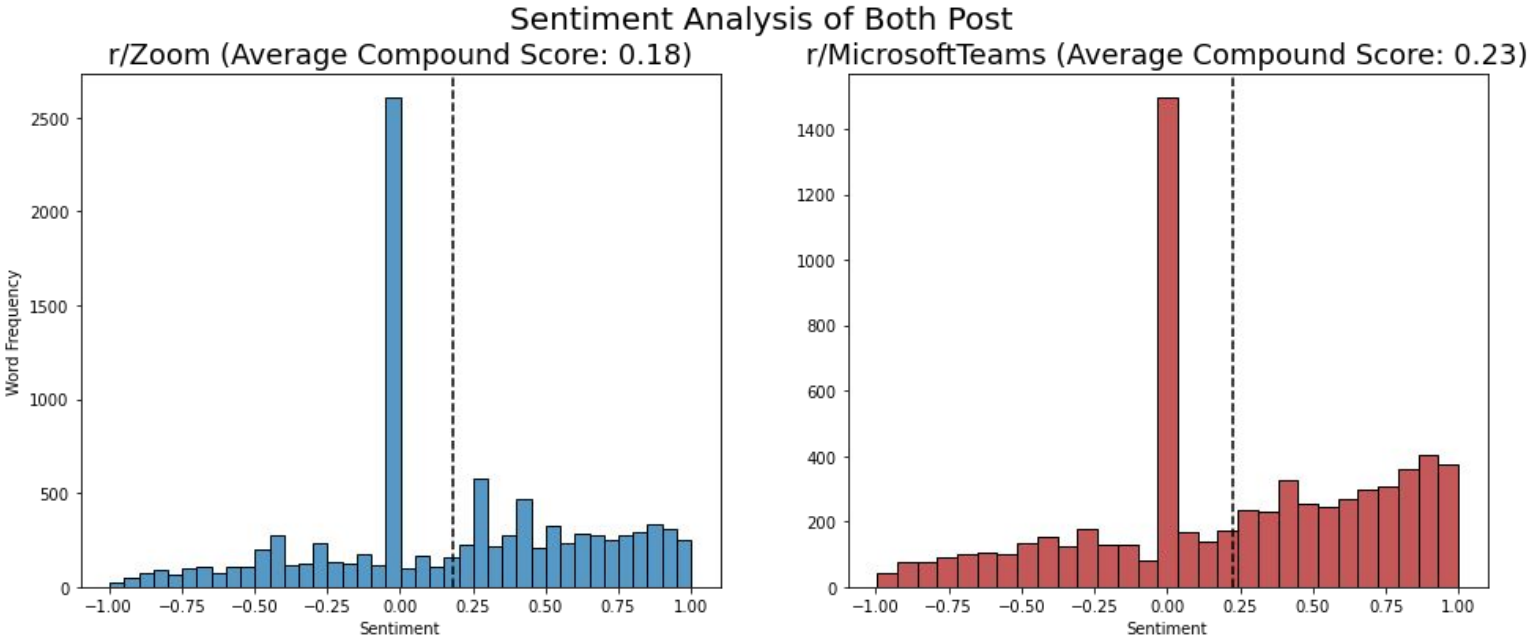


Analyzing Words in Selftext



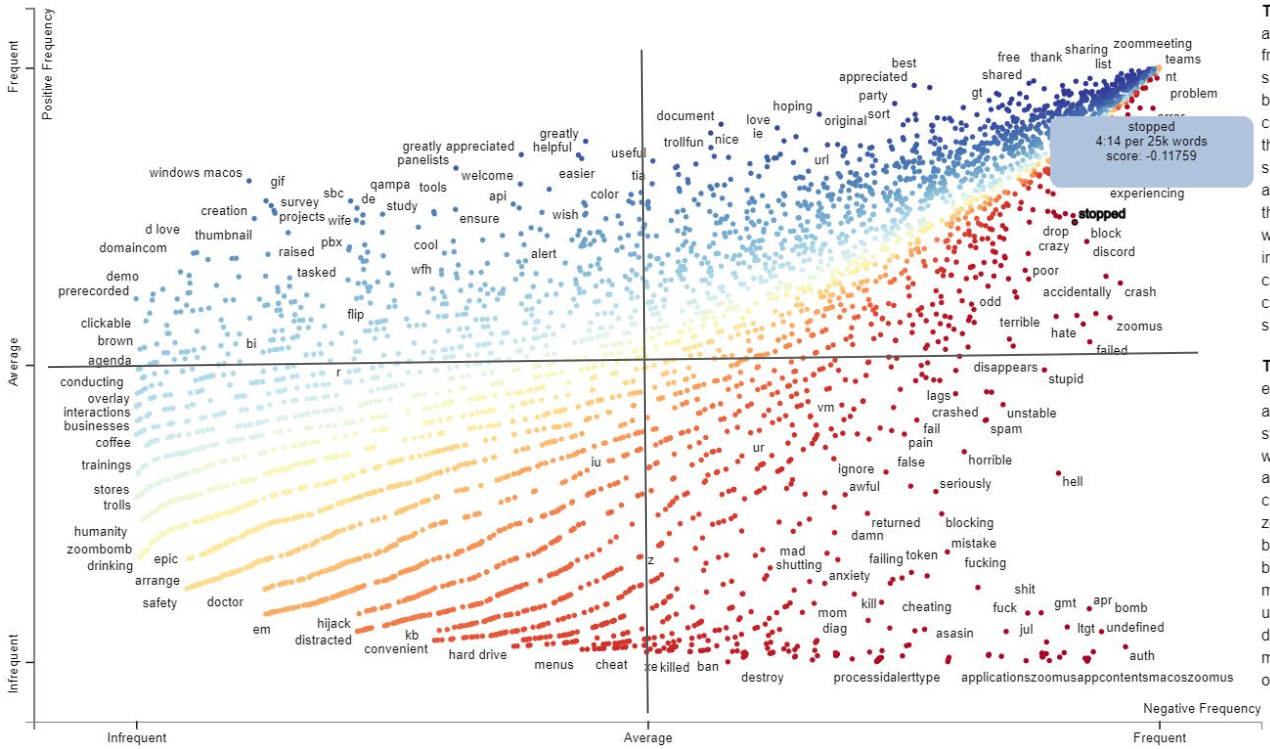


Sentiment Analysis



Observation: Based on the average compound score computed, it seems to indicate that there are more positive posting on Team as compared to Zoom (this can also be observed in the distribution on the right.

