



Sarcasm Detection

Diana Baumgartner-Zhang, Amy Tzu-Yu Chen, David Roesler, Juliana McCausland

Background

- Sarcasm: “the use of remarks that clearly mean the opposite of what they say”
- The nature of sarcasm makes it inherently difficult to detect
 - Negatively impacts the performance of sentiment classification models (which are widely used in industry)





Task: iSarcasmEval

- Binary classification: Given a text, determine whether or not it is sarcastic
- Previous work relies on distant supervision (using hashtags like #sarcasm) which can lead to noisy data, or manual labeling, which brings up the question of annotator agreement
- This task uses a dataset with annotations by the tweet authors themselves
- English and Arabic datasets

tweet	sarcastic	rephrase	sarcasm	irony	satire	understatement	overstatement	rhetorical_question
@WalesOnline Riveting news.	1	What a boring tweet.	1.0	0.0	0.0	0.0	0.0	0.0
My (extended) fam was discussing going on a trip instead of getting presents for Christmas, like okay I'm bout it... Yeah they were talking about doing to the dells. I'll take my god damn presents thank you very much	1	Everyone needs gas-x after having beans.	1.0	0.0	0.0	0.0	0.0	0.0
okay but like the say so song ain't that bad. I'm unironically jam to it lmao	1	The say so song was good after the first 5 listens but after that it got annoying	0.0	1.0	0.0	1.0	0.0	0.0
if you see me crying in the self-service car wash in my rosati's uniform, no you didn't ❤️	1	No.	1.0	0.0	0.0	0.0	0.0	0.0
Tinnitus is my favourite thing ever I love it. So loud so atmospheric https://t.co/mNJw5GvPzy	1	My tinnitus is so annoying	1.0	0.0	0.0	0.0	0.0	0.0
thanks @SovietWomble for ratting my location out on stream and then embarrassing me 😂 it made my day though https://t.co/3b7Cjql2GO	1	It wasn't very nice to rat out my location	1.0	0.0	0.0	0.0	0.0	0.0
love living in a capitalist society where i look forward to getting a surgery where i'm legit GETTING AN INTERNAL BODY PART REMOVED bc it means i'll get a few days off from work xoxo	1	I hate that in order to get time off of work, I have to do something drastic like getting surgery	1.0	0.0	0.0	0.0	0.0	0.0
i love ingesting poison and then wondering why my tummy is rumbling the next morning	1	Every time I drink I wake up with an upset stomach	1.0	0.0	0.0	0.0	0.0	0.0
Beautician and the Beast (1997). Best Timothy Dalton film of all time.	1	I would say what I said above: Beautician and the Beast isn't Timothy Dalton's best film.	1.0	0.0	1.0	0.0	0.0	0.0
@FeeyaCruz Ahahahaha the Feeya I know and love, miss u 😭	1	'Oh Feeya, you haven't changed"	1.0	0.0	0.0	0.0	0.0	0.0
absolutely love waking up to the fire alarm at 7 am 🚒	1	I hate waking up to the fire alarm at 7 am.	0.0	1.0	0.0	0.0	0.0	0.0
fully vaxxed feeling very girlboss	1	fully vaxxed not feeling very girlboss	1.0	0.0	0.0	0.0	0.0	0.0

