



Analyzing Context and User Information in Online Sarcasm Detection

Members: Tusheet Sidharth Goli and Mohan Dodda

College of Computing, Georgia Institute of Technology

Introduction

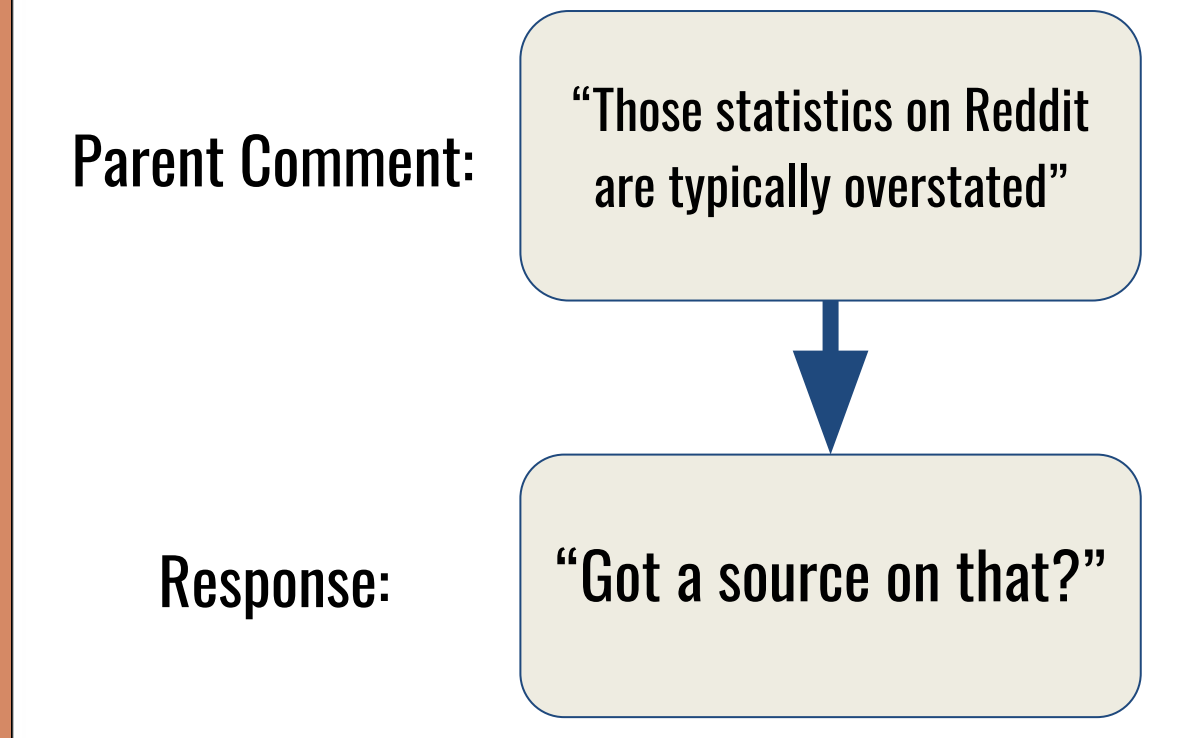
- Project Goal**
- Identify sarcasm in conversations
 - Examine context and circumstances that provoke sarcastic responses
 - Analyze linguistic cues and user attributes that correlate to high sarcastic behaviour
- Motivation**
- Help distinguish between incorrect information and sarcasm
 - Help people prevent sarcastic responses in their comments
 - Sarcasm can give the feeling being left out
 - Sarcastic responses not normally desired (negativity)
 - Prevent borderline cyber-bullying
 - Gain a deeper understanding of human psychology and contribute to linguistic analysis

Data

SARC Reddit Dataset

label	comment	author	subreddit	score	ups	downs	date	created_utc	parent_comment
0	NC and NH.	Trumpbart	politics	2	-1	-1	2016-10-10	2016-10-16 23:55:23	Yeah, I get that argument. At this point, I'd ...
1	You do know west teams play against west teams...	Shbshb906	nba	-4	-1	-1	2016-11-11	2016-11-01 00:24:10	The blazers and Mavericks (The wests 5 and 6 s...
2	They were underdogs earlier today, but since G...	Creepeth	nfl	3	3	0	2016-09-09	2016-09-22 21:45:37	They're favored to win.
3	This meme isn't funny none of the "new york ni...	icebrotha	BlackPeopleTwitter	-8	-1	-1	2016-10-10	2016-10-18 21:03:47	deadass don't kill my buzz
4	I could use one of those tools.	cush2push	MaddenUltimateTeam	6	-1	-1	2016-12-12	2016-12-30 17:00:13	Yep can confirm I saw the tool they use for th...