Scattertext Visualization - Part 1



Positive document count: 13,110; word count: 735,707
Negative document count: 3,921; word count: 255,680

Scattertext Visualization - Part 2

Term: stopped

Positive frequency:
4 per 25,000 terms
3 per 1,000 docs
Some of the 48 mentions:

r/Zoom
sound not working on screen sharing i got it to work the other night with an online game called jackbox and everything went smoothly besides the game sound being too loud and not being able to hear people speaking now the sound has **stopped** working even when i check the share computer sound button i cant figure out why its not working any suggestions im logged in on my computer for screen sharing and ipad for video same account but i had no issues the other night ive also logged into a family members device to further test the sound not working

r/Zoom
employees using work zoom to record sexy videos of themselves then saved to cloud hi i am the administrator of our zoom account and i saw that there was a new recording on our business account when i loaded the recording i saw that the recording was of my coworker and of her boyfriend outside of hours she who is was doing a sexy dance and it was clear they were zoom sexing and she had recorded it i am guessing on accident cause she is bad at computers and i am on the phone with her all day fixing it problems for her i **stopped** watching went back to the admin account and deleted her recording and didnt say anything to anyone in retrospect maybe i should have what is your policy about recording meetings this was so awkward i am thinking about sending a message to all staff and letting them know that due to security concerns we can no longer record meetings and please do not use your company zoom outside for personal calls what do you think

r/Zoom
my microphone **stopped** working during a call i just had my first class via zoom at first all worked perfectly well both camera and headset then the professor split us into groups and had us work in breakout rooms which too worked fine however upon returning to the main room my mic didnt work anymore or rather the others only got some sort of buzzing noise when i talked over the rest of the session i did get several pop up notifications that i was apparently in muted mode and that the host had set everyone to mute none of the others were affected by this however any ideas what went wrong there

r/Zoom
urgent how to make attendees muted and video **stopped** and not allowed to turn on video and audio please help i am assisting a zoom meeting where we wanted guests to enter zoom with video and audio both muted but bunch of people were able to turn on their video we had turn off audio and video upon entry is this a matter of setting up a webinar and if so what setting do i use to make it so attendees enter zoom unmuted video off and not able to turn on video without hosts and cohosts permission please walk me through this as our talk did not go as planned and had bunch of people talking when it was only suppose to be so they were nonvideo participants and not allowed to show video

r/Zoom
screen goes black on zoom while screen sharing other participants can still seehear me hi im posting here to see if someone has had

Negative frequency:
14 per 25,000 terms
13 per 1,000 docs
Some of the 63 mentions:

r/Zoom
virtual background wo green screen **stopped** working i had been doing virtual backgrounds with green screens for a while on my laptop but since a recent update it stopped working and gives the message your computer doesnt meet the requirements when i try to disable green screen this is after a recent update as far as i can see laptop meets requirements lenovo yoga x anyone has seen this

r/Zoom
snap camera **stopped** working after recent zoom update deleted

r/Zoom
best way to use my s as a webcam on my desktop had problems with droidcam so i had no problems up until a few days ago except once my earphones **stopped** working when my cheap logitech broke im guessing the usb cable has problems so i tried using my phone as a webcam but i never got it to work using droid cam ever since i cant even hear from it properly i figured out it was the the droid mechanical voice filter that i didnt even set it to but i cant make sense of what they are saying like delay robotic vocie makes everything so bad it creeped me the hell out at first iv cam you have to pay for which now doesnt seem like a bad idea because i rather not buy a new webcam tlr what are the best ways to use galaxy s as a webcam to connect to the computer wireless would be ideal but whatever works i have a desktop with no wifi starting to feel desperate at this point thanks

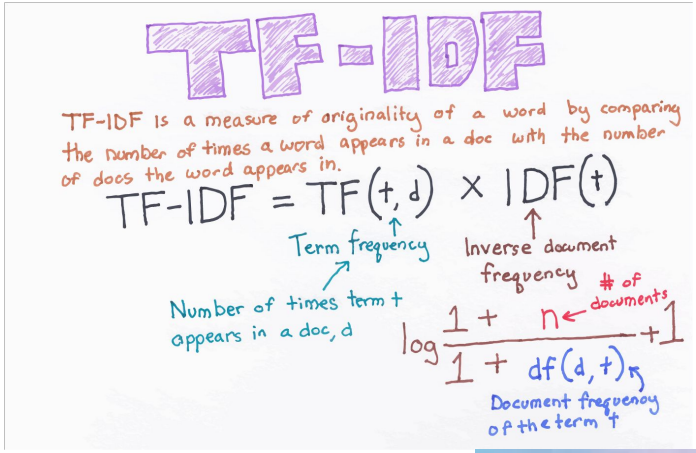
r/Zoom
zoom and split screen view on android hi i was on zoom and my split screen view on my galaxy just **stopped** working can teachers see this additional screen

r/Zoom
is there a way to get the virtual backgrounds feature working without a greenscreen if only my cpu driver doesnt fit the minimal requirements so i tried updating the drivers on my laptop and at first it worked like a charm but then the screen on my laptop **stopped** working and i had to downgrade back to my old driver that was made for this laptop but the version of it is too old to run the feature and it has problems with premiere as well the laptop is an asus rog glvw

r/Zoom
music recital got zoombombed people from my school decided to have our trim music honors society induction ceremony over zoom and the first half of the concert consisted of a random person joining under different names screaming and making inappropriate sounds during peoples musical performances drawing all over peoples videos of their musical performances and writing shut up and other inappropriate things on the video i dont know how they were writing and drawing on someones video and at one point even screensharing their phone and looking inappropriate things up on google one of the students even had to start their song all over again because the zoombomber continued to scream over their performance eventually we **stopped** the meeting and started a new

