

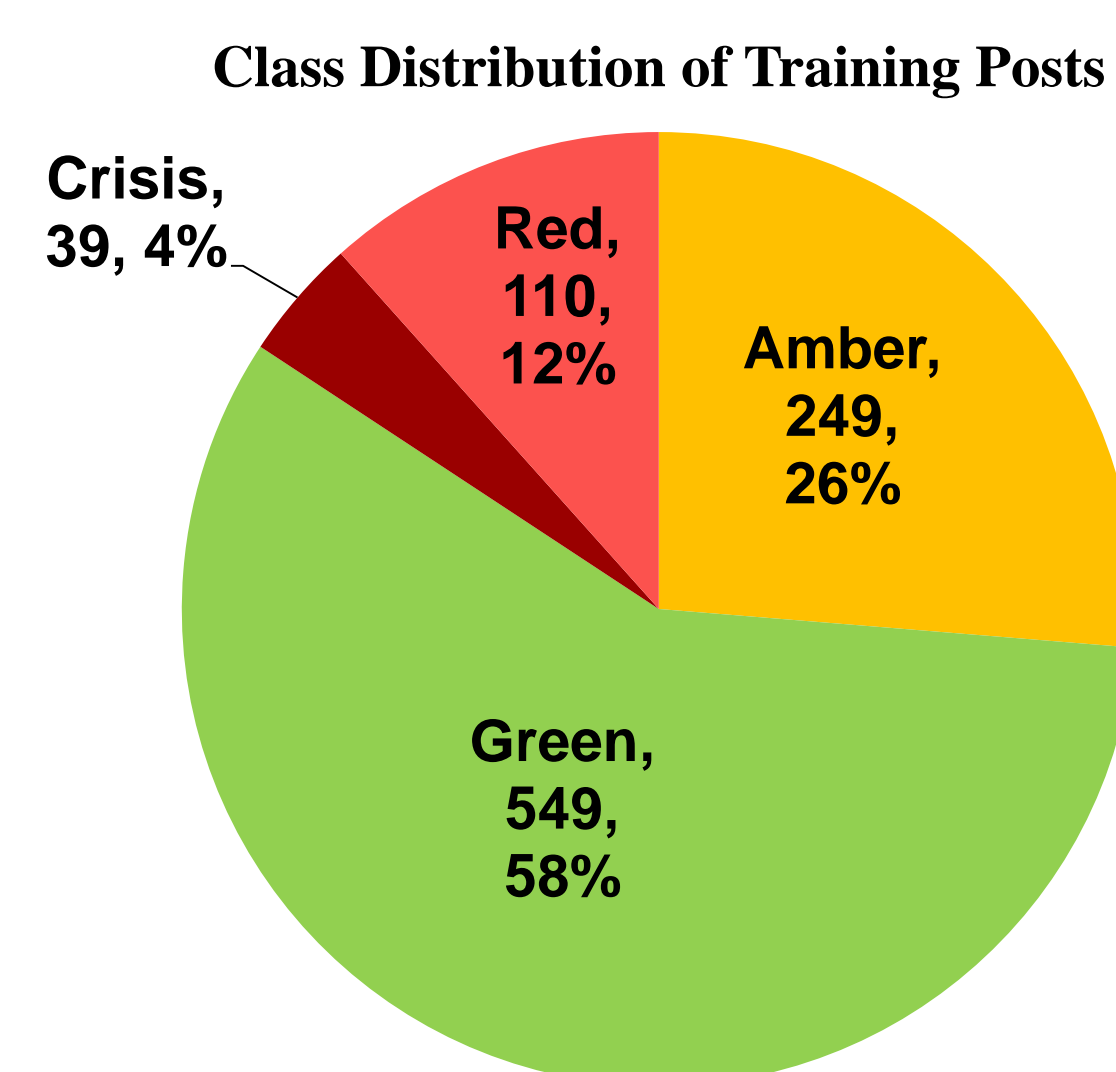
Abstract

- Triaging mental health forum posts helps moderators to focus on individuals who need urgent attention
- Our model is based on a regularized logistic regression classifier with various sets of lexical, psycholinguistic and topic modeling features.
- Our approach identified cases that require moderator attention with a F1 score of over 80%, showing the effectiveness of the proposed model.



