Analysing Twitter Data on Rhino Conservation using Natural Language Processing

Swapnil Mallick
Department of Computer Science
University of Southern California
Los Angeles, CA 90007
smallick@usc.edu

Venkatesh Dharmaraj
Department of Computer Science
University of Southern California
Los Angeles, CA 90007
vdharmar@usc.edu

Kiran Sampat
Department of Computer Science
University of Southern California
Los Angeles, CA 90007
ksampat@usc.edu

Samarth Ganesh Khaire
Department of Computer Science
University of Southern California
Los Angeles, CA 90007
khaire@usc.edu

Team Unicornis (Team-1)

Abstract

This study addresses the complex issue of illegal wildlife trade, focusing on rhino poaching and its profound impact on global biodiversity. Our research leverages the power of social media data to gauge public sentiment and identify trends in rhino conservation efforts. We analyzed a dataset comprising over 19 million tweets, with a focus on English-language content related to rhino conservation. Various toolkits, including Pattern, TextBlob, VADER, and Twitter-roBERTa-base, were utilized to gauge the sentiment expressed in the tweets. In parallel, topic modeling techniques such as LDA and BTM were applied to gain deeper insights into the prevalent discussion topics and identify the user categories participating in these conversations. These methodologies yielded satisfactory outcomes, providing valuable insights into the critical aspects of rhino conservation discussions on Twitter. Furthermore, this study shed light on the extent to which pre-trained natural language processing tools can be relied upon to draw conclusions about conservation related issues.

1 Introduction

Illegal wildlife trade is a pervasive and highly challenging issue concerning global biodiversity preservation and international collaboration [35]. It is one of the most profitable transnational organized crimes [55]. The commerce of wildlife has roots that stretch back to ancient times [3], with early Egyptian and Greek civilizations recording commercial dealings involving wildlife, a tradition that has persisted throughout history [26]. Instances of extensive and unsustainable commercial utilization of wildlife were observed even during the Roman Empire [3], and concerns regarding the unsustainable exploitation of wildlife for conservation purposes can be traced back to at least the 1960s [32].

Illegal wildlife trade is causing immense harm to wildlife species globally, with poachers, traffickers, and well-organized criminal groups relentlessly pursuing financial gain without regard for the consequences to meet consumer demand [54]. The trafficking and unsustainable trade of wildlife products such as elephant ivory, rhino horn, pangolin scales, tiger bone, bear bile, and rosewood are leading to unprecedented declines in a wide range of wildlife species, including some of the world's most iconic creatures as well as lesser-known species.

Asia, with a particular focus on China and Southeast Asia, is recognized as a critical region for both the procurement and the utilization of illegal wildlife [5, 27]. Rhino horns, elephant ivory, and tiger products remain in high demand among consumers, particularly in Asia. One of the primary motivations cited for the use of rhino horn is its perceived health benefits [25]. Rhino horn is believed to possess detoxifying properties, which stems from Chinese traditional medicine practices. Practitioners in this tradition characterize rhino horn as having bitter, acidic, and salty attributes, which are thought to be effective in lowering body temperature and purging toxins from the body [11]. Vietnam, a nation where the native Javan rhinoceros (*R. s. annamiticus*) was officially declared extinct in 2011, has been acknowledged as one of the globe's major hubs for the illegal trade of African rhino horns. In just the initial five months of 2015, Vietnamese authorities confiscated a minimum of 100 kilograms of illegally imported rhino horns [25]. In Vietnam, an unfounded belief that rhino horn possesses cancer-curing properties resulted in extensive poaching in South Africa [9] and drove the price of a single rhino horn to as much as $300,000 [13].

The rhino poaching crisis we are currently facing commenced in 2008, witnessing a surge in the number of rhinos slaughtered for their horns across Africa until 2015. Fortunately, there has been a decline in poaching incidents across the continent since the peak of 1,349 in 2015 [53]. However, it's crucial to note that at least one rhino still falls victim to poaching daily, emphasizing the urgent need for further action. South Africa, home to the majority of the world's rhino population, has borne the brunt of poaching activities, enduring over 1,000 rhino deaths annually from 2013 to 2017. South Africa continues to grapple with significant losses to poaching, with 448 rhinos illegally killed in 2022, a slight decrease from the 451 killed in 2021. In anticipation of World Rhino Day on September 22, 2023, authorities in Africa estimate that by the end of 2022, the continent was home to 23,290 rhinos, marking a 5.2% increase compared to 2021 [49]. While this is a positive trend, it's essential to note that a rhino is still poached in the country every 20 hours [53].

Animals, including rhinos, play a vital role in biodiversity and ecosystem dynamics. They create habitats for other species, offering opportunities for future global restoration and rewilding initiatives. Thriving wild rhino populations can also have a positive impact on the livelihoods and well-being of local communities, attracting tourists from around the world, generating employment opportunities, and contributing to economic development.

Numerous global and regional organizations are committed to wildlife conservation. One of these is the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), an international treaty with the primary goal of ensuring the survival of wildlife in the face of trade [7]. The UNEP World Conservation Monitoring Centre has established the CITES Trade Database, which encompasses more than 25 million trade transaction records (as of 2023) involving living animals, plants, and their derivatives [51, 50]. This valuable database plays a crucial role in monitoring the extent of illegal wildlife trade and offers a comprehensive overview of significant trade activities between countries.

In today's age of information technology, personal computers and the widespread availability of affordable smartphones, social media services and platforms like Facebook, Instagram, Twitter, Flickr, WhatsApp, WeChat and Sina Weibo have collectively attracted over two billion users worldwide. This accounts for approximately 29% of the global population and encompasses around 67% of all internet users worldwide [33]. Hence, social media presents a potential avenue for devising creative approaches to promote social change and combat the loss of biodiversity [34]. Notably, social media has demonstrated its capacity to influence people's actions [40, 44, 36].

Leveraging the extensive reach and impact of social media could enhance conservation initiatives and mobilize pro-environmental actions on a broad scale to benefit nature conservation efforts [46]. Twitter provides a valuable means to gather and assess public sentiment regarding the boycott of wildlife trade or the call for government bans on such trade [31]. It has emerged as a potent instrument for data collection, offering insights into the public's perception of this issue. Consequently, social media offers a fresh and distinct vantage point for the analysis of trends in rhino trade, differing significantly from conventional data sources.

This study is designed to demonstrate the utility of social media data in monitoring public opinion regarding illegal rhino poaching and trade. To fulfill the research objectives, our study has been divided of three main components:

1. Understanding public sentiment towards rhino conservation to assist in the prioritization of conservation efforts. Positive sentiment can mobilize support, while negative sentiment can pinpoint areas requiring immediate attention and intervention.
2. Categorizing social media posts by topic to aid conservationists in identifying the most critical issues and priorities concerning animal conservation.
3. Analyzing and classifying Twitter user profiles to gain insights into which user demographics are actively involved in tweeting about rhino conservation.

To garner vital insights into public sentiment as well as to monitor the evolving trends surrounding rhino conservation efforts, we aim to use Natural Language Processing methods to analyze tweets related to rhinos and see how useful these can be in rhino conservation efforts.

2 Related Works and Background

The widespread illegal wildlife trade poses a significant threat to biodiversity conservation. Historically, researchers have sought to analyze this phenomenon by conducting global surveys on wildlife trade [1]. Additionally, there have been endeavors to assess the impact of the 1989 CITES ivory trade ban on reducing both ivory demand and elephant poaching, while also exploring innovative approaches to decrease ivory demand [4]. Region-specific wildlife trade has also been a focus of research. For instance, there has been a study quantifying the rise and distribution of illegal trade in Temminck's ground pangolin in Zimbabwe between October 2010 and June 2015 [30]. Furthermore, efforts have been made to evaluate the effectiveness of measures taken to combat this illegal trade, including seizures, legal penalties, and convictions carried out by the Zimbabwe Parks and Wildlife Management Authority, magistrates, and public prosecutors in Zimbabwe.

Sentiment analysis [24, 6] and opinion mining have found applications across various domains. However, their utilization in conservation science remains limited [36], particularly in studies focused on evaluating people's attitudes, opinions, and reactions to species and conservation-related events [19]. Previous research, which didn't employ natural language processing techniques, has examined how conservation topics are portrayed in television, movies [10], news media [15], online search trends [22], and social media [12, 28]. Only a small number of reliable studies have ventured into the analysis of social media data related to conservation using natural language processing methods.

Using a case study approach, a content and sentiment analysis of tweets was conducted about seven South African national parks (Table Mountain, Karoo, Addo Elephant, Pilanesberg, West Coast, Tsitsikamma and Golden Gate Highlands) posted over 14 months on Twitter [45]. The analysis aimed to assess the following aspects: the topics discussed, the authors of the tweets, the timing of the tweets, the emotions conveyed, and the potential managerial implications. To gain insights into public perceptions and interests regarding these seven parks, a quantitative content analysis of the tweet text was conducted using a culturomics approach [20]. First, word counts for all tweets were obtained using Google word cloud. Then using the Tidytext package in R [23, 8], place names, special characters, connecting and stop words were removed. Then the frequencies of variants or plural forms of words referring to similar terms were merged. Finally, tweets were classified and coded into broad topics relating to national parks based on the use of specific terms. Chi-square tests in R [8] were used to test for significant differences in the frequency with which topics were discussed in tweets among the seven parks.

In their study [39], Anna Hausmann et al. utilized automated natural language processing techniques to evaluate the sentiment expressed by visitors when they described their experiences in Instagram posts that were geolocated within four national parks in South Africa. To conduct this analysis, they opted to employ the NRC Word-Emotion Lexicon [14], which is a general lexicon-based approach. This approach involves classifying text based on a predefined dictionary (lexicon) containing words associated with different sentiment polarities. By assessing the occurrence of these words in the text and their corresponding values as annotated in the lexicon, the authors calculated a cumulative value for each sentiment and emotion category for each social media post.

Another work introduces novel application methods to automatically collect and analyze textual data on species of conservation concern from digital platforms [43]. In this work, news articles were compiled together from two channels, namely, Google News and Twitter. After the collation of articles, all data were filtered for duplicates and irrelevant articles using machine learning techniques. Deduplication was performed using a vector comparison algorithm and thus takes into account text similarity of articles rather than just webpage links or titles. Meanwhile, irrelevant articles were removed using a neural network trained to identify such articles from relevant ones. Once the articles are selected, Named Entity Recognition [21, 48] was applied to tag articles and particular sentences with specific information such as names of persons, prices, locations, etc. For the purpose of this study, six types of entities were extracted using the en_core_sm model for English.

Using elephants (Loxodonta Africana, Loxodonta cyclotis, and Elephas maximus) as test species, Hammond et al. conducted a content analysis of tweets about elephants posted to Twitter during 2019 [47]. To classify the emotional sentiment of each remaining tweet, we utilized a well-regarded sentiment analysis tool, VADER: Valence Aware Dictionary and sEntiment Reasoner [17], using the R package vader [42]. Apart from that, a quantitative content analysis of a random sample of tweets was undertaken to understand what was discussed outside of the main events. The authors identified eight topics occurring within our dataset by reading a random sample of tweets. These topics were "elephant welfare concerns," "general conservation messages," "habitat loss," "human-elephant conflict," "poaching and wildlife trade," "tourism," "trophy hunting and culling," and "videos and non-conservation news."

Yufang Gao et al. used mass media coverage as a proxy for macro-level public opinion to analyze the media framing of elephant ivory in 6394 Chinese newspaper articles published from 2000 to 2021 and thus determine the effects of wildlife policies on public opinion [52]. Using Latent Dirichlet Allocation topic modeling [2], we identified 8 topics about elephant ivory and grouped them into 3 frames: ivory arts and culture, ivory crimes, and elephant conservation. Studies showed that topics related to ivory crimes remained the most prevalent in news articles over the past two decades. Contrary to popular belief, results also indicated that Chinese macro-level public opinion on ivory had become more negative following the CITES approval of ivory importation and less negative after the ivory ban announcement, at least for certain periods.

3 Data

3.1 Data Description

The raw datasets that we are using consist of two tables, the first table contains tweet information of 19,606,315 tweets, in different languages, and the second table contains twitter user bio information, that corresponds to the twitter data, which has 452,552 unique users.

Each tweet contains a unique tweet_id, the author who posted it ('author_id') and the language it is written in ('lang') along with the time ('created_at') that the tweet was posted on the platform. Some tweets are also found to be composed of multiple languages, though as explained in the following sections, our models only utilised the tweets in English, for better consistency. In the user description dataset, each user is uniquely identified by a 'user_id'. Each user entry also contains information such as the username, account description, user defined location, the date the account was created, the total number of followers, the total number of accounts followed, and whether the account is verified or not. However, we refrained from utilizing a significant portion of user information due to privacy considerations. The entire dataset can be found at the following Google Drive link here.

After processing we reduced the total number of tweets to 6,296,568 and after analysing this tweet data itself we found there are 2,120,163 unique users who have tweeted, and out of this we have user bios for 288,217 which is around 13.5% of users. All the data remains unlabelled.

3.2 Data Scraping

Given the multilingual nature of the dataset, our focus in this study is solely on English-language tweets. Within the dataset, a substantial portion is comprised of tweets unrelated to the topic of animal rhinos, including references to sports teams known as 'Rhinos.' To address this, a set of 38 selected keywords is utilized to filter out tweets pertaining to animal rhinos, ensuring the precision and relevance of our analysis. As mentioned previously, 6,296,568 English tweets were scraped from the dataset using the scrap words.

3.3 Data Statistics

Stages	Number of Tweets
Initially	19,606,315
After filtering English Tweets	16,383,449
After Scraping	6,296,568

Table 1: Data Statistics of User Tweets

	Number of Users
Total User Bios	452,552
User Bios related to Scraped Tweets	288,217
After Analysis of only Tweet Data without User Bio Data	2,120,163

Table 2: Data Statistics of Unique Twitter Users

3.4 Data Visualisation

In this section we will discuss the interesting findings after analysing the dataset, in the form of visualisations.

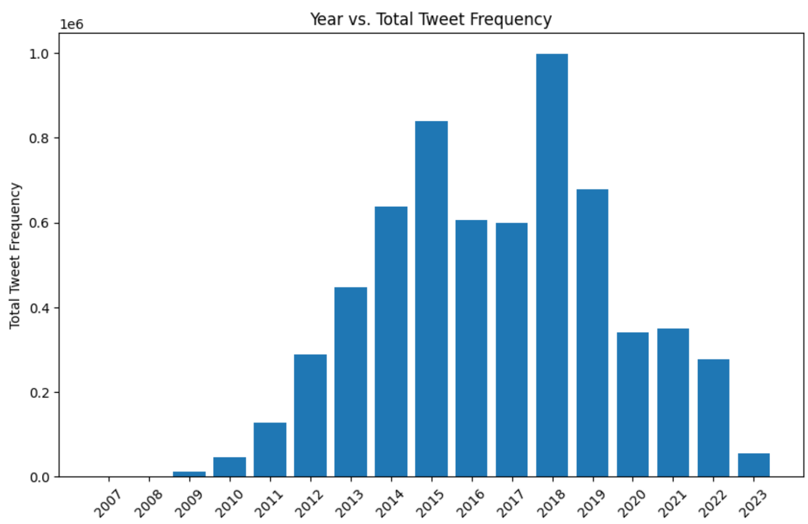


Figure 1: Distribution of tweets

The dataset's total distribution of tweets is shown in Figure 1, which displays the volume of collected tweets per year. Interestingly, the highest frequencies were found between the years 2014 and 2019, with the highest peak in 2018.

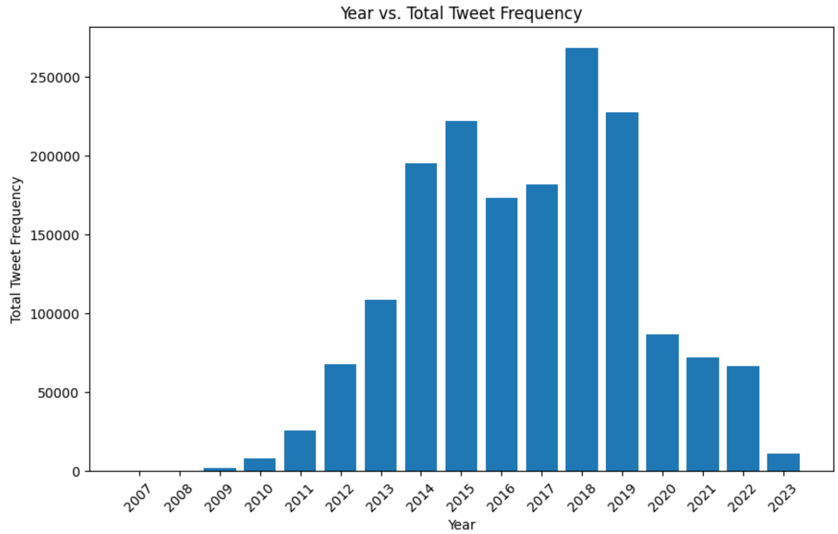


Figure 2: Distribution of Conservation-related tweets

Figure 2 focuses on the subset of conservation-related tweets, indicating the rhino conservation engagement level. Similarly to Figure 1, we found that there was significant activity in conservation-related Tweets between 2014 and 2019. These conservation-related tweets are scrapped using the keywords 'save','conserve/conservation/conserving','awareness','vulnerable','extinction', and 'endanger'.

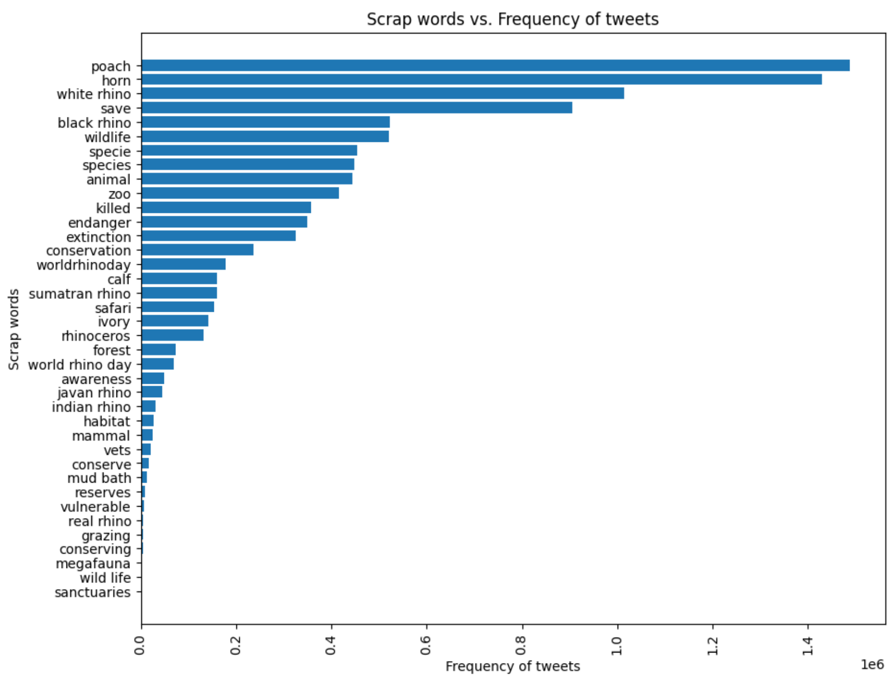


Figure 3: Frequency of scrap words across the tweets

The frequency of scrap words that appear within the dataset is depicted in Figure 3. This is very important, as it reveals the most mentioned words in the context of rhino conservation, across the entire dataset.

3.5 User Activity Data

To prepare our analysis on rhino conservation Twitter user activity, we categorized user engagement levels. This categorization helped us create visual representations of the data: a pie chart showing how often users tweeted and a bar chart identifying the most active users in terms of tweet counts.

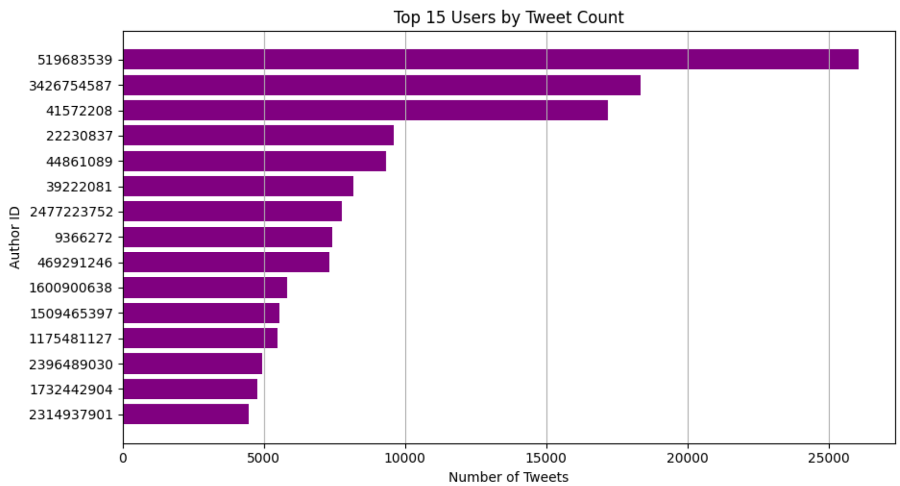


Figure 4: Shows the Top 15 Users by Tweet Count

From the bar chart, which can be seen in Figure 4, the analysis further uncovers that the top 3 users in terms of tweeting activity, originated from Twitter accounts associated specifically with rhino conservation groups. This information was found after matching the author IDs in the tweet dataset with the IDs in the User Bios dataset. Notably, profiles such as HelpingRhinos, SudanTheRhino, and SaveTheRhinos were among the most active in tweet contributions, all possessing more than 15,000 tweets each, with the top contributor, having over 25,000 tweets, a finding that aligns with expectations.

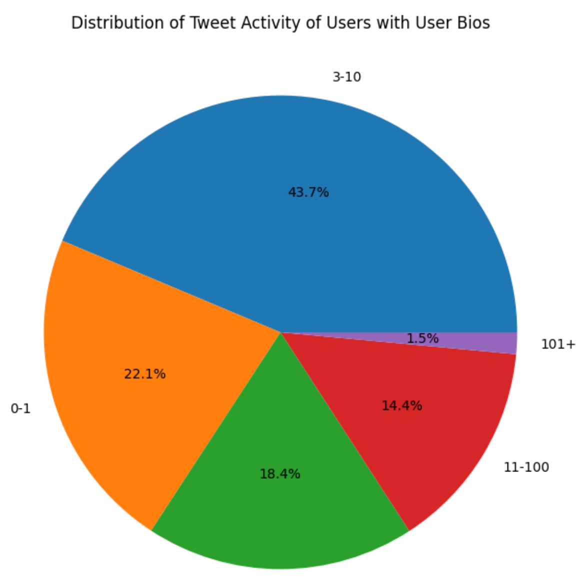


Figure 5: Shows the Distribution of Users having User Bios by Tweet Activity

Figure 5 illustrates the distribution of tweets for users who have user bios, accounting for approximately 288k users.