Modelling & Evaluation

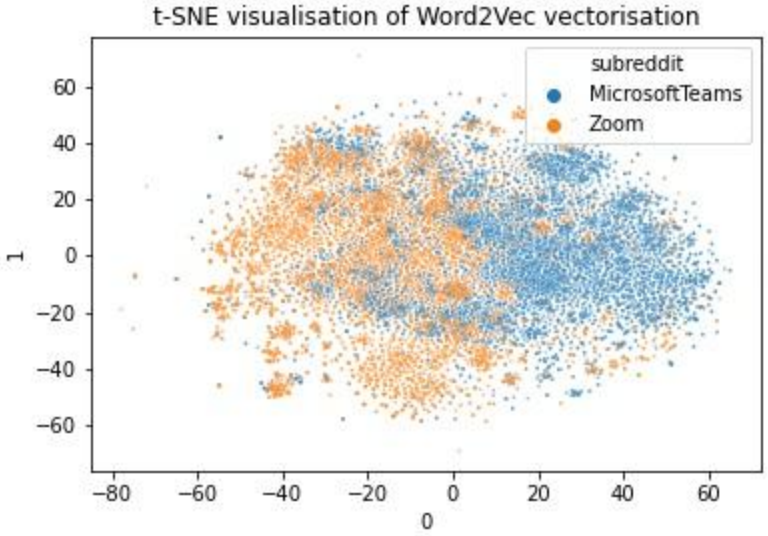
Text Vectorization

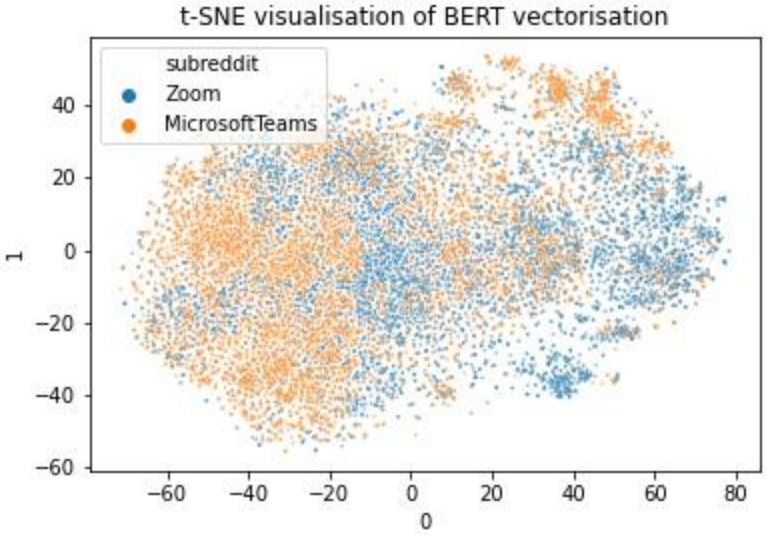


Text Vectorization - Example

	Dim 1	Dim 2	…	Dim 768
I like cats	0.216	0.582	..	-1.424
Cats like me	0.671	0.149	…	1.472

Further EDA





We can also employ t-SNE for dimensionality reduction. We observe that the clusters are denser when using Word2Vec and clusters are better separated.

zoom

Further EDA

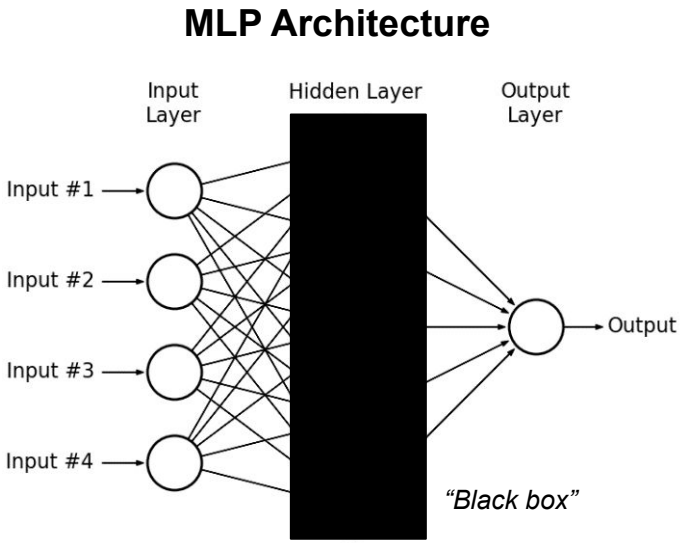
We will experiment with 3 classifiers:

- Logistic Regression (LR)
- Random Forest (RF)
- MultiLayer Perceptron Classifier (MLP)

Baseline:

