IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 1 November 2023, accepted 3 December 2023, date of publication 13 December 2023,
date of current version 21 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3342154

RESEARCH ARTICLE

Multi-Aspect Annotation and Analysis of Nepali Tweets on Anti-Establishment Election Discourse

KRITESH RAUNIYAR[1], SWETA POUDEL[2], SHUVAM SHIWAKOTI[3], SURENDRABIKRAM THAPA[4], (Member, IEEE), JUNAID RASHID[5], JUNGEUN KIM[6], MUHAMMAD IMRAN[7], AND USMAN NASEEM[8]

[1]Department of Computer Science and Engineering, Delhi Technological University, New Delhi 110042, India
[2]Kathmandu Engineering College, Tribhuvan University, Kathmandu 44600, Nepal
[3]Department of Software Engineering, Delhi Technological University, New Delhi 110042, India
[4]Department of Computer Science, Virginia Tech, Blacksburg, VA 24060, USA
[5]Department of Data Science, Sejong University, Seoul 05006, Republic of Korea
[6]Department of Software, Kongju National University, Cheonan 31080, Republic of Korea
[7]Institute of Innovation, Science and Sustainability, Federation University, Brisbane, QLD 4000, Australia
[8]College of Science and Engineering, James Cook University, Cairns, QLD 4814, Australia

Corresponding authors: Junaid Rashid (junaid.rashid@sejong.ac.kr) and Jungeun Kim (jekim@kongju.ac.kr)

This work was supported in part by the Basic Science Research Program under Grant 2020R1I1A3069700, and in part by the Technology Development Program of MSS under Grant S3033853.

ABSTRACT In today's social media-dominated landscape, digital platforms wield substantial influence over public opinion, particularly during crucial political events such as electoral processes. These platforms become hubs for diverse discussions, encompassing topics, reforms, and desired changes. Notably, in times of government dissatisfaction, they serve as arenas for anti-establishment discourse, highlighting the need to analyze public sentiment in these conversations. However, the analysis of such discourse is notably scarce, even in high-resource languages, and entirely non-existent in the context of the Nepali language. To address this critical gap, we present **N**epal **A**nti **E**stablishment discourse **T**weets (NAET), a novel dataset comprising 4,445 multi-aspect annotated Nepali tweets, facilitating a comprehensive understanding of political conversations. Our contributions encompass evaluating tweet relevance, sentiment, and satire, while also exploring the presence of hate speech, identifying its targets, and distinguishing directed and non-directed expressions. Additionally, we investigate hope speech, an underexplored aspect crucial in the context of anti-establishment discourse, as it reflects the aspirations and expectations from new political figures and parties. Furthermore, we set NLP-based baselines for all these tasks. To ensure a holistic analysis, we also employ topic modeling, a powerful technique that helps us identify and understand the prevalent themes and patterns emerging from the discourse. Our research thus presents a comprehensive and multi-faceted perspective on anti-establishment election discourse in a low-resource language setting. The dataset is publicly available, facilitating in-depth analysis of political tweets in Nepali discourse and further advancing NLP research for the Nepali language through labeled data and baselines for various NLP tasks. The dataset for this work is made available at https://github.com/rkritesh210/NAET.

INDEX TERMS Natural language processing, social media analytics, sentiment analysis, topic modeling, Nepali election discourse.

I. INTRODUCTION

Social media has become a prominent platform for engaging in discussions about major global topics, including politics. This trend extends to political matters as well, with politicians utilizing social media to promote their campaigns and ideologies, while the general public leverages it to express their opinions. The ease of use and anonymity provided by social media promotes the production of political discourse and influences the type of discourse produced [1]. Especially

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Asif.

during agitated times, such as protests and social movements demanding political changes, social media experiences a rise in the amount of political discourse produced [2]. The discourse produced may include varying tones, aggressiveness, and sentiments of the people. This creates an opportunity as well as a need for the study of how people express their views and ideas involving political matters in an informal setting like social media.

In nations with political instability, elections emerge as pivotal moments parking intense discourse on social media. Nepal, being one of such countries, has experienced extreme political instability over just 15 years of gaining democracy [3]. The end of the monarchy on May 28, 2008, and the formation of a democratic multi-party system marked a new era for the country [4]. The perception of having a fair opportunity in the democratic process rose among the people, however, the election and the established Constituent Assembly were dominated by the previously established big parties [5]. After the end of the autocratic Shah dynasty, Nepal had some opportunities to develop, but the increasing intra and inter-party conflicts led to political polarization. As a result, corruption escalated, and the country plunged into a state of political and constitutional crises [3]. The lasting political crises and frequent changes in the government led to distrust among the people of the established government, old political parties, and their leaders. As a result, as the campaign period leading to the local elections on May 13, 2022 [6] approached, the citizens expressed their dissatisfaction with the current political landscape in Nepal. They yearned for greater participation of youth to bring about meaningful changes in the political sector. This led to the commencement of various anti-establishment movements on Twitter and Facebook that appealed to the voters not to vote for the older political parties and their elderly established leaders who have consistently failed to bring political stability to the country [7]. Twitter was flooded with hashtags like #enoughisenough, #nonotagain, #wewantchange, etc. about anti-establishment movements. These anti-establishment movements became highly prominent in the Nepalese general election which was held on 20 November, 2022 especially when new parties like Rashtriya Swatantra Party came into the picture. To study the effects of tweets related to such movements, we analyzed tweets from July 27, 2022, to February 7, 2023.

In recent times, social media, especially Twitter, has housed several such movements [8] and has become a turning point to bring about a significant change. The same holds true for anti-establishment movements in Nepal as citizens in great numbers stood with the movement and spoke against the established failing political regime. These movements escalated to the point where the Election Commission of Nepal released a statement on October 25, urging the public to abstain from making "negative" remarks about prominent leaders, or risk potential legal consequences [9]. However, this only escalated the movement further as the public questioned the integrity of the nation's poll authority. Later, the Supreme Court of Nepal ruled in favor of anti-establishment movements and issued an interim order against the commission not to take any action against the campaign members as it was against the Freedom of Speech of the nation's citizens [10].

Vigorous movements like anti-establishment movements with such a great number of participants are certain to receive varying sentiments from the public. Hate speech against individuals and groups is also inevitable. These sentiments of the public are best captured in their native language as it also includes certain socio-cultural aspects of the country. Nepali - the most spoken language in Nepal [11], is also the major language for social media discourse in Nepal. The complex and monographically rich nature of the Nepali language makes Natural Language Processing (NLP) tasks particularly challenging for this language [12]. Unlike languages such as English, which has ample resources and a plethora of NLP studies [13], [14], research in the field of low-resource languages such as Nepali is scarce [15], [16]. There have been some works involving sentimental analysis in Nepali [12], [17], [18]. These works have focused on a single aspect of sentiment analysis such as hate speech detection or abuse analysis. However, given the complex narratives of speech in social media, the classification of speech within a single aspect is not enough to capture the layers of sentiment that people try to communicate. To address this issue, we developed a dataset containing 4,445 tweets related to the anti-establishment movements which is annotated through a multi-aspect annotation schema.

Within this comprehensive multi-aspect annotation framework, each tweet undergoes a nuanced evaluation across multiple dimensions, enabling a deeper understanding of the sentiments being conveyed. The dimension of "Relevance" gauges the extent to which a tweet is intrinsically tied to the movement, accounting for context and content. Additionally, the "Sentiment" category delves into whether a tweet conveys positive, negative, or neutral sentiment about the subject matter. The recognition of the prevalent use of satire in online conversations, addressed under the "Satire" dimension, identifies instances where tweets employ satirical elements to convey sentiment. In instances where hate is expressed, the "Hate Speech" dimension identifies tweets containing such language or sentiments. Moreover, within the "Directed/Undirected Hate" sub-category, we determine whether hate speech targets specific entities or adopts a more general tone. This information is further enhanced in the "Targets" sub-category, which pinpoints whether individuals, organizations, or entire communities are the focus in cases of directed hate speech. Lastly, the "Hope Speech" aspect evaluates whether a tweet conveys hope within its content.

Our multi-aspect annotation schema is designed to address subtle differences in the sentiments expressed by tweets. There is often a thin line between satire and hate speech. A few words or some background context may completely change the intended sentiment. Manual annotations performed by native speakers of the Nepali language ensure that such thin lines and subtle differences are addressed properly. To analyze our dataset, we performed benchmarks with popular machine learning and deep learning models. Further, we also performed topic modeling for a more detailed analysis of the dataset. With the rapid growth of social media, the development of reliable and efficient NLP tools has become important now more than ever. Thus, we believe that our dataset in a low-resource language is a vital contribution to the field of NLP and serves as a stepping stone for further research in NLP for the Nepali language. Our contributions are:

1. We developed a new dataset called **N**epal **A**nti **E**stablishment discourse **T**weets **(NAET)**, which includes 4,445 manually annotated tweets with a multi-aspect annotation schema related to Nepali election discourse.
2. Multi-aspect annotation had a total of 7 primary classes i.e., 'Relevance', 'Sentiment Analysis', 'Satire', 'Hate Speech', 'Direction of Hate Speech', 'Targets of hate speech', and 'Hope Speech'.
3. We perform a thorough analysis of the dataset with techniques like topic modeling. Using various textual models, we have established benchmarks for different tasks. Our benchmarks demonstrate the potential for development in the automated detection of different types of speech in Nepali.

The relevance of the study holds importance to the fields of political discourse analysis and social media studies. Assessing the existence of confidence, encouragement, and enthusiasm in regard to prospective political developments, enhancements, or modifications, it offers useful insights into how individuals participate in political debate on social media networks. Recognizing such behaviours can shed insight into the dynamics of political communication and public emotions in the digital era, supporting researchers, lawmakers, and political analysts in understanding the developing landscape of political engagement and its consequences for current society.

II. RELATED WORK

Online hate speech poses a significant concern within the realm of social media. It capitalizes on the gaps and vulnerabilities present in the regulations that characterize the majority of social media platforms. The primary contributors to this situation are offensive remarks made during user interactions or in shared content [19]. The widespread use of the internet has increased access to political information, prompting debates about the significance of online dialogues in the context of politics [20]. Hate speech has become more prevalent in political discourse, which has an impact not only on the reputations of individual politicians but also on the functioning of society as a whole [21].

A. POLITICAL DISCOURSE IN THE INTERNET

The analysis of discourse on the Internet has received significant attention in the political areas of the countries with high language resources, such as English, and Japanese. Takikawa et al. [22] addressed echo chamber issues in the Japanese Twitter political field by analyzing the structure of tweets in combination with large-scale social network analysis and natural language processing. With 3,200 tweets from Twitter, they discovered six unique communities in the area that represented a wide range of political ideologies in Japan. Johnson et al. [23] analyzed the usefulness of ideological expressions as a characteristic for forecasting the content of a political tweet using Probabilistic Soft Logic (PSL) models on 2,050 congressional tweets, labeled using 17 possible frames. From each frame, small sentences that were often repeated were further selected and combined into more substantial phrases. Similarly, Zakharov et al. [13] presented an annotation scheme, tagset, and task for conversational discourse parsing that was created specifically for non-convergent talks. A total of 16,000 discussions were collected from Reddit. The annotations were considered valid only if two annotators agreed upon the label. The final dataset consisted of a total of 10,559 posts out of which 9,620 were labeled with 17,964 tags (31 unique tags). Solovev et al. [21] implemented hierarchical regression models to examine whether politicians who exhibit particular traits are more likely to be the targets of hate speech. They used Twitter's historical API to collect the timelines of every politician in the 117th U.S. Congress and collected 199,294 tweet histories excluding replies and retweets. Additionally, up to 250 replies from each tweet were queried which resulted in a total number of 8,362,555 replies. Moreover, Macrohon et al. [14] proposed a semi-supervised sentiment analysis using multinomial Naive Bayes of tweets during the 2022 Philippine Presidential Election. A total of 150,792 raw tweets were collected from Twitter API. After removing duplicates and retweets, a total of 114,851 tweets were annotated. 83.90% of tweets were negative polarity followed by 13.49% and 2.60% positive and neutral tweets, respectively. Johnson et al. [24] provided a weakly supervised strategy for assessing politicians' positions on a wide range of issues by examining how issues are articulated in their tweets and temporal activity patterns. With an average of 3,000 tweets per politician, there are 99,161 tweets for the entire group of 32 politicians participating in the 2016 U.S. presidential election. These studies highlight the use of Natural Language Processing (NLP) in political discourse and politics in general.

B. NATURAL LANGUAGE PROCESSING IN NEPALI LANGUAGE

Natural language processing (NLP) has made tremendous advances in recent years. New possibilities for data analysis and information extraction have been made possible by ground-breaking developments in NLP, which has enabled machines to interpret human language with very high precision [25]. Although NLP has vastly advanced in high language resources, in the context of the Nepali language, NLP is a relatively new field of involvement. The Nepali Spell Checker and Thesaurus, which were released in 2005, were some of the first NLP resources ever created for the Nepali language [26]. Below, we discuss a few works in the Nepali language.

Gupta et al. [27] provided two key approaches for Nepali text sentiment analysis. In the first approach, they established a method wherein terms with emotional connotations are found in texts written in Nepali are used to determine the mood of the document. Additionally, in the second approach, annotated Nepali text data were used to build a machine learning-based text classifier to categorize the content. Similarly, Prabha et al. [28] suggested a deep learning-based Part of Speech (POS) tagger for Nepali text utilizing Recurrent Neural Networks (RNN), Long Short- Term Memory Networks (LSTM), Gated Recurrent Units (GRU), and their bidirectional versions.

Moreover, Shrestha et al. [29] put together a rule-based anaphora resolution module, a named entity recognition classifier, and a part-of-speech tagger for the Nepali language. With a sentiment Corpus of 3,490 sentences from Nepali News Media texts, manually annotated, they developed a Machine Learning based sentiment classifier. Beyond the development of methodologies and techniques, particularly in deep learning-based research and applications, the availability of improved word embeddings in any language plays a crucial role in propelling the progress of computational linguistics within that language. Koirala et al. [30] presented NPVec1 which consisted of 25 state-of-art word embeddings for the Nepali language derived from a large corpus using GloVe, Word2Vec, fastText, and BERT. The embeddings were made publicly available to the NLP community.

In recent times, transformer-based models, trained on extensive corpora, have demonstrated impressive potential in the field of Natural Language Processing (NLP) [31]. Leveraging their capability to grasp context and language nuances effectively, these models exhibit versatility and good accuracy across multiple languages. Such transformer-based models can be categorized as either monolingual, trained on a corpus specific to a single language, or multilingual, trained on data from various languages. Nevertheless, despite this versatility, multilingual models like Multilingual BERT [32], XLM [33], and XLM-RoBERTa [34] have not yielded promising outcomes when applied to the Nepali language. To address this, Timilsina et al. [35] proposed NepBERTa, a Nepali Language model trained on a dataset of 0.8 Billion words collected from 36 popular news sites in Nepal. They also introduced the first Nepali Language Understanding Evaluation (Nep-gLUE) benchmark system. They made their Nep-gLUE along with the pre-trained models publicly available for further research.

