

# Classification of Emotions for Text Messages

By: Alyse Su (DSI-16)



# WHO ARE WE AND WHAT DO WE DO?

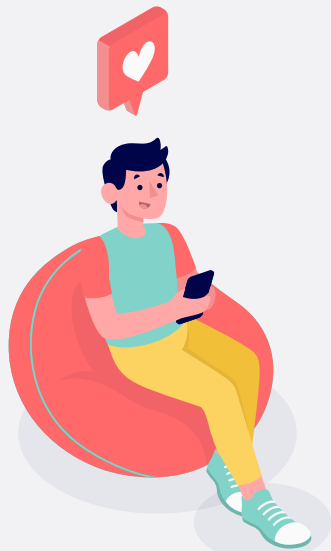
## WHO ARE WE

Team of data scientist in a communications department in a renowned Social Media Platform provider.

## PROJECT OF INTEREST

Build a casual chat sentiment analyzer as a value-added product for users of our communications' platform

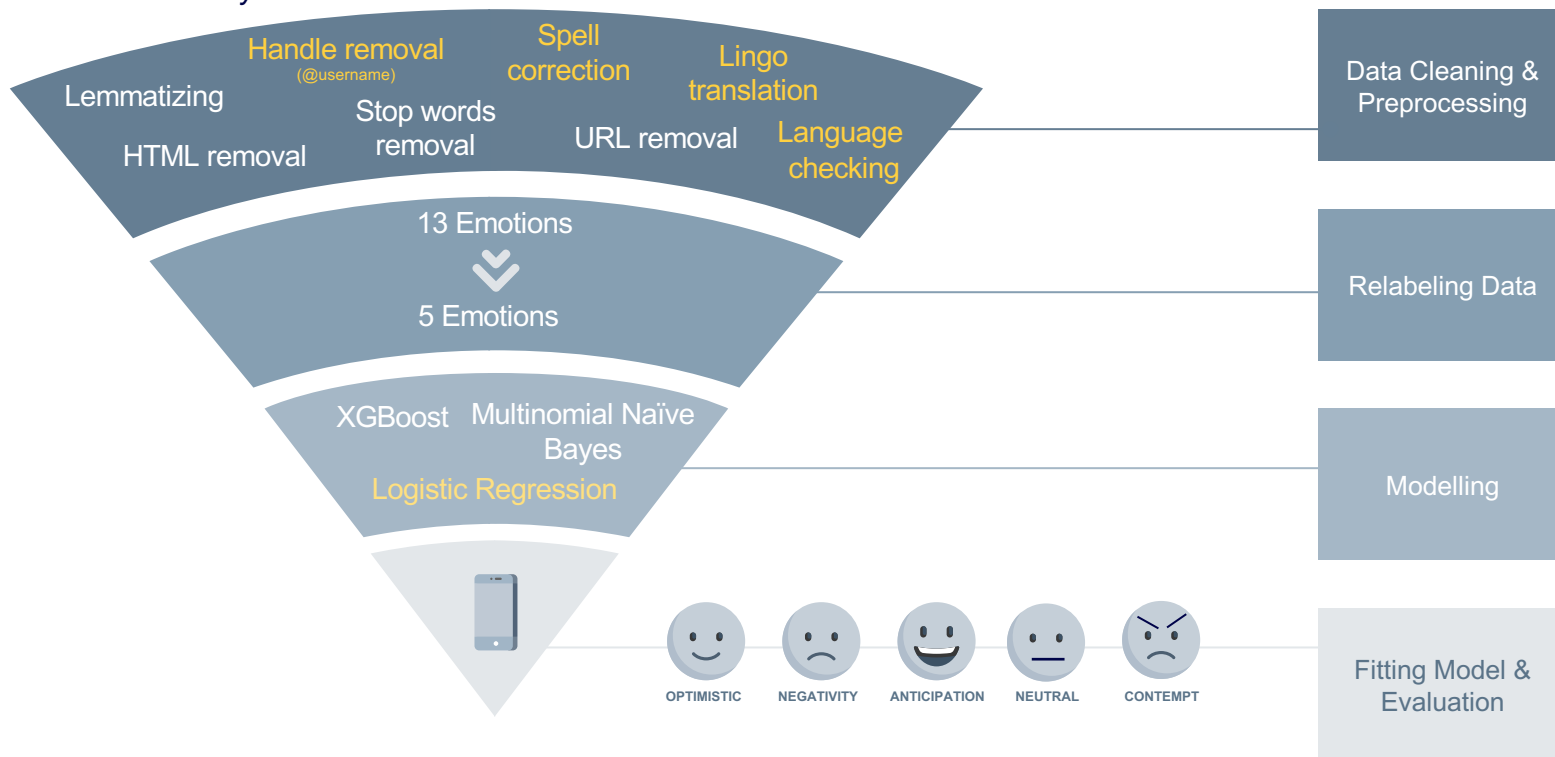
In collaboration with the Software Engineering team