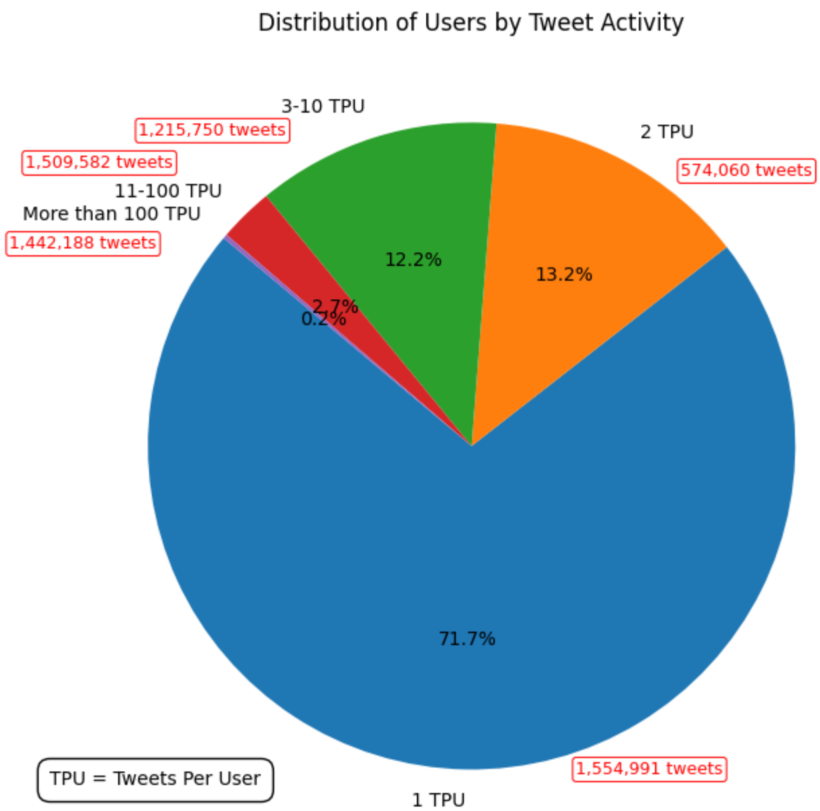


Figure 6: Shows the Distribution of Users by Tweet Activity

The pie chart, as can be seen in Figure 6, shows three metrics, the first metric indicates the different user activity levels, i.e. the number and/or range of tweets per user, which are split into the following sections: once, twice, between three and ten times, between 11 and 100 times, and over 101 times tweets per user within the dataset, which is abbreviated as ”TPU”.

The second metric denotes the percentage of users in each section out of the total number of users, these are the percentages shown in black text in Figure 6. Some very interesting statistics from this, are that the majority of users, i.e. 71.7% (1,554,991 out of 2,120,163), were only single contributors, meaning they were only recorded to have Tweeted once each (1 TPU). Secondly, only 0.2% of users (4,670 out of 2,120,163), were recorded to have Tweeted more than 100 times each (more than 100 TPU). This suggests that while engagement on the topic is widespread, it is often only limited to a singular expressions or instances of interest for certain users.

Finally, the third metric illustrates the total number of tweets contributed by the users in each section, which is the text shown in red in Figure 6. Fascinating statistics from this metric, are that the sections, except the section for 2 tweets per user, each appear to have around 1.5 million tweets in total, and the section for 2 tweets per user has around 500,000 tweets in total. To confirm this, summing all these figures in red, for the number of tweets contributed by section, gives us the total number of tweets in the dataset, which is exactly 6,296,568.

This visual representation reveals an especially striking insight, as it essentially illustrates the Pareto Principle in a unique context. It highlights how a very small percentage of total users, 0.2% (the users with more than 100 tweets each), can have an almost equivalent impact in terms of tweet volume as a vastly larger group of the total users, 71.7% (the users with only single tweet contribution), with both subsets of users contributing 1,554,991 and 1,442,188 tweets respectively. This demonstrates the significant influence of a few highly active users on social media platforms, underscoring the role of power users in shaping online discourse. The comparison between these two vastly different groups, in terms of both size and contribution, provides a fascinating look into user behavior and content creation on social media platforms, and especially in relation to rhino conservation.

3.6 Hashtags Analysis

Analyzing hashtags in rhino-related tweets is important for rhino conservation in gauging public sentiment and awareness, enabling conservation organizations to tailor their outreach efforts. It can help us identify influential trends and campaigns for targeted engagement and fundraising.

In this work, hashtags used along with the frequencies of tweets containing the hashtag have been analyzed. Since different forms of the same word are used as hashtags across the tweets, the 'Stemming' methodology is carried out to group or convert the words in the hashtag to their base or root form.

Before Stemming, there were 173,870 unique hashtags in the tweets. However, after applying stemming, these hashtags were grouped or converted into 134,944 root forms. The hashtag with the root '#rhino' is been found to be most common among the tweets with their presence in 949479 tweets.

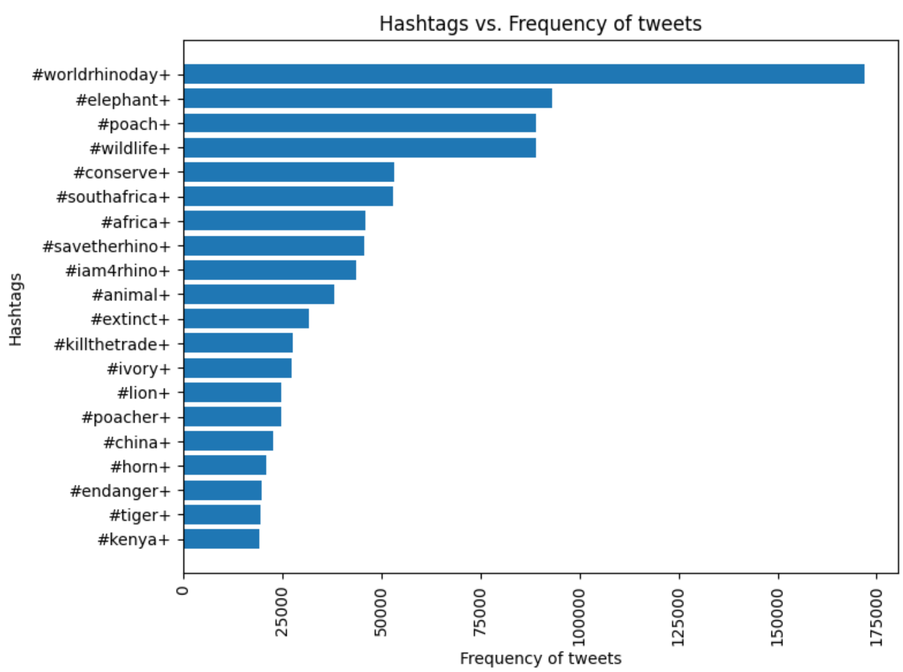


Figure 7: Frequency distribution of top 20 hashtags except the hashtag set '#rhino+'

Figure 7 displays the frequencies of the top 20 hashtags used, excluding '#rhino+'. It is noteworthy that the hashtags associated with conservation awareness, such as "#conserve," "#extinct," "#killthetrade," "#endanger," and "#savetherhino," are prevalent among these hashtags. Additionally, tweets related to other wildlife, such as '#elephant,' '#lion,' and '#tiger,' are commonly tagged alongside rhino-related content.

4 Approaches and Result Analysis

4.1 Sentimental Analysis on Tweets

In this study, we conducted sentiment analysis on approximately 6.2 million English Tweets pertaining to rhino conservation. The objective was to gain insights into the emotional expressions conveyed in these tweets. Twitter, being a widely-used social media platform with a diverse user base, serves as a valuable source for assessing public awareness and engagement in animal conservation issues. By analyzing sentiment, we aimed to identify positive sentiment as an indicator of public support and detect negative sentiment as an area where increased education and outreach efforts may be necessary.

Our initial approach involved utilizing K-means clustering to categorize tweets based on their sentiment. Additionally, we employed voting method using three pretrained models to generate the labels for each tweet and evaluated the performance of Twitter-roBERTa-base.

4.1.1 Data Preprocessing

Data preprocessing for the English rhino-related tweets involves several essential tasks:

1. Eliminating special characters, URLs, and emojis from the text;
2. Converting all words to lowercase for consistency;
3. Normalizing the text by addressing spelling and abbreviation errors;
4. Tokenizing the text to break it down into individual words;
5. Removing stopwords that do not carry significant meaning for analysis

4.1.2 Sentiment Analysis Using K-Mean clustering

We applied K-Means clustering as our baseline model to group the tweets into three clusters. We used TF-IDF (Term Frequency-Inverse Document Frequency) for feature extraction. However, the results have not met our expectations. We observed that tweets sharing common terms like 'poaching' have been grouped together, even though they express different sentiments. As an illustration, tweets such as "funded programme fails to curb poaching of wild rhino in south africa" and "trust campaign launched in namibia to create awareness and save rhinos from poaching" have been clustered together, despite the fact that they express markedly distinct sentiments.

4.1.3 Sentiment Analysis Using Voting Method

We have used the following toolkits/models for Sentiment Analysis:

- **Pattern:** Pattern is a text data-mining tool that has a text-processing component. Its sentiment() function returns a tuple with the polarity and subjectivity of a given text, respectively, where polarity is a value between -1.0 and +1.0 and subjectivity is a value between 0.0 and 1.0.
- **TextBlob:** A Python library that provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation etc.
- **VADER:** VADER(Valence Aware Dictionary for sEntiment Reasoning) [17] is an NLTK module that provides sentiment scores based on the words used. It is a rule-based sentiment analyzer specifically attuned to sentiments expressed in social media, in which the terms are generally labeled as per their semantic orientation as either positive or negative. Vader not only monitors for valence shifters but also assesses emoticons, slang, punctuation, and capitalization. For example, the sentence "I LOVE rhinos" will be assigned a more positive compound score (0.713) than "I love rhinos" (0.637), where scoring ranges from -1 (very negative) to +1 (very positive). The relevant tweets were classified into three categories - "Positive", "Negative" and "Neutral".

- **Twitter-roBERTa-base:** A RoBERTa-base model trained on 124M tweets from January 2018 to December 2021, and finetuned for sentiment analysis with the TweetEval [38] benchmark.

Since our data is not labelled, we had to generate the labels using a voting method. Initially, the label for each tweet is generated using **Pattern, TextBlob** and **VADER**. The final label for each tweet is decided by taking the label occurring the most number of times. Eg, if the three labels generated for a tweet are "POSITIVE", "POSITIVE" and "NEUTRAL", then the final label is considered "POSITIVE"; if there is a tie (eg, NEUTRAL, POSITIVE, NEGATIVE) then that tweet has not been taken into consideration. Once we generated labels foe each of the tweets, the performance of Twitter-roBERTa-base was then evaluated based on the following evaluation metrics: **Accuracy, Precision, Recall, F1-score**.

Our Findings:
Table 3 shows the performance of the Twitter-roBERTa-base model:

Evaluation Metric	Score
Accuracy	57.67%
Precision	0.63
Recall	0.58
F1 Score	0.56

Table 3: Performance Evaluation of Twitter-roBERTa-base

Based on the table above, it may seem that the Twitter-roBERTa-base model fails to provide satisfactory results for classification on the given dataset. However since the data has been annotated using pretrained models, we think that there are cases where a tweet has been labelled incorrectly using the voting method. Eg, a tweet stating "nine poachers arrested in Kenya..." has been labelled as "NEGATIVE" by Pattern and VADER but "POSITIVE" by TextBlob. So, this causes the final label of the tweet to be "NEGATIVE" when it should not have been so. These cases affect the scores even if the Twitter-roBERTa-base model manages to predict the sentiment of the tweet correctly.

4.1.4 Interpretation Using SHAP Scores:

Because of this obscurity in the evaluation of the Twitter-roBERTa-base model, SHAP (SHapley Additive exPlanations) values [29] were used to get a consistent and objective explanation of how each feature impacts the model's prediction. SHAP has specific support for natural language models like those in the Hugging Face transformers library. By adding coalitional rules to traditional Shapley values we can form games that explain large modern NLP model using very few function evaluations [41]. One of the fundamental properties of SHAP values is that they always sum up to the difference between the game outcome when all players are present and the game outcome when no players are present. For machine learning models this means that SHAP values of all the input features will always sum up to the difference between baseline (expected) model output and the current model output for the prediction being explained. The easiest way to visualize this is through a waterfall plot.

Tweet 1: "This baby orphan Black Rhino is an endangered species and needs your help. #phinda #ProtectThePlanet" (Figures 8, 9)

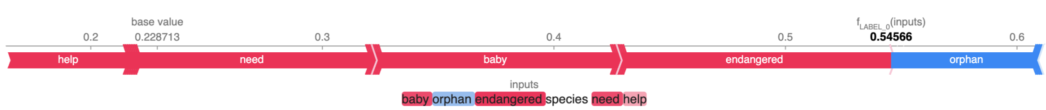


Figure 8: Tweet 1 SHAP Score Plot

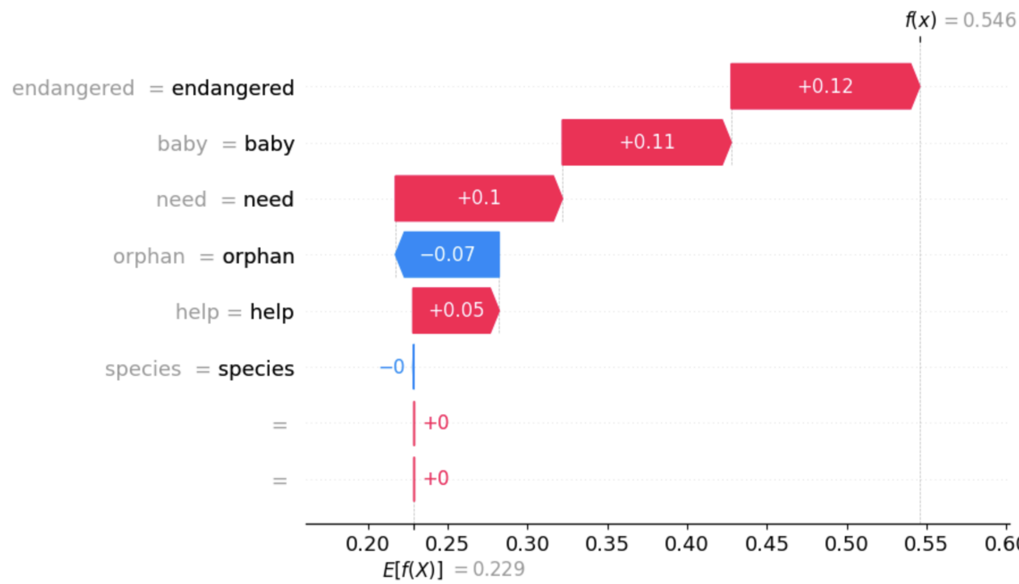


Figure 9: Tweet 1 Waterfall Plot

Tweet 2: “Rhino at risk of an early death due to loneliness and isolation at #Lahore Zoo!” (Figures 10, 11)

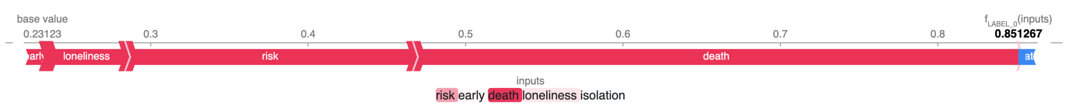


Figure 10: Tweet 2 SHAP Score Plot

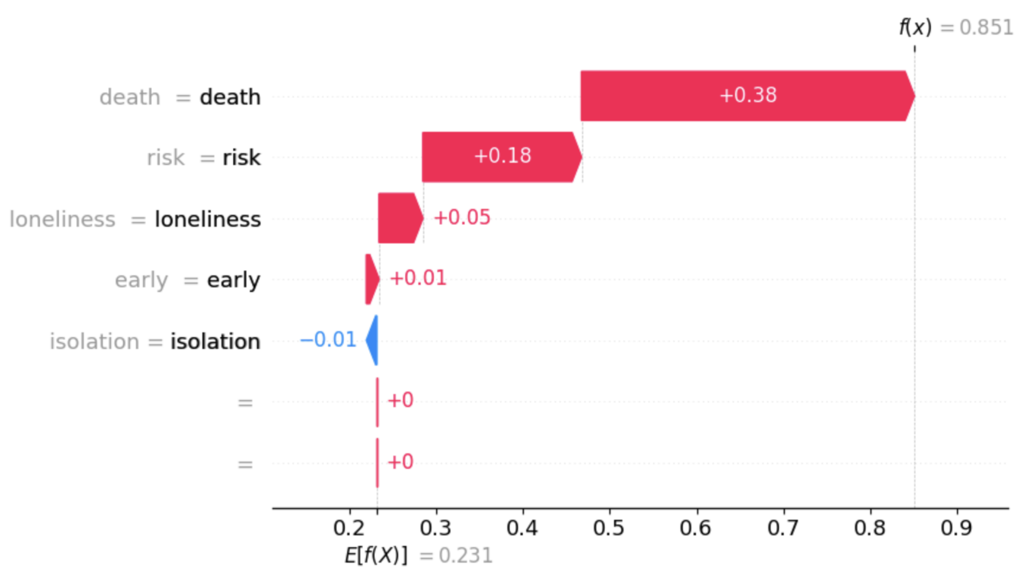


Figure 11: Tweet 2 Waterfall Plot

Tweet 3: “KRUGER NATIONAL PARK RHINO POACHER IN COURT It is alleged that Kipampa is part of a syndicate involved with poaching in the Kruger National Park. Full story: https://t.co/B79FgEDtPg #TzaneenVoice” (Figures 12, 13)

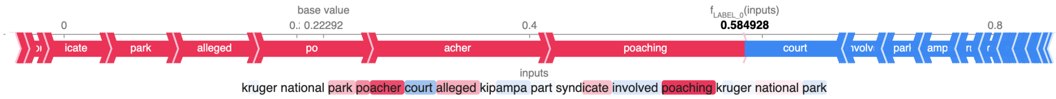


Figure 12: Tweet 3 SHAP Score Plot

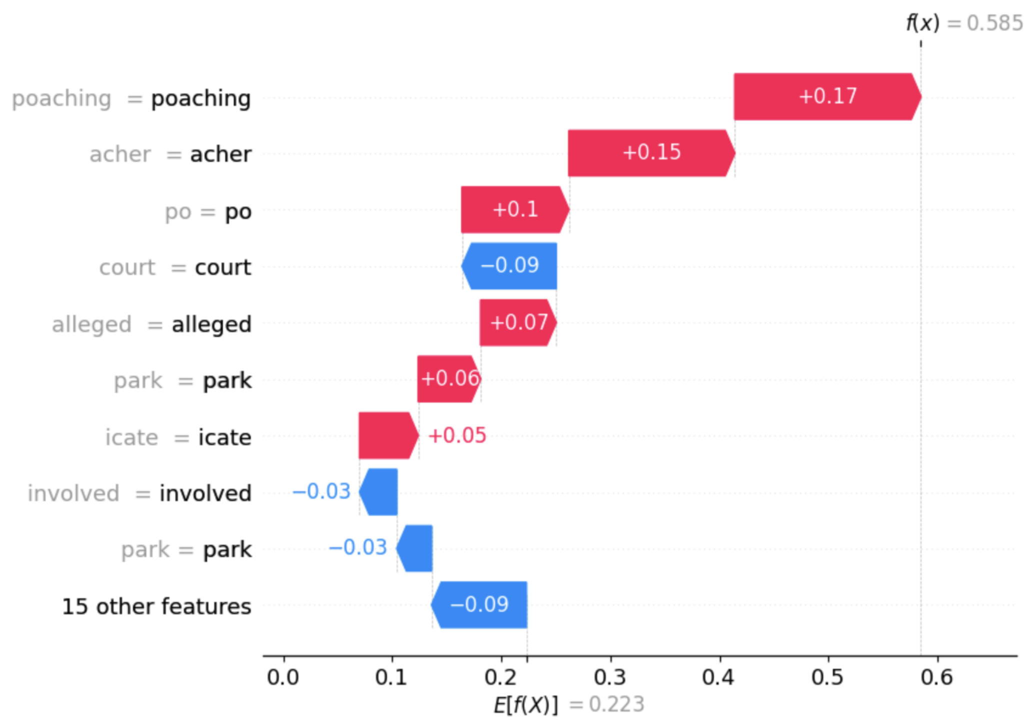


Figure 13: Tweet 3 Waterfall Plot

Tweet 4: “The West African Black Rhino has been officially declared extinct. It was hunted for its horn. Shame on our species.” (Figures 14, 15)

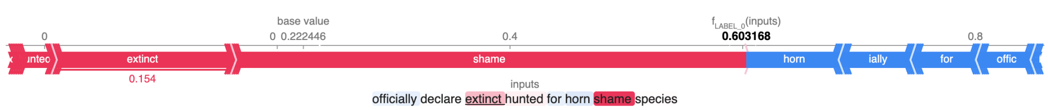


Figure 14: Tweet 4 SHAP Score Plot

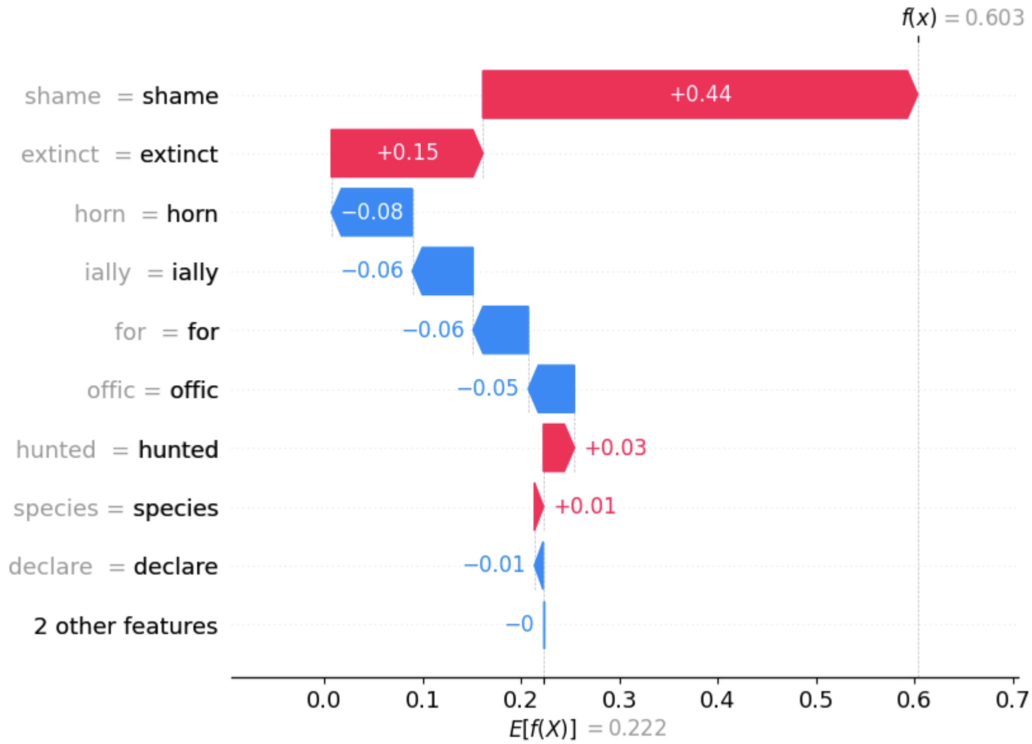


Figure 15: TWEET 4 Waterfall Plot

4.2 Topic Modeling

Topic Modeling (TM) is a Natural Language Processing (NLP) Technique used to identify hidden themes within an extensive collection of textual data, helping to categorize and extract meaningful topics. Various techniques in Topic Modeling (TM) can automatically extract topics from both short texts and standard long-text data. These methods yield reliable results across diverse text analysis domains, including probabilistic latent semantic analysis (PLSA), latent semantic analysis (LSA), and latent dirichlet allocation (LDA). However, numerous existing TM methods face challenges in effectively learning from short texts. Additionally, within Online Social Network (OSN) platforms, TM approaches encounter various issues with short textual data. These challenges encompass slang usage, data sparsity, spelling and grammatical errors, unstructured data, inadequate word co-occurrence information, and the presence of non-meaningful and noisy words [37]. Yan et al. have introduced a short-text TM method known as the biterm topic model (BTM). This approach leverages word correlations or embeddings to enhance the capabilities of TM [16].