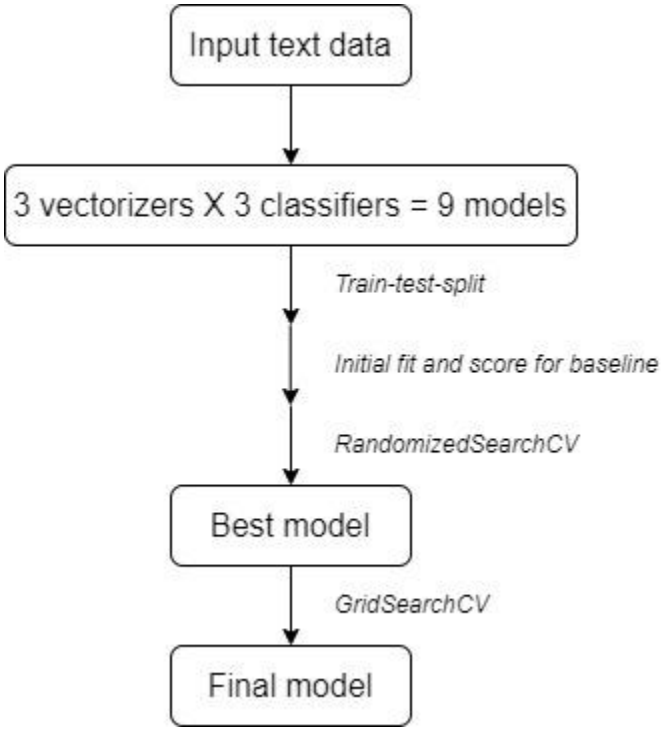
Checking whether submission contains the word “microsoft” or “teams”

Baseline score: **87%**

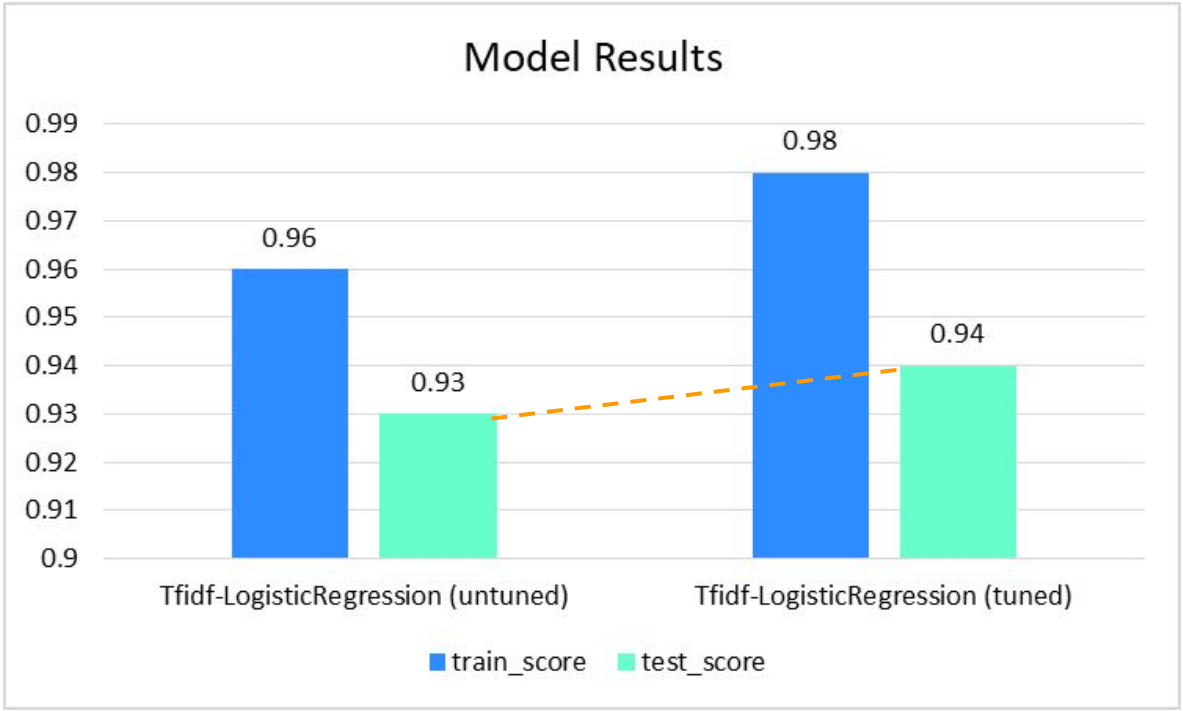


zoom

Data Modeling

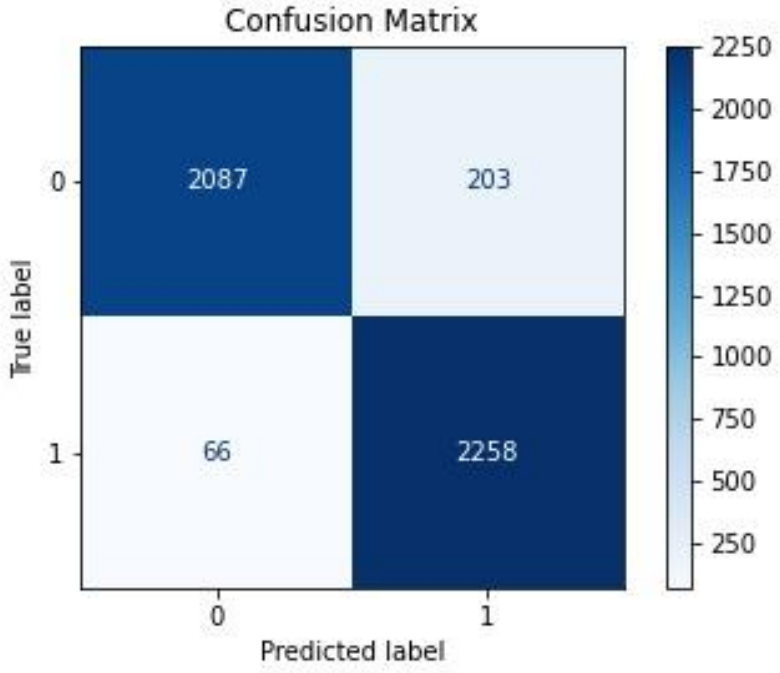


Model Evaluation



We managed to improve our score by ~1% which is decent, given that our untuned model already has 93% accuracy!