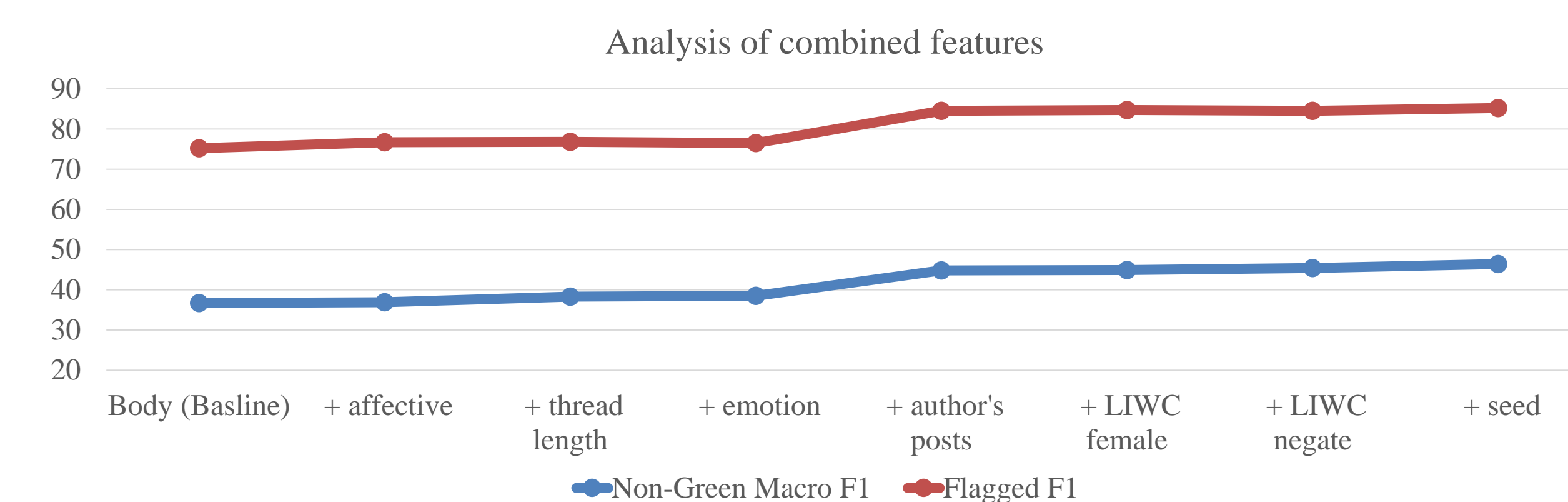
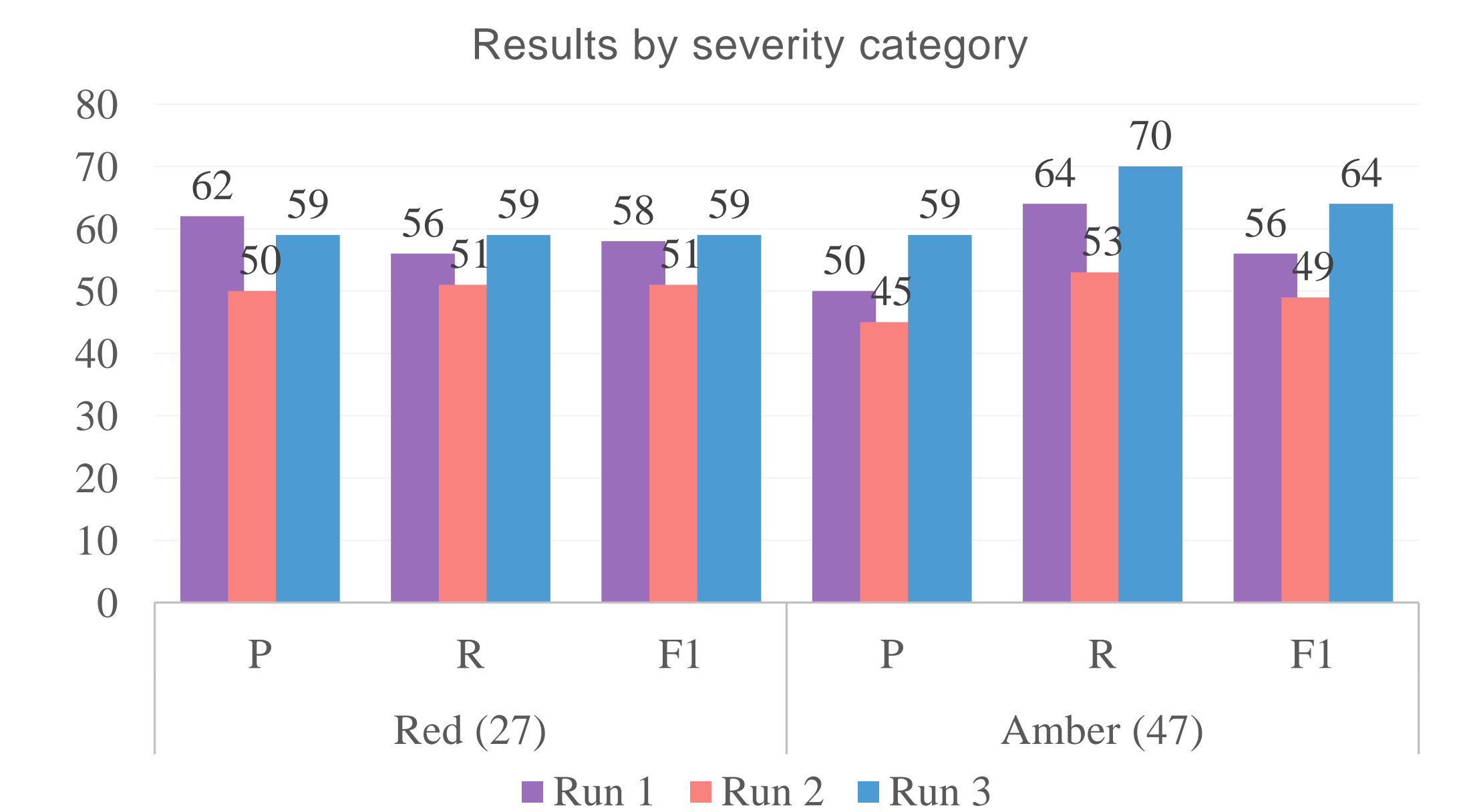