# OVERALL METHODOLOGY

40K tweet data from  
a similar study

±2k tweet data (anger/hate)  
to supplement main dataset



# CLEANING AND PREPROCESSING

## BASIC CLEANING:

- HTML, URL, HANDLE REMOVAL
- STOPWORDS, LEMMATIZING

## ADDITIONAL CLEANING:

- LANGUAGE CHECKER
- SPELLING CHANGES
- LINGO TRANSLATION
- RECLASSIFICATION OF LABELS

- Using Spacy Spell Checker to filter for non-English content
- Parsing texts with 'unknown' language for SymSpell's spell correction

For variations

of 'haha'

```
def replace_hah(string):  
    results_1 = re.sub(r"\S+hah\S+", 'haha', string)  
    results_2 = re.sub(r"\S+aha\S+", 'haha', results_1)  
    results_3 = re.sub(r"hah\S+", 'haha', results_2)  
    results_vf = re.sub(r"hah", 'haha', results_3)  
    return results_vf
```

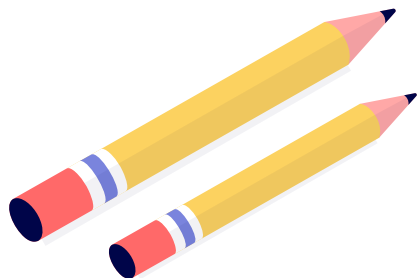
Snippet of

Lingo Dictionary

	short	long
7721	zomfg	oh my fucking god
7722	zomg	Oh my God
7723	zomgzorrz	oh my god
7724	zup	what's up?