In this work, LDA and BTM are used for TM on tweets and user bios. Both the models are evaluated quantitatively by means of the Coherence Score and qualitatively by means of the dominant topics identified.

In topic modeling, coherence scores quantify the interpretability of topics. Topics, represented by the top N words with the highest probabilities, are evaluated based on the similarity of these words. These metric helps to distinguish between topics that are semantically interpretable topics and topics that are artifacts of statistical inference.

In this work, the coherence scores of both LDA and BTM are evaluated using Gensim Library. Two coherence metrics, namely CV (the default in the Gensim library) and UMass, are employed for this evaluation. CV is based on a sliding window, one-set segmentation of the top words and an indirect confirmation measure that uses normalized pointwise mutual information (NPMI) and the cosine similarity whereas UMass is based on document co-occurrence counts, a one-preceding segmentation, and a logarithmic conditional probability as confirmation measure. In this study, the UMass metric is used as the primary metric due to its suitability for document co-occurrence-based confirmation, providing a more robust measure of topic coherence in the context of our analysis. Both metrics are calculated considering topics represented by the top 5, 10, and 20 words.

4.2.1 Data Preprocessing for Topic Modelling

The steps involved in topic modelling for tweets and user bios are as follows,

1. Exclude mentions (user references) from the text.
2. Eliminate special characters and URLs from the text.
3. Exclude texts that are not in English and consist of more than 30% non-standard English words.
4. Change the text to lowercase.
5. Produce part-of-speech (POS) tags for text and retain only the words with specific tags (based on the analysis carried out).
6. Filter out most common words as well anonymous/irrelevant words from the text.
7. Lastly, Lemmatize each filtered word.

Tweets containing more than 30% of non-standard words are removed to improve data quality, consistency, and interoperability. Removing the most common and anonymous/irrelevant helps focus more on specific and meaningful nouns, which leads to more informative and distinct topics, improving the quality and interpretability of the results. Retaining words with specific POS tags in topic modeling is important for enhancing relevance, reducing noise, improving interpretability, and addressing data quality issues. The selective consideration of words with specific tags also contributes to reduced computation time. The choice of specific tags depends on the goals of the analysis and the characteristics of the data being used. Martin et al. study found that lemmatizing and limiting the news corpus to the nouns offers advantages in topic coherence and speed, compared to topic modeling the raw corpus of SJMN articles, or lemmatizing alone [18].

Here, POS tagging and lemmatization are performed using spaCy.

4.2.2 Topic Modeling on Tweets

Topic modeling on Twitter data would be a promising avenue for understanding discussions about rhinos and their conservation. This method facilitates a nuanced understanding of trends and key topics within the Rhino-related discourse on Twitter. During preprocessing, common words like 'rhino', 'animal', 'specie' etc., and irrelevant words like 'rt', and 'amp' are removed from the tweets.

4.2.2.1 Baseline Approach: The K-means clustering technique is the baseline approach for topic modeling with tweets. While K-means is not inherently a topic modeling algorithm, in this work, it is employed to cluster tweets based on the similarity of their feature vectors, often representing word frequency or presence. Only nouns (NOUN) and proper nouns(PROPN) from the tweets are considered. The feature extraction process used TF-IDF (Term Frequency-Inverse Document Frequency) with a feature dimensionality of 100, encompassing unigrams and bigrams.

Under this baseline framework, evaluating the K-means clustering method entails a systematic investigation of cluster numbers ranging from 5 to 15 for randomly selected 100k tweets, with a particular emphasis on assessing cluster quality using silhouette scores. Notably, a consistent pattern emerges within the confines of the baseline methodology, as silhouette scores steadily rise with an increasing number of clusters. The observed increase in cluster size may be attributed to potential over-segmentation, which, in turn, impacts the generalizability of the topic modeling process. Careful consideration of the optimal cluster number is essential to balance granularity and general applicability in clustering tasks.

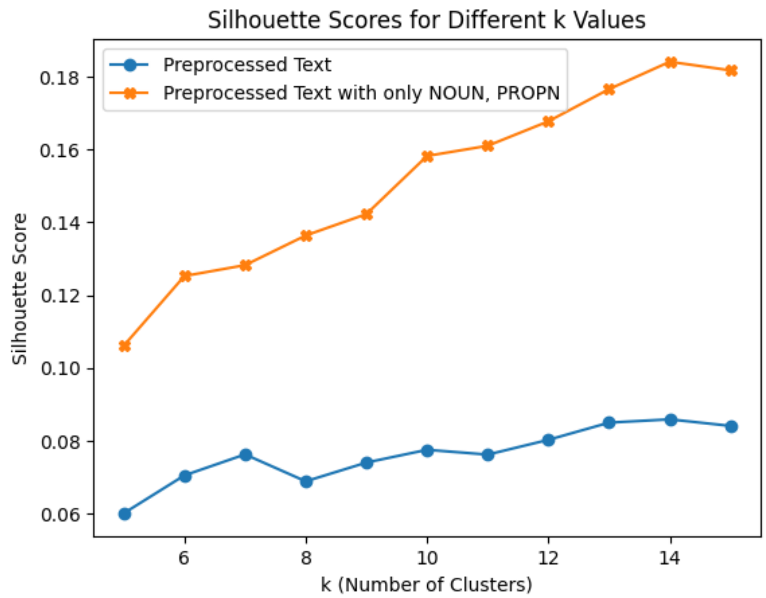
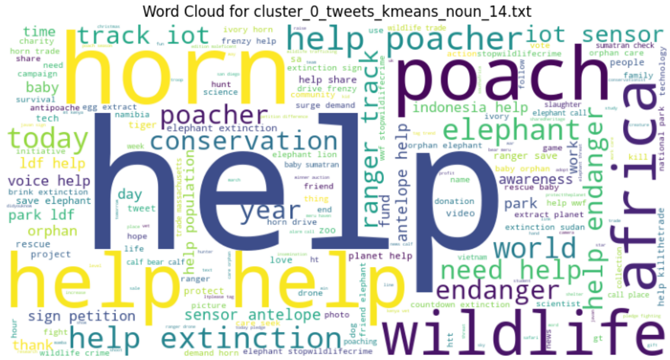


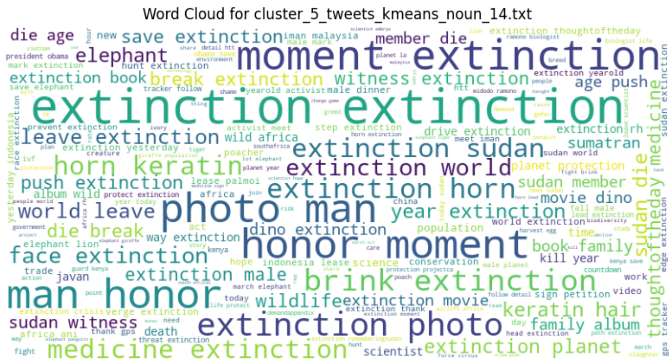
Figure 16: Silhouette Scores Vs. Number of Clusters for Topic Modeling on Tweets using K-means Clustering

Figure 16 illustrates the contrast in silhouette scores between scenarios where only Nouns and Proper Nouns are taken into account and scenarios where all parts of speech are considered. From the figure, it can be seen that using only nouns and proper nouns helps in improving the silhouette scores significantly. However, The Silhouette Scores of the clusters still remain low. The highest Silhouette score achieved is approximately 0.184, observed when considering only Nouns and Proper Nouns and utilizing 14 clusters.

After analyzing the clusters, the model groups tweets about 'horns and trade' into a single cluster, tweets about 'South Africa' into a single cluster, and tweets about 'extinction' into a single cluster. However, some clusters lack informativeness, such as one cluster dedicated to tweets containing the word 'help'. This can be observed from Figure 17. The clusters generated by the model demonstrate limitations in terms of informativeness and interpretability. Specifically, the quality of clusters, as determined by k-means, does not meet the desired standards for meaningful and useful topic delineation.



(a)



(b)

Figure 17: Word clouds of sample clusters from Topic Modeling on Tweets Using K-means with k=14 considering only Nouns and Proper Nouns.

4.2.2.2 Major Approaches: TM models like LDA (models the word occurrences in a document) and BTM (models the biterm occurrences in a corpus) has been utilized in modeling the $\tilde{6}$.2M Tweets. Each of the model is been modeled in two ways

1. Considering only Nouns (NOUN) and Proper Nouns (PROPN)
2. Considering only Nouns (NOUN), Proper Nouns (PROPN), Verbs (VERB) and Adjectives (ADJ)

Considering Nouns and Proper nouns is useful in identifying/ modeling the primary context of the tweets and helps in reducing noise, and maintaining simplicity and interpretability in the generated topics. In the context of Rhino Conservation, the consideration of not only nouns and proper nouns but also verbs and adjectives becomes imperative. Verbs such as 'hunting' and 'poaching' play a pivotal role in understanding the threats faced by rhinos, while adjectives such as 'white', 'black', and 'dehorned' provide crucial descriptors that contribute to the overall context of Rhino Conservation efforts. For this reason, both models LDA and BTM are also modelled by considering only nouns, proper nouns, verbs, and adjectives.

Both models were modeled to uncover nine topics. Leveraging LDA, we experimented with topic counts ranging from 5 to 15. The final choice of nine topics was determined by selecting the number of topics that yielded the highest coherence score. Table 4 depicts the result of quantitative analysis for topic modeling on Tweets with LDA and BTM. It is observed that the BTM model quantitatively outperforms the LDA with respect to both UMass and CV Coherence measures.

Tables 5 and 6 present the topics identified by considering Nouns, Proper Nouns, Verbs, and Adjectives for LDA and BTM. Meanwhile, Tables 7 and 8 showcase the topics identified by considering Nouns and Proper Nouns for LDA and BTM. While both models, LDA and BTM, reveal similar topics, each model introduces some distinct dominant topics. For instance, Table 6 highlights the topic of Rhino Horns - Medicine by BTM, while Table 5 emphasizes the topic of the Loneliness Struggle of Rhinos by LDA.

The inclusion of Verbs and Adjectives alongside Nouns and Proper Nouns proves to be valuable in pinpointing topics related to specific news and rhino-related discussions. Notably, both models captured event-specific topics, such as the Last White Male Northern Rhino Extinction and the Black African Rhino Extinction. The qualitative enhancement achieved by considering Verbs and Adjectives, in addition to Nouns and Proper Nouns, underscores the models' capability to identify a broader spectrum of topics, encompassing both general and event-specific discussions.

Figure 18 and 19 depict the year-wise tweet distribution of each topic. From both graphs, it can be observed that the topic "Last White Male Northern White Rhino Extinction" reaches a peak in 2018, and this peak is significantly higher compared to all other topics. The surge is attributed to the declaration of the extinction of the last male white rhino named Sudan on March 19, 2018. Notably, in 2014, the second last white male rhino died, and in 2015, the last male white rhino was guarded 24 hours. Additionally, 2014 and 2015 marked periods with a higher number of rhinos being poached.

While the West African Black Rhino was declared extinct in 2011, the news garnered widespread attention through tweets and retweets in 2015. This phenomenon is evident in Figures 18 and 19 under the topic "Black African Rhino Extinction". Additionally, the BTM model identifies an event-specific topic "Horn Stealing/Veterinarians." As depicted in Figure 19, it is evident that this topic garnered significant discussion primarily in 2019. This surge in discussion can be attributed to numerous tweets and retweets regarding a rhino whose horn was poached by hunters. The incident gained widespread attention due to accompanying images and clips depicting the rhino's cries when veterinarians attempted to save its life.

Model (POS Tags)	**UMass**			**CV**		
	5	**10**	**20**	**5**	**10**	**20**
LDA (NOUN, PROPN)	-4.9942	-6.0526	-7.7187	0.4595	0.2897	0.2751
LDA (NOUN, PROPN, VERB, ADJ)	-3.5453	-4.5691	-5.7713	0.5541	0.4063	0.2802
BTM (NOUN, PROPN)	-2.7616	-3.5083	-4.3234	**0.6288**	0.4395	0.2451
BTM (NOUN, PROPN, VERB, ADJ)	**-2.3550**	**-2.8250**	**-3.6007**	0.6162	**0.5444**	**0.3984**

Table 4: Topic Coherence Scores considering topn = 5, 10, 20 words per topic for Tweets.

Dominant Topics	Top 10 Related Words	Top 5 Frequent Words	Example Tweet
Topic 1 (Last White Male Northern Rhino Extinction)	white, last, male, world, protect, extinction, leave, sudan, htt, kenya	white, last, male, world, sudan	"With the passing away of the last male northern white rhino, there are now just two female northern white rhinos alive"
Topic 2 (Trophy Hunting/ Elephants)	elephant, hunter, need, pay, see, more, be, trophy, knowlton, corey	elephant, lion, black, poacher, tiger	"Black Rhino Extinct? Texas Hunter Corey Knowlton Pays $350k To Kill A Rhino And Save The Species http://t.co/S2DFE39FKA #BanTrophyHunting"
Topic 3 (Horn/ Ivory Trade)	horn, ivory, trade, say, level, stuff, ban, think, vietnam, man	horn, trade, ivory, ban, south	"Legal rhino horn and ivory trade should benefit Africa, says Swaziland government - The Guardian https://t.co/r170fxW0ER"
Topic 4 (Loneliness Struggle/ Sumatran Rhino)	save, help, loneliness, thank, sumatran, work, follow, ht, tweets, today	save, help, sumatran, zoo, thank	"Rhino at risk of an early death due to loneliness and isolation at #Lahore Zoo! Please do something to save him"
Topic 5 (Endangered or orphan Rhinos/ News)	endanger, find, wildlife, orphan, news, calf, week, old, post, dead	wildlife, endanger, calf, black, news	"This baby orphan Black Rhino is an endangered species and needs your help. #phinda #ProtectThePlanet"
Topic 6 (Poachers/ National Park)	kill, poacher, africa, south, baby, picture, poach, shoot, stop, park	poacher, kill, south, africa, poach	"KRUGER NATIONAL PARK RHINO POACHER IN COURT It is alleged that Kipampa is part of a syndicate involved with poaching in the Kruger National Park. Full story: https://t.co/B79FgEDtPg #TzaneenVoice"
Topic 7 (Black African Rhino Extinction)	black, african, extinct, hunt, declare, west, armed, arrest, sad, innocent	black, extinct, african, horn, declare	"The West African Black Rhino has been officially declared extinct. It was hunted for its horn. Shame on our species."
Topic 8 (Rhino Conservation/ Texas Hunting)	shame, hour, conservation, look, hope, day, texas, have, endangered, get	horn, save, have, life, day	"Your show on the rhino hunt was irresponsible and inaccurate! #KillingIsNOTConservation! #SHAME! https://t.co/SKL1nsVdzf,"
Topic 9 (Poaching/ Anti-poaching)	guard, year, poach, poaching, survive, ranger, namibia, victim, record, Mozambique	poach, year, poaching, horn, white	"In 2022, South Africa's anti-poaching efforts led to a decline in #rhino poaching numbers."

Table 5: LDA Topic Modeling on Tweets - NOUN, PROPN, VERB, ADJ

Dominant Topics	Top 10 Related Words	Top 5 Frequent Words	Example Tweet
Topic 1 (Wildlife Poaching/ Conservation/ World Rhino Day)	poach, wildlife, poaching, conservation, save, day, world, help, new, great	poach, save, poaching, conservation, day	"Today is World Rhino Day! Poaching is devastating Rhino populations in Africa and Asia. Let's help save these amazing creature"
Topic 2 (Other Wildlife Species)	elephant, save, tiger, have, help, lion, wild, live, india, other	save, elephant, help, lion, tiger	"Nepal has made progress in recent years to conserve its charismatic mammals like tigers and rhinos, but similar attention now needs to be paid to orchid conservation. Otherwise, we may soon find that these flowers will only exist in pictures. https://t.co/vNt4I2YZnJ"
Topic 3 (Black African Rhino Extinction)	black, year, extinct, african, hunt, kill, horn, white, declare, see	black, extinct, hunt, year, declare	"The West African Black Rhino has been officially declared extinct. It was hunted for its horn. Shame on our species."
Topic 4 (Rhino Horn - Medicine)	horn, black, make, have, white, think, medicine, get, people, go	horn, white, black, make, have	"We're getting into "rhino horn medicine" territory."
Topic 5 (Rhino Killing/ Poachers/ Other Wildlife Species)	poacher, kill, south, horn, africa, lion, poach, eat, elephant, park	poacher, south, kill, horn, africa	"Poachers killed more than 190 rhinos in Assam between 2000 and 2021 but none was killed last year, according to data shared"
Topic 6 (Horn/ Ivory Trade)	horn, trade, ivory, elephant, save, south, africa, ban, stop, china	horn, trade, elephant, save, ivory	"Legal rhino horn and ivory trade should benefit Africa, says Swaziland government - The Guardian https://t.co/r170fxW0ER"
Topic 7 (Last White Male Northern Rhino Extinction)	white, last, male, world, sudan, die, extinction, leave, photo, say	white, last, male, world, sudan	" With the passing away of the last male northern white rhino there are now just two female northern white rhino alive"
Topic 8 (Horn Stealing/ Veterinarian)	horn, save, life, try, hunter, steal, cry, time, veterinarian, man	horn, save, life, cry, try	"Rhino cries while veterinarians try to save its life after hunters steal its horn ... I feel sick inside."
Topic 9 (Baby/ Calf/ Orphan Rhinos)	baby, zoo, calf, black, white, mother, bear, orphan, poacher, park	baby, zoo, calf, black, white	"During their visit, the King and Queen announced the name of a southern white rhino calf born at Colchester Zoo"

Table 6: BTM Topic Modeling on Tweets - NOUN, PROPN, VERB, ADJ

Dominant Topics	Top 10 Related Words	Top 5 Frequent Words	Example Tweet
Topic 1(Elephants/Wildlife conservation)	elephant, poach, africa, protect, ranger, endanger, south, conservation, wildlife, news	africa, elephant, south, poacher, wildlife	"Human-wildlife conflicts are common across Africa as human populations grow and compete with nature for space. In 2021, due to the deployment of 37 rangers under the @biglifeafrica no rhino or elephant poaching incidents were reported. https://t.co/daFB8j3tja https://t.co/rYv9NxAbjZ"
Topic 2 (Sudan Rhinos/ Botswana Rhinos/ Poaching)	poacher, sudan, work, mother, ht, look, san, rate, protection, botswana	poacher, sudan, world, mother, today	"The last male northern white rhino on Earth! Please support the rangers who protect Sudan from poachers. https://t.co/66ir4LT3uv"
Topic 3 (Hunting)	hour, hunter, help, male, texas, shoot, follow, knowlton, jonathan, sign	help, sign, petition, male, hunter	"No one knows, cares; loves wildlife more than a hunter." Riiiight Corey Knowlton Black Rhino??? #TrophyHuntingLies http://t.co/CHf3lyvr1K
Topic 4 (Medicinal Purpose/ Tiger)	year, htt, place, jones, post, medicine, tiger, hand, stop, end	year, horn, tiger, end, india	"Sadly different places actually believe in odd things like this, it's embedded in their culture and their ancestry. Just like how Rhinos are still being hunted for their horns used in Chinese medicine."
Topic 5 (Guards/ National Parks)	guard, day, park, month, national, man, face, word, pembient, die	park, day, national, guard, poacher	"In India all the rhinos are protected in sanctuaries where there are 24x7 patrol guards and range officers, it's not hard to come across this"
Topic 6 (Others 1)	world, save, orphan, hope, kenya, people, enjoy, tweets, time, camera	world, horn, life, time, cry	"Camera Trap Tuesday Black Rhino Look who popped up on our camera traps! One of our inquisitive black rhino orphans."
Topic 7 (Calf/ Loneliness Struggle/ Others)	baby, loneliness, calf, white, week, today, zoo, video, share, name	zoo, baby, calf, bear, white	"A Rhino Calf's Prayer, by YvdM http://t.co/vfOBx6i5Ad. Capturing the devastation, terror and loneliness an orphaned baby rhino may feel."
Topic 8 (Horny/ Ivory Trade)	horn, hunt, ivory, trade, victim, ban, vietnam, police, d***, sale	horn, trade, ivory, hunt, ban	"Legal rhino horn and ivory trade should benefit Africa, says Swaziland government - The Guardian https://t.co/r170fxW0ER"
Topic 9 (Extinction)	shame, extinction, picture, photo, head, kruger, rh, attack, print, rip	extinction, photo, goodbye, moment, man	"countdown to extinction: only 6 northern white rhinos left on earth. shame on you, human race. http://t.co/DyizbYEoii"