Machine translation systems play a pivotal role in the realm of Natural Language Processing (NLP) across various languages. In this context, Laskar et al. [36] used Neural Machine Translation (NMT) with an attention mechanism to develop a translation model that could translate texts between two similar languages Hindi and Nepali. Despite some notable efforts in NLP focused on Nepali, the progress in this domain lags behind the advancements witnessed in more widely recognized languages. Much ground remains to be covered in order to bridge the gap and bring Nepali NLP on par with its counterparts. As is often the case with low-resource languages, a key hindrance to advancing NLP for Nepali is the lack of high-quality data.

C. THE PRESSING NEED OF DATA IN NEPALI LANGUAGE

Within this section, we discuss some existing datasets within the Nepali language, emphasizing the growing need for substantial data resources in Nepali. Additionally, we explore the pivotal role our dataset plays in bridging this data gap, propelling the advancement of Natural Language Processing (NLP) in Nepali to new heights.

Lamsal [37] compiled a Nepali text corpus of over 90 million running words (6.5+ million sentences). Furthermore, they also created 300-dimensional word vectors for more than 500 thousand Nepali words/phrases. For data collection, they used popular Nepali news portals such as Kantipur, Nagariknews, and Setopati. The text corpus covered domains such as Health, Literature, Finance, News, Entertainment, Sports, and Technology. Similarly, Senapati et al. [38] collected data from 18 different domains which included short stories, blogs, and news articles. They used some of the previously established annotation schemes to manually annotate the data with necessary information like parts-of-speech (POS), named entity, chunking information, and other morphological details like number, gender, person, etc. The dataset consisted of 309 sentences and approximately 4,700 words. They used the dataset to develop an Anaphora Resolution system in Nepali using machine learning.

Additionally, Singh et al. [39] introduced a novel NER (Named Entity Recognition) dataset in Nepali, namely NepaliNER. They collected the data from various Nepali news websites and contained texts from multiple categories like politics, business, crime, sports, tourism, and arts. Their NER schema contained 3 major classes - Person, Location, and Organization. They also annotated their dataset for POS tagging but didn't perform manual annotations for it. They used the Nepali National Corpus (NCC) [40] to train a BiLSTM model and used the model to annotate the NepaliNER dataset with POS tags. Adding to the contribution in datasets, Sitaula et al. [41] designed a dataset by crawling news documents from popular Nepali online news portals like Kantipur, Ratopati, and Nagarik to collect a total of 35,651 documents across 17 news categories. The categories included Art, Bank, Blog, Business, Diaspora, Entertainment, Filmy, Health, Hollywood-Bollywood, Koseli, Literature, Music, National, Opinion, Society, Sports and World. They used various classification techniques to perform classification tasks across these 17 categories. Moreover, Shahi and Pant [42] created a Nepali News Corpus with data crawled from different national news portals and contained a total of 4,964 documents spanning 20 different categories. The identified categories were Agriculture, Automobiles, Bank, Blog, Business, Economy, Education, Employment, Entertainment, Health, Interview, Literature, Migration, Opinion, Politics, Society, Sports, Technology, Tourism, and World. They performed baselines for the classification tasks using various machine learning and deep learning models.

Sentimental analysis is a major task in NLP involving any language. Singh et al. [18] introduced a new Nepali sentiment analysis dataset (NepSA) by collecting comments from popular Nepali YouTube channels and videos. They extracted 3,068 comments from 37 different YouTube videos of 9 different YouTube channels and used the binary sentiment polarity schema to divide the comments into 6 aspect categories - General, Profanity, Violence, Feedback, Sarcasm, and Out-of-scope. Further, they also classified the comments into 4 target entities - Person, Organization, Location, and Miscellaneous. They used the dataset for abusive sentiment detection in Nepali social media. In addition to this, Sitaula et al. [17] prepared a sentimental analysis dataset in Nepali, namely NepCOV19Tweets containing tweets related to COVID-19. They used 4 manual annotators to annotate 33,247 tweets into 3 categories - positive, neutral, and negative. They employed various machine learning and deep learning methods to perform benchmark classifications. Contributing more with respect to sentiment analysis, Niraula et al. [12] collected 7,462 records (comments, posts, articles) from various social media platforms like Facebook, Twitter, YouTube, Blogs, and News Portals, and annotated them for sentimental analysis. Their analysis primarily focused on offensive language detection. They annotated the collected records for 4 categories: Sexist, Racist, Other-Offensive, and Non-Offensive. They used various machine learning and transformer-based models for binary (Offensive vs. Non-Offensive) and multi-class (all 4 classes) sentiment classification.

Apart from this, there have also been some works in medical NLP for the Nepali language. Adhikari et al. [16] used the data from DemetiaBank [43] and performed manual translations to create a Nepali Alzheimer's disease dataset of transcripts from 168 Alzheimer's disease (AD) patients and 98 Control normal (CN) participants. The dataset consisted of a total of 499 transcripts with 255 transcripts belonging to AD patients and 244 belonging to CN participants. They developed a representation of the information included in textual data and presented an NLP-based framework for the early diagnosis of AD patients using Nepali transcripts.

Over time, the field of NLP in Nepali has witnessed certain advancements. However, given the morphologically rich nature of the language [12] and the complex sentence structure [16], NLP in Nepali is particularly challenging and demands vigorous research. NLP advancements in any low-resource language like Nepali are usually restricted by the lack of pre-training data, resource uniformity, and computing resources [35]. The cited works in this section serve as valuable resources that contribute to the ongoing advancement of NLP for Nepali language. Yet, thus far, the focus in Nepali NLP, particularly in sentiment analysis, has been primarily on singular aspects. These encompass offensive language detection, hate speech identification, or rudimentary sentiment classification.

However, a more comprehensive dataset that explores multiple aspects of sentiment analysis and comprehensively captures the diverse spectrum of sentiments in social media discourse remains a gap in the realm of Nepali NLP. Our dataset addresses this issue as we introduce a novel Nepali language dataset, meticulously annotated through a multi-aspect annotation schema. Notably, our dataset also encompasses the detection of hope speech—a previously unexplored segment in Nepali NLP. Furthermore, it's important to note that our dataset is contextualized within a political context. We believe such a dataset will not only enhance the overall landscape of Nepali NLP but will also stimulate the growth of regional languages, empower Nepali users, and narrow the digital language divide by infusing Nepali data into NLP research and applications. Table 1 compares speech datasets in different aspects for multiple languages to give insight into the present condition of the data.

III. DATASET

With growing dissatisfaction with established political parties, the people started to tweet against them seeking change. This is something really important in a political discourse and has to be studied properly. Addressing this issue, we collected tweets spanning from 27 July 2022 to 07 February 2023 using Twitter API that focused on hashtags like '#NoNotAgain', '#EnoughIsEnough', '#wewantchange', '#VoteForChange', and their equivalents in the Nepali language. These campaigns, which first appeared on social media sites like Twitter and Facebook, implore people not to support the more established political parties and their senior established leaders who have repeatedly failed to bring about political stability in the nation.

A. CRITERIA FOR FILTERING THE TWEETS

An essential part of annotating is filtering tweets to exclude unnecessary or misleading information that can skew the findings of the study. Thus, we created a variety of filtering criteria based on the following factors to make sure that our dataset was relevant. We purposefully filtered out tweets published in languages other than Nepali to ensure that the focus was on debates pertaining to the Nepalese election. As long as most of the tweet was written in the Nepali language, our selection technique kept tweets with a few non-Nepali words or phrases. With this strategy, we successfully maintained the key components of the dialogues around

TABLE 1. Summary of other available datasets in the literature.

Works	Data Source	Size	Language	Context	Objective
Sitaula et al. [17]	Twitter	33,247	Nepali	COVID 19	Sentiment Analysis
Niraula et al. [12]	Twitter, Facebook, YouTube (YT)	7,462	Nepali	General discourse	Sentiment Analysis, Hate Speech, Target: Sexist, Racist
Singh et al. [18]	YT	3,068	Nepali	General discourse	Sentiment Analysis, Hate Speech, Target: Person, Organization, Location, and Misc.
Shrestha et al. [29]	News Portals	3,490	Nepali	Nepali News Media	Sentiment Analysis
Zakharov et al. [13]	Reddit	10,559	English	General Discourse	Sentiment Analysis
Macrohon et al. [14]	Twitter	114,851	English, Tagalog	Election	Sentiment Analysis
Mdhaffar et al. [44]	Facebook	17,000	Tunisian	Tunisian Dialects	Sentiment Analysis
Al-Hassan et al. [45]	Twitter	11,000	Arabic	General discourse	Hate Speech, Target: Religious, Racism, and Sexism
NAET (Ours)	**Twitter**	**4,445**	**Nepali**	**Election**	**Hate Speech, Targets of Hate Speech, Direction of Hate Speech, Hope Speech, Sentiment Analysis, Satire**

the anti-establishment movement theme, resulting in a more authentic portrayal of the continuing discussion on social media. Similarly, to ensure the accuracy and reliability of the annotation process, we made a deliberate choice to exclude tweets that lacked clear context or a clear understanding of the Nepalese political discourse.

B. ANNOTATION PROCESS

The annotation process is a vital step in creating a labeled dataset for any analysis or model development task. It involves assigning specific labels or categories to data points, such as tweets in this case. Four annotators with different backgrounds were taken to annotate the data. Annotators had at least 13 years of formal education with Nepali as their language. Annotators had prior experience in annotating datasets in Nepali, Hindi, and English. Accurate and consistent annotations are pivotal to ensure the reliability and validity of the dataset, as well as the results derived from it. To develop a clear and comprehensive set of annotation guidelines, a thorough understanding of the data and the target categories is necessary. The annotation guidelines should be designed in a way that minimizes ambiguity and provides clear instructions to the annotators. Iterative revisions of the guidelines and pilot testing are often conducted to ensure that all annotators have a common understanding of the task and can apply the guidelines consistently. Therefore, the annotators may encounter difficulties during the process. To address this, researchers overseeing the project are crucial. Clarifications and discussions can help resolve uncertainties and improve the overall quality of annotations. Considering the challenge of annotating tweets with text, we developed a 3-phase annotation procedure.

1) 3-PHASE ANNOTATION PROCESS

In order to ensure a consistent and accurate labeling of the dataset, it is imperative to have well-defined annotations. This is significant since the labeled data will serve as the foundation for any analysis or model development findings. Results may be erroneous or unreliable if the annotations are inconsistent or imprecise. Additionally, precise annotations guarantee that any inferences generated from the data are legitimate and that the dataset is reflective of the underlying phenomena being examined. So, we annotate in three phases. We utilized Fleiss' Kappa (κ) as our inter-rater agreement metric in order to objectively evaluate the inter-annotator agreement.

Instructions for annotating the data were developed initially. Iterative revisions were made to the guidelines until they were understood by every annotator. We used a three-phase annotation process to make sure the annotation guidelines were clear. A pilot test was conducted during the first stage to ensure that everyone learned the annotation. The guidelines were checked again in the second step to make sure they were still sufficiently clear. Issues in the annotation were resolved in the third step.

(i) **Pilot Run:** In order to make sure that everyone has the same understanding regarding the annotation guidelines, we execute a pilot annotation for 50 tweets as the first stage of the process. It is crucial that everyone understands what constitutes different types of speech (satire, hope, hate, sentiment) since it might be difficult to classify tweets and because doing so can be stressful. Annotators had some misunderstandings. As a result of a few annotators' demands for unambiguous annotation, we updated the guidelines. To clear up all the ambiguity, the annotation guidelines were then updated.

(ii) **Revised Instructions:** To make sure that the guidelines, which had been updated after the first stage, were enough to understand. All 4 annotators finished the second phase of the annotation of 150 tweets. The modified guidelines were handed to the annotators and they were instructed to annotate the tweets. This stage was crucial to ensure that the updated guidelines were understandable and the annotators could reliably recognize different forms of speech.

(iii) **Conflict Resolution:** The disagreements discovered in the second phase of annotation were the topic of a group discussion by the annotators during the third stage. In order to reach a consensus, they tried to sort out any inconsistencies in their annotations. Disputes were resolved during this stage, and it was vital to guarantee that each tweet had uniform labels. The group discussion also helped to make the instructions more clear and find any remaining ambiguities or inconsistencies.

C. ANNOTATION GUIDELINES

To ensure the reliability and uniformity of the NAET dataset, a crucial step in the annotation process is providing comprehensive guidelines for the annotation team. These guidelines serve as a roadmap, outlining the criteria and considerations for annotating tweets related to anti-establishment election discourse in Nepali. The annotation process began with an in-depth training session for all annotators, where they were introduced to the context of political conversations in the Nepali language. They were provided with examples, along with explanations of the nuances involved in identifying tweets that contribute to anti-establishment discourse.

To maintain consistency, the guidelines discouraged annotators from making subjective judgments or injecting personal opinions while labeling tweets. The focus remained on the tweet's objective content and its alignment with anti-establishment discourse. Annotators were told to maintain an impartial stance and adhere to the predefined criteria. Furthermore, the guidelines underscored the importance of addressing ambiguities through regular discussions and feedback sessions. Annotators collaborated with expert reviewers to resolve any uncertainties or edge cases, ensuring a coherent and accurate dataset. The expert reviewers had a minimum of 13 years of formal education, and their primary language of instruction was Nepali. Expert reviewers had experience in annotation and releasing datasets. Throughout the annotation process, inter-annotator agreement checks were conducted to assess the consistency among annotators. This quality assurance measure helped identify potential discrepancies and reinforced uniformity in the labeling process. By following these comprehensive guidelines, the annotation team could effectively label tweets, contributing to the creation of a high-quality dataset. Table 2 and Table 3 present examples of tweets representing each class label across all the assigned tasks with their translation.

1) ANNOTATION GUIDELINES FOR RELEVANCE

In the context of the NAET dataset creation, the relevance annotation process involved distinguishing tweets that genuinely contribute to anti-establishment election discourse in the Nepali language. To achieve this, it was essential to filter out tweets that merely utilized anti-establishment hashtags for spamming or unrelated purposes. Annotators were told to be vigilant in identifying tweets that exploited anti-establishment hashtags as a means of generating spam or unrelated content. These tweets typically lacked substantive political discourse and consisted of repetitive, noisy, and irrelevant information. Such tweets were labeled as "Non Relevant" to maintain the dataset's quality and ensure it genuinely reflects meaningful political discussions. By adhering to these guidelines and diligently identifying and discarding spam tweets, the relevance annotation process contributed to a high-quality dataset.

2) ANNOTATION GUIDELINES FOR SENTIMENT ANALYSIS

Sentiment analysis plays a crucial role in understanding the emotional nuances and attitudes expressed in tweets related to anti-establishment election discourse in Nepali language. The annotation process required annotators to meticulously assess each tweet's content and determine whether it conveys a "negative," "positive," or "neutral" sentiment toward the political context.

a: IDENTIFYING NEGATIVE SENTIMENT

Tweets with negative sentiment expressed dissatisfaction, criticism, or pessimism towards the political landscape, government, or specific political figures. Annotators were told to look for words or phrases conveying discontent, anger, frustration, disappointment, or disapproval related to anti-establishment issues.

b: IDENTIFYING POSITIVE SENTIMENT

Tweets composed of positive sentiment reflected optimism, approval, or enthusiasm towards potential political changes, new leaders, or emerging parties within anti-establishment discourse. Annotators were instructed to look for words or phrases expressing hope, support, confidence, or excitement for political developments.

c: IDENTIFYING NEUTRAL SENTIMENT

Neutral sentiment tweets did not exhibit any strong emotional tone and provided factual information or statements without

TABLE 2. Examples of annotated tweets corresponding to every class label for annotation aspects of relevance, satire, hope speech and sentiment.