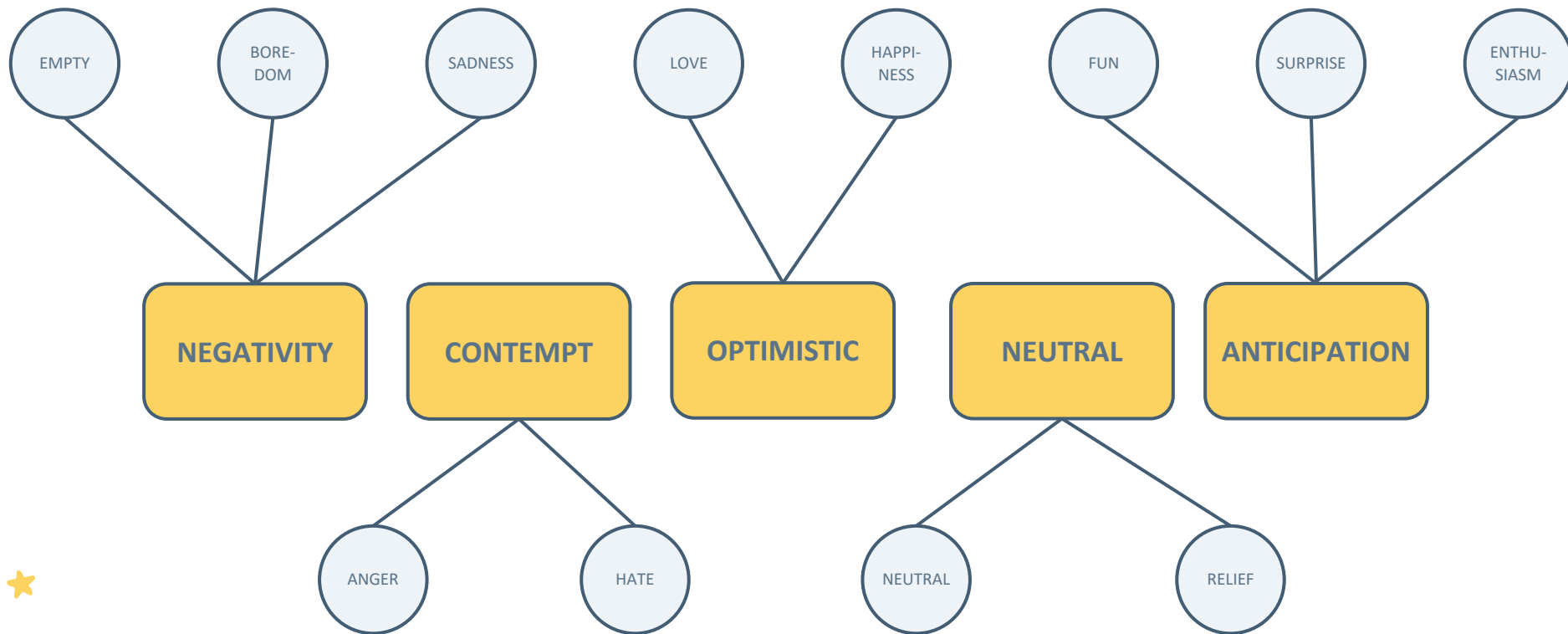
- Keyword: "Happy"
- Keyword: "Thanks"

NEUTRAL → HAPPY



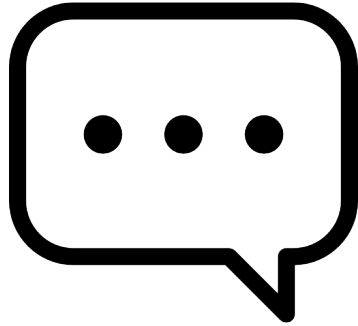
# GROUPING EMOTIONS

○ : ORIGINAL LABELS  
■ : NEW LABELS



# FEATURE EXTRACTION

EXTRACTING NUMBER OF PUNCTUATIONS IN EACH TEXT MESSAGE



## ELLIPSIS

Omitted section of text



## QUESTION MARKS

Doubt or uncertainty



## EXCLAMATION MARKS

Loudness or excitement



# MODELLING



**LOGISTIC  
REGRESSION**

Excellent for  
feature  
inference



**XGBOOST**

Modified tree-  
based model  
that prevents  
overfitting

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

**MULTINOMIAL  
NAÏVE-BAYES**

Statistically  
driven &  
known for NLP  
purposes



# MODELLING RESULTS

## BASELINE MODEL

```
msg_emotion
1      0.351789
2      0.235196
3      0.223409
4      0.113611
5      0.075995
dtype: float64
```

- Baseline accuracy: 35%

I'M 16% BETTER  
THAN BASELINE!

## AFTER GRID SEARCH & MODEL TUNING

#	Models Tested	Accuracy	F1	Precision	Recall
1	Logistic Regression	0.51	0.48	0.50	0.51
2	XGBoost	0.50	0.48	0.49	0.50
3	Multinomial Naive Bayes	0.48	0.46	0.46	0.48



