

Twitter Sentiment and Intrinsic Attention with Fastai

Blake Samaha

TABLE OF CONTENTS



- 01 THE DATA**
Overview and Insights
- 02 METHODOLOGY**
Neural Nets and NLP with Fastai
- 03 RESULTS**
Model Performance and Intrinsic Attention
- 04 CONCLUSIONS**
Takeaways and Future Work

INTRODUCTION - Sentiment Understanding

- Customer Service
 - Brand Monitoring
 - Product Analysis
- Legal Documents
 - So many documents so little time to read
- Voice of your Employees
 - Understand employees' needs
- Natural Language Processing (NLP) Market to Exhibit 32.4% CAGR; Increasing Technological Advancement to Drive Growth: Fortune Business Insights™





01

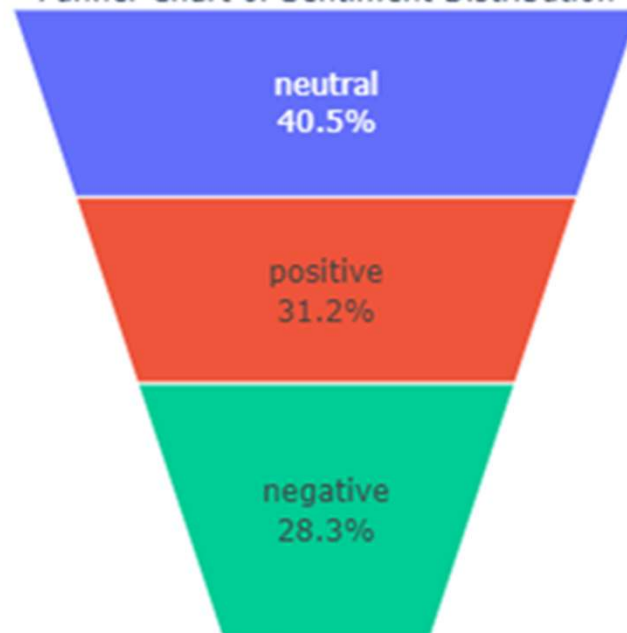
THE DATA

Overview and Insights

Distribution of Data

Data set containing
~30k Tweets

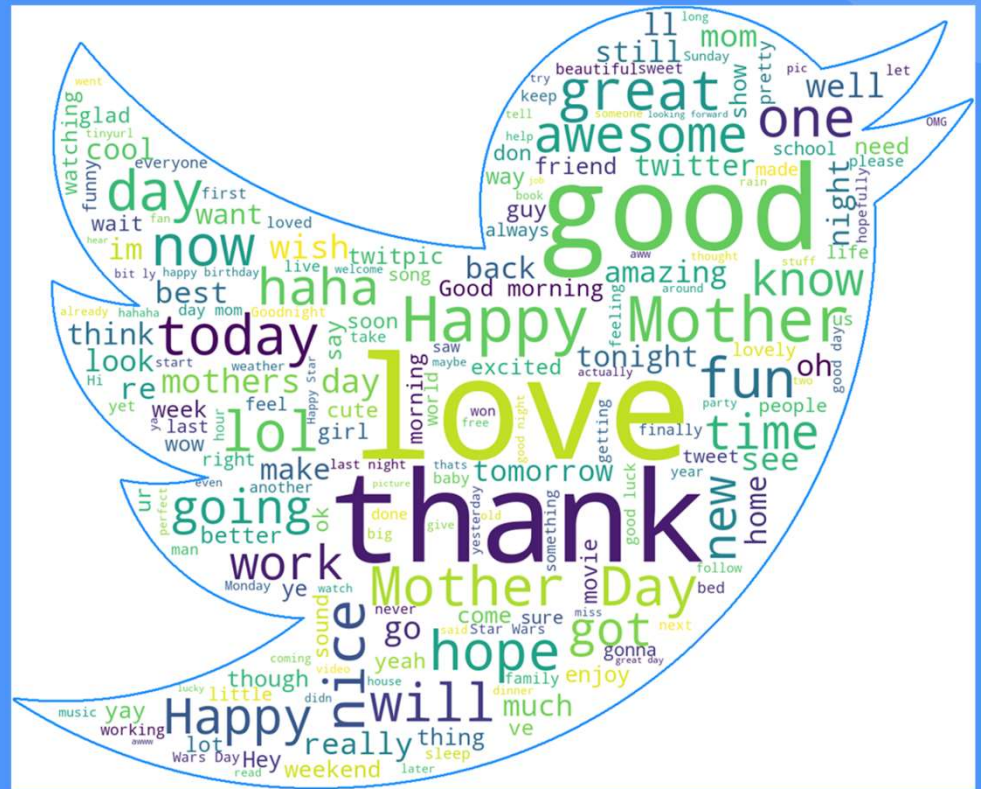
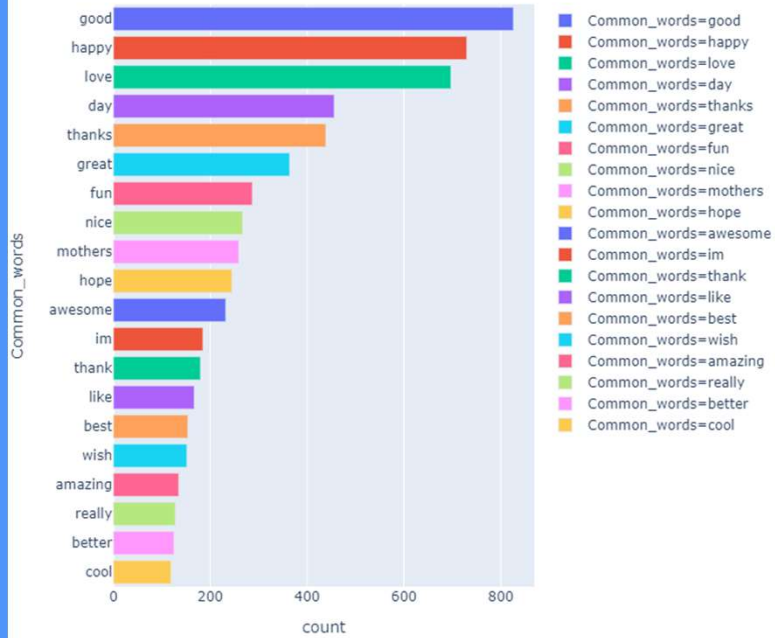
Funnel-Chart of Sentiment Distribution



