Twitter User Account Classification to Gain Insights into Communication Dynamics and Public Awareness During Tampa Bay’s Red Tide Events

Andrey Skripnikov, Tania Roy, Fehmi Neffati, Melvin Adkins
New College of Florida
5800 Bay Shore Rd
Sarasota, Florida
USA, 34243

Marcus Beck
Tampa Bay Estuary Program
Tampa, Florida
USA

Abstract
This study presents an innovative approach to analyzing environmental challenges, focusing on the localized impacts of toxic algal blooms, specifically the dinoflagellate Karenia brevis on Florida’s Gulf Coast, commonly known as “red tide”. Despite the extensive influence of social media in public discourse, its potential in environmental awareness remains largely untapped. Our research exploits Twitter data to examine communication trends and public understanding of red tide issues in the Tampa Bay area from 2018 to 2022. For that study period, we collected 63K tweets from 30K accounts that mentioned terms related to red tide. Our methodology involves a tiered labeling process to obtain over 15K labeled accounts. In the initial tier, we employ predefined dictionaries for account groups to establish preliminary class designations, streamlining the subsequent labeling tiers, one of which is aided by preliminary machine learning classification. Having used several text classification algorithms and feature preprocessing approaches, Support Vector Machine with Bidirectional Encoder Representations from Transformers (BERT) yielded the best cross-validation performance in both accuracy (90%) and versatility (unweighted F1 score of 0.67). Lastly, we creatively leveraged the Term Frequency-Inverse Document Frequency (TF-IDF) method to study the terms that most distinguish each user category from the rest.

1 Introduction
Although strides have been made in advancing the capability for physical measurements indicative of certain environmental issues (e.g. microscopic analysis of water samples, sediment and fish tissue analysis, etc), the crowdsourcing dimension hasn’t been given as much attention. Over the past decade social media platforms have grown substantially, offering potentially unique opportunities to assess complementary sources of information that can aid management response to, or knowledge of, local environmental concerns. For large-scale disaster events that receive extensive worldwide media coverage (e.g. hurricanes, typhoons, earthquakes, floods), research has focused on social media communication patterns across various platforms, including Twitter (Eismann, Posegga, and Fischbach 2016), Facebook (Bhuvana and Aram 2019), Instagram (Sherchan et al. 2017), and Flickr (Liu et al. 2008), demonstrating the value of leveraging such data. Not as well-studied are issues where the impacts and media coverage are more localized, such as toxic algal blooms caused by the dinoflagellate Karenia brevis, more commonly known as “red tide”, that impact Florida’s Gulf Coast with regularity. Although not as severe in terms of physical damage and loss of human life as the more acute disasters, red tide causes large die-offs of marine organisms (Steininger 1996), which in turn significantly hinder local tourism and the coastal economy. Moreover, the aerosolized toxins produced by K. brevis cause respiratory or skin irritation for humans, occasionally resulting in hospitalizations (Backer et al. 2005). Notable recent blooms in southwest Florida have occurred in 2018 and 2021, with the former bringing some of the most severe impacts in history, while the latter was likely exacerbated in Tampa Bay related to inorganic nutrient sources from an industrial wastewater release [X, X]. In light of this, the local governments, resource managers and environmental agencies are increasingly interested in ways to gauge public’s awareness about red tide, and social media could help fill that gap.

Our project will focus on Twitter, recently rebranded as "X" - a social media platform with over 450 million monthly active users (as of 2022) who can communicate their thoughts by sharing short strings of text, which we’ll refer to as "tweets". Very little research has evaluated the use of Twitter data to evaluate public concern and awareness during red tide events. [X] demonstrated strong correlations between localized Twitter activity and the beach/ocean conditions observed in southwest Florida during the 2018-2019 red tide event, while Mascareño et al. 2020 used Twitter to study social connections during the 2016 red tide event at Chiloe island, Chile.

The goal of this work is to develop a data-driven classification methodology for identifying Twitter account type based on user description. This classification will allow a more detailed assessment of public knowledge and awareness of different user groups (such as media, government, general public, etc) about red tide, potentially aiding re-
rational algorithms to achieve a good classification performance.

2 Background

In recent years, a surge of interest has emerged in the dynamics of information dissemination, particularly concerning public interest topics such as health and environmental issues. This section consolidates various research findings on information sharing, public engagement, and the impacts of specific events, such as the red tide outbreaks and the COVID-19 pandemic. The lessons learned from information dissemination through online platforms such as Twitter are discussed below.