Tasks	Class	Example	Translation
Relevance	Relevant	आज बिहानको घरदैलो। हाम्रो प्रयास तपाईं/तपाईंका छोराछोरीलाई आफ्नै देशमा बस्न योग्य बनाउने हो। तपाईंको साथ, सहयोग बिनाअ सम्भव छ! भोट बाँडिएर पुरानै विकृतिको पोकोले संसदमा प्रवेश नपाओस्। देशको राजनीतिलाई ट्र्याकमा ल्याउनैपर्छ। #wewantchange #EnoughIsEnough	This morning's doorstep. Our endeavor is to enable you/your children to live in their own country. Without your help, it is impossible! By dividing the votes, the old distortions should not enter the parliament. The country's politics must be brought back on track. #wewantchange #EnoughIsEnough
	Non- Relevant	आज २०७९ माघ १५ आइतबारको हिमालय टाइम्स पत्रिका पढ्नुहोस् ।हाम्रो अ नलाइन मा क्लिक गरेर ताजा समाचार पढ्न सक्नुहुन्छ । #HimalayaTimes #News #Cartoon #VoteForChange	Read the Himalaya Times newspaper of Sunday, Magh 15, 2079. You can read the latest news by clicking on our online page. #Himalaya_Times #News #Cartoon #VoteForChange
Satire Detection	Satire	चुनावमा मासुभात बाँड्न तत्पर नेताहरु कोरोना भ्याक्सिन बाँड्ने बेलामा कुन दुलोमा पसेका थिए कुन्नि ? भोट परिवर्तन का लागि होस् मासु र नोटका लागि होईन #EnoughIsEnough	Where were the leaders, who are now eager to serve meat and rice, when they had to distribute COVID vaccines? May the vote be for change not for meat and money. #EnoughIsEnough
	No Satire	राजनीतिमा नैतिकता हुन्छ कि हुदैन? नैतिकता हुन्छ भने जनताबाट अ स्वीकृत भएपछि तपाईंहरुले राजीनामा दिनु पर्छ कि पर्दैन ? #NoNotAgain #Rabidai	Is there morality in politics or not? If it is moral, should you resign after being rejected by the people or not? #NoNotAgain #Rabidai
Hope Speech Detection	Hope Speech	सोच बदलौ देश आफै बदलिन्छ ! राम्रा उम्मेदवारलाई भोट गरौं, चिनेकालाई होइन । इतिहास नदोहोरिऔं र यस पटक हाम्रो देशको लागि राम्रो उम्मेदवार छनोट गरौं। #wewantchange #VoteForChange	Change your thinking, and the country itself will change! Let's vote for the good candidate, not the ones you know. Let's not repeat the history and choose a good candidate for our country this time. #wewantchange #VoteForChange
	No Hope Speech	जस्ले देशलाई १० बर्ष पछाडी धकलियो उस्लाई देशको अ र्थतन्त्रको चिन्ता भन्नेअ भिव्यक्ती त अ लि अ पच हुने रैछ । #EnoughIsEnough #KPOli	Those who pushed the country back 10 years are worried about the country's economy. #EnoughIsEnough #KPOli
Sentiment Detection	Positive	लोकतन्त्रमाथि हामीलाई पूर्ण विश्वास छ र हरेक परिवर्तन 'ब्यालेट'बाट सुरु हुनुपर्छ भन्ने हामीलाई लाग्छ । पुराना नेताले जति गरे त्यसको लागि धन्यबाद तर अ ब परिवर्तन आवश्यक । #EnoughIsEnough #change	We have full faith in democracy and we think that every change should start with 'ballot'. Thank you to the old leaders for what they have done but now a change is needed. Let's hope that they can be honored in the next election. #EnoughIsEnough #change
	Negative	ल है गोर्खाली हो, योअ वसरवादी घुमन्ते लाई यस पली लात हानेर चितवन नै फर्काइदिनु पर्यो। जनताको नेता, गोर्खाको विकास गर्न आका हैन मात्र सांसद बन्न आका यहाँ। #VoteForChange	Hey Gorkhali, this opportunistic nomad had to be kicked and sent back to Chitwan. The leader of the people, he is not here to develop Gorkha, but he is here to become a Member of Parliament (MP). #VoteForChange
	Neutral	आजदेखि देशभर चुनावी चहलपहल सुरु भएको छ । के तपाई अ सल उम्मेदवारलाई भोट दिनुहुन्छ? '#NoNotAgain' '#EnoughIsEnough'	From today election is starting all around the country. Are you going to vote for a good candidate? '#NoNotAgain' '#EnoughIsEnough'

expressing explicit positivity or negativity. The annotators were told to look for the tweets that presented unbiased observations, news updates, or statements without overt emotional language. They also considered tweets that discuss political events or issues in an objective manner without taking a clear stance.

TABLE 3. Examples of annotated tweets corresponding to aspects of hate speech detection along with direction and targets of hate speech.

Tasks	Class	Example	Translation
Hate Speech Detection	Hate Speech	एक पटक यो हत्याराको भाषण सुनौं त कसरी १५-१६ वर्षका युवाहरुको दिमाग भुटेको रहेछ। लड्न र उक्साएर मर्न तयार पारेको भाषण। युद्धमा होमेर १७ हजार मान्छे मार्यो आफ्नी छोरीलाई विदेश पढ्न पठाएर अ हिले सांसदमा उठाको छ। #VoteForChange	Let's listen to the speech of this murderer once and see how the minds of 15-16 year old youths were brainwashed. A speech prepared to fight and provoke. He killed 17 thousand people in the war, he sent his daughter to study abroad and made her a candidate for election. #VoteForChange
	No Hate Speech	यदि तपाईंको क्षेत्रमा कुनै पनि उम्मेदवारले शक्तिको अ नुचित प्रयोग गरि मतदानलाई प्रभावित पारेका छन भने हामी सार्वजनिक चाँसोका मुद्दा (PIL) मार्फत सर्वोच्च अ दालत जान सक्छौं। #EnoughIsEnough	If any candidate in your constituency has improperly used power to influence the polls, we can approach Supreme Court through Public Interest Litigation (PIL). #EnoughIsEnough
Targets of Hate Speech Detection	Individual	देउवा, ओली, प्रचण्ड, माधव नेपाल, र झलनाथ खनाल लगातार तिनै निकम्मा बुढाबाट अ ब शासित हुनु छैन। आफ्नो त जिन्दगी बर्बाद भयो भयो आगामि पुस्ताकै जिन्दगी बर्बाद भएको अ ब हेर्नु छैन। #Deuba #wewantchange #Prachanda #Oli	We don't want to be ruled by same old Deuba, Oli, Prachanda, Madhav Nepal and Jhalnath Khanal continuously. Our life is wasted, we will not see the lives of future generations wasted. #Deuba #wewantchange #Prachanda #Oli
	Organization	मेरो एक भोट खेर जाओ तर एमाले/माओवादी/गठबन्धन लाई भोट हाल्दिन । यो यिनीहरु भोट हालेर पनि मेरो भोट खेर नै गएको थियो। #VoteForChange	I will waste one of my votes but not vote for UML/Maoist/Alliance. Even after voting for them, my vote was wasted.#VoteForChange
	Community	अ क्सर नेपाली जनताहरु भेडा र झोले हुन्। राम्रो प्रतिनिधिलाई कहिल्यै जिताउने काम गर्दैनन् भनेर चियापसल/जमघटमा बहस गर्ने हाम्रा सो कल्ड शिक्षित र बुद्धिजीवी राजधानीका जनताहरु। आज बालेन शाहलाई जिताउन लौरो चिन्हमा भोट हाले होलान् त? #NoNotAgain	Nepali people are often like a sheep and non-verbal. Our so-called educated and intellectual capital people who argue in tea shops/gatherings that a good representative will never win. Did they vote for Balen Shah today in the stick symbol? #NoNotAgain
Direction of Hate Speech Detection	Directed	जे होस प्रचण्ड ले चाहिँ हार्नै पर्छ। एकदम न मीठो हार हुनु पर्छ।यस्तो न बिचार न सिद्धान्त। छिमेकीले जन्माएको उसैको लागि हुन्छ भन्ने प्रमाणित गरिसकेको महान राजनैतिक नेता नेपाल लाई चाहिएको हैन। #EnoughIsEnough #wewantchange	Whatever happens, Prachanda must lose. There should be a very bitter defeat. Neither thought nor principle. Nepal does not need a great political leader who has proved that what is born by the neighbour is for him. #EnoughIsEnough #wewantchange
	Undirected	कुहिएको, सडेका, बुद्धि भाँडिएका, विदेशीका दलालहरुलाई जनताले चिनिसके। अ ब भागेर दिल्ली जाउ, नोएडामा राजनिति शुरु गर। तिमीहरु जस्ता यो देशका भष्मासुर हरुलाई जनताले भोट हाल्ने छैनन्। #VoteForChange #NoNotAgain	The people came to know the decayed, rotten, brainless, foreign brokers. Now run away to Delhi, and start politics in Noida. People will not vote for monsters like you. #VoteForChange #NoNotAgain

It is important to note that sentiment analysis is context-dependent, and annotators were made to consider the overall theme and intent of each tweet to make accurate judgments. Tweets sometimes contained multiple sentiments or ambiguous expressions, requiring careful assessment, and, when necessary, discussions with expert reviewers were done to reach a consensus.

3) ANNOTATION GUIDELINES FOR SATIRE ANALYSIS

Satire analysis in the NAET dataset aimed at identifying tweets that employed satire or ironic humor to critique or comment on anti-establishment election discourse in Nepali. To effectively annotate for satire, annotators analyzed each tweet's content, paying close attention to the use of irony, sarcasm, or parody. The guidelines provided clear criteria and examples to ensure consistent and accurate labeling.

(i) **Satirical Content:** Tweet was labeled as ‘‘satirical’’ when it employed humor, wit, or exaggeration to comment on political events, figures, parties, or electoral situations. Satirical tweets often convey implied meanings that contrast with the literal interpretation, aiming to offer criticism or ridicule with a humorous undertone.

(ii) **Identifying Satire:** Annotators recognized the subtle and indirect nature of satire, where the tweet may not explicitly express criticism but uses clever wordplay or context to convey a critical or humorous perspective. Understanding the implied meanings and intended humor was crucial for accurate labeling.

(iii) **Avoiding Misinterpretation:** Annotators were guided to avoid misinterpreting tweets expressing legitimate frustration or criticism as satirical, focusing instead on the humorous and ironic elements present in the tweet.

(iv) **Cultural Context:** The guidelines accounted for cultural references and specific cultural norms that may impact the perception of satire. Annotators considered the context in which the satire is presented to ensure cultural relevance and accuracy.

The guidelines offered illustrative examples of satirical tweets within the context of anti-establishment discourse. Annotators referred to these examples to gain a deeper understanding of the varying forms of satire present in political conversations. The accurate identification of satirical content enhances the NAET dataset’s value and empowers researchers to explore the interplay of satire and politics in the digital sphere.

4) ANNOTATION GUIDELINES FOR HATE SPEECH ANALYSIS

Hate speech analysis aims to accurately identify and distinguish tweets that contain hate speech from those that do not. The following detailed guidelines provide a comprehensive approach to the hate speech analysis:

Hate Speech: Hate speech is defined as any language that promotes violence, hostility, discrimination, or prejudice based on attributes such as race, ethnicity, religion, gender, or political affiliations. Tweets containing derogatory or offensive content targeting individuals or groups based on such attributes were labeled as ‘‘hate speech.’’ It included content that propagated harmful stereotypes or incited animosity towards specific groups. Annotators carefully assessed the language and context of each tweet to determine whether it qualified as hate speech. They looked for explicit instances of derogatory terms, slurs, or offensive language targeting individuals or communities. Annotators considered the intent behind the language used in the tweet. Hate speech often seeks to demean, insult, or harm a particular group, while non-hate tweets may express strong opinions or criticisms without promoting violence or discrimination. The annotators also evaluated the tweet’s tone and emotional content. Hate speech often exhibits anger, hostility, or animosity towards the targeted group, while non-hate tweets may express disagreements in a respectful manner.

Non-Hate Speech: Annotators have to distinguish between tweets that engage in healthy political discussions and those that incite hatred or harm. Non-hate tweets were distinct from hate speech in that they did not promote violence, hostility, discrimination, or prejudice based on any attributes mentioned above. Non-hate tweets expressed strong opinions, criticisms, or disagreements related to anti-establishment discourse without resorting to offensive language or attacking specific groups. Annotators were told to be cautious in distinguishing non-hate tweets from hate speech, as certain tweets contained subtle expressions of discontent without crossing into hate speech territory.

By adhering to these comprehensive guidelines for hate speech and non-hate tweet analysis, annotators contributed to a reliable and accurate identification of hate speech in the context of anti-establishment election discourse in Nepali.

5) ANNOTATION GUIDELINES FOR DIRECTION OF HATE SPEECH:

In the context of anti-establishment election discourse, hate speech can take two distinct directions: directed hate and undirected hate. Annotators assessed each tweet containing hate speech and categorized it based on the direction of the hateful expression.

a: DIRECTED HATE

In the realm of political discourse, directed hate speech serves as a potent tool to attack specific individuals, organizations, or entities. Such targeted expressions aim to demean, vilify, or incite harm toward their subjects based on their attributes, actions, or affiliations. In the context of anti-establishment election discourse, directed hate speech may be employed to discredit political opponents, delegitimize opposing views, or mobilize support for particular ideologies.

Annotators looked for explicit use of derogatory terms, slurs, or offensive language specifically aimed at individuals, organizations, or communities. They evaluated the tweet’s context to identify if it singles out and attacks a specific subject based on its attributes or actions. The annotators also considered the intent behind the language used in the tweet and whether it aims to inflict harm or incite animosity towards the mentioned subject. They also assessed the emotional content of the tweet to determine if it exhibits hostility or animosity toward the targeted individual or group.

b: UNDIRECTED HATE

In contrast to directed hate speech, undirected hate expressions lack specific targets and may encompass more generalized negative statements. These expressions may include prejudiced or hostile language towards a community, ethnic group, religious group, or any collective entity without singling out particular individuals or organizations.

Annotators distinguished between undirected hate speech from other strong expressions of discontent or criticism within the anti-establishment discourse. Identifying undirected hate speech is crucial in understanding the broader prevalence of hateful language and its potential impact on social cohesion and political discussions. The annotators looked for tweets that contained generalized negative statements without targeting specific individuals or organizations. They also evaluated the language used in the tweet to identify whether it exhibits a broader expression of animosity or prejudice without pinpointing particular targets.

By adhering to these guidelines for directed and undirected hate speech annotation, annotators categorized tweets based on the direction of hateful expressions within the anti-establishment election discourse in Nepali.