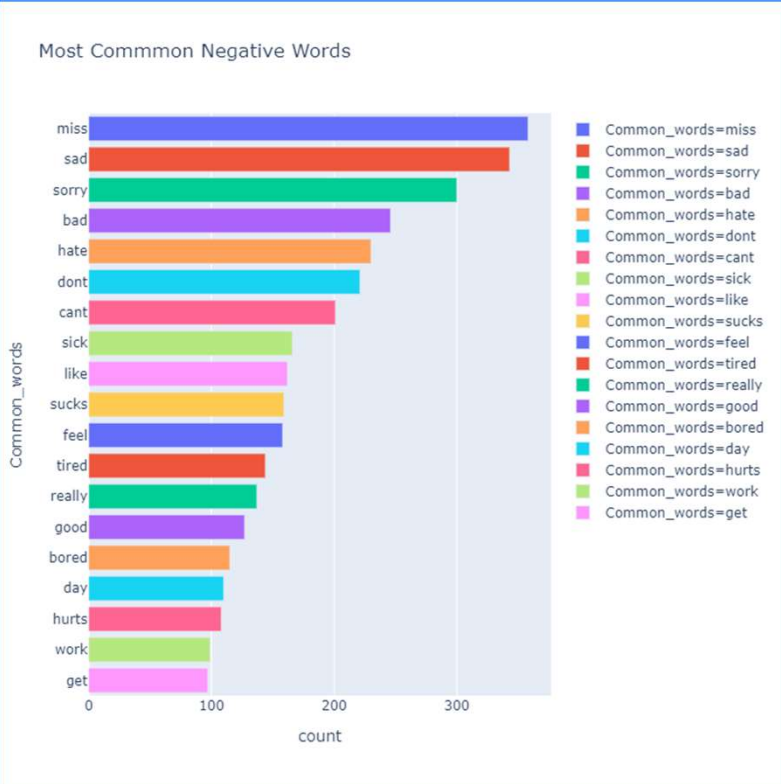
Common words associated with positive tweets

Most Common Positive Words

Common_words	count
good	820
happy	750
love	700
day	450
thanks	430
great	380
fun	280
nice	260
mothers	250
hope	230
awesome	220
im	180
thank	180
like	170
best	160
wish	150
amazing	140
really	130
better	120
cool	110