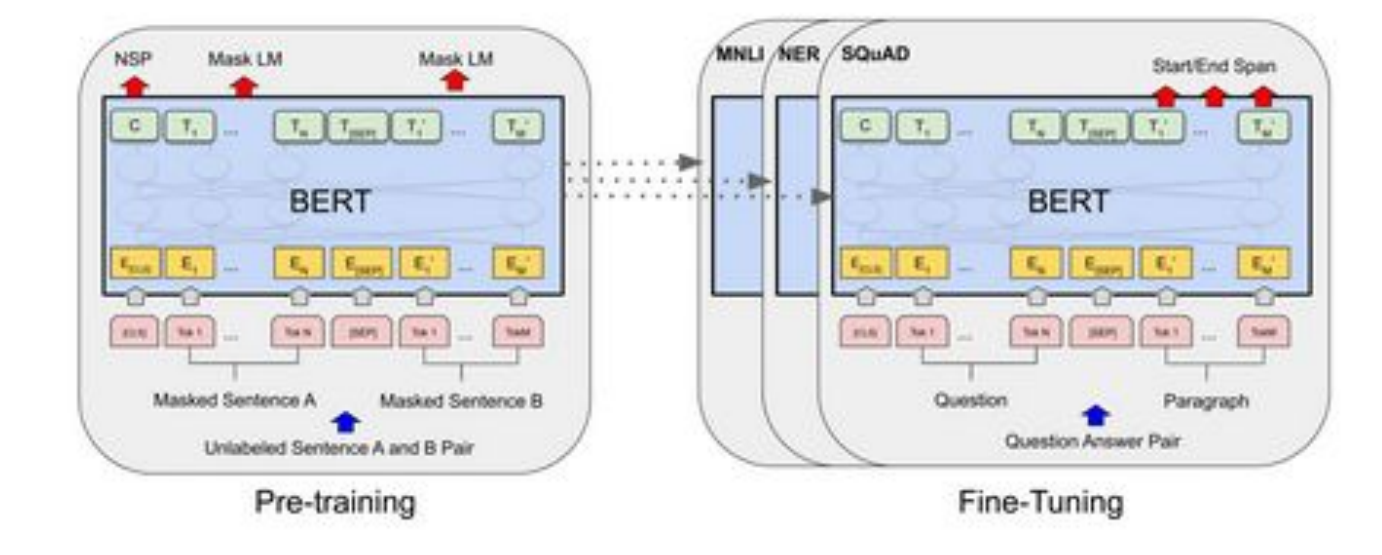
- Contains parent comment and response (which may or may not be sarcastic)



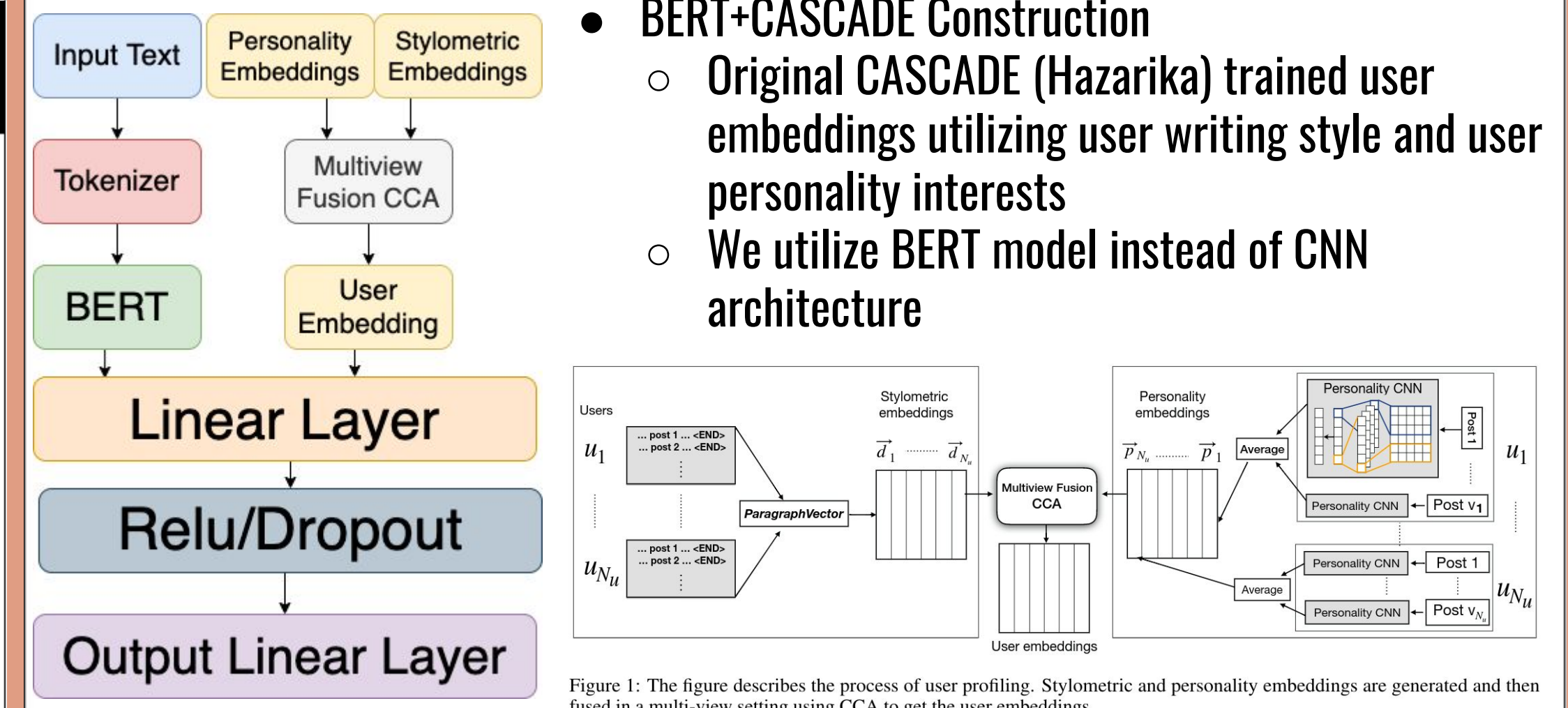
Statistics vs. comment type	Sarcastic	Non-Sarcastic
average response size (words)	10.33	10.59
average parent comment size	24.21	24.56
mean upvotes	5.22	5.78
mean downvotes	-0.13	-0.17
mean score	6.40	7.37
Number of Documents	505368	505405
label	1	0