Table 7: LDA Topic Modeling on Tweets - NOUN, PROPN

Dominant Topics	Top 10 Related Words	Top 5 Frequent Words	Example Tweet
Topic 1 (World Rhino day/ Conservation)	world, horn, day, elephant, conservation, save, hunt, africa, endanger, wildlife	world, day, horn, save, conservation	””Anti-Rhino poaching activist, Wayne Bolton’s, cycling adventure is almost complete. His journey is expected to end on World Rhino Day on the 22nd of September” #WorldRhinoDay #RhinoConservation https://t.co/I4Dbv5nm7o”
Topic 2 (Poachers/ Wildlife news)	poacher, work, hope, wildlife, news, kenya, people, follow, horn, elephant	work, hope, poacher, kenya, wildlife	”It’s too late. There are 5 of these particular rhinos left anywhere, and they’re not interested in mating. I hope all poachers rot in hell.”
Topic 3 (Last Northern White Rhino/ Others)	year, end, age, ice, survive, era, poacher, meteor, sudan, str	year, end, age, guard, sudan	”The end of an era.. The northern white rhino which survived 55 million years and saw ice ages, earthquakes, meteor strikes and was testament to innumerable historical changes on the planet could not survive humans.”
Topic 4 (Baby Rhinos/ Last Northern White Rhino and Photos)	horn, poacher, baby, photo, mother, extinction, wildlife, world, assam, park	photo, extinction, horn, baby, goodbye	”Saying goodbye to a species, the very last male Northern White Rhino. A powerful photo of 2018. https://t.co/QynOebPIVd”
Topic 5 (Elephant/ Wildlife Conservation)	elephant, horn, wildlife, poach, city, africa, conservation, drone, trade, day	elephant, horn, poach, wildlife, drone	”Support this anti-poaching campaign and help save #elephants and #rhinos! #conservationbiology https://t.co/yswspeh6yM”
Topic 6 (Horn Stealing/ Veterinarian)	horn, life, man, hunter, steal, cry, time, veterinarian, tiger, f***	horn, life, man, hunter, steal	”For ISU veterinarian and vet students, an endangered black rhinos pregnancy is a (very) big deal - https://t.co/k3LTxSqzIN”
Topic 7 (Baby/Calf Rhinos)	zoo, baby, calf, world, park, bear, sudan, day, sumatran, white	zoo, world, baby, calf, sudan	”During their visit, the King and Queen announced the name of a southern white rhino calf born at Colchester Zoo”
Topic 8 (Poachers/South Africa)	poacher, africa, south, horn, lion, poach, arrest, park, elephant, reserve	poacher, africa, south, horn, poach	”$25 billion offered to South Africa to crack down on Rhino poachers (Natives starving forced to poach to feed villages) ”
Topic 9 (Horn/ Ivory Trade)	horn, trade, elephant, ivory, ban, africa, wildlife, hunt, sign, trophy	horn, trade, elephant, hunt, ivory	”Legal rhino horn and ivory trade should benefit Africa, says Swaziland government - The Guardian https://t.co/r170fxW0ER”

Table 8: BTM Topic Modeling on Tweets - NOUN, PROPN

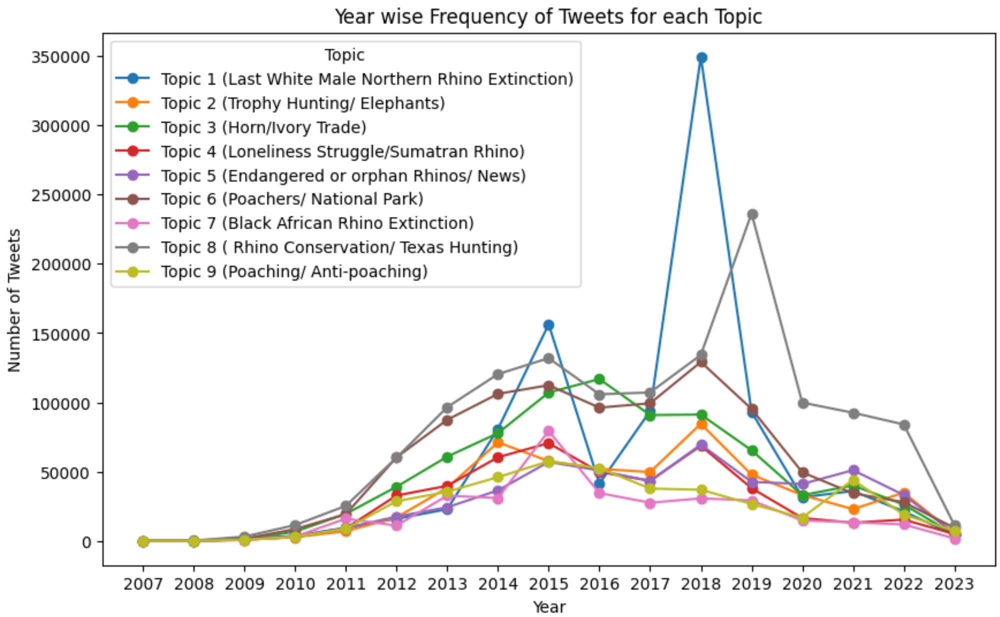


Figure 18: Year-wise Frequency of Tweets for each Topic found through LDA Model (NOUN, PROPN, VERB, ADJ)

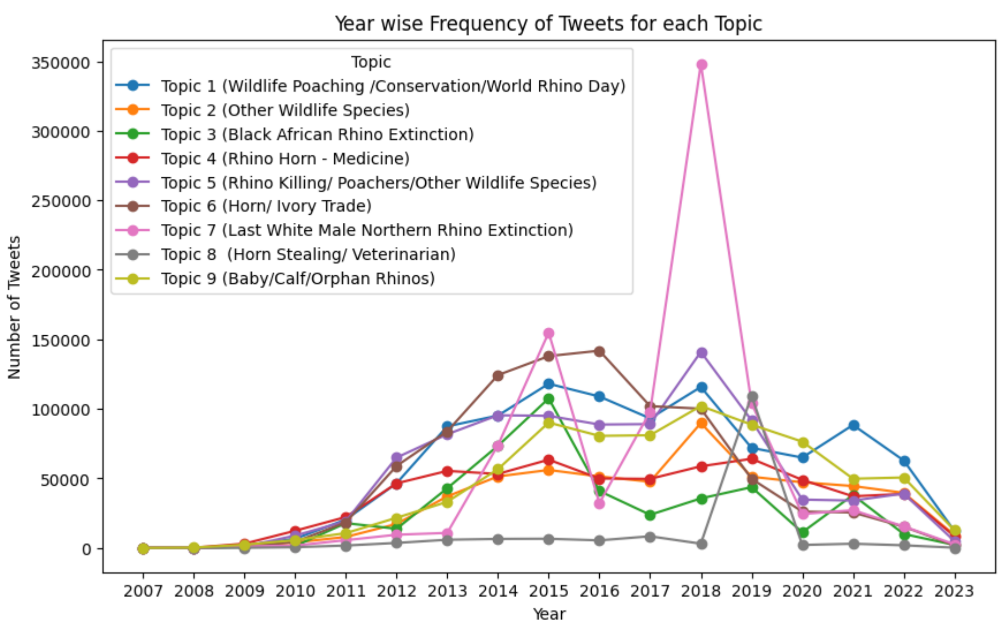


Figure 19: Year-wise Frequency of Tweets for each Topic found through BTM Model (NOUN, PROPN, VERB, ADJ)

4.2.3 Topic Modeling on User Bios

Topic modeling on user bios can reveal shared themes and interests from user profiles, providing insights into user characteristics and affinities. This information is versatile and applicable for community analysis. Specifically, only user bios associated with users in the scraped dataset are considered for this process. This selective approach aims to exclude users who did not tweet about animal rhinos, ensuring a focused exploration of topics related to the target subject matter. During preprocessing, common words like 'love', 'like', or 'fan' etc., and anonymous words like 'rt' or 'dm' etc. are removed from the user bios.

Although there are approximately 2.12 million unique users who have tweeted about rhinos, data on the bios of only around 288k unique users who tweeted about rhinos is available.

4.2.3.1 Baseline Approach: Similar to topic modeling on tweets, K-means clustering is employed as the baseline model for user bios. The baseline approach considers only Nouns and Proper Nouns in user bios. Feature extraction is performed using TF-IDF with a feature dimensionality of 50, encompassing only unigrams.

The evaluation of the K-means clustering method within this baseline framework involves a systematic investigation of cluster numbers ranging from 3 to 7. A particular emphasis is placed on assessing cluster quality using silhouette scores. The highest silhouette score achieved is approximately 0.37, observed when considering only Nouns and Proper Nouns and utilizing 7 clusters.

All the clusters did not provide significant insights into users' occupations, such as whether they are conversationalists or journalists etc. This limitation may arise from the fact that not all users specify their occupations in their bios; instead, many users include their hobbies.

Figure 20 displays word clouds representing sample clusters. In Figure 20(a), it can be observed that user bios in this cluster are predominantly grouped based on the term 'animal.' Conversely, in Figure 20(b), user bios in this cluster are primarily grouped based on the term 'thing,' which does not provide meaningful insights into the users.

The clusters generated by k-means lack distinctiveness and fail to provide comprehensive descriptions of users.

4.2.3.2 Major Approaches: Similar to TM on Tweets, LDA and BTM models are used for TM on user Bios. Each of the model is been modeled in two ways

1. Considering only Nouns (NOUN) and Proper Nouns (PROPN)
2. Considering only Nouns (NOUN), Proper Nouns (PROPN), and Verbs (VERB)

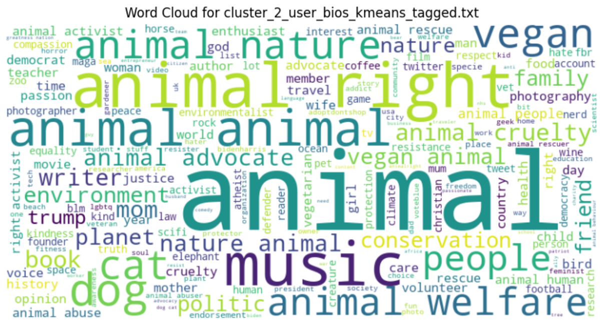
The first approach, focusing solely on Nouns and Proper Nouns, is designed to identify and extract essential primary entities. This method can provide valuable insights into user attributes such as occupation and key terms associated with their bios.

Whereas, the second approach, which includes Verbs in addition to Nouns and Proper Nouns, offers a more comprehensive understanding of user bios. Verbs have the capacity to reveal actions, interests, and activities associated with users, thereby enriching the topics with dynamic elements related to user behavior.

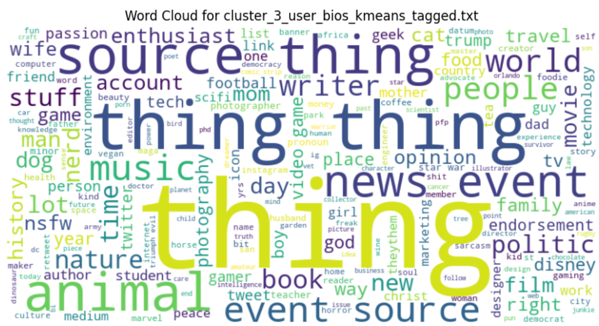
By employing both approaches, the topic modeling process gains versatility. This dual methodology allows for a nuanced exploration of user characteristics and interests encapsulated in bios, offering a more thorough and detailed representation of the diverse facets of user profiles.

Both models were modeled to uncover six topics. Leveraging LDA, we experimented with topic counts ranging from 3 to 10. The final choice of six topics was determined by selecting the number of topics that yielded the highest coherence score. Table 13 depicts the result of quantitative analysis for topic modeling on User Bios with LDA and BTM. It is observed that the BTM model quantitatively outperforms the LDA with respect UMass Coherence measures.

Tables 11 and 12 present the topics identified by considering Nouns, Proper Nouns, and Verbs for LDA and BTM repectively. Meanwhile, Tables 9 and 10 showcase the topics identified by considering Nouns and Proper Nouns for LDA and BTM repectively. Most of the topics identified by LDA and BTM exhibit similarity. Qualitatively, considering only Nouns and Proper Nouns yields more satisfactory results compared to the inclusion of Verbs. The top-related words tend to be populated with common verbs, including unwanted ones such as 'have,' 'do,' 'get,' etc. This presence of generic verbs may contribute to the observed improvement in coherence scores. BTM is slightly to provide better dominant topics like Information Providers/ Journalist, Writers, Convervation/ Environmnet Enthusiast compared to LDA (Table 10).



(a)



(b)

Figure 20: Word clouds of sample clusters from Topic Modeling on User Bios with number of clusters as seven.

Dominant Topics	**Top 10 Related Words**	**Top 5 Frequent Words**	**Example User Bio**
Topic 1 (Tweeters/ Others)	thing, time, account, tweet, twitter, opinion, endorsement, view, way	thing, time, account, tweet, twitter	“Lightly following all things Royal. Retweets are not endorsements”
Topic 2 (Political Enthusiast/Veteran)	god, FORMER_US_PRESIDENT, man, maga, family, country, truth, patriot, peace, president	god, man, FORMER_US_PRESIDENT, maga, family	“I love God, Country our Constitution, Bill of Rights, Liberal turned Conservative. I voted for FORMER_US_PRESIDENT and I will vote for FORMER_US_PRESIDENT again”
Topic 3 (Family-Centric/ Sports Enthusiast)	father, husband, dad, football, music, jesus, rugby, cricket, player, coach	father, husband, football, dad, music	“Husband, father, brother, coach, politically independent, self-employed, cursed Cincinnati sports fan, environment wacko, in that order.”
Topic 4 (Media, News or Business Professional/ Organization)	news, world, business, medium, story, editor, journalist, community, director, founder.	news, world, business, medium, africa	“Former Senior TV News Journalist at SABC. Deputy News Editor and Head of Politics at You FM”
Topic 5 (Others)	writer, game, video, stuff, student, day, engineer, nft, enthusiast, designer	writer, game, nft, enthusiast, day	“Libertarian, Philosopher, Poet, Writer, Gamer, Native Arizonan; Strength of will means letting the waves crash upon you while being unchanged”
Topic 6 (Family-Centric/ Animal Lovers)	animal, music, people, dog, mom, cat, wife, nature, book, mother	animal, music, nature, people, dog	“Animal lover, animal advocate, wildlife enthusiast”

Table 9: LDA Topic Modeling on User Bio - NOUN, PROPN

Dominant Topics	Top 10 Related Words	Top 5 Frequent Words	Example User Bio
Topic 1 (Others)	animal, music, nature, science, history, world, family, politic, dog, view	animal, music, nature, science, history	“ I love most music, all Elvis Presley, all animals, avid Reader, Science, History, football, hockey, am supporter of our military & Troops. Devoted to God”
Topic 2 (Conservationist / Information provider/ Journalist)	news, travel, world, africa, business, medium, conservation, view, twitter, journalist	news, world, travel, business, africa	“Wildlife Conservation Society (East Africa), follow me for news on African, Asian and Marine conservation, plus wildlife updates from the field, views my own.”
Topic 3 (Political Enthusiast/ Animal Lovers/ theist)	animal, god, right, FORMER_US_PRESIDENT, people, maga, world, truth, country, family	animal, god, people, right, world	“CA girl. Loves God, All animals and FORMER_US_PRESIDENT”
Topic 4 (Family-Centric/ Pet Lovers)	dog, mom, music, animal, cat, wife, time, writer, husband, family	dog, mom, animal, wife, music	“Family girl. My Passions are my family, Pets & Travel”
Topic 5 (Writers/Entertainment Pursuits)	writer, game, music, book, nsfw, thing, account, stuff, video, movie	thing, writer, game, music, account	“British Asian. She/They. Writer, gamer. Clarion West 2012 grad”
Topic 6 (Animal Lovers/ Others)	animal, member, writer, music, right, science, dog, mom, nature, teacher	animal, member, writer, teacher, music	“Retired teacher, gardener, animal lover, environmentalist”

Table 10: BTM Topic Modeling on User Bio - NOUN, PROPN

Dominant Topics	Top 10 Related Words	Top 5 Frequent Words	Example User Bio
Topic 1 (Others)	make, thing, time, have, world, do, live, travel, day, get	make, world, thing, have, time	”there are only two things i love in this world: everybody and television.”
Topic 2 (Entertainment Pursuits)	music, game, book, enthusiast, movie, video, rugby, gamer, nerd, ig	music, game, book, enthusiast, movie	”HUGE Dallas Cowboys Fan; MASSIVE Supporter of NHS; Love Diversity & Culture; Food & Music; HUGE Gamer Supporter (JazzyFaux) TOGETHER we can CONQUER the World”
Topic 3 (Political Enthusiast/Family-Centric/Animal Lovers)	animal, FORMER_US_PRESIDENT, maga, dog, mom, family, wife, country, god, mother	animal, dog, right, mom, cat	”Denali Alaska 2015 Antarctica 2013 advocate Intellectual disabled US Civil war, US/Australian politics, love animals Anti FORMER_US_PRESIDENT living in Wodi Wodi land”
Topic 4(Media/ Information Providers/ Journalist)	writer, author, medium, director, editor, journalist, manager, creator, photographer, consultant	writer, author, africa, medium, photographer	”progressive, writer, journalist, editor, educator. Interests: History, politics, psychology, mythology, ethics.”
Topic 5 (News Providers/ Ohers)	news, account, twitter, follow, business, nft, community, science, post, use	news, account, follow, twitter, world	”All latest local News, statistics, match updates ,opinion, and many more.”
Topic 6(Family-Centric / theist)	father, god, husband, view, child, dad, jesus, endorsement, state, university	view, father, husband, god, endorsement	”Husband,Father, Conservative God fearing Patriot!”

Table 11: LDA Topic Modeling on User Bio - NOUN, PROPN, VERB

Dominant Topics	Top 10 Related Words	Top 5 Frequent Words	Example User Bio
Topic 1 (Information Providers/ Journalist)	news, medium, endorsement, view, world, business, twitter, writer, journalist, tweet	news, view, tweet, endorsement, twitter	"Egyptian Blogger / Journalist / Photographer. Just another talkative Egyptian RT is not Endorsement. My views do not represent my employer"
Topic 2 (Political Enthusiast/ Animal Lovers/ theist)	god, animal, FORMER_US_PRESIDENT, maga, family, country, man, father, music, time	god, FORMER_US_PRESIDENT, animal, time, maga	"Former County Chairman (elected),county manager, business owner. Conservative Christian,President FORMER_US_PRESIDENT #MAGA"
Topic 3 (Others)	travel, music, africa, news, world, tour, business, book, photographer, food	travel, music, africa, world, photographer	"Human, woman, US citizen with Constitutional Rights, world traveler, environmental & humanitarian volunteer"
Topic 4 (Family- Centric/ Animal Lovers)	animal, mom, dog, cat, music, wife, mother, right, blm, democrat	animal, mom, dog, cat, wife	"caring, kind, helpful, good sense of humor. Dog mom to Riley Australian Terrier who is very demanding. Lover of all animals."
Topic 5 (Writers/Entertainment Pursuits)	writer, game, music, nsfw, thing, book, account, stuff, enthusiast, video	game, thing, account, writer, nsfw	"Writer/story editor Documentary tv/ GS & IMAX, Strategic Communications/Marketing Content Creator/ Neurodivergent"
Topic 6 (Conservation/ Environment Enthusiast)	animal, science, nature, conservation, right, world, climate, people, health, environment	animal, nature, world, science, conservation	"Forester, Conservationists, Wildlifer, Climate Advocate, Traveler, Volunteer, Event Organizer, Team Work, NPU Ambassador, National Coordinator"

Table 12: BTM Topic Modeling on User Bio - NOUN, PROPN

Model (POS Tags)	**UMass**			**CV**		
	5	**10**	**20**	**5**	**10**	**20**
LDA (NOUN, PROPN)	-3.4086	-3.9009	-4.6429	0.6291	0.5395	0.4023
LDA (NOUN, PROPN, VERB)	-3.2527	-3.8306	-4.2100	**0.6490**	**0.5399**	**0.4485**
BTM (NOUN, PROPN)	-3.2924	-3.6045	-3.9943	0.6190	0.5117	0.4060
BTM (NOUN, PROPN, VERB)	**-3.1432**	**-3.6032**	**-3.9846**	0.6329	0.5349	0.4432

Table 13: Topic Coherence Scores considering topn = 5, 10, 20 words per topic for User Bios.