6) ANNOTATION GUIDELINES FOR TARGETS OF HATE SPEECH

The annotation process for identifying the targets of hate speech involved categorizing the hate speech into three main target categories: **organization, individual,** and **community**. Annotators analyzed the content of each tweet to determine the specific target category accurately. The following guidelines provided a detailed approach to annotating the targets of hate speech:

Organization: Hate speech may target specific organizations, institutions, companies, or any formal entity. Annotators were told to be vigilant in recognizing tweets that contain hateful content directed toward such organizations. The hate speech may be based on attributes, ideologies, actions, or any other characteristic associated with the organization. So, annotators looked for explicit mentions of organizations or keywords that indicate the target as an organization.

Individual: Hate speech may be directed at specific individuals, including public figures, politicians, activists, or any person referenced in the tweet. It aims to harm, demean, or incite hostility towards the targeted individual based on personal attributes, beliefs, actions, or any other characteristic. Annotators examined the tweet for direct references to individuals or personal attacks. They also looked for mentions of names, titles, or any identifiers that indicate the target as an individual. They analyzed the language and tone used in the tweet to determine if the content was hateful toward a specific person.

Community: Hate speech may generalize negative sentiments towards a particular community, ethnic group, religious group, or any collective entity. The content expresses prejudice, discrimination, or hostility towards the entire community based on their shared characteristics, beliefs, or cultural backgrounds. Annotators were told to be attentive to language that portrayed generalized negativity towards a community or group of people. They looked for keywords, slurs, or derogatory terms that indicate the target as a community.

In cases where the tweet contains hate speech targeting multiple categories (organization, individual, and community), annotators were told to prioritize the primary target based on the tweet's content and intent. If there was ambiguity in categorizing the target, annotators engaged in discussions with expert reviewers to reach a consensus and ensure consistent annotation.

The accurate identification of hate speech targets provides valuable insights into the diverse nature and impact of hate speech on organizations, individuals, and communities within the context of anti-establishment election discourse in the Nepali language. With such comprehensive guidelines, we were able to build a well-annotated dataset.

7) ANNOTATION GUIDELINES FOR HOPE SPEECH ANALYSIS

Hope speech analysis for the NAET dataset aimed to identify and differentiate tweets that contain expressions of hope from those that do not. In the context of anti-establishment election discourse, hope speech reflects the aspirations, expectations, and positive outlook towards potential political changes, new leaders, or emerging parties. In the scope of our annotation, hope speech refers to language that expresses optimism, positive expectations, and aspirations for political developments, reforms, or changes. Annotators were asked to look for tweets that convey positive sentiments, express optimism, or anticipate favorable outcomes in the context of the political landscape. The annotators considered the tweet's context and whether it envisions a brighter future, improvement in governance, or positive shifts in policies and leadership. They also evaluated the language's tone and emotional content, looking for expressions of confidence, encouragement, and enthusiasm.

a: DIFFERENTIATING HOPE SPEECH FROM OTHER SENTIMENTS

Accurate differentiation of hope speech from other kinds of sentiments is crucial to ensure the dataset's integrity and provide meaningful insights into the political discourse. Annotators were told to be cautious in identifying and distinguishing hope speech from various sentiments to maintain the dataset's precision and relevance.

(i) **Distinguish from General Positivity:** Hope speech revolves around positive expectations and aspirations for political developments. Annotators were told to avoid labeling tweets as hope speech solely based on general positive expressions, such as happiness, delight, or personal achievements. Instead, they were told to focus on language that specifically addresses political changes or advancements.

(ii) **Separate from Mere Support:** Mere expressions of support for political figures or parties do not constitute hope speech. Annotators were told to be attentive to tweets that primarily indicate allegiance to specific entities without expressing optimism for broader political improvements.

(iii) **Consider the Context:** The tweet's context plays a pivotal role in accurately identifying hope speech.

Annotators were told to assess whether the positive sentiment is directly related to political changes, governance, or reforms, rather than being unrelated to the anti-establishment discourse.

(iv) **Mind Negation and Sarcasm:** The annotators were told to be cautious in identifying negation, where seemingly positive statements may be negated to express doubts or cynicism regarding the political landscape. Additionally, consider instances of sarcasm, where positive language may actually convey negative or ironic undertones toward political developments.

(v) **Focus on Aspirations for Change:** Annotators were guided to look for expressions of confidence, encouragement, and enthusiasm towards potential political advancements, improvements, or transformation. It is important that the speeches reflect a sense of aspiration for a better political future.

(vi) **Differentiate from Desperation:** Expressions of hope may sometimes emerge from a place of desperation or dissatisfaction with the current political scenario. Annotators were told to focus on the underlying optimistic tone and expectations for positive changes.

By adhering to these comprehensive guidelines for hope speech analysis, annotators can accurately identify and categorize tweets that contain expressions of hope within the anti-establishment election discourse in Nepali. Properly annotated hope speech data offers valuable insights into the prevalence of positive sentiment and aspirations within political discussions, contributing to a more nuanced understanding of public attitudes towards potential political changes and new emerging figures or parties.

D. DATASET STATISTICS AND ANALYSIS

A total of 4,445 tweets were labeled for relevance with 4,252 of them labeled as 'Relevant' and 193 as 'Non Relevant'. For the relevant tweets, there were a total of 6 aspects of annotation i.e., **Hate speech, Targets** of Hate speech, **Direction** of hate speech, **Sentiment, Satire**, and **Hope speech**. The hate speech was annotated as 'Hate' and 'No Hate' with total tweets of 833 (19.60%) and 3,419 (80.40%) respectively. The Targets of the hate speech were 'Individual', 'Community', and 'Organization'. 'Individual' comprised 320 (59.47%) tweets of hate speech, 'Community' had 101 (18.77%), and 'Organisation' had 117 (21.76%) tweets of hate speech. Additionally, it is important to highlight that hate speech and hope speech typically contain a substantially greater number of words compared to no-hate speech and no-hope speech respectively. The Sentiment had three categories: 'Neutral', 'Positive', and 'Negative'. 'Neutral' comprised 2,169 (51.01%) tweets of total sentiment tweets, 'Positive' had 927 (21.80%), and 'Neutral' had 1,156 (27.19%) tweets of total sentiment tweets. The Satire class was divided into two labels: 'Satire', and 'No-Satire'. 'Satire' made up 867 (20.39%) and "No-Satire" had 3,385 (79.61%) of the total satire class. Accordingly, Hope Speech had two labels i.e., 'Hope', and 'No-Hope'. 'Hope' was composed of 556 (10.58%) and 'No-Hope' had 3,709 (89.44%) tweets of total hope speech. Finally, the Direction of hate speech had two labels i.e., 'Directed', and 'Undirected'. 'Directed' had 538 (64.58%) tweets and 'Undirected' had 295 (35.42%) tweets of hate speech labeled as 'Hate'. Table 4 presents the data statistics, including the average character count and average word count.

1) INTER-ANNOTATOR AGREEMENT

We utilized Fleiss' Kappa to determine the level of inter-annotator agreement between four annotators. Compared to basic percent agreement calculations [46], it is a more reliable indicator of inter-rater agreement. The Fleiss' Kappa for the two-class annotation of "Hate vs. No-Hate" is 0.71. "Hope vs. No-Hope" involves two-class annotation and has a Fleiss' Kappa of 0.74. Similarly, the two-class annotation of "Satire vs. Non-Satire" has Fleiss' Kappa of 0.72. Also, the two-class annotation of "Directed vs. Non-Directed" hate speech which has Fleiss' Kappa of 0.79. Similarly, the 3-class annotation of "Sentiment: Neutral vs Positive vs Negative" has Fleiss' Kappa of 0.81. Finally, the 3-class annotation of "Targets: Individual vs Community vs Organisation" has Fleiss' Kappa of 0.74.

2) TOPIC MODELING

The notion of topic modeling is made up of subjects such as 'words', 'documents', and 'corpora'. A 'word' is regarded as the fundamental component of discrete data in a text and is defined as a piece of vocabulary that is indexed for each distinct word in the document. The term 'document' refers to a grouping of N words. A corpus is a group of M documents, while corpora are the plural version of the corpus. While topic' refers to the distribution of some fixed vocabulary. Simply explained, each document in the corpus has a different proportion of the topics mentioned based on the terms used in it [47]. In recent years, machine learning and natural language processing fields have paid much attention to probabilistic graphical models. Latent Dirichlet Allocation (LDA), among other statistical topic models, offers a potent framework for describing and condensing the content of enormous document collections [48]. LDA has had a significant impact on the domains of statistical machine learning and natural language processing, and it has swiftly emerged as one of the most well-liked probabilistic topic modeling techniques in machine learning [49].

Let us assume we have a corpus, which is a collection of M documents, denoted by $D = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \ldots, \mathbf{w}_M\}$. $\mathbf{w} = \{w_1, w_2, w_3, \ldots, w_N\}$, where w_n is the n^{th} word in the sequence, represents a document as a sequence of N words and the corresponding vocabulary of a word, indexed by $\{1, 2, \ldots, V\}$. For each document $\mathbf{w}$ in a corpus D, LDA presupposes the following generative process:

1) Select $N \sim \text{Poisson}(\xi)$ and $\theta \sim \text{Dir}(\alpha)$.
2) For each N-word, w_n :

TABLE 4. Statistics of the NAET dataset. The values in parenthesis for average characters and words show statistics after preprocessing.

Tasks	Labels	Tweets	Avg. Char	Avg. Words
Hate Speech	Hate	833 (19.60%)	178.66 (120.18)	26.62 (18.50)
	No-Hate	3,419 (80.40%)	149.82 (95.95)	22.27 (14.78)
Sentiment	Neutral	2,169 (51.01%)	138.77 (86.71)	20.64 (13.43)
	Positive	927 (21.80%)	172.36 (114.15)	25.57 (17.46)
	Negative	1,156 (27.19%)	172.89 (116.12)	25.76 (17.83)
Satire	Satire	867 (20.39%)	149.89 (99.85)	22.76 (15.62)
	No-Satire	3,385 (79.61%)	156.94 (100.87)	23.22 (15.48)
Hope Speech	Hope	556 (10.56%)	181.45 (120.83)	26.42 (18.32)
	No-Hope	3,696 (89.44%)	151.58 (100.87)	22.63 (15.08)
Targets	Individual	320 (59.47%)	177.04 (121.57)	26.55 (18.86)
	Community	101 (18.77%)	186.61 (119.30)	27.73 (18.64)
	Organization	117 (21.76%)	185.77 (124.56)	28.19 (19.05)
Direction	Directed	538 (64.58%)	180.78 (121.80)	27.14 (18.86)
	Undirected	295 (35.42%)	174.68 (117.25)	25.63 (17.83)

a. Select a topic, $z_n \sim$ Multinomial (θ).
b. Select a word w_n from $p(w_n \mid z_n, \beta)$, a multinomial probability conditioned on the topic z_n

Where ξ, α, and β represent Average Document Length (Poisson Distribution Parameter), Dirichlet Distribution Parameter for Topic Distribution θ, and Word Distribution for Each Topic respectively.

The probability density of the (k-1)-simplex for a k-dimensional Dirichlet random variable is as follows:

$$p(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \theta_1^{\alpha_i - 1} \ldots \theta_k^{\alpha_k - 1} \tag{1}$$

The joint distribution of a topic mixture θ, a collection of N topics z, and a set of N words w, given the parameters α and β, is given by:

$$p(\theta, \mathbf{z}, w|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta) p(w_n|z_n, \beta) \tag{2}$$

and,

$$p(\theta, \mathbf{z}, w|\alpha, \beta) = \int p(\theta|\alpha) \left(\prod_{n=1}^{N} \sum_{z_n} p(z_n|\theta) p(w_n|z_n, \beta) \right) d\theta \tag{3}$$

Topic modeling for our dataset is divided into numerous stages in order to produce precise results. The LDA approach was used in the research to simulate the topic and generate clusters for each topic discourse. First, we take our raw unannotated dataset and perform data cleaning to remove unnecessary characters like hashtags and mentions. We also remove the duplicate data. To improve the data's relevance and better prepare it for our model, we also removed the stopwords. After that, tokens were made and fitted to the LDA model. The result was observed and the topic number was adjusted to obtain a relevant result. Finally, the result was drawn and the topic name was given to each cluster.

After data pre-processing and tokenizing the corpus, the topics were generated. The cluster was observed carefully and it was noted that no clusters should overlap each other. By doing so, the same words did not repeatedly come across two or more clusters. The evaluation of the topic modeling was done by using the coherence score. When calculating coherence scores, high-scoring words in a topic's vocabulary are compared in terms of their semantic closeness. The more positive value of the coherence score, the more correct topics will be generated. However, this is not always the case as clusters can overlap and generate a high coherence score. In our analysis, as given in Figure 1, the best coherence score of 0.504 was observed for the number of topics as 10. However, as seen in Figure 2, we have 10 as the choice of topics generated clusters that overlapped with each other. The next choice for the number of topics could have been 7 since the coherence score spiked there but as seen in Figure 3, the problem of overlapping persisted in 7 topics as well. After many trials, our experiments showed that using 4 as the number of topics generated good clusters that had no overlapping. Figure 4 shows the visualization of the clusters with 4 as the number of topics selected and a coherence score of 0.374. We also experimented with perplexity which measures how well the model has learned the underlying patterns and structure of the data. However, we did not include it in our analysis since perplexity is not very helpful in determining the ideal number of topics [50].

Table 5 gives the list of 4 topics along with the most prevalent words in each topic. Also, Figure 5 gives the word cloud overview of the four topics. Next, based on the qualitative evaluation, we describe and name the theme of each of the four topics.

TABLE 5. Top-10 prevalent words from each topic given by topic modeling.

Topic	Domain	Prevalent Words
First Topic	Political Parties and Leaders	नेता (Leader), आयोग (Commission), चुनावमा (In the election), एमाले (UML - Political party), पार्टी (Party), प्रचण्ड (Prachanda- A political leader), देउवा (Deuba - Political leader), राजनीति (Politics), सरकार (Government), ओली (Oli - A political leader)
Second Topic	Governance	भोट (Vote), देश (Country), चुनाव (Election), पार्टी (Party), जनताको (Of the people), प्रधानमन्त्री (Prime Minister), नेपाली (Nepali), सरकार (Government), राजनीतिक (Political), बिचार (Thought/Consideration)
Third Topic	Socio-political Dynamics	युवा (Youth), नेपाल (Nepal), सत्ता (Power), सम्मान (Respect), उम्मेदवार (Candidate), स्वतन्त्रता (Freedom), नेताले (By the leader), मतदान (Voting), देश (Country), नेताहरु (Leaders)
Fourth Topic	Democratic Processes	निर्वाचन (Election), ⊠धिकार (Right), आयोगको (Of the commission), ⊠भिव्यक्ति (Expression), जनतालाई (To the people), नेतृत्व (Leadership), पाउने (To get), योगदान (Contribution), पार्टीका (Of the party), बहिस्कार (Boycott)

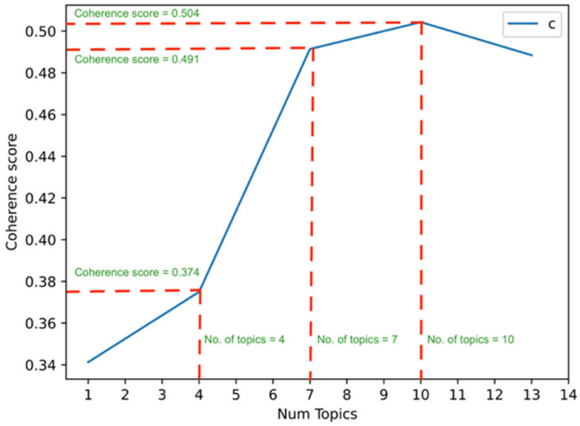


FIGURE 1. Number of topics compared to coherence score.

