Assignment III – CS 6474 Social Computing

Tyra Silaphet

4/8/2025

Introduction

This assignment is focused on replicating and extending key ideas from Althoff et al. (2014), who investigated the factors that contribute to the success of altruistic requests in online communities, specifically on Reddit's "Random Acts of Pizza" (RAOP) subreddit. Their study demonstrated that beyond the content of the request, factors such as *who* is asking, *how* they ask, and the social context surrounding the post all play a role in determining whether someone receives help.

Part 1: Classification Models

To explore these dimensions, I implemented four classification models to predict whether a pizza request posted on RAOP would be successful or not. The dataset included 5,671 posts from December 2010 to September 2013, each labeled with whether the requester received a pizza. The models were designed to test distinct types of features to isolate different psychological mechanisms that could drive altruism. By comparing their performance, the goal is to identify which features most effectively predict successful requests.

a) **Model 1: N-grams**

   Features: Top 500 unigrams and top 500 bigrams

   How you ask matters (linguistic style)

b) **Model 2: Activity and Reputation**

   Features: Activity and Reputation

   Who asks matters (trustworthiness)

c) **Model 3: Narratives**

   Features: Normalized counts of words from 5 narrative dictionaries

   Context matters (narrative context)

d) **Model 4: Moral Foundations**

   Features: Normalized counts of words from 5 moral dimensions

   Why you deserve it matters (moral reasoning)

As specified in the assignment requirements, I trained all models using a Support Vector Machine (SVM) classifier with a linear kernel and default parameters, and I split the dataset into a training set (90%) and a test set (10%) for consistency. The code used to implement these models can be found in the accompanying file titled "Part1Code.ipynb."

The table below summarizes the performance of each model:

Model	Accuracy	Precision	Recall	F1	Specificity	AUC
1. N-grams	0.637	0.347	0.536	0.421	0.671	0.620
2. Activity and Reputation	0.801	0.603	0.564	0.583	0.879	0.839
3. Narratives	0.590	0.218	0.257	0.236	0.699	0.495
4. Moral Foundations	0.731	0.240	0.043	0.073	0.956	0.464

Table 1. *Performance metrics for four SVM classifiers predicting "Random Acts of Pizza" (RAOP) subreddit request success using different feature sets.*

Part 2: Model Performance Discussion

The performance of the four models reveals critical insights into what drives successful altruistic requests on social media. Among them, Model 2 (Activity and Reputation) performed the best, with an accuracy of 0.801, F1 score of 0.583, and AUC of 0.839. This indicates that user behavior and reputation data, such as account age, comment and post history, and upvote/downvote metrics, are strong predictors of whether a request will be fulfilled. Unlike language-based models, these features are less prone to noise and ambiguity and reliably reflect the requestor's credibility and engagement within the RAOP community.

In contrast, Model 4 (Moral Foundations) performed the worst overall, with an AUC of just 0.464 and a recall of only 0.043. One possible reason for this is that moral language (references to fairness, loyalty, or care) is often subtle or implied rather than clearly stated in pizza requests. Because the model relies on matching words from a fixed dictionary, it may miss these nuances. This method also tends to misclassify neutral or ambiguous posts as negative, which most likely explains its high specificity (0.956) but very low recall. In other words, the model was good at identifying unsuccessful requests, but it had a hard time recognizing the ones that were actually successful.

Model 1 (N-grams) showed moderate performance (AUC: 0.620; F1: 0.421), suggesting that basic linguistic patterns still carry some predictive value. Politeness cues like "please" and "thank you," as well as words reflecting need, such as "hungry" or "unemployed," likely helped the model identify successful requests. Still, it fell short of Model 2, reinforcing the idea that *who* is asking may matter more than *how* they phrase their request. Model 3 (Narratives) performed slightly worse, with an AUC of 0.495. The narrative dictionaries it relied on, covering themes like "family," "student," and "money," were often too broad to capture specific, meaningful signals. Some categories, like "desire," included general terms such as "game" or "movie," which didn't consistently align with compelling or urgent requests. As a result, the model had trouble recognizing stories that would actually connect with readers and indicate a successful request.

The difference in performance across the four models highlights an important pattern: models that rely on clear, frequent signals, like user behavior (Model 2) or common word patterns (Model 1), tend to perform better. In contrast, Models 3 and 4, which are based on more abstract semantic categories, struggled because they rely on rigid keyword lists and lack the ability to interpret context. Between the two, Model 3 performed slightly better than Model 4 in terms of recall (0.257 vs. 0.043). This may be because narrative terms like “student” or “family” are more specific and commonly used in these requests than broader moral concepts. Although both models aim to capture deeper meaning, their dictionary-based approaches often miss important nuances like tone, emotional appeal, or subtle persuasion. This shows the limitations of trying to analyze complex language using simple keyword matching, especially in the messy and informal context of real-world online posts.