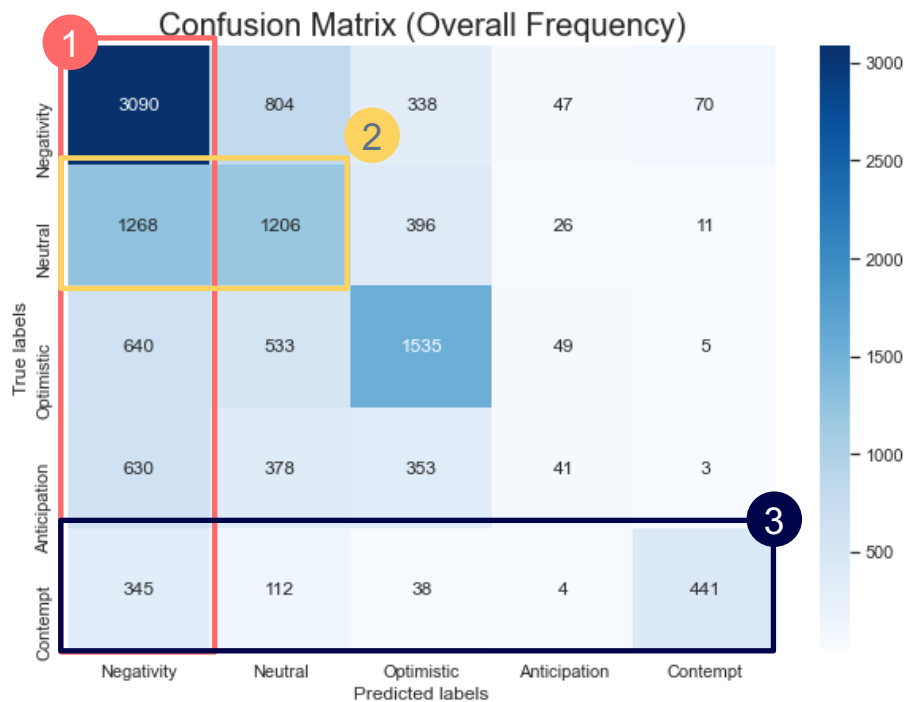
QUALITATIVE  
INTERPRETATION IS KEY!





# UNDERSTANDING MISCLASSIFICATION

## CONFUSION MATRIX



1 General Misclassification INTO Negativity class

2 High Misclassification between Negativity-Neutral pairs

3 Misclassification from Contempt to Negativity

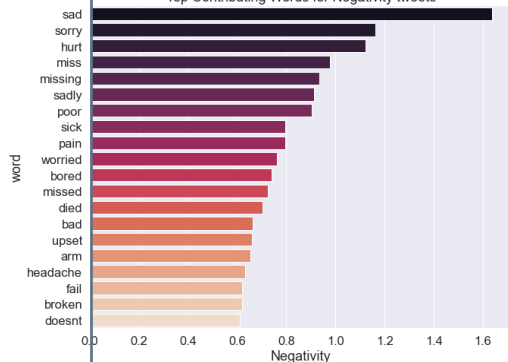
# TOP CONTRIBUTING WORDS (NEGATIVITY & CONTEMPT)

## NEGATIVITY



NEGATIVITY

Top Contributing Words for Negativity tweets

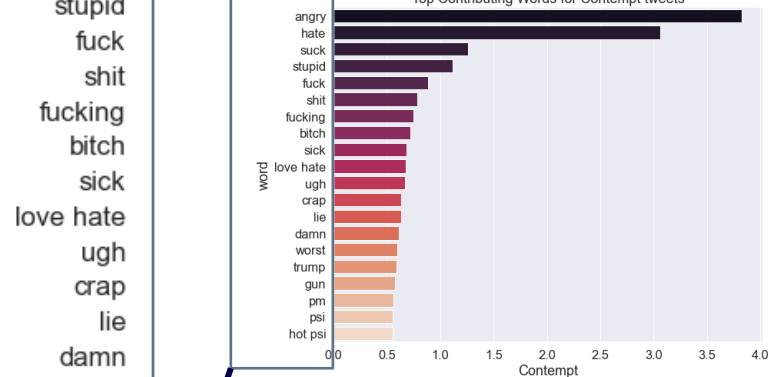


## CONTEMPT



CONTEMPT

Top Contributing Words for Contempt tweets



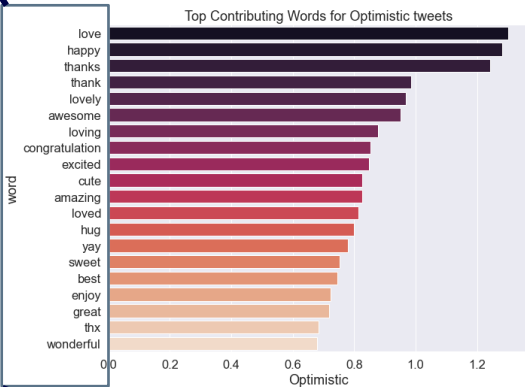
# TOP CONTRIBUTING WORDS (OPTIMISTIC & ANTICIPATION)