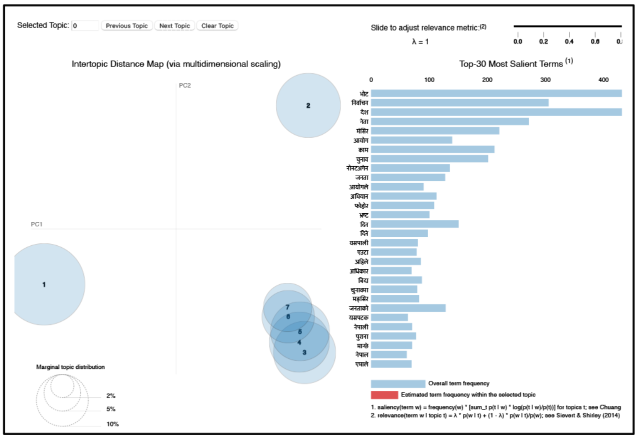

Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0 0.2 0.4 0.6 0.8 1.0
Intertopic Distance Map (via multidimensional scaling)
Top-30 Most Salient Terms (1)
0 100 200 300 400
PC1
PC2
1
2
3
4
5
6
7
Marginal topic distribution
2%
5%
10%
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

FIGURE 3. Discarded topic modeling (no. of topics = 7).

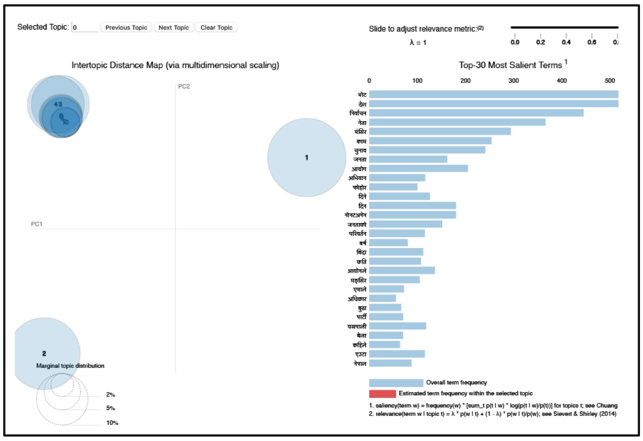

Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0 0.2 0.4 0.6 0.8 1.0
Intertopic Distance Map (via multidimensional scaling)
Top-30 Most Salient Terms 1
0 100 200 300 400 500
PC1
PC2
1
2
Marginal topic distribution
2%
5%
10%
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

FIGURE 2. Discarded topic modeling (no. of topics = 10).

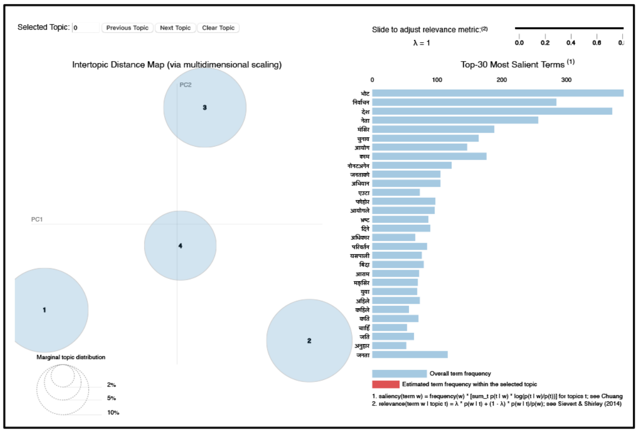
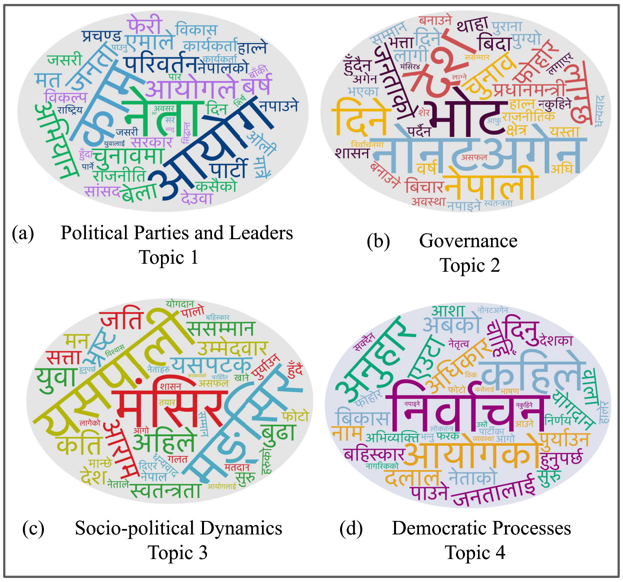

Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0 0.2 0.4 0.6 0.8 1.0
Intertopic Distance Map (via multidimensional scaling)
Top-30 Most Salient Terms (1)
0 100 200 300
PC1
PC2
1
2
3
4
Marginal topic distribution
2%
5%
10%
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

FIGURE 4. Selected topic modeling (no. of topics = 4).

(i) **Political Parties and Leaders:** The first cluster of words included frequently used names of the political parties like एमाले (UML) and political leaders such as प्रचण्ड (Prachanda), देउवा (Deuba) and ओली (Oli).

(ii) **Governance:** The second topic represented the process, systems, and government that make decisions, enact laws, and guarantee a fair and effective functioning of its affairs. Words like भोट (Vote), प्रधानमन्त्री (Prime Minister), सरकार (Government), राजनीतिक (Political) portrayed key aspect of democratic governance.

(iii) **Sociopolitical Dynamics:** The third cluster represented the intricate interaction between social and political forces within a society or group. Words like



(a) Political Parties and Leaders Topic 1

(b) Governance Topic 2

(c) Socio-political Dynamics Topic 3

(d) Democratic Processes Topic 4

FIGURE 5. Wordcloud for each topic given by topic modeling.

युवा (Youth), सम्मान (Respect), स्वतन्त्रता (Freedom), नेताहरु (Leaders) constituted sociopolitical dynamics of a country.

(iv) **Democratic Process:** The fourth topic included words such as निर्वाचन (Election), आयोगको (Of the commission), and जनतालाई (To the people) which hinted towards the democratic processes within a country.

E. EXPLORATORY DATA ANALYSIS

As a part of exploratory data analysis, we find the top 10 words in our dataset along with the top 10 words in different annotation aspects. We use TF-IDF to extract the relevant words according to the weight. The statistical technique TF- IDF [51] is important for determining a word's importance within a corpus of documents. The TF-IDF score has two parts: the TF (Term Frequency) component, which shows how frequently a term appears in a specific document, and the IDF (Inverse Document Frequency) component, which shows how common or uncommon the word is over the entire set of documents. We calculate the TF-IDF score by multiplying the TF and IDF scores together.

$$tf\ idf(t,d,D) = tf(t,d) * idf(t,D) \tag{4}$$

where,

$$tf(t,d) = log\,(1 + freq\,(t,d)) \tag{5}$$

$$idf(t,d) = log\left(\frac{N}{count\ (d \in D : t \in d)}\right) \tag{6}$$

Simple explanation of the formula:

$$TF = log\left(\frac{X}{Y}\right) \tag{7}$$

$$IDF = log\left(\frac{A}{B}\right) \tag{8}$$

$$TF - IDF = TF * IDF \tag{9}$$

where,

X = Number of occurrences of the particular word in the document

Y = Total number of words in the document

A = Count of the documents in the corpus

B = Word appears in numerous documents in the corpus

The terms Vote (भोट), Country (देश), Leader (नेता), and Election (निर्वाचन) are highly significant in the majority of tasks. Every individual word is accompanied by its corresponding translation and an associated TF-IDF score. Table 6 displays the information for the categories: All Posts from the dataset, Non-Hate Speech Posts, and Hate Speech Posts. Table 7 represents the top words for the Targets of Hate Speech classes. Table 8 illustrates the prominent terms associated with the Sentiment Analysis tasks. Table 9 represents the noteworthy terms linked with the Hope Speech category and Satirical Posts. Table 10 shows the significant terms associated with the No- Satirical Posts and the category of Hope Speech. In Figure 6, a basic depiction of the word found in our dataset is provided in the form of wordcloud.

The histogram for the total number of characters before pre-processing and after pre-processing is shown in Figure 7 (a) and Figure 7 (c) respectively. Similarly, the histogram for the total number of words before pre-processing and after pre-processing is shown in Figure 7 (b) and Figure 7 (d) respectively. We can see the reduction in character and word counts as we removed Twitter mentions (@), hashtags (#), and stopwords like Or (वा), This (यो), And (र), and You (तिमी). The average number of words and characters per tweet across all classes is shown in Table 4. Diving deeper into the insights provided by TF-IDF, we can assess the discourse in terms of the unique characteristics provided by the frequency of words. Along with it, the length of tweets can also provide major insights into the kind of discourse. For example, it is clear that tweets marked as 'Hate' include many more words on average compared to tweets marked as 'No Hate'. This is consistent with the observation that most of the 'No Hate' tweets that analysts observed were just helpful tweets. On the other hand, people frequently engaged in lengthy discussions about how the present political structure has weaknesses in the 'Hate' tweets. To ascertain the fundamental causes for this variation in tweet length, more research is required.

IV. BENCHMARKS AND ANALYSIS

A. BASELINES

1) **Naive Bayes (NB):** NB is a probabilistic machine learning algorithm commonly used for classification [52]. It is based on Bayes' theorem which is used to find the probability of an event occurring given the probability of another event that has already occurred. It earns its name "naive" from an assumption that, given a class label, all features are independent. Although a simple algorithm, Naive Bayes is efficient and works surprisingly well on tasks such as text classification and sentimental analysis.

TABLE 6. Top-10 most frequent words in the overall dataset and also for class belonging to hate speech and no hate speech.

All Posts			No Hate Speech Posts			Hate Speech Posts		
Words	**Translation**	**TF-IDF**	**Words**	**Translation**	**TF-IDF**	**Words**	**Translation**	**TF-IDF**
भोट	Vote	0.1084	भोट	Vote	0.1133	भोट	Vote	0.1109
देश	Country	0.0668	देश	Country	0.0641	देश	Country	0.0898
नेता	Leader	0.0499	नेता	Leader	0.0473	नेता	Leader	0.0706
निर्वाचन	Election	0.0424	निर्वाचन	Election	0.0427	सबै	All	0.0616
आफ्नो	Own	0.0377	आफ्नो	Own	0.0387	जनता	People	0.0525
सबै	All	0.0350	मंसिर	Mangsir	0.0360	निर्वाचन	Election	0.0519
मात्र	Only	0.0338	मात्र	Only	0.0352	दिन	Day	0.0396
चुनाव	Election	0.0335	चुनाव	Election	0.0342	फेरि	Again	0.0385
काम	Work	0.0333	काम	Work	0.0329	पटक	Times	0.0376
पटक	Times	0.0330	पटक	Times	0.0329	चुनाव	Election	0.0369

TABLE 7. Top-10 most frequent words in each hate speech target class. The TF-IDF scores are given for each word.

Target: Individual			Target: Community			Target: Organization		
Words	**Translation**	**TF-IDF**	**Words**	**Translation**	**TF-IDF**	**Words**	**Translation**	**TF-IDF**
भोट	Vote	0.1011	भोट	Vote	0.1106	निर्वाचन	Election	0.1942
देश	Country	0.0757	सबै	All	0.0934	आयोग	Commission	0.1463
काम	Work	0.0590	देश	Country	0.0922	एमाले	UML	0.1070
प्रचण्ड	Prachanda	0.0571	नेता	Leader	0.0921	भोट	Vote	0.1053
दिन	Day	0.0568	सम्म	Until	0.0633	जनता	People	0.0898
ओली	Oli	0.0518	जनता	People	0.0581	चुनाव	Election	0.0729
नेता	Leader	0.0538	केही	Something	0.0548	देश	Country	0.0648
प्रधानमन्त्री	Prime Minister	0.0489	भेडा	Sheep	0.0498	सबै	All	0.0641
गरेर	By doing	0.0463	फेरि	Again	0.0476	माओवादी	Maoist	0.0598
पटक	Times	0.0402	पछि	Next	0.0465	गठबन्धन	Alliance	0.0428

TABLE 8. Top-10 most frequent words in each sentiment class. The TF-IDF scores are given for each word.

Sentiment: Neutral			Sentiment: Positive			Sentiment: Negative		
Words	**Translation**	**TF-IDF**	**Words**	**Translation**	**TF-IDF**	**Words**	**Translation**	**TF-IDF**
भोट	Vote	0.0892	भोट	Vote	0.1106	भोट	Vote	0.1166
निर्वाचन	Election	0.0519	देश	Country	0.1030	देश	Country	0.0903
देश	Country	0.0430	नेता	Leader	0.0573	नेता	Leader	0.0655
नेता	Leader	0.0428	आफ्नो	Own	0.0559	जनता	People	0.0493
होला	May	0.0393	पटक	Times	0.0489	निर्वाचन	Election	0.0462
चुनाव	Election	0.0369	जनता	People	0.0436	सबै	All	0.0457
सबै	All	0.0335	परिवर्तन	Change	0.0418	चुनाव	Election	0.0398
आफ्नो	Own	0.0323	मात्र	Only	0.0405	आफ्नो	Own	0.0371
मंसिर	Mangsir	0.0319	मंसिर	Mangsir	0.0400	काम	Work	0.0367
पटक	Times	0.0291	मत	Vote	0.0353	पटक	Times	0.0315

2) **Decision Tree (DT):** Decision Tree is a non-linear supervised learning algorithm that can be used for both regression and classification problems. This method uses a tree-like structure to represent a problem using the CART (Classification and Regression Tree) algorithm, where each node represents a decision based on a feature, branches represent the possible values of that feature, are prone to overfitting, especially when the tree becomes too deep or complex. This is where ensemble methods like Random Forest and Gradient Boosting come into play which combine multiple decision trees to improve the performance.

3) **XGBoost:** XGBoost (Extreme Gradient Boosting) [54] is a powerful machine learning algorithm that comes from the family of gradient boosting methods. Gradient boosting builds multiple weak learners (like

TABLE 9. Top-10 most frequent words in hope speech and non-hope speech posts along with satirical posts. The TF-IDF scores are given for each word.

Hope Speech Posts			No Hope Speech Posts			Satire Posts		
Words	Translation	TF-IDF	Words	Translation	TF-IDF	Words	Translation	TF-IDF
भोट	Vote	0.1749	भोट	Vote	0.1035	भोट	Vote	0.0948
देश	Country	0.1054	देश	Country	0.0633	नेता	Leader	0.0695
आफ्नो	Own	0.0751	नेता	Leader	0.0507	देश	Country	0.0692
परिवर्तन	Changes	0.0699	निर्वाचन	Election	0.0476	चुनाव	Election	0.0522
नेता	Leader	0.0582	चुनाव	Election	0.0348	सबै	All	0.0468
मंसिर	Mangsir	0.0572	काम	Work	0.0345	मंसिर	Mangsir	0.0350
सबै	All	0.0523	आफ्नो	Own	0.0337	पछि	Next	0.0335
मात्र	Only	0.0510	सबै	All	0.0334	निर्वाचन	Election	0.0327
पटक	Times	0.0410	पटक	Times	0.0327	मात्र	Only	0.0323
मत	Vote	0.0359	जनता	People	0.0323	जस्तो	Like that	0.0314

TABLE 10. Top-10 most frequent words in non-satirical posts along with posts related to direction of hate speech. The TF-IDF scores are given for each word.