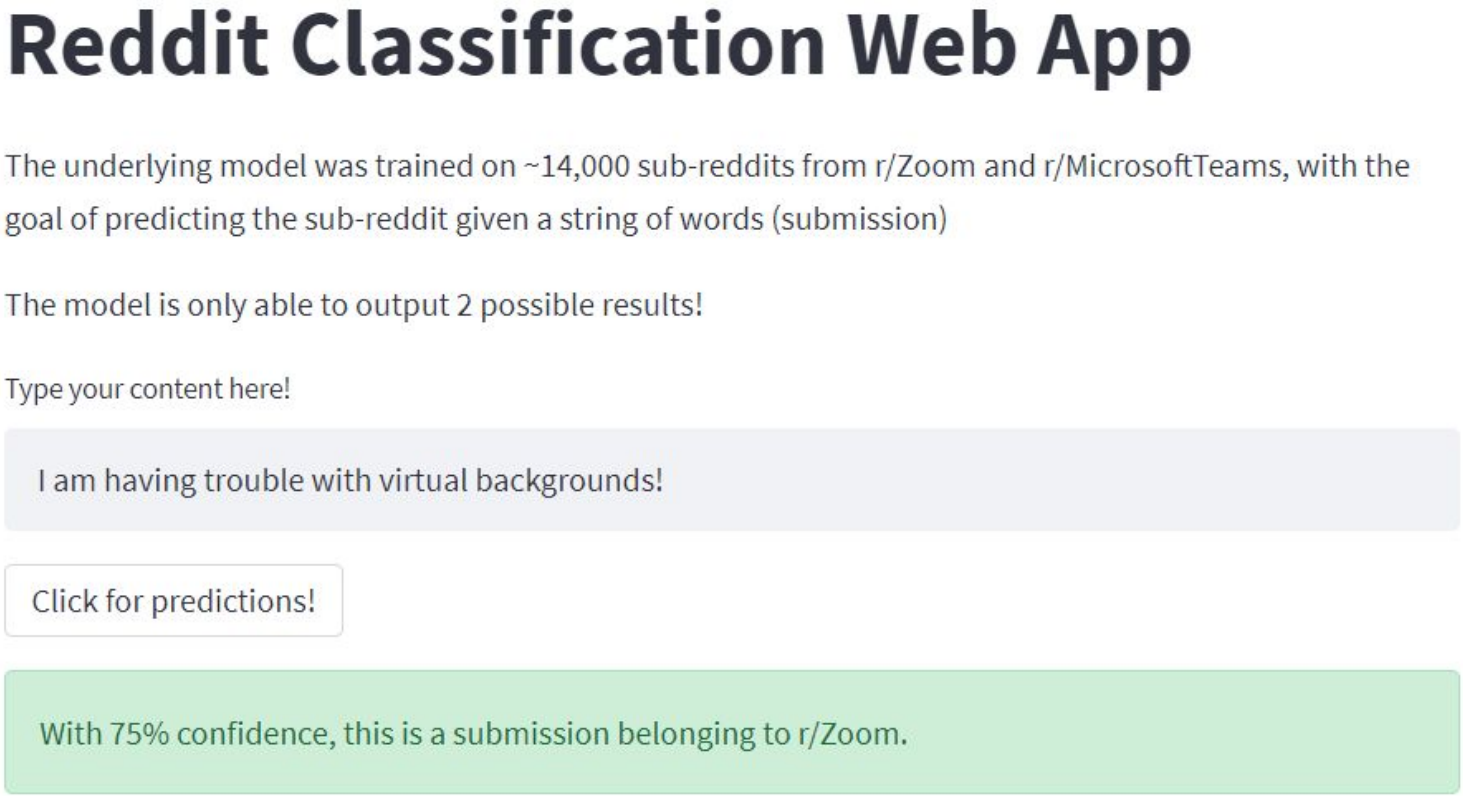
Model Evaluation



	Confidence
Predicting MST	88%
Predicting Zoom	84%

zoom

Deploying the Model



Conclusion

Recommendation for Software Development Team

Pain points for users

1. Stopped
2. Drop
3. Crash
4. Reinstall
5. Error

Keep an eye on

zoom

Recommendation for Digital Marketing Team

Top words for Zoom

1. Zoom
2. Password
3. Host
4. Join
5. Participant
6. Class
7. Virtual
8. Breakout
9. Room
10. Id

Top words for MST

1. Team
2. Guest
3. Microsoft
4. Channel
5. User
6. Assignment
7. Call
8. Notification
9. Chat
10. Feature

Zooming ahead

Refining our current model

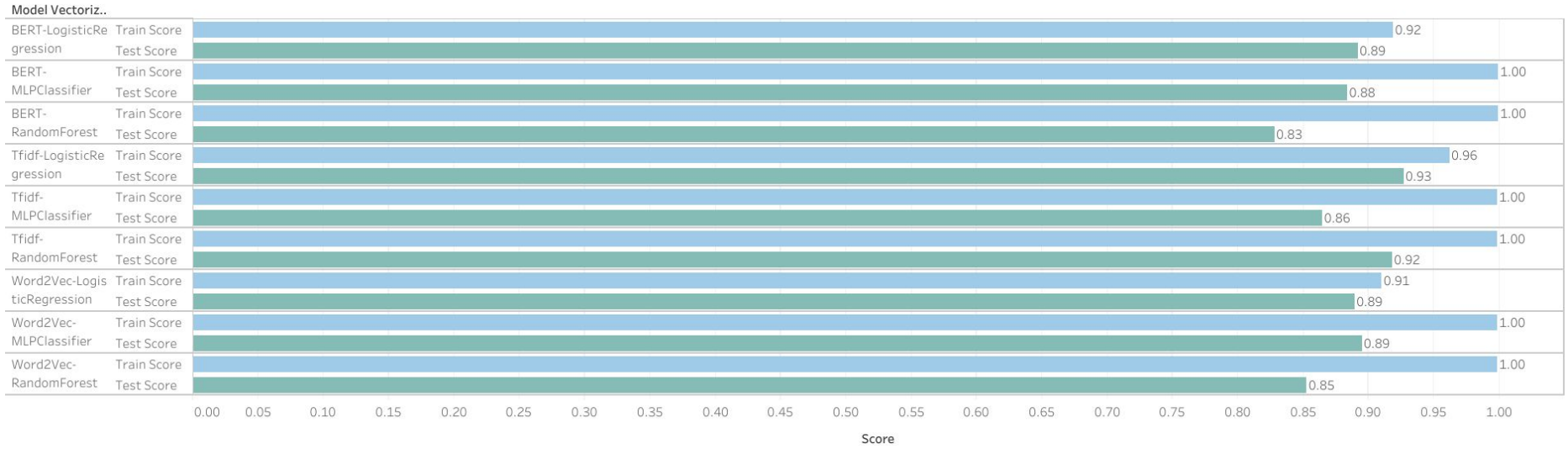