4.3 Additional Analysis

4.3.1 Sentiment Analysis of Example Tweets Per Topic:

Tables 14 and 15 present the sentiment of example tweets as predicted by the Twitter-roBERTa-base model under each topic identified by LDA and BTM:

Topics	Example Tweet	Predicted Sentiment
Topic 1 (Last White Male Northern Rhino Extinction)	“With the passing away of the last male northern white rhino, there are now just two female northern white rhinos alive”	NEUTRAL
Topic 2 (Trophy hunting/ Elephants)	“Black Rhino Extinct? Texas Hunter Corey Knowlton Pays $350k To Kill A Rhino And Save The Species http://t.co/S2DFE39FKA #BanTrophyHunting”	NEURTAL
Topic 3 (Horn/ Ivory Trading)	“Legal rhino horn and ivory trade should benefit Africa, says Swaziland government - The Guardian https://t.co/r170fxW0ER”	NEUTRAL
Topic 4 (Loneliness Struggle/ Sumatran Rhino)	“Rhino at risk of an early death due to loneliness and isolation at #Lahore Zoo! Please do something to save him”	NEGATIVE
Topic 5 (Endangered or Orphan Rhinos/ News)	“This baby orphan Black Rhino is an endangered species and needs your help. #phinda #ProtectThePlanet”	NEGATIVE
Topic 6 (Poachers/ National Park)	“KRUGER NATIONAL PARK RHINO POACHER IN COURT It is alleged that Kipampa is part of a syndicate involved with poaching in the Kruger National Park. Full story: https://t.co/B79FgEDtPg #TzaneenVoice”	NEGATIVE
Topic 7 (Black African Rhino Extinction)	“The West African Black Rhino has been officially declared extinct. It was hunted for its horn. Shame on our species.”	NEGATIVE
Topic 8 (Rhino Conservation/ Texas Hunting)	“Your show on the rhino hunt was irresponsible and inaccurate! #KillingIsNOTConservation! #SHAME! https://t.co/SKL1nsVdzf,”	NEGATIVE
Topic 9 (Poaching/ Anti-poaching)	”In 2022, South Africa’s anti-poaching efforts led to a decline in #rhino poaching numbers.”	NEGATIVE

Table 14: Sentiment Analysis of Tweets Under Each Topic Provided by LDA

Topics	Example Tweet	Predicted Sentiment
Topic 1 (Wildlife Poaching/ Conservation/ World Rhino Day)	"Today is World Rhino Day! Poaching is devastating Rhino populations in Africa and Asia. Let's help save these amazing creature"	NEUTRAL
Topic 2 (Other Wildlife Species)	"Nepal has made progress in recent years to conserve its charismatic mammals like tigers and rhinos, but similar attention now needs to be paid to orchid conservation. Otherwise, we may soon find that these flowers will only exist in pictures. https://t.co/vNt4I2YZnJ"	NEUTRAL
Topic 3 (Black African Rhino Extinction)	"The West African Black Rhino has been officially declared extinct. It was hunted for its horn. Shame on our species."	NEGATIVE
Topic 4 (Rhino Horn - Medicine)	"We're getting into "rhino horn medicine" territory."	NEUTRAL
Topic 5 (Rhino Killing/ Poachers/ Other Wildlife Species)	"Poachers killed more than 190 rhinos in Assam between 2000 and 2021 but none was killed last year, according to data shared"	NEGATIVE
Topic 6 (Horn/ Ivory Trade)	"Legal rhino horn and ivory trade should benefit Africa, says Swaziland government - The Guardian https://t.co/r170fxW0ER"	NEUTRAL
Topic 7 (Last White Male Northern Rhino Extinction)	"With the passing away of the last male northern white rhino there are now just two female northern white rhino alive"	NEUTRAL
Topic 8 (Horn Stealing/ Veterinarian)	"Rhino cries while veterinarians try to save its life after hunters steal its horn ... I feel sick inside."	NEGATIVE
Topic 9 (Baby/ Calf/ Orphan Rhinos)	"During their visit, the King and Queen announced the name of a southern white rhino calf born at Colchester Zoo"	NEUTRAL

Table 15: Sentiment Analysis of Tweets Under Each Topic Provided by BTM

4.3.2 Tweet Distribution Per Topic For Both LDA and BTM:

Figures 21, 22 present the tweet distribution for each topic generated by LDA whereas Figures 23, 24 depict the same for each topic generated by BTM. These histograms show the count of POSITIVE, NEGATIVE and NEUTRAL tweets under each topic given by LDA and BTM.
LDA: Tweets are majorly NEGATIVE under topics ”Horn/Ivory Trading”, ”Poachers/National Park” and ”Poaching/Anti-poaching”; POSITIVE under the topics ”Loneliness Struggle/Sumatran Rhino” and ”Rhino Conservation/Texas Hunting”; evenly distributed under the topics ”Last White Male Northern Rhino Extinction” and ”Trophy Hunting/Elephants”.
BTM: Tweets are majorly NEGATIVE under topics ”Black African Rhino Extinction”, ”Rhino Killing/Poacher/Other Wildlife Species”, ”Horn/Ivory Trade” and ”Horn Stealing/Veterinarian”; POSITIVE under the topics ”Wildlife Poaching/ Conservation/ World Rhino Day” and ”Other Wildlife Species”; evenly distributed under the topics ”Last White Male Northern Rhino Extinction” and ”Rhino Horn - Medicine”.

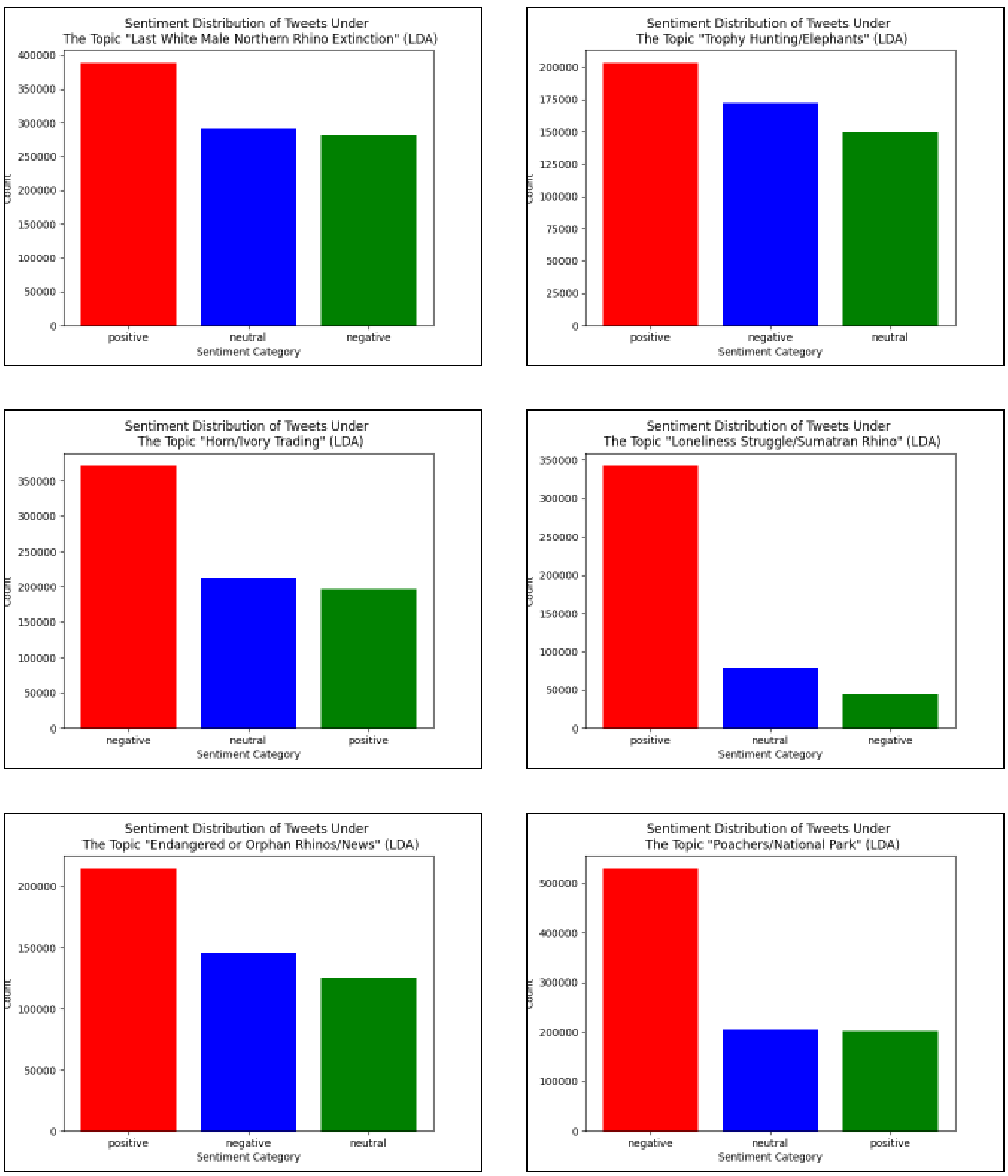


Figure 21: Tweet Distribution Under Each Topic Given By LDA

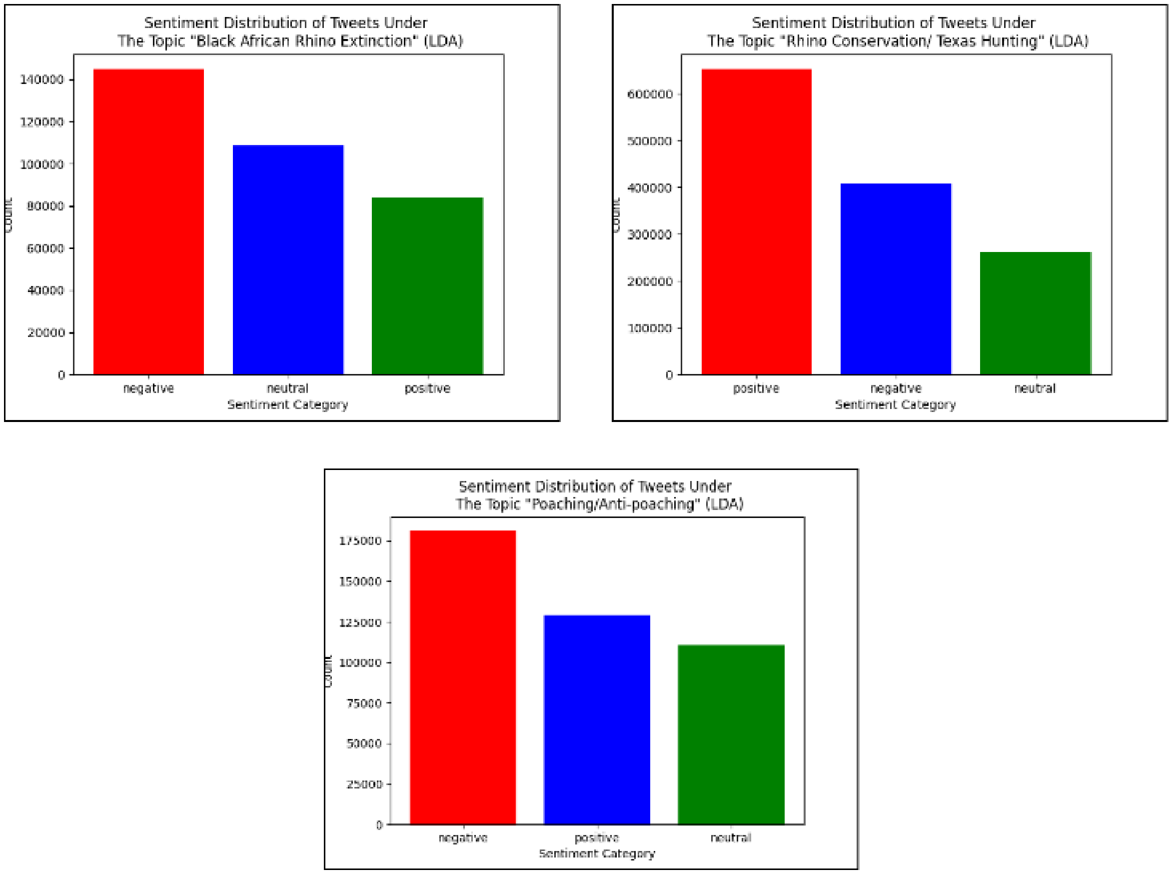


Figure 22: Tweet Distribution Under Each Topic Given By LDA

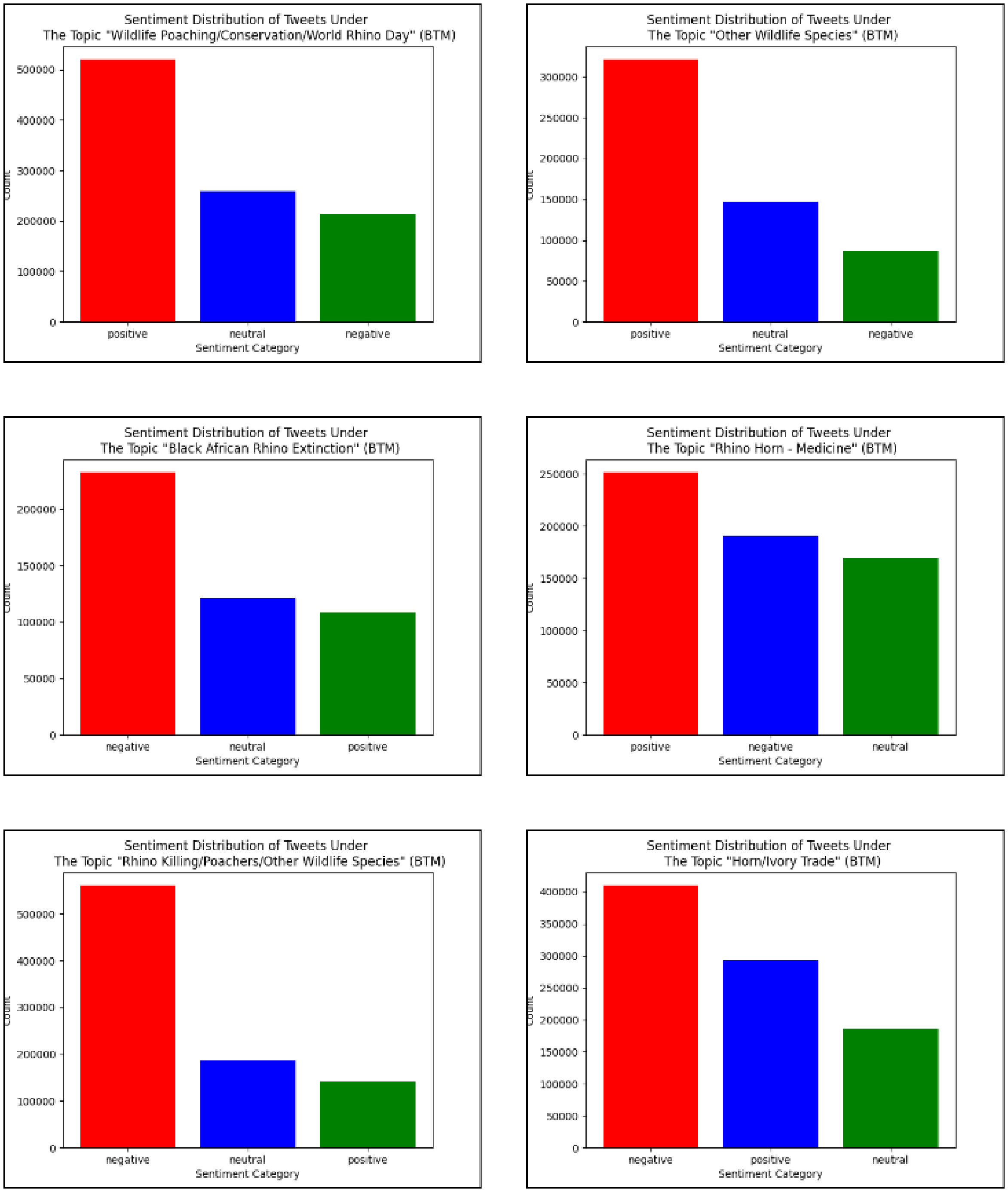


Figure 23: Tweet Distribution Under Each Topic Given By BTM

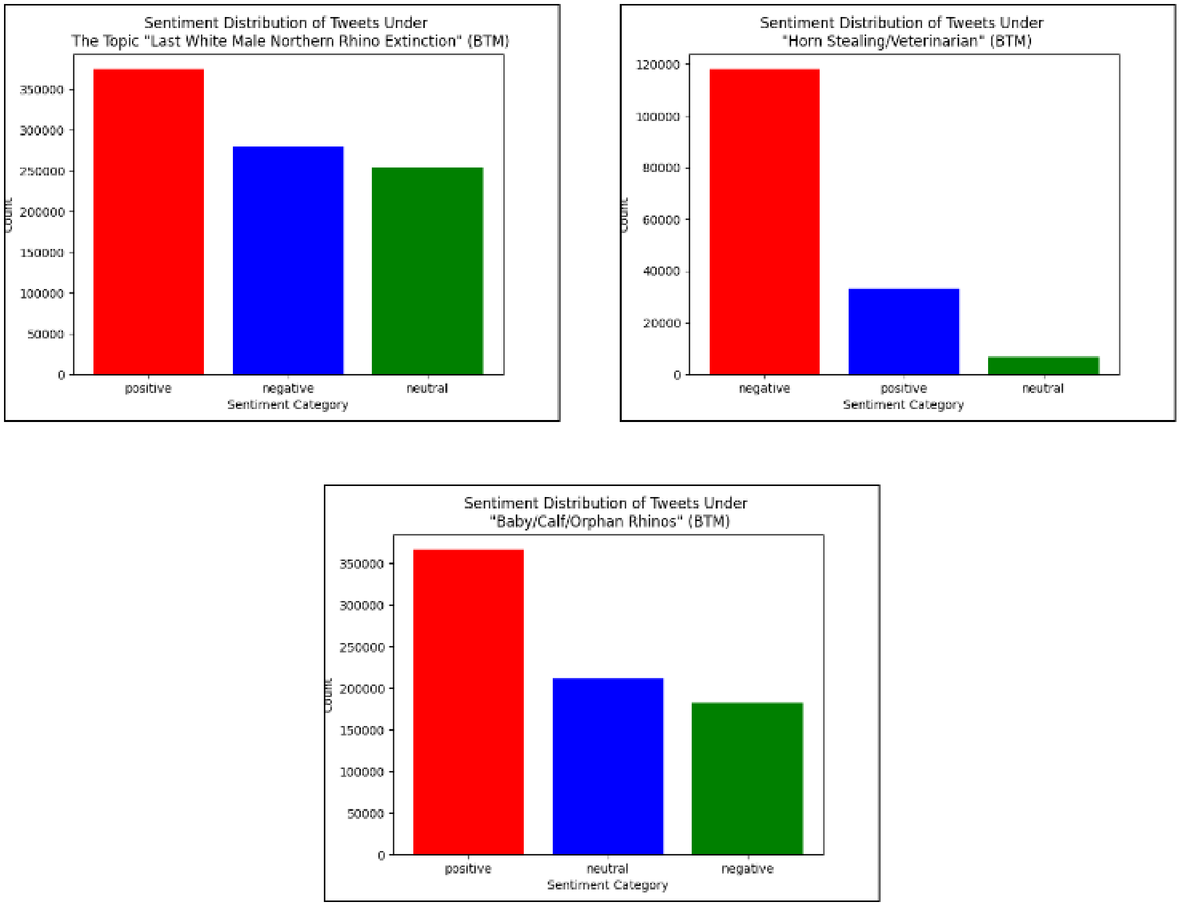


Figure 24: Tweet Distribution Under Each Topic Given By BTM

4.3.3 Topic Modeling on Tweets - Users with 10 or Fewer Tweets

Exploring topics from users with fewer tweets allows for a deeper understanding of how novice or less active users engage with and contribute to discussions on Rhinos. This exploration is crucial for tailoring outreach strategies to effectively reach and engage this specific user segment.

As illustrated in Figure 6, 97% of the users have tweeted less than or equal to 10 times, contributing to only 47% of all tweets, which amounts to approximately 3.3 million tweets. BTM is been employed considering Nouns, Proper Nouns, Verbs and Adjectives to identify the topics. The number of topics to be identified by BTM is set to six. The quantitative result analysis of the model is shown in Table 16. The identified topics are presented in Table 17. Most of the tweets by these users revolve around specific events, such as ’Last White Male Rhino Extinction,’ ’Black African Rhino Extinction,’ and ’Horn Stealing/Veterinarian.’ In comparison, there are fewer tweets discussing rhinos and their conservation in a more general context.

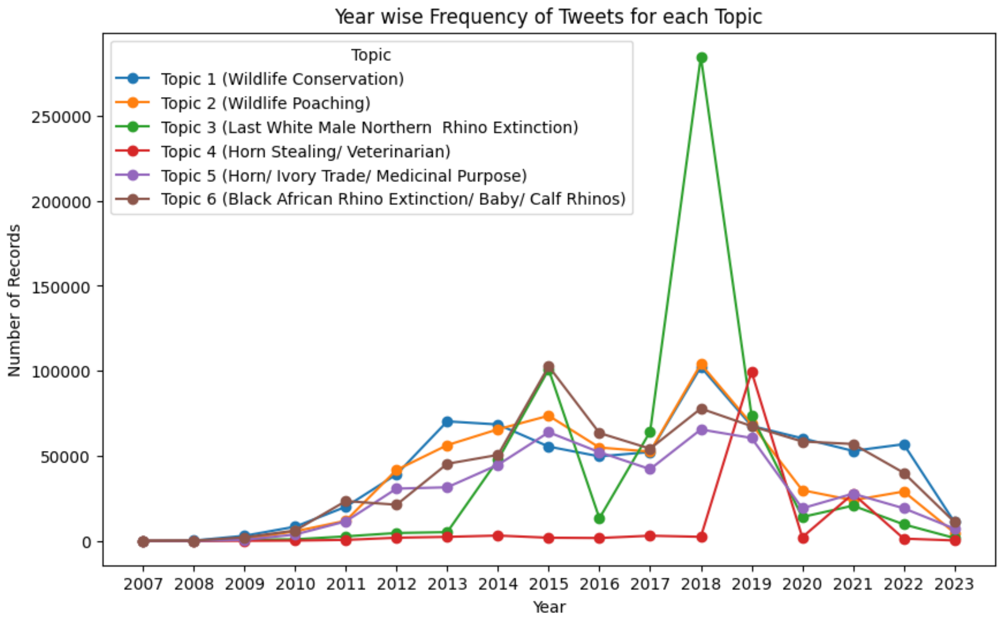


Figure 25: Year-wise Frequency of Tweets for Topic Modeling on Tweets - Users with 10 or Fewer Tweets (BTM Model - NOUN, PROPN, VERB, ADJ)

Figure 25 depicts the year-wise distribution of tweets for these specific topics. Notably, topics like 'Last White Male Rhino Extinction,' 'Black African Rhino Extinction,' and 'Horn Stealing/Veterinarian' exhibit similar patterns to those shown in Figure 19.