id	tweet	sarcastic	rephrase	dialect
1	ضبط شخص بدلولم انتحل صفة طبيب بلد مافيش حد فيها بيشغل بشهادته	1	شخص ينتحل صفة طبيب ويفتتح عيادة فى بلد فاشلة ضاعت فيها الرقابة	nile
2	مش معنى انك قولتلى رايك يبقى أنا هعمل بيه طب افرض انت حمار استحمر أنا كما زيك	1	مش لازم دائما اعمل برأيك	nile
3	اية المهلبية دي يصحبي	1	ما هذا الجمال	nile
4	الحديث قياس فيه الفضة و فيه النحاس	1	لسانك ترجمان قلبك	magreb
5	ده فاكسر نفسه باشا و بيه كمان	1	ده مغرور و شايف نفسه علاناس	nile
6	حامل في السادس وماكتتش ملاحظ ده قبل الجواز طب وبالنسبة لليلة الدخلة	1	انت مغفل وتزوجت امرأة حامل ولم تلاحظ	nile
7	الدنيا عاملة زى اوضة وصالة كل ما امشى فى حته الاقى ايمن فى وشى	1	الدنيا صغيرة ودايما نلاقى فيها حد نعرفه	nile
8	مهما دورت ولقيت عمرك ما هتلاقى حد يحس بيك غير كركشك	1	محدث هينفعك غير بطنك انما اى حد تاني مش هيحس بيك	nile
9	فلوس ايه يا امى اللى احوشها من شغلى ده أنا مافيش فلوس اروح الشغل بكرة	1	مرتبى قليل ومش معايا فلوس حتى المواصلات للشغل	nile
10	وراء كل رجل عظيم امراه تقول له انت لست عظيم	1	المراه تحاول احباط زوجها وهو بمنطلق العند يحاول ان يكون عظيما	msa
11	نفسى افهم مين الي نزل الكائن ده كتالوج الحريم	1	دى وحشة أوى علشان تبقي فى المجلات ومشهورة	nile
12	سنة أغسطس بيكون فيها فجوة زمنية مجهولة الأيام بتبقى تصاعدي ليوم 15 وترجع تنازلي ثاني ليوم 15	1	شهر اغسطس بيدو طويلا جدا	nile
13	للصوص مفيش اتعس منهم الناس قاعدة في البيوت ولو حيو يتويوا فالمساجد مقفولة	1	الكرونا وقفت حال اللصوص	nile
14	اللي تعرف ديه اقتله	1	حاول ان تجد نقاط ضعف خصمك	nile
15	مبروك كسبت جايزة اويل في الذكاء	1	انت غبي	nile
16	هيورينا ايام اسود من قرن الخروب	1	هيورينا ايام صعبة وشديدة	nile
17	"🤔" فينجر طلب منى الخضوع لاختبار قبل ضمي للأرسنال، فى البداية وافقت ثم أجبت كما يلي زلاتان لا يقوم بالاختبارات	1	مدرب فريق ارسنال يطلب من اللاعب ابراهيموفيتش الدخول فى اختبارات قبل الانضمام لكنه يرفض	msa
18	فيه ٣ في القاهرة حابسين نفسهم في بيوتهم حد يقولهم ينزلوا علشان نكمل لمتنا	1	القاهرة زحمة ومفيش اهتمام بالعزل المنزلي	nile
19	لي قرا قرا بكري	1	خلص وقت الدراسة	magreb



Baseline Models

Preprocessing:

- TfidfVectorizer with ngram_range set to (2, 6)

Models:

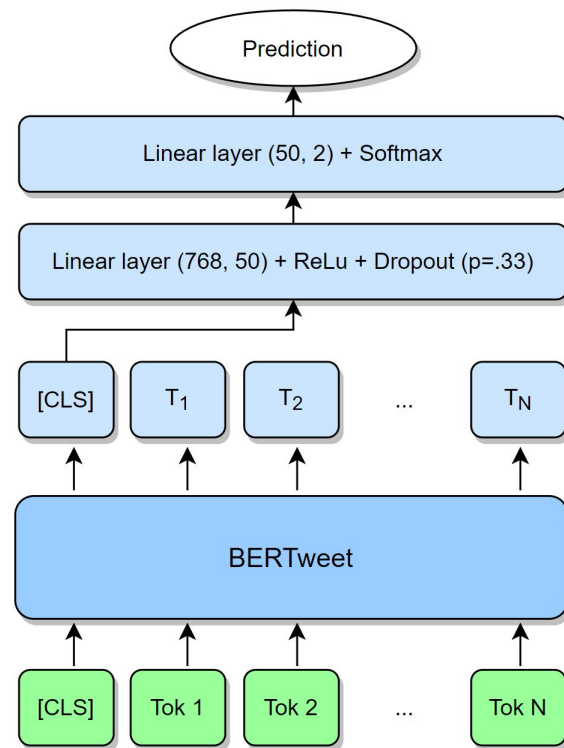
1. RandomForest classifier
2. Light Gradient Boosting Machine (LightGBM) classifier