Training our model to recognise words unique to subreddit.

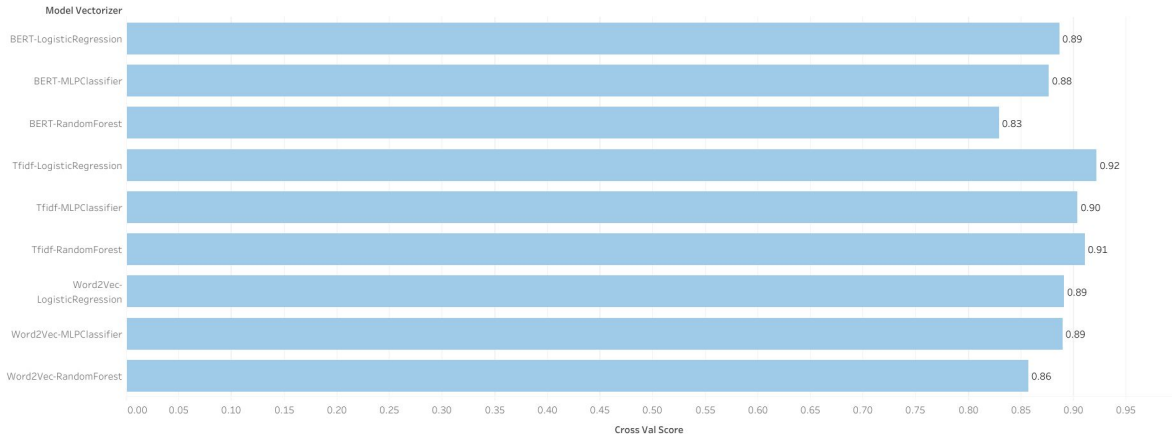
Running the refined model on other competitor pairing

Zoom vs Google, Zoom vs Skype

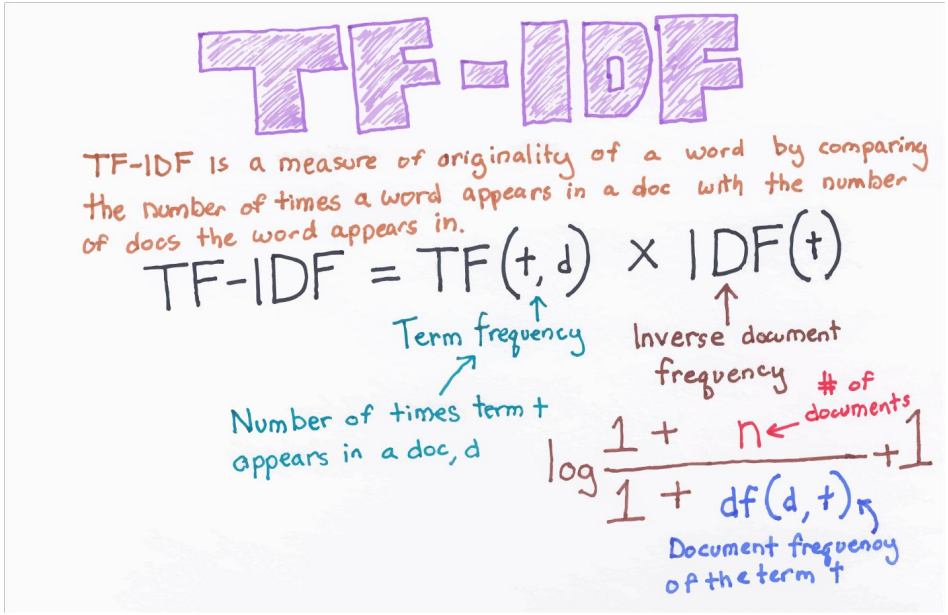
Thank You

Annex


Model Vectoriz..
BERT-LogisticRegression
BERT-MLPClassifier
BERT-RandomForest
Tfidf-LogisticRegression
Tfidf-MLPClassifier
Tfidf-RandomForest
Word2Vec-LogisticRegression
Word2Vec-MLPClassifier
Word2Vec-RandomForest
Train Score
Test Score
0.92
0.89
1.00
0.88
1.00
0.83
0.96
0.93
1.00
0.86
1.00
0.92
0.91
0.89
1.00
0.89
1.00
0.85
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00
Score