## OPTIMISTIC



OPTIMISTIC

Top Contributing Words for Optimistic tweets

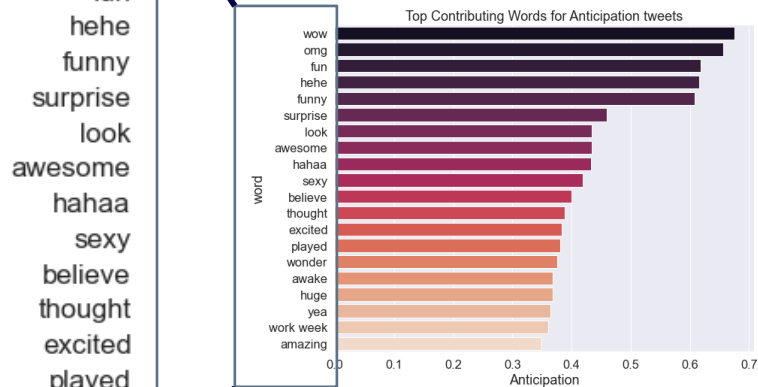


## ANTICIPATION



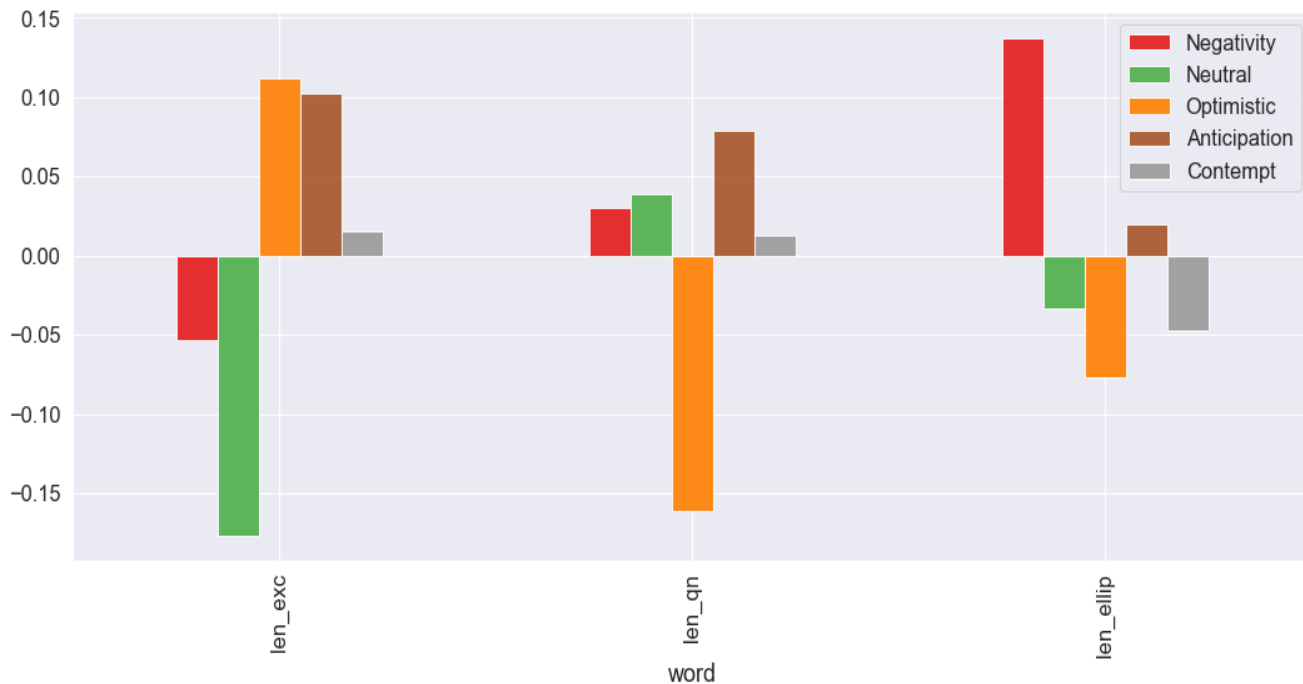
ANTICIPATION

Top Contributing Words for Anticipation tweets



# EFFECT OF PUNCTUATIONS

Effect of Punctuations on Tone of tweets



Positive correlation with:  
**Optimistic & Anticipation**



Negative correlation with:  
**Optimistic**




Positive correlation with:  
**Negativity**




# Misclassified Classes:

## Optimistic-Negativity Pair

**PREDICTED**



"Took a math test today. The day before the test, the teacher says bring your calculator. Luke was confident. Teacher decided no calcs."



**ACTUAL**

**PREDICTED**



Good luck with the design....remember the little people  
Heading to bed, goodnight!



**ACTUAL**

**CONCLUSION: ORIGINALLY DATA COULD BE BETTER LABELLED**



# Misclassified Classes:

Contempt-Optimistic Pair

PREDICTED




I missed this little ferocious one. This is my cat when she was a kitten  
[image link]




ACTUAL

PREDICTED



Shout out to my moms for waking me up so prematurely! Preciate it



ACTUAL

CONCLUSION: ORIGINALLY DATA COULD BE BETTER LABELLED

CONCLUSION: SARCASM IS A KEY CHALLENGE IN NLP SENTIMENT ANALYSIS





# HOW HAVE OUR MODEL FARED IN PREDICTING EMOTIONS OF TEXTS?

## EXAMPLE: ANTICIPATION

112	Hahahahaha Shall we play amongst us?? The latest hottest game that everyone is playing now
76	Wow so hardworking man
118	right now when I sit on it , it is even very squeaky every night my back aches And it is not pleasant. Lol. Nonsensical chats. So funny haha.
82	🤔 Don't go too overboard ok Take some time off to travel the earth Or perhaps just lepak And watch some movies With your girlfriend Huh! I'm your best friend how can I not know that you have a girlfriend???
48	So excited Shall we go for pre drinks??? We can go to Lego Land or watch star wars together Perhaps that would be a fun filled night