2.1 Red Tide Communication: Significance and Challenges

Red tide in Sarasota County, Florida, led to increased hospital diagnoses and economic losses, including an annual $22 million due to Harmful Algal Blooms (Wright 2016, Ferreira et al. 2021, Science 2022). In 2018, Airbnb rental rates dropped by $0.45/day with each detection of K. Brevis, impacting tourism. Staugler et al. (2021) found misalignment in public agencies’ communication strategies and public expectations, with social media misinformation exacerbating the issue (Staugler, Simoniello, and Monaghan). This mirrors the COVID-19 scenario, where despite using Twitter for information dissemination, vaccine opposition increased by 80% (Bonnevie et al. 2020), highlighting the need for better management of social media-driven public concerns, particularly in environmental crises.

2.2 Twitter as a Platform and use of Machine Learning

Our work builds on the integration of social media data in crisis communication strategies as highlighted in studies by Krimsky et al., Gilani et al., and Pennacchiotti et al. Krimsky et al.’s 2021 study on red tide in Florida revealed significant but varied social media usage: 2.5% regional reliance, with a preference for Twitter (11.7%) and Facebook (38%). Gilani et al. utilized a Random Forest classifier to distinguish between automated bots and human users on Twitter, leveraging human-annotated datasets for validating algorithm accuracy. This highlighted the importance of human judgment alongside automated methods (Krimsky, Montes, and Johns 2021, Tang et al. 2021, Gilani, Kochmar, and Crowcroft 2017). Similarly, Pennacchiotti et al. employed Gradient Boosted Decision Trees to extract demographic features from Twitter profiles, finding that additional tweet context improved prediction accuracy. These studies collectively demonstrate the crucial role of sophisticated data analysis, machine learning, and human insight in optimizing social media’s potential for effective crisis communication (Pennacchiotti and Popescu 2011).

3 Methodology / Technical details

3.1 Data Collection

To collect relevant historical Twitter data, we used the Full-archive search under the Academic Research track of Twitter API (no longer available as of this writing). To create search queries, we developed the search term dictionary for red tide, and also separate search term dictionaries to indicate reference to the five Florida counties of interest: Hillsborough, Pinellas, Manatee, Sarasota, and Pasco. We started with a list of terms recommended by an advisory committee of local government representatives and resource management professionals, then manually evaluated the API query results for those terms to confirm relevancy, removing or modifying keywords that yielded too many matches unrelated to red tide or the five counties in the Tampa Bay area. In addition, after an array of terms was confirmed to yield quality results, new candidate keywords were tested via “exclusionary” search: query the API for tweets that contained those keywords and none of the other established dictionary terms. The final search term dictionary for red tide was: {“red tide”, “red tides”, “red algae”, “#redtide”, “karenia brevis”, “kbrevis”, “#kareniabrevis”}. This resulted in 52,476 tweets and retweets relevant to red tide and Tampa Bay area. Moreover, to capture potentially relevant discussions, we collected all replies to those tweets. For that, we leveraged Twitter API functionality allowing one to search for full reply threads using a unique conversation identifier. That added another 11,106 unique threads.

3.2 Account Types

Alongside the tweet contents, Twitter API also allowed to pull information about the users posting the tweets, such as their account name and description. Our data set included 30,490 unique accounts, and we were interested in categorizing them based on the contents of the description field, which has a 280-character free-text entry format.

Five account categories were defined based on interests of our advisory committee: “Academia/Research”, “Government”, “Media”, “Tourism” and “Other” (mostly general public). Keyword dictionaries for the first four categories were created using domain knowledge and suggestions of our advisory committee, followed by an extensive procedure of vetting certain terms analogous to the dictionary development from Section 3.1. For example, terms like “college” or “university” were removed from the initial “Academia/Research” dictionary as they yielded many accounts of users who were simply alumni of the school, while a word like “adjunct” was discovered in some of the descriptions and served as a useful addition to the dictionary.

The following category definitions were created after manual review of at least 50 accounts per category by three annotators (referred to as Annotator 1, 2, and 3, respectively). “Academia/Research” includes those either working as faculty in academia, or those employed full-time in a position involving scientific research for academic publishing purposes. “Government” consists of those employed by the government. “Media” includes those working in the news industry, providing original news coverage, having to do with
creative content generation, such as writing, speaking, or content-editing. "Tourism" includes businesses and establishments strongly relying on tourism or coastal resources. "Other" implies accounts in none of the previous categories, expected to consist vastly of the general public. See Table 1 for specific examples of occupations and descriptors for each category.