Model Vectorizer
BERT-LogisticRegression
BERT-MLPClassifier
BERT-RandomForest
Tfidf-LogisticRegression
Tfidf-MLPClassifier
Tfidf-RandomForest
Word2Vec-LogisticRegression
Word2Vec-MLPClassifier
Word2Vec-RandomForest
0.89
0.88
0.83
0.92
0.90
0.91
0.89
0.89
0.86
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95
Cross Val Score

Text Vectorization and Further EDA - TfidfVectorizer



	I	like	cats	me
I like cats	0.1	0.1	0.1	0
Cats like me	0	0.1	0.1	0.1

Text Vectorization and Further EDA - Word2Vec

	Dim 1	Dim 2	…	Dim 300
I like cats	0.01	0.01	…	0.01
Cats like me	0.01	0.01	…	0.01

Text Vectorization and Further EDA - Bi-directional Encoder Representations from Transformers

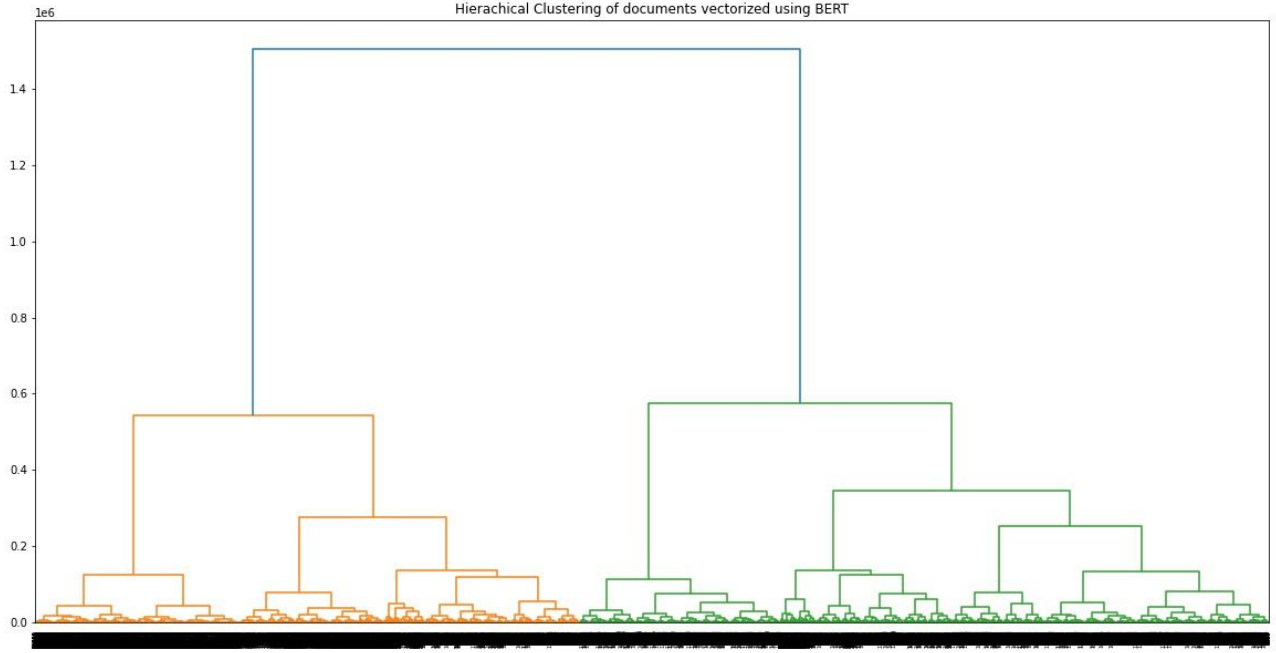
	Dim 1	Dim 2	…	Dim 768
I like cats	0.01	0.01	…	0.01
Cats like me	0.01	0.01	…	0.01

Exploring Misclassifications

	predictions	true	probability	text_nostop
640	0	1	0.467682	hi schedule recurring meeting phone app show online account anyone know sync these tia app sync
9965	1	0	0.835560	please help camera hidden share screen participant get bored
10197	1	0	0.815036	aspect ratiorequirements virtual background
4410	0	1	0.480724	sit window left side table take call face look like halfmoon one side lighted side lighted course could shift 90 degree face window need refer desktop pc time make sense keep alternating window back pc every time take video call idea one problem thought getting ring light seem many option webcam ring light even ring light seem good enough illuminate face compensate uneven lighting deal uneven lighting taking video call
7775	1	0	0.510970	suggestion good way prevent student cheating multiple choice form quiz example failing student finish difficult physic quiz 7 second score 100 something definitely i give openended question would like make testing process somewhat manageable right cheating
479	0	1	0.425665	mute presenter shared sound presenter know shared sound