No Satire Posts			Direction of Hate: Directed			Direction of Hate: Undirected		
Words	Translation	TF-IDF	Words	Translation	TF-IDF	Words	Translation	TF-IDF
भोट	Vote	0.1187	भोट	Vote	0.1092	देश	Country	0.1121
देश	Country	0.0694	देश	Country	0.0795	भोट	Vote	0.1103
निर्वाचन	Election	0.0478	निर्वाचन	Election	0.0658	नेता	Leader	0.0829
नेता	Leader	0.0466	नेता	Leader	0.0626	सबै	All	0.0575
आफ्नो	Own	0.0437	सबै	All	0.0626	आफ्नो	Own	0.0495
जनता	People	0.0361	जनता	People	0.0582	फेरि	Again	0.0483
मात्र	Only	0.0355	काम	Work	0.0468	चुनाव	Election	0.0435
पटक	Times	0.0351	दिन	Day	0.0461	जनता	People	0.0423
काम	Work	0.0339	एमाले	UML	0.0420	पटक	Times	0.0339
सबै	All	0.0324	प्रधानमन्त्री	Prime Minister	0.0326	जनताको	Of the people	0.0326

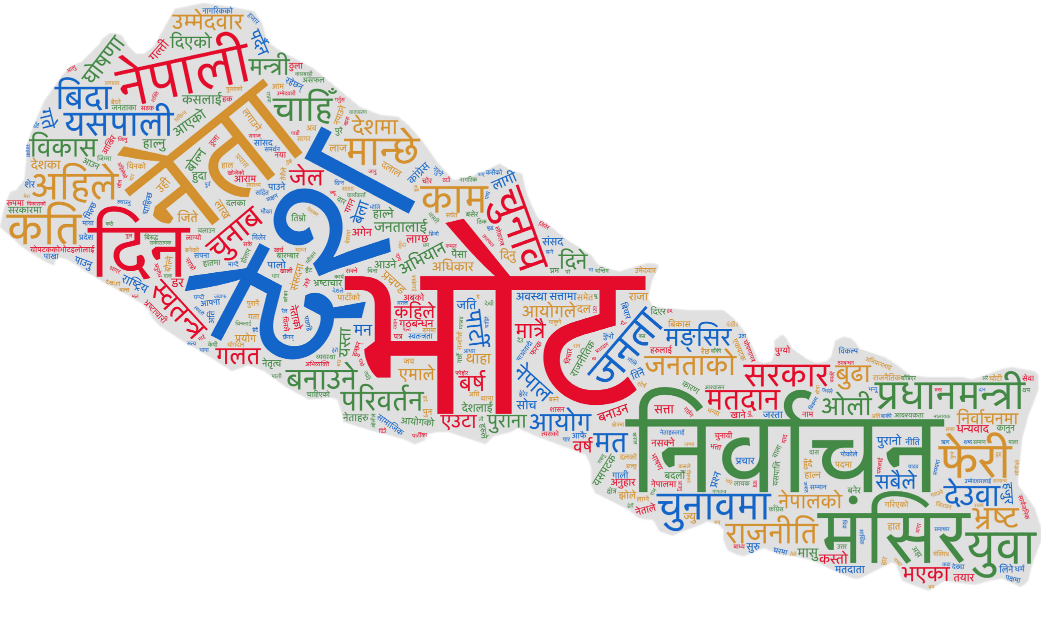
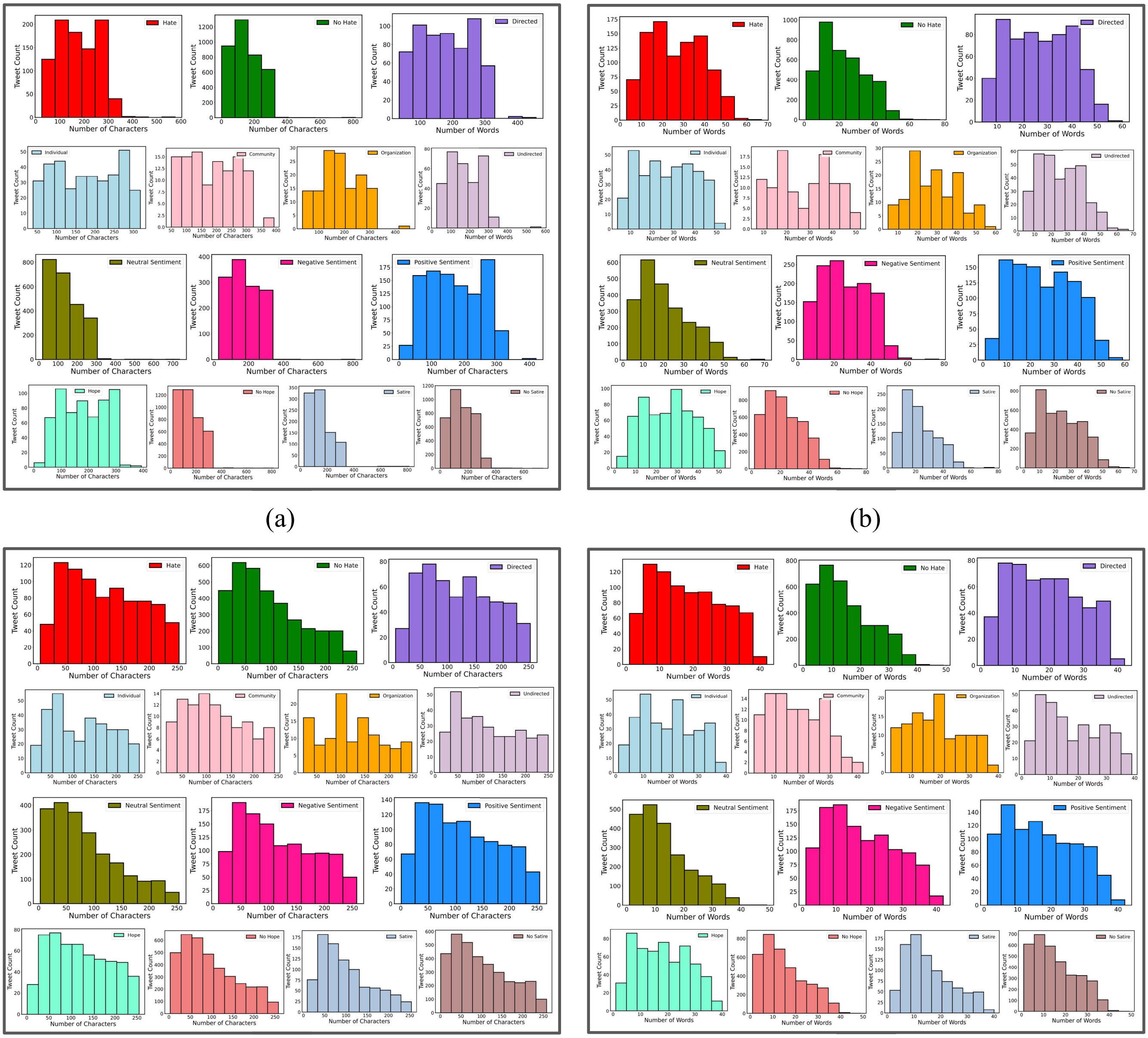


FIGURE 6. Visualization of frequent words in the complete dataset through a wordcloud representation.

decision trees) sequentially, where each subsequent learner corrects the errors of its predecessors. This iterative process leads to a more accurate and robust final model. Due to its highly useful features like parallel processing and handling missing values, XGBoost has become a very popular method that is used for a variety



(a)

(b)

(c)

(d)

FIGURE 7. Histogram obtained before and after pre-processing for different tasks. (a) Histogram of number of characters per tweet before pre-processing, (b) Histogram of number of words per tweet before pre-processing, (c) Histogram of number of characters per tweet after pre-processing, (d) Histogram of number of words per tweet after pre-processing.

of learning tasks such as classification, regression, and ranking problems.

4) **Random Forest (RF):** The bagging method is extended by the RF algorithm, which uses feature randomness in addition to bagging to produce a non-correlated forest of decision trees. A low level of correlation across decision trees is ensured by feature randomization, also known as feature bagging, which creates a random collection of features. RF merely choose a portion of those value splits, whereas decision trees take into account all possible value splits. There are three key hyperparameters for algorithms based on random forests that must be set prior to training. Node size, tree count, and sampled feature count are a few of them [55]. This algorithm can then be utilised to address classification or regression tasks.

5) **Support Vector Machine (SVM):** The objective of SVM is to choose the optimum hyperplane that splits diverse input data classes while enhancing their separation. Using a kernel function to transform the data elements into a higher-dimensional space where they can be linearly separated. Support vectors, crucial for developing the hyperplane, are then determined to be the data points closest to the decision boundary [56].

The SVM seeks to maintain the greatest distance between the support vectors while limiting classification error.

6) **Logistic Regression (LR):** Predictive analytics and categorization frequently employ this kind of statistical model. Based on a given dataset of independent variables, logistic regression calculates the likelihood that an event will occur, such as voting or not voting. The measured variable's range is 0 to 1, so the result is a probability. A logit transform is performed to the odds in LR, which is the possibility that it will succeed divided by the possibility of failure. Most frequently, maximum likelihood estimation (MLE) is used to estimate the beta parameter or coefficient, in this model [57]. In order to find the most appropriate for the log odds, this approach iteratively evaluates various beta values.
7) **AdaBoost:** AdaBoost or 'Adaptive Boosting' [58] is a popular machine learning algorithm that is used for classification and regression tasks. Like XGBoost, AdaBoost is also an ensemble technique but unlike XGBoost it doesn't use gradient boosting. The main idea behind AdaBoost is that it combines the predictions of weak learners to create a strong powerful ensemble model. Although AdaBoost is a very effective method, it can be sensitive to noisy data and might cause overfitting if the weak learners are too complex. Proper data pre-processing and careful selection of weak learners are ways to mitigate this issue.
8) **BERT:** BERT (Bidirectional Encoder Representations from Transformers) [59] is a framework for natural language processing (NLP) that is built upon the transformer architecture. Unlike previous language models that read text sequentially in one direction(left-to-right or right-to-left), BERT is designed to take the entire input sequence from both directions simultaneously. This allows the model to have a contextual understanding of each word based on the surrounding words on both sides, resulting in a more robust representation of the language. BERT is pre-trained on a massive amount of text data using a self-supervised approach and this pre-training phase enables BERT to capture general language patterns and relationships. Pre-trained BERT model can be fine-tuned on task-specific data to adapt representations of specific tasks. BERT has been ground-breaking in NLP and has set new state-of-the-art performances in various NLP tasks such as classification, named entity recognition, sentiment analysis, question answering, and more. Since BERT is pre-trained on language data, there are various language-specific and multilingual BERT models created by researchers. In this work, for benchmark purposes, we use 4 different types of BERT models in the Nepali language pre-trained on different data.

DistillBERT (Nepali): DistilBERT [60] is a variant of BERT that aims to reduce the model's size and computational complexity while retaining most of its performance. It does this by using knowledge distillation to train a smaller model to approximate the behaviour of a larger, pre-trained BERT model. The Nepali DistillBERT model [61] that we used for benchmarking is available in the Hugging Face library and was trained on the OSCAR Nepali corpus [62].

RoBERTa (Nepali): RoBERTa (Robustly Optimized BERT Pretraining Approach) [63] is a variant of the BERT model and was designed to address some of the limitations of the original BERT by improving performance and efficiency. Like BERT, RoBERTa also uses a self-attention mechanism to process input sequences but the key difference between BERT and RoBERTa is that RoBERTa was trained on a much larger dataset in a more effective training method. RoBERTa was trained on 160GB of text data which is 10 times more compared to BERT which used 16GB of text data. Unlike BERT which used random masking of words during training, RoBERTa introduces a dynamic masking technique that ensures a more robust and generalized representation of words in the model. For our benchmark purpose, we used the roberta-base-ne model [64] that was trained on the Nepali CC-100 dataset [65], [66] which contains 12 million sentences in the Nepali language.

NepBERTa: NepBERTa [35] is a BERT-based Natural Language Understanding (NLU) model trained on an extensive Nepali corpus with about 0.8 billion Nepali words. Using BERT as the base model, hyper-parameter tuning was done to obtain the optimal parameters for the model to perform well in the Nepali language.

NepaliBERT: NepaliBERT [67] is BERT based model for the Nepali language trained from 6.7 million lines of raw Nepali texts. The dataset for training was formed by combining a large-scale Nepali corpus [37] and the OSCAR Nepali corpus [62].

NepNewsBERT: NepNewsBERT [68] is a Masked Language Model(MML) for the Nepali language trained on new data scrapped from popular Nepali news websites. The dataset contains about 10 million sentences in the Nepali language.

All four BERT-based models except the DistillBERT (Nepali) were only available as fill masks. Since we needed the models for classification tasks, we modified each model for downstream classification and used them accordingly.

B. PERFORMANCE METRICS

The performance was evaluated based on three performance metrics which were Macro Precision Score ($\text{Pre}_{\text{macro}} \uparrow$), Macro Mean Absolute Error (MMAE↓), and F1-Macro scores ($\text{F1}_{\text{macro}} \uparrow$). In the context of precision and F1 scores, optimal performance is indicated by values approaching 1. Conversely, when considering the MMAE metric, optimal performance is indicated by values that approach 0.

1) PRECISION SCORE (PS)

In the realm of classification tasks, the precision score serves as a metric designed to assess the correctness of positive predictions generated by a model. It quantifies the ratio of true positive predictions to the total instances that the model categorizes as positive. A precision score [69] signifies the model's adeptness at minimizing the occurrence of false positive predictions. In essence, it signifies the model's ability to accurately recognize positive instances during its predictive process. Mathematically,

$$PS = \frac{True\ Positive}{(False\ Positive + True\ Positive)} \tag{10}$$

2) MEAN ABSOLUTE ERROR

MAE measures the average absolute difference between the predicted and actual values of the target variable. It provides an indication of how close the predictions are to the actual values. MAE score [70] is computed by taking the average of the absolute values of the errors. The absolute value, a mathematical operation, ensures that a numerical value is treated as positive regardless of its sign. The calculation of the MAE value can be expressed as:

$$Mean\ Absolute\ Error = \frac{1}{n}\sum_{i=1}^{n} |x_i - x| \tag{11}$$

where,
$|x_i - x| = absolute\ error$
$n = number\ of\ error$

3) F1-SCORE

The F1 score [69] integrates precision and recall, offering a harmonized assessment of a model's effectiveness. By incorporating both the true positive rate (recall) and the positive predictive value (precision), the F1 score presents a unified metric that encapsulates the model's overall accuracy.