### 3.3 Tiered Data Labeling

With the prohibitive amount of distinct accounts in our collected data set (30,490), we have attempted using Amazon Mechanical Turk (Dupuis, Renaud, and Searle 2022) to crowdsourse labels for accounts. We provided category descriptions similar to Table 1 and recruited each participant to label 300 accounts. These participants had no domain expertise. After evaluating the results from first 20 participants, inconsistency in provided labels made evident that the account classification task, unlike a more intuitive labeling task like image classification, posed significant challenges to those lacking our domain knowledge. That prompted us to carry out the labeling ourselves via a tiered approach.

First, a random subsample of 8,000 accounts was selected for manual categorization. To streamline the labeling process, we executed a tiered approach (Figure 1). Tier 1 used automated term-matching based on account type keyword dictionaries discussed in Section 3.2 to give each account a preliminary category designation. It lemmatized (Bird, Loper, and Klein 2009) the user description fields, counted the number of description terms to match with each of the first four account type’s keyword dictionaries, and assigned it to the category(ies) with most matches. Descriptions with none of the keywords from said dictionaries were assigned to the "Other" category (mostly general public). Note that for preliminary account labeling we also considered using unsupervised topic modeling approaches like Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003), but quickly realized that dictionary-based approach was much more appropriate in our case due to the specific nature of the user groups of interest outlined to us by the committee of regional stakeholders. Categorization produced by LDA wasn’t able to pick up on such groups as well as the basic keyword dictionary matching approach described above. For Tier 2, the accounts were grouped by Tier 1 designation ("Academia/Research", then "Government", etc), allowing Annotator 1 to manually label those in sequence of the groupings. Depending on the description, preliminary designation was either left unchanged, or re-assigned a more appropriate category. Although one could argue that the preliminary Tier 1 designations biased the Tier 2 labeling, (Syrdal et al. 2001) provided a controlled study indicating such bias for trained annotators to be negligible. Moreover, this streamlined grouping approach increased the speed of labeling, also agreeing with (Syrdal et al. 2001). It helped Annotator 1 anticipate categories based on text patterns yet also be vigilant about false positives (e.g. someone retired from the occupation of interest, like "former reporter", "ex mayor", etc). In Tier 3, Annotator 2 resolved unclear cases identified by Annotator 1 in Tier 2. Examples of unclear cases: whether to consider freelance writers as media (only if they cover news), or real estate agents as tourism (yes, given the prevalence of vacation rentals in Florida).

Having labeled the 8,000 accounts we found strong class imbalance in favor of the "Other" category (90%), prompting us to enhance the representation of smaller categories for the sake of improved versatility of classification performance across all classes. For that, a "biased sampling" method was considered, where we leveraged the keyword dictionaries discussed in Section 3.2 in order to obtain all the accounts that had at least one such keyword. We have then carried out the same Tiers 1-3 from Figure 1 to manually label those accounts, hereby enhancing our training data with 2,322 additional labeled cases (for a total of 10,322), increasing prevalence of underrepresented classes to 20%.

Afterwards, for Tier 4 (Figure 1) we have trained a preliminary machine learning algorithm to classify the remaining 20K accounts (we picked SVM with unigram Bag-of-Words feature representation, for more detail see Sections 3.4 and 4.1). Only 3% of those accounts (677) were classified as one of the non-"Other" categories, and we manually went through them to verify the labels. Subsequently, we also looked at all accounts having at least a 0.15 probability for any class besides "Other" (805 such cases), hand labeling them as well. To further augment our training data in a time-efficient manner, we manually went through all accounts with descriptions of fewer than 30 characters (Tier 5), labeling them according to the streamlined process used back in Tiers 1-2. It added 1.5K labeled examples to our training set, not including 3,798 accounts with empty description (categorized as "Other", but not used for text classification due to not being informative for that purpose).

### 3.4 Machine Learning Algorithms for Account Type Classification

With the primary goal of classifying the free-text entries in the user account description field, we used naive Bayes (NB), random forest (RF) and support vector machines
4 Results

The tiered labeling approach described in Section 3.3 resulted in 11,915 manually classified accounts. That set excludes 3,798 accounts with empty descriptions (labeled “Other” by default) and 155 accounts with descriptions not written in English. Our training set label distribution was as follows: 539 academia/research, 128 government, 1,624 media, 194 tourism, 9,441 - other (mostly general public).