Model (POS Tags)	**UMass**			**CV**		
	5	**10**	**20**	**5**	**10**	**20**
BTM (NOUN, PROPN, VERB, ADJ)	-2.1118	-2.5463	-3.7297	0.6783	0.5991	0.3950

Table 16: Topic Coherence Scores considering topn = 5, 10, 20 words per topic for Tweets - Users with 10 or Fewer Tweets

Dominant Topics	Top 10 Related Words	Top 5 Frequent Words	Example Tweet
Topic 1 (Wildlife Conservation)	save, have, white, help, elephant, tiger, day, world, live, make	save, help, white, have, day	”Save our tigers, our pride. Otherwise our future generation will address the Tiger in the same way as today we are addressing the Tapanauli Orungutan, Northern White Rhino and many more.”
Topic 2 (Wildlife Poaching)	poacher, south, kill, lion, africa, elephant, poach, eat, african, reserve	poacher, south, kill, africa, lion	”There is a lot wrong with the anti poaching scene in general. One of the most hardcore units i worked for, who have never ever lost a rhino in their area, are absolutely brilliant elephant poachers themselves and make millions doing it.”
Topic 3 (Last White Male Northern Rhino Extinction)	white, last, northern, male, sudan, world, die, extinction, earth, leave	white, last, northern, male, world	“The last male Northern White Rhino died yesterday. The species has existed for at least 55 million years and survived”
Topic 4 (Horn Stealing/ Veterinarian)	horn, save, try, life, hunter, steal, cry, veterinarian, come, time	save, horn, life, cry, try	”Rhino cries while veterinarians try to save its life after hunters steal its horn ... I feel sick inside.”
Topic 5 (Horn/ Ivory Trade/ Medicinal Purpose)	horn, f***, trade, poacher, kill, medicine, china, make, ivory, mother	horn, poacher, trade, kill, f***	”Legal rhino horn and ivory trade should benefit Africa, says Swaziland government - The Guardian https://t.co/r170fxW0ER”
Topic 6 (Black African Rhino Extinction/ Baby/ Calf Rhinos)	black, zoo, baby, horn, extinct, calf, white, park, african, first	black, baby, zoo, extinct, white	” I’m sorry to tell you its too late. The black African rhino has been declared #extinct #extinctionisnow #ExtinctionIsForever”

Table 17: BTM Topic Modeling on Tweets: Users with 10 or Fewer Tweets - NOUN, PROPN, VERB, ADJ

5 Conclusion

In conclusion, the exploratory data analysis conducted on rhino-related tweets unveiled notable trends. The volume of such tweets exhibited a comparative increase from 2014 to 2019, reaching a peak in 2018. Poaching and Rhino horn-related discussions emerged as prominent themes within these conversations. Hashtag analysis further underscored the prevalence of conservation-related tags, indicating a significant focus on efforts to protect rhinos. Additionally, the co-occurrence of hashtags with other animal-related terms suggests a broader context of wildlife conservation discussions on the platform.

Our current sentiment analysis approach, based on a voting method, falls short of our expectations. To enhance the data annotation process, we are considering two potential improvements: one is employing models that have been pretrained on Twitter data for voting method, and the other involves exploring the option of manual labeling. The results we have obtained so far do not provide a clear indication of the effectiveness of relying on pretrained NLP tools for assessing the sentiment of tweets in conservation-related discussions.

Our approach to topic modeling, focusing on retaining words associated with specific tags, proved effective in enhancing the informativeness of topics in both Tweets and User Bios. The quantitative analysis revealed that BTM outperformed LDA in coherence scores for both Tweets and User Bios. Qualitative analysis further highlighted similarities in identified topics between LDA and BTM, yet each model presented distinct dominant topics, offering valuable insights. This combined quantitative and qualitative evaluation underscores the effectiveness of BTM in extracting meaningful and coherent topics from text data.

The topic modeling analysis revealed a dual nature of discussions on rhinos, encompassing both general themes such as 'Poaching' and 'Horn/Ivory Trade', as well as event-specific topics like the 'Last White Male Northern Rhino Extinction' and 'Horn Stealing/Veterinarian.' Notably, a substantial proportion of tweets were event-based, with the extinction of the Last White Male Northern Rhino in 2018 emerging as the most trending topic in the context of rhinos.

Remarkably, a small fraction of users, constituting only 3% contributed significantly and were responsible for 53% of the overall tweets. Conversely, approximately 97% of users engaging in rhino-related discussions were infrequent tweeters, posting ten or fewer tweets. These infrequent users predominantly shared retweets, and their tweets were predominantly event-specific. This distribution sheds light on the concentration of engagement within a minority of active users, with the majority participating in the discourse around specific events. Also, There is been a significant drop in tweets after 2019 related to rhinos.

Despite the limited dataset comprising user bios of approximately 13.5% of total unique users, the topic modeling analysis on user bios has provided insights into the diverse backgrounds of individuals engaging in discussions about rhinos on social media. Among those identified, the user categories encompass animal lovers, journalists, writers, conservationists, political enthusiasts, and so on.

Overall, this multifaceted analysis contributes to a richer comprehension of the online narrative surrounding rhinos, emphasizing the importance of considering temporal trends, key themes, and user engagement patterns. These findings not only inform our understanding of public interest in rhino conservation but also offer valuable insights for conservation organizations and policymakers seeking to leverage online platforms for effective outreach and awareness.

5.1 Limitations

The limitations of this work could include,

1. Despite preprocessing the data, certain unwanted characters/words such as 'htt' and 'ht' etc. are still present and observable in the Topic Modeling results.
2. Topic Modeling on User Bios is performed with limited data (around 13.5% of the actual number of users).
3. The method used for annotating data for sentiment analysis is not reliable.

4. Cross-analysis couldn't be carried out due to limitations with Sentiment Analysis and the limited availability of User Bios.

5.2 Possible extensions

Some of the Future scope of this work could include,

1. Implementing stricter data scraping techniques by considering the exclusion of scrap words.
2. Exploring more or different preprocessing techniques for improved data preparation.
3. Enhancing topic modeling, especially on user bios when considering verbs, by eliminating common verbs.
4. Conducting topic modeling on both users and tweets with a higher number of topics and tuning hyperparameters.
5. Obtaining more particular topics in context with the initial one, topic modeling can be done within the tweets that correspond to a topic on the initial modeling.
6. Performing topic modeling on tweets based on a specific type or category of users.
7. Using models specifically pretrained using twitter data for voting method.
8. Fine-tuning the roBERTa model using the given dataset for Sentiment Analysis.

6 Contributions

- **Swapnil Mallick:**
  - Prepared data by initially segregating tweets that are in English language, implemented data preprocessing for sentiment analysis, implemented baseline method of K-means for sentiment analysis, generated labels for sentiment analysis using voting method, evaluated the performance of Twitter-roBERTa-base, tried interpreting the behavior of Twitter-roBERTa-base model prediction using SHAP values, generated sentiment of example tweets under each topic generated by both LDA and BTM, calculated tweet distribution under each topic generated by LDA and BTM.
  - Responsible for writing the following sections of the report: Abstract, Introduction (1), Related Works and Background (2), Data Description (3.1) under Data (3), Sentiment Analysis of Tweets (4.1 - 4.1.1, 4.1.2, 4.1.3, 4.1.4) under Approaches and Result Analysis (4), Sentiment Analysis of Example Tweets per Topic (4.3.1), Tweet Distribution Per Topic For Both LDA and BTM (4.3.2) under Additional Analysis (4.3), parts of Conclusion (5), parts of Limitations (5.1) and parts of Possible Extensions (5.2).
- **Venkatesh Dharmaraj:**
  - Data Scraping for rhino related tweets; EDA and Hashtag Analysis; Data Preprocessing for Topic Modeling; Topic modeling on Tweets and User Bios Using K-means, BTM and LDA; Topic modeling on tweets for users who tweeted 10 or fewer times; Report - Data Scraping (3.2), Data Statistics (3.3), Data Visualization (3.4), part of User Activity Data (3.5), Hashtag Analysis (3.6) under Data (3), Topic Modeling (4.2 - 4.2.1, 4.2.2, 4.2.3) under Approaches and Result Analysis (4), Topic Modeling on Tweets - Users with 10 or Fewer Tweets (4.3.3) under Additional Analysis (4.3), parts of Conclusion (5), parts of Limitations (5.1) and parts of Possible Extensions (5.2).
- **Kiran Sampat:**
  - Extracted user bios that posted tweets related to rhinos; Carried out user tweet analysis; Assisted with report formatting; Report - part of User Activity Data (3.5) under Data (3).
- **Samarth Khaire:**
  - Worked on scrapping the tweets related to animal rhino; Tried to implement topic modeling for user bios.

References

[1] C Margules and MB Usher. “Criteria used in assessing wildlife conservation potential: a review”. In: *Biological conservation* 21.2 (1981), pp. 79–109.

[2] David M Blei, Andrew Y Ng, and Michael I Jordan. “Latent dirichlet allocation”. In: *Journal of machine Learning research* 3.Jan (2003), pp. 993–1022.

[3] J Donald Hughes. “Europe as consumer of exotic biodiversity: Greek and Roman times”. In: *Landscape Research* 28.1 (2003), pp. 21–31.

[4] Daniel Stiles. “The ivory trade and elephant conservation”. In: *Environmental Conservation* 31.4 (2004), pp. 309–321.

[5] TRAFFIC. “What’s Driving the Wildlife Trade? A Review of Expert Opinion on Economic and Social Drivers of the Wildlife Trade and Trade Control Efforts in Cambodia, Indonesia, Lao PDR and Vietnam”. In: *East Asia and Pacific Region Sustainable Development Discussion Papers. East Asia and Pacific Region Sustainable Development Department, World Bank, Washington, DC.* (2008).

[6] Rudy Prabowo and Mike Thelwall. “Sentiment analysis: A combined approach”. In: *Journal of Informetrics* 3.2 (2009), pp. 143–157.

[7] Gail Emilia Rosen and Katherine F Smith. “Summarizing the evidence on the international trade in illegal wildlife”. In: *EcoHealth* 7 (2010), pp. 24–32.

[8] R Developement Core Team. “R: A language and environment for statistical computing”. In: *(No Title)* (2010).

[9] Tom Milliken and Jo Shaw. “The South Africa-Viet Nam rhino horn nexus”. In: (2012).

[10] Gregg Mitman. *Reel nature: America’s romance with wildlife on film*. University of Washington Press, 2012.

[11] Julie Ayling. “What sustains wildlife crime? Rhino horn trading and the resilience of criminal networks”. In: *Journal of International Wildlife Law & Policy* 16.1 (2013), pp. 57–80.

[12] Bram Büscher and Jim Igoe. “‘Prosuming’conservation? Web 2.0, nature and the intensification of value-producing labour in late capitalism”. In: *Journal of Consumer Culture* 13.3 (2013), pp. 283–305.

[13] Gwynn Guilford. “Why does a rhino horn cost $300,000? Because Vietnam thinks it cures cancer and hangovers”. In: *The Atlantic* 15 (2013). URL: http://www.theatlantic.com/business/archive/2013/05/why-does-a-rhinohorn-cost-300-000-because-vietnam-thinks-it-cures-cancer-and-hangovers/275881/.

[14] Saif M Mohammad and Peter D Turney. “Nrc emotion lexicon”. In: *National Research Council, Canada* 2 (2013), p. 234.

[15] Bret A Muter et al. “Australian and US news media portrayal of sharks and their conservation”. In: *Conservation Biology* 27.1 (2013), pp. 187–196.

[16] Xiaohui Yan et al. “A Biterm Topic Model for Short Texts”. In: (2013), pp. 1445–1456. URL: https://doi.org/10.1145/2488388.2488514.

[17] C. Hutto and Eric Gilbert. “VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text”. In: *Proceedings of the International AAAI Conference on Web and Social Media* 8.1 (May 2014), pp. 216–225. DOI: 10.1609/icwsm.v8i1.14550. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14550.

[18] Fiona Martin and Mark Johnson. “More Efficient Topic Modelling Through a Noun Only Approach”. In: (Dec. 2015). Ed. by Ben Hachey and Kellie Webster, pp. 111–115. URL: https://aclanthology.org/U15-1013.

[19] Margreet Drijfhout et al. “Sentiment Analysis: ready for conservation”. In: *Frontiers in Ecology and the Environment* 14.10 (2016), pp. 525–526.

[20] Richard J Ladle et al. “Conservation culturomics”. In: *Frontiers in Ecology and the Environment* 14.5 (2016), pp. 269–275.

[21] Guillaume Lample et al. “Neural Architectures for Named Entity Recognition”. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Ed. by Kevin Knight, Ani Nenkova, and Owen Rambow. San Diego, California: Association for Computational Linguistics, June 2016, pp. 260–270. DOI: 10.18653/v1/N16-1030. URL: https://aclanthology.org/N16-1030.

[22] Le TP Nghiem et al. “Analysis of the capacity of Google Trends to measure interest in conservation topics and the role of online news”. In: *PloS one* 11.3 (2016), e0152802.

[23] Julia Silge and David Robinson. “tidytext: Text mining and analysis using tidy data principles in R”. In: *Journal of Open Source Software* 1.3 (2016), p. 37.

[24] Maite Taboada. “Sentiment analysis: An overview from linguistics”. In: *Annual Review of Linguistics* 2 (2016), pp. 325–347.

[25] V Dao Truong, Nam VH Dang, and C Michael Hall. “The marketplace management of illegal elixirs: illicit consumption of rhino horn”. In: *Consumption Markets & Culture* 19.4 (2016), pp. 353–369.

[26] Daan P Van Uhm. *The illegal wildlife trade: Inside the world of poachers, smugglers and traders*. Vol. 15. Springer, 2016.

[27] *World wildlife crime report*. Tech. rep. United Nations Office on Drugs and Crime (UNODC), 2016. URL: http://www.unodc.org/unodc/en/data-and-analysis/wildlife.html.

[28] Roberta Hawkins and Jennifer J Silver. “From selfie to# sealfie: Nature 2.0 and the digital cultural politics of an internationally contested resource”. In: *Geoforum* 79 (2017), pp. 114–123.

[29] Scott M Lundberg and Su-In Lee. “A Unified Approach to Interpreting Model Predictions”. In: *Advances in Neural Information Processing Systems 30*. Ed. by I. Guyon et al. Curran Associates, Inc., 2017, pp. 4765–4774. URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

[30] Chris R Shepherd et al. “Taking a stand against illegal wildlife trade: the Zimbabwean approach to pangolin conservation”. In: *Oryx* 51.2 (2017), pp. 280–285.

[31] Lindall R Kidd et al. “Tweeting for their lives: Visibility of threatened species on twitter”. In: *Journal for Nature Conservation* 46 (2018), pp. 106–109.

[32] EJ Milner-Gulland. “Documenting and tackling the illegal wildlife trade: change and continuity over 40 years”. In: *Oryx* 52.4 (2018), pp. 597–598.

[33] Yinglin Wu et al. “Using social media to strengthen public awareness of wildlife conservation”. In: *Ocean & Coastal Management* 153 (2018), pp. 76–83.

[34] Sandra Díaz et al. “Pervasive human-driven decline of life on Earth points to the need for transformative change”. In: *Science* 366.6471 (2019), eaax3100.

[35] Michael ‘t Sas-Rolfes et al. “Illegal wildlife trade: Scale, processes, and governance”. In: *Annual Review of Environment and Resources* 44 (2019), pp. 201–228.

[36] Tuuli Toivonen et al. “Social media data for conservation science: A methodological overview”. In: *Biological Conservation* 233 (2019), pp. 298–315.

[37] Rania Albalawi, Tet Hin Yeap, and Morad Benyoucef. “Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis”. In: *Frontiers in Artificial Intelligence* 3 (2020). ISSN: 2624-8212. DOI: 10.3389/frai.2020.00042. URL: https://www.frontiersin.org/articles/10.3389/frai.2020.00042.

[38] Francesco Barbieri et al. “TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification”. In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association for Computational Linguistics, Nov. 2020, pp. 1644–1650. DOI: 10.18653/v1/2020.findings-emnlp.148. URL: https://aclanthology.org/2020.findings-emnlp.148.

[39] Anna Hausmann et al. “Understanding sentiment of national park visitors from social media data”. In: *People and Nature* 2.3 (2020), pp. 750–760.

[40] Magdalena Lenda et al. “Effects of amusing memes on concern for unappealing species”. In: *Conservation Biology* 34.5 (2020), pp. 1200–1209.

[41] Scott M. Lundberg et al. "From local explanations to global understanding with explainable AI for trees". In: *Nature Machine Intelligence* 2.1 (2020), pp. 2522–5839.

[42] K Roehrick. *vader: Valence Aware Dictionary and sEntiment Reasoner (VADER)*. 2020.

[43] Ritwik Kulkarni and Enrico Di Minin. "Automated retrieval of information on threatened species from online sources using machine learning". In: *Methods in Ecology and Evolution* 12.7 (2021), pp. 1226–1239.

[44] Magdalena Lenda et al. "Misinformation, internet honey trading and beekeepers drive a plant invasion". In: *Ecology Letters* 24.2 (2021), pp. 165–169.

[45] Joy Rumbidzai Mangachena and Catherine Marina Pickering. "Implications of social media discourse for managing national parks in South Africa". In: *Journal of Environmental Management* 285 (2021), p. 112159.

[46] Jordanna N Bergman et al. "Evaluating the benefits and risks of social media for wildlife conservation". In: *Facets* 7.1 (2022), pp. 360–397.

[47] Niall L Hammond, Amy Dickman, and Duan Biggs. "Examining attention given to threats to elephant conservation on social media". In: *Conservation Science and Practice* 4.10 (2022), e12785.

[48] Jing Li et al. "A Survey on Deep Learning for Named Entity Recognition". In: *IEEE Transactions on Knowledge and Data Engineering* 34.1 (Jan. 2022), pp. 50–70. DOI: 10.1109/tkde.2020.2981314. URL: https://doi.org/10.1109%2Ftkde.2020.2981314.

[49] *African Rhino Numbers Are Increasing Despite Poaching*. Tech. rep. IUCN, 2023. URL: https://www.iucn.org/press-release/202309/african-rhino-numbers-are-increasing-despite-poaching.

[50] *CITES Trade Database*. Tech. rep. CITES.ORG, 2023. URL: https://trade.cites.org/.

[51] *CITES Trade Database surpasses 25 million trade transaction records*. Tech. rep. CITES.ORG, 2023. URL: https://cites.org/eng/news/cites_trade_database_surpasses_25_million_trade_transaction_records.

[52] Yufang Gao et al. "Tracking Chinese newspaper coverage of elephant ivory through topic modeling". In: *Conservation Biology* (2023).

[53] *Rhino Poaching Stats*. Tech. rep. Save The Rhino, 2023. URL: https://www.savetherhino.org/rhino-info/poaching-stats/.

[54] *TRAFFIC Illegal Wildlife Trade*. Tech. rep. TRAFFIC, Nov. 2023. URL: https://www.traffic.org/about-us/illegal-wildlife-trade/.

[55] *UNODC Transnational Organized Crime*. Tech. rep. United Nations Office on Drugs and Crime, Nov. 2023. URL: https://www.unodc.org/toc/en/crimes/organized-crime.html.

CHIP: Contrastive Hierarchical Image Pretraining

Team J.A.R.V.I.S. aka Rural Notice

Abhishek Ajmera amajmera@usc.edu

Arpit Mittal arpitm@usc.edu

Harshil Jhaveri hjhaveri@usc.edu

Swapnil Mallick smallick@usc.edu

Abstract

Few-shot object classification is the task of classifying objects in an image with limited number of examples as supervision. We propose a one-shot/few-shot classification model that can classify an object of any unseen class into a relatively general category in an hierarchically based classification. Our model uses a three-level hierarchical contrastive loss based ResNet152 classifier for classifying an object based on its features extracted from Image embedding, not used during the training phase. For our experimentation, we have used a subset of the ImageNet (ILSVRC-12) dataset that contains only the animal classes for training our model and created our own dataset of unseen classes for evaluating our trained model. Our model provides satisfactory results in classifying the unknown objects into a generic category which has been later discussed in greater detail.

1 Introduction

Recent research in the field of few shot object classification models [11], [8], [13] presume that one of the target classes is present in the query image. These models are not equipped to handle cases where the query image does not contain any of the target class objects along with incapability to categorize the classes. This limits the utility of these models making them incompetent for generic object classification tasks. We try to solve this problem by introducing a one-shot/few-shot learning model with Contrastive Learning [3], [9], [5] approach that can classify any unseen class into a relatively general category (Figure 1). For our project, we aim to classify only animal classes into more general categories. Our model exploits the class hierarchy using contrastive loss using CNNs with respect to the parent embeddings. Contrastive Learning is a method that is known to improve the performance of tasks by contrasting query samples against target labels. In doing so, this method helps to learn both the common attributes between data classes as well as the features that differentiate a data class from another. [1]

2 Related Work

Recently, a considerable amount of research has been conducted to add the aspect of generalization in few-shot learning models. However, most of these proposed networks, in spite of coming up with different novel approaches, fail to achieve good performance in generic few-shot object classification.

In order to improve generalizability, Jiawei Yang et al. [12] have used contrastive learning to annotate unlabelled data followed by latent augmentation methods comprising of k-means selection for training instances. Their key interest is to transfer semantic diversity from the training set to the augmented set. The features in the augmented sets are extrapolated and computed as a dot product and the most useful features are selected based on the magnitude of this dot product. However, the main problem with this approach is that it may be impactful only towards datasets with fewer classes but may not show significant improvements in generalization in datasets with relatively more classes.

Alayrac et al. [1] achieve state-of-the-art results with regards to huge corpora of interleaved visual and text data with disparity in the method that context is sent for each example pair. A visual encoder is used to produce fixed length tokens which is then fed as additional input to weighted attention nets for embeddings generated from the text vectors. A frozen pretrained ResNet-50 architecture is used to preprocess image and video data. Since this architecture heavily utilizes transfer learning from large language models, it inherits their inductive biases. Moreover since the model aims at performing a series of tasks including classification, it lags in its performance as compared to other contrastive

[1] View the source code on GitHub: https://github.com/harshiljhaveri/CHIP/tree/main

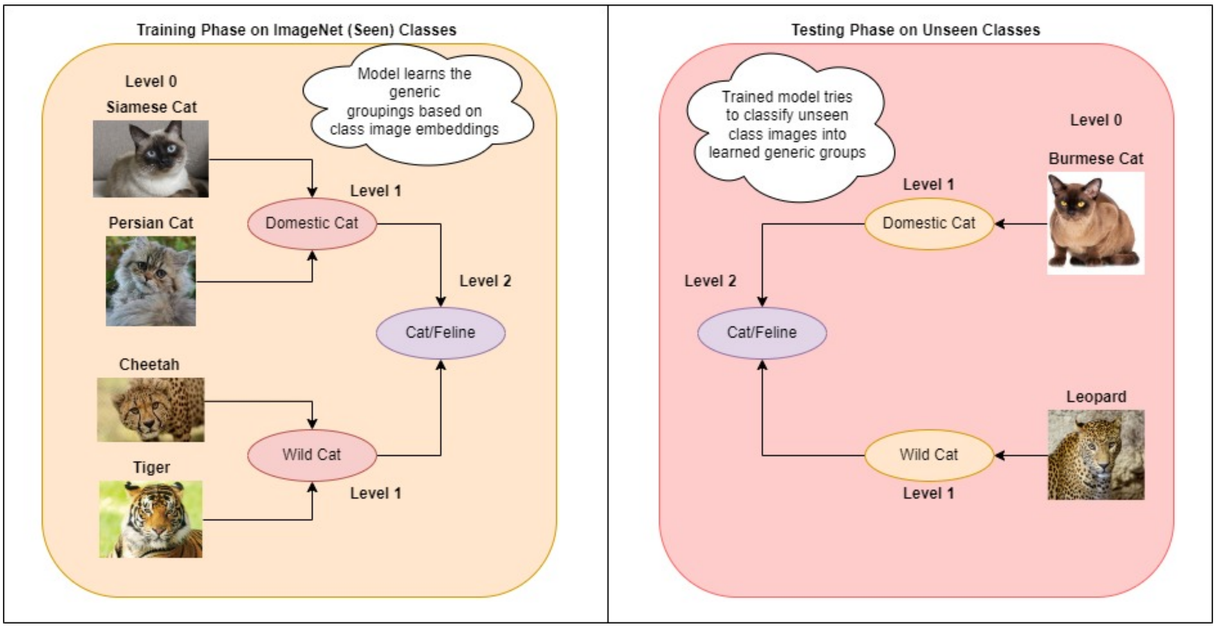


Figure 1: Pictorial representation of the aim of our project

models.
Another attempt [7] to achieve generalization aims at addressing the issues of overlooked classes during classification of multi-label images into a single entity. At first, this method decomposes higher level images into smaller patches and annotates each of these using a fine grained vision transformer. Then a weighted tokening method is carried out to assign labels with minimum losses to each patch. Consequently, similarities are computed between query image patches and support images to assign closest classes. Tokens are assigned to each class based on temperature scaling while minimizing the cross entropy loss. Masks are applied to each patch in order to generalize the constraint and help it fit within the limited support image class options. However, the main drawback of this method is, the transformer is unable to define strict decision boundaries where high feature similarity exists between classes and training data is limited in size.
We have tried to come up with a novel method to solve the problem of generalization in one-shot/few-shot learning that has been discussed in Section 4 in greater detail.