115	I mean If you're an imposter you kill people But if you are a crew you try to hide and determine the imposter?? Heard is fun from my friends Why don't you consider streaming your game play online? Can perhaps earn some money I mean you are already super high profile to begin with With 10k followers on your instagrm I'm sure there is a place for you!
66	nah he has turned around and became a better person less abusive more tolerant nicer person altogether. Amazing
114	Yes! We should all call our friends and play together! Haha Im sure it will be super fun playing this game together My friends and i are addicted to it I am ditching dota already Like all of us are ditching dota We are only playing amongst us I am already earning big bucks. Dont need the additional stress of keeping up my image both online and offline Not easy you know Fake people are disgusting My worst enemy
96	Wow our make up are from there? Surprise surprise 🤔 Why are you surprised Dumb of you not to know that???



Disclaimer: None of these messages were truly exchanged.

Messages were generated for testing purposes only.



# HOW HAVE OUR MODEL FARED IN PREDICTING EMOTIONS OF TEXTS?

## EXAMPLE: NEGATIVITY

		msg
65		Your boss has been giving shit to you lately isn't it
24		And how I need to see a doctor all the time
30		So pissed off.
104	sure. now it feels like i am being interviewed or interrogated by you. I have been a systems engineer for 4 years before moving into a more administrative role	
51	Let's surprise our friends??? I'm so so so looking forward to our gathering next week	
108	I mean not easy to get the job but once I am a pilot my life is great No la we are always vigilant everytime we fly Autopilot is just a saying Work hours are long man I fly to New York 20 hours you knw I have to be awake for at least 15 hours during these flightd Yes I have been	
37	Damn You rich kid I can only afford to eat at ananas cafe you knw I brought my mum to hawker centre so I can treat her to a mesl	
103	We should talk about more neutral conversations Perhaps you tell me what specially you are doing at your job?	
7	Miss boyfriend ah Hahahahahha	
68	Shouldn't work in this kind of environment Not good for your mental health and wellbeing	



Disclaimer: None of these messages were truly exchanged.