Section 4.1 describes performance evaluation of various machine learning models and feature transformation approaches we considered. The best-performing model is subsequently used to classify the remaining 13K accounts, with the entire set of 30K accounts being leveraged to compare the communication tendencies and terms that distinguish different account types.

4.1 Machine Learning for Account Classification

We ran several models with different N-gram settings (N=1,2, both), classifiers (NB, RF and SVM) and feature extractors (BOW, TF-IDF). Unigram representation (N=1) was either outperforming or at least similar to the bigram (N=2) and mixed (N=1&2) feature formats in terms of accuracy and F1 scores. As for feature extractors, the TF-IDF representation did not outperform the BOW format, despite the latter being a simpler method. Both of these findings were consistent regardless of the classifier, pointing to the BOW unigram representation as a better choice when compared to higher order N-grams and TF-IDF.
When comparing the 10-fold cross-validation performance of RF, NB and SVM on the training set of 9,538 accounts, SVM yielded the best results, especially in terms of F1 scores (Figure 2). In particular, for the unigram BOW models, SVM showed the highest macro average F1 score of 0.61 (no higher than 0.55 for NB, RF), with superior performance on virtually every individual class, particularly the government (0.37, others - below 0.10). Moreover, when considering BERT for feature extractor, SVM improved even further. The macro average F1 score rose to 0.67, outperforming any other approach on each individual class (most classes had F1 score over 0.60, none lower than 0.40), pointing to superior classification versatility. Lastly, SVM BERT also exhibited highest accuracy (0.90) and weighted F1 score (0.89), both indicative of a better overall classification performance.

![Performance Scores by Metric and Classifier](image)

Figure 2: Classification performance metrics comparing several model types (NB, RF, SVM) for unigram Bag-of-Words feature transformation, and also BERT for the best-performing SVM model. Dictionary-based classification approach is used as the baseline.

When applied to the test set of 2,385 accounts, the performance metrics for SVM BERT held up. It had macro average F1 score of 0.72, all individual class F1 scores ≥ 0.58, overall accuracy and weighted F1 scores ≥ 0.90. It showed model’s ability generalize beyond the training data, prompting us to use it for classifying the remaining 13K accounts.

4.2 Statistical Analysis of Group Differences

Left plot on Figure 3 illustrates the 95% confidence intervals for average character lengths of the original posts by accounts from each respective group. Government accounts had the longest tweets, averaging 160–170 characters, while the general public had much shorter tweets (< 120 characters). That difference was both statistically significant (see 95% confidence intervals) and practically important (Cohen’s d of 0.6). Media, academia and tourism accounts had moderate tweet lengths averaging 120 – 150 characters.

Right plot on Figure 3 contains word cloud of terms that most effectively differentiate each user category, created via the TF-IDF method detailed in 3.5. For the general public accounts ("Other"), words like "ton", "dead" and "beach" refer to the tons of dead marine life washing ashore, while "bloom" stems from the toxicity of algal blooms, pointing to public awareness of red tide’s nature and impacts. Terms "money", "covid" and "governor" suggest the general public’s concerns about government activities and spending during red tide events. In the meantime, government accounts tended to use leadership-related terminology ("mayor", "issue", hashtag "protectingtogether") and official language in talking about big-picture aspects ("impact", "public"). Media utilized matter-of-fact reporting terms like "story", "update", "impact", "concentration" (of red tide), etc. Academics and researchers tended to employ scientific terminology ("algal", "marine") and mention research-related concepts ("science", "research", "lab"), despite us not having used tweet contents for the account classification task. Lastly, tourism accounts discussed impacts to local businesses and tourist attractions ("business", "trout", "beach", "park"), although at times emphasizing the positive messaging (e.g. "sunglass" comes from the sunglasses emoji). Term "bad" being prevalent across several account categories indicates the negative nature of red tide’s impacts.

