

# Feature Ablation

- *Feature Set 1* refers to the original feature extraction set.
- *Feature Set 2* is customized for comments on social platforms, including the number of function words, the number of punctuation marks, the type-token ratio, average sentence length, readability score, the uppercase ratio, number of slang terms, number of URLs, whether there is a subject, and the type of the subject.
- *Feature Set 3* was obtained by modifying *Feature Set 2*, but with the number of slang terms and the number of URLs removed. Instead, part of speech n-gram features were added.

**What do the results imply?** A plausible explanation of this trend is that social media markers actually *decrease* performance on topically homogeneous sets but have a negligible positive effect on performance in the case of more heterogeneous sets.