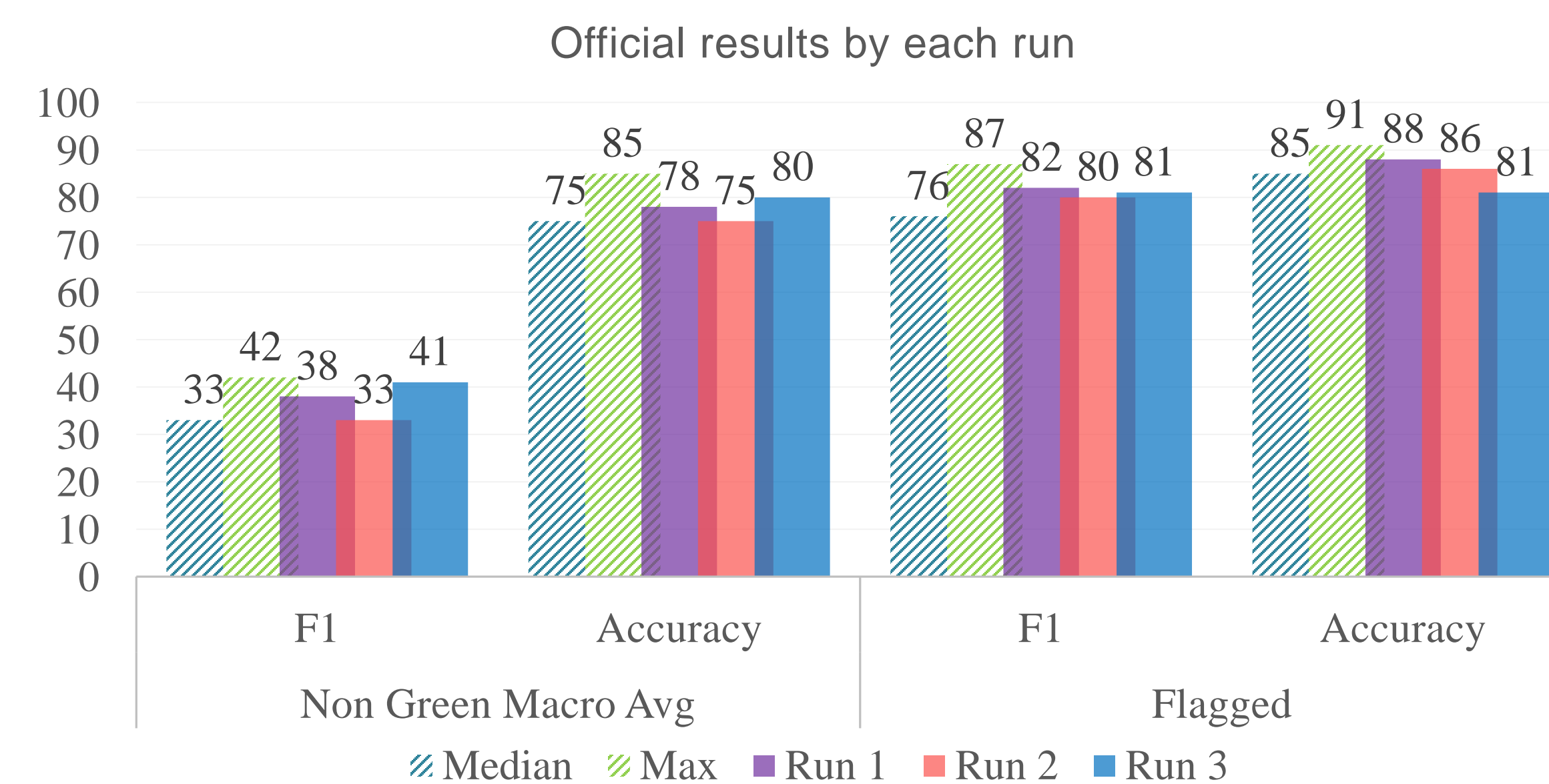
Method