From a mathematical standpoint, the F1 score is computed as the harmonic mean derived from the precision and recall values:

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{12}$$

In our implementation, we use Macro Precision ($\text{Pre}_{\text{macro}}$), Macro Mean Absolute Error (MMAE), and F1-Macro scores (F1_{macro}). Each of these is an unweighted average of individual metrics for each label.

C. ANALYSIS

Table 11 and Table 12 provide an overview of the performance of different algorithms across diverse tasks. Altogether, we performed the baseline on a total of 6 tasks viz. Hate Speech Detection, Direction of Hate Speech Detection, Targets of Hate Speech, Sentiment Analysis, Satire Detection, and Hope Speech Detection. A comprehensive assessment was conducted employing a total of 12 distinct models. Within this set of models, 7 were machine learning (ML) algorithms, while the remaining 5 models were deep learning (DL) models, specifically employing the Transformer architecture. For the ML-based model, 85% of the data was used as the training set and the remaining 15% was used as the test set. For DL models, 70% of the data was used as training data and validation and the test set had an equal split of 15%. Table 13 shows the data splitting of train, validation, and test. For 'Hate Speech Detection', NepNewsBERT model gave the highest precision score and F1-score of 0.633 and 0.640 respectively, whereas the SVM model gave the best MMAE score of 0.175. For the detection of 'Direction of Hate Speech', ML based algorithm performed well where XGBoost gave the precision score of 0.665 and F1-score of 0.669. Among the various models examined, the SVM model showed the most favourable MMAE score of 0.151. In the realm of identifying the 'Targets of Hate Speech', the highest precision score was 0.667 which was archived by DistillBERT (Nepali). MMAE score of 0.105 was obtained by the SVM model, while the XGBoost model achieved the highest precision score of 0.667. In the context of discerning the 'Sentiment Analysis', the DistillBERT (Nepali) model emerged as a standout performer, attaining the highest precision score of 0.465. Furthermore, the SVM model showcased a low MMAE score of 0.641.

Notably, the XGBoost model exhibited the highest F1-score of 0.467. Regarding the task of 'Satire Detection', the DistillBERT (Nepali) model stood out by achieving the highest precision score and F1-score of 0.600 and 0.608, respectively. Moreover, the SVM model demonstrated a notably low MMAE score of 0.272. Finally, for the 'Hope Speech Detection', both the RoBERTa (Nepali) and NepNewsBERT showed the highest precision score of 0.589. Additionally, the SVM model impressively showcased a low MMAE score of 0.268. It was found that NepBERTa gave the highest F1-score of 0.600. Also, Topic Modeling was used to identify prevalent themes and patterns within the discourse, providing a holistic perspective on anti-establishment election discourse in the context of a low-resource language like Nepali.

In Figure 8, it was interesting to see that the tweet without hate were misclassified as hateful tweet. Such misclassifications are an important part of the study to make models robust. Figure 8 also shows that the misclassified tweet contains some words with negative connotations like "devils" which might have influenced the ability of the model to accurately distinguish between the classes. Further analysis on the explainability of the models is warranted in order to investigate the causes of misclassification. Identification of the most significant words and cues that models leverage in classification would help us to probe the model and make it more accurate.

A noteworthy discovery that points to the need for additional study to create models that can reliably differentiate between different kinds of speech. Overall, the findings point to the need for more research and advancement in the field of multi-aspect speech recognition in the Nepali language, with the usage of transformer-based models being one potentially effective strategy.

TABLE 11. Baseline results for hate speech detection along with direction and targets of hate speech with different algorithms.

Model	Hate vs Non Hate			Direction			Targets		
	Pre$_{\text{macro}}$↑	MMAE↓	F1$_{\text{macro}}$↑	Pre$_{\text{macro}}$↑	MMAE↓	F1$_{\text{macro}}$↑	Pre$_{\text{macro}}$↑	MMAE↓	F1$_{\text{macro}}$↑
NB	0.535	0.359	0.526	0.552	0.424	0.547	0.473	0.507	0.488
XGBoost	0.558	0.414	0.563	**0.665**	0.325	**0.669**	0.593	0.427	**0.594**
AdaBoost	0.557	0.427	0.562	0.605	0.384	0.608	0.478	0.586	0.488
DT	0.571	0.415	0.575	0.593	0.411	0.589	0.553	0.614	0.544
RF	0.523	0.300	0.499	0.606	0.329	0.607	0.530	0.132	0.562
SVM	0.519	**0.175**	0.487	0.571	**0.151**	0.535	0.446	**0.105**	0.455
LR	0.546	0.188	0.537	0.606	0.306	0.605	0.482	0.341	0.515
DistillBERT (Nepali)	0.621	0.662	0.417	0.598	0.390	0.600	**0.667**	0.616	0.352
RoBERTa (Nepali)	0.582	0.370	0.556	0.620	0.370	0.619	0.621	0.600	0.514
NepaliBERT	0.593	0.355	0.576	0.555	0.445	0.555	0.400	0.659	0.360
NepNewsBERT	**0.633**	0.312	**0.640**	0.580	0.461	0.512	0.311	0.735	0.336
NepBERTa	0.595	0.362	0.595	0.567	0.440	0.558	0.370	0.626	0.371

TABLE 12. Baseline results for sentiment analysis, satire detection and hope speech detection with different algorithms.

Model	Sentiment			Satire			Hope Speech		
	Pre$_{\text{macro}}$↑	MMAE↓	F1$_{\text{macro}}$↑	Pre$_{\text{macro}}$↑	MMAE↓	F1$_{\text{macro}}$↑	Pre$_{\text{macro}}$↑	MMAE↓	F1$_{\text{macro}}$↑
NB	0.412	0.681	0.399	0.542	0.331	0.531	0.515	0.444	0.508
XGBoost	0.430	0.761	0.432	0.523	0.427	0.525	0.527	0.437	0.530
AdaBoost	0.418	0.777	0.419	0.556	0.421	0.560	0.555	0.408	0.565
DT	0.366	0.885	0.366	0.558	0.431	0.561	0.556	0.432	0.560
RF	0.327	0.700	0.355	0.515	0.422	0.487	0.524	**0.268**	0.519
SVM	0.412	**0.641**	0.398	0.506	**0.272**	0.455	0.412	0.641	0.398
LR	0.445	0.705	0.450	0.505	0.439	0.459	0.501	0.485	0.479
DistillBERT (Nepali)	**0.465**	0.717	**0.467**	**0.600**	0.375	**0.608**	0.465	0.717	0.467
RoBERTa (Nepali)	0.451	0.767	0.418	0.542	0.446	0.541	**0.589**	0.367	0.599
NepaliBERT	0.381	0.833	0.380	0.531	0.454	0.521	0.565	0.453	0.551
NepNewsBERT	0.434	0.756	0.420	0.564	0.422	0.567	**0.589**	0.360	0.599
NepBERTa	0.424	0.751	0.421	0.558	0.435	0.560	0.597	0.318	**0.600**

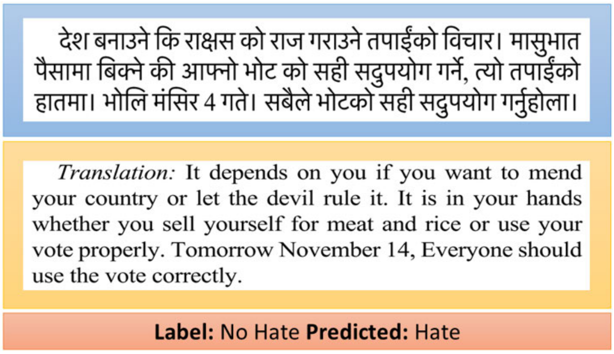
D. LIMITATIONS AND CONSIDERATION

In this paper, we provide a large dataset for an in-depth analysis of Nepali tweets pertaining to an anti-establishment electoral discourse. We propose baselines for assessing the discourse in terms of hate, hope, satire, and sentiment. Additionally, the targets of any directed hate speech were identified. But there are certain limitations to our effort. Our dataset only includes tweets from a certain period of time connected to the election in Nepal, it might not always be indicative of the types (hate, hope, and satire) of speech and might not represent the same sentiment in other contexts. Additionally, only tweets from a single microblogging network are included in our sample. Also, target annotation method is focused on general categories (Individuals, Organizations, and Communities), it might not be able to identify targets that are more specialized or complex. Furthermore, because annotation is a subjective process, various annotators can have different views on tweets. The baselines we offer are predicated on a constrained set of features, and it's feasible that different features or architectural configurations might result in better performance. It is also crucial to highlight that technology for multi-aspect analyzing speech and identifying targets may present moral issues, such as possible prejudice. These moral issues ought to be taken into account and resolved in the creation and application of such technology.

TABLE 13. Train/validation/test split for different tasks for transformer-based models.

Tasks	Train	Validation	Test
Hate Speech Detection	2,976	638	638
Targets for Hate Speech	376	81	81
Direction of Hate Speech	583	125	125
Satire Detection	2,976	638	638
Hope Speech Detection	2,976	638	638
Sentiment Analysis	2,976	638	638

The work has a few more limitations, with a particular focus on data imbalance. This is noticeable when addressing



देश बनाउने कि राक्षस को राज गराउने तपाईंको विचार। मासुभात पैसामा बिक्ने की आफ्नो भोट को सही सदुपयोग गर्ने, त्यो तपाईंको हातमा। भोलि मंसिर 4 गते। सबैले भोटको सही सदुपयोग गर्नुहोला।

Translation: It depends on you if you want to mend your country or let the devil rule it. It is in your hands whether you sell yourself for meat and rice or use your vote properly. Tomorrow November 14, Everyone should use the vote correctly.

Label: No Hate **Predicted:** Hate

FIGURE 8. Example of misclassification by models.

the identification of hate speech, as there is comparatively less hate speech in real-world discourse. The volume of non-hate tweets compared to hate speech tweets can undermine the accuracy of algorithms. The next study directions should address these gaps through various approaches like oversampling minority groups or applying advanced techniques like cost-sensitive learning. Additionally, the study's concentration on anti-establishment political discourse in the Nepali language restricts the generalization of findings. Further studies should study the cross-lingual and cross-cultural scalability of hate speech detection algorithms to boost adaptability and consider more low-resource languages for a broader overview of online hate speech across varied linguistic contexts.

V. CONCLUSION

In this paper, we offer the NAET dataset, a crucial resource for developing and evaluating models to analyze speech in the context of Nepali election discourse. The Nepali language is significantly understudied in academic research on artificial intelligence. Thus, this dataset can serve as a valuable resource in NLP studies. Since the dataset follows a multi-aspect annotation scheme, this could result in the subjective nature of the dataset. However, the thorough annotation guidelines and high inter-annotator agreements ensure the quality of the dataset. Our work explores multiple aspects of speech analysis namely, sentimental analysis, hate speech detection with targets, satire detection, and hope speech detection. The benchmarks set by our work also help future researcher to continue developing newer architectures to beat the baseline results. As part of future research, it is our goal to develop cutting-edge NLP models for the Nepali language that specialize in each of these tasks. It is our belief that the NAET dataset can serve as a helpful starting point for further Nepali language annotation efforts, enabling the incorporation of annotation features beyond those presented in this work. Finally, we hope that our dataset will contribute to the development of effective speech analysis techniques in Nepali, promoting more inclusive and courteous online conversations. We promote low-resource language that addresses multiple aspects of discourse in the Nepali language.

REFERENCES

[1] E. García-Sánchez, P. R. Benetti, G. L. Higa, M. C. Alvarez, and E. Gomez-Nieto, "Political discourses, ideologies, and online coalitions in the Brazilian congress on Twitter during 2019," *New Media Soc.*, vol. 25, no. 5, pp. 1130–1152, May 2023.

[2] J. T. Jost, P. Barberá, R. Bonneau, M. Langer, M. Metzger, J. Nagler, J. Sterling, and J. A. Tucker, "How social media facilitates political protest: Information, motivation, and social networks," *Political Psychol.*, vol. 39, no. S1, pp. 85–118, Feb. 2018.

[3] M. A. Hossain. (2023). *The Political Instability in Nepal and Its Geopolitical Implications*. Accessed: 2023. [Online]. Available: https://southasiajournal.net/the-political-instability-in-nepal-and-its-geopolitical-implications/

[4] M. of Foreign Affairs. (2022). *History of Nepal*. Accessed: Apr. 5, 2023. [Online]. Available: https://mofa.gov.np/about-nepal/history-of-nepal/

[5] IFES. (2023). *Federal Democratic Republic of Nepal*. Accessed: Apr. 10, 2023. [Online]. Available: https://www.electionguide.org/elections/id/1499/

[6] Setopati. (2022). *Local Election on May 13*. Accessed: 2023. [Online]. Available: https://en.setopati.com/political/157891

[7] S. Karki. (2022). *Many Nepalis Say no, Not Again*. Accessed: Mar. 8, 2023. [Online]. Available: https://www.nepalitimes.com/news/many-nepalis-say-no-not-again

[8] A. Gautam, P. Mathur, R. Gosangi, D. Mahata, R. Sawhney, and R. R. Shah, "#MeTooMA: Multi-aspect annotations of tweets related to the MeToo movement," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 14, 2020, pp. 209–216.

[9] A. O. A. Gurung. (2022). *Outrage as Poll Authority Cracks Down on Negative Campaigning*. Accessed: Apr. 20, 2023. [Online]. Available: https://kathmandupost.com/politics/2022/10/29/outrage-as-poll-authority-cracks-down-on-negative-campaigning

[10] T. K. Post. (2022). *Supreme Court Issues Interim Order in Favour of 'No Not Again' Campaign*. Accessed: Jul. 14, 2023. [Online]. Available: https://kathmandupost.com/national/2022/11/06/supreme-court-issues-interlocutory-interim-order-not-to-take-action-against-no-not-againcampaigners

[11] R. Khanal, "Linguistic geography of Nepalese languages," *3rd Pole, J. Geography Educ.*, vol. 10, pp. 45–54, Dec. 2019.

[12] N. B. Niraula, S. Dulal, and D. Koirala, "Offensive language detection in nepali social media," in *Proc. 5th Workshop Online Abuse Harms (WOAH)*, 2021, pp. 67–75.

[13] S. Zakharov, O. Hadar, T. Hakak, D. Grossman, Y. B.-D. Kolikant, and O. Tsur, "Discourse parsing for contentious, non-convergent online discussions," in *Proc. Int. Conf. Web Social Media*, vol. 15, 2021, pp. 853–864.

[14] J. J. E. Macrohon, C. N. Villavicencio, X. A. Inbaraj, and J.-H. Jeng, "A semi-supervised approach to sentiment analysis of tweets during the 2022 Philippine presidential election," *Information*, vol. 13, no. 10, p. 484, Oct. 2022.

[15] S. Thapa, S. Adhikari, U. Naseem, P. Singh, G. Bharathy, and M. Prasad, "Detecting Alzheimer's disease by exploiting linguistic information from Nepali transcript," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2020, pp. 176–184.

[16] S. Adhikari, S. Thapa, U. Naseem, P. Singh, H. Huo, G. Bharathy, and M. Prasad, "Exploiting linguistic information from Nepali transcripts for early detection of Alzheimer's disease using natural language processing and machine learning techniques," *Int. J. Hum.-Comput. Student*, vol. 160, Apr. 2022, Art. no. 102761.

[17] C. Sitaula, A. Basnet, A. Mainali, and T. B. Shahi, "Deep learning-based methods for sentiment analysis on nepali COVID-19-related tweets," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–11, Nov. 2021.