Methods

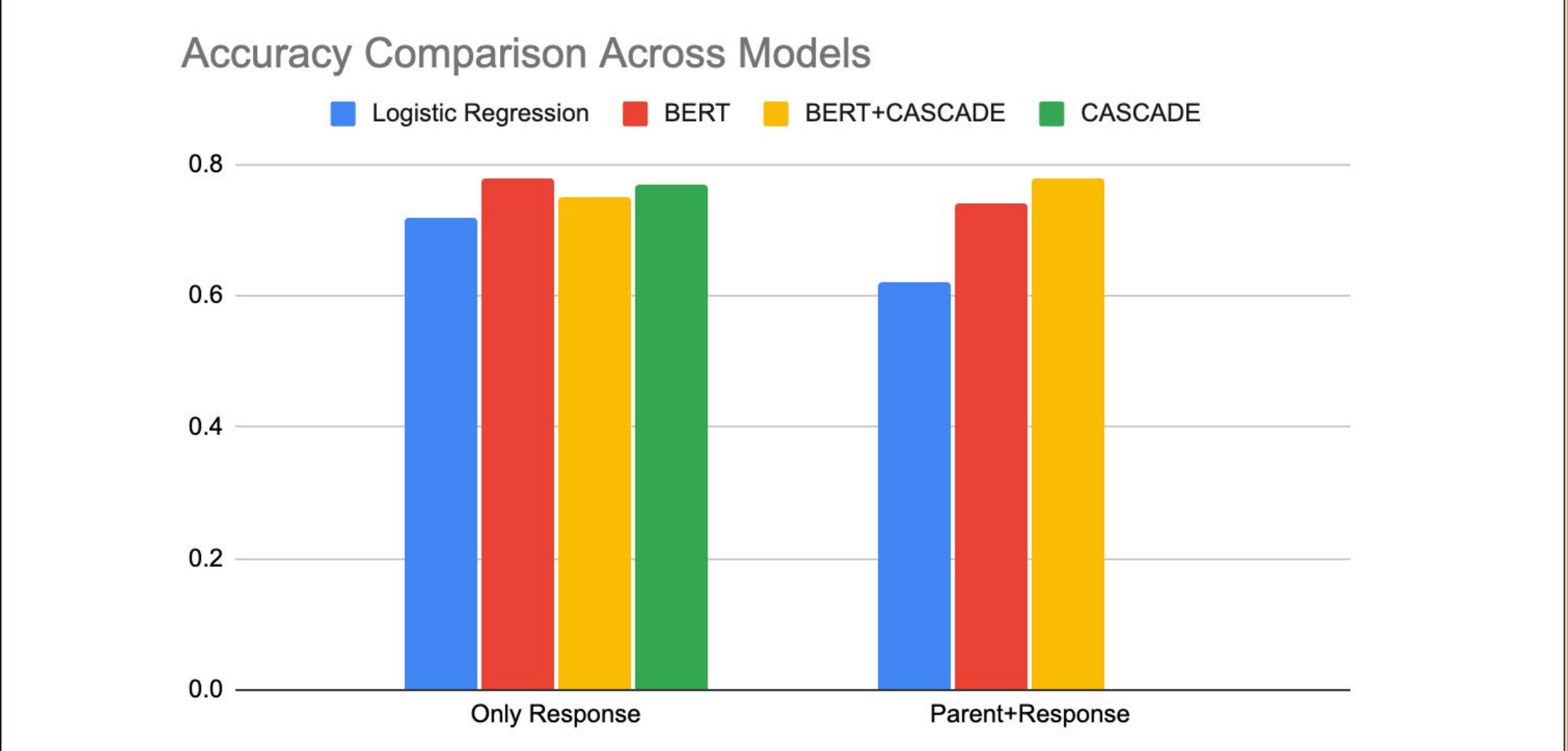
- Logistic Regression (Khodak), fine-tuned BERT model and BERT+CASCADE model on response, parent comments, and parent+response
- Analyzed data further in context and user based features subjectivity, number of special characters, word/sentence length, and profanity and user writing/personality style, subreddit
- Train feature based to determine which features most impact our results



