



# Sarcasm Detection

Group 8

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



## Task Description

- SemEval 2022 - iSarcasmEval: Intended Sarcasm Detection In English and Arabic
- SubTask A: Given a text, determine whether it is sarcastic or non-sarcastic;
- Sarcasm: a form of verbal irony when there is a discrepancy between the literal and intended meanings
- Primary task: English
- Adaptation task: Arabic

# Task Data Review - English

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 <a href="https://t.co/mNJw5GvPzy">https://t.co/mNJw5GvPzy</a>	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 <a href="https://t.co/3b7Cjql2GO">https://t.co/3b7Cjql2GO</a>	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

# Task Data Review - Arabic

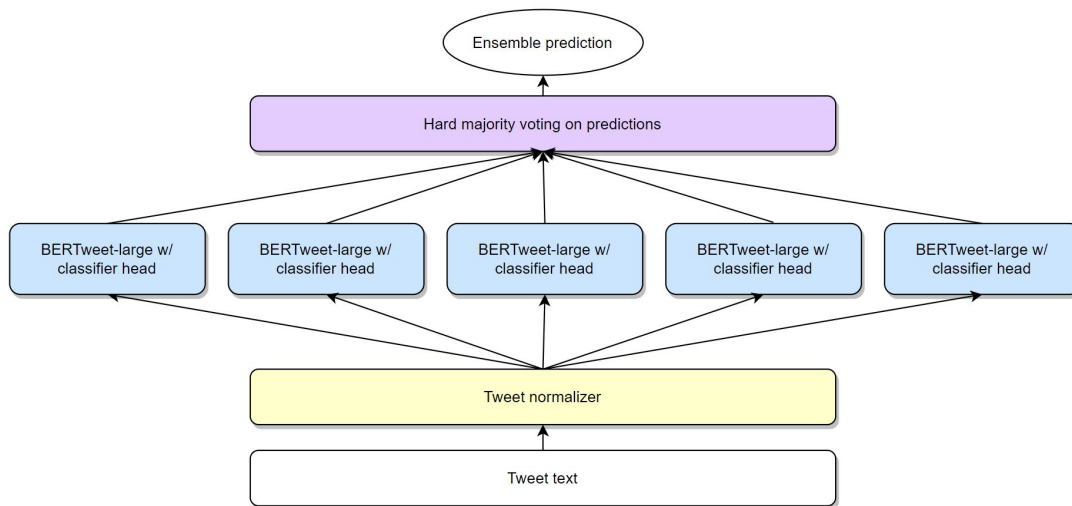
id	tweet	sarcastic	rephrase	dialect
1	ضبط شخص بدلولم انتحل صفة طبيب بلد مافيش حد فيها بيشغل بشهادته	1	شخص ينتحل صفة طبيب ويفتتح عيادة فى بلد فاشلة ضاعت فيها الرقابة	nile
2	مش معنى انك قولتلى رايك يبقى أنا هعمل بيه طب افرض انت حمار استحمر أنا كما زيك	1	مش لازم دائما اعمل برأيك	nile
3	اية المهلبية دي يصحبي	1	ما هذا الجمال	nile
4	الحديث قياس فيه الفضة و فيه النحاس	1	لسانك ترجمان قلبك	magreb
5	ده فاكدر 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 الحرير	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 حيوا 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 الزكاء	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



## Revised Approach on Primary Task

- We change the encoder component of our system from a BERTweet-base(768-dimensional outputs) to a BERTweet-large model(1024-dimensional outputs) . The two BERTweet models have the same structure
- Threshold moving (Provost & Fawcett, 2001): Select the probability threshold for class decision that maximizes the metric of interest (positive class F1 in our case)
- Augment training data w/ tweets from the English sarcastic tweet dataset produced by Ptáček et al. (2014)
- Tried to reuse LightGBM model (baseline) to be a part of ensemble, but did not include it in the final system
- Ensemble of 5 BERTweet-large models trained on original training data + random sampling of augment data, with hard voting to determine final predictions.

# Revised Approach on Primary Task





## Results - Primary Task

Evaluation set	Model	Accuracy	Precision	Recall	F1	Macro F1
dev.	BERTweet-large	.81	.82	.79	.80	.81
dev.	BERTweet-large w/ data augmentation + thresholding	.81	.77	.87	.82	.81
dev.	Ensemble w/ data augmentation + thresholding	.81	.78	.86	.82	.81
test	BERTweet-large	.88	.58	.59	.58	.76
test	BERTweet-large w/ data augmentation + thresholding	.90	.67	.63	<b>.65</b>	.80
test	Ensemble w/ data augmentation + thresholding	.90	.67	.61	.64	.79
test	iSarcasmEval task A best reported result				.61	



# Adaptation Task Approach

- Modifying modeling framework for primary task to perform sentiment analysis on Arabic tweets
- Attempted fine-tuning of two different pre-trained BERT models on Arabic, with similar hyperparameters to the settings of English Bert models:
  - [CAMELBERT-mix](#) (Inoue et al., 2021)
  - [MARBERT](#) (Abdul-Mageed et al., 2021)





## Results - Adaptation Task

Evaluation set	Model	Accuracy	Precision	Recall	F1	Macro F1
dev.	CAMeLBERT-mix	0.86	0.83	0.91	0.86	0.86
dev.	MARBERT	0.86	0.82	0.93	0.87	0.86
test	CAMeLBERT-mix	0.49	0.2	0.88	0.33	0.46
test	MARBERT	0.44	0.2	0.94	0.32	0.42
test	iSarcasmEval task A best reported result				.56	



# Learnings from Issues and Successes

## Issues and Challenges:

- None of us knew Arabic so we could not really figure out what went wrong with adaptation task
- We hit “ceiling” when trying to improve the primary task model

## Successes:

- Reusable modeling framework that allows us to start training new versions of models as well as adaptation task
- Early successes on primary task that allows us to experiment many modeling techniques



## Related Readings

**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 + Emojis:** Filik, R., Turcan, A., Thompson, D., Harvey, N., Davies, H., & Turner, A. (2016). Sarcasm and emoticons: Comprehension and emotional impact. Quarterly Journal of Experimental Psychology, 69(11), 2130–2146. <https://doi.org/10.1080/17470218.2015.1106566>

**Thresholding:** Sheng, V. S., & Ling, C. X. (2006). Thresholding for making classifiers cost-sensitive. Proceedings of the 21st National Conference on Artificial Intelligence - Volume 1, 476–481.

**Augmentation dataset:** Ptáček, T., Habernal, I., & Hong, J. (2014). Sarcasm Detection on Czech and English Twitter. Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, 213–223. <https://aclanthology.org/C14-1022>

**CAMELBERT-mix:** Inoue, G., Alhafni, B., Baimukan, N., Bouamor, H., & Habash, N. (2021). The Interplay of Variant, Size, and Task Type in Arabic Pre-trained Language Models (arXiv:2103.06678). arXiv. <https://doi.org/10.48550/arXiv.2103.06678>

**MARBERT:** Abdul-Mageed, M., Elmadany, A., & Nagoudi, E. M. B. (2021). ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 7088–7105. <https://doi.org/10.18653/v1/2021.acl-long.551>

**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.