[18] O. M. Singh, S. Timilsina, B. K. Bal, and A. Joshi, "Aspect based abusive sentiment detection in Nepali social media texts," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Dec. 2020, pp. 301–308.

[19] I. Mollas, Z. Chrysopoulou, S. Karlos, and G. Tsoumakas, "ETHOS: An online hate speech detection dataset," 2020, *arXiv:2006.08328*.

[20] M. E. Wojcieszak and D. C. Mutz, "Online groups and political discourse: Do online discussion spaces facilitate exposure to political disagreement?" *J. Commun.*, vol. 59, no. 1, pp. 40–56, 2009, doi: 10.1111/j.1460-2466.2008.01403.x.

[21] K. Solovev and N. Pröllochs, "Hate speech in the political discourse on social media: Disparities across parties, gender, and ethnicity," in *Proc. ACM Web Conf.*, Apr. 2022, pp. 3656–3661.

[22] H. Takikawa and K. Nagayoshi, "Political polarization in social media: Analysis of the 'Twitter political field' in Japan," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 3143–3150.

[23] K. Johnson, I.-T. Lee, and D. Goldwasser, "Ideological phrase indicators for classification of political discourse framing on Twitter," in *Proc. 2nd Workshop NLP Comput. Social Sci.*, 2017, pp. 90–99.

[24] K. Johnson and D. Goldwasser, "Identifying stance by analyzing political discourse on Twitter," in *Proc. 1st Workshop NLP Comput. Social Sci.*, 2016, pp. 66–75.

[25] U. Naseem, I. Razzak, S. K. Khan, and M. Prasad, "A comprehensive survey on word representation models: From classical to state-of-the-art word representation language models," *ACM Trans. Asian Low-Resource Lang. Inf. Process.*, vol. 20, no. 5, pp. 1–35, Sep. 2021.

[26] B. K. Bal, "Towards building advanced natural language applications: An overview of the existing primary resources and applications in nepali," in *Proc. 7th Workshop Asian Lang. Resour.*, 2009, pp. 165–170.

[27] C. P. Gupta and B. K. Bal, "Detecting sentiment in nepali texts: A bootstrap approach for sentiment analysis of texts in the nepali language," in *Proc. Int. Conf. Cognit. Comput. Inf. Processing(CCIP)*, Mar. 2015, pp. 1–4.

[28] G. Prabha, P. V. Jyothsna, K. K. Shahina, B. Premjith, and K. P. Soman, "A deep learning approach for part-of-speech tagging in Nepali language," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2018, pp. 1132–1136.

[29] B. B. Shrestha and B. K. Bal, "Named-entity based sentiment analysis of Nepali news media texts," in *Proc. 6th Workshop Natural Lang. Process. Techn. Educ. Appl.*, 2020, pp. 114–120.

[30] P. Koirala and N. B. Niraula, "NPVec1: Word embeddings for Nepali–construction and evaluation," in *Proc. 6th Workshop Represent. Learn. NLP*, 2021, pp. 174–184.

[31] U. Naseem, I. Razzak, K. Musial, and M. Imran, "Transformer based deep intelligent contextual embedding for Twitter sentiment analysis," *Future Gener. Comput. Syst.*, vol. 113, pp. 58–69, Dec. 2020.

[32] J. D. Kenton, C. Ming-Wei, and L. K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.

[33] A. Conneau and G. Lample, "Cross-lingual language model pretraining," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 7059–7069.

[34] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 8440–8451.

[35] S. Timilsina, M. Gautam, and B. Bhattarai, "NepBERTa: Nepali language model trained in a large corpus," in *Proc. 2nd Conf. Asia–Pacific Chapter Assoc. Comput. Linguistics 12th Int. Joint Conf. Natural Lang. Process.*, 2022, pp. 273–284.

[36] S. R. Laskar, P. Pakray, and S. Bandyopadhyay, "Neural machine translation: Hindi–Nepali," in *Proc. 4th Conf. Mach. Transl.*, 2019, pp. 202–207.

[37] R. Lamsal, "A large scale Nepali text corpus," IEEE Dataport, Tech. Rep., 2020. [Online]. Available: https://ieee-dataport.org/open-access/large-scale-nepali-text-corpus

[38] A. Senapati, A. Poudyal, P. Adhikary, S. Kaushar, A. Mahajan, and B. N. Saha, "A machine learning approach to anaphora resolution in Nepali language," in *Proc. Int. Conf. Comput. Perform. Eval. (ComPE)*, Jul. 2020, pp. 436–441.

[39] O. M. Singh, A. Padia, and A. Joshi, "Named entity recognition for nepali language," in *Proc. IEEE 5th Int. Conf. Collaboration Internet Comput. (CIC)*, Dec. 2019, pp. 184–190.

[40] Y. P. Yadava, A. Hardie, R. R. Lohani, B. N. Regmi, S. Gurung, A. Gurung, T. McEnery, J. Allwood, and P. Hall, "Construction and annotation of a corpus of contemporary nepali," *Corpora*, vol. 3, no. 2, pp. 213–225, Nov. 2008.

[41] C. Sitaula, A. Basnet, and S. Aryal, "Vector representation based on a supervised codebook for Nepali documents classification," *PeerJ Comput. Sci.*, vol. 7, p. e412, Mar. 2021.

[42] T. B. Shahi and A. K. Pant, "Nepali news classification using Naïve bayes, support vector machines and neural networks," in *Proc. Int. Conf. Commun. Inf. Comput. Technol. (ICCICT)*, Feb. 2018, pp. 1–5.

[43] J. T. Becker, F. Boiler, O. L. Lopez, J. Saxton, and K. L. McGonigle, "The natural history of Alzheimer's disease: Description of study cohort and accuracy of diagnosis," *Arch. Neurol.*, vol. 51, no. 6, pp. 585–594, 1994.

[44] S. Medhaffar, F. Bougares, Y. Estève, and L. Hadrich-Belguith, "Sentiment analysis of Tunisian dialects: Linguistic ressources and experiments," in *Proc. 3rd Arabic Natural Lang. Process. Workshop*, 2017, pp. 55–61.

[45] A. Al-Hassan and H. Al-Dossari, "Detection of hate speech in Arabic tweets using deep learning," *Multimedia Syst.*, vol. 28, no. 6, pp. 1963–1974, Dec. 2022.

[46] R. Falotico and P. Quatto, "Fleiss' Kappa statistic without paradoxes," *Quality Quantity*, vol. 49, no. 2, pp. 463–470, Mar. 2015.

[47] E. S. Negara, D. Triadi, and R. Andryani, "Topic modelling Twitter data with latent Dirichlet allocation method," in *Proc. Int. Conf. Electr. Eng. Comput. Sci. (ICECOS)*, Oct. 2019, pp. 386–390.

[48] D. Mimno, H. Wallach, E. Talley, M. Leenders, and A. McCallum, "Optimizing semantic coherence in topic models," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2011, pp. 262–272.

[49] Z. Tong and H. Zhang, "A text mining research based on LDA topic modelling," in *Proc. Comput. Sci. Inf. Technol.*, May 2016, pp. 201–210.

[50] M. Hasan, A. Rahman, M. R. Karim, M. S. I. Khan, and M. J. Islam, "Normalized approach to find optimal number of topics in latent Dirichlet allocation (LDA)," in *Proc. Int. Conf. Trends Comput. Cognit. Eng.* Cham, Switzerland: Springer, 2021, pp. 341–354.

[51] S. Qaiser and R. Ali, "Text mining: Use of TF-IDF to examine the relevance of words to documents," *Int. J. Comput. Appl.*, vol. 181, no. 1, pp. 25–29, Jul. 2018.

[52] K. P. Murphy, "Naive Bayes classifiers," *Univ. Brit. Columbia*, vol. 18, no. 60, pp. 1–8, Oct. 2006.

[53] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, "An introduction to decision tree modeling," *J. Chemometrics*, vol. 18, no. 6, pp. 275–285, Jun. 2004.

[54] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 785–794.

[55] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[56] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *J. Mach. Learn. Res.*, vol. 2, pp. 45–66, Nov. 2001.

[57] S. T. Indra, L. Wikarsa, and R. Turang, "Using logistic regression method to classify tweets into the selected topics," in *Proc. Int. Conf. Adv. Comput. Sci. Inf. Syst. (ICACSIS)*, Oct. 2016, pp. 385–390.

[58] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," in *Proc. Eur. Conf. Comput. Learn. Theory*. Cham, Switzerland: Springer, 1995, pp. 23–37.

[59] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.

[60] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," 2019, *arXiv:1910.01108*.

[61] D. Shrestha. (2021). *DistillBERT(Nepali)*. Accessed: Jul. 15, 2023. [Online]. Available: https://huggingface.co/dexhrestha/Nepali-DistilBERT

[62] P. J. O. Suárez, B. Sagot, and L. Romary, "Asynchronous pipeline for processing huge corpora on medium to low resource infrastructures," in *Proc. 7th Workshop Challenges Manag. Large Corpora*, 2019, pp. 1–9.

[63] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," 2019, *arXiv:1907.11692*.

[64] A. Chaudhary. (2021). *RoBERTa (Nepali)*. Accessed: Jun. 10, 2023. [Online]. Available: https://huggingface.co/amitness/roberta-base-ne

[65] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," 2019, *arXiv:1911.02116*.

[66] G. Wenzek, M.-A. Lachaux, A. Conneau, V. Chaudhary, F. Guzmán, A. Joulin, and E. Grave, "CCNet: Extracting high quality monolingual datasets from web crawl data," 2019, *arXiv:1911.00359*.

[67] R. Ghimire. (2022). *NepaliBERT*. Accessed: May 25, 2023. [Online]. Available: https://huggingface.co/Rajan/NepaliBERT

[68] S. Pudasaini. (2021). *NepNewsBERT*. Accessed: May 23, 2023. [Online]. Available: https://huggingface.co/Shushant/NepNewsBERT

[69] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *Proc. Eur. Conf. Inf. Retr.* Cham, Switzerland: Springer, 2005, pp. 345–359.

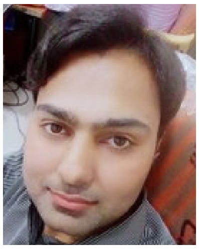
[70] W. Wang and Y. Lu, "Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model," *IOP Conf. Ser., Mater. Sci. Eng.*, vol. 324, Jan. 2018, Art. no. 012049.

KRITESH RAUNIYAR received the B.Tech. degree in computer engineering from Delhi Technological University, India. During the B.Tech. degree, he was a research assistant. He received the Mitacs GRI as a Visiting Scholar with the University of Regina, Canada, in 2022. During the Mitacs summer internship, he has conducted research in federated learning. He has published papers at various conferences and has been involved in several projects related to computer vision, LLMs, and low-resource NLP.

SWETA POUDEL received the B.Tech. degree in computer engineering from the Kathmandu Engineering College, Nepal. Following the completion of the B.Tech. degree, she started a professional career as a software engineer, a role in which she continues to serve to this day. She has contributed to various projects and has published papers at a few conferences.

SHUVAM SHIWAKOTI received the B.Tech. degree in software engineering from Delhi Technological University, India. During the B.Tech. study, he did a few internships as a software engineer and a research assistant. He was a part of various projects, some of which were published at conferences. His research interests include natural language processing (NLP), large language models, research in low-resource languages, medical NLP, computer vision, and medical imaging.

SURENDRABIKRAM THAPA (Member, IEEE) received the B.Tech. degree in software engineering from Delhi Technological University, India, and the M.S. degree in computer science from the Department of Computer Science, Virginia Tech, Blacksburg, USA. He is currently a Research Faculty with Virginia Tech, where he works primarily on deep learning. He was a Visiting Scholar with the University of Technology Sydney, Australia, in 2020. During the M.S. study, he was funded by the National Science Foundation (NSF) Grant and various government agencies. He has published research papers at various reputed conferences and journals. His research interests include natural language processing (NLP) computational social sciences and computer vision applications. He has served as a program committee member for numerous conferences and workshops. He has been serving as a reviewer for several reputed journals and conferences.



JUNAID RASHID received the B.S. and M.S. degrees in computer science from the COMSATS Institute of Information Technology, Wah Campus, Pakistan, in 2014 and 2016, respectively, and the Ph.D. degree in computer science from the University of Engineering and Technology, Taxila, Pakistan, in 2020. Since November 2020, he has been an Honorary Senior Postdoctoral Research Fellow with the Center for AI and Data Science, Edinburgh Napier University, U.K., and received the Fellowship, in 2022. He worked as a Research Professor and Postdoctoral Researcher at Kongju National University, Korea. He is currently working as an Assistant Professor at Sejong University, Korea. He has published research papers in prestigious journals and conferences. His research interests include data science, machine learning, natural language processing, topic modeling, text mining, information retrieval, big data, pattern recognition, fuzzy systems, medical informatics, biomedical text analytics, and software engineering. He received the fully-funded scholarship for the Ph.D. degree. He has served/been serving as a reviewer for various reputed journals and conferences. He has been an Academic Editor of *PLOS One*.

JUNGEUN KIM received the Ph.D. degree in knowledge service engineering from the Korea Advanced Institute of Science and Technology (KAIST). He is currently an Assistant Professor with Kongju National University (KNU). Before joining KNU, he was a Senior Researcher with the Artificial Intelligence Research Laboratory, Electronics and Telecommunications Research Institute (ETRI). His research interests include data mining, artificial intelligence, big data analysis with distributed processing platforms, and open data platforms.

MUHAMMAD IMRAN is a Senior Lecturer with the Institute of Innovation, Science and Sustainability, Federation University, Australia. He is the Founding Leader of the Wireless Networks and Security (WINS) Research Group. He has completed many international collaborative research projects with reputable universities. He has published more than 300 research articles in peer-reviewed and highly reputable international conferences and journals. He is included in the 2022 Clarivate Highly Cited Researchers list in the category of computer science. His research interests include mobile and wireless networks, the Internet of Things, big data analytics, cloud/edge computing, and information security. Consecutively, he was awarded as an Outstanding Associate Editor of *Future Generation Computer Systems* (FGCS) and IEEE Access, in 2018, 2019, and 2021, respectively. He served as the Editor-in-Chief for the *European Alliance for Innovation (EAI) Transactions on Pervasive Health and Technology* and an Associate Editor for *IEEE Communications Magazine*. He is serving as an Associate Editor for top-ranked international journals, such as IEEE Network, *FGCS*, and IEEE Access.

USMAN NASEEM is a Lecturer with the College of the School of Science and Engineering, James Cook University, Australia. Prior to persuading his research, he worked in leading ICT companies for over ten years. His research interests include natural language processing (NLP) and computational social science, focusing on understanding human communication in social contexts and developing socially aware language technologies. He publishes and serves with the Program Committee, including the Area Chair for several top-tier venues, including ACL, EMNLP, NAACL, COLING, WSDM, Webconf, and AAAi. He also delivered several invited talks and a tutorial on recommender systems at Webconf 2023. His work in NLP has attracted attention from the World Health Organization and earned him the Nepean Blue Mountains Local Health District (NBMLHD) Board Chair's Quality Award, in 2021. He also received the prestigious IEEE Best Transactions Paper Award, in 2022.

• • •