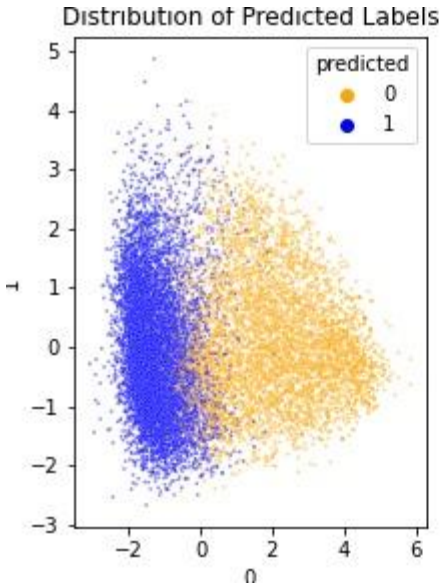
zoom

Further EDA



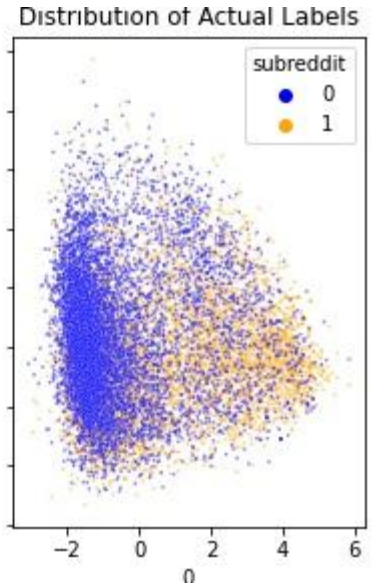
Agglomerative clustering shows that there are 2 distinct clusters in our word vectors (BERT), although this may not necessarily correspond to our Zoom and MST sub-reddits.

Further EDA



The clustering earlier has nicely separated the data into 2 clusters, although the overall accuracy is not high.

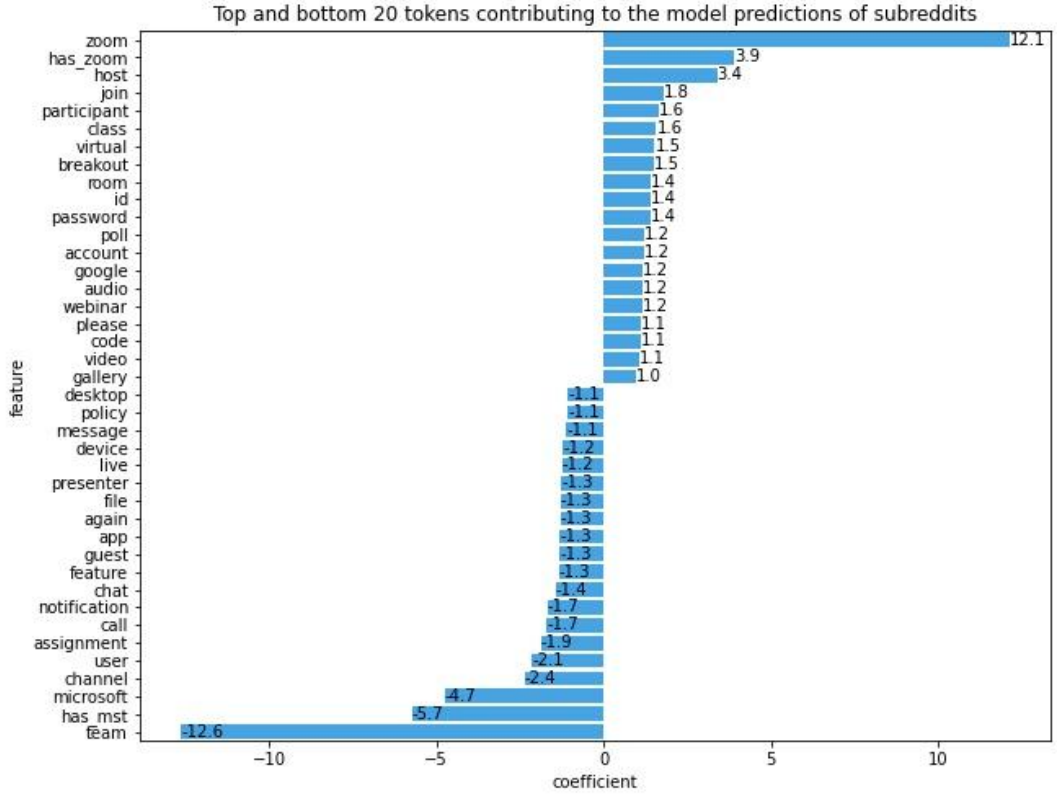
The labels have also been “inverted”, although we can be fairly certain that 0 (actual) = 1 (predicted)



We see a rather messy overlap of points from Zoom (blue) and MST (yellow) texts, although there is a strong Zoom cluster on the left. This is expected given that we have “over-simplified” the data

zoom

Understanding the Final Model



Example of Negative & Positive Posting about Zoom

Negative Posting about Zoom:

zoom troubleshooting help alright i have a weird issue i have one student in a class of that can join the zoom meeting but is then dropped seconds after no other student is experiencing this issue the teacher is not having any network lag the student can access and watch other classes just fine student is using a chromebook student is using a hotspot student hasnt indicated this has been a problem for the last month issue only occurs at am class father is adamant that the issue is with zoom or the teacher but i believe that can be ruled out since no other student is having the same problem the device can be ruled out as a problem since connection with other classes is fine my thought is the hotspot they are currently using a sprint mifi unlimited plan after some quick research i found that sprint deprioritizes connections after theyve used gb in a billing cycle i am assuming the am time is a busy network for the tower any other potential ideas what it could be

Positive Posting about Zoom:

trying to host workouts on zoom and i cant figure out how to play spotify in my meeting without screen share apparently there is a program called loopback that does exactly this but im hoping someone knows of a free option thanks in advance

Example of Negative & Positive Posting about Teams

Negative Posting about Team:

problems when logging in on laptop with error code whenever i want to start ms teams i always get a error code and it asks me to restart the application whenever i restart my application it again gives me the same error if i log in on my other computer via google chrome or with a clean install it will work and i dont seem to have any problems does someone know whats the problem here

Positive Posting about Team:

generate report of all microsoft teams users hello is it possible to create a report of all active microsoft teams users from the teams admin portal ideally i would like to create a report that includes usernamesemails of all the microsoft teams users thanks