So, is language alone enough to predict the success of an altruistic request? To some extent, yes. Model 1 shows that even basic textual patterns can offer moderate predictive power. However, the relatively weak performance of Models 3 and 4 shows that higher-level linguistic framing, such as storytelling or moral appeals, doesn’t translate well into predictive features without more nuanced analysis. Overall, language matters, but it’s not the most important factor. Model 2’s strong performance makes it clear that trust, credibility, and community engagement are far more decisive. People may be more willing to help when they recognize a requestor as active, credible, or respected, regardless of how emotionally appealing or well-written the post is. These behavioral signals act as implicit indicators of authenticity, which may increase the likelihood that others view the request as genuine and deserving of support. Unlike text, which can be crafted or embellished, engagement history is harder to fake, making it a stronger indicator of intent and reliability.

Another thing I noticed is that while textual models struggled to match the predictive power of Model 2, they still offer complementary insights. For example, a person’s use of narrative framing or moral language might not be predictive on its own but could enhance behavioral models when combined. Requests with both a strong reputation signal and emotionally resonant language may stand out more to potential donors. This calls for potential feature interaction, where subtle language cues might only be meaningful in certain behavioral or social context, something not captured in models evaluated separately. For future work, it may be valuable to combine the strengths of Models 1 and 2 to build hybrid models that better capture both what is said and who is saying it. This could lead to more robust predictions that reflect both emotional appeal and community dynamics.

Part 3: Comparative Discussion

When comparing my models to those in Table 4 of Althoff et al. (2014), there are clear similarities in goals and feature types. Both studies aim to predict the success of altruistic requests on the "Random Acts of Pizza" subreddit using a combination of textual and social features. Althoff et al. experimented with logistic regression models using linguistic features (e.g., gratitude, evidentiality, reciprocity), social attributes (e.g., status, karma), and temporal factors. Similarly, I implemented four SVM-based models that isolate the predictive power of linguistic style (Model 1), user activity and reputation (Model 2), narrative framing (Model 3), and moral reasoning (Model 4). A key difference in our approaches is that Althoff et al. combined multiple feature categories in their best models, while I evaluated each feature set independently to better understand their individual impact.

Another important difference is the choice of model and data split. Althoff et al. used logistic regression, while I used Support Vector Machines (SVMs) with a linear kernel, which may have captured more complex patterns in the data. Additionally, I used a 90/10 train-test split, whereas their study used a 70/30 split with cross-validation. This difference in partitioning, along with the use of a fixed random seed in my implementation, may have contributed to higher performance metrics by reducing test variance and better preserving signal in the training set. Despite these variations, we arrived at the same conclusion. Althoff et al. found that social features like status and karma significantly boosted predictive power, and my results reinforce this finding through Model 2's strong performance.

In terms of results, Althoff et al.'s best-performing model, which combined social, temporal, and textual features, achieved an AUC of 0.672. Their text-only model, based on targeted linguistic features, reached 0.625. In contrast, my Model 2 (Activity and Reputation) outperformed both with an AUC of 0.839, suggesting that even when evaluated on its own, user metadata can be highly effective for prediction. This performance gain may also be attributed to the more granular nature of my features, such as subreddit-specific engagement, comment/post history, and account age, compared to the broader social metrics (e.g., karma) used in the original study.

While my behavioral model exceeded the performance of Althoff's full-feature model, my language-based models performed similarly or worse. Model 1 (N-grams) achieved an AUC of 0.620, closely matching their linguistic model. However, Model 3 (Narratives) and Model 4 (Moral Foundations) lagged significantly, with AUCs of 0.495 and 0.464 respectively. These models relied on dictionary-based keyword matching, which was less effective than the original paper's use of topic modeling and LIWC-informed feature engineering. These results highlight the limitations of static lexicons for capturing nuanced communication strategies like storytelling or moral appeals, especially in short, informal online posts.

Ultimately, both my study and Althoff et al.'s work support a common conclusion: while linguistic style and framing carry some predictive value, user reputation and behavioral signals are far more reliable indicators of success in altruistic online communities. My findings suggest that even without combining features, detailed user metadata alone can outperform more complex integrated models.

Conclusion

This assignment helped me better understand the factors that drive successful altruistic requests in online spaces. After building and evaluating four different classification models, it became clear that user behavior and reputation features were the most powerful predictors of success. While language-based features had some predictive value, it wasn't surprising that Model 2 (Activity and Reputation) outperformed the rest. Quantitative indicators like account age, comment history, and post engagement are far more consistent and difficult to manipulate compared to how someone chooses to phrase their request.

What really struck me, though, was how this finding connects to a broader theme we discussed in class. We've recently talked about how innocuous, everyday digital interactions like Facebook likes can be used in ways that invade privacy or feed into manipulative systems. The Cambridge Analytica case taught us to be more cautious about how our digital footprints can be monetized or exploited. But in the context of RAOP, those same interactions—those footprints—are what build trust and enable generosity. The more active, consistent, and genuine someone is in the community, the more likely they are to receive help. That contrast really stood out to me: the very behaviors we're warned to guard can actually be the foundation of meaningful support in the right context.

It made me realize that digital interaction isn't inherently good or bad. It just depends on how platforms are structured and what the data is used for. In a space like RAOP, where generosity is the goal, your interaction history becomes a signal of credibility, not just a data point to be sold. This also reinforces the idea that we need to design online systems more thoughtfully, especially when data is involved. For me, this assignment was more than just an exercise in modeling, but it was a chance to reflect on how algorithms, behavior, and social computing collide in ways that can be both powerful and deeply human.