5 Discussion

A tiered labeling approach, followed by implementing state of the art machine learning models and feature extraction methods on the labeled data, led to several findings of note. Despite its simpler nature, the unigram Bag-of-Words feature transformation method performed on par or better than bigrams and TF-IDF representations, likely due to many user descriptions simply listing self-descriptive nouns or adjectives, e.g. "Professor, grandmother, Democrat, francophile" or "Award winning Documentarian, Seasoned Journalist and Author". Unlike a more narrative writing style, that format didn’t seem to benefit from higher orders of N-grams as much. Support Vector Machine tended to perform better than both Naive Bayes and Random Forest, regardless of feature extraction approach chosen, proving to be more flexible in classifying such text format. Lastly, BERT feature extractor improved the best-performing SVM model even further, confirming the value of a more thoroughly contextualized sentence-encoding.

In Section 4.2 we demonstrated the value of distinguishing between Twitter account categories to obtain insights into red tide communication tendencies in the Tampa Bay area. In particular, media tended to report using tweets of moderate length and matter-of-fact terminology (e.g. "Birds sickened from eating fish killed by red tide - FOX 13 News, Tampa Bay"). General public mostly used short tweets and voiced concerns about government’s handling of the situation ("How about money for red tide clean up?"). Govern-
ment, on the other hand, averaged longer tweets, providing extensive information to keep citizens updated on conditions and advisories (“Manatee County residents, @ManateeGov has a great #RedTide resource providing updates on current conditions and beach status reports…”). Lastly, academics and tourism-related accounts mostly discussed the aspects most relevant to their primary interests (red tide research for the former, impacts to businesses/tourism for the latter), using tweets of moderate length.

Besides the distinct communication dynamics of various groups, we could also gauge the levels of public awareness about red tide. Consistent use of words like “ton”, “dead”, “beach” and “bloom” across several groups, but more importantly the general public, indicated the understanding of red tide’s nature (it’s an algal bloom) and its most tangible impact in public’s eyes (dead marine life on beaches).

All of the aforementioned could aid local governments and resource managers in several ways. First, the results allow specific insight into how different user groups respond to environmental conditions. Secondly, it offers an understanding of the narratives among the general public, which could help crafting communication strategies to directly address public concerns. Lastly, if a certain targeted marketing campaign is launched in an explicit attempt to educate and increase awareness about the issue, shifts in terminology used by general public accounts could be studied to evaluate the effectiveness of such a campaign.

6 Conclusion, Limitations and Future Work

Our work employed a tiered labeling approach for categorizing Twitter user accounts into distinct groups based on user descriptions. We subsequently implemented a handful of machine learning models and feature extraction methods, leading to a respectable text classification performance. The value of Twitter account categorization was demonstrated on the example of distinguishing term “user frequency” analysis (based on TF-IDF word clouds). While having limitations, this study offers potentially valuable insights for regional government in understanding public awareness and communication patterns around a localized environmental issue like red tide in Tampa Bay. As part of future work, we are considering other environmental topics pertinent to the area such as a wider topic of Harmful Algal Blooms (HAB), or various spill events (industrial, sewage, oil).

One limitation of the work is the fact that Twitter/X user account groups may not be fully representative of the respective communities (e.g. academics on Twitter might differ from academics in general). Moreover, due to the recent changes to the Twitter/X platform, users might have switched to alternatives such as Mastodon (Mastodon 2016) or Threads (Meta 2023), which could result in shifted discussion landscape compared to the time frame of our analysis (2018-2022). Although our results are promising, more future work is required to study whether our findings are generalizable across platforms and time frames.

Finally, despite a respectable classification performance of our SVM BERT model, there are avenues for further improvement. First, given that user is generally expected to start their description with words that they self-identify with the most, a text classification weighting earlier words more heavily could be used in order to account for that. Second, while BERT has been shown to be an elite feature transformation approach in natural language processing, other large language models could be considered. Among those: permutation language model XLNet (Yang et al. 2019), Google’s Pathways Language Model (PaLM) (Chowdhery et al. 2023) or Generative Pre-Trained Transformer (GPT) (Yenduri et al. 2023). Lastly, deep learning architectures like Long-Short Term Memory (LSTM) and Convolutional Neural Networks (CNN) be considered for further improvement of predictive performance (Schmidhuber 2015).

7 Acknowledgments

This project was supported through the Tampa Bay Environmental Restoration Fund (TBERF). We would also like thank the advisory committee for their provided expertise.


[Staugler, Simoniello, and Monaghan 2020] Staugler, E. A.; Simoniello, C.; and Monaghan, P. Insights from tourism, media, small business hospitality industry and public information officer professionals on key elements of red tide messaging and modes of communication.


[Wright 2016] Wright, P. 2016. 10 things to know about florida’s harmful algae blooms.