- General approach
 - Supervised multi-class classification
 - Regularized Logistic Regression
 - Boosting prediction probabilities for critical categories
 - Evaluation
 - 4-fold cross validation, 3 submitted runs
 - Baseline: Classifier with Unigram/Bigram features
 - Data: Forum posts from ReachOut.com
 - 1188 annotated posts (947 train, 241 test)
 - 65K+ unlabeled posts
 - Features:
 - Psychological:
 - LIWC¹, DepecheMood², MPQA³, Subjectivity
 - Topic Modeling
 - LDA to identify topics of the posts
 - Linguistic features
 - Terms in the body of the post (n-grams)
 - Contextual
 - Author's prior posts, surrounding posts, subjectivity, LIWC affective features, last sentence to capture last mental state of the user
 - Textual Statistics:
 - Number of posts in the thread, critical seed words



Results

Run	Features	Boost
1	Body, author's posts, subject, emotion, thread length, LIWC (affective, female), and seed terms.	Crisis + 0.2
2	Body, author's posts, emotion, thread length, LIWC (affective, female, negate), and seed terms	Crisis + 0.3 Red + 0.2 Amber + 0.1
3	Body, author's posts, emotion, thread length, LIWC (affective, female, negate), seed terms, and the last sentence	Crisis + 0.3 Red + 0.2



Conclusions

- We identified *non-green* posts with F1-score of over 80%.
- We improved over the baseline in virtually all categories.
- The most effective features were psycholinguistic, contextual and sentence level affective features.
- Boosting the classifier predictions for the critical categories resulted in further improvements.
- F1-score for *crisis* (28%), *red* (48%), and *amber* (67%) correlate with their distribution in the dataset; lack of training data in critical categories is the main reason of lower performance on these categories.

¹ Pennebaker, J.W., Boyd, R. L., Jordan, K. & K. Blackburn (2015). The development and psychometric properties of liwc2015. *UT Faculty/Research Words*.

² Staiano, J. and M. Guerini (2014). Depechemood: a lexicon for emotion analysis from crowd annotated news, in *Proceedings of ACL 2014*, 427-433.

³ Wilson, T. Wiebe, J., and P. Hoffmann (2005). Recognizing contextual polarity in phrase-level sentiment analysis, in *Proceedings of the conference on EMNLP 05*, 347-354.