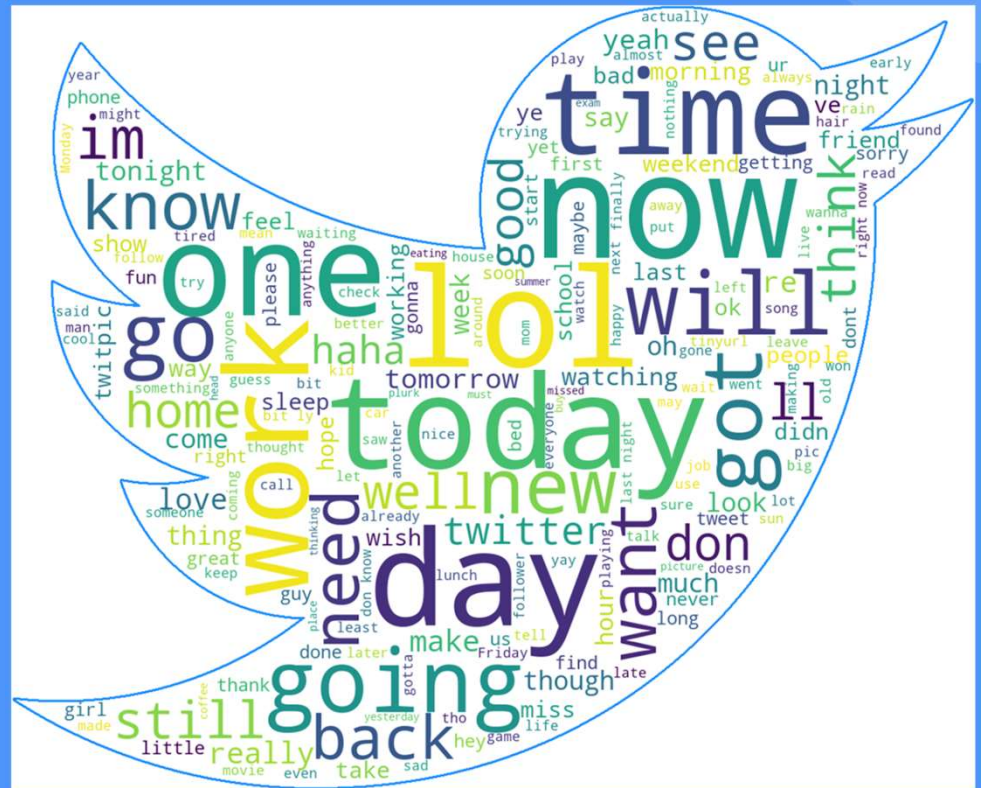
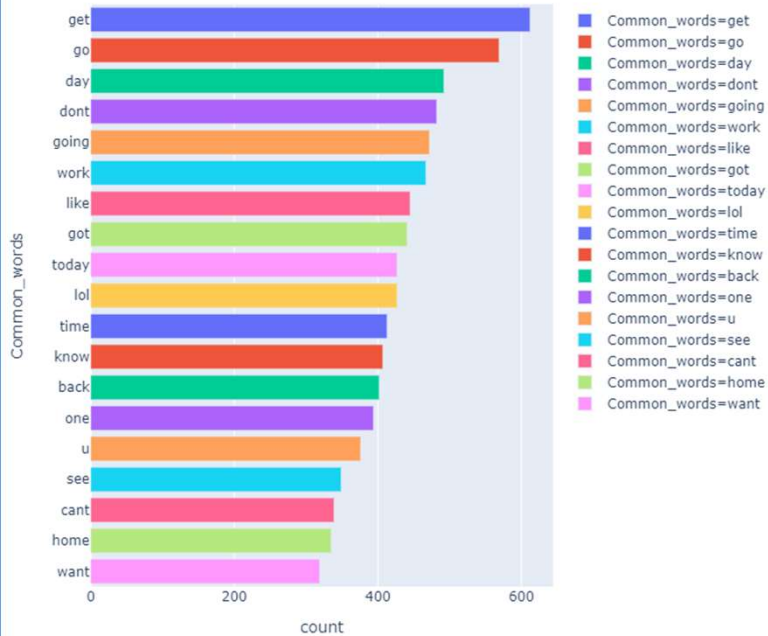


Common words associated with negative tweets



Common words associated with neutral tweets

Most Common Neutral Words



A large blue Twitter bird logo is centered on the slide. The number '02' is written in white inside the bird's wing.