Improved Model: System Architecture

BERTweet language model

Classifier head attached to BERTweet base model

- Takes CLS token representation output from BERTweet encoder and makes a prediction



Approach

Tweets contain informal grammar, irregular vocabulary, emojis

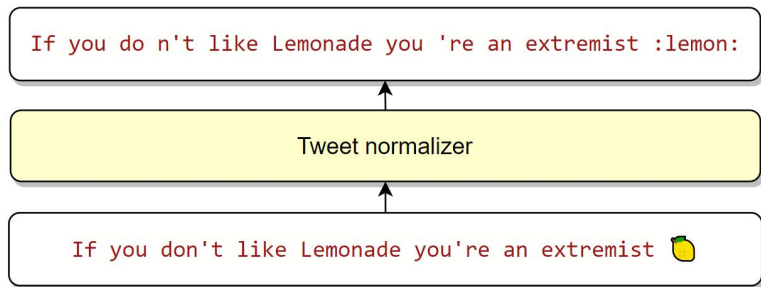
- BERTweet ideal for dealing with this type of data (trained on 850M tweets)
- Preserved emoji info with normalization

Balanced classes

- Included all positive samples and random sample of equal number of negative samples

PyTorch and HuggingFace for fine-tuning

Evaluated using F1 score



 PyTorch





Results

Model	Accuracy	Precision	Recall	F1	Macro F1
Random forest	.53	.53	.66	.59	0.53
LightGBM	.53	.52	.76	.62	0.50
BERT-base (cased)	.58	.57	.58	.58	.58
RoBERTa-base	.69	.69	.67	.68	.69
BERTweet-base	.73	.76	.69	.72	.73
iSarcasmEval top F1 (test set)				.61	



Results

Biased	Loss1: Without Weights		Loss2: W1=1/#NS, W2=1/#S	
	Val	Test	Val	Test
B0	0.5784	0.4487	0.5944	0.4519
B1	0.5714	0.4548	0.5738	0.479
B2	0.5767	0.465	0.5951	0.4917
B3	0.601	0.4626	0.5931	0.4817
B4	0.5954	0.4727	0.6025	0.4769
B5	0.5874	0.5142	0.5803	0.5008
B6	0.5858	0.4791	0.5624	0.4884
B7	0.5957	0.5016	0.5637	0.5034
B8	0.584	0.492	0.5814	0.5
B9	0.5723	0.487	0.5554	0.49

Table 7: Results of the BERT model for subtask A (English) on external datasets.



Problems

- Environments
 - Different environments produce different results
- Slowness of patas/dryas :)
- Finding language model
 - BERT, RoBERTa less successful with tweet data

Problems: Noise in Training Set

Tweets with both positive and negative labels:

tweet	sarcastic	rephrase
Whoever's toddler ass sprayed the entire toile...	1	I hope whoever keeps peeing all over the bathr...
@AsdaServiceTeam imagine your delivery being 2...	1	It's not acceptable for you to just refund my ...
@AlexShawESPN @robwishart @GNev2 Does he order...	1	I don't think he drank 11 beers every hour
Replace Pelosi #Nancy	1	He should not have been elected president and ...
if i saw a capybara in person id probably thro...	1	I think capybaras are scary
Whoever's toddler ass sprayed the entire toile...	0	NaN
@kevinabstract ye	0	NaN
@kevinabstract ye	0	NaN
@AsdaServiceTeam imagine your delivery being 2...	0	NaN
@AsdaServiceTeam imagine your delivery being 2...	0	NaN
@AsdaServiceTeam imagine your delivery being 2...	0	NaN
@AlexShawESPN @robwishart @GNev2 Does he order...	0	NaN
going to class! https://t.co/VgCWGI9YTG	0	NaN
going to class! https://t.co/VgCWGI9YTG	0	NaN
Replace Pelosi #Nancy	0	NaN
Great effort @Patriots We'll get em all next year	0	NaN
Great effort @Patriots We'll get em all next year	0	NaN
if i saw a capybara in person id probably thro...	0	NaN
Omg flashbacks to emma blackery girl balls mer...	0	NaN
Omg flashbacks to emma blackery girl balls mer...	0	NaN