3 Dataset

The ImageNet [10] dataset consists of 14,197,122 images annotated according to the WordNet hierarchy and divided into 1000 classes. The ImageNet dataset is a part of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC-12) which is a benchmark in object category classification and detection of images. Since the ImageNet dataset contains classes other than that of animals, we have segregated the images of all the animal classes from the dataset for our experiment. In doing so, we have got 366 animal classes with 1300 images in each class. We have used images from these classes in our training phase.
For testing our model, we have created our own dataset which contains 20 unseen animal classes that are not present in the ImageNet dataset. This dataset contains the following classes - *American Bobtail cat, American Paint horse, British Shorthair cat, Burmese cat, Camarillo White horse, Catfish, Crow, Cuckoo, Deer, Friesian horse, Giraffe, Ibis, Kingfisher, Kiwi, Parrot, Puffbird, Ragdoll cat, Rhinoceros, Sparrow and Tuna.*

4 Proposed Architecture

Our proposed architecture can be broadly divided into three phases:

4.1 Create Target Parent Image Embeddings (Phase 1):

The first phase of our architecture involves generating mean image embeddings of each ImageNet animal class using the pre-trained ResNet-152 [6] model and using unsupervised K-Means clustering to learn the class hierarchical structure (Figure 2).

Algorithm 1: Creating Target Parent Image Embeddings ($Phase$ 1)

Input : Training Set T, Resnet-152 Image classifier M, Model pretrained weights Θ

```
/* Load M with Θ and remove last FC layer to get image embeddings from M,
   Embedding Dataframe D                                                  */
```

Function `optimalKforCluster` *(mean_image_embedding, mink, maxK)*:
  for *k in range(mink, maxK)* **do**
    KMeansMap [k] = Kmeans(mean_image_embedding, k)
  end
  $optimalK$ = $silhouetteAnalysis$(KMeansMap)
  return $optimalK$
end $funtion$
foreach *Class C in T* **do**
  foreach *Image I in C* **do**
    $imageEmbeddings$ = M(Theta, I)
  end
  $mean\_image\_embedding\_for\_class$ = $mean$(imageEmbeddings)
  $Level0ParentEmbeddingsperClass, D[C, 0]$ = mean_image_embedding_for_class
end
$Level1, 2ParentEmbeddingsperClass, D[C, 1]$ = $KMeans$(D[C,level], $optimalKforCluster$(D[C, level], minK, maxK))
return D

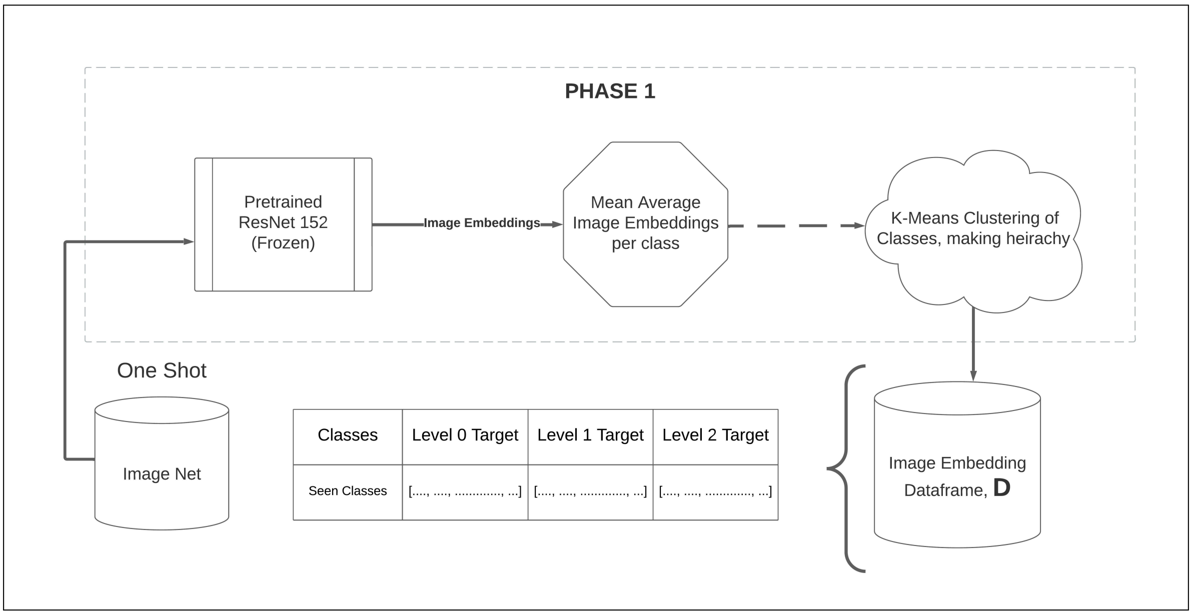


Figure 2: Phase 1 Network Architecture

Algorithm 2: One-Shot Hierarchical Model Learning ($Phase$ 2)

```
Input   : Training Data Tr, Validation Data Va, Test Data Te, Target Parent Embedding
          Dataframe D, Resnet-152 Image classifier M, Model pre-trained weights Θ
Tr = One random image from each class
Va = 5 random image from each class
Te = 5 random image from each class
pretrainedResnet152Model = M(Θ)
Function contrastiveLoss (imageEmbed, targetEmbed):
    dist = CosineDistance(imageEmbed, targetEmbed)
    negDist = ( margin – dist ).relu()
    posDist = dist
    loss = Concat(posDist, negDist).mean
    return lossK
end funtion
Function fineTuneModel (model, trainData, valData, numEpochs):
    for epoch in range(numEpochs) do
        for image, targetEmbed in trainData do
            embed = model(image)
            loss = contrastiveLossFn(embed, targetEmbed)
            backPass
            updateGrads
        end
        for image, targetEmbed in valData do
            embed = model(image)
            loss = contrastiveLossFn(embed, targetEmbed)
        end
        saveModelforEachEpoch()
    end
    return optimalPreTrainedModelForValData
end funtion
Function testModel (model, testData):
    for image, targetEmbed in testData do
        imageEmbed = model(image)
        loss = contrastiveLossFn(embed, targetEmbed)
        cosineSim = CosineSimilarity(imageEmbed, targetEmbed)
    end
    return mean(loss), mean(cosineSim)
end funtion
Level_HeirachialModel = fineTuneModel(new M(Θ), (Tr, level_target_embedding), (Va,
  level_target_embedding), numEpoch, 0)
return Level_HeirachialModel
```

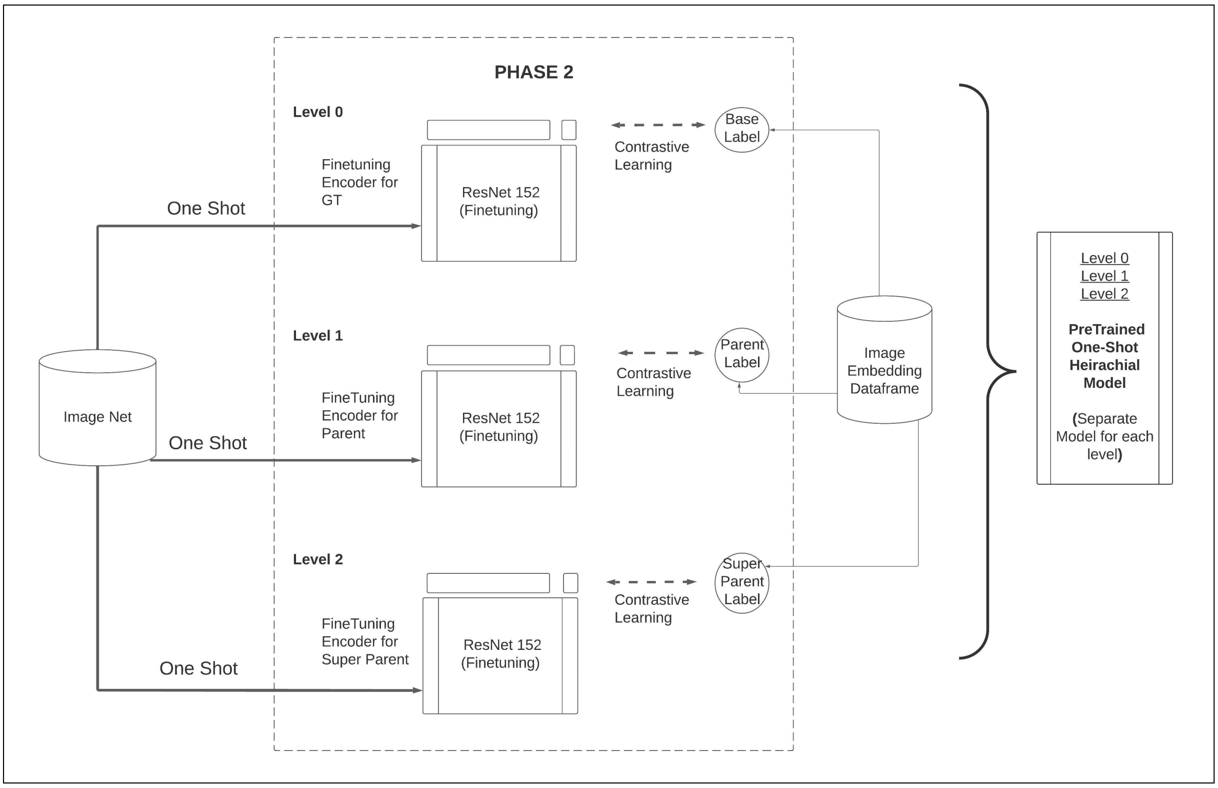


Figure 3: Phase 2 Network Architecture

The steps in Phase 1 are as follows:

1. At first, the mean image embeddings of each of the 366 ImageNet animal classes is computed by loading the ResNet-152 model with its pre-trained weights. So, we get 366 mean embeddings for 366 classes. They would be the target embeddings for the leaf node in our hierarchy and represent as Level 0 embeddings .
2. Using these mean embeddings from 1, these 366 animal classes are then grouped using K-Means clustering and the centroid embedding of each of the clusters is also calculated. Now we can associate each of the 366 classes to a cluster and its corresponding cluster centroid embedding. They would be the target embeddings for the parent node in our hierarchy and represent as Level 1 embeddings .
3. Next using the centroid embeddings from the previous step, these already formed clusters are further grouped into clusters using K-Means clustering and the centroid embeddings of these clusters are also computed. They would be the target embeddings for the super-parent node in our hierarchy and represent as Level 2 embeddings . In order to get the optimal number of clusters, Silhouette Analysis has been done both the times.
4. Finally, a mapping dataframe has been created that stores all the 366 classes, their corresponding mean class embeddings (Level 0) and maps each of these classes to their corresponding Level 1 and Level 2 clusters and their cluster centroid embeddings.

4.2 One-Shot Hierarchical Model Learning (Phase 2):

In the second phase of our architecture, a three-layered pre-trained ResNet-152 model is fine-tuned using one-shot learning approach (Figure 3). The steps of Phase 2 are as follows:
1. In this phase, we have three separate models for three levels. We use ResNet-152 encoder for Level 0 Ground Truth. The Level 1 encoder is finetuned on parent clusters whereas the Level 2 encoder is finetuned on the super-parent clusters.
2. One image is taken from each of the 366 ImageNet animal classes and fed into all the three models (or layers).
3. For each image, an embedding is generated and Constrastive loss is calculated between the generated image embedding and the corresponding target embedding retrieved from the mapping dataframe generated in Phase 1.
4. Then we backpropagate and learn the respective

Algorithm 3: Few Shot Learning on Unseen Data ($Phase$ 3)

Input : Unseen Data UD , Pre-trained Novel Models for Hierarchical Embeddings - $Level\_0\_HeirachialModel$, $Level\_1\_HeirachialModel$, $Level\_2\_HeirachialModel$

Function `assignTargetLabelToNewClass` *(classEmbeddings, level)*:

  $distanceMatrix$ = $CosineDistances$(classEmbeddings, D[level])

  $targetLabels$ = embedding with min distance for each class embeddings

  return $targetLabels$

end $funtion$

Function `targetLabelsForUnseenClasses` *(unseenClassData)*:

  $classEmbeddings$ = $Level\_HeirachialModel$(unseenClassData)

  $targetLables$ = $assignTargetLableToNewClass$(classEmbeddings, level)

  return $targetLables$

end $funtion$

$D[UD, level]$ = $targetLabelsForUnseenClasses$(UD)

$Level\_UnseenModel$ = $fineTuneModel$(Level_HeirachialModel , D[UD,level], numepochs, level)

return $Level\_UnseenModel$

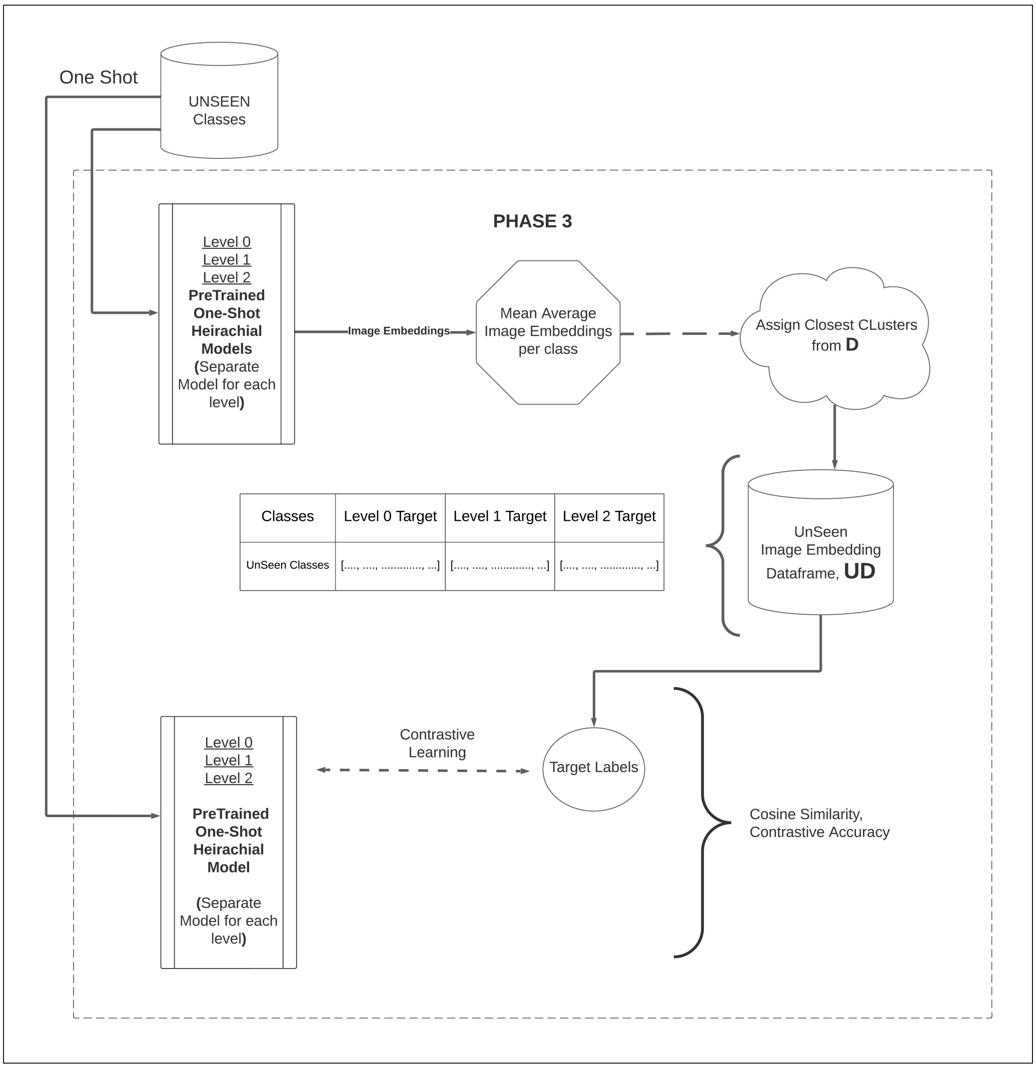


Figure 4: Phase 3 Network Architecture

weights.
5. Steps 2-4 are repeated for a set number of epochs and this is done for all the three separate models. Thus, we get pre-trained one-shot hierarchical model for each of the three levels.

4.3 Few Shot Learning on Unseen Data (Phase 3)

In this final phase, we use the pre-trained ResNet-152 model from Phase 2 and the clusters obtained from Phase 1 to classify our unseen data of 20 animal classes into similar clusters (Figure 4). This phase consists of the following steps:
1. First we feed the unseen class images to the three pre-trained ResNet-152 models from Phase 2 and obtain the mean class embedding for each level.
2. Then we assign target labels to each of the 20 unseen classes by calculating the cosine distance between the mean class embedding of the unseen classes and the mean embeddings obtained in Phase 1 dataframe. The class with the minimum cosine distance is assigned as the class of the unseen data. This done separately for each level.
3. Then we store the mapping of the unseen class embeddings in a new dataframe.
4. Following this, the three pre-trained one-shot models are finetuned using the mapping dataframe of the unseen classes over a set number of epochs.

5 Results

Phase 1: With 1300 images per class, it was computationally expensive to perform clustering and figuring out the optimal K for clustering. We performed 3 methods to get our clusters :
1. Mean Embedding of 1300 images per class.
2. Majority Pooling for 1300 data-points per class.
3. Mean Embedding of 10 images per class in succession, and then performing majority pooling of 130 embedding per class.
We perform a Silhouette analysis for the appropriate number of clusters and overall belongingness of each feature to a cluster. Other popular techniques like top-k accuracy and Cluster Purity were not used since we attempted to cluster features based on similarity rather than comparing with ground truth labels. With comparable Max Silhouette scores for all 3 choices, and method 1 being the most computationally feasible, we proceeded with Mean embedding of all images in a class.

Level 1 results: We implemented K-Means on a range of 50 to 200 clusters for Level 1 clustering and then performed silhouette analysis to evaluate the quality of the clustering results and selected the optimal k for level 1 of the hierarchy which came out to be 88 as in Figure 5.

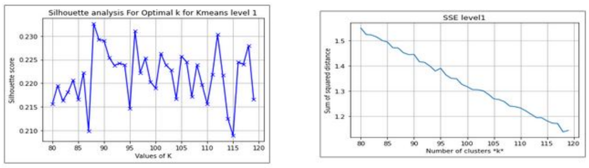


Figure 5: Level 1 results

Level 2 results: We decided to have minimum 7 clusters for level 2, and implemented K-Means on a range of 7 to 20 and based on the silhouette Analysis and Elbow Method, we decided to select 8 as the number of clusters for level 2.

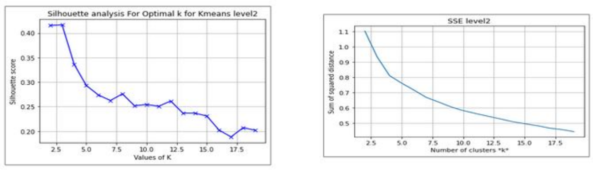


Figure 6: Level 2 results

For Phase 2 and Phase 3, we tried few shot, one shot and training on whole ImageNet data-set. One Shot training gave us much better performance, and we believe this is due to Variance in feature embeddings of the images when our ResNet-152 learns to classify images based on Feature Extraction with Contrastive Loss.

Phase 2: We used Cosine Similarity and Contrastive Accuracy to evaluate Phase 2 using validation and test data. Cosine similarity, in contrast to Euclidean distance, uses dot product rather than the magnitude of distances between the spatial features. As shown in Figure 7, we received an impressive accuracy of around 94% for each level and cosine similarity of over 0.85 for Level 0 and Level 1, and over 0.74 for Level 2. We believe, the drop in Level 2 cosine similarity is due to the fact that Level 2 is classifying more generalized features of the images at Super-Parent level (Table 1)

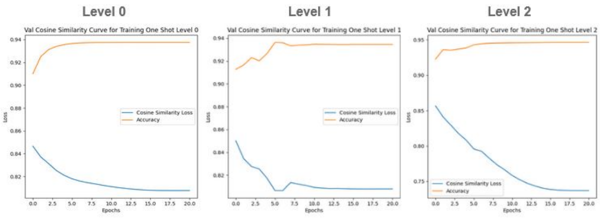


Figure 7: Phase 2 results

	Avg Accuracy	Avg Cosine Similarity
Level 0	93.8	81.5
Level 1	93.9	82
Level 2	95	74.5

Table 1: Phase 2 results for Seen Classes.

Phase 3: The evaluation metrics used for embeddings are Cosine Similarity for Pairwise Similarity between feature embedding of unseen images and the target embedding from Phase 2. While accuracy based on a threshold of pairwise similarity and precision-recall were calculated, they provided highly skewed results and hence were not considered further. Figure 8 shows the Training loss and Cosine Similarity for the 3 levels.