Proposed Architecture - BERT+CASCADE



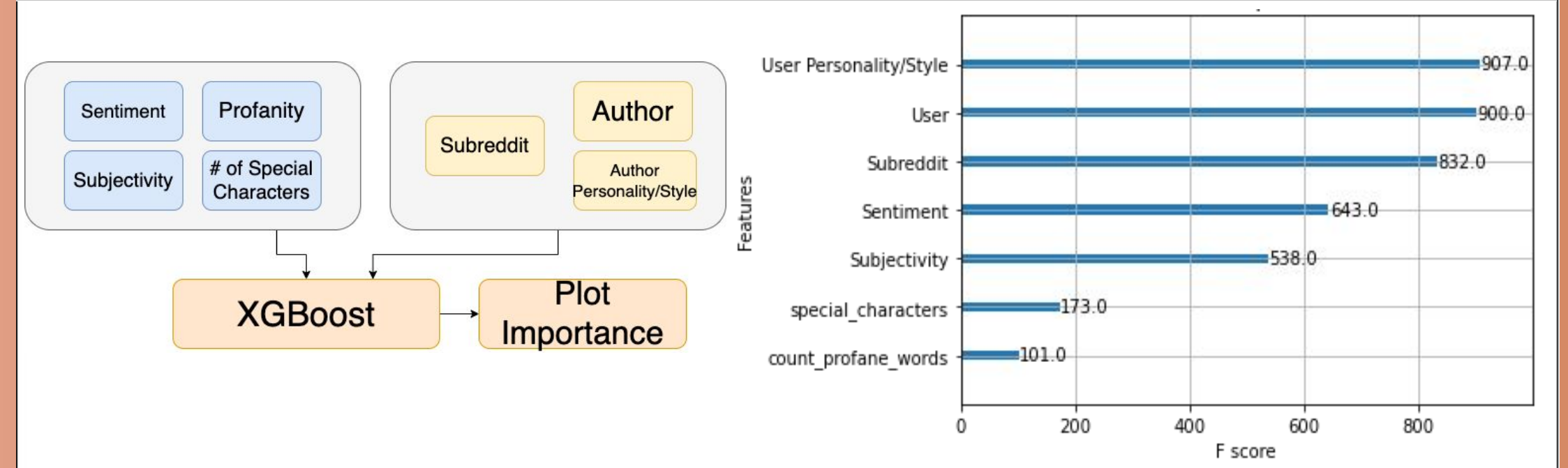
BERT+CASCADE Results



BERT+CASCADE Results Discussion

- BERT+CASCADE produces worse results with only response compared to BERT variant. For parent response, BERT+CASCADE was better than BERT
- User info properly complements parent comment context

Feaure Model Results



- We train a XGBoost model to predict sarcasm utilizing context, background, and user related features
- We find most that the significant features are subreddit and sentimentality
- User history and writing style are major user related features

Conclusion and Future Goals

- Conclusion**
- We showed BERT+CASCADE provides minimal improvement
 - Learning from just response from BERT much higher than user contexts - user contexts act as noise
 - Display various user and context related features that influence sarcasm
 - User personality and Subreddit shows to have most influence in predicting sarcasm more than standard context related features
- Future Goals**
- Train regression models to generate longer dimension features for subjectivity, word/sentence length, to better capture their features
 - Embed context and subreddit features into BERT+CASCADE to have it utilize contextual information.

References

[1] Hazarika, D., Poria, S., Gorantla, S., Cambria, E., Zimmermann, R., and Mihalcea, R. 2018. CASCADE: Contextual sarcasm detection in online discussion forums. In Proceedings of the 27th International Conference on Computational Linguistics, pages 1837— 1848. Association for Computational Linguistics.

[2] Khodak, M., Saunshi, N., Vodrahalli, K. A large self-annotated corpus for sarcasm. In Proceedings of the Linguistic Resource and Evaluation Conference (LREC), 2018.