Messages were generated for testing purposes only.

# HOW HAVE OUR MODEL FARED IN PREDICTING EMOTIONS OF TEXTS?

## EXAMPLE: CONTEMPT

	msg
26	I am angry Pissed off at how you are so capable unlike me A useless chap
95	Btw do you know about mining mica products in a rural part of india? Our make up products are from there Very very lethal and deathly Incredibly horrifying Child abuse and stuff I feel like im trying to quarrel with myself I HATE work I hate life I hate everyone! <del>!!!</del> everything!
28	Angry angry angry Sleeping whole day
25	wow don't be upset dont be angry
18	Super lazy and unsuccessful Really hate it that she is so so sucky in what she does and lazes around all day
14	Awesome will always know that you're here to support me Anyways I really hate my dog My dog has the worst sucky shit ever He poops everywhere around the house And doesn't pee at the right placr He pees even on my bed and sofa CAN YOU BELIEVE IT?????!!!! Stupid dog! Good for you This fake conversation is not easy to sustain What should we talk about How is your job
56	Why? Your temper sucks man I really cannot stand it It is pissing people off <del>Amf</del> hell
54	Lol You coward So annoying man Anyhow also get angry
29	why are you angry? yearh tweeter is giving me a headache! i am currently scraping data from the internet and it is NOT easy at all

\*@&(#@\*!#

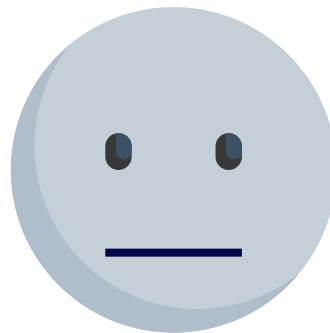


Disclaimer: None of these messages were truly exchanged.

Messages were generated for testing purposes only.

# Final Evaluation

## QUANTITATIVE STANDPOINT:

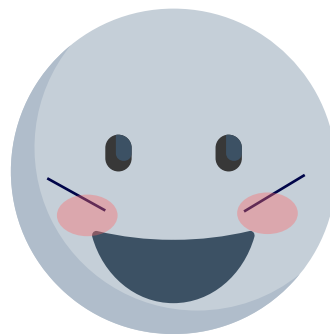


NEUTRAL

Accuracy score only  
51%... sigh



## QUALITATIVE STANDPOINT:



OPTIMISTIC!

Hey! If you look  
at the predicted  
emotions of your  
chat messages,  
they're not too  
bad after all!



# **LIMITATIONS & NEXT STEPS**

## **LIMITATIONS**

1. Lack of open-source chat conversations.
  - Costly to obtain emotion labels for chat conversations even if conversations are available.
2. Use of original tweet data that exhibit degrees of mislabeling by manual annotators that skew our dataset.
3. Insufficient data for certain emotion classes (ie. Contempt, Anticipation)
4. General NLP difficulties: Dealing with Sarcasm, Stretched words (ie. Tireeeeeed), dual meanings, spelling mistakes and more.

## **NEXT STEPS:**

1. Passover prototype to Software Engineering team for initial implementation
2. Conduct the experiment for a second time with newly scraped tweets (by hashtags), ensuring balanced dataset.
  1. Account for key spelling mistakes, stretched words, commonly misspelled words (ie 'haaahaa').
  2. Explore more emotion classes (ie. Disgust, Horror) as we scrape more data from Twitter.

# TESTING TIME!

Input Text:

Text:

Hey a|

Censoring previous search results...

Hey all, how have you been?