02

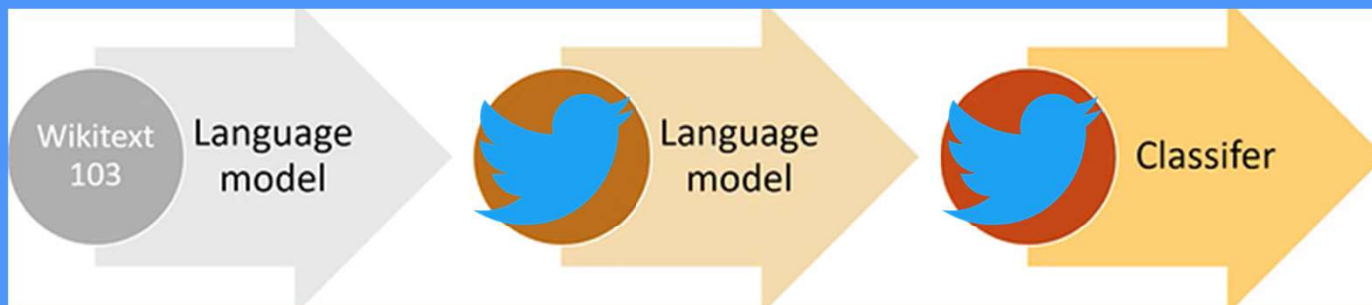
METHODOLOGY

Neural Nets and NLP with
Fastai

METHODS



ULMFiT - Universal Language Model Fine Tuning



Our Tweet based language model

```
[ ] learn.predict('I really', n_words=7)  
👤 'I really want to go see hannah montana'
```

Intrinsic Attention - What is it?

- Essentially a ranking of words based upon significance to its classification
- Allows quick, at a glance, identification of keywords



Our Kaggle Dilemma and Intrinsic Attention

	textID	text	selected_text	sentiment
0	cb774db0d1	I'd have responded, if I were going	I'd have responded, if I were going	neutral
1	549e992a42	» Sooo SAD I will miss you here in San Diego!!!	<u>Sooo SAD</u> «	negative
2	088c60f138	my boss is bullying me...	bullying me	negative
3	9642c003ef	what interview! leave me alone	leave me alone	negative

Fastai's built-in Intrinsic Attention method

```
[ ] text_interp.show_intrinsic_attention(train_df['text'][1])
```

```
☞ xxbos xxmaj sooo xxup sad i will miss you here in xxmaj san xxmaj diego ! ! !  
time: 235 ms
```

Obtaining Selected Text using Intrinsic Attention

	textID	text	selected_text	sentiment
0	cb774db0d1	I'd have responded, if I were going	I'd have responded, if I were going	neutral
1	549e992a42	Sooo SAD I will miss you here in San Diego!!!	Sooo SAD	negative
2	088c60f138	my boss is bullying me...	bullying me	negative
3	9642c003ef	what interview! leave me alone	leave me alone	negative

```
selected_text_grabber(text_interp.intrinsic_attention(train_df['text'][1]))
```

```
'sooo sad' time: 229 ms
```

A large blue Twitter bird logo is centered on the slide. The number '03' is written in white inside the bird's wing. The word 'RESULTS' is written in white inside the bird's body. Below 'RESULTS', the text 'Model Performance and Intrinsic Attention' is written in a smaller white font. There are two light blue decorative shapes: one in the top-left corner and one in the bottom-right corner, both consisting of a quarter-circle and a rectangular section.

03

RESULTS

Model Performance and
Intrinsic Attention

Model Performance

Average Overall
Accuracy
~76%

Confusion matrix

Actual \ Predicted	negative	neutral	positive
negative	1217	278	48
neutral	379	1630	235
positive	80	256	1373

Intrinsic Attention

Desired Output

	textID	text	selected_text	sentiment
0	cb774db0d1	I'd have responded, if I were going	I'd have responded, if I were going	neutral
1	549e992a42	Sooo SAD I will miss you here in San Diego!!!	Sooo SAD	negative
2	088c60f138	my boss is bullying me...	bullying me	negative
3	9642c003ef	what interview! leave me alone	leave me alone	negative

Our Output

	textID	text	selected_text	sentiment
0	cb774db0d1	I'd have responded, if I were going	I'd have responded, if I were going	neutral
1	549e992a42	Sooo SAD I will miss you here in San Diego!!!	sooo sad	negative
2	088c60f138	my boss is bullying me...	my boss	negative
3	9642c003ef	what interview! leave me alone	leave	negative



04

CONCLUSIONS

Takeaways and Future
Work

Takeaways

Fastai

- Powerful and efficient library built on pytorch, but takes some time to understand the inner-workings.
- Highly recommend the fast.ai course!

Twitter

- Unique dialect
 - Character limit
 - Slang
 - Typos
- A lot of non alphanumeric characters
 - URLs

FUTURE WORK

- Improvements to Vocab creation
 - Experiment with stop words...etc
- Experiment with different Tokenizers
 - BERT, ROBERTA, GPT-2...etc
- Refinement of our “Selected Text” function
 - Words in different parts of the sentence may have similar weights
- Attempting additional tuning parameters or starting from a different pretrained model
- Explore alternative methods to intrinsic attention and compare results