Problems: Data Annotations

Tweets annotated as sarcastic:

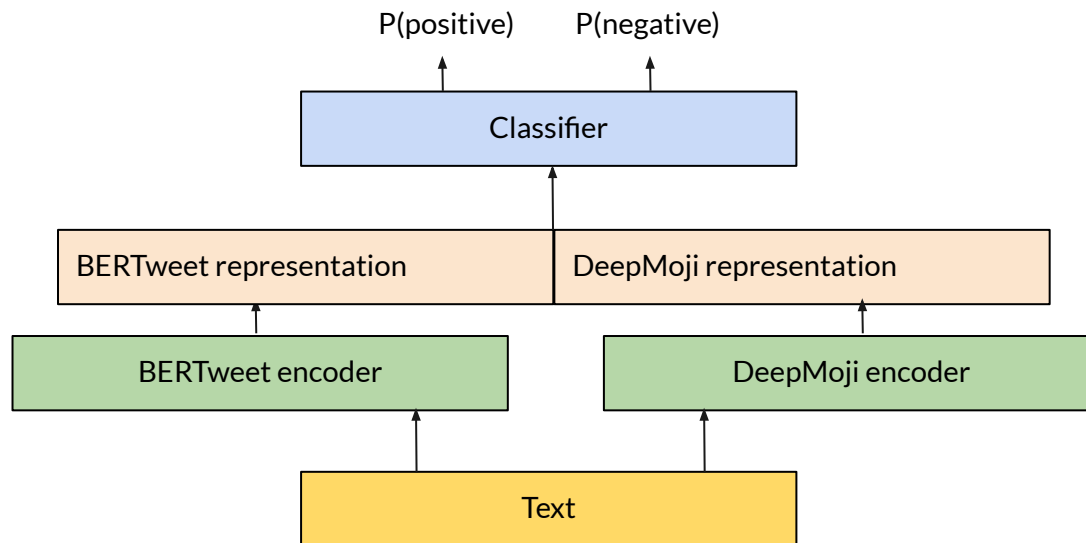
Original Tweet	Rephrase (from tweet author)	Labels
"You like ketchup but not tomatoes???"	It seems childish to enjoy ketchup but not tomatoes	sarcasm
"CLB cover art makes me want to pound my knees with a hammer"	I do not like the certified loverboy artwork	overstatement

Tweets annotated as not sarcastic:

Original Tweet	Our interpretation
"Oh cool i drive to school in less than 12 hours and I'm frantically looking up fun facts about the final destination franchise for absolutely no reason it's fine everything's fine"	Everything is not fine. The author is anxious after reading about final destination right before having to drive long distance
"Ah yes. After a decent birthday party it is now time to stress out about money again."	The 'ah yes' carries a connotation of pleasantness, making this statement seem ironic/sarcastic

Plans for D3:

Fusing BERTweet representations with affective representations:





Plans for Adaptation Task

Arabic dataset

Potential problems and considerations:

- Dialectal variation
- Orthographic inconsistency
- Orthographic ambiguity
- Need to account for these features in pre-processing - CaMeL Tools?
- None of us know Arabic :)



Related Reading

BERTweet:

Nguyen, D. Q., Vu, T., & Tuan Nguyen, A. (2020). BERTweet: A pre-trained language model for English Tweets. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, 9–14.

<https://doi.org/10.18653/v1/2020.emnlp-demos.2>

Sarcasm Disrupting NLP:

Liu, B. (2010). Sentiment analysis and subjectivity. Handbook of Natural Language Processing, 2(2010), 627–666.

DeepMoji:

Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 1615–1625.

4th place iSarcasmEval2022 Submission:

Nigam, S. K., & Shaheen, M. (2022). Plumeria at SemEval-2022 Task 6: Robust Approaches for Sarcasm Detection for English and Arabic Using Transformers and Data Augmentation. ArXiv:2203.04111 [Cs]. <http://arxiv.org/abs/2203.04111>