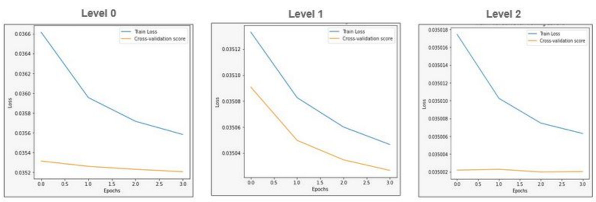


Figure 8: Phase 3 results

	Avg Loss	Avg Cosine Similarity
Level 0	0.035117637	0.92117435
Level 1	0.03500075	0.5109769
Level 2	0.03300126	0.5102735

Table 2: Phase 3 results for Unseen Classes.

There is a significant drop in cosine similarity for levels 1 and 2 when compared to level 0, which may be due to learning more generalized feature mapping. We also noticed a drop in Cosine Similarity for unseen classes, as compared to seen classes. We believe the reason for the drop in both cases is due to the nature of the unseen classes, i.e., the models were not pretrained on unseen classes to learn their features, as compared to pretrained model we used in Phase 2, which already had pretrained weights on ImageNet.

We tested our model on 20 unseen classes, and 14 of them were classified to their correct parent and super parent.

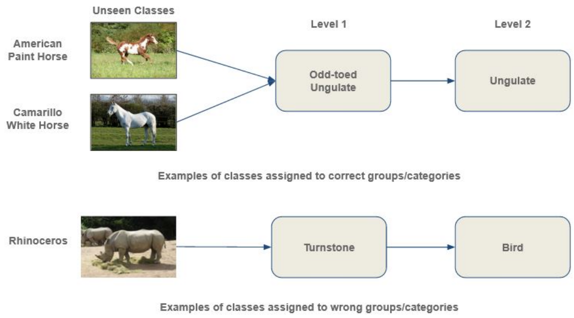


Figure 9: Phase 3 results

6 Conclusion and Future Work

In conclusion, the technique of using ground truth embeddings by clustering images based on feature similarity has been useful to the task of comparing feature extractions if image embeddings. Compared to the commonly performed classification tasks, comparing embedding features using contrastive learning required task adaptive ground truth labels. Thorough experimentation further proves that developing separate functions for each level of the hierarchy is better than using base levels to infer higher level labels. Further, not all evaluation metrics that commonly work for image classification tasks work in this domain since the points of comparison are pairwise similarities and cluster appropriateness rather than singular labels.

The proposed architecture has a lots of scope of improvement in each each of its phases. For Phase 1, we can try a more apt approach for clustering the Images based on Image Features, either by reducing its dimensions using PCA, using density based clustering like DBSCAN [4] or HDBSCAN [2] or other methods to cluster high dimension data. For Phase 2, we can use motivation from CLIP [9] model , and introduce Multi-Modal architecture , adding Textual, Prompt, and contextual features as well while training Level hierarchical models. For Phase 3, we can Better feature extractions and then modulate it with our Phase 2 models, similar to encoder-decoder architecture in transformers, to have the models learn features of our unseen classes in a broader way to improve similarity between our generalized Feature mapping of Hierarchical embeddings.

References

[1] Jean-Baptiste Alayrac et al. "Flamingo: a Visual Language Model for Few-Shot Learning". In: *Advances in Neural Information Processing Systems*. Ed. by Alice H. Oh et al. 2022. URL: https://openreview.net/forum?id=EbMuimAbPbs.

[2] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. "Density-based clustering based on hierarchical density estimates". In: *Advances in Knowledge Discovery and Data Mining: 17th Pacific-Asia Conference, PAKDD 2013, Gold Coast, Australia, April 14-17, 2013, Proceedings, Part II 17*. Springer. 2013, pp. 160–172.

[3] Ting Chen et al. "A simple framework for contrastive learning of visual representations". In: *International conference on machine learning*. PMLR. 2020, pp. 1597–1607.

[4] Martin Ester et al. "A density-based algorithm for discovering clusters in large spatial databases with noise." In: *kdd*. Vol. 96. 34. 1996, pp. 226–231.

[5] Ju He et al. "Transfg: A transformer architecture for fine-grained recognition". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 36. 1. 2022, pp. 852–860.

[6] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[7] Markus Hiller et al. "Rethinking Generalization in Few-Shot Classification". In: *Advances in Neural Information Processing Systems (NeurIPS)*. Ed. by Alice H. Oh et al. 2022. URL: https://openreview.net/forum?id=p_g2nHlMus.

[8] Subhash Chandra Pujari, Annemarie Friedrich, and Jannik Strötgen. "A multi-task approach to neural multi-label hierarchical patent classification using transformers". In: *Advances in Information Retrieval: 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28–April 1, 2021, Proceedings, Part I 43*. Springer. 2021, pp. 513–528.

[9] Alec Radford et al. "Learning transferable visual models from natural language supervision". In: *International conference on machine learning*. PMLR. 2021, pp. 8748–8763.

[10] Olga Russakovsky et al. "Imagenet large scale visual recognition challenge". In: *International journal of computer vision* 115 (2015), pp. 211–252.

[11] Jonatas Wehrmann, Ricardo Cerri, and Rodrigo Barros. "Hierarchical multi-label classification networks". In: *International conference on machine learning*. PMLR. 2018, pp. 5075–5084.

[12] Jiawei Yang et al. "Towards Better Understanding and Better Generalization of Low-shot Classification in Histology Images with Contrastive Learning". In: *International Conference on Learning Representations (ICLR)*. 2022.

[13] Zhiwei Yang et al. "HiTRANS: a hierarchical transformer network for nested named entity recognition". In: *Findings of the Association for Computational Linguistics: EMNLP 2021*. 2021, pp. 124–132.

Contributions

Abhishek Ajmera: Implemented basic k-means clustering and prepossessing steps for images with Swapnil Mallick. Performed integration of Phase 1 and Phase 2. Ran a section of Phase 2 on CARC. Helped in making the final presentation with Swapnil Mallick and contributed to the final report

Arpit Mittal: Designing Model Architecture for all the phases with Harshil, including research on Few Shot Transfer Learning and integrating it with our Model. Running All baseline Models for our Image-Net data set, figuring out appropriate Pretrained Model to be used for IMage Feature extraction in Phase 1 and Model training in Phase 2 and 3. Took mean embedding of each class with three different techniques for Phase 1 as defined in Results, and implemented Kmeans and Agglomerative clustering with different Hyper-parameters for various range of K, and figuring out the best K for our levels and making the Clusters for Level 0,1 and 2. Image preprocessing, including transformation and finetuning of our pretrained Model with one-shot/few-Shot/full data-set using contrastive Loss and Cosine similarity in Phase 2 for all the three levels end to end with output as Pretrained Model for Phase 3. Analysis of Unseen classes and there image embeddings with our current clusters, and performing Scaling and transformations on them w.r.t. imagenet feature space for cluster allocation for Phase 3. Basic implementation of Phase 3 to use Phase 2 results and depict Cross-validation scores for Validation and Test Data. Designing Algorithms for all 3 phases, with architecture Structure.

Harshil Jhaveri: Devised the designs for model architectures along with Arpit. Started with running all baseline pretrained models to generate embeddings for Phase 0 and Phase 1. Wrote the model from scratch for Phase 2 comprising of retraining the pretrained model's entire process of feature extraction. This was followed by appropriately judging model performance and metrics. Rescaled and generated embeddings for Phase 3 and ran the model to generate metrics. Conducted extensive research on our model appropriateness for the new classes and used the most apt metrics for the space and correspondingly preprocessed the test data. Finally mapped the unseen data in the same space and delivered results. Wrote scripts for embedding calculation and clustering allocation.

Swapnil Mallick: Segregated all the animal classes from the ImageNet dataset to prepare our training dataset. Prepared a dataset with 20 unseen animal classes for testing our model. Implemented basic K-Means clustering and preprocessing of images with Abhishek Ajmera. Generated the mapping dataframes (both on seen and unseen data) used for finetuning the ResNet-152 model in Phase 2 and Phase 3. Prepared the final presentation with Abhishek Ajmera. Prepared part of the final report (Abstract, Introduction, Related Works, Dataset, Proposed Architecture).

A Multi-modal Approach to Speech Emotion Recognition

Jiayu Zhou, Paras Sibal, Rohit Bernard, Sanjana Mallikarjuna, Swapnil Mallick
University of Southern California
{zhoujiay,psibal,rpbernar,moodbagi,smallick}@usc.edu

1 Introduction

Speech Emotion Recognition (SER) attempts to identify emotions from speech irrespective of the semantic content. The domain of emotion recognition from natural language is quite extensive. Emotions can be recognised from multiple sources ranging from audio data, facial images to the utterances or text.

Emotion recognition has found great importance in the field of human-computer interaction. This objective motivates the use of automatic emotion recognition systems. The advancements in the field of emotion recognition have made it useful for monitoring psycho-physiological state, improving customer service and forensics.

Having said that, emotion recognition from natural language is still considered to be a challenging task. This is because human emotion can be very subjective. Sometimes, it can be difficult for even human beings to accurately identify the emotion attached to a particular conversation. Thus, considerable research work is still being carried out to devise automatic emotion recognition systems that can efficiently identify the emotion in a conversation.

In this paper, an attempt has been made to come up with a novel multi-modal architecture for emotion recognition from natural language. Different modalities can be used when it comes to emotion recognition. For our work, we are using audio and text as the modalities and IEMOCAP (Busso et al., 2008) as the primary source of data. First, we try to build models that are capable to identify emotions separately from each of the two modalities. Consequently, we try to concatenate these two models to see the effect use of different modalities has on emotion recognition. Adding different modalities makes our emotion recognition model more powerful and robust.

2 Related Work

Attempts to build efficient systems for emotions recognition has been an active area of research over the years. But the main focus of most of the previous attempts have been on either text or audio data separately. Research on multi-modal approach to emotion recognition is still in its latent stages. One of the reasons for insufficient research may be the lack of reliable and complete datasets. However, IEMOCAP can be considered as a benchmark dataset in this regard.

Though IEMOCAP provides as a good source of data with multiple modalities, a significant amount of preliminary research on IEMOCAP dataset has been focussed on speech (or audio) data only. Majority of the architectures are based on neural networks. (Han et al., 2014) is considered to be one of the earliest works in this regard. In their work, the authors have proposed an MLP based architecture to solve the problem. For every speech segment, an emotion state probability distribution is produced. These segment level features are then fed to the proposed model.

Another seminal work (Lee and Tashev, 2015) suggests a Bidirectional Long Short-Term Memory (Bi-LSTM) architecture. For every frame, 32 features are extracted - F0 (pitch), voice probability, zero-crossing rate, 12-dimensional Mel-frequency cepstral coefficients (MFCC) with log energy, and their first-time derivatives. These 32-dimesional vectors are expanded to 800-dimensional vectors and fed into the Bi-LSTM model which contains 2 hidden layers with 128 Bi-LSTM nodes.

(Hazarika et al., 2018) have proposed a multi-modal architecture tested on IEMOCAP dataset. The model is based on Gated Recurrent Unit (GRU). The authors have used different combinations of modalities and compared the accuracy of their model to see how different modalities affect the performance of the model. In our work, we propose a novel method for multi-modal emotion recognition and achieve either similar or superior performance.

3 IEMOCAP

3.1 Existing dataset

The dataset used to build the multi-modal architecture is USC's Interactive Emotional Dyadic Motion Capture Database (IEMOCAP) dataset. To obtain the data, we requested USC's SAIL laboratory. IEMOCAP is an audio-visual dataset used in emotional learning. It contains 12-hour record of simulated scenarios from ten actors. The scenarios are designed to draw out a set of emotions. There are 10039 utterance files generated from ten actors. Six annotators classify each utterance based on their emotions into ten categories: anger, disgust, excited, fear, frustration, happiness, neutral, sadness, surprise, and other. The files were annotated in such a way that each utterance was annotated by three annotators. For each utterance, there is audio-video file and transcriptions. One utterance can have as many as four emotions. Because it is possible that one annotator labels an utterance with more than one emotion. This shows that nature of human interaction from the hearing can lead to different meaning.

3.2 Dataset creation

From the IEMOCAP dataset, we extract the data required for our modelling. For each utterance, both audio as well as the transcriptions are used. Labels given by the three annotators are converted to percentages. For example, if one utterance was labelled 'anger' by first annotator, 'frustrated' by second annotator and 'anger' by third annotator, then the utterance was assigned values 0.66 anger and 0.33 frustration. The dataset also had one label assigned. The assessment for each utterance is decided by the simple majority vote. For simplicity and for better performance of the multi-modal architecture - 5 emotion categories - Anger, Frustration, Happiness, Neutral and Sadness were considered for model training and testing. Excited was converted to Happiness and utterances with other labels were ignored. On training our model using 8 categories vs 5 categories, 5 classes gave us a more stable model with lesser variance.

Emotion	No. of Utterances
Anger	1103
Frustration	1849
Happiness	1636
Neutral	1708
Sadness	1084

Table 1: No. of Utterances of each category

3.3 Data Split

Our experiment required splitting of data into training, validation and testing set. The IEMOCAP dataset does not have these splits. To train and validate our model, and finally to make predictions on the test data, we split the dataset using sklearn library's train-test-split into 70% training data, 20% validation data and 10% testing data. There is no overlapping occurrence between the sets. Splitting is such that there are approximately same proportion of data of each category in each of the splits.

4 Data Preprocessing

For the development of a speech emotion analyzer, preprocessing is the crucial phase to tune the audio data to create feature vectors used as an input to the model. Nowadays, most researchers utilize deep learning techniques for SER using Mel-frequency cepstrum coefficients and Mel-scale filter bank speech spectrogram as input features to their models. Similarly, we utilized the transfer learning strategies for SER using speech spectrograms to extract the salient and discriminate features.

The audio data obtained from the raw dataset had non-uniform dimensions varying in audio length. The raw data was sampled to have uniform standardized dimensions by adding zero paddings to shorter audio clips to get an equal number of features. Furthermore, we checked for the signal-to-noise ratio to assure that there was less ambient distorted noise in the data. Moreover, the sampling rate for each audio file was set to 22 kHz. This allowed each audio sample to get all features that helped classify the audio file while keeping noise minimum.

When we do speech recognition tasks, Mel-frequency cepstrum coefficients (MFCCs) are the state-of-the-art features of the training model. For this project, we generated mel spectrograms using tuned data clips with the help of python librosa library and converted them to mel scales so that the perpetual difference becomes smaller to deal with human frequencies. Furthermore, the log-mel spectrogram, given as input to the model, is normalized by the mean and variance of the training set. Moreover, we compressed the mel spectrogram into a short-time spectrum to extract only the essential coefficients, which only correspond to human frequency ranges.

In addition, we performed Data Augmentation by adding noise, pitch, stretch, and shift to the original signal. This immensely improved the accuracy of our model.

5 Design

In this paper, we compare the accuracies achieved by text-only based models, audio-only based models, and our novel multimodal model. In order to compare these accuracies, our multimodal architecture combines the results generated by two separate models, each trained using a different modality.

5.1 Emotion Recognition from text

In order to predict emotions from text, we fine tune BERT (Devlin et al., 2019) on our dataset to perform 5-class text classification. We make use of the pre-trained transformer model 'bert-base-cased'. To fine tune the model, we use CrossEntropy loss and Adam optimizer over 15 epochs. The results obtained by the text-only model were documented.

5.2 Emotion Recognition from audio

In order to predict emotions from audio, we train a Convolutional Neural Network using VGG frame from scratch. The model consists of 5 Convolutional layers with Max Pooling, 2 Linear layers, LeakyReLU activation, Dropout, and a Softmax head. The model is trained to classify 5 classes of emotions using the MFCC generated from the audio files. We use CrossEntropy loss and Adam optimizer over 15 epochs. The results obtained by the text-only model were documented.

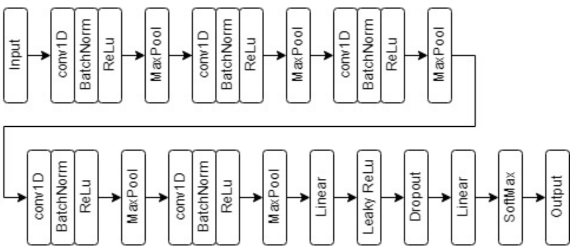


Figure 1: CNN in VGG Frame Architecture

5.3 Multimodal Design

To create our multimodal architecture, we combine our fine-tuned BERT model with our CNN in parallel, and combine the results of these two models. For each datapoint, both models produce 5x1 vectors containing the probabilities for each class. We pass each vector through a Linear layer, after which they are summed up and passed through a Softmax layer to generate the final predictions. This multimodal architecture is trained again using CrossEntropy loss and Adam optimizer over 15 epochs. The results obtained by the text-only model were documented.

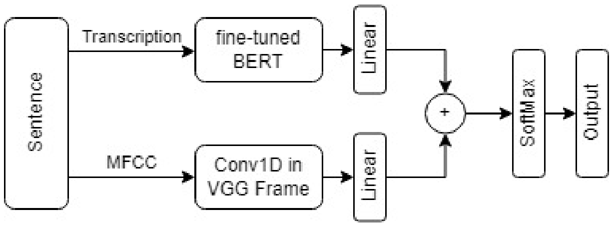


Figure 2: Multimodal model Architecture

6 Experiment

6.1 Experimental Setup

We evaluate our proposed model on IEMOCAP dataset, and consider five broad emotion categories - Anger, Frustration, Happiness, Neutral and Sadness for examining the performance of our architecture.

Baseline Methods: We compare our proposed model with different state-of-the-art multi-modal emotion recognition architectures:

- **ICON:** Interactive COnversational memory Network (ICON) (Hazarika et al., 2018) is a multi-modal emotion recognition model. It uses Gated Recurrent Unit (GRU) to hierarchically model self- and inter-speaker emotional influences into global memories.

- **bc-LSTM:** (Poria et al., 2017) have proposed a bidirectional LSTM with hierarchical fusion. The model uses context features from uni-modal LSTM's. The features are concatenated and fed as input to the final LSTM.

Evaluation Protocols: We use 20% of our training set as the validation set for hyper-parameter tuning. We train each of the three models for 15 epochs. We use Adam optimizer to train our models with a learning rate of 1.5e-4 and a weight decay of 1e-3.

Hyper-parameter	Value
Optimizer	Adam
Learning Rate	1.5e-4
Weight Decay	1e-3
Loss Function	CrossEntropy
Activation	ReLU, LeakyReLU
Dropout	0.6
# Epochs	15

Table 2: Hyper-Parameters

For audio-only model, we have used different architectures of CNN before coming up with the final model. We also used Wav2Vec (Schneider et al., 2019) method. For text-only model, we used different transformers like, BERT and RoBERTa (Liu et al., 2019) to see which model performs better.

6.2 Results and Discussion

We compare the performance of our proposed multi-modal architecture with that of the previously mentioned baseline models and find some interesting results. The results obtained from our experiment can be found in the table below.

Modality	**Models**		
	ICON	bc-LSTM	Our Model
Text	58.3	**73.6**	72.4
Audio	50.7	**57.1**	47.8
Text + Audio	63.8	75.6	**76.9**

Table 3: Comparison of accuracy of proposed model

When we take text as the modality, our model performs significantly better than ICON and comes pretty close to the performance of bc-LSTM. However, our model fails to outperform both the baseline models when audio is chosen as the modality. Finally if we use both text and audio as modalities then we find that our model performs notably better than ICON and achieves a small improvement over bc-LSTM.

If we examine our own model, we find that the multi-modal architecture performs slightly better than the text-only model. This, further, indicates that using multiple modalities helps in improving the performance of automatic emotion recognition systems.

7 Conclusion

In this paper, we have attempted to design a novel multimodal approach to Emotion Recognition, and compare the effectiveness of multimodality against any single modality. We can see that our multimodal model performs better than the baseline models. The results of the experiment show that we can achieve a good accuracy by fine-tuning a pre-trained model to predict emotion solely from text. However, a from-scratch, purely audio-based model does not perform as well as a text-based model. But, combining the predictions of both models before the final layer of the network does help increase the performance, which shows that speech audio does have some useful information that cannot be picked up from text alone, and thus showing that multiple modalities of data help achieve better accuracy than any single modality.

Team responsibilities.

Our project had various aspects that needed to work upon, from the preprocessing to the model implementation. Each individual carried out their respective task and helped other team members accomplish the results. Paras Sibal and Sanjana Mallikarjuna were in charge of Data Preparation, Data Preprocessing, and Data Augmentation. Rohit Bernard was in charge of the speech-to-emotion classification model and Swapnil Mallick for the text-to-emotion classification model. Jiayu Zhou carried out the final task of designing the multimodal architecture to combine the results from the speech-to-emotion and text-to-emotion classification models. In addition, each team member worked on a separate speech-to-emotion classification model based on their understanding to find the perfect fit model for our project.

References

Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N. Chang, Sungbok Lee, and Shrikanth S. Narayanan. 2008. IEMOCAP: interactive emotional dyadic motion capture database. *Lang. Resour. Evaluation*, 42(4):335–359.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA,* *June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.

Kun Han, Dong Yu, and Ivan Tashev. 2014. Speech emotion recognition using deep neural network and extreme learning machine. In *INTERSPEECH 2014, 15th Annual Conference of the International Speech Communication Association, Singapore, September 14-18, 2014*, pages 223–227. ISCA.

Devamanyu Hazarika, Soujanya Poria, Rada Mihalcea, Erik Cambria, and Roger Zimmermann. 2018. ICON: interactive conversational memory network for multimodal emotion detection. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 2594–2604. Association for Computational Linguistics.

Jinkyu Lee and Ivan Tashev. 2015. High-level feature representation using recurrent neural network for speech emotion recognition. In *INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association, Dresden, Germany, September 6-10, 2015*, pages 1537–1540. ISCA.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Soujanya Poria, Erik Cambria, Devamanyu Hazarika, Navonil Majumder, Amir Zadeh, and Louis-Philippe Morency. 2017. Context-dependent sentiment analysis in user-generated videos. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 873–883. Association for Computational Linguistics.

Steffen Schneider, Alexei Baevski, Ronan Collobert, and Michael Auli. 2019. wav2vec: Unsupervised pre-training for speech recognition. In *Interspeech 2019, 20th Annual Conference of the International Speech Communication Association, Graz, Austria, 15-19 September 2019*, pages 3465